Analytic solution of the two-star model with correlated degrees

Maíra Bolfe

Physics Department, Federal University of Santa Maria, 97105-900 Santa Maria, Brazil

Fernando L. Metz [*]

Physics Institute, Federal University of Rio Grande do Sul, 91501-970 Porto Alegre, Brazil
and London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, United Kingdom

Edgar Guzmán-González and Isaac Pérez Castillo [†]

Departamento de Física, Universidad Autónoma Metropolitana-Iztapalapa, San Rafael Atlixco 186, Ciudad de México 09340, México

Exponential random graphs are important to model the structure of real-world complex networks. Here we solve the two-star model with degree-degree correlations in the sparse regime. The model constraints the average correlation between the degrees of adjacent nodes (nearest neighbors) and between the degrees at the end-points of two-stars (next nearest neighbors). We compute exactly the network free energy and show that this model undergoes a first-order transition to a condensed phase. For non-negative degree correlations between next nearest neighbors, the degree distribution inside the condensed phase has a single peak at the largest degree, while for negative degree correlations between next nearest neighbors the condensed phase is characterized by a bimodal degree distribution. We calculate the degree assortativities and show they are nonmonotonic functions of the model parameters, with a discontinuous behavior at the first-order transition. The first-order critical line terminates at a second-order critical point, whose location in the phase diagram can be accurately determined. Our results can help to develop more detailed models of complex networks with correlated degrees.

I. INTRODUCTION

Random graphs constitute the main tool to model the complex behavior of large empirical networks observed in social, technological, and biological systems [1,2]. In random graph models a network is typically represented by nodes that interact through edges. Random graph theory leads to important insights into the structure of networks as well as on the dynamical processes occurring on them, such as the spreading of diseases [3,4], the stability of ecosystems to perturbations [5,6], and the dynamics of sparsely connected neurons [7,8].

A fruitful approach to network modeling consists of measuring a set of observables in an empirical network and then building an ensemble of random graphs that matches these features in an average sense [9,10]. The probability distribution of graph configurations is derived by maximizing the network entropy subject to the constraints dictated by the empirical observations [11]. The resulting family of models, known as *exponential random graph* (ERG) models, aims to reproduce a set of empirical patterns while keeping other network properties entirely random. ERGs were introduced in the pioneering work of Holland and Leinhardt [12], and they soon became popular models in social network analysis [13–17].

There are at least two main motivations to study ERGs. First, they serve as benchmark models to distinguish between random and nonrandom traits in the structure of real-world networks [18–20]. Analytic solutions of ERGs give detailed information on the expected values of structural observables and their fluctuations, which can be directly compared with data from real-world networks. Second, ERG models may exhibit degenerate configurations and phase transitions, i.e, abrupt changes in the macroscopic properties of the graph ensemble. Discontinuous phase transitions can be a serious limitation in the generation of ERGs, since they prevent that certain configurations, with the desired structural features, are sampled. Analytic solutions of ERGs predict the existence and location of phase transitions in the parameter space.

ERG models with specific constraints have been widely studied through numeric [21–24] and analytic [10,11,25–33] techniques from statistical mechanics. The two-star model is probably the simplest of ERGs that undergoes a phase transition [11,26]. In this model, the graph ensemble is constrained by the average number of edges and the average number of two-stars (a two-star is a pair of edges that share a common node). The two-star model has been originally solved in the high-connectivity regime [11,26], using mean-field techniques, and more recently in the sparse regime [30], where each node is adjacent to a finite number of others.

Here we take the theory of ERGs one step further by solving the two-star model with degree-degree correlations [34,35], which is an important structural feature of networks.

---

[*]fmetzfmetz@gmail.com
[†]iperez@izt.uam.mx

In general, the degrees in real-world networks are not independent, but they can be positively or negatively correlated with each other, as quantified by the Pearson correlation coefficient [34]. Nodes with similar degrees have positive degree correlations, whereas nodes with highly distinct degrees have negative degree correlations. For instance, the degrees of adjacent or nearest neighbor nodes in social networks are positively correlated [34,35], but these correlations become negative for pairs of nodes connected through paths with more than one edge [36,37].

Earlier works have focused on nearest neighbor degree correlations [34,35] and their impact on dynamical processes on networks, such as the spreading of diseases [38,39] and the synchronization of coupled oscillators [40–42]. However, models that only take into account local properties [43,44] do not reproduce certain global features of networks, such as their community structure or the distribution of the shortest path length. In fact, recent works [36,37,45–48] have shown that long-range degree correlations between nodes separated by more than one edge are important for the organization of networks at a global level. For instance, results suggest that the fractal structure of scale-free networks requires long-range degree correlations between hubs [44]. Long-range degree correlations have been also observed in the airport transportation network of the United States [36], transcriptional regulatory networks [36], coauthorship networks [37], and the Twitter network [48]. Therefore, the solution of ERG models that incorporate degree correlations in a systematic way represents significant progress in network modeling.

In this work we solve the two-star model with degree-degree correlations between nearest neighbors and between next nearest neighbors in the sparse regime. The free energy is exactly calculated thanks to the introduction of an upper cutoff in the degree sequence. We show that the phase diagram of the model exhibits a first-order critical line, surrounded by a metastable region, in which the graph sampling process may get stuck in a local minimum of the free energy. The first-order transition separates a phase characterized by an approximate Poisson degree distribution from a condensed phase, where the degree distribution strongly depends on the degree correlations. We quantify the degree correlations through the degree assortativity corresponding to nearest neighbor nodes [34,35] and to next nearest neighbor nodes, located at the end-points of two-stars. When the degree assortativity of next nearest neighbors is non-negative, the degree distribution in the condensed phase is peaked at the maximum degree; when the assortativity of next nearest neighbors is negative, the condensed phase is characterized by a bimodal degree distribution. Both assortativities are non-monotonic functions of the model parameters and exhibit a nonanalytic behavior at the first-order transition. The main theoretical findings are well corroborated by Monte Carlo simulations.

In the next section we introduce the generic framework of ERG models. In Sec. III we define the main structural properties of interest, including the two assortativity coefficients. Section IV explains how the model is analytically solved using conventional techniques of statistical mechanics, and how the structural properties follow from the free energy. The numerical results, obtained from the solutions of the saddle-point equations, are discussed in Sec. V, while in the last section

we present some final remarks. The Appendix discusses the symmetry properties of the order-parameter functions.

II. EXPONENTIAL RANDOM GRAPH MODELS

A graph configuration of a network with N nodes can be represented by a realization of the $N \times N$ adjacency matrix $\boldsymbol{C}$. The entries of $\boldsymbol{C}$ fully encode the network topology, i.e., the matrix element C_{ij} tells whether there is an edge joining nodes i and j. We consider undirected and simple random graphs [1], which means that $\boldsymbol{C}$ is a symmetric matrix with all diagonal elements equal to zero. If there is an edge connecting nodes i and j, then we set $C_{ij} = 1$, whereas $C_{ij} = 0$ otherwise. The degree K_i of a node i

$$K_i = \sum_{j=1}^{N} C_{ij} \tag{1}$$

counts the number of edges attached to i, and the sequence $K_1, \ldots, K_N$ contains important information about the network structure. In this work we consider random graph models in which the maximum degree that may appear in a graph configuration is $k_{\max}$. The cutoff $k_{\max}$ is a model parameter, independent of N, which can be freely adjusted. As we will see below, the introduction of $k_{\max}$ allows us to compute exactly the network properties in the limit $N \to \infty$.

The probability $\mathcal{P}_N(\boldsymbol{C})$ to observe a certain graph configuration $\boldsymbol{C}$ follows the Boltzmann-like form [11]

$$\mathcal{P}_N(\boldsymbol{C}) = \frac{e^{-\mathcal{H}_N(\boldsymbol{C})}}{\mathcal{Z}_N} \prod_{i=1}^{N} \Theta(k_{\max} - K_i), \tag{2}$$

where $\Theta(x) = 1$ if $x \geqslant 1$, and $\Theta(x) = 0$ otherwise. The graph Hamiltonian $\mathcal{H}_N(\boldsymbol{C})$ depends on the network constraints, and $\mathcal{Z}_N$ is the graph partition function

$$\mathcal{Z}_N = \sum_{\boldsymbol{C}} e^{-\mathcal{H}_N(\boldsymbol{C})} \prod_{i=1}^{N} \Theta(k_{\max} - K_i). \tag{3}$$

The sum $\sum_{\boldsymbol{C}}$ runs over all possible realizations of the adjacency matrix. To study the stability of the graph configurations for $N \to \infty$, we need to compute the free energy density

$$f = -\lim_{N \to \infty} \frac{1}{N} \ln \mathcal{Z}_N, \tag{4}$$

which plays the role of a generating function for the graph structural properties.

III. NETWORK OBSERVABLES

The topology of graphs sampled from $\mathcal{P}_N(\boldsymbol{C})$ can be characterized by a set of structural observables. In this work we only consider global observables, which are obtained by averaging a local quantity over the entire network.

An important quantity to probe the network structure is the empirical degree distribution

$$p_k(\boldsymbol{C}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{k, K_i}, \tag{5}$$

which gives the probability that a randomly chosen node has degree k. The density of edges $\ell(\boldsymbol{C})$ and the density of two-stars $s(\boldsymbol{C})$ are given by [11,30]

$$\ell(\boldsymbol{C}) = \frac{1}{2N} \sum_{ij=1}^{N} C_{ij} = \frac{1}{2N} \sum_{i=1}^{N} K_i, \tag{6}$$

$$s(\boldsymbol{C}) = \frac{1}{2N} \sum_{ijn=1}^{N} (1 - \delta_{in}) C_{ij} C_{jn} = \frac{1}{2N} \sum_{i=1}^{N} \left(K_i^2 - K_i\right). \tag{7}$$

A two-star (or path of length two) is a set with three different nodes $\{i, j, k\}$ such that $C_{ij} C_{jk} = 1$.

Degree correlations are commonly quantified by the degree assortativity coefficient [34–37]. This is a global observable defined as the Pearson correlation coefficient between the degrees of two nodes. Here we characterize the degree-degree correlations by means of two assortativity parameters: the standard assortativity $A^{(1)}(\boldsymbol{C})$ measures the degree correlations between adjacent nodes, while the assortativity $A^{(2)}(\boldsymbol{C})$ measures the degree correlations between nodes at the endpoints of two-stars. In other words, $A^{(1)}$ ($A^{(2)}$) quantifies degree correlations between nearest neighbors (next nearest neighbors).

