Agent-based model for friendship in social networks

H. M. Singer,* I. Singer,† and H. J. Herrmann‡
Computational Physics, IfB, Eidgenössische Technische Hochschule, CH-8093 Zürich, Switzerland
(Received 25 April 2008; revised manuscript received 7 June 2009; published 13 August 2009)

A model is proposed to understand the structuring of social networks in a fixed setting such as, for example, inside a university. The friendship formation is based on the frequency of encounters and mutual interest. The model shows distinctive single-scale behavior and reproduces accurately the measurable experimental quantities such as clustering coefficients, degree distribution, degree correlation, and friendship distribution. The model produces self-organized community structures and can be described as a network of densely interconnected networks. For the friendships, we find that the mutual interest is the dominant factor, which optimizes the network and that the number of encounters determines the statistically relevant distributions.

DOI: 10.1103/PhysRevE.80.026113 PACS number(s): 89.75.Hc, 89.65.Ef, 87.23.Ge

I. INTRODUCTION

In statistical physics complex networks have recently attracted a considerable interest [1,2]. Barabási and Albert (BA) [3] introduced a prototypical growing network model (BA model), which exhibits scale-free properties for the degree distribution $P(k)\sim k^{-3}$. The main ingredients for this model are growth and preferential attachment, which seem to be able to explain and describe various observations in social science. Preferential attachment corresponds to the Mathew effect that the "rich get richer." However, friendship is fundamentally different from the behavior of other social networks in that they are single-scale networks and show a small-world effect [4]. It has been observed that preferential attachment is sufficient to establish a power-law behavior in the growing model; however, for nongrowing networks with a constant number of nodes the degree distribution is unstable and converges to a Gaussian distribution upon saturation [1]. It was shown however that also in a fixed-setting network scale-free behavior can be found [5–8]. In these models random weights were introduced. González *et al.* [9] presented a model based on a system of moving particles, which by colliding form links between each other. While this model was able to recover degree distribution, clustering coefficients, and other quantities for a large database of empirical friendship networks, it has also two drawbacks: (i) it is not apparent why the simulations are carried out in a two-dimensional space and (ii) a collision in this system automatically leads to a "friendship," which is generally not the case with a random encounter.

In this paper we present a model for a friendship network, which reproduces known quantities of empirical networks such as degree distribution, clustering coefficients, and friendship distributions. Similar to Ref. [9], the community structure emerges naturally, without the need of prelabeling the community for each agent as in Ref. [10].

II. MODEL DESCRIPTION

Our model defines N_a agents with no initial connections. This setting is comparable to a large group of students enrolling at the same time in a college or a high school. At every step two agents are chosen to have an encounter. In the beginning the encounters are random; however, in time, due to preferential selection, eventually some agents form connections—friendships—which are based on two criteria: (i) the number of contacts with the same individual and (ii) the mutual interest. We have assumed here that friendship is reciprocal. An extension of the model to unidirectional (profiting) connections will be discussed elsewhere. In time, the number of contacts with different agents increases up to a largest number, which is an individual property of the agent. Further contacts are then only possible by "forgetting" previous contacts. This is similar to the concept of aging of Amaral *et al.* [4]. Despite the similarity, there is however an important difference: in Ref. [4] the agents were aging until they died. In contrast, here, the agent will accept new connections throughout the simulation, however, at the expense of dropping the old ones as will be shown below in detail.

Each agent i possesses several individual properties: (i) the maximal acquaintance parameter λ_i, which relates to the maximal possible number of other agents with which it can have contacts (in network terms this corresponds to the maximal possible degree of a particular node); (ii) the actual number of contacts with different agents k_i, which corresponds to the degree; and (iii) the "specific degree" or popularity $\Pi_i=k_i/\Sigma_j k_j$, which defines the probability to be selected for being added into another agent's contact list, where Π_i corresponds to the preferential attachment of the BA model. (iv) Each agent keeps a list of the encounters with other agents. This list contains the total number of encounters, n_{ij}, between agent i and any particular agent j. The list also contains the relative desire p_{ij} to meet a given agent j again.

The friendship f_{ij} between two agents i and j is defined as a function of the total number of contacts n_{ij},

$$f_{ij}=f(n_{ij})=1-e^{-n_{ij}}, \tag{1}$$

where we have chosen an exponential saturation, meaning that another agent will become a better friend the more often the two agents meet; however, after many encounters the total number does not play such an important role any more. A visual representation of such a friendship network with 100 agents is shown in Fig. 1. One could imagine that a

*hsinger@solid.phys.