Free-energy density functional for Strauss's model of transitive networks

Diego Escribano [1] and José A. Cuesta [1,2,*]

[1]*Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain*
[2]*Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain*

Ensemble models of graphs are one of the most important theoretical tools to study complex networks. Among them, exponential random graphs (ERGs) have proven to be very useful in the analysis of social networks. In this paper we develop a technique, borrowed from the statistical mechanics of lattice gases, to solve Strauss's model of transitive networks. This model was introduced long ago as an ERG ensemble for networks with high clustering and exhibits a first-order phase transition above a critical value of the triangle interaction parameter where two different kinds of networks with different densities of links (or, alternatively, different clustering) coexist. Compared to previous mean-field approaches, our method describes accurately even small networks and can be extended beyond Strauss's classical model—e.g., to networks with different types of nodes. This allows us to tackle, for instance, models with node homophily. We provide results for the latter and show that they accurately reproduce the outcome of Monte Carlo simulations.

I. INTRODUCTION

Networks are currently one of the most useful theoretical tools of analysis [1], one that finds applications in many different fields, such as biology [2–5], sociology [6], economics [7,8], or technology [9,10]. Modeling using networks is a two-step process. First of all, we need to identify which elements of a system can play the role of nodes, and which connections, interactions, or relations between pairs of them can play the role of links. In the World Wide Web (WWW), these two elements could be the web pages and the hyperlinks; in a social environment, the individuals and their relationships; in the cell, the proteins and their interactions. The result of this first modeling step is a snapshot of the system cast as a network.

But in most cases this network is just an instance, a single realization of a more general set of networks that we could have obtained if we had modeled other similar systems (another portion of the WWW, another group of people, another cell). Often, the network is literally a snapshot because the system evolves in time, so at different instants we observe different realizations of the network. In general, when it comes to modeling through networks, what we look for is a model whose generic features—whichever they may be—reproduce those of the particular instances that we observe. In other words, we look for an *ensemble* of networks from which our particular network is a typical element—and this is the second, and most important, modeling step.

A network ensemble (or random graph model) is specified by two elements: the set $\mathcal{G}$ of possible realizations of the network, and a probability distribution $P(G)$ on this set ($G \in \mathcal{G}$).

It is only natural to write $P(G)$ as

$$P(G) = \Xi^{-1} e^{-H(G)}, \quad \Xi = \sum_{G \in \mathcal{G}} e^{-H(G)}, \quad (1)$$

where the analogy to the Gibbs distribution in statistical mechanics justifies referring to $H(G)$ as the Hamiltonian of the ensemble. This sort of Gibbs's ensembles for networks appeared in the early 1980s under the name of exponential random graphs (ERGs) in the context of social network modeling [11,12] and were inspired by previous work on Markov random fields [13,14]. Hammersley-Clifford's theorem [14,15] provides the conditions under which the probability distribution of a Markov random field has the form (1) of a Gibbs's ensemble.

The Hamiltonian $H(G)$ of an ERG is specified as [16]

$$H(G) = -\sum_{\mu=1}^{r} \lambda_\mu \omega_\mu(G), \quad (2)$$

where $\{\omega_1(G), \ldots, \omega_r(G)\}$ is a set of observables on the network (or graph) G. For instance, $H(G) = \lambda L(G)$ with $L(G)$ defined as the number of links (edges) of the graph is the well-known Erdős-Rényi model [1,16]. The so-called "conjugate" parameters λ_μ are determined by fixing the averages of the observables,

$$\langle \omega_\mu \rangle = \sum_{G \in \mathcal{G}} \omega_\mu(G) P(G) = \frac{\partial}{\partial \lambda_\mu} \ln \Xi. \quad (3)$$

The obvious connection between ERGs and statistical mechanics allows us to obtain these ensembles in a different way—one that sheds light on their meaning. The distribution (1) can be obtained by maximizing the entropy functional $S = -\sum_G P(G) \ln P(G)$ subject to the constraint that $P(G)$ must have specific values of the averages of a certain set of

*cuesta@math.uc3m.es

observables [17]. According to the Bayesian interpretation [18], the probability distribution obtained this way is the most agnostic one, given the information we have—namely, the values of the specified averages. In other words, any other distribution having the same averages would incorporate spurious information that we do not know to be true or false for our system. In this sense, it is optimal in that it maximizes our ignorance beyond the data we have.

Simple models involving single-link observables, such as the (directed or undirected) Erdős-Rényi model or the reciprocity model, have a simple closed-form solution [16,17]. However, as soon as the observables involve two or more links, the models become more difficult to analyze—but also more interesting. The simplest model in which links interact is the two-star model [17,19]. This model exhibits a first-order phase transition when the interaction is strong enough from a low-density to a high-density phase, which can be accurately obtained, in the thermodynamic limit of very large number of nodes using a mean-field approximation.

But perhaps the equivalent to the Ising model for ERGs is Strauss's model of transitive networks [12,20]. This model enhances the clustering of the networks by introducing an interaction associated with triangles. As the two-star model, Strauss's can be studied in mean-field approximation, and it exhibits a similar phase transition.

Ever since it was discovered, researchers have been intrigued by the nature of the phase transition in Strauss's model [17,19–26] and have explored extensions of it in pursuit of ensembles either with more realistic features [25,27–31] or more amenable to analytic calculations [32,33]. What most of these studies seem to imply is that, for certain sets of parameters, there are values of some observables (e.g., the mean number of links or the clustering) that no graph in the ensemble can attain. Furthermore, in this regime the typical graphs of the ensemble abruptly change from sparse to dense upon a slight variation of the control parameter. These facts cast serious doubts on the usefulness of Strauss's model [20], or even of ERGs, in general [16], to reproduce the features of real-life networks. For this reason, alternative models have been proposed exhibiting topological and dynamical features more akin to those of real networks [27–30,34]—even though they still produce graphs that are hard to tune. More recent models choose to impose a given degree distribution as a control mechanism [5,24,31–33], but the phase transition remains and clustered networks are still produced.

The aim of this paper is twofold. First of all, by introducing the language of lattice gases [35] we will show that there is no qualitative difference between the phase transition exhibited by Strauss's model and the condensation transition of an Ising lattice gas [36]. Thermodynamics teaches us how to interpret states that have intermediate densities between a liquid and a gas. Likewise, thermodynamics will provide a description of the sort of networks that we must expect for those "forbidden" values of the observables in Strauss's model.

Second, we will address this problem using a density-functional formalism especially tailored for lattice gases [37,38]. This formalism provides a method to construct a mean-field-like free energy of the system from which everything else can be derived. It also has the advantage that the nonhomogeneous counterpart of Strauss's model can be solved with no extra effort. Networks in which nodes of different types interact in different ways are of this kind, and using them we can study, e.g., the effect of homophily in social networks.

The paper is organized as follows. Section II introduces Strauss's model and its interpretation as a lattice gas of links. The version of the model we will be dealing with is intrinsically inhomogeneous insofar as the interaction parameters are all link dependent. In Sec. III we use a density-functional formalism to obtain the free energy of the system. In Sec. IV we calculate the free energy assuming that every link has equal probability of occurring. We discuss the thermodynamic limit as well as the well-known phase transition that Strauss's homogeneous model exhibits. The lattice-gas viewpoint we are adopting reveals that this transition is akin to a condensation in fluids, and this analogy allows us to discuss the nature of the system in the region of coexistence. It is one of the main points of this paper because it stands for the validity of the model even in the coexistence region—which has been questioned so far. We end the section by comparing our results to those of Park and Newman [20] for finite networks where the present approach proves to be more accurate (they both coincide in the thermodynamic limit). Finally, Sec. V discusses some of the differences that inhomogeneities (e.g., homophily when there are different kinds of nodes) introduced in the system as a way of illustrating the ability of the functional developed here to deal with this sort of situations. We conclude in Sec. VI with a summary and a brief discussion.

II. STRAUSS'S MODEL AND ITS LATTICE-GAS INTERPRETATION

Strauss's model is an ERG ensemble of undirected graphs with N nodes defined by the Hamiltonian,

$$-H(G) = \phi L(G) + \frac{\gamma}{N} T(G), \qquad (4)$$

where $L(G)$ is the number of links (edges) in the graphs and $T(G)$ the number of triangles (clustering). In this model, a positive ϕ enhances the creation of links and a positive γ enhances the formation of triangles. (Note that we are using different sign conventions than those used in previous works [12,16,17,19,20] in order to agree with those commonly adopted in the statistical mechanics of lattice gases). The factor N^{-1} in front of $T(G)$ accounts for the fact that there are $O(N^3)$ triangles in the network compared to $O(N^2)$ links. With this factor both terms are comparable for large N if the constants are $O(1)$. The model was introduced by Strauss [12] to describe graphs with a clustering higher than that of a typical Erdős-Rényi graph.