In what follows, the indexes k and l refer to degrees. For a single graph instance, the assortativities are defined as

$$A^{(r)}(\boldsymbol{C}) = \frac{\sum_{kl=0}^{\infty} kl \, W_{k,l}^{(r)}(\boldsymbol{C}) - \left[\sum_{k=0}^{\infty} k W_k^{(r)}(\boldsymbol{C})\right]^2}{\sum_{k=0}^{\infty} k^2 W_k^{(r)}(\boldsymbol{C}) - \left[\sum_{k=0}^{\infty} k W_k^{(r)}(\boldsymbol{C})\right]^2}, \tag{8}$$

with $r = 1, 2$. The quantity

$$W_{k,l}^{(1)}(\boldsymbol{C}) = \frac{\sum_{ij=1}^{N} C_{ij} \delta_{k,K_i} \delta_{l,K_j}}{\sum_{ij=1}^{N} C_{ij}} \tag{9}$$

is the probability that a randomly chosen edge joins two nodes with degrees k and l, while

$$W_{k,l}^{(2)}(\boldsymbol{C}) = \frac{\sum_{ijn=1}^{N} (1 - \delta_{in}) C_{ij} C_{jn} \delta_{k,K_i} \delta_{l,K_n}}{\sum_{ijn=1}^{N} (1 - \delta_{in}) C_{ij} C_{jn}} \tag{10}$$

is the probability that a randomly chosen two-star has degrees k and l at its end-points. The marginal distributions

$$W_k^{(r)}(\boldsymbol{C}) = \sum_{l=0}^{\infty} W_{k,l}^{(r)}(\boldsymbol{C}) \quad (r = 1, 2) \tag{11}$$

have the explicit forms

$$W_k^{(1)}(\boldsymbol{C}) = \frac{1}{2N \, \ell(\boldsymbol{C})} \sum_{i=1}^{N} K_i \delta_{k,K_i}, \tag{12}$$

$$W_k^{(2)}(\boldsymbol{C}) = \frac{1}{2N \, s(\boldsymbol{C})} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} C_{ij} K_j - K_i \right) \delta_{k,K_i}, \tag{13}$$

which can also be written as follows:

$$W_k^{(1)}(\boldsymbol{C}) = \frac{1}{2 \, \ell(\boldsymbol{C})} k \, p_k(\boldsymbol{C}), \tag{14}$$

$$W_k^{(2)}(\boldsymbol{C}) = \frac{\ell(\boldsymbol{C})}{s(\boldsymbol{C})} \sum_{l=0}^{\infty} (l - 1) W_{k,l}^{(1)}(\boldsymbol{C}). \tag{15}$$

Equation (12) shows that the contribution of a node to $W_k^{(1)}$ is weighted according to its degree, while the weight of node i to the distribution $W_k^{(2)}$ is determined by the number of edges attached to the neighbors of i, except from the links coming from i itself. This is intuitive, as a node i with a large second-order degree $\sum_{j=1}^{N} C_{ij} K_j - K_i$ is the end-point of a large number of two-stars.

The assortativities $A^{(1)}$ and $A^{(2)}$ give the same type of statistical information. Networks with statistically independent degrees satisfy

$$W_{k,l}^{(r)}(\boldsymbol{C}) = W_k^{(r)}(\boldsymbol{C}) W_l^{(r)}(\boldsymbol{C}), \tag{16}$$

and, consequently, $A^{(r)}(\boldsymbol{C}) = 0$. Networks with $A^{(r)}(\boldsymbol{C}) > 0$ are positively correlated or assortative, which means that nodes connected through edges or two-stars are likely to have similar degrees. Finally, networks with $A^{(r)}(\boldsymbol{C}) < 0$ are negatively correlated or disassortative, meaning that nodes with large degrees preferentially connect through edges or two-stars to nodes with small degrees.

Equation (8) is not practical to calculate $A^{(r)}$. To prepare the ground for the analytic computation of the assortativities, let us derive more convenient expressions for $A^{(1)}$ and $A^{(2)}$. Substituting Eqs. (9), (10), (12), and (13) in Eq. (8), we rewrite $A^{(1)}$ and $A^{(2)}$ as follows:

$$A^{(1)}(\boldsymbol{C}) = \frac{\Lambda_{11}(\boldsymbol{C}) - \left(\frac{1}{2N\ell(\boldsymbol{C})} \sum_{i=1}^{N} K_i^2\right)^2}{\frac{1}{2N\ell(\boldsymbol{C})} \sum_{i=1}^{N} K_i^3 - \left(\frac{1}{2N\ell(\boldsymbol{C})} \sum_{i=1}^{N} K_i^2\right)^2}, \tag{17}$$

$$A^{(2)}(\boldsymbol{C}) = \frac{\chi(\boldsymbol{C}) - [\Sigma(\boldsymbol{C})]^2}{\frac{\ell(\boldsymbol{C})}{s(\boldsymbol{C})} \Lambda_{21}(\boldsymbol{C}) - \frac{1}{2Ns(\boldsymbol{C})} \sum_{i=1}^{N} K_i^3 - [\Sigma(\boldsymbol{C})]^2}, \tag{18}$$

with

$$\Sigma(\boldsymbol{C}) = \frac{\ell(\boldsymbol{C})}{s(\boldsymbol{C})} \Lambda_{11}(\boldsymbol{C}) - \frac{1}{2Ns(\boldsymbol{C})} \sum_{i=1}^{N} K_i^2. \tag{19}$$

The object $\Lambda_{qr}(\boldsymbol{C})$ defines higher-order moments of nearest neighbor degrees

$$\Lambda_{qr}(\boldsymbol{C}) = \frac{1}{2N\ell(\boldsymbol{C})} \sum_{ij=1}^{N} C_{ij} K_i^q K_j^r \quad (q, r \geqslant 1), \tag{20}$$

and $\chi(\boldsymbol{C})$ is the correlation between next nearest neighbor degrees

$$\chi(\boldsymbol{C}) = \frac{1}{2Ns(\boldsymbol{C})} \sum_{ijn=1}^{N} (1 - \delta_{in}) C_{ij} C_{jn} K_i K_n. \tag{21}$$

Equations (17) and (18) hold for a single realization of $\boldsymbol{C}$ and they show that $A^{(1)}$ and $A^{(2)}$ are ultimately given in terms of moments of the joint distribution of degrees at different pairs of nodes.

In the limit $N \to \infty$, the fluctuations of intensive variables vanish and a single realization of an intensive quantity coincides with its ensemble averaged value. Thus, we naturally assume that the assortativities and all other global

observables of interest display such self-averaging behavior when $N \rightarrow \infty$, and the assortativities become

$$\langle A^{(1)} \rangle = \frac{\langle \Lambda_{11} \rangle - \left(\frac{\langle K^2 \rangle}{2\langle \ell \rangle} \right)^2}{\frac{\langle K^3 \rangle}{2\langle \ell \rangle} - \left(\frac{\langle K^2 \rangle}{2\langle \ell \rangle} \right)^2}, \tag{22}$$

$$\langle A^{(2)} \rangle = \frac{\langle \chi \rangle - \left[\frac{\langle \ell \rangle}{\langle s \rangle} \langle \Lambda_{11} \rangle - \frac{\langle K^2 \rangle}{2\langle s \rangle} \right]^2}{\frac{\langle \ell \rangle}{\langle s \rangle} \langle \Lambda_{21} \rangle - \frac{\langle K^3 \rangle}{2\langle s \rangle} - \left[\frac{\langle \ell \rangle}{\langle s \rangle} \langle \Lambda_{11} \rangle - \frac{\langle K^2 \rangle}{2\langle s \rangle} \right]^2}, \tag{23}$$

where $\langle \mathcal{G} \rangle$ denotes the ensemble average of an arbitrary random function $\mathcal{G}(\boldsymbol{C})$ for $N \rightarrow \infty$

$$\langle \mathcal{G} \rangle = \lim_{N \rightarrow \infty} \sum_{\boldsymbol{C}} \mathcal{G}(\boldsymbol{C}) \mathcal{P}_N(\boldsymbol{C}). \tag{24}$$

All ensemble averages in Eqs. (22) and (23) can be calculated from the free energy f, which works as a generating function for the moments of degrees.

IV. ANALYTIC SOLUTION OF THE TWO-STAR MODEL WITH CORRELATED DEGREES

In this section we present the Hamiltonian of the two-star model with correlated degrees and we explain how to calculate, in the limit $N \rightarrow \infty$, the free energy f and the structural observables using standard tools from statistical mechanics.

A. The Hamiltonian of the model

The ERG model is defined by the Hamiltonian

$$\mathcal{H}(\boldsymbol{C}) = -\sum_{r=1}^{Q} \alpha_r \sum_{i=1}^{N} F_r(K_i) - \frac{\gamma}{2} \sum_{ij=1}^{N} C_{ij} D(K_i, K_j)$$
$$- \frac{\beta}{2} \sum_{ijk=1}^{N} (1 - \delta_{ik}) C_{ij} C_{jk} K_i K_k + \ln N \sum_{i<j} C_{ij}, \tag{25}$$

where $D(k,l)$ and $F_1(k), \ldots, F_Q(k)$ are arbitrary functions of the degrees, while β, γ, and $\alpha_1, \ldots, \alpha_Q$ are conjugate parameters that enforce the corresponding global constraints. The function $D(k,l)$ fulfills $D(k,l) = D(l,k)$.

From left to right, the first term in Eq. (25) enforces Q global constraints involving single-site functions $F_1(k), \ldots, F_Q(k)$ of the degrees; the second term introduces a global constraint with $D(k,l)$ defined at pairs of adjacent nodes; the third term couples the degrees of the next nearest neighbor nodes located at the end-points of a two-star; finally, due to the logarithmic scaling with N, the fourth term in Eq. (25) ensures that networks sampled from $\mathcal{P}_N(\boldsymbol{C})$ are sparse [11], i.e., the probability of having an edge between two nodes is proportional to $1/N$ and the degrees $K_1, \ldots, K_N$ remain finite in the limit $N \rightarrow \infty$.

