Properties of a random attachment growing network

László Zalányi,[1,2] Gábor Csárdi,[1,2] Tamás Kiss,[1,2] Máté Lengyel,[1,2] Rebecca Warner,[2,3] Jan Tobochnik,[2,3,*]
and Péter Érdi[1,2,3]

[1]*Department of Biophysics, KFKI Research Institute for Particle and Nuclear Physics of the Hungarian Academy of Sciences,
Budapest, Hungary*
[2]*Center for Complex Systems Studies, Kalamazoo College, Kalamazoo, Michigan 49006, USA*
[3]*Physics Department, Kalamazoo College, Kalamazoo, Michigan 49006, USA*

In this study we introduce and analyze the statistical structural properties of a model of growing networks which may be relevant to social networks. At each step a new node is added which selects k possible partners from the existing network and joins them with probability δ by undirected edges. The "activity" of the node ends here; it will get new partners only if it is selected by a newcomer. The model produces an infinite-order phase transition when a giant component appears at a specific value of δ, which depends on k. The average component size is discontinuous at the transition. In contrast, the network behaves significantly different for $k=1$. There is no giant component formed for any δ and thus in this sense there is no phase transition. However, the average component size diverges for $\delta \geqslant \frac{1}{2}$.

I. INTRODUCTION

There are many kinds of networks including probably the most influential network of all, the World Wide Web [1]. This network is a popular one to analyze because of its size and easy accessibility for statistical analysis. However, there are many other networks that share some of the properties of the Web and some that do not. Among these networks we find social networks [2–4], collaboration nets [5–8], industrial and business related networks [5,6,9], transportation nets [10] and many biological related nets such as food, ecological, and protein interaction networks [11–14], and neural networks [15].

The mathematical description of networks started with the fundamental works of Erdős and Rényi [16,17], which in the absence of reliable data on large networks were rarely compared to real networks. Recently, the computational boom has provided us an increasing number of types of networks and more data on these networks. One of the most exciting discoveries is the scale-free structures of certain evolving networks [18–20]. These nets have power-law degree distribution, where only a few vertices have many connections to the others and the rest of the graph is rarely connected. To explain the origin of this scale-free structure of networks Barabási *et al.* [21,22] suggested the mechanism of preferential attachment and emphasized the key role of growth. In their model the probability of a new node connecting to an existing node is proportional to the degree of the target node. Variations on this model include networks where there is aging of nodes, nonlinear attachment probabilities, and rewiring are allowed [23–26].

Probably the most obvious feature of real networks that is missing from most of the models studied by mathematicians and physicists are characteristics of individual nodes in real networks which influence the connection probability. Thus, if

the nodes represent individual persons, it is obvious that in many circumstances two people are more likely to become connected in some form of relationship because of the nature of their individual characteristics. Our model is motivated by the need to incorporate this idea. A similar idea was used in a preferential attachment model by Bianconi and Barabási [27] who assigned to each new node a fitness parameter. In their model a larger fitness parameter may overcompensate the smaller probability of attachment.

In our study we propose a simple model of growing networks whose statistical properties are identical to a more complicated model containing nodes with distinct characteristics. We will calculate the edge distribution of the growing network, the distribution of cluster sizes, and the emergence of a giant cluster. We will also show how the number of attempted connections made when a new node is added determines the position and type of the phase transition as well as the cluster-size distribution.

II. THE MODEL

We first consider a social network model where each node has individual characteristics or traits. Each node that is added to the network is assigned a permanent set of random traits which could be coded as an ordered binary string or vector of length L. When a node is added it chooses randomly $k \in N$ possible partners from the already existing nodes, or if there are less than $k+1$ (because the simulation has not yet reached time step $k+2$) it chooses all the existing nodes as possible partners. A trait distance between the new node and one of its possible partners is calculated based on their trait vectors $(\vec{t}_1, \vec{t}_2)$ using a distance measure, $D(\vec{t}_1, \vec{t}_2)$, such as the Hamming distance. Then a connection is formed between the two nodes with a probability determined from a given probability distribution over the distance function $p(D)$. Different functions $p(D)$ correspond to different sociopsychological situations. Thus, if we wish to model the case where people are more likely to link together

---

*Corresponding author. Email address: jant@kzoo.edu

if they have similar traits, then $p(D)$ would be a monotonically decreasing function of D. For this case, the simplest $p(D)$ would be to form a link if D is below some threshold. This procedure is repeated for each possible partner of the new node. Thus, each new node can have initially up to k links with the other existing nodes. Existing nodes can have more than k links as more nodes are added to the network and link up with the existing nodes. There are no multiple links between pairs of nodes.

Because each node is given a random trait vector, and the nodes to link to are also chosen randomly, many properties of the network simply depend on the probability δ that two chosen nodes will link together:

$$\delta = \sum_{D} p(D(\vec{t}_1, \vec{t}_2)) r(D(\vec{t}_1, \vec{t}_2)), \qquad (1)$$

where $r(D)$ is the probability of the distance D between two nodes, and the sum is over all possible distance values. Thus, the model is reduced to the following procedure. At each time step we add a node to the network, and attempt to link with k existing nodes which are chosen at random. An actual connection is made with a probability δ. The asymptotic behavior of the network in the limit of large time t does not depend on the initial condition of starting with a single isolated vertex.