Here we will deal with a nonhomogeneous version of this model. In order to write down its Hamiltonian we need to introduce some notation. Let $\mathcal{N}$ denote the set of all nodes of an undirected graph G, $\mathcal{N}_2$ the set of all subsets of two elements of $\mathcal{N}$ (hence, the potential links, which will be denoted by their indices $\{i_1, i_2\}$), and $\mathcal{N}_3$ the set of all subsets of three elements of $\mathcal{N}$ (hence, the potential triangles, which will be denoted $\{i_1, i_2, i_3\}$). If $|\mathcal{N}| = N$, then $|\mathcal{N}_2| = \binom{N}{2}$, $|\mathcal{N}_3| = \binom{N}{3}$. We also denote $\{\tau_{ij}\}$ the adjacency matrix of the graph G, where $\tau_{ij} = 1$ if the link $\{i, j\}$ exists in G and $\tau_{ij} = 0$
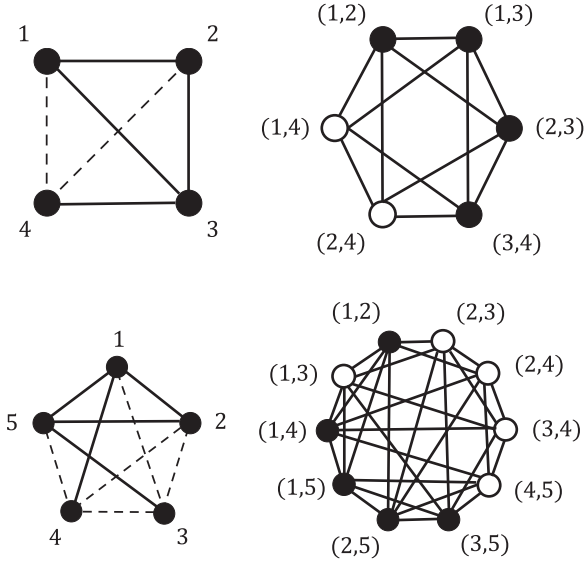
FIG. 1. Dual networks for the complete graphs of four (up) and five nodes (down). In the dual networks, links are represented as nodes, and two of these nodes are neighbors if they share a node of the original one. (For an alternative representation see Fig. 1 of Ref. [35].)

otherwise. Then, in terms of these variables a triangle $\{i, j, k\}$ exists if and only if $\tau_{ij}\tau_{jk}\tau_{ki} = 1$. Thus, the Hamiltonian of the nonhomogeneous version of Strauss's model can be written as

$$-H(G) = \sum_{\{ij\} \in \mathcal{N}_2} \phi_{ij}\tau_{ij} + \sum_{\{ijk\} \in \mathcal{N}_3} \frac{\gamma_{ijk}}{N} \tau_{ij}\tau_{jk}\tau_{ki}. \qquad (5)$$

The parameters ϕ_{ij} and γ_{ijk} are local versions of those of Hamiltonian (4)—hence, a positive ϕ_{ij} enhances the creation of the link $\{i, j\}$ and similarly a positive γ_{ijk} enhances the formation of the triangle $\{i, j, k\}$.

The variables τ_{ij} play the role of "particles" sitting on the links of the *complete* graph over the set of nodes $\mathcal{N}$. If $\tau_{ij} = 1$, it means that the link $\{ij\}$ is occupied by a particle, whereas if $\tau_{ij} = 0$, the link is empty. Thus, G can also be interpreted as the configuration of up to $\binom{N}{2}$ such particles in the complete graph. Under this interpretation ϕ_{ij} can be regarded as a (local) chemical potential and, therefore, Ξ in (1) would play the role of the grand partition function of this system [36–38].

As particles occupy the links, rather than the nodes, of a complete graph, the "space" where these particles live in is weird. As a matter of fact, in this dual network every two particles are either neighbors or second neighbors to each other. The reason is that if links $\{ij\}$ and $\{kl\}$ are not neighbors (i.e., have no common nodes) then they are both neighbors to a common link (e.g., $\{ik\}$). Figure 1 illustrates these dual networks for the complete graphs of four and five nodes. Each link is neighbor to $2(N-2)$ other links, and second neighbor to the remaining $\binom{N-2}{2}$.

The grand potential of this system $\Omega = -\ln \Xi$ is a function of all conjugate fields $\boldsymbol{\phi} = \{\phi_{ij}\}$ and $\boldsymbol{\gamma} = \{\gamma_{ijk}\}$ from which the probability that link $\{ij\}$ is occupied (henceforth,

density) can be obtained as

$$-\frac{\partial \Omega}{\partial \phi_{ij}} = \langle \tau_{ij} \rangle = \rho_{ij}. \qquad (6)$$

A Legendre transform on the grand potential yields the free energy,

$$F(\boldsymbol{\rho}, \boldsymbol{\gamma}) = \sum_{\{ij\}} \phi_{ij}(\boldsymbol{\rho})\rho_{ij} + \Omega(\boldsymbol{\phi}, \boldsymbol{\gamma}), \qquad (7)$$

where $\phi_{ij}(\boldsymbol{\rho})$ is obtained by solving (6) for fixed $\boldsymbol{\rho}$. Differentiating F with respect to the densities,

$$\frac{\partial F}{\partial \rho_{ij}} = \phi_{ij} + \sum_{\{kl\}} \frac{\partial \phi_{kl}}{\partial \rho_{ij}} \rho_{kl} + \sum_{\{kl\}} \frac{\partial \Omega}{\partial \phi_{kl}} \frac{\partial \phi_{kl}}{\partial \rho_{ij}},$$

so if we take (6) into account we finally find

$$\frac{\partial F}{\partial \rho_{ij}} = \phi_{ij}, \qquad (8)$$

which is the usual equation for the chemical potential.

It can be proven that, given the free-energy density functional $F(\boldsymbol{\rho})$ of a system, its equilibrium density is the unique density profile that minimizes the functional $\Omega(\boldsymbol{\rho}) \equiv F(\boldsymbol{\rho}) - \boldsymbol{\phi} \cdot \boldsymbol{\rho}$ [39, Appendix B] (alternatively, it minimizes $F(\boldsymbol{\rho})$ at constant mean density). Thus, Eq. (8), which is dual to (6), is the expression of this variational principle. Hence, its solution provides the values of the densities for a given set of chemical potentials $\boldsymbol{\phi}$.

III. FUNDAMENTAL-MEASURE APPROXIMATION

The technique we will use to find an approximation to the free energy of Strauss's model is know in the field of lattice gases as *fundamental-measure theory* [37,38]. It is a reformulation of the well-known cluster variation method [36]. The idea is to decompose the system in overlapping clusters and express the free energy as a sum of the free energies of those clusters, controlling for overcounting.

In the case of Strauss's model, the geometry of the Hamiltonian suggests that the simplest possible clusters are triangles. Thus, as a first approximation the free energy is obtained as the sum of the contributions to the free energy of all the triangles within the complete graph. However, in doing so every link participates in $N-2$ triangles, so we need to subtract $N-3$ times the contribution to the free energy of all links. In other words, the fundamental-measure approximation to the free energy of this model will be

$$F(\boldsymbol{\rho}, \boldsymbol{\gamma}) = \sum_{\{ijk\}} \Phi_3(\rho_{ij}, \rho_{jk}, \rho_{ki}, \gamma_{ijk})$$
$$-(N-3) \sum_{\{ij\}} \Phi_2(\rho_{ij}), \qquad (9)$$

where Φ_2 and Φ_3 are the free energies of a single link and a single triangle, respectively.

The expression for Φ_2 is easy to obtain. Denoting $z_{ij} \equiv e^{\phi_{ij}}$, the grand partition function for a single link $\{ij\}$ is simply $\Xi_2 = 1 + z_{ij}$. Thus,

$$\rho_{ij} = z_{ij} \frac{\partial}{\partial z_{ij}} \ln \Xi_2 = \frac{z_{ij}}{1 + z_{ij}},$$

from which,

$$z_{ij} = \frac{\rho_{ij}}{1 - \rho_{ij}}, \quad \Xi_2 = \frac{1}{1 - \rho_{ij}}.$$

Substituting these expressions in the Legendre transform $\Phi_2 = \rho_{ij} \ln z_{ij} - \ln \Xi_2$, we end up with

$$\Phi_2(\rho_{ij}) = \rho_{ij} \ln \rho_{ij} + (1 - \rho_{ij}) \ln(1 - \rho_{ij}), \quad (10)$$

which is simply the free energy of an ideal lattice gas.

The calculation of Φ_3 is rather more involved, and is deferred to Appendix A. Introducing the shorthand,

$$\zeta_{ijk} \equiv \exp(\gamma_{ijk}/N) - 1 = \frac{\gamma_{ijk}}{N} + O\left(\frac{1}{N^2}\right), \quad (11)$$

its expression turns out to be

$$\Phi_3 = \Phi_2(\rho_{ij}) + \Phi_2(\rho_{jk}) + \Phi_2(\rho_{ki}) + \rho_{ij} \ln\left(1 - \frac{\rho_{ijk}}{\rho_{ij}}\right)$$

$$+ \rho_{jk} \ln\left(1 - \frac{\rho_{ijk}}{\rho_{jk}}\right) + \rho_{ki} \ln\left(1 - \frac{\rho_{ijk}}{\rho_{ki}}\right)$$

$$- 2 \ln(1 - \rho_{ijk}), \quad (12)$$

where ρ_{ijk} is one of the real solutions of the cubic equation

$$\zeta_{ijk}(\rho_{ij} - \rho_{ijk})(\rho_{jk} - \rho_{ijk})(\rho_{ki} - \rho_{ijk}) = \rho_{ijk}(1 - \rho_{ijk})^2. \quad (13)$$

This "triangle" density is related to $T_{ijk} = \langle \tau_{ij}\tau_{jk}\tau_{ki} \rangle$, the probability that nodes i, j, k form a triangle, as (see Appendix B)

$$T_{ijk} = \frac{1 + \zeta_{ijk}}{\zeta_{ijk}} \rho_{ijk} = \frac{1}{1 - e^{-\gamma_{ijk}/N}} \rho_{ijk}. \quad (14)$$