We obtain the two-star model with correlated degrees by setting

$$F_r(k) = \delta_{r,1} k + \delta_{r,2} k^2, \quad D(k,l) = kl. \tag{26}$$

The motivation to solve the ERG model described by the Hamiltonian of Eq. (25) is twofold. First, the generic form

of Eq. (25) allows to calculate higher-order moments of the joint distribution of $K_1, \ldots, K_N$ by taking derivatives of f with respect to the conjugate parameters. Such higher-order moments are needed to determine the assortativities $A^{(1)}$ and $A^{(2)}$ of the two-star model [see Eqs. (22) and (23)]. Second, although here we discuss explicit results for the two-star model with correlated degrees, the flexible Hamiltonian of Eq. (25) allows to explore a variety of situations by combining the simultaneous effect of different constraints.

B. The calculation of the free energy

In this subsection we solve the model defined by Eq. (25). The aim is to calculate the free energy f in the limit $N \rightarrow \infty$, from which ensemble averages of the network observables readily follow.

The graph partition function reads

$$\mathcal{Z}_N = \left(\prod_{i<j} \sum_{C_{ij}=0,1} \right) \left[\prod_{i=1}^{N} \Theta(k_{\max} - K_i) \right] e^{-\ln N \sum_{i<j} C_{ij}}$$
$$\times \exp \left(\sum_{r=1}^{Q} \alpha_r \sum_{i=1}^{N} F_r(K_i) + \gamma \sum_{i<j} C_{ij} D(K_i, K_j) \right)$$
$$\times \exp \left(\frac{\beta}{2} \sum_{ijk=1}^{N} (1 - \delta_{ik}) C_{ij} C_{jk} K_i K_k \right). \tag{27}$$

We remind the reader that $K_1, \ldots, K_N$ depend on the matrix elements $\{C_{ij}\}$ according to Eq. (1). With the purpose of linearizing the exponent of Eq. (27) with respect to $\boldsymbol{C}$, we rewrite the above expression using Kronecker δ's

$$\mathcal{Z}_N = \left(\prod_{i<j} \sum_{C_{ij}=0,1} \right) \left[\prod_{i=1}^{N} \sum_{k_i=0}^{N-1} \delta_{k_i, K_i} \Theta(k_{\max} - k_i) \right]$$
$$\times \exp \left(-\ln N \sum_{i<j} C_{ij} \right)$$
$$\times \exp \left(\sum_{r=1}^{Q} \alpha_r \sum_{i=1}^{N} F_r(k_i) + \gamma \sum_{i<j} C_{ij} D(k_i, k_j) \right)$$
$$\times \exp \left(\frac{\beta}{2} \sum_{ijr=1}^{N} (1 - \delta_{ik}) C_{ij} C_{jr} k_i k_r \right).$$

Using the integral representation

$$\delta_{k_i, K_i} = \int_0^{2\pi} \frac{du_i}{2\pi} e^{iu_i(k_i - K_i)} \tag{28}$$

and substituting $K_i = \sum_{j=1}^{N} C_{ij}$, the partition function can be written as

$$\mathcal{Z}_N = \left(\prod_{i<j} \sum_{C_{ij}=0,1} \right) \sum_{k_1, \ldots, k_N=0}^{k_{\max}} \int_0^{2\pi} \left(\prod_{i=1}^{N} \frac{du_i}{2\pi} \right)$$
$$\times \exp \left(i \sum_{i=1}^{N} k_i u_i - i \sum_{i<j} C_{ij}(u_i + u_j) - \ln N \sum_{i<j} C_{ij} \right)$$

$$\times \exp \left(\sum_{r=1}^{Q} \alpha_r \sum_{i=1}^{N} F_r(k_i) + \gamma \sum_{i<j} C_{ij} D(k_i, k_j) \right)$$

$$\times \exp \left[\frac{\beta}{2} \sum_{j=1}^{N} \left(\sum_{i=1}^{N} C_{ij} k_i \right)^2 - \frac{\beta}{2} \sum_{ij=1}^{N} C_{ij} k_i \right]. \quad (29)$$

It is still not possible to sum over the graph configurations, as the exponent in the above equation contains a quadratic term in $\boldsymbol{C}$. We linearize this term via the exact identity

$$\exp \left[\frac{\beta}{2} \sum_{j=1}^{N} \left(\sum_{i=1}^{N} C_{ij} k_i \right)^2 \right]$$

$$= \int_{-\infty}^{\infty} \left(\prod_{j=1}^{N} Dx_j \right) \exp \left(\sqrt{\beta} \sum_{i<j} C_{ij}(x_j k_i + x_i k_j) \right), \quad (30)$$

with the Gaussian measure

$$Dx_j = \frac{dx_j}{\sqrt{2\pi}} e^{-\frac{1}{2} x_j^2}. \quad (31)$$

Equation (30), known as the Hubbard-Stratonovich transformation, simply follows from a Gaussian integral [49]. Substituting Eq. (30) in Eq. (29) and summing over all graph configurations, we arrive at an expression for $N \gg 1$

$$\mathcal{Z}_N = \sum_{k_1,\dots,k_N=0}^{k_{\max}} \int_0^{2\pi} \left(\prod_{i=1}^{N} \frac{du_i}{2\pi} \right) \int_{-\infty}^{\infty} \left(\prod_{j=1}^{N} Dx_j \right)$$

$$\times \exp \left(i \sum_{i=1}^{N} k_i u_i + \sum_{r=1}^{Q} \alpha_r \sum_{i=1}^{N} F_r(k_i) \right)$$

$$\times \exp \left(\frac{1}{2N} \sum_{ij=1}^{N} e^{-i(u_i+u_j)+\mathcal{W}_{\gamma,\beta}(k_i,x_i;k_j,x_j)} \right), \quad (32)$$

where

$$\mathcal{W}_{\gamma,\beta}(k, x; l, x') = \gamma D(k, l) - \frac{\beta}{2}(k^2 + l^2)$$

$$+ \sqrt{\beta} x' k + \sqrt{\beta} x l. \quad (33)$$

The last step is to decouple sites and reduce Eq. (32) to a single-site problem. This is achieved by introducing $k_{\max} + 1$ functional order-parameters

$$\rho_k(x) = \frac{1}{N} \sum_{i=1}^{N} \delta_{k,k_i} \delta(x - x_i) e^{-iu_i}, \quad k = 0, \dots, k_{\max}, \quad (34)$$

through the following identity:

$$1 = \int \left(\prod_{k=0}^{k_{\max}} \mathcal{D}\rho_k \mathcal{D}\hat{\rho}_k \right) \exp \left(i \sum_{k=0}^{k_{\max}} \int dx \rho_k(x) \hat{\rho}_k(x) \right)$$

$$\times \exp \left(-\frac{i}{N} \sum_{i=1}^{N} \hat{\rho}_{k_i}(x_i) e^{-iu_i} \right), \quad (35)$$

where the functional integration measure is formally defined as $\mathcal{D}\rho_k \mathcal{D}\hat{\rho}_k = \lim_{|\mathcal{X}| \to \infty} \prod_{x \in \mathcal{X}} d\rho_k(x) d\hat{\rho}_k(x)/2\pi$, with $\mathcal{X}$ representing the set of all possible values of x obtained

after discretization ($|\mathcal{X}|$ is the size of $\mathcal{X}$). Inserting Eq. (35) in Eq. (32), we obtain

$$\mathcal{Z}_N = \int \left(\prod_{k=0}^{k_{\max}} \mathcal{D}\rho_k \mathcal{D}\hat{\rho}_k \right) \exp \left(i \sum_{k=0}^{k_{\max}} \int dx \rho_k(x) \hat{\rho}_k(x) \right)$$

$$\times \exp \left(\frac{N}{2} \sum_{k,l=0}^{k_{\max}} \int dx dx' \rho_k(x) \rho_l(x') e^{\mathcal{W}_{\gamma,\beta}(k,x;l,x')} \right)$$

$$\times \left(\sum_{k=0}^{k_{\max}} \int_0^{2\pi} \frac{du}{2\pi} \int_{-\infty}^{\infty} Dx e^{iku+\sum_{r=1}^{Q} \alpha_r F_r(k)-\frac{i}{N} \hat{\rho}_k(x) e^{-iu}} \right)^N. \quad (36)$$

By rescaling the conjugate order parameters as $\hat{\rho}_k(x) \to iN\hat{\rho}_k(x)$ and integrating over u, we find a compact expression for $\mathcal{Z}_N$ when $N \gg 1$

$$\mathcal{Z}_N = \int \left(\prod_{k=0}^{k_{\max}} \mathcal{D}\rho_k \mathcal{D}\hat{\rho}_k \right) \exp \left(-N\mathcal{F}[\rho_k, \hat{\rho}_k] \right), \quad (37)$$

in which

$$\mathcal{F}[\rho_k, \hat{\rho}_k] = \sum_{k=0}^{k_{\max}} \int dx \rho_k(x) \hat{\rho}_k(x)$$

$$- \frac{1}{2} \sum_{k,l=0}^{k_{\max}} \int dx dx' \rho_k(x) \rho_l(x') e^{\mathcal{W}_{\gamma,\beta}(k,x;l,x')}$$

$$- \ln \left(\sum_{k=0}^{k_{\max}} \frac{1}{k!} \int_{-\infty}^{\infty} Dx [\hat{\rho}_k(x)]^k e^{\sum_{r=1}^{Q} \alpha_r F_r(k)} \right). \quad (38)$$

We have neglected the factor appearing in the integration measure of Eq. (37) due to the rescaling $\hat{\rho}_k(x) \to iN\hat{\rho}_k(x)$, since this factor yields a subleading contribution to the free-energy for large N. Equations (37) and (38) determine the leading contribution to $\ln \mathcal{Z}_N$ in the limit $N \to \infty$.

Since we introduced the finite cutoff $k_{\max}$ in the degree sequence, $\mathcal{F}[\rho_k, \hat{\rho}_k]$ is independent of N and, in the limit $N \to \infty$, the integral in Eq. (37) can be solved through the saddle-point method, according to which $\mathcal{Z}_N$ is dominated by the set of functions $\{\rho_k^*(x), \hat{\rho}_k^*(x)\}$ that minimize $\mathcal{F}[\rho_k, \hat{\rho}_k]$. The fact the degrees are bounded in the present model is a crucial difference with respect to Ref. [30], which ensures the application of the saddle-point method and the convergence of the partition function for $N \to \infty$. Thus, combining Eqs. (37) and (4), the free energy f is directly given by

$$f = \mathcal{F}[\rho_k^*, \hat{\rho}_k^*], \quad (39)$$

where $\{\rho_k^*(x), \hat{\rho}_k^*(x)\}$ fulfills the saddle-point equations

$$\hat{\rho}_k^*(x) = \sum_{l=0}^{k_{\max}} \int dx' \rho_l^*(x') e^{\mathcal{W}_{\gamma,\beta}(k,x;l,x')}, \quad (40)$$

$$\rho_k^*(x) = \frac{\frac{1}{k!} k [\hat{\rho}_k^*(x)]^{k-1} e^{-\frac{1}{2} x^2 + \sum_{r=1}^{Q} \alpha_r F_r(k)}}{\sum_{l=0}^{k_{\max}} \frac{1}{l!} \int_{-\infty}^{\infty} dx [\hat{\rho}_l^*(x)]^l e^{-\frac{1}{2} x^2 + \sum_{r=1}^{Q} \alpha_r F_r(l)}}, \quad (41)$$

with $k = 0, \dots, k_{\max}$. The solutions of the self-consistent Eqs. (40) and (41), together with the free energy, Eq. (39), fully characterize the stability and the structural properties

of infinitely large ERGs defined by Eq. (25). We remark that Eqs. (39), (40), and (41) are exact in the limit $N \to \infty$.