Although frequently structural properties of a network of nodes with trait vectors depend only on δ, there are other properties which will depend on the detailed form of $p(D)$ and the nature of the trait vectors. Examples of such properties include the distribution of traits in different parts of the network and the correlation of traits with distance in the network. For example, one can imagine a very simple network of nodes representing men and women. In one network the probability of forming a link is independent of sex and in the other persons prefer to link up with members of the opposite sex. As long as the mean probability of two chosen nodes linking together is the same in the two scenarios the structural properties of the two networks will be the same, but the distribution of men and women within the network will be quite different in the two cases. In this paper we confine ourselves to the structural properties of networks and are considering these other nonstructural properties in our current research.

III. AGE DEPENDENCE OF THE EXPECTED NUMBER OF EDGES, AND THE EDGE DISTRIBUTION

The expected number of edges at a node is approximately

$$K_N(t) = \overbrace{\delta k}^{\text{initial connections}} + \overbrace{\sum_{s=N+1}^{t} \delta \frac{k}{s-1}}^{\text{later connections}}$$

$$= \delta k (1 + H_{t-1} - H_{N-1}), \qquad (2)$$

where $N > k$ is the time step when it was created (the smaller N is, the older the node is), t is the total simulation time, δ is the probability that two nodes form a connection, k is the maximum number of initial connections of a newly created node, and H_n is the nth harmonic number given by the formula $H_n = \Sigma_{i=1}^{n} 1/i$ for $n > 0$, and $H_0 = 0$. This equation shows that the number of edges of a node heavily depends on the age of the node.

Equation (2) slightly overestimates the number of connections for the oldest nodes in the network in two respects. First, the above formula assumes that a node always has k possible initial connections. However, multiple connections between a pair of nodes are not allowed, and there are less than k available partners for the initial connections of a node created before or in the kth time step (overestimation of initial connections). Second, the term for the late connections assumes that a node has a $k/(m-1)$ chance of being selected as the partner of the mth node (which chooses k possible partners out of $m-1$ already existing nodes). However, for a node created in time step $N < k$, this term yields a probability of being chosen greater than 1 between time steps $N+1$ and k (where $m-1 < k$) that is unacceptable again because multiple connections between a pair of nodes are not allowed (overestimation of late connections). Below is the formula correcting these errors, but we will use the simpler, uncorrected formula in the remaining part of our paper because the errors are negligible.

$$K_N(t) = \delta \min(k, N-1) + \sum_{s=N+1}^{t} \delta \min\left(\frac{k}{s-1}, 1\right)$$

$$= \delta k (1 + H_{t-1} - H_{N-1})$$

$$- \delta \Bigg(\underbrace{\max(k-N+1, 0)}_{\text{initial connections}} + \underbrace{k \max(H_{k-1} - H_{N-1}, 0) - \max(k-N, 0)}_{\text{late connections}} \Bigg)$$

$$\underbrace{}_{\text{correction for oldest nodes}}$$

$$= \begin{cases} \delta \left[k \left(\alpha_t - \ln(k-1) - \dfrac{1}{2(K-1)} \right) - 1 \right] & \text{if } N \leq k+1 \\[2ex] \delta k \left(\alpha_t - \ln(N-1) - \dfrac{1}{2(N-1)} \right) & \text{if } N > k+1, \end{cases} \qquad (3)$$
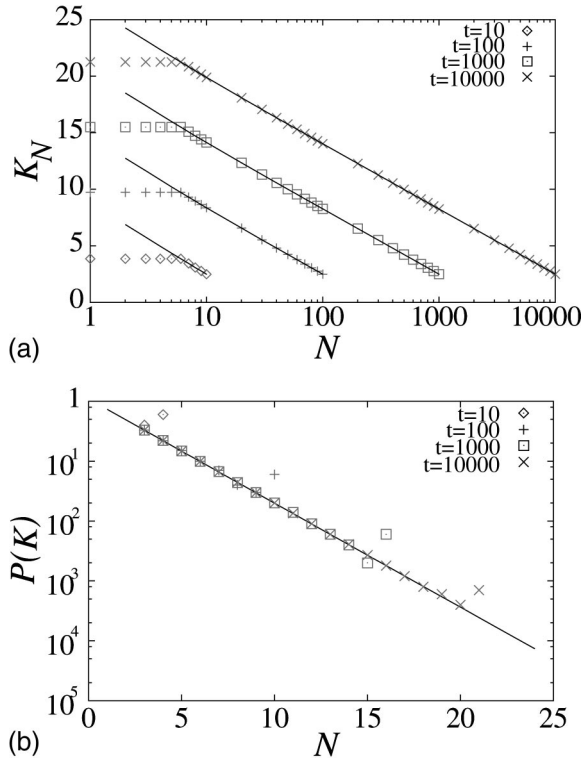
FIG. 1. (a) Expected number of edges of a node (K_N) as a function of the age of the node (N) at different ages of the network (t). *Symbols* are numerically calculated values from Eq. (3), showing that the first $k+1$ nodes have the same number of connections at any t, whereas there is an exponential breakdown in the expected number of edges for nodes created later than these. *Lines* represent approximations by Eq. (4): the number of edges of the first k nodes are overestimated because the correction term in Eq. (3) was ignored (see text). Note that the x axis is logarithmic. (b) Edge distribution of the network. *Symbols* represent numerically calculated distributions, where the numbers of edges of individual nodes were obtained from Eq. (3). These numbers were binned into integer values and the relative frequencies of occurrences in each bin were plotted. The *line* represents the approximate distribution given by Eq. (6), showing that it is valid for the edge distribution at large values of t (for $t \gtrsim 100$). The mismatch between approximated and actual distributions at the highest connection numbers is due to the same reason as in (a). Note that y axis is logarithmic. Parameters were $\delta = 0.5$, $k = 5$.