If we now substitute (10) and (12) into (9) and take into account that

$$\sum_{\{ijk\}} (A_{ij} + A_{jk} + A_{ki}) = (N - 2) \sum_{\{ij\}} A_{ij} \quad (15)$$

for any link-dependent magnitude A_{ij}, we finally get

$$F = \sum_{\{ij\}} [\rho_{ij} \ln \rho_{ij} + (1 - \rho_{ij}) \ln(1 - \rho_{ij})]$$

$$+ \sum_{\{ijk\}} \left[\rho_{ij} \ln\left(1 - \frac{\rho_{ijk}}{\rho_{ij}}\right) + \rho_{jk} \ln\left(1 - \frac{\rho_{ijk}}{\rho_{jk}}\right)\right.$$

$$\left. + \rho_{ki} \ln\left(1 - \frac{\rho_{ijk}}{\rho_{ki}}\right) - 2 \ln(1 - \rho_{ijk})\right]. \quad (16)$$

IV. HOMOGENEOUS NETWORKS

A. Free energy

We can recover Strauss's original model by assuming $\rho_{ij} = \rho$ for every link $\{ij\}$ and $\gamma_{ijk} = \gamma$ for every triangle $\{ijk\}$. Then the free energy *per link* $f(\rho, \gamma) = \binom{N}{2}^{-1} F(\boldsymbol{\rho}, \boldsymbol{\gamma})$ will be

$$f = \rho \ln \rho + (1 - \rho) \ln(1 - \rho)$$

$$+ (N - 2)\left[\rho \ln\left(1 - \frac{\rho_{\mathrm{T}}}{\rho}\right) - \frac{2}{3} \ln(1 - \rho_{\mathrm{T}})\right], \quad (17)$$
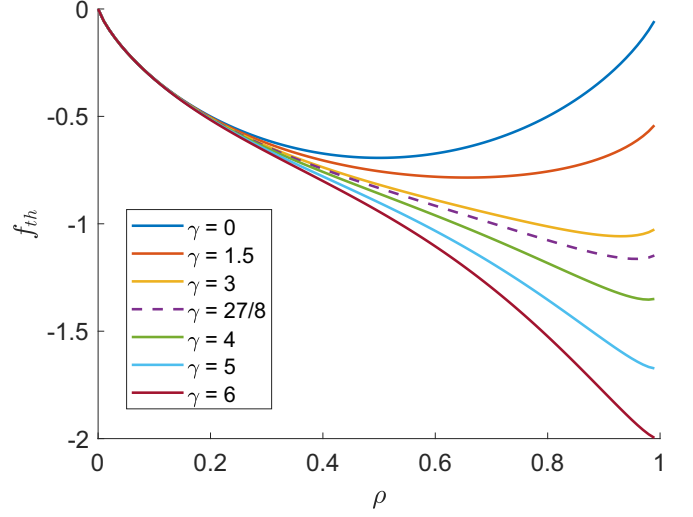
FIG. 2. Thermodynamic free energy of the model for different values of γ below and above the critical value γ_c (γ increases from top to bottom). The curves illustrate the onset of the concavity as γ grows past $\gamma_c = 27/8$, which is represented by the purple dashed line.

where ρ_{T} is the only real root of

$$\zeta(\rho - \rho_{\mathrm{T}})^3 = \rho_{\mathrm{T}}(1 - \rho_{\mathrm{T}})^2. \quad (18)$$

With the change of variable,

$$t = \frac{\rho - \rho_{\mathrm{T}}}{1 - \rho_{\mathrm{T}}}, \quad \rho_{\mathrm{T}} = \frac{\rho - t}{1 - t}, \quad (19)$$

the cubic equation (18) can be rewritten as

$$t^3 + \frac{t}{\zeta(1 - \rho)} - \frac{\rho}{\zeta(1 - \rho)} = 0. \quad (20)$$

This equation has only one real root, which is given by the formula [40, pp. 102–103]

$$t = \frac{2}{\sqrt{3\zeta(1 - \rho)}} \sinh\left[\frac{1}{3} \sinh^{-1}\left(\frac{3}{2}\rho\sqrt{3\zeta(1 - \rho)}\right)\right]. \quad (21)$$

B. Thermodynamic limit

In the thermodynamic limit $N \to \infty$ one can see, either from (18) or directly from (21), that the (thermodynamic) free energy becomes simply

$$f_{\mathrm{th}} = \rho \ln \rho + (1 - \rho) \ln(1 - \rho) - \frac{\gamma \rho^3}{3}. \quad (22)$$

This free energy is convex as long as

$$\frac{\partial^2 f_{\mathrm{th}}}{\partial \rho^2} = \frac{1}{\rho(1 - \rho)} - 2\gamma\rho > 0, \quad (23)$$

which is equivalent to $2\gamma\rho^2(1 - \rho) < 1$. As the maximum value of $\rho^2(1 - \rho)$ is $4/27$ (reached at $\rho_c = 2/3$), the condition above implies $\gamma < 27/8$. So, the values $\gamma_c = 27/8$, $\rho_c = 2/3$, mark a critical point, above which the system exhibits a first-order phase transition.

Figure 2 illustrates the concavity that f_{th} develops as γ increases past γ_c. It is the fingerprint of a condensation transition in lattice gases [36] because a concave free energy implies
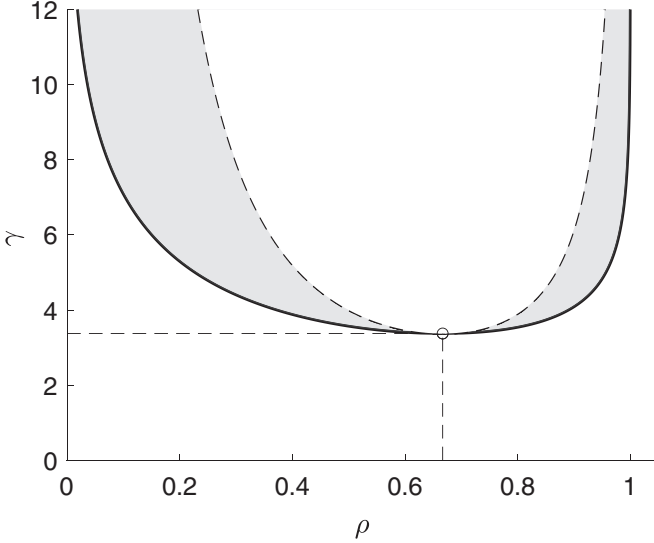
FIG. 3. In the upper region delimited by the solid curve the system is not homogeneous but separated in two coexisting phases whose respective densities are given by the values of curve at the corresponding γ. The dashed line represents the spinodal (25), i.e., the curve at which the compressibility vanishes—hence, the thermodynamic free energy changes from convex to concave. The circle where both curves meet marks the critical point. Within the shaded region the system may be trapped in a metastable homogeneous state.

thermodynamic instability (the compressibility is negative). The homogeneous "fluid" separates in two phases, each of a different density, in thermodynamic equilibrium. The fraction occupied by each phase must be such that the overall density matches the prescribed one.

Thermodynamic equilibrium means "chemical" equilibrium (equality of chemical potentials) and "mechanical" equilibrium (equality of pressures). The first condition implies $f_\rho(\rho_1, \gamma) = f_\rho(\rho_2, \gamma)$; the second condition implies $\rho_1 f_\rho(\rho_1, \gamma) - f(\rho_1, \gamma) = \rho_2 f_\rho(\rho_2, \gamma) - f(\rho_2, \gamma)$. Both conditions are summarized in the equation,

$$f_\rho(\rho_1, \gamma) = f_\rho(\rho_2, \gamma) = \frac{f(\rho_2, \gamma) - f(\rho_1, \gamma)}{\rho_2 - \rho_1}, \qquad (24)$$

which represent Maxwell's double tangent construction [41]. For f_{th}, the solution of these equations is represented in Fig. 3. Given any $\rho_1 < \rho < \rho_2$, there will be a fraction x of the graph of density ρ_1 and a fraction $1 - x$ of density ρ_2 such that $\rho = x\rho_1 + (1 - x)\rho_2$.

On the other hand, the condition $f_{\rho\rho}(\rho, \gamma) = 0$ marks the points where the compressibility vanishes—i.e., where the system is no longer mechanically stable. This curve is known as the *spinodal* (see Fig. 3). According to (23), this curve is

$$\gamma = \frac{1}{2\rho^2(1 - \rho)}. \qquad (25)$$

Within the region between the coexistence curve and the spinodal (the shaded area of Fig. 3) the system can still be prepared in a homogeneous—but metastable—state. This explains the origin of the hysteresis usually observed in first-order phase transitions—this one in particular [20].

It is difficult to guess the nature of the phase transition that this system undergoes above the critical point. Recall that any two links are separated by no more than one intermediate neighbor. The very notion of space breaks down in such a system, so the picture of the usual condensation transition where gas and liquid occupy different portions of the volume, has no reasonable counterpart in a complete graph. Nonetheless, the transition may be illustrated by computing a histogram of the number of links belonging to a given number of triangles [24]. As a way of illustration we have obtained such histograms by performing Monte Carlo simulations using the dynamics of Kawasaki [42], which preserves the number of links—hence, the density ρ. In this dynamics, a Monte Carlo step amounts to first removing a link at random and then creating a link also at random. The results, obtained for three different values of the interaction γ (below, just above, and well above the critical point) and three different densities, are depicted in Fig. 4 . In each of these simulations we perform 5×10^5 Monte Carlo steps. When $\gamma < \gamma_c$ the histograms show a single peak that shifts to the right and shrinks as the density increases, whereas if $\gamma > \gamma_c$ the distribution exhibits two very neat peaks, one at high values and the other one at lower values of the number of triangles. Obviously, the links forming each of the two peaks belong to each of the two—low- and high-density—phases. Figure 4 reveals that networks within the coexisting region do exist, but they have different structural properties than those outside this region. We will return to this point in Sec. VI.

