Solution of the two-star model of a network

Juyong Park and M. E. J. Newman
Department of Physics, University of Michigan, Ann Arbor, Michigan 48109-1120, USA
(Received 25 May 2004; published 29 December 2004)

The p-star model or exponential random graph is among the oldest and best known of network models. Here we give an analytic solution for the particular case of the two-star model, which is one of the most fundamental of exponential random graphs. We derive expressions for a number of quantities of interest in the model and show that the degenerate region of the parameter space observed in computer simulations is a spontaneously symmetry-broken phase separated from the normal phase of the model by a conventional continuous phase transition.

DOI: 10.1103/PhysRevE.70.066146 PACS number(s): 89.75.Hc, 05.50.+q, 64.60.Fr, 87.23.Ge

I. INTRODUCTION

There has in recent years been a surge of interest within the physics community in the properties of networks, including the internet, the worldwide web, and social and biological networks of various kinds [1–4]. Work has been divided between studies of specific real-world networks, along with the development of measures and algorithms for their analysis, and the creation of models to predict and explain network behavior. It is on models that we focus here.

Network modeling goes back at least as far as the well-known random graph or Bernoulli graph, studied by Solomonoff and Rapoport in the early 1950s [5] and famously by Erdős and Rényi [6] a decade later. The random graph, however, is a poor model for most real-world networks, as has been argued by many authors [1,4,7], and so other models have been developed. Recent attention has focused particularly on generalized random graphs such as the configuration model [8–10] and on generative models, particularly models of growing networks [2,4,11,12]. There is, however, another class of network models that, while widely used and valuable, has attracted little attention in the physics community, namely, the class of "exponential random graphs" or "p-star models." Building on early statistical work by Besag [13], exponential random graphs were first studied in the 1980s by Holland and Leinhardt [14], and later developed extensively by Strauss and others [15,16]. Today, they are commonly used as a practical tool by statisticians and social network analysts [17–19].

Despite their widespread adoption, few analytic results are known for exponential random graphs: most work has made use of computer simulation to fit models to observational data and evaluate model predictions. Exponential random graphs, however, are ideally suited to study using the techniques of statistical physics. Recently, physicists have examined exponential random graph models of network assortativity [20,21] and transitivity [22]. Here we take a different approach and show how physics techniques can be used to derive analytically the behavior of one of the most fundamental of exponential random graph models, the two-star model. We view this solution not only as a calculation of interest in its own right, but also as a demonstration of the way in which physics techniques can be fruitfully applied to problems from other fields.

II. THE MODEL

The exponential random graph is an ensemble model. One defines an ensemble consisting of the set of all simple undirected graphs with n vertices and no self-edges (i.e., networks with either zero or one edge between each pair of distinct vertices) and one specifies a probability $P(G)$ for each graph G in this ensemble. Properties of the model are calculated as averages over the ensemble. Let us define the *graph Hamiltonian*, also referred to by statisticians as a *log odds ratio*, to be $H(G)=F-\ln P(G)$. Here F (usually called the free energy) is any convenient origin for the measurement of the Hamiltonian, such as, for instance, the logarithm of the probability of the empty graph (i.e., the probability of n vertices with no edges). Then

$$P(G)=\frac{e^{-H(G)}}{Z}, \quad Z=e^{-F}=\sum_G e^{-H(G)}. \tag{1}$$

Z is the graph partition function and many quantities of interest can be calculated from it, or alternatively from the free energy.

So far, this model is entirely general, but progress is made by assuming the Hamiltonian to be a linear combination of scalar graph observables, such as number of edges, degree sequences, or clustering coefficients. In this paper we study one of the simplest nontrivial cases, the two-star model, for which $H(G)=\theta_1 m(G)+\theta_2 s(G)$, where θ_1 and θ_2 are independent parameters, $m(G)$ is the number of edges in the graph, and $s(G)$ is the number of "two-stars." A two-star is a pair of edges that share a common vertex. By adding a term coupling to the number of two-stars, this model gives us the ability either to encourage or to discourage the appearance in the network of vertices with high degree, the number of two-stars around a vertex increasing quadratically with degree, so that high-degree vertices are more strongly affected by the value of θ_2 than low-degree ones.