C. The equations for the structural observables

In the limit $N \to \infty$, the ensemble averages of the network observables, defined in section III, follow from the derivatives of f with respect to the model parameters. Let us illustrate this fact by deriving the analytic expression for the degree distribution. If we set $F_1(l) = \delta_{k,l}$ for arbitrary integers $0 \leqslant k, l \leqslant k_{\max}$, then the ensemble average degree distribution $\langle p_k \rangle$ is determined from

$$\langle p_k \rangle = -\frac{\partial f}{\partial \alpha_1}, \tag{42}$$

which follows from Eqs. (2) and (4). From the explicit form of f, Eqs. (38) and (39), we get

$$\langle p_k \rangle = \frac{\frac{1}{k!} \int_{-\infty}^{\infty} Dx [\hat{\rho}_k^*(x)]^k e^{\sum_{r=1}^{Q} \alpha_r F_r(k)}}{\sum_{l=0}^{k_{\max}} \frac{1}{l!} \int_{-\infty}^{\infty} Dx [\hat{\rho}_l^*(x)]^l e^{\sum_{r=1}^{Q} \alpha_r F_r(l)}}. \tag{43}$$

This is a common strategy to calculate ensemble averages in statistical mechanics, namely, one performs the derivative of the free energy with respect to a parameter that is coupled to a certain observable in the Hamiltonian. Note that $F_1(k)$ in Eq. (43) is not necessarily given by $F_1(k) = \delta_{k,l}$. In other words, after the choice $F_1(k) = \delta_{k,l}$ has served the purpose to obtain an expression for $\langle p_k \rangle$, we are free to choose $F_1(k)$ as we please.

Following an analogous procedure, the equations for the ensemble averages of all other observables introduced in Sec. III are obtained in the limit $N \to \infty$. The density of links and the density of two-stars read

$$\langle \ell \rangle = \frac{\langle K \rangle}{2}, \tag{44}$$

$$\langle s \rangle = \frac{1}{2} (\langle K^2 \rangle - \langle K \rangle), \tag{45}$$

where the moments $\langle K^n \rangle$ $(n = 1, 2, \dots)$ of $\langle p_k \rangle$ are determined from

$$\langle K^n \rangle = \sum_{k=0}^{k_{\max}} k^n \langle p_k \rangle. \tag{46}$$

The moments $\langle \Lambda_{qr} \rangle$ of the joint degree distribution at adjacent nodes read

$$\langle \Lambda_{qr} \rangle = \frac{1}{2\langle \ell \rangle} \sum_{kl=0}^{k_{\max}} k^q l^r \int dx dx' \rho_k^*(x) \rho_l^*(x') e^{\mathcal{W}_{\gamma,\beta}(k,x;l,x')}, \tag{47}$$

and the average degree correlation $\langle \chi \rangle$ at the end-points of two-stars is given by

$$\langle \chi \rangle = \frac{1}{2\langle s \rangle} \sum_{kl=0}^{k_{\max}} \int dx dx' \rho_k^*(x) \rho_l^*(x')$$
$$\times \left(\frac{1}{\sqrt{\beta}} xl - k^2 \right) e^{\mathcal{W}_{\gamma,\beta}(k,x;l,x')}. \tag{48}$$

Once we determine $\{\rho_k^*(x), \hat{\rho}_k^*(x)\}$ from the solutions of the saddle-point Eqs. (40) and (41), Eqs. (43)–(48) allow us to

compute the assortativities and characterize the network structure in the limit $N \to \infty$. In case Eqs. (40) and (41) have more than a single solution, the structural observables are calculated from the solution that corresponds to the global minimum of $\mathcal{F}[\rho_k, \hat{\rho}_k]$.

V. RESULTS

The exact equations derived in the previous section describe ERGs with the generic Hamiltonian of Eq. (25) in the limit $N \to \infty$. In this section we solve these equations and study the effect of degree correlations in the phase diagram of the two-star model.

A. The two-star model without degree correlations

In this section we present results for the two-star model in the absence of degree correlations [9–11,26,30], where the average density of edges and the average density of two-stars are the only constraints. The model undergoes a discontinuous transition as a function of the control parameters both in the dense regime [11,26] and in the more realistic case of sparse networks [30]. In the latter case, Ref. [30] reports a discontinuous behavior in the structural parameters, but the stability of the macroscopic states and the corresponding phase diagram remain elusive. We complement the work of Ref. [30] by constructing the full phase diagram of the two-star model in the sparse regime.

The Hamiltonian of the two-star model

$$\mathcal{H}(\boldsymbol{C}) = -\alpha_1 \sum_{i=1}^{N} K_i - \alpha_2 \sum_{i=1}^{N} K_i^2 + \ln N \sum_{i<j} C_{ij} \tag{49}$$

is recovered from Eq. (25) by setting $\beta = \gamma = 0$, $Q = 2$, and $F_r(k) = \delta_{r,1} k + \delta_{r,2} k^2$. In this case, $\hat{\rho}_k^*(x)$ becomes independent of k and x, and we set $\hat{\rho}_k^*(x) \equiv \mu$. The fixed-point equation for μ follows from Eq. (41)

$$\mu = \frac{\sum_{k=0}^{k_{\max}-1} \frac{1}{k!} \mu^k e^{\alpha_1(k+1)+\alpha_2(k+1)^2}}{\sum_{k=0}^{k_{\max}} \frac{1}{k!} \mu^k e^{\alpha_1 k + \alpha_2 k^2}}, \tag{50}$$

and the free energy assumes the form

$$f = \mathcal{F}(\nu)|_{\nu=\mu_*}, \tag{51}$$

where μ_* is the global minimum of the function $\mathcal{F}(\nu)$

$$\mathcal{F}(\nu) = \frac{1}{2} \nu^2 - \ln \left(\sum_{k=0}^{k_{\max}} \frac{1}{k!} \nu^k e^{\alpha_1 k + \alpha_2 k^2} \right). \tag{52}$$

The degree distribution is obtained from Eq. (43)

$$\langle p_k \rangle = \frac{\frac{1}{k!} \mu_*^k e^{\alpha_1 k + \alpha_2 k^2}}{\sum_{l=0}^{k_{\max}} \frac{1}{l!} \mu_*^l e^{\alpha_1 l + \alpha_2 l^2}}. \tag{53}$$

In the limit $k_{\max} \to \infty$, Eqs. (50) and (53) are equivalent to the main equations in Ref. [30]. However, in contrast to Ref. [30], the series in Eqs. (50)–(53) contain a finite number of terms. After μ_* is determined from the solutions of Eqs. (50) and (51), f and $\langle p_k \rangle$ can be evaluated for any α_1 and α_2.

Figure 1 shows the functional behavior of $\mathcal{F}(\nu)$. For the combinations of α_1 and α_2 shown in Fig. 1, $\mathcal{F}(\nu)$ exhibits two minima, which reflects the existence of a metastable region
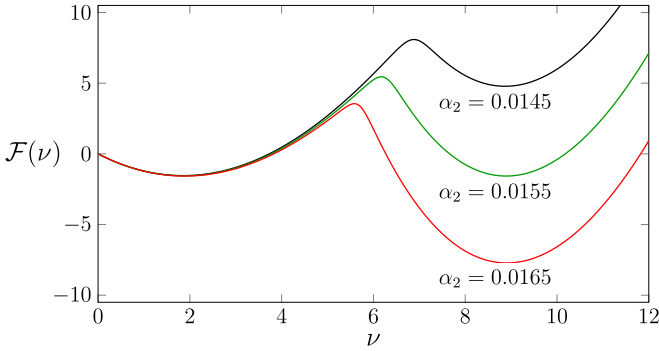
FIG. 1. The function $\mathcal{F}(\nu)$ [see Eqs. (37) and (38)] for the two-star model [see Eq. (49)] with maximum degree $k_{\max} = 80$, $\alpha_1 = 0.5$, and different α_2. The values of ν that minimize $\mathcal{F}(\nu)$ are solutions of Eq. (50). The global minimum determines the free energy and the structural parameters.

in the phase diagram. Each minimum corresponds to a stable fixed-point solution of Eq. (50) and, consequently, to a certain macroscopic state of the two-star model. The global minimum yields the leading contribution to the partition function for $N \rightarrow \infty$, from which one determines the graph structural properties. The values of (α_1, α_2) along which the depths of the minima become equal identify a first-order critical line. For fixed α_1, $\mathcal{F}(\nu)$ has a single minimum if α_2 is either sufficiently large or small. This particular situation is not shown in Fig. 1.

The free energy allows us to characterize the stability of the different phases and construct the phase diagram in the plane (α_1, α_2). The phase diagram for $k_{\max} = 80$, shown in Fig. 2(a), exhibits a metastable region enclosing a first-order critical line, which terminates at a critical point. The inset in Fig. 2(a) shows the continuous phase transition of μ along the first-order critical line. Figures 2(b) and 2(c) illustrate the typical profile of the degree distribution $\langle p_k \rangle$ in each phase. Clearly, the first-order transition corresponds to an abrupt condensation of $\langle p_k \rangle$ onto the maximum degree $k = k_{\max}$. Below the first-order critical line, $\langle p_k \rangle$ is closer to a Poisson distribution, whereas above the critical line $\langle p_k \rangle$ has a peak at $k = k_{\max}$ and the graph samples are approximately regular.