C. Finite networks

We can use an asymptotic expansion in N to obtain ρ_T from (21). The first two terms are

$$\rho_T = \frac{\gamma \rho^3}{N}\left[1 + \frac{\gamma(1 - 6\rho^2 + 4\rho^3)}{2N} + O\left(\frac{1}{N^2}\right)\right], \qquad (26)$$

and, consequently, the free energy can be expanded as

$$f = \rho \ln \rho + (1 - \rho)\ln(1 - \rho) - \frac{\gamma \rho^3}{3}\left[1 - \frac{2}{N}\right.$$
$$\left. + \frac{\gamma(1 - 3\rho^2 + 2\rho^3)}{2N} + O\left(\frac{1}{N^2}\right)\right]. \qquad (27)$$

The critical point will then be a solution of the equations $f_{\rho\rho} = f_{\rho\rho\rho} = 0$, which yields

$$\gamma_c(N) = \frac{27}{8}\left[1 + \frac{45}{16N} + O\left(\frac{1}{N^2}\right)\right],$$
$$\rho_c(N) = \frac{2}{3} + O\left(\frac{1}{N^2}\right). \qquad (28)$$

The numerical solution for $\gamma_c(N)$ is depicted in Fig. 5 along with the asymptotic expansion above. As for $\rho_c(N)$, within numerical resolution we find that its value is always 2/3. It is noteworthy that the curve of $\gamma_c(N)$ diverges somewhere between $N = 4$ and $N = 3$.

Despite that the uniform free energy (27) predicts a critical point and a first-order phase transition for arbitrary N (as low as $N = 4$, see Fig. 5), we know that this is not possible; in other words, this phase transition is not real. The reason why the free energy exhibits a concavity for some values
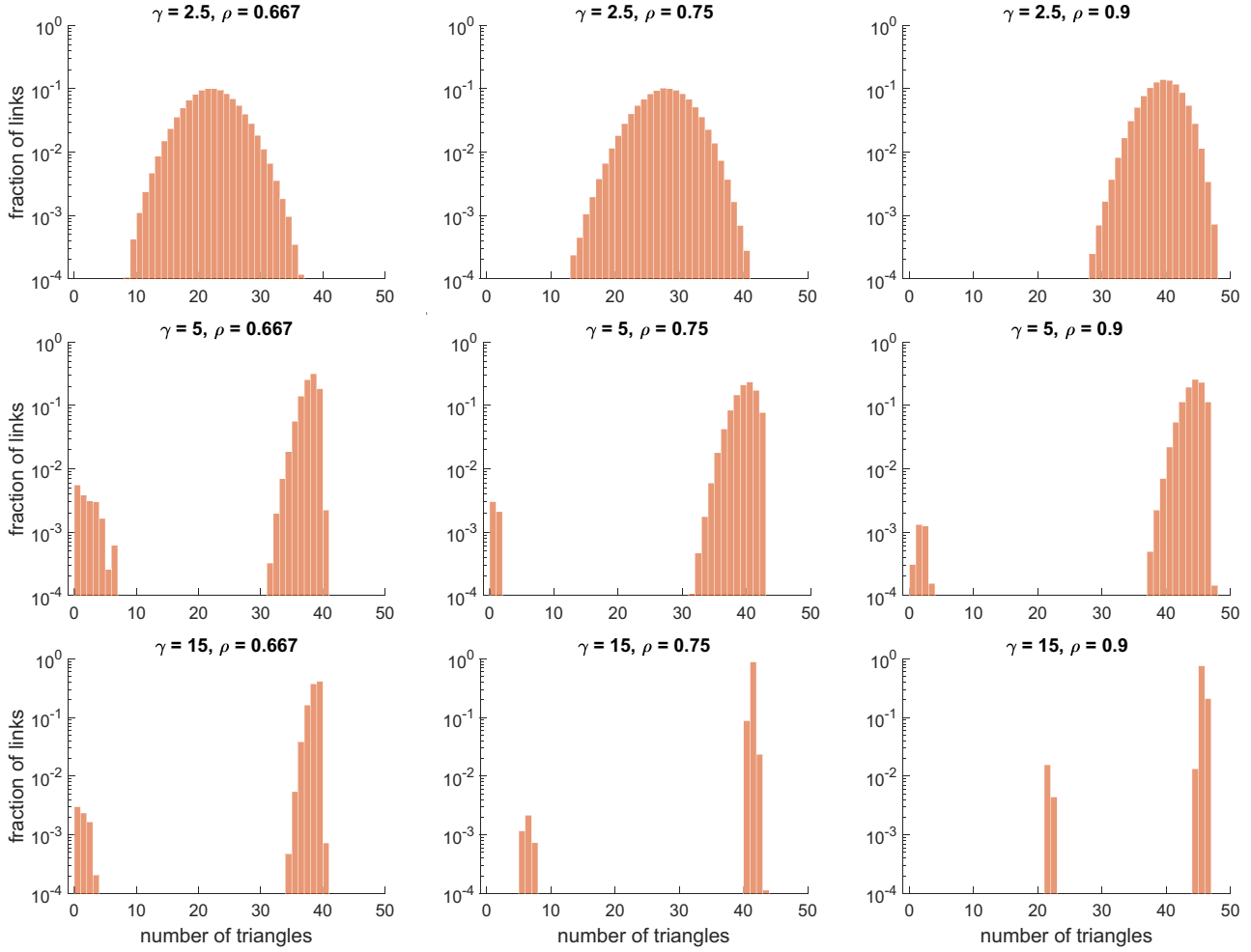
FIG. 4. Each of these nine panels depicts the average fraction of links belonging to a given number of triangles as obtained from Monte Carlo simulations—using Kawasaki dynamics—of a Strauss network with $N = 50$ nodes for three values of the interaction parameter γ (below, just above, and well above the critical point) and three different densities.
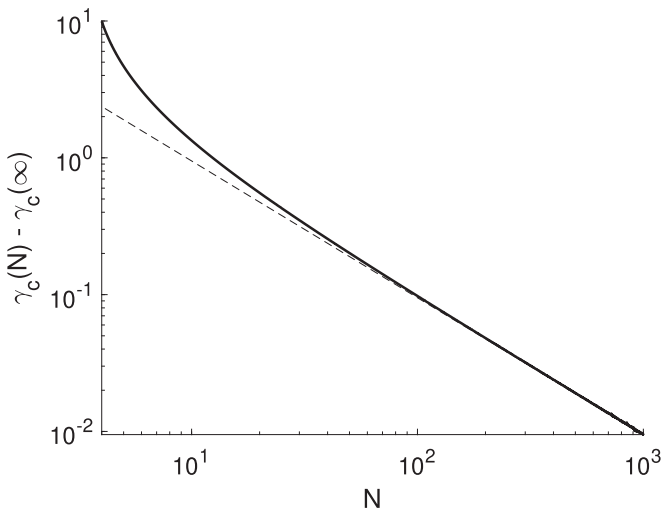

FIG. 5. Difference between the critical value $\gamma_c(N)$ for a network with N nodes and its limit for $N \to \infty$ as a function of N. The solid line is obtained by numerically solving the equations for the critical point; the dashed line arises from the asymptotic expression (28).

of ρ when $\gamma > \gamma_c(N)$ is that the equilibrium solution is not truly uniform for any value of the density—even though for some densities it is indistinguishable from a uniform one. The transition from the regions where the solution is almost uniform to those in which the structure is like those shown in Fig. 4 (third row) is continuous—albeit probably abrupt. Thus, in the approximation we are using here, it shows up as a phase transition (i.e., the convex envelope of the free energy as a function of ρ represents a good approximation of the true free energy of the system). In order to illustrate this point we compare our approximate free energy for $N = 4$ with the exact one (obtained in Appendix C) in Fig. 6.