Let us denote by k_i the degree of vertex i. Then

$$m(G)=\tfrac{1}{2}\sum_i k_i, \quad s(G)=\tfrac{1}{2}\sum_i k_i(k_i-1), \tag{2}$$

and hence we can write the Hamiltonian in the form

$$H = -\frac{J}{n-1}\sum_i k_i^2 - B\sum_i k_i, \tag{3}$$

where the "coupling constant" $J=-\frac{1}{2}(n-1)\theta_2$ and the "field" $B=\frac{1}{2}(\theta_2-\theta_1)$. The factor $(n-1)$ in the definition of J is not strictly necessary, but it makes the equations simpler later on.

There are a number of analytic techniques from statistical mechanics that can be brought to bear on problems like this. As discussed elsewhere [23], the two-star model can be regarded as a type of Ising model on the edge-dual of a fully connected graph, and can thus usefully be treated using mean-field theory or perturbation theory [22,23]. Alternatively, one can use the Hubbard-Stratonovich transform and saddle-point expansions to derive nonperturbative results [20]. Here we make use of the latter approach to solve the two-star model.

III. ANALYTIC APPROACH

Our goal is to calculate the partition function Z, Eq. (1), or equivalently the free energy. First, we introduce auxiliary fields ϕ_i on the vertices of the graph using the Hubbard-Stratonovich relation

$$\exp[Jk_i^2/(n-1)] = \sqrt{\frac{(n-1)J}{\pi}} \times \int_{-\infty}^{\infty} d\phi_i \exp[-(n-1)J\phi_i^2 + 2J\phi_i k_i], \tag{4}$$

which gives

$$Z = \left[\frac{(n-1)J}{\pi}\right]^{n/2} \int \mathcal{D}\boldsymbol{\phi} \exp\left(-(n-1)J\sum_i \phi_i^2\right) \times \sum_G \exp\left(\sum_i (2J\phi_i + B)k_i\right), \tag{5}$$

where $\mathcal{D}\boldsymbol{\phi}$ indicates the path integral over the fields $\{\phi_i\}$ and we have interchanged the order of the integral and the sum over graphs G.

The sum over graphs can now be performed by defining the symmetric adjacency matrix σ_{ij} equal to 1 if there is an edge between vertices i and j and zero otherwise. Then, noting that $k_i=\Sigma_j\sigma_{ij}$, we can write

$$\sum_i (2J\phi_i + B)k_i = \sum_{ij} (2J\phi_i + B)\sigma_{ij} = \sum_{i<j} [2J(\phi_i + \phi_j) + 2B]\sigma_{ij}. \tag{6}$$

Since σ_{ij} is symmetric, its values for $i<j$ completely define the graph, and hence

$$\sum_G \exp\left(\sum_i (2J\phi_i + B)k_i\right) = \prod_{i<j}\sum_{\sigma_{ij}=0}^{1} e^{[2J(\phi_i+\phi_j)+2B]\sigma_{ij}} = \prod_{i<j}(1 + e^{2J(\phi_i+\phi_j)+2B}). \tag{7}$$

Substituting this result into Eq. (5), we then get

$$Z = \int \mathcal{D}\boldsymbol{\phi}\, e^{-\mathcal{H}(\phi)}, \tag{8}$$

where the effective Hamiltonian $\mathcal{H}$ is

$$\mathcal{H}(\phi) = (n-1)J\sum_i \phi_i^2 - \frac{1}{2}\sum_{i\neq j}\ln(1 + e^{2J(\phi_i+\phi_j)+2B}) - \frac{1}{2}n\ln[(n-1)J]. \tag{9}$$

Thus we have transformed our network model into a field theory of a continuous scalar field on n sites, which can be solved using a variety of methods. The simplest mean-field approach is to ignore fluctuations and assume ϕ_i always to be equal to its most probable value, which occurs at the saddle point

$$\frac{\partial\mathcal{H}}{\partial\phi_i} = 0 = 2(n-1)J\phi_i - J\sum_{j(\neq i)}\{\tanh[J(\phi_i+\phi_j)+B]+1\}. \tag{10}$$