using $H_n \sim \ln n + \gamma + 1/2n$, where $\gamma = -\int_0^\infty e^{-x} \ln x\, dx \sim 0.5772$ is the Euler-Mascheroni constant [28], and $\alpha_t = 1 + \ln(t-1) + 1/2(t-1)$.

Note that the first $k+1$ nodes are expected to have the same number of connections (because K_N does not depend on N in their case), and the edge number starts breaking down exponentially for nodes created after time step $k+1$ [Figs. 1 and 2(a)]. This means that this growth mechanism is identical to that where the first $k+1$ nodes are created in the same time step.

We now wish to determine the edge distribution $P(X)$ equal to the probability that a node picked at random has on average X edges. We return to Eq. (2) ignoring the correction term in Eq. (3), and write the formula for $K_N(t)$ in the simpler form [Figs. 1 and 2(a)]:
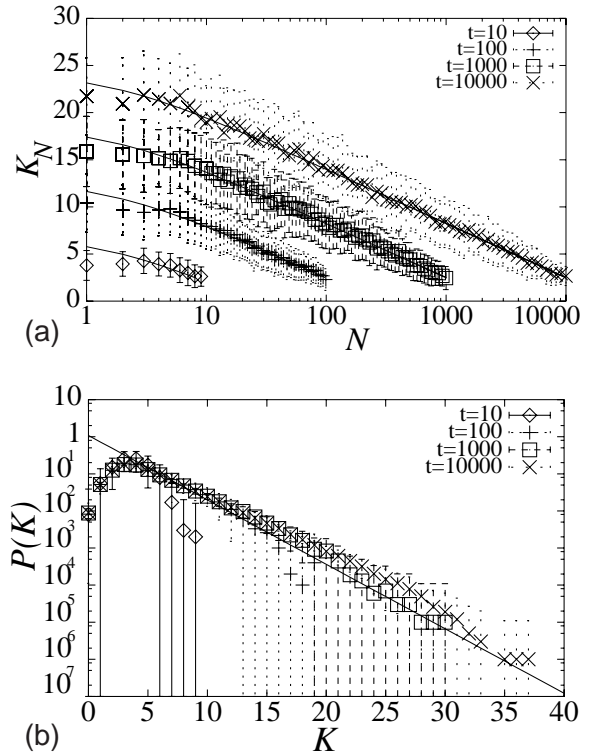
FIG. 2. (a) Numerical simulation of the average expected number of edges of a node (K_N) as a function of the age of the node (N). *Symbols* are results of numerical simulations, *line* is the graph of Eq. (4). Also see notes of Fig. 1(a). (b) Numerical simulation of the edge distribution of the network (*symbols*). *Line* is the graph of Eq. (6). Average relative frequency of individual number of edges and their standard deviation were calculated. Deviation of simulation from analytical results at high number of edges is a result of the finite size of simulated networks due to dispersion of expected number of edges around its expected value as shown in part *A* of this figure. As age of the network increases this deviation disappears and simulation results approach analytical approximation for longer interval. At low number of edges deviation results from neglecting the correction term in Eq. (4) Averages and standard deviations were calculated from 100 simulations. Parameters were $\delta = 0.5$, $k = 5$ as in Fig. 1.

$$K_N(t) \simeq \delta k \left(\alpha_t - \ln(N-1) - \frac{1}{2(N-1)} \right), \qquad (4)$$

where α_t is the same quantity used in Eq. (3). Using Eq. (4), neglecting the term $1/2(N-1)$ in Eq. (4) for N large enough, and knowing that the age distribution of nodes is uniform, we analytically approximate the edge distribution of the network with the following exponential

$$P(K(t)=X) = \frac{1}{\delta k t} e^{-X/\delta k + \alpha_t}. \qquad (5)$$

We used the standard transformation rule for random variables, $P(N) = P(K_N) |dK_N/dN|$ with $P(N) = 1/t$. For sufficiently large t, due to the definition of α_t, this can be effectively approximated by a distribution which is independent of t [Figs. 1 and 2(b)]:

$$P(X) = \frac{1}{\delta k} e^{-X/\delta k + 1}. \tag{6}$$

We can also determine a slightly different degree or edge distribution which is the percentage of nodes with m edges. Denote by $d_m(t)$ the expected number of nodes with degree m at time t. The number of isolated nodes $d_0(t)$ will increase by $(1-\delta)^k$, which is the probability of the addition node not connecting to any existing node, and decrease on average by $k\delta d_0(t)/t$:

$$d_0(t+1) = d_0(t) + (1-\delta)^k - k\delta \frac{d_0(t)}{t}. \tag{7}$$

The formula for the expected number of nodes of degree $m > 0$ is a bit complicated. For ($1 \leq m \leq k$) there are two ways to increase d_m: either selecting degree $m-1$ nodes for connection with the new node or the new node having exactly m edges. For ($m > k$), the new node cannot contribute to d_m. The decrease will be proportional to the probability of choosing a degree m node for attachment.

$$1 \leq m \leq k: \quad d_m(t+1) = d_m(t) + k\delta \frac{d_{m-1}(t)}{t} + \binom{k}{m}$$

$$\times \delta^m (1-\delta)^{k-m} - k\delta \frac{d_m(t)}{t}, \tag{8}$$

$$m > k: \quad d_m(t+1) = d_m(t) + k\delta \frac{d_{m-1}(t)}{t} - k\delta \frac{d_m(t)}{t}. \tag{9}$$

These equations are correct as $t \to \infty$, and numerical simulations show that $d_m(t) \sim p_m t$. Substituting this form into the equations for $d_m(t)$ we obtain

$$m \leq k: \quad p_m = \delta^m \sum_{j=0}^{m} \binom{k}{j} \frac{(1-\delta)^{k-j}}{(1+k\delta)} \left(\frac{k}{1+k\delta}\right)^{m-j}, \tag{10}$$

$$m > k: \quad p_m = p_k \left(\frac{k\delta}{1+k\delta}\right)^{m-k}. \tag{11}$$

This degree distribution p_m decays exponentially consistent with our previous result for $P(X)$.

IV. CRITICAL BEHAVIOR

A. Cluster-size distribution

In some network models, such as the preferential attachment models, all the nodes belong to a single cluster. For such models the focus is on the degree distribution and the distance between nodes in the network. However, our network can contain a number of disconnected clusters of nodes. Then the key questions become what is the cluster-size distribution and is there a phase transition between a collection of finite size clusters and the appearance of a giant cluster much larger than the rest. The transition is similar to that in percolation, with our parameter δ playing the role of the site occupation probability in a percolation model. The key difference between our model and percolation models is that our nodes do not sit on a lattice structure, and there is thus no geometric constraints. The definition of a giant cluster in our model is somewhat different than a spanning cluster in percolation models. Nevertheless, some of the behavior is similar.

Our model is similar to one by Callaway *et al.* [29] where an infinite-order phase transition was found. In that model after a node was added to the network, two nodes were picked at random and connected with probability δ. Our model is more general in that we consider the effect of making more than one link at any given time. Also, in our model the new links are between the added node and existing nodes, whereas in the model by Callaway *et al.* the new links are between any two nodes in the network.

To determine the cluster distribution we use a procedure similar to the one we used to calculate the degree distribution. The cluster number $N_j(t)$ denotes the expected number of clusters of size j. On average, at each time step, $(1-\delta)^k$ isolated nodes arrive at the network and $k\delta N_1(t)/t$ nodes will be chosen for attachment reducing N_1. Thus, N_1 is described by

$$N_1(t+1) = N_1(t) + (1-\delta)^k - k\delta \frac{N_1(t)}{t}. \tag{12}$$

For $j > 1$ new clusters of size j come from connecting the new node to a cluster of size $j-1$ or if $k > 1$ using the new node to make connections between smaller clusters whose sizes add up to j. Reducing N_j will be $jk\delta N_j(t)/t$ nodes from clusters of size j connecting to the new node. Thus, we have

$$N_2(t+1) = N_2(t) + \binom{k}{1} \delta(1-\delta)^{k-1} \frac{N_1(t)}{t} - k\delta \frac{2N_2(t)}{t}, \tag{13}$$

$$\vdots$$

$$N_j(t+1) = N_j(t) + \left(\sum_{r=1}^{\min(k,j-1)} \binom{k}{r} \delta^r (1-\delta)^{k-r} \sum_{\substack{z_1+\ldots+z_r=j-1 \\ z_i \geq 1, i \leq r}} \frac{z_1 N_{z_1}(t)}{t} \frac{z_2 N_{z_2}(t)}{t} \cdots \frac{\left(j-1-\sum_{i=1}^{r-1} z_i\right) N_{(j-1-\sum_{i=1}^{r-1} z_i)}(t)}{t} \right)$$