D. Comparison with Park and Newman's mean-field calculations

A fair question to ask is how does the present theory compares with the mean-field calculations of Park and Newman [20]. In spirit, this theory is also mean-fieldlike, but clearly its construction follows a very different approach. On the other hand because of the high dimensionality of this system one
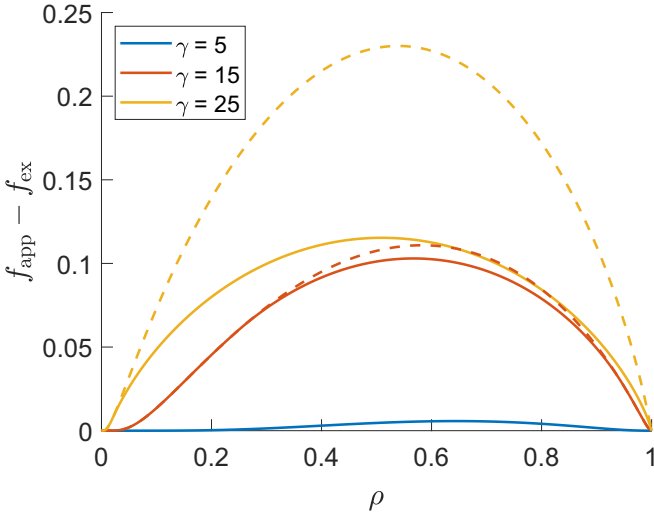
FIG. 6. Dashed lines represent the difference between the approximate free energy (f_{app}), given by Eq. (27), and the exact free energy (f_{ex}), given by Eq. (C4) for the complete graph with $N = 4$ nodes and for three values of the interaction parameter γ (below, just above, and well above the pseudocritical point). As the free energy has an unphysical concave region above the critical point, we also represent in solid lines the difference between the *convex envelope* of the approximate free energy and the exact one since this convex envelope is a better approximation to the real free energy.

expects that in the thermodynamic limit it becomes exact [20], so it would be desirable that, if not for all N, at least, in this limit both theories coincide. Figure 3(b) of Ref. [20] shows the expected number of triangles T (among other things) as a function of the interaction parameter γ/N, for $\phi = -0.53$ and $N = 500$. We can obtain T and ϕ as a function of ρ and γ through Eqs. (8) and (14), respectively. From these two calculations, we can obtain parametrically the curve $T(\gamma)$ for fixed ϕ and for different values of N. The discrepancy between our results and those of Park and Newman is shown in Fig. 7 [20]. Figure 7(a) illustrates that the difference between the predictions of both theories decreases with system size—so that they both coincide in the thermodynamic limit. However, for very small networks their predictions differ significantly (e.g., for $N = 10$, the discrepancy may be as high as $\sim 20\%$). Figure 7(b) compares the predictions of both theories for $N = 10$ along with Monte Carlo simulations performed using a Metropolis-Hastings algorithm [43]. Observable magnitudes are averaged over 10^6 configurations of the Markov chain, taken one every 5×10^4 steps. This figure highlights the higher accuracy of the current theory in calculating results for small networks.

Given that most real networks are large, the discussion of this section may seem as an academic issue of little practical relevance. Nevertheless, small networks of about 20–30 nodes are common, e.g., in social science, anthropology, or biology. Thus, one can find instances of these small networks in studies of different social organizations [44–47], in bands of hunter gatherers [48,49], or in groups of social animals [50–52]. For these studies, the improved accuracy provided by the present approach might not be negligible.
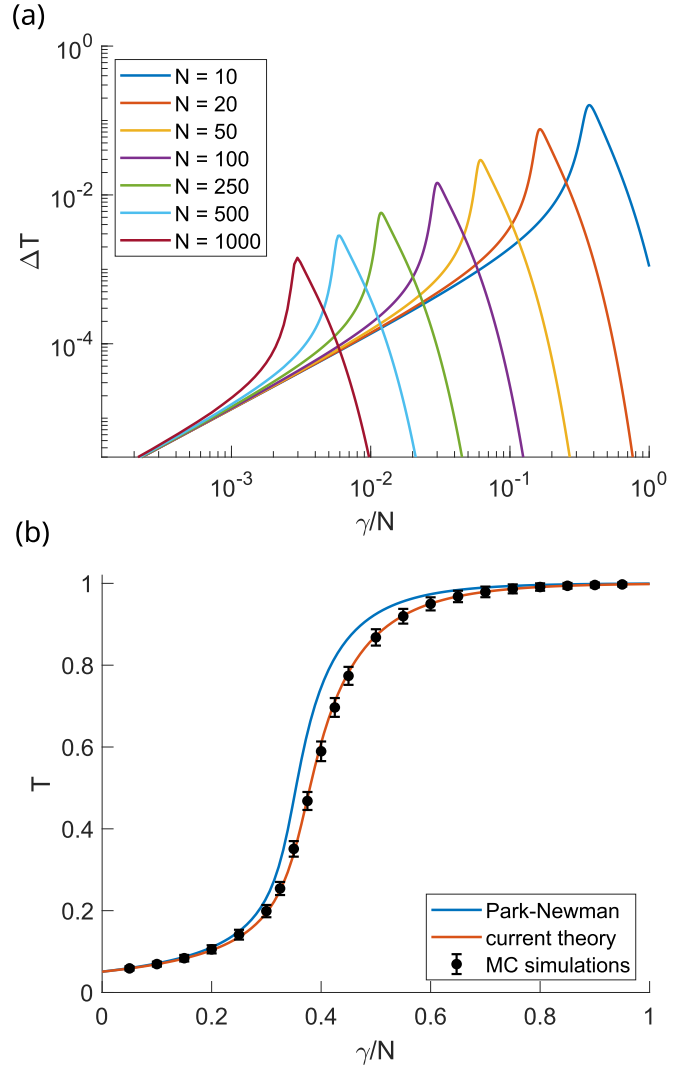

FIG. 7. (a) Difference between the expected number of triangles T as obtained from Park and Newman's mean-field calculations and from the current theory [20]. (b) Expected number of triangles for $N = 10$ according to both theories along with Monte Carlo simulations. Error bars represent the standard deviation of the number of triangles along the simulations. In both panels $\phi = -0.53$.

V. NONHOMOGENEOUS NETWORKS: HOMOPHILY

Having an expression for the free energy of the nonhomogeneous Strauss's model allows us to tackle other interesting cases. Particularly important is the case where there are different types of nodes in the network with different interaction parameters. This case can model, e.g., homophily in a social network where like nodes are more prone to form links or triangles than different nodes are [53]. A particular version of this model has already been used to study segregation on Strauss networks where triangles are both favored or disfavored [54].

Suppose we have two types of nodes in the network, A and B. Since the underlying graph is a complete graph, the actual location of these nodes is irrelevant, only how many of each type there are matters. So let us assume that there are N_A of type A and $N_B = N - N_A$ of type B. Accordingly,

$\binom{N_A}{2}$ links are homophilic of type AA, $\binom{N_B}{2}$ of type BB, and $N_A N_B$ are of mixed type. Likewise, there will be $\binom{N_A}{3}$ homophilic triangles of type AAA, $\binom{N_B}{3}$ of type BBB, $\binom{N_A}{2}N_B$

mixed triangles of type AAB, and $\binom{N_B}{2}N_A$ of type ABB. Hence, the free energy of the system can be obtained as

$$
\begin{aligned}
F = & \sum_{X=A,B} \binom{N_X}{2}[\rho_{XX} \ln \rho_{XX} + (1 - \rho_{XX}) \ln(1 - \rho_{XX})] \\
& + N_A N_B [\rho_{AB} \ln \rho_{AB} + (1 - \rho_{AB}) \ln(1 - \rho_{AB})] \\
& + \sum_{X=A,B} \binom{N_X}{3} \left[3\rho_{XX} \ln \left(1 - \frac{\rho_{XXX}}{\rho_{XX}} \right) - 2 \ln(1 - \rho_{XXX}) \right] + \sum_{X=A,B} \sum_{Y \neq X} \binom{N_X}{2} \\
& \times N_Y \left[\rho_{XX} \ln \left(1 - \frac{\rho_{XXY}}{\rho_{XX}} \right) + 2\rho_{XY} \ln \left(1 - \frac{\rho_{XXY}}{\rho_{XY}} \right) - 2 \ln(1 - \rho_{XXY}) \right],
\end{aligned}
\tag{29}
$$

where $\rho_{XY} = \rho_{YX}$ is the density of links of type XY, and the densities associated with the triangles are the solutions of

$$
\zeta_{XXY}(\rho_{XX} - \rho_{XXY})(\rho_{XY} - \rho_{XXY})^2 = \rho_{XXY}(1 - \rho_{XXY})^2.
\tag{30}
$$

In order to reduce the number of parameters of the model we will, henceforth, assume that it is only homophily, and not the nature of the nodes, what determines interactions. This means that there are only two values of the interaction parameter instead of four, namely, $\gamma_{AAA} = \gamma_{BBB} \equiv \gamma_+$, $\gamma_{AAB} = \gamma_{BBA} \equiv \gamma_-$. Furthermore, in the thermodynamic limit, the solution to (30) is

$$
\rho_{XXY} = \rho_{XX} \rho_{XY}^2 \frac{\gamma_\pm}{N} + O\left(\frac{1}{N^2} \right),
\tag{31}
$$

where the subindex of $\gamma_\pm$ depends on whether $X = Y$ (+) or $X \neq Y$ (−). In this same limit, and setting $N_A = uN, N_B = (1 - u)N$, the free energy per link $f \equiv \binom{N}{2}^{-1} F$ turns out to be

$$
\begin{aligned}
f = & u^2 [\rho_{AA} \ln \rho_{AA} + (1 - \rho_{AA}) \ln(1 - \rho_{AA})] \\
& + (1 - u)^2 [\rho_{BB} \ln \rho_{BB} + (1 - \rho_{BB}) \ln(1 - \rho_{BB})] \\
& + 2u(1 - u)[\rho_{AB} \ln \rho_{AB} + (1 - \rho_{AB}) \ln(1 - \rho_{AB})] - \frac{\gamma_+}{3} [u^3 \rho_{AA}^3 + (1 - u)^3 \rho_{BB}^3] \\
& - \gamma_- u(1 - u)\rho_{AB}^2 [u\rho_{AA} + (1 - u)\rho_{BB}].
\end{aligned}
\tag{32}
$$