This has a symmetric solution $\phi_i=\phi_0$ for all i with

$$\phi_0 = \frac{1}{2}[\tanh(2J\phi_0 + B) + 1]. \tag{11}$$

This quantity has a simple physical interpretation. The mean degree $\langle k\rangle$ of a vertex in the graph is given by the derivative of the free energy thus:

$$\langle k\rangle = \frac{1}{n}\sum_i \langle k_i\rangle = \frac{1}{n}\frac{\partial F}{\partial B} = \frac{1}{2n}\sum_{i\neq j}\langle\tanh[J(\phi_i+\phi_j)+B]+1\rangle_\phi, \tag{12}$$

where $\langle\cdots\rangle_\phi$ indicates an average in the $\boldsymbol{\phi}$ ensemble of Eq. (8). Making the mean-field assumption of Eq. (11), this becomes

$$\langle k\rangle = (n-1)\phi_0, \tag{13}$$

and hence ϕ_0 is simply proportional to the mean degree of a vertex, within the mean-field approximation. The quantity $\langle k\rangle/(n-1)$ is called the "connectance" of the graph—it is the fraction of possible edges that are actually present and is a measure of the mean density. So we could also say that ϕ_0 is equal to the connectance. This allows us to interpret Eq. (11) very directly. For $J\leqslant 1$, this equation has only a single solution, but for $J>1$ we have three coexisting solutions when B is sufficiently close to $-J$. Only the outer two solutions are stable, giving us a bifurcation at $J_c=1$ corresponding to a continuous phase transition at this point to a symmetry-broken state exhibiting two phases, one of high density (typically nearly a complete graph) and one of low density. We show a plot of the solution of Eq. (11) in the main panel of Fig. 1.

Along the line $B=-J$ the Hamiltonian (3) is symmetric with respect to the interchange of edges and "holes"—the absence of edges between vertex pairs. In the inset to Fig. 1 we show the solution for the connectance as a function of J along this symmetric line and the plot shows the bifurcation clearly.

To move beyond the mean-field result, we make use of the method of stationary phase. Expanding the effective Hamil-

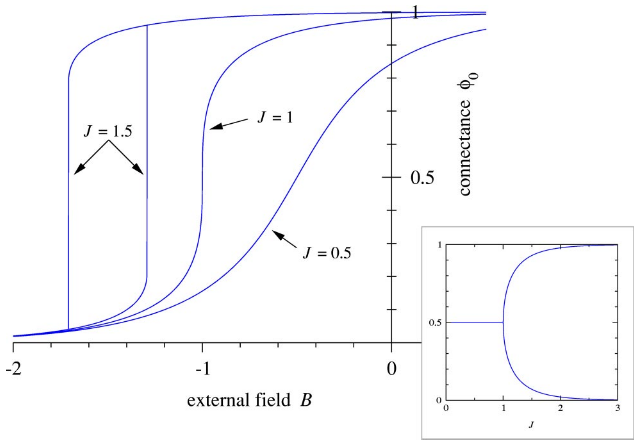

FIG. 1. The mean-field solution for the connectance $\phi_0 = \langle k\rangle/(n-1)$ in the two-star model from Eq. (11), for values of the coupling J below, at, and above the phase transition. For the case $J=1.5$ we are in the symmetry-broken phase and the hysteresis loop corresponding to the high- and low-density phases of the system is clearly visible. Inset: the bifurcation of the connectance as a function of J along the symmetric line $B=-J$.

tonian (9) about the mean-field solution to leading order we have

$$\mathcal{H} = \mathcal{H}(\boldsymbol{\phi}_0) + \boldsymbol{\phi}'\mathbf{M}\boldsymbol{\phi}' + O(\boldsymbol{\phi}^3), \tag{14}$$

where $\boldsymbol{\phi}' \equiv \boldsymbol{\phi} - \boldsymbol{\phi}_0$ and $\mathbf{M}$ is the Hessian matrix of second derivatives of $\mathcal{H}$ with respect to $\boldsymbol{\phi}$, evaluated at $\boldsymbol{\phi}_0$. Changing variables to $\boldsymbol{\xi} = \mathbf{Q}\boldsymbol{\phi}'$, where $\mathbf{Q}$ is the matrix of eigenvectors of $\mathbf{M}$, $\mathbf{M}$ is diagonalized and