The behavior of the structural properties across the phase transition is a subject of practical interest. Figure 3 shows the density of links and the density of two-stars as a function of α_2 for different α_1. Both quantities are discontinuous at the first-order transition. The discontinuity becomes gradually smaller as we increase α_1, until it finally disappears at the critical point, i.e., $\langle \ell \rangle$ and $\langle s \rangle$ are continuous and monotonic functions of α_2 provided $\alpha_1 \gtrsim 1.42$. Since $\beta = \gamma = 0$, both assortativities are zero in this model. The theoretical results of Fig. 3 are well corroborated by data obtained from Monte Carlo simulations that sample graphs from the distribution of Eq. (2). Monte Carlo methods to generate ERGs are thoroughly discussed in Ref. [9].

B. Degree correlations between nearest neighbors

In this subsection we analyze the role of nearest neighbor degree correlations on the phase diagram of the two-star model. We consider an ERG model that allows us to tune
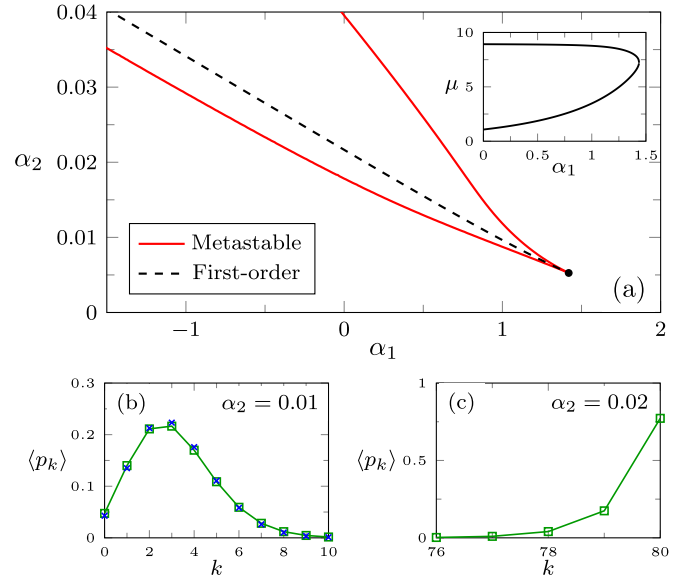

FIG. 2. (a) Phase diagram of the two-star model with maximum degree $k_{\max} = 80$ [see Eq. (49)]. The dashed black curve is the first-order critical line and the solid red curves delimit the metastable region, within which the free energy has two minima. The inset shows the two stable solutions of Eq. (50) as a function of α_1 along the dashed curve. The two solutions for μ merge continuously at a critical point, identified by the black dot in (a). (b) and (c) show the degree distribution (green squares) for fixed $\alpha_1 = 0.5$, $k_{\max} = 80$, and a value of α_2 inside each phase. The blue crosses in (b) denote a Poisson distribution with the same mean degree.

the density of links, the density of two-stars, and the degree correlations between adjacent nodes. The model is defined by the Hamiltonian

$$\mathcal{H}(\boldsymbol{C}) = -\alpha_1 \sum_{i=1}^{N} K_i - \alpha_2 \sum_{i=1}^{N} K_i^2$$
$$- \gamma \sum_{i<j}^{N} C_{ij} K_i K_j + \ln N \sum_{i<j} C_{ij}, \qquad (54)$$

which is recovered from Eq. (25) by setting $\beta = 0$, $Q = 2$, $F_r(k) = \delta_{r,1}k + \delta_{r,2}k^2$, and $D(k, l) = kl$. Most of the results
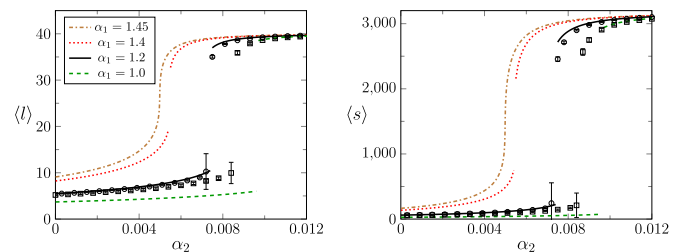

FIG. 3. Theoretical results (different line styles) for the average density of edges $\langle \ell \rangle$ and the average density of two-stars $\langle s \rangle$ as a function of α_2 for the two-star model [see Eq. (49)] with maximum degree $k_{\max} = 80$ and different α_1. The symbols are obtained from the average over 10^5 graph samples generated through Monte Carlo simulations with $\alpha_1 = 1.2$ and two system sizes: $N = 200$ (squares) and $N = 800$ (circles).
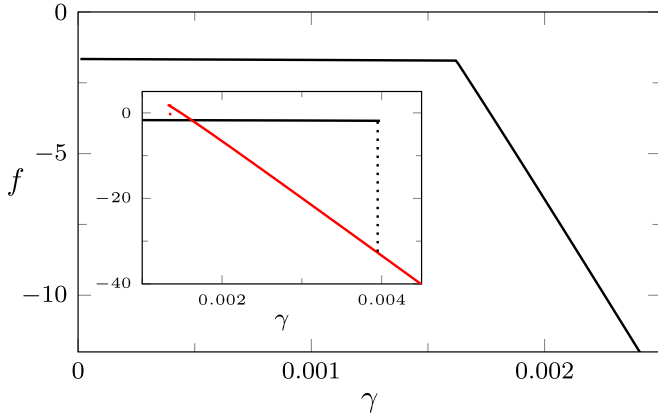
FIG. 4. The main panel shows the free energy as a function of γ for an exponential random graph model with nearest neighbor degree correlations [see Eq. (54)], maximum degree $k_{\max} = 30$, $\alpha_1 = 0.6$, and $\alpha_2 = 0$. The nonanalytic point of the free energy marks the first-order critical point. The inset displays the two branches of the function f calculated from Eq. (58). The red solid line and the black solid line in the inset represent the function f corresponding to each one of the two stable solutions of Eq. (57). The metastable region, where Eq. (57) has two stable solutions, is delimited by the dotted vertical lines.

in this section examine the effect of nearest neighbor degree correlations when $\alpha_2 = 0$, since the results for $\alpha_2 \neq 0$ are qualitatively similar. Equation (54) for $\alpha_2 = 0$ is the Hamiltonian of the Erdös-Rényi model [1] with degree correlations between adjacent nodes. We show below that degree correlations induce a first-order condensation transition in the simple case of Erdös-Rényi random graphs.

The function $\hat{\rho}_k^*(x)$ is independent of x for $\beta = 0$. By writing

$$\int dx \rho_k^*(x) \equiv \rho_k^*, \quad \hat{\rho}_k^*(x) \equiv \hat{\rho}_k^*, \tag{55}$$

the quantities $\{\rho_k^*, \hat{\rho}_k^*\}_{k=0,\ldots,k_{\max}}$ solve

$$\hat{\rho}_k^* = \sum_{l=0}^{k_{\max}} \rho_l^* e^{\gamma k l}, \tag{56}$$

$$\rho_k^* = \frac{\frac{1}{k!} k (\hat{\rho}_k^*)^{k-1} e^{\alpha_1 k + \alpha_2 k^2}}{\sum_{l=0}^{k_{\max}} \frac{1}{l!} (\hat{\rho}_l^*)^l e^{\alpha_1 l + \alpha_2 l^2}}, \tag{57}$$

while the free energy follows from

$$f = \frac{1}{2} \sum_{k=0}^{k_{\max}} \rho_k^* \hat{\rho}_k^* - \ln \left(\sum_{k=0}^{k_{\max}} \frac{1}{k!} (\hat{\rho}_k^*)^k e^{\alpha_1 k + \alpha_2 k^2} \right). \tag{58}$$

Equation (57) represents a system of $k_{\max} + 1$ coupled fixed-point equations that can be solved by iteration.

The free energy is obtained from the global minimum of the $k_{\max} + 1$-dimensional surface $\mathcal{F}(\rho_0, \ldots, \rho_{k_{\max}})$ [see Eq. (38)], which follows from the solutions of Eq. (57). Figure 4 depicts f as a function of γ for $\alpha_2 = 0$ and fixed α_1. The free energy exhibits a nonanalytic point, marking a first-order transition, at which the derivative of f with respect to γ is discontinuous. The inset in Fig. 4 shows the two branches of f,
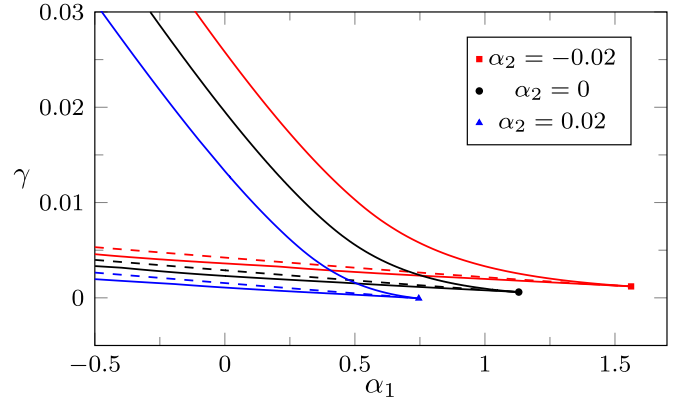

FIG. 5. Phase diagram of the two-star model with nearest neighbor correlated degrees [see Eq. (54)], maximum degree $k_{\max} = 30$, and different values of α_2. The dashed curve for each α_2 marks the first-order critical line and the solid curves delimit the metastable region, within which the free energy has two minima. For each value of α_2, the dashed curve and the two solid curves terminate at the critical point identified by the corresponding symbol. The critical points are approximately given by $(\alpha_1, \gamma) = (1.56, 1.2 \times 10^{-3})$ for $\alpha_2 = -0.02$, $(\alpha_1, \gamma) = (1.13, 6.1 \times 10^{-4})$ for $\alpha_2 = 0$, and $(\alpha_1, \gamma) = (0.746, -6 \times 10^{-5})$ for $\alpha_2 = 0.02$.

each one corresponding to a minimum of $\mathcal{F}$ or a stable fixed-point of Eq. (57). In the metastable region, $\mathcal{F}(\rho_0, \ldots, \rho_{k_{\max}})$ has two minima, one of them is local (metastable), while the other is global (stable). The structural properties of the graph are evaluated from the solution $\{\rho_k^*\}_{k=0}^{k_{\max}}$ at the global minimum of $\mathcal{F}$.