$$- k\delta \frac{j N_j(t)}{t}. \tag{14}$$

The first sum in Eq. (14) determines the number of sums in the next term. Each of these sums represent a cluster that is melted into the j sized cluster. These equations are valid for $t \to \infty$, where the probability of closed loops tends to zero. The giant cluster, if there exists one, is an exception in which connection of nodes in loops is not negligible. Thus, Eq. (14) holds only for the finite sized clusters in the network. This property lets us to determine a generating function which we can use to find the size of the giant cluster. Our simulations show that solutions of Eqs. (12)–(14) are of the steady state form $N_j(t) = a_j t$. Using this form in Eqs. (12)–(14), we find
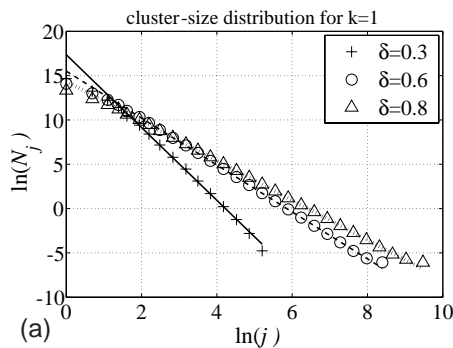
$$a_1 = \frac{(1-\delta)^k}{1+k\delta}, \tag{15}$$

$$a_2 = \frac{\binom{k}{1}\delta(1-\delta)^{k-1}a_1}{(1+2k\delta)}, \tag{16}$$

$$a_j = \frac{1}{1+jk\delta}\left[\sum_{r=1}^{\min(k,j-1)} \binom{k}{r}\delta^r(1-\delta)^{k-r} \right.$$

$$\times \sum_{\substack{z_1+\cdots+z_r=j-1 \\ z_i \geq 1, i \leq r}} \left(j-1-\sum_{i=1}^{r-1} z_i \right)$$

$$\left. \times a_{(j-1-\Sigma_{i=1}^{r-1}z_i)} \prod_{l=1}^{r-1} (z_l a_{z_l}) \right]. \tag{17}$$

Generally, we cannot obtain a simpler equation for the cluster-size distribution a_j, except for $k=1$. Substituting $k=1$ into the Eqs. (15)–(17) we obtain after some algebra the general result

$$a_j = (1-\delta)\delta^{j-1}(j-1)! \prod_{m=1}^{j} \frac{1}{1+m\delta}, \tag{18}$$

which can be written in the form

$$a_j = \frac{(1-\delta)\Gamma(1/\delta)}{\delta^2} \frac{\Gamma(j)}{\Gamma(j+1+1/\delta)}, \tag{19}$$

where $\Gamma(x)$ denotes the γ function. Equation (19) shows that the cluster-size distribution for $k=1$ always follows a power-law distribution. This result is confirmed by simulations shown in the left graph of Fig. 3. Distributions of cluster sizes for $k=2$ (right graph of Fig. 3), in contrast to $k=1$ show power-law behavior only near the phase transition.

B. Position of the phase transition

Figure 4 shows the simulation results for S, the ratio of the average size of the largest cluster to the total number of nodes versus the connection probability δ. The figure suggests that there is a smooth transition in the appearance of S at a specific value of δ between $\delta=0$ and $\delta=0.2$, which depends on the parameter k. To predict the position of a possible phase transition δ_c [29], we will use a generating function for the cluster-size distribution [30]. To derive the generating function we use the iterative Eqs. (15)–(17). The generating function will be of the form

$$g(x) = \sum_{j=1}^{\infty} b_j x^j, \tag{20}$$

where

$$b_j = j a_j, \tag{21}$$

is the probability that a randomly chosen node is from a cluster of size j. Multiplying both sides of Eqs. (15)–(17) by jx^j, and summing over j we derive a differential equation for $g(x)$
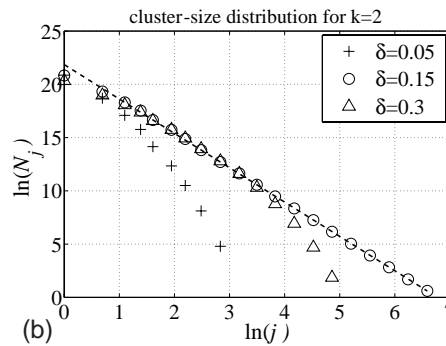

FIG. 3. Cluster-size distribution for different δs and $k=1$ (left), $k=2$ (right). Solid, dashed, and dotted lines are obtained from a least squares fit for the interval $11 > \ln(N_j) > -4$ (a) and $20 > \ln(N_j) > 2$ (b) indicating the power-law behavior of the distributions. Simulation data were obtained by averaging over 500 runs of 10^7 time steps and are shown on a log-log plot. Note that in (b) simulations for $\delta=0.05$ and $\delta=0.3$ distributions do not follow a power law. In Sec. IV B it is shown that there is a phase transition near $\delta=0.146$.
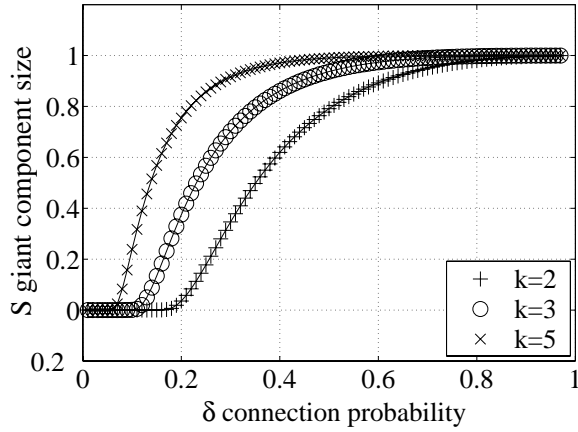
FIG. 4. Giant cluster size S as a function of δ and k. Symbols are from simulations of the growing network for 10^6 time steps averaged over 30 runs. Lines are from the analytical calculations.