$$\mathcal{H} = \mathcal{H}(\boldsymbol{\phi}_0) + \sum_i \lambda_i \xi_i^2 + O(\boldsymbol{\xi}^3), \tag{15}$$

with λ_i being the ith eigenvalue of $\mathbf{M}$. Substituting into Eq. (8) and observing that the Jacobian of the variable change $|\mathbf{Q}|=1$, the path integral becomes a product of independent Gaussian integrals and $Z = e^{-\mathcal{H}(\phi_0)}/\sqrt{|\mathbf{M}|}$, or equivalently

$$F = \mathcal{H}(\boldsymbol{\phi}_0) + \tfrac{1}{2}\ln|\mathbf{M}|, \tag{16}$$

where $|\mathbf{M}|$ is the determinant of $\mathbf{M}$.

Notice that the peak in the Boltzmann factor of Eq. (8) becomes increasingly narrow as n becomes large because of the leading $(n-1)$ in the effective Hamiltonian (9), and hence we expect the higher-order terms in Eq. (14) to become negligible in this limit by comparison with the quadratic term. Thus we expect that the stationary phase approximation will be highly accurate for large networks, and we show below that this is indeed the case.

The elements of the Hessian matrix have the values

$$M_{ij} = \begin{cases} -4J^2\phi_0(1-\phi_0) & \text{for } i \neq j, \\ (n-1)[2J - 4J^2\phi_0(1-\phi_0)] & \text{for } i = j, \end{cases} \tag{17}$$

giving

$$|\mathbf{M}| = [2(n-1)J]^n[1-2J\phi_0(1-\phi_0)]^{n-1}[1-4J\phi_0(1-\phi_0)]. \tag{18}$$

Then, making use of Eqs. (9) and (11), we arrive at the solution for the free energy

$$F = n(n-1)J\phi_0^2 - \tfrac{1}{2}n(n-1)\ln(1+e^{4J\phi_0+2B}) + \tfrac{1}{2}(n-1)\ln[1-2J\phi_0(1-\phi_0)], \tag{19}$$

where we have kept leading order corrections to the mean-field result but dropped terms of order a constant and smaller that vanish in the large n limit.

From the free energy we can calculate expected values of a variety of properties of the model. For instance the mean degree $\langle k\rangle$ and the mean squared degree $\langle k^2\rangle$ are given by derivatives with respect to B and J and are equal to

$$\langle k\rangle = (n-1)\phi_0 + \frac{2J\phi_0(1-\phi_0)(1-2\phi_0)}{[1-4J\phi_0(1-\phi_0)][1-2J\phi_0(1-\phi_0)]}, \tag{20}$$

$$\langle k^2\rangle = (n-1)^2\phi_0^2 + \frac{(n-1)\phi_0(1-\phi_0)(1-4J\phi_0^2)}{[1-4J\phi_0(1-\phi_0)][1-2J\phi_0(1-\phi_0)]}. \tag{21}$$

The leading order term in each case is the same as the mean-field result, so that in the limit of large n both $\langle k\rangle$ and $\langle k^2\rangle$ take their mean-field values. The variance of the degree $\langle k^2\rangle - \langle k\rangle^2$ on the other hand is zero within the mean-field approximation because of the cancellation of the leading terms but nonzero beyond mean field:

$$\langle k^2\rangle - \langle k\rangle^2 = (n-1)\frac{\phi_0(1-\phi_0)}{1-2J\phi_0(1-\phi_0)}. \tag{22}$$

From consideration of Fig. 1 one might expect this quantity to diverge at the phase transition, but in fact it does not, having merely a cusp at that point. In Fig. 2 we show the form of this function along the symmetric line $B=-J$ as a function of J. The figure also shows the results of Monte Carlo simulations of the two-star model for the same parameter values and, as we can see, agreement between the simulations and the analytic solution is excellent.