The phase diagram (α_1, γ) of the ERG model defined by Eq. (54), and obtained from the analysis of f, is shown in Fig. 5. For each value of α_2, the phase diagram has a first-order critical line, surrounded by a metastable region, which ends at a critical point. The first-order critical line marks anew the condensation transition, above which the degree distribution $\langle p_k \rangle$ has a peak at $k = k_{\max}$. The profile of $\langle p_k \rangle$ below and above each dashed line in Fig. 5 is qualitatively similar to, respectively, Figs. 2(b) and 2(c). For $\gamma > 0$, adjacent nodes tend to have similar degrees, which strongly favors the formation of a regular random graph, driving the ERG model to the condensed phase even in the case of $\alpha_2 = 0$.

The nearest neighbor assortativity $\langle A^{(1)} \rangle$ is discontinuous at the first-order transition. Figure 6 shows $\langle A^{(1)} \rangle$ as a function of γ for different α_1 and $\alpha_2 = 0$. The discontinuity of $\langle A^{(1)} \rangle$ becomes smaller as α_1 increases, until it vanishes continuously at the critical point. In fact, the two solutions for $\langle A^{(1)} \rangle$ corresponding to each phase merge into a single value as we approach the critical point along the corresponding dashed curve in Fig. 5. The assortativity $\langle A^{(1)} \rangle$ is a nonmonotonic function of γ that vanishes when $|\gamma| \to \infty$. The latter property can be understood from the ground state configurations of the Hamiltonian. For $\gamma \to -\infty$, the minimum of Eq. (54) is attained when all degrees are zero; for $\gamma \to \infty$, Eq. (54) is minimized for all degrees equal to $k_{\max}$. Since $\beta = 0$, nodes at the end-points of two-stars are uncorrelated and $\langle A^{(2)} \rangle = 0$. Figure 6 also shows results generated through Monte Carlo simulations of finite random graphs, which confirms our theoretical findings for $N \to \infty$.

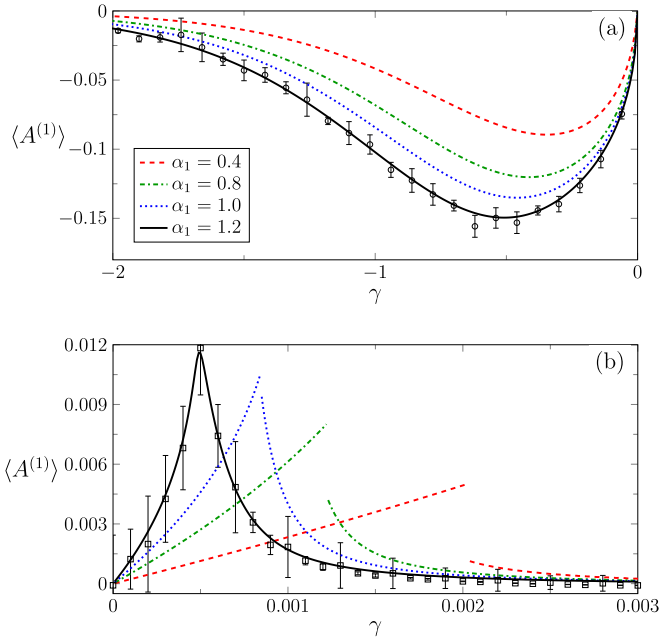FIG. 6. Theoretical results (different line styles) for the degree assortativity $\langle A^{(1)} \rangle$ of adjacent nodes as a function of γ for an exponential random graph model with nearest neighbor correlated degrees [see Eq. (54)], maximum degree $k_{\max} = 30$, and $\alpha_2 = 0$. The symbols are obtained from the average over 10^4 graph samples generated through Monte Carlo simulations with $\alpha_2 = 0$, $\alpha_1 = 1.2$, $N = 1500$ (circles), and $N = 5000$ (squares).

C. Degree correlations between next nearest neighbors

In this subsection we focus on the competition between nearest neighbor degree correlations and next nearest neighbor degree correlations. We consider the Hamiltonian

$$\mathcal{H}(\boldsymbol{C}) = \ln N \sum_{i<j} C_{ij} - \gamma \sum_{i<j} C_{ij} K_i K_j$$
$$- \frac{\beta}{2} \sum_{ijk=1}^{N} (1 - \delta_{ik}) C_{ij} C_{jk} K_i K_k, \qquad (59)$$

that results from Eq. (25) by setting $\alpha_r = 0$ for $r = 1, \ldots, Q$. We will not present explicit results for nonzero values of α_1 and α_2 [see Eq. (49)], since changing the density of links or two-stars does not modify the overall qualitative picture discussed below.

In comparison to $\beta = 0$, it is far more challenging to solve the saddle-point Eqs. (40) and (41) and determine the functions $\rho_0^*(x), \ldots, \rho_{k_{\max}}^*(x)$ for $\beta \neq 0$, due to the exponential or oscillatory behavior of the integrands. We calculate numerically the integrals in Eqs. (40) and (41) by discretizing $\rho_0^*(x), \ldots, \rho_{k_{\max}}^*(x)$ over $x \in \mathbb{R}$. The discretized version of the saddle-point equations is iterated until each function $\rho_k^*(x)$ converges to a stationary functional form. In the Appendix, we discuss some useful symmetry properties of $\rho_0^*(x), \ldots, \rho_{k_{\max}}^*(x)$ for $\beta < 0$. We set $k_{\max} = 10$ throughout this subsection.

First, we consider the ERG model of Eq. (59) for $\beta > 0$. In this regime the order-parameters $\rho_0^*(x), \ldots, \rho_{k_{\max}}^*(x)$ are real-
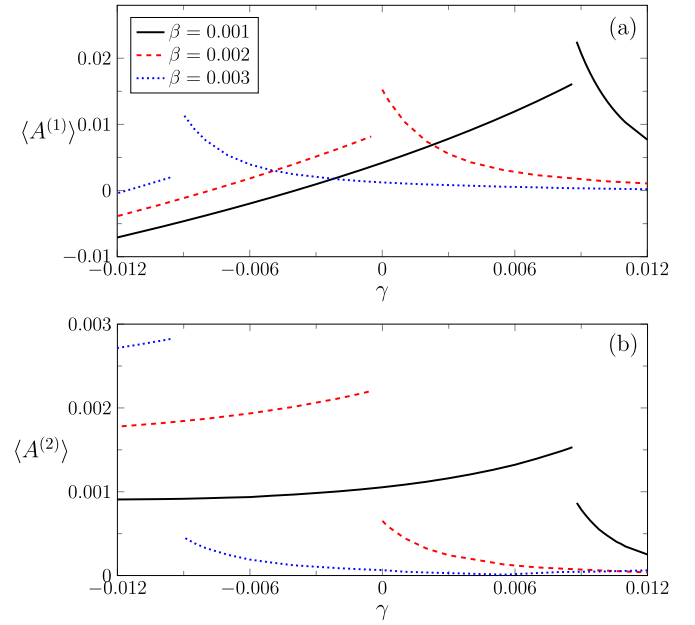

FIG. 7. Degree assortativities $\langle A^{(1)} \rangle$ and $\langle A^{(2)} \rangle$ corresponding, respectively, to nearest neighbor nodes and next nearest neighbors nodes of an exponential random graph model with degree-degree correlations [see Eq. (59)], $\beta > 0$, and maximum degree $k_{\max} = 10$.

valued functions. Figure 7 shows that the degree assortativities $\langle A^{(1)} \rangle$ and $\langle A^{(2)} \rangle$ have a discontinuous behavior as a function of γ, and the model undergoes once more a first-order transition at $\gamma = \gamma_c(\beta)$. For $\gamma < \gamma_c(\beta)$, $\langle p_k \rangle$ is approximately given by a Poisson distribution, while for $\gamma > \gamma_c(\beta)$ the degree distribution has a single peak at $k = k_{\max}$. Figure 8 illustrates the behavior of $\langle p_k \rangle$ and $\rho_0^*(x), \ldots, \rho_{k_{\max}}^*(x)$ inside the condensed phase. For $\gamma \to \infty$ or $\beta \to \infty$, the degree distribution converges to $\langle p_k \rangle = \delta_{k,k_{\max}}$ and both assortativities are zero. Overall, Figs. 7 and 8 show that positive next nearest neighbor degree correlations do not change the phase diagram qualitatively.

Let us now present results for $\beta < 0$, where $\rho_0^*(x), \ldots, \rho_{k_{\max}}^*(x)$ are complex-valued functions of $x \in \mathbb{R}$. In the Appendix, we demonstrate that $\mathrm{Re}\rho_k^*(x)$ is an even function and $\mathrm{Im}\rho_k^*(x)$ is an odd function in the regime $\beta < 0$. It is interesting to note that, for $\beta < 0$, we impose conflicting
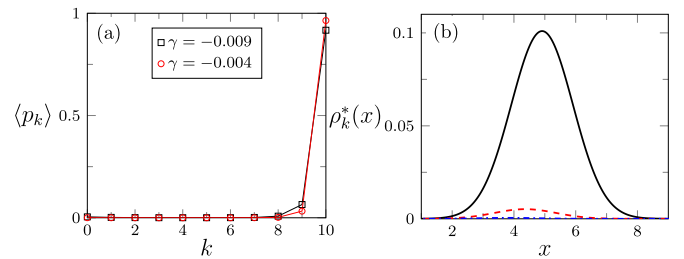

FIG. 8. Degree distribution $\langle p_k \rangle$ and the order-parameter functions $\{\rho_k^*(x)\}$ of an exponential random graph model with degree-degree correlations [see Eq. (59)], $\beta = 0.003$, maximum degree $k_{\max} = 10$, and different values of γ in the condensed phase. (b) shows the functional behavior of $\rho_{10}^*(x)$ (black solid line), $\rho_9^*(x)$ (red dashed line), and $\rho_8^*(x)$ (blue dot-dashed line) for $\gamma = -0.004$. The other components $\rho_0^*(x), \ldots, \rho_7^*(x)$ are approximately zero.
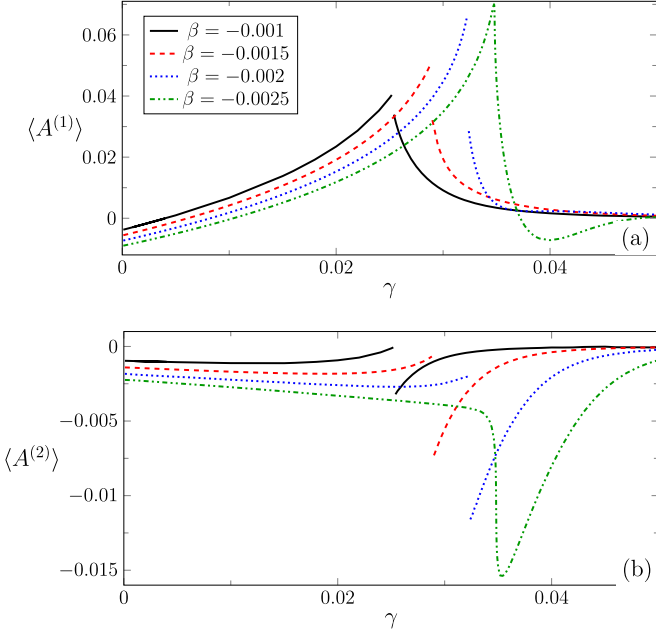
FIG. 9. Degree assortativities $\langle A^{(1)} \rangle$ and $\langle A^{(2)} \rangle$ corresponding, respectively, to nearest neighbor nodes and next nearest neighbors nodes of an exponential random graph model with degree-degree correlations [see Eq. (59)], $\beta < 0$, and maximum degree $k_{max} = 10$.

constraints in the generation of graph samples, since negative values of β favor dissimilar degrees at the end-points of two-stars, leaving the degrees at the central nodes of two-stars in a frustrating situation. The appearance of these frustrated configurations should influence the graph structure.