$$g = -k\delta g' + x(1-\delta)^k + \sum_{i=1}^{k} \binom{k}{i} \delta^i (1-\delta)^{k-i}$$
$$\times (x^2 g' g^{i-1} i + x g^i). \quad (22)$$

Rearranging for g' we obtain

$$g' = \frac{(1-\delta)^k - g/x + \sum_{i=1}^{k} \binom{k}{i} \delta^i (1-\delta)^{k-i} g^i}{k\delta - x \sum_{i=1}^{k} \binom{k}{i} \delta^i (1-\delta)^{k-i} g^{i-1} i}, \quad (23)$$

which can be further simplified to

$$g' = \frac{-g/x + [1+(g-1)\delta]^k}{k\delta - xk\delta(1+(g-1)\delta)^{k-1}}. \quad (24)$$

The generating function for the finite size clusters is exactly one at $x=1$ when there is no giant cluster in the network and $g(1)<1$ otherwise. Hence

$$S = 1 - g(1). \quad (25)$$

Without an analytic solution for Eq. (24), we calculate S numerically by integrating Eq. (24) with the initial condition $[x, g(x)] = (x_0, x_0(1-\delta)^k/[1+k\delta])$ where x_0 is small. This is equivalent to starting with a cluster of only one node. In Fig. 4 there are results from direct simulations of the model (symbols) and solid lines from the integration of the generating function. The agreement is good which verifies the approximations.

To discuss the phase transition location we first consider the cases $k>1$. Consider the expected value that a randomly chosen node belongs to a finite size cluster. We can determine this quantity in terms of the generating function $g(x)$

$$\langle s \rangle = \frac{g'(1)}{g(1)}. \quad (26)$$

For those values of δ where no giant cluster exists, $\delta < \delta_c$, $g(1)=1$, and both the numerator and denominator of Eq. (24) goes to zero as $x \to 1$. Using l'Hôpital's rule we derive a quadratic equation for $g'(1)$. The solution of this equation is

$$g'(1) = \frac{1 - 2k\delta \pm \sqrt{(2k\delta-1)^2 - 4k(k-1)\delta^2}}{2k(k-1)\delta^2}, \quad (27)$$

for $g(1)=1$. Because as $\delta \to 0$ all clusters will have size 1, one can show that the correct solution of Eq. (27) is the one with the negative sign. In addition from Eq. (27) we can find the location of the phase transition. It is the value of δ where the solution of Eq. (27) becomes complex:

$$\delta_c = \frac{1 - \sqrt{1-1/k}}{2}. \quad (28)$$

In the region where there is a giant cluster $\delta > \delta_c$, Eq. (24) becomes as $x \to 1$,

$$g' = \frac{-g + (1+(g-1)\delta)^k}{k\delta - k\delta(1+(g-1)\delta)^{k-1}}, \quad (29)$$

which is still not solvable analytically. Making the approximation $(1 \pm a)^k \approx 1 \pm ka$ when $a \ll 1$, we can simplify Eq. (29) close to δ_c:

$$g'(1) \approx \frac{k\delta - 1}{k\delta^2(1-k)}, \quad (30)$$

where $g(1)<1$, $\delta > \delta_c$, and $[g(1)-1]\delta \ll 1$. In Fig. 5 we show the simulation results and the above derived theoretical functions for $g'(1)$. We can see that for $\delta < \delta_c$, where we have an explicit expression for $g'(1)$ in terms of the parameters k and δ the fit is very good. For $\delta > \delta_c$ the fit is good close to the phase transition point, where the approximation $(g-1)\delta \ll 1$ holds. Although below δ_c the description of $g'(1)$ is very good, it seems that the location of the phase transition and the value of the function $g'(1)$ above δ_c is somewhat different than the data. Also if we carefully check Fig. 5 at the jumps, we find that the larger the jump the less accurate the theory seems to be. This can be explained as follows. At the critical point the average size of finite clusters jumps, hence much larger clusters appear in the network. As we can only simulate for a finite time large (but not the giant) clusters are underrepresented. The weights of them computed from the simulation data are less than they would be in an infinitely long simulation. Away from the transition regime fewer finite size clusters remain beside the giant cluster in the network, and thus the distribution can be specified better.