A divergence does occur in the variance of the number of edges in the network at the phase transition. This quantity, which plays the role of a susceptibility for the model, is given to leading order by

$$\langle m^2\rangle - \langle m\rangle^2 = \frac{\partial^2 F}{\partial B^2} = (n-1)\frac{2\phi_0(1-\phi_0)}{1-4J\phi_0(1-\phi_0)}. \tag{23}$$

This diverges as $|J-J_c|^{-1}$ as we approach the transition along the symmetric line $B=-J$. [One might imagine that the variances of k and m in Eqs. (22) and (23) would be proportional to one another, but this is not the case—notice that the denominators of the two equations differ in the multiple of J.]

By constrast with the case of conventional statistical mechanics, the critical point itself is not usually a focus of interest in network models—there is no reason why a real-world network should be near this special point and in most

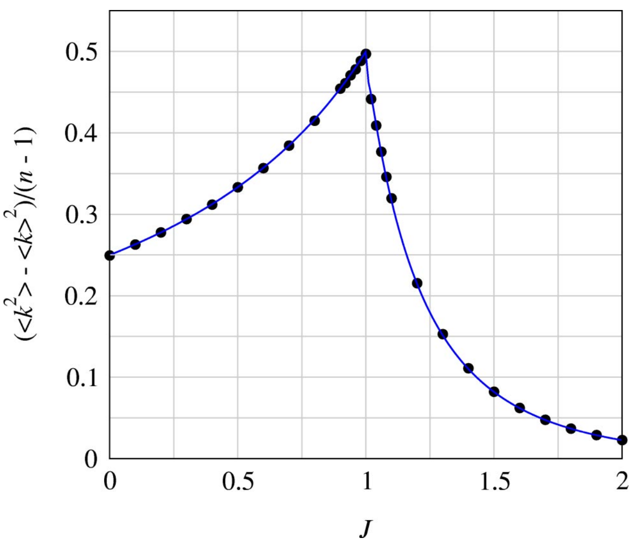

FIG. 2. The variance of vertex degree in the two-star model as a function of the coupling J along the symmetric line $B=-J$. The phase transition is marked by a cusp in the variance, but no divergence. The solid line represents the analytic solution, Eq. (22), in the large system size limit, and the points are the results of Monte Carlo simulations of the model for $n=1000$.

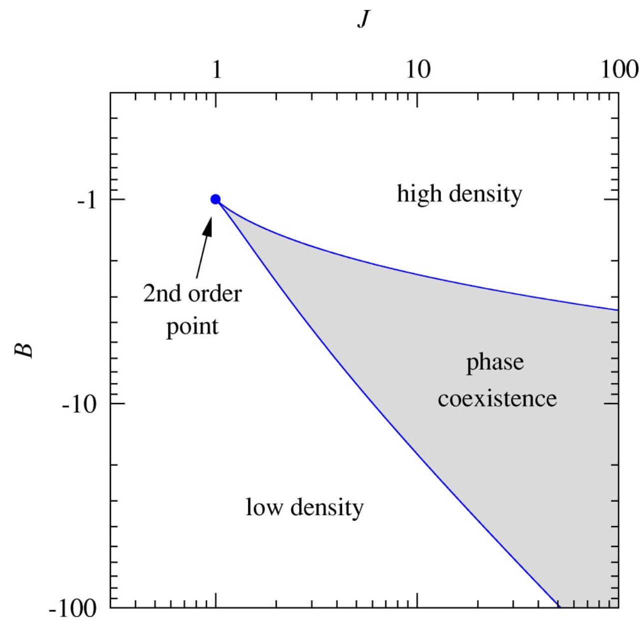

FIG. 3. The phase diagram for the two-star model. The shaded region indicates the hysteretic region in which both high- and low-density phases are possible.

cases one is not at liberty to tune network parameters to make it so. Still, it is reasonable to ask whether the network has any unusual structure in the critical region. Normally, however, it will not. Criticality is a property of the ensemble of graphs rather than any single graph, just as it is a property of the ensemble in conventional statistical mechanics. No individual member of the ensemble necessarily has any unusual form, but the ensemble as a whole has a critical structure: there are strong "fluctuations" in the number of edges from one member to another. In practical cases where we only observe a single member of the ensemble however, such fluctuations would not be apparent.