In Fig. 9 we present $\langle A^{(1)} \rangle$ and $\langle A^{(2)} \rangle$ as a function of γ for $\beta < 0$. The ERG model undergoes a first-order transition at the critical point $\gamma = \gamma_c(\beta)$. For $\gamma < \gamma_c(\beta)$, the degree correlations do not considerably affect $\langle p_k \rangle$, which is closer to a Poisson distribution, similar to Fig. 2(b). For $\gamma > \gamma_c(\beta)$, negative degree correlations between next nearest neighbors have an important effect in the graph structure and $\langle p_k \rangle$ can exhibit a bimodal shape, as illustrated in Fig. 10. For increas-
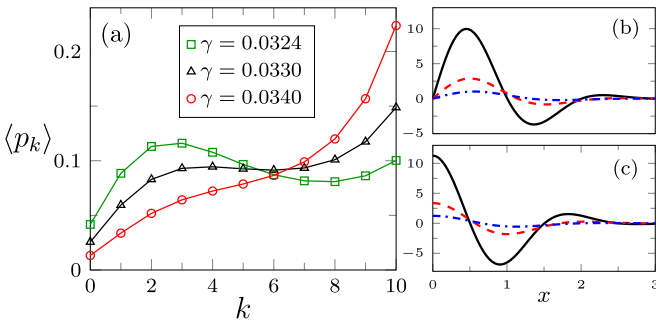

FIG. 10. Degree distribution $\langle p_k \rangle$ and the order-parameter functions $\{\rho_k^*(x)\}$ of an exponential random graph model with degree-degree correlations [see Eq. (59)], $\beta = -0.002$, maximum degree $k_{max} = 10$, and different values of γ above the first-order transition. (b) and (c) show, respectively, the functional behavior of $\mathrm{Im}\rho_k^*(x)$ and $\mathrm{Re}\rho_k^*(x)$ for $\gamma = 0.0324$ and three values of k: $k = 10$ (black solid lines), $k = 9$ (red dashed lines), and $k = 8$ (blue dot-dashed lines). The other components $\rho_0^*(x), \ldots, \rho_7^*(x)$ are approximately zero.
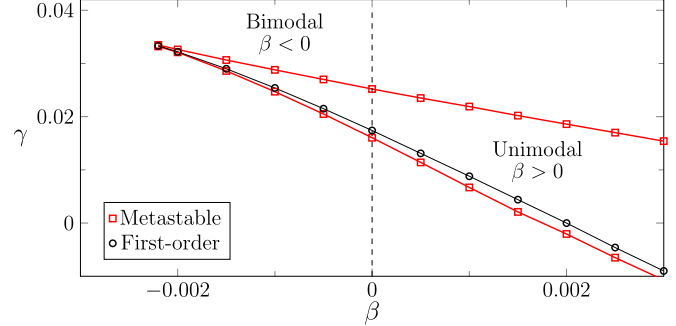

FIG. 11. Phase diagram of an exponential random graph model with degree-degree correlations [see Eq. (59)] and maximum degree $k_{max} = 10$. The symbols are theoretical results obtained from the solutions of Eqs. (39)–(41), and the solid lines are just a guide to the eye. The phase diagram has a first-order critical line surrounded by a metastable region, where the free energy has two minima. Above the first-order critical line, the degree distribution is bimodal for $\beta < 0$, and it has a single peak at the largest degree for $\beta > 0$. The degree distribution approaches $\langle p_k \rangle = \delta_{k,k_{max}}$ as $\gamma \to \infty$.

ing $\gamma > \gamma_c(\beta)$, the weight $\langle p_{k_{max}} \rangle$ gradually increases until we attain $\langle p_k \rangle = \delta_{k,k_{max}}$ for $\gamma \to \infty$. Nevertheless, the graph structure in the condensed phase for $\beta < 0$ is qualitatively distinct from the regime $\beta > 0$, which is also attested by the functional behavior of the order parameters, presented in Figs. 10(b) and 10(c) [as a comparison, see Fig. 8(b)]. As shown by Fig. 9, the first-order transition disappears for β smaller than a certain threshold, which marks the terminating point of the first-order critical line in the plane (β, γ).

The competition between nearest neighbor and next nearest neighbor degree correlations is summarized in Fig. 11, which depicts the phase diagram of the model defined by Eq. (59). The phase diagram exhibits a metastable region around a first-order critical line $\gamma_c(\beta)$ that terminates at a negative value of β. For $\gamma < \gamma_c(\beta)$, $\langle p_k \rangle$ is closer to a Poisson distribution. For $\gamma > \gamma_c(\beta)$ and $\beta > 0$, we have $\langle A^{(2)} \rangle > 0$ and the degree distribution has a single peak at $k = k_{max}$. For $\gamma > \gamma_c(\beta)$ and $\beta < 0$, we have $\langle A^{(2)} \rangle < 0$ and $\langle p_k \rangle$ can exhibit two peaks, one of them located at $k = k_{max}$, and an additional peak at a smaller degree.

VI. FINAL REMARKS

In this paper we have solved the two-star model with degree-degree correlations in the sparse regime. The model allows us to generate random graphs with prescribed degree correlations between adjacent nodes and between nodes at the end-points of two-stars. By introducing an upper cutoff in the degree sequence, we have exactly calculated the network free energy, from which we derived complete phase diagrams and characterized the graph structure in the different phases.

In terms of the degree distribution $\langle p_k \rangle$, the phase diagram of the model is characterized by three distinct regions. There is a phase where $\langle p_k \rangle$ is approximately given by a Poisson distribution, reminiscent from the structure of Erdös-Rényi random graphs [1]. The phase diagram also exhibits a condensed phase, where the shape of $\langle p_k \rangle$ strongly depends on the degree correlations. If the degree assortativities are non-negative in-

side the condensed phase, then $\langle p_k \rangle$ has a single peak at the maximum degree and the graph is approximately regular. If the degree assortativity of next nearest neighbors is negative inside the condensed phase, then $\langle p_k \rangle$ is given by a bimodal distribution, with one maximum at the largest degree and an additional maximum at a smaller degree. While the Poisson and standard condensed phases appear even in the absence of degree-degree correlations, the existence of a bimodal degree distribution is a genuine effect of negative degree correlations between next nearest neighbor nodes. This result reveals the importance of long-range degree correlations, beyond nearest neighbor nodes, in shaping the network structure.

We have shown that the model undergoes a first-order transition between the Poisson phase and the condensed phase. The first-order critical line is surrounded by a metastable region, where the free energy has two minima, each one corresponding to a phase. For combinations of model parameters inside the metastable region, algorithms to sample finite graphs from this ERG model may get stuck in a local minimum of the free energy [9]. In addition, the jump of the structural observables across the first-order critical line prevents us from generating graph samples with structural parameters in a certain range. Taken together, these features represent serious limitations of the present ERG model as an effective tool to model real-world networks. The analytic solution of the model for $N \to \infty$ and the construction of its phase diagrams have practical relevance, as these results allow us to estimate the metastable regions in the parameter space of finite graphs.

The present paper constitutes a first step towards controlling the generation of ERGs with correlated degrees. Overall, the results for the assortativities, the degree distribution, and the phase diagrams allow us to identify the regime of parameters where the model can be useful to reproduce certain properties of empirical networks. A drawback of the present model is that the degree distribution in each phase does not bear any resemblance to the broad degree distributions found in real-world networks. With the purpose of improving the model, it would be interesting to solve it with a hard constraint in the degree sequence or with a prescribed degree distribution [10]. This work also opens the perspective to explore systematically the role of short-range and long-range degree correlations in dynamical processes occurring on tree-like networks, since in this case the equations for the dynamics are typically determined only by the degree distribution [50]. Finally, we point out that the free energy of ERG models can be mapped in the cumulant generating function of certain structural observables of Erdös-Rényi random graphs [51–53]. Therefore, the results of the present paper can be readily applied to study analytically the large deviations of higher-order topological properties of Erdös-Rényi random graphs in the limit $N \to \infty$ [54].

APPENDIX: SYMMETRY PROPERTIES OF THE ORDER-PARAMETER FUNCTIONS FOR $\beta < 0$

In this Appendix we obtain the symmetry properties of the order parameters $\rho_0^*(x), \ldots, \rho_{k_{\max}}^*(x)$ under the transformation $x \to -x$. These properties allow us to simplify the saddle-point Eqs. (40) and (41) and the computation of the structural observables introduced in Sec. III.