The convexity of this function is linked to the positive definiteness of its Hessian matrix (entries are ordered as AA, BB, and AB)

$$
H = \begin{pmatrix} \dfrac{u^2}{\rho_{AA}(1 - \rho_{AA})} - 2\gamma_+ u^3 \rho_{AA} & 0 & -2\gamma_- u^2(1 - u)\rho_{AB} \\[2mm] 0 & \dfrac{(1 - u)^2}{\rho_{BB}(1 - \rho_{BB})} - 2\gamma_+(1 - u)^3 \rho_{BB} & -2\gamma_- u(1 - u)^2 \rho_{AB} \\[2mm] -2\gamma_- u^2(1 - u)\rho_{AB} & -2\gamma_- u(1 - u)^2 \rho_{AB} & \dfrac{2u(1 - u)}{\rho_{AB}(1 - \rho_{AB})} - 2\gamma_- u(1 - u)\bar{\rho}(u) \end{pmatrix},
\tag{33}
$$

where we have introduced the shorthand notation $\bar{\rho}(u) \equiv u\rho_{AA} + (1 - u)\rho_{BB}$. This translates into the positiveness of the first two diagonal elements plus $\det H > 0$, in other words,

$$
\chi_{AA} \equiv \frac{1}{\rho_{AA}(1 - \rho_{AA})} - 2\gamma_+ u\rho_{AA} > 0,
\tag{34}
$$

$$
\chi_{BB} \equiv \frac{1}{\rho_{BB}(1 - \rho_{BB})} - 2\gamma_+(1 - u)\rho_{BB} > 0,
\tag{35}
$$

and, removing trivial positive factors,

$$
\chi_{AA}\chi_{BB}\chi_{AB} - 2u(1 - u)\gamma_-^2 \rho_{AB}^2 (\chi_{AA} + \chi_{BB}) > 0,
\tag{36}
$$

with

$$
\chi_{AB} \equiv \frac{1}{\rho_{AB}(1 - \rho_{AB})} - \gamma_- \bar{\rho}(u).
\tag{37}
$$

Without loss of generality we may assume $0 \leqslant u \leqslant 1/2$. With this assumption, inequalities (34) and (35) hold for any set of densities provided

$$
\gamma_+ < \frac{27}{8(1 - u)}.
\tag{38}
$$

For any γ_+ satisfying this constraint, the critical value of γ_- is obtained as the smallest value for which inequality (36) breaks down for some set of densities. This curve is represented in Fig. 8 for several values of u.

Within the region enclosed by these curves, the function (32) correctly describes the system for given values of the two interaction parameters and the three densities. Outside this region, there are values of the densities for which the network is no longer stable as a uniform system; instead, it
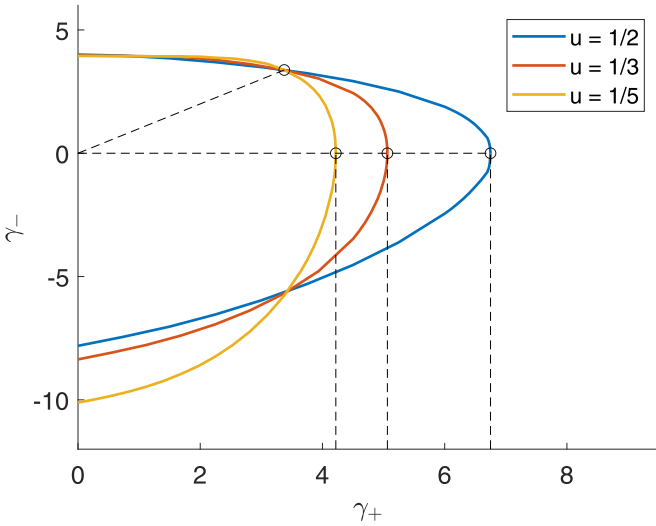
FIG. 8. Critical curves γ_- vs γ_+ for different values of $0 \leqslant u \leqslant 1/2$. The free energy of the nonhomogeneous model is convex only for the points on the left of the curve. The curves reach their rightmost values of γ_+ (marked with circles and vertical dashed lines) for $\gamma_+ = 27/8(1-u)$, $\gamma_- = 0$. The oblique dashed line is $\gamma_+ = \gamma_-$. It meets all critical curves at one point (marked with a circle): $\gamma_+ = \gamma_- = 27/8$, the critical point of the homogeneous system.

fractionates into coexisting subnetworks with different values of those densities.

A detailed analysis of the phase coexistence in the ternary mixture of links that describes this system goes beyond the scope of this paper and will be dealt with in a subsequent study. The real purpose of this section is to illustrate how the functional we have derived can handle variations of the original system, such as this one.

In order to validate the expression of the free energy (32) we have performed Monte Carlo simulations (same method as before) to calculate the fractions of the different kinds of triangles as functions of γ_+ for a system with $N = 50$ nodes, for two values of $u = N_A/N$ (1/2 and 2/5), and for fixed values of the other parameters ($\gamma_-/N = 0.04$ and $\phi_{AA} = \phi_{BB} = \phi_{AB} = -0.25$). Analytic expressions for those fractions of triangles are obtained from (14). As in the case of uniform nodes, the agreement between theory and simulations (Fig. 9) suggests that the expression of the free energy (32) might be exact in the thermodynamic limit.

VI. DISCUSSION AND CONCLUSIONS

In this paper, we have solved approximately Strauss's model of transitive networks using a technique specific of the statistical physics of lattice gases—density-functional theory. The solution we have found is more accurate than a standard mean-field approximation for small systems but coincides with it (and probably with the exact solution) in the thermodynamic limit of infinitely many nodes. The model exhibits a first-order phase transition for triangle interactions above a critical threshold γ_c. For $\gamma > \gamma_c$, upon increasing the probability that links are created the system crosses a region where two solutions are possible—one with a low and one with a high fraction of links (density). Because of this fact,
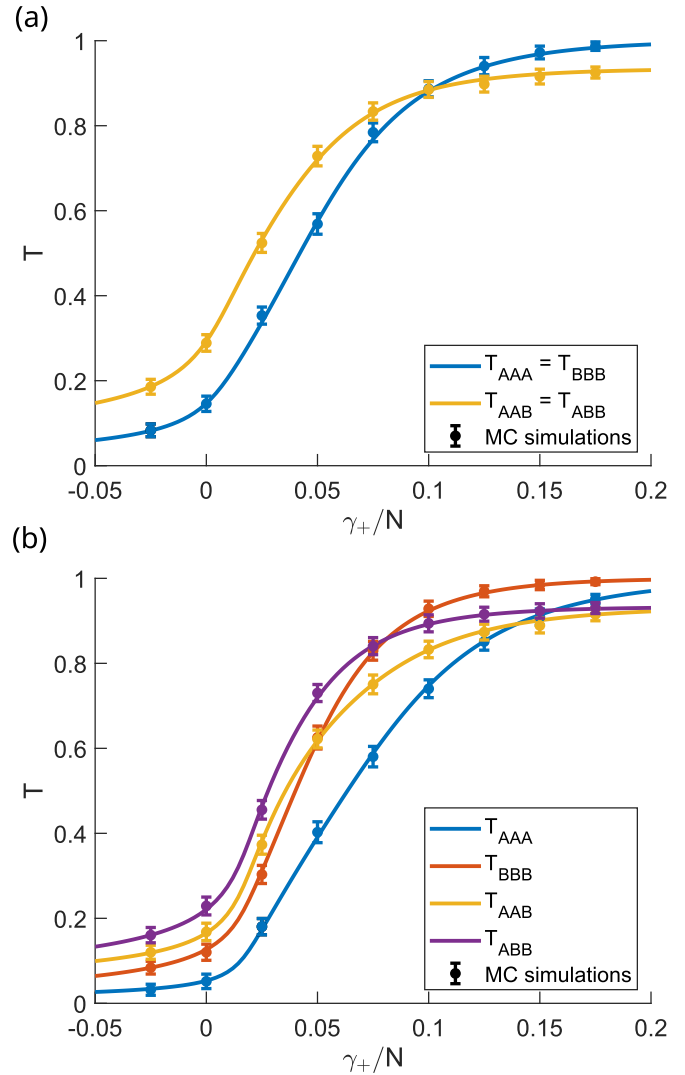

FIG. 9. Fractions of the different kinds of triangles vs γ_+/N for a system with $N = 50$ nodes of two different types A and B. The fraction of A nodes is (a) $u = 1/2$ or (b) $u = 2/5$. Solid lines are the curves obtained from (14), and bullet points are the Monte Carlo results. Error bars represent the standard deviation of the number of triangles along the simulations. In both panels $\gamma_-/N = 0.04$ and $\phi_{AA} = \phi_{BB} = \phi_{AB} = -0.25$.

this model has been deemed unsuitable to produce networks with intermediate fractions of links.

The density-functional formalism that we have employed reveals that the canonical ensemble (constant density and "temperature," i.e., triangle interaction) is the natural description for this system if we want to access this forbidden intermediate states. In this ensemble, the system behaves as a fluid undergoing a condensation transition. The two (low- and high-density) phases are akin to a gas and a liquid, and at those intermediate densities both phases coexist in chemical and mechanical equilibrium. A histogram of links belonging to a given number of triangles shows that links in the graph form two separated groups, each associated with one of these two phases. Hence, graphs within this coexisting region have a different structure to those out of it.

Under this interpretation, the problem of generating graphs with this model in the "inaccessible" coexisting region amounts to performing Monte Carlo simulations using Kawasaki dynamics, which keeps the number of links constant. The idea of accessing these intermediate states by controlling for an extensive parameter has already been suggested [28,34]—although the preferred control variable has always been the number of triangles.