One can also ask whether the network described by the two-star model possesses a giant component. Molloy and Reed [8] have demonstrated that a network without degree correlations possesses a giant component if and only if $\langle k^2\rangle > 2\langle k\rangle$. We can evaluate this criterion using Eqs. (20) and (21), and find that for all values of the system parameters the network possesses a giant component in the limit of large n.

In Fig. 3 we show the phase diagram for the two-star model as a function of the parameters J and B. The critical point is at $J=1$, $B=-1$, and beyond this point there are high- and low-density phases separated by a phase coexistence region. In the coexistence region the phase of the model depends on its history in a manner characteristic of hysteretic systems. Some studies of exponential random graphs have considered the case in which the number of edges in the graph is fixed, a "conserved-order-parameter" version of the current model [20]. In such a case, the phase coexistence region will correspond to true coexistence; low free-energy states of the system will be states in which the system prefers simultaneously to have some high-degree "hub" vertices that connect to essentially all others and some of lower degree, rather than being uniform everywhere. Such "degenerate" behavior has been observed since the earliest numerical experiments on exponential random graphs [14–16,24]. Here we see that this behavior is the precise network analog of the phase separation phenomenon known to physicists from many other systems.

IV. CONCLUSIONS

In this paper, we have given a nonperturbative analytic solution of one of the oldest of network models, the two-star model, which is perhaps the simplest nontrivial model of the class known as exponential random graphs and has been long studied in the social sciences. The model turns out to be perfectly suited to solution by the methods of statistical physics, and among other things the solution shows the degenerate behavior of the model in certain parameter regimes to be the result of a symmetry breaking between high- and low-density phases, which are separated from the "normal" region of the model by a continuous phase transition.

The exponential random graphs are, we believe, an important class of network models, which have largely been neglected despite the high level of interest in networks in the last few years. We hope that others will also take up the study of these models, using either methods like those discussed here or other methods yet to be described.

ACKNOWLEDGMENTS

The authors thank Julian Besag, Mark Handcock, and Pip Pattison for useful conversations. This work was supported in part by the National Science Foundation under Grant No. DMS-0234188 and by the James S. McDonnell Foundation.

[1] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[5] R. Solomonoff and A. Rapoport, Bull. Math. Biophys. **13**, 107 (1951).
[6] P. Erdős and A. Rényi, Publ. Math. Inst. Hungarian Acad. Sci. **5**, 17 (1960).
[7] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[8] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[9] W. Aiello, F. Chung, and L. Lu, in *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing* (Association of Computing Machinery, New York, 2000), pp. 171-180.
[10] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[11] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[12] P. L. Krapivsky and S. Redner, in *Statistical Mechanics of Complex Networks*, edited by R. Pastor-Satorras, J. Rubi, and A. Díaz-Guilera, Lecture Notes in Physics, Vol. 625 (Springer, Berlin, 2003), pp. 3-22.
[13] J. E. Besag, J. R. Stat. J. R. Stat. Soc. Ser. B. Methodol. **36**, 192 (1974).
[14] P. W. Holland and S. Leinhardt, J. Am. Stat. Assoc. **76**, 33 (1981).
[15] D. Strauss, SIAM Rev. **28**, 513 (1986).
[16] O. Frank and D. Strauss, J. Am. Stat. Assoc. **81**, 832 (1986).
[17] S. Wasserman and P. Pattison, Psychometrika **61**, 401 (1996).
[18] C. Anderson, S. Wasserman, and B. Crouch, Soc. Networks **21**, 37 (1999).
[19] T. A. B. Snijders, J. Soc. Struct. **3**, 2 (2002).
[20] J. Berg and M. Lässig, Phys. Rev. Lett. **89**, 228701 (2002).
[21] J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).
[22] Z. Burda, J. Jurkiewicz, and A. Krzywicki, Phys. Rev. E **69**, 026106 (2004).
[23] J. Park and M. E. J. Newman, e-print cond-mat/0405566.
[24] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Phys. Rev. E **69**, 046117 (2004).