For $\beta < 0$, $\{\rho_k^*(x), \hat{\rho}_k^*(x)\}$ are complex-valued functions of $x \in \mathbb{R}$. This can be seen from Eq. (40), which can be written for $\beta < 0$ as

$$\hat{\rho}_k^*(x) = \sum_{l=0}^{k_{\max}} U_{kl} e^{i\sqrt{|\beta|}xl} \int_{-\infty}^{\infty} dx' \rho_l^*(x') e^{i\sqrt{|\beta|}x'k}, \quad \text{(A1)}$$

where

$$U_{kl} = e^{\gamma D(k,l) + \frac{|\beta|}{2}(k^2 + l^2)}. \quad \text{(A2)}$$

If we take the complex-conjugate $\overline{(\ldots)}$ of Eq. (41) and make the transformation $x \to -x$, then we get

$$\overline{\rho_k^*}(-x) = \frac{1}{\overline{\mathcal{T}}} \frac{k}{k!} [\overline{\hat{\rho}_k^*}(-x)]^{k-1} e^{-\frac{1}{2}x^2 + \sum_{r=1}^{Q} \alpha_r F_r(k)}, \quad \text{(A3)}$$

with $\overline{\mathcal{T}}$

$$\overline{\mathcal{T}} = \sum_{k=0}^{k_{\max}} \frac{1}{k!} \int_{-\infty}^{\infty} dx [\overline{\hat{\rho}_k^*}(-x)]^k e^{-\frac{1}{2}x^2 + \sum_{r=1}^{Q} \alpha_r F_r(k)}. \quad \text{(A4)}$$

By taking the complex-conjugate of Eq. (A1), the function $\overline{\hat{\rho}_k^*}(-x)$ fulfills

$$\overline{\hat{\rho}_k^*}(-x) = \sum_{l=0}^{k_{\max}} U_{kl} e^{i\sqrt{|\beta|}xl} \int_{-\infty}^{\infty} dx' \overline{\rho_l^*}(-x') e^{i\sqrt{|\beta|}x'k}. \quad \text{(A5)}$$

Since Eqs. (40) and (41) for $\{\rho_k^*(x), \hat{\rho}_k^*(x)\}$ are the same as Eqs. (A3) and (A5) for $\{\overline{\rho_k^*}(-x), \overline{\hat{\rho}_k^*}(-x)\}$, we conclude that

$$\rho_k^*(x) = \overline{\rho_k^*}(-x) \quad \text{(A6)}$$

and

$$\hat{\rho}_k^*(x) = \overline{\hat{\rho}_k^*}(-x) \quad \text{(A7)}$$

for arbitrary x. This implies that $\{\text{Re}\rho_k^*(x), \text{Re}\hat{\rho}_k^*(x)\}$ and $\{\text{Im}\rho_k^*(x), \text{Im}\hat{\rho}_k^*(x)\}$ are, respectively, even and odd functions of x.

Let us use the above symmetry properties to simplify the order-parameter equations. By setting $\hat{\rho}_k^*(x)$ in polar form

$$\hat{\rho}_k^*(x) = r_k(x) e^{i\varphi_k(x)}, \quad \text{(A8)}$$

with $\varphi_k(x) \in (-\pi, \pi]$, and noting that

$$r_k(-x) = r_k(x) \quad \text{(A9)}$$

and

$$\varphi_k(-x) = -\varphi_k(x), \quad \text{(A10)}$$

we readily obtain $\text{Im}\mathcal{T} = 0$ from Eq. (A4). Hence Eq. (41) can be written as

$$\text{Re}\rho_k^*(x) = \frac{1}{\text{Re}\mathcal{T}} \frac{k}{k!} [r_k(x)]^{k-1} \cos[(k-1)\varphi_k(x)]$$
$$\times e^{-\frac{1}{2}x^2 + \sum_{r=1}^{Q} \alpha_r F_r(k)}, \quad \text{(A11)}$$

$$\mathrm{Im}\rho_k^*(x) = \frac{1}{\mathrm{Re}\mathcal{T}}\frac{k}{k!}[r_k(x)]^{k-1}\sin\left[(k-1)\varphi_k(x)\right]$$
$$\times e^{-\frac{1}{2}x^2 + \sum_{r=1}^{Q}\alpha_r F_r(k)}. \quad \text{(A12)}$$

The above equations are coupled to Eq. (40), which can be simplified for $\beta < 0$ using Eqs. (A6) and (A7)

$$\mathrm{Re}\hat{\rho}_k^*(x) = 2\sum_{l=0}^{k_{\max}}U_{kl}\cos\left(\sqrt{|\beta|}xl\right)\int_0^\infty dx'Y_{kl}(x'), \quad \text{(A13)}$$

$$\mathrm{Im}\hat{\rho}_k^*(x) = 2\sum_{l=0}^{k_{\max}}U_{kl}\sin\left(\sqrt{|\beta|}xl\right)\int_0^\infty dx'Y_{kl}(x'), \quad \text{(A14)}$$

where

$$Y_{kl}(x) = \mathrm{Re}\rho_l^*(x')\cos\left(\sqrt{|\beta|}x'k\right)$$
$$- \mathrm{Im}\rho_l^*(x')\sin\left(\sqrt{|\beta|}x'k\right). \quad \text{(A15)}$$

The fixed-point functions $\{\rho_k^*(x)\}$ that solve Eqs. (A11)–(A14) determine the free energy f and all the structural parameters for $\beta < 0$. Using the symmetry properties of Eqs. (A6) and (A7), it is straightforward to verify from Eq. (38) that $f \in \mathbb{R}$ for $\beta < 0$.

[1] M. E. J. Newman, *Networks: An Introduction* (OUP, Oxford, 2010).

[2] S. N. Dorogovtsev and J. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (OUP, Oxford, 2013).

[3] P. V. Mieghem, Europhys. Lett. **97**, 48004 (2012).

[4] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, Phys. Rev. Lett. **109**, 128702 (2012).

[5] J. Grilli, T. Rogers, and S. Allesina, Nat. Commun. **7**, 12031 (2016).

[6] I. Neri and F. L. Metz, Phys. Rev. Res. **2**, 033313 (2020).

[7] N. Brunel, Comput. Neurosci. **8**, 183 (2000).

[8] S. Ostojic, Nat. Neurosci. **17**, 594 (2014).

[9] A. C. C. Coolen, A. Annibale, and E. Roberts, *Generating Random Networks and Graphs* (Oxford University Press, Oxford, 2017).

[10] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, Nat. Rev. Phys. **1**, 58 (2019).

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

[12] P. W. Holland and S. Leinhardt, J. Am. Stat. Assoc. **76**, 33 (1981).

[13] O. Frank and D. Strauss, J. Am. Stat. Assoc. **81**, 832 (1986).

[14] D. Strauss, SIAM Rev. **28**, 513 (1986).

[15] S. Wasserman and P. Pattison, Psychometrika **61**, 401 (1996).

[16] C. J. Anderson, S. Wasserman, and B. Crouch, Soc. Networks **21**, 37 (1999).

[17] T. A. Snijders, Ann. Rev. Sociol. **37**, 131 (2011).

[18] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).

[19] S. Maslov, K. Sneppen, and A. Zaliznyak, Physica A **333**, 529 (2004).

[20] T. Squartini and D. Garlaschelli, New J. Phys. **13**, 083001 (2011).

[21] P. Akara-Pipattanaa, T. Chotibuta, and O. Evnin, arXiv:2102.11477 [cond-mat.dis-nn].

[22] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Phys. Rev. E **69**, 046117 (2004).

[23] D. Jeong, M. Y. Choi, and H. Park, J. Phys. A **40**, 9723 (2007).

[24] A. Gorsky and O. Valba, J. Complex Networks **8** (2020).

[25] J. Berg and M. Lässig, Phys. Rev. Lett. **89**, 228701 (2002).

[26] J. Park and M. E. J. Newman, Phys. Rev. E **70**, 066146 (2004).

[27] Z. Burda, J. Jurkiewicz, and A. Krzywicki, Phys. Rev. E **69**, 026106 (2004).

[28] Z. Burda, J. Jurkiewicz, and A. Krzywicki, Phys. Rev. E **70**, 026106 (2004).

[29] A. Annibale, A. C. C. Coolen, L. P. Fernandes, F. Fraternali, and J. Kleinjung, J. Phys. A **42**, 485001 (2009).

[30] A. Annibale and O. T. Courtney, J. Phys. A **48**, 365001 (2015).

[31] A. C. C. Coolen, J. Phys.: Conf. Ser. **699**, 012022 (2016).

[32] F. A. López and A. C. C. Coolen, J. Phys. A **53**, 065002 (2020).

[33] F. A. López and A. C. C. Coolen, J. Phys. Complex. **2**, 035010 (2021).

[34] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[35] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).

[36] M. Mayo, A. Abdelzaher, and P. Ghosh, Comput. Soc. Networks **2**, 4 (2015).

[37] Y. Fujiki, T. Takaguchi, and K. Yakubo, Phys. Rev. E **97**, 062308 (2018).

[38] V. M. Eguíluz and K. Klemm, Phys. Rev. Lett. **89**, 108701 (2002).

[39] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).

[40] F. Sorrentino, M. di Bernardo, G. H. Cuéllar, and S. Boccaletti, Physica D **224**, 123 (2006).

[41] I. Sendiña-Nadal, I. Leyva, A. Navas, J. A. Villacorta-Atienza, J. A. Almendral, Z. Wang, and S. Boccaletti, Phys. Rev. E **91**, 032811 (2015).

[42] S. Jalan, A. Kumar, A. Zaikin, and J. Kurths, Phys. Rev. E **94**, 062202 (2016).

[43] C. Orsini, M. M. Dankulov, P. Colomer-de Simón, A. Jamakovic, P. Mahadevan, A. Vahdat, K. E. Bassler, Z. Toroczkai, M. Boguñá, G. Caldarelli, S. Fortunato, and D. Krioukov, Nat. Commun. **6**, 8627 (2015).

[44] Y. Fujiki, S. Mizutaka, and K. Yakubo, Eur. Phys. J. B **90**, 126 (2017).

[45] D. Rybski, H. D. Rozenfeld, and J. P. Kropp, Europhys. Lett. **90**, 28002 (2010).

[46] A. Arcagni, R. Grassi, S. Stefani, and A. Torriero, Eur. J. Oper. Res. **262**, 708 (2017).

[47] S. Mizutaka and T. Hasegawa, J. Phys.: Complexity **1**, 035007 (2020).

[48] Y. Fujiki and K. Yakubo, Phys. Rev. E **101**, 032308 (2020).

[49] J. Negele and H. Orland, *Quantum Many-Particle Systems*, Advanced Book Classics (Basic Books, New York, 1988).

[50] E. Guzmán-González, I. P. Castillo, and F. L. Metz, Phys. Rev. E **101**, 012133 (2020).

[51] F. L. Metz and I. Pérez Castillo, Phys. Rev. Lett. **117**, 104101 (2016).

[52] F. L. Metz and I. P. Castillo, Phys. Rev. E **100**, 012305 (2019).

[53] C. Giardinà, C. Giberti, and E. Magnanini, J. Stat. Phys. **182**, 23 (2021).

[54] H. Chen, F. Huang, G. Li, and H. Zhang, arXiv:1907.13330 [cond-mat.stat-mech].