Whether the graphs produced by Strauss's model are a suitable model for some real networks is still an open question. It is true that the peculiar structure of these graphs has never been observed so far, but it is also true that the existence of the phase transition in Strauss's model (and its extensions) is an unavoidable consequence of the specific interaction among its links. We still do not know whether a real system with that particular kind of interactions can be observed or even devised.

The density-functional formalism that we have developed here can further be applied to systems in which the interaction constants are link or triangle dependent. This way we can study systems in which nodes have different types and interactions depending on the type of the nodes involved. Homophily is one of the situations that can be so described. The analysis of the simplest example of homophilic interactions shows that homophily favors the stability of uniform networks (networks with uniform density) by increasing the value of the critical point. The predictions for this case have been validated with Monte Carlo simulations, which—as in the case of uniform networks—seem to suggest that the free energy here obtained might be exact in the thermodynamic limit. Further studies are needed to achieve a full characterization of the complex phase behavior of a system like this.

Finally, we would like to emphasize that the main contribution of this paper is to provide a formalism that can be extended to tackle other ERG models—its advantage with respect to more standard mean-field approaches being that it provides a systematic procedure to deal with them. Hopefully, it may become a useful tool to analyze this class of network models.

APPENDIX A: FREE ENERGY OF A SINGLE TRIANGLE

For a triangle,

$$\Xi_3 = (1 + z_{ij})(1 + z_{jk})(1 + z_{ki}) + \zeta_{ijk} z_{ij} z_{jk} z_{ki}, \qquad \text{(A1)}$$

hence,

$$\rho_{ij} = \frac{z_{ij}(1 + z_{jk})(1 + z_{ki}) + \zeta_{ijk} z_{ij} z_{jk} z_{ki}}{\Xi_3},$$

$$\rho_{jk} = \frac{z_{jk}(1 + z_{ij})(1 + z_{ki}) + \zeta_{ijk} z_{ij} z_{jk} z_{ki}}{\Xi_3}, \qquad \text{(A2)}$$

$$\rho_{ki} = \frac{z_{ki}(1 + z_{ij})(1 + z_{jk}) + \zeta_{ijk} z_{ij} z_{jk} z_{ki}}{\Xi_3}. \qquad \text{(A3)}$$

It will prove convenient to introduce

$$\rho_{ijk} \equiv \frac{\zeta_{ijk} z_{ij} z_{jk} z_{ki}}{\Xi_3}. \qquad \text{(A4)}$$

Now, dividing (A1) by Ξ_3 we get

$$\frac{(1 + z_{ij})(1 + z_{jk})(1 + z_{ki})}{\Xi_3} = 1 - \rho_{ijk}. \qquad \text{(A5)}$$

On the other hand, (A3) can be rewritten as

$$\rho_{ij} - \rho_{ijk} = \frac{z_{ij}(1 + z_{jk})(1 + z_{ki})}{\Xi_3},$$

$$\rho_{jk} - \rho_{ijk} = \frac{z_{jk}(1 + z_{ij})(1 + z_{ki})}{\Xi_3}, \qquad \text{(A6)}$$

$$\rho_{ki} - \rho_{ijk} = \frac{z_{ki}(1 + z_{ij})(1 + z_{jk})}{\Xi_3}. \qquad \text{(A7)}$$

Multiplying them out and using (A4) and (A5) leads to Eq. (13). Also, using (A5) in (A7) we obtain

$$\rho_{ij} - \rho_{ijk} = \frac{z_{ij}}{1 + z_{ij}}(1 - \rho_{ijk}),$$

$$\rho_{jk} - \rho_{ijk} = \frac{z_{jk}}{1 + z_{jk}}(1 - \rho_{ijk}), \qquad \text{(A8)}$$

$$\rho_{ki} - \rho_{ijk} = \frac{z_{ki}}{1 + z_{ki}}(1 - \rho_{ijk}),$$

whose solutions are

$$z_{ij} = \frac{\rho_{ij} - \rho_{ijk}}{1 - \rho_{ij}}, \quad z_{jk} = \frac{\rho_{jk} - \rho_{ijk}}{1 - \rho_{jk}}, \quad z_{ki} = \frac{\rho_{ki} - \rho_{ijk}}{1 - \rho_{ki}}.$$

Substituting these expressions in (A4) and using (13) we obtain

$$\Xi_3 = \frac{(1 - \rho_{ijk})^2}{(1 - \rho_{ij})(1 - \rho_{jk})(1 - \rho_{ki})}. \qquad \text{(A9)}$$

Thus, $\Phi_3 = \rho_{ij} \ln z_{ij} + \rho_{jk} \ln z_{jk} + \rho_{ki} \ln z_{ki} - \ln \Xi_3$ becomes (12).

APPENDIX B: PROBABILITY OF FORMING A TRIANGLE

The probability that nodes i, j, k form a triangle can be obtained from the grand potential as

$$T_{ijk} \equiv \langle \tau_{ij} \tau_{jk} \tau_{ki} \rangle = -N \frac{\partial \Omega}{\partial \gamma_{ijk}}. \qquad \text{(B1)}$$

But inverting the Legendre transform (7),

$$\Omega = F - \sum_{\{ij\}} \phi_{ij} \rho_{ij}, \qquad \text{(B2)}$$

where ρ_{ij} depends $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$ through (8). Thus,

$$T_{ijk} = \sum_{\{lm\}} \phi_{lm} N \frac{\partial \rho_{lm}}{\partial \gamma_{ijk}} - N \frac{\partial F}{\partial \gamma_{ijk}}. \qquad \text{(B3)}$$

Now,

$$\frac{\partial F}{\partial \gamma_{ijk}} = \sum_{\{lm\}} \frac{\partial F}{\partial \rho_{lm}} \frac{\partial \rho_{lm}}{\partial \gamma_{ijk}} + \left(\frac{\partial F}{\partial \gamma_{ijk}} \right)_{\boldsymbol{\rho}},$$

where the last partial derivative is taken at constant $\boldsymbol{\rho}$. Thus, substituting into (B3) and using (8) we obtain

$$T_{ijk} = -N \left(\frac{\partial F}{\partial \gamma_{ijk}} \right)_{\boldsymbol{\rho}} = -(1 + \zeta_{ijk}) \left(\frac{\partial F}{\partial \zeta_{ijk}} \right)_{\boldsymbol{\rho}}. \quad \text{(B4)}$$

Notice that the free energy depends on ζ_{ijk} only through ρ_{ijk} via Eq. (13), therefore,

$$T_{ijk} = -(1 + \zeta_{ijk}) \left(\frac{\partial F}{\partial \rho_{ijk}} \right)_{\boldsymbol{\rho}} \left(\frac{\partial \rho_{ijk}}{\partial \zeta_{ijk}} \right)_{\boldsymbol{\rho}}$$

$$= (1 + \zeta_{ijk}) \rho_{ijk} \left[\frac{1 - 3\rho_{ijk}}{\rho_{ijk}(1 - \rho_{ijk})} + \frac{1}{\rho_{ij} - \rho_{ijk}} \right.$$

$$\left. + \frac{1}{\rho_{jk} - \rho_{ijk}} + \frac{1}{\rho_{ki} - \rho_{ijk}} \right] \left(\frac{\partial \rho_{ijk}}{\partial \zeta_{ijk}} \right)_{\boldsymbol{\rho}}.$$

On the other hand, taking logarithms of (13) and differentiating with respect to ζ_{ijk} at constant $\boldsymbol{\rho}$ we get

$$\frac{1}{\zeta_{ijk}} = \left[\frac{1 - 3\rho_{ijk}}{\rho_{ijk}(1 - \rho_{ijk})} + \frac{1}{\rho_{ij} - \rho_{ijk}} \right.$$

$$\left. + \frac{1}{\rho_{jk} - \rho_{ijk}} + \frac{1}{\rho_{ki} - \rho_{ijk}} \right] \left(\frac{\partial \rho_{ijk}}{\partial \zeta_{ijk}} \right)_{\boldsymbol{\rho}},$$

which finally leads to (14).

APPENDIX C: STRAUSS'S MODEL FOR SMALL NETWORKS

Let $\boldsymbol{\tau}$ denote a vector of components τ_ν, where ν is a subset of two elements of $\{1, 2, \dots, N\}$, and $W \equiv \{0, 1\}^N$. The grand

partition function is defined as

$$\Xi = \sum_{\boldsymbol{\sigma} \in W} \exp \left(\phi \sum_{i<j}^{N} \tau_{ij} + \frac{\gamma}{N} \sum_{i<j<k}^{N} \tau_{ij} \tau_{jk} \tau_{ik} \right)$$

$$= \sum_{L=0}^{\binom{N}{2}} \sum_{T=0}^{\binom{N}{3}} Q(L, T) e^{\phi L + \gamma T/N}, \quad \text{(C1)}$$

where $Q(L, T)$ is the number of configurations $\boldsymbol{\tau}$ with L links and T triangles.