Although the formalism using the generating function can be done for $k=1$, the meaning of a giant cluster is problematic. In Sec. IV A we showed that the size distribution of clusters for $k=1$ always follows a power law which means there is no obvious border between the "giant" cluster and smaller clusters. There is not a sharp break between the largest and the next largest cluster. The physical reason for this is that clusters grow only by the addition of newly added
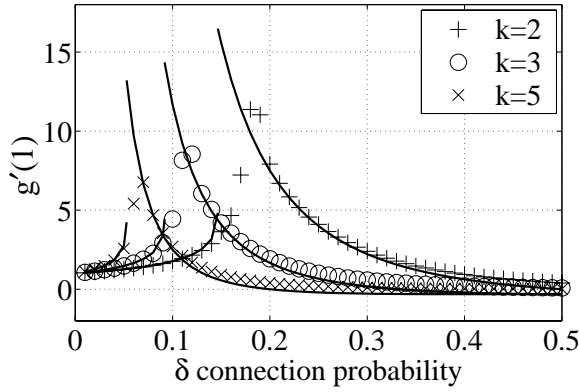
FIG. 5. Discontinuity in $g'(1)$ for different values of k. Solid lines are theoretical, and symbols are results from the simulations of growing networks for 10^6 time steps, averaged over 30 runs.

nodes. This is different than the case for $k>1$ and in percolation models where clusters can also grow by a link combining two clusters. In this sense no giant cluster appears in the network except for $\delta=1$. Equation (24) becomes

$$g'(x)=\frac{(1-\delta)-g/x+\delta g}{\delta(1-x)}, \qquad (31)$$

which becomes $\frac{0}{0}$ in the limit $x\to 1$ with $g(1)=1$. Applying l'Hôpital's rule yields

$$g'(1)=\frac{1}{1-2\delta}. \qquad (32)$$

At $\delta=\frac{1}{2}$, $g'(1)\to\infty$, which means the average size of finite clusters approaches infinity. From the definition of $g(x)$ in Eq. (20) and the power-law cluster-size distribution for a_j, it follows that $g(1)=1$ for any $\delta\neq 1$. To see that $g'(1)\to\infty$ as $x\to 1$ for $\delta>\frac{1}{2}$, we consider the sum form of the generating function in Eq. (20). For large j, $a_j\approx 1/j^{(1+1/\delta)}$, Eq. (19), and $g'=\sum_{j=1}^{\infty}j^2 a_j x^{j-1}$, which can not be summed for $\delta\geqslant\frac{1}{2}$.

When $\delta<\frac{1}{2}$, the probability of a new node not joining a cluster is higher than joining, and thus the weight of small clusters is higher than that of larger clusters, and hence the average size remains finite. As $\delta\to\frac{1}{2}$, the probability of forming clusters increases and so do the weight of large clusters.

C. Infinite-order transition

To show the nature of our phase transitions [29], we numerically integrated Eq. (24) for different values of k near the corresponding critical δ_c. In Fig. 6 the linear parts of the $\ln[-\ln(S)]$ plots suggest that
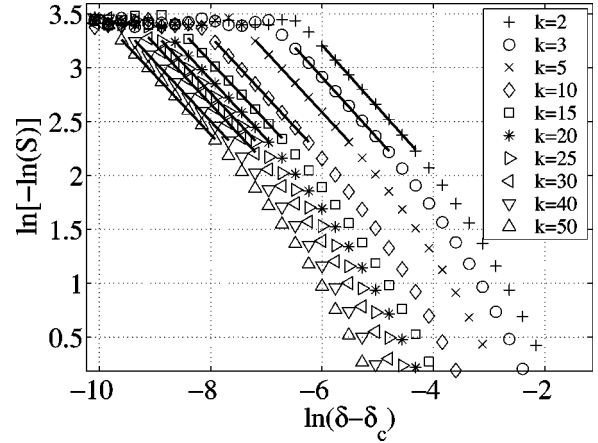

FIG. 6. Numerical calculation of the giant cluster size close to but above the phase transition. Least-squares fitted solid straight lines suggest $S(\delta)\sim e^{\alpha(\delta-\delta_c)^\beta}$. The flat ends of the curves on the top appear due to the limit of the accuracy of numerical integration.

$$S(\delta)\sim e^{\alpha(\delta-\delta_c)^\beta} \qquad \text{as } \delta\to\delta_c, \qquad (33)$$

and because all derivatives of S vanish at δ_c, the transition is of infinite order.

Table I contains the parameters of the fitted straight lines in Fig. 6. As the calculations were done close to the numerical limit and referring to the similar results in Ref. [29] we conjecture that β equals $-\frac{1}{2}$ for all k. This result suggests that the mechanism of the transition is common and the number of possible partners for each node to link to determines the speed of emergence of the giant cluster S. These results are in accord with Eq. (30), the average cluster size decrease is approximately independent of k, but the size of the jump and the rate of decrease is driven by k.