Let us set $N = 4$ and compute the values of $Q(L, T)$. Clearly, $Q(L, 0) = \binom{6}{L}$ for $L = 0$–2, but $Q(L, T) = 0$ otherwise. Furthermore, $Q(5, 2) = 6$, $Q(6, 4) = 1$, and $Q(5, T) = Q(6, T) = 0$ otherwise. As for $L = 3, 4$, there are configurations with either $T = 0$ or $T = 1$. Thus, $Q(3, 1) = 4$ and $Q(3, 0) = \binom{6}{3} - 4 = 16$. On the other hand, $Q(4, 0) = 3$ and $Q(4, 1) = \binom{6}{4} - 3 = 12$. Accordingly, if we denote $x \equiv e^\phi$, $y \equiv e^{\gamma/4}$,

$$\Xi = 1 + 6x + 15x^2 + 16x^3 + 3x^4 + 4x^3(1 + 3x)y$$

$$+ 6x^5 y^2 + x^6 y^4. \quad \text{(C2)}$$

The density can be obtained as

$$\rho = \frac{1}{6} \frac{\partial}{\partial \phi} \ln \Xi = \frac{x}{6} \frac{\partial}{\partial x} \ln \Xi, \quad \text{(C3)}$$

and from that,

$$f = \frac{F}{6} = \rho \ln x - \frac{1}{6} \ln \Xi. \quad \text{(C4)}$$

Thus, fixing the interaction γ (i.e., fixing y) we can obtain parametrically as $0 < x < \infty$, the curve $f(\rho, \gamma)$.

---

[1] M. E. J. Newman, *Networks*, 2nd ed. (Oxford University Press, Oxford, UK, 2018).

[2] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, The large-scale organization of metabolic networks, Nature (London) **407**, 651 (2000).

[3] M. Pascual, Computational ecology: From the complex to the simple and back, PLoS Comp. Biol. **1**, e18 (2005).

[4] J. Iranzo, M. Krupovic, and E. V. Koonin, The double-stranded DNA virosphere as a modular hierarchical network of gene sharing, mBio **7**, e00978-16 (2016).

[5] N. Pospelov, S. Nechaev, K. Anokhin, O. Valba, V. Avetisov, and A. Gorsky, Spectral peculiarity and criticality of a human connectome, Phys. Life Rev. **31**, 240 (2019).

[6] T. A. B. Snijders, Statistical models for social networks, Annu. Rev. Sociol. **37**, 131 (2011).

[7] S. Goyal, *Connections: An Introduction to the Economics of Networks* (Princeton University Press, Princeton, NJ, 2007).

[8] M. O. Jackson, *Social and Economic Networks* (Princeton University Press, Princeton, NJ, 2008).

[9] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Dynamical and Correlation Properties of the Internet, Phys. Rev. Lett. **87**, 258701 (2001).

[10] R. Carvalho, L. Buzna, F. Bono, E. Gutiérrez, W. Just, and D. Arrowsmith, Robustness of trans-European gas networks, Phys. Rev. E **80**, 016106 (2009).

[11] P. W. Holland and S. Leinhardt, An exponential family of probability distributions for directed graphs, J. Am. Stat. Assoc. **76**, 33 (1981).

[12] D. Strauss, On a general class of models for interaction, SIAM Rev. **28**, 513 (1986).

[13] F. Spitzer, Markov random fields and Gibbs ensembles, Am. Math. Mon. **78**, 142 (1971).

[14] J. Besag, Spatial interaction and the statistical analysis of lattice systems, J. R. Stat. Soc. B **36**, 192 (1974).

[15] J. Honerkamp, *Statistical Physics: An Advanced Approach with Applications* (Springer, Berlin, 1998).

[16] A. Fronczak, Exponential random graph models, in *Encyclopedia of Social Network Analysis and Mining*, edited by R. Alhajj and J. Rokne (Springer, New York, 2018), pp. 810–826.

[17] J. Park and M. E. J. Newman, Statistical mechanics of networks, Phys. Rev. E **70**, 066117 (2004).

[18] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, UK, 2003).

[19] J. Park and M. E. J. Newman, Solution of the two-star model of a network, Phys. Rev. E **70**, 066146 (2004).

[20] J. Park and M. E. J. Newman, Solution for the properties of a clustered network, Phys. Rev. E **72**, 026136 (2005).

[21] J. Jonasson, The random triangle model, J. Appl. Probab. **36**, 852 (1999).

[22] Z. Burda, J. Jurkiewicz, and A. Krzywicki, Network transitivity and matrix models, Phys. Rev. E **69**, 026106 (2004).

[23] S. Chatterjee and P. Diaconis, Estimating and understanding exponential random graph models, Ann. Stat. **41**, 2428 (2013).

[24] M. V. Tamm, A. B. Shkarin, V. A. Avetisov, O. V. Valba, and S. K. Nechaev, Islands of Stability in Motif Distributions of Random Networks, Phys. Rev. Lett. **113**, 095701 (2014).

[25] V. Avetisov, M. Hovhannisyan, A. Gorsky, S. Nechaev, M. Tamm, and O. Valba, Eigenvalue tunneling and decay of quenched random network, Phys. Rev. E **94**, 062313 (2016).

[26] M. Yin, A detailed investigation into near degenerate exponential random graphs, J. Stat. Phys. **164**, 241 (2016).

[27] P. Holme and B. J. Kim, Growing scale-free networks with tunable clustering, Phys. Rev. E **65**, 026107 (2002).

[28] M. E. J. Newman, Random Graphs with Clustering, Phys. Rev. Lett. **103**, 058701 (2009).

[29] D. Foster, J. Foster, M. Paczuski, and P. Grassberger, Communities, clustering phase transitions, and hysteresis: Pitfalls in constructing network ensembles, Phys. Rev. E **81**, 046115 (2010).

[30] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, Statistical physics of exchangeable sparse network ensembles, Phys. Rev. E **90**, 042806 (2014).

[31] F. A. López and A. C. C. Coolen, Transitions in random graphs of fixed degrees with many short cycles, J. Phys. Complex. **2**, 035010 (2021).

[32] F. A. López, P. Barucca, M. Fekom, and A. C. C. Coolen, Exactly solvable random graph ensemble with extensively many short cycles, J. Phys. A: Math. Theor. **51**, 085101 (2018).

[33] F. A. López and A. C. C. Coolen, Imaginary replica analysis of loopy regular random graphs, J. Phys. A: Math. Theor. **53**, 065002 (2020).

[34] J. C. Miller, Percolation and epidemics in random clustered networks, Phys. Rev. E **80**, 020901(R) (2009).

[35] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Statistical mechanics of topological phase transitions in networks, Phys. Rev. E **69**, 046117 (2004).

[36] D. A. Lavis and G. M. Bell, *Statistical Mechanic of Lattice Systems*, Vol. 1 (Springer, Berlin, 1999).

[37] L. Lafuente and J. A. Cuesta, Density Functional Theory for General Hard-Core Lattice Gases, Phys. Rev. Lett. **93**, 130603 (2004).

[38] L. Lafuente and J. A. Cuesta, Cluster density functional theory for lattice models based on the theory of Möbius functions, J. Phys. A **38**, 7461 (2005).

[39] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 4th ed. (Academic, Cambridge, 2013).

[40] G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra* (CRC, London, 1997).

[41] K. Huang, *Statistical Mechanics*, 2nd ed. (Wiley, New York, 1987).

[42] K. Kawasaki, Kinetics of Ising models, in *Phase Transitions and Critical Phenomena*, Vol. 2, edited by C. Domb and M. S. Green (Academic, New York, 1972), Chap. 11, pp. 443–501.

[43] T. A. B. Snijders, Markov chain Monte Carlo estimation of exponential random graph models, J. Soc. Struct. **2**, 1 (2002).

[44] G. Huitsing and R. Veenstra, Bullying in classrooms: Participant roles from a social network perspective, Aggressive Behav. **38**, 494 (2012).

[45] M. G. Everett and S. P. Borgatti, Networks containing negative ties, Soc. Networks **38**, 111 (2014).

[46] C. Stadtfeld, K. Takács, and A. Vörös, The emergence and stability of groups in social networks, Soc. Networks **60**, 129 (2020).

[47] D. Escribano, V. Doldán-Martelli, F. J. Lapuente, J. A. Cuesta, and A. Sánchez, Evolution of social relationships between first-year students at middle school: From cliques to circles, Sci. Rep. **11**, 11694 (2021).

[48] A. E. Page, N. Chaudhary, S. Viguier, M. Dyble, J. Thompson, D. Smith, G. D. Salali, R. Mace, and A. B. Migliano, Hunter-gatherer social networks and reproductive success, Sci. Rep. **7**, 1153 (2017).

[49] A. B. Migliano, F. Battiston, S. Viguier, A. E. Page, M. Dyble, R. Schlaepfer, D. Smith, L. Astete, M. Ngales, J. Gomez-Gardenes, V. Latora, and L. Vinicius, Hunter-gatherer multilevel sociality accelerates cumulative cultural evolution, Sci. Adv. **6**, eaax5913 (2020).

[50] C. Kasper and B. Voelkl, A social network analysis of primate groups, Primates **50**, 343 (2009).

[51] A. Ilany, A. Barocas, L. Koren, M. Kam, and E. Geffen, Structural balance in the social networks of a wild mammal, Anim. Behav. **85**, 1397 (2013).

[52] D. Escribano, V. Doldán-Martelli, K. A. Cronin, D. B. Haun, E. J. van Leeuwen, J. A. Cuesta, and A. Sánchez, Chimpanzees organize their social relationships like humans, Sci. Rep. **12**, 16641 (2022).

[53] M. McPherson, L. Smith-Lovin, and J. M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. **27**, 415 (2001).

[54] V. Avetisov, A. Gorsky, S. Maslov, S. Nechaev, and O. Valba, Phase transitions in social networks inspired by the Schelling model, Phys. Rev. E **98**, 032308 (2018).