V. DISCUSSION

The present model was intended to gain insight into the evolution of various social networks by considering mechanisms that account for heterogeneity in the population of participating entities. To analyze the statistical properties of the generated network we simplified the model. We found that the structure of the network dramatically changes when the number of possible links to a newly added node increases from $k=1$ to $k=2$. With $k=1$ the network does not form a giant cluster but the average cluster size goes to infinity (at $\delta=\frac{1}{2}$) in contrast to $k\geqslant 2$, where the giant cluster appears in an infinite-order phase transition and the average cluster size jumps discontinuously but remains finite. The size of the jump corresponds to how slowly the giant cluster overcomes

TABLE I. The parameter values (α and β) of the fitted lines in Fig. 6. Taking into account that we were at the border of the maximal numerical accuracy and that the fit is short we presume $\beta=-\frac{1}{2}$.

k	2	3	5	10	15	20	25	30	40	50
α	-0.25	-0.5	-0.75	-1.14	-1.35	-1.52	-1.64	-1.77	-1.9	-2.02
β	-0.577	-0.569	-0.557	-0.554	-0.551	-0.552	-0.551	-0.554	-0.551	-0.55

the other competitive large clusters. However, there is no transition for $k=1$, where none of the clusters can absorb other clusters. The distribution of the size of finite clusters always follows an exponential distribution, both below and above the critical point for $k>1$, while the model studied in Refs. [31,29] is in a critical state below and at the critical point and exhibits an exponential distribution of cluster size above the transition as in a Berezinskii-Kosterlitz-Thouless phase transition. Thus, even though there are disconnected clusters as in our model, there are significant differences in the behavior of the cluster-size distribution.

Our model is similar to a previous model of Callaway *et al.* [29], but there are essential differences in several points due to nature of the growth algorithm: in the model of Callaway *et al.*, network growth and connection formation are independent while in our model only newly added nodes form connections. Also, in our model multiple connections might be formed in one time step depending on parameters k and δ. This difference is well reflected in the generating function derived for the two models.

The structural properties of our model are more relevant to many social networks than other growth models such as preferential attachment because the degree distribution is exponential which is closer to real social systems and because there are clusters of nodes which represents the reality of social systems where people usually form various communities which are relatively isolated from each other. As long as the distribution of nodal traits are random, the structural properties which we have discussed in this paper do not depend on the nature of the traits and thus our network model should be relevant to any social network. The next step is to analyze the distribution of traits on a social network. This will vary depending on how the attachment rule depends on the values of these traits even though the structural properties of the network remains the same. We will discuss the distribution of traits on a network in a future publication.

VI. CONCLUSIONS

We introduced a model of growing social networks and analyzed its statistical properties. Our analytical calculations showed that these growing networks exhibit exponential degree distributions. We gave an explicit description of the expected number of edges which showed an exponential dependence on the age of a node. We also showed that emergence of a giant cluster and the cluster-size distribution strongly depend on the number of possible initial partners. Numerical simulations suggested that the generated networks have scale-free cluster distributions only at the phase transition point. In all other regions of the phase space the cluster distribution was exponential. In the absence of an exact solution for Eq. (24), we showed numerical results suggesting that the order of the phase transition is infinite, which is similar to the results found in Ref. [29].

[1] P. Baran, Report No. RM-3420-PR 1964 (unpublished).

[2] S. Milgram, Psychol. Today **2**, 60 (1967).

[3] J. Scott, *Social Network Analysis: A Handbook* (SAGE Publications, London, 1991).

[4] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge, 1994).

[5] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 400 (1998).

[6] D.J. Watts, *Small Worlds* (Princeton University Press, Princeton, 1999).

[7] M.E.J. Newmann, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2002).

[8] A.L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, Physica A **311**, 590 (2002).

[9] T. Lux and M. Marchesi, Nature (London) **397**, 493 (1999).

[10] J. Banavar, A. Maritan, and A. Rinaldo, Nature (London) **399**, 130 (1999).

[11] H. Jeong, S.P. Mason, A.-L. Barabási, and Z.N. Oltvai, Nature (London) **411**, 41 (2001).

[12] S.A. Kauffman, J. Theor. Biol. **22**, 437 (1969)

[13] J.M. Montoya and R.V. Solé, Santa Fe Institute, Technical Report No. 00-10-059, 2000 (unpublished); J.M. Montoya and R.V. Solé, e-print cond-mat/0011195 .

[14] R.V. Solé and J.M. Montoya, Santa Fe Institute, Technical Report No. 00-11-060, 2000 (unpublished); R.V. Solé and J.M. Montoya, e-print cond-mat/0011196.

[15] O. Sporns, Corrosion (Houston) Complexity **8**, 56 (2002).

[16] P. Erdős and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).

[17] P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).

[18] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).

[19] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **281**, 69 (2000).

[20] S. Redner, Eur. Phys. J. B **4**, 131 (1998).

[21] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[22] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).

[23] S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E **62**, 1842 (2000).

[24] S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E **63**, 056125 (2001).

[25] P.L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).

[26] P.L. Krapivsky, G.J. Rodgers, and S. Redner, Phys. Rev. Lett. **86**, 5401 (2001).

[27] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).

[28] J.H. Conway and R.K. Guy, *The Book of Numbers* (Springer-Verlag, New York, 1996), pp. 143 and 258–259.

[29] D.S. Callaway, J.E. Hopcroft, J.M. Kleinberg, M.E.J. Newman, and S.H. Strogatz, Phys. Rev. E **64**, 041902 (2001).

[30] H.S. Wilf, *Generatingfunctionology*, 2nd ed. (Academic Press, London, 1994).

[31] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, Phys. Rev. E **64**, 066110 (2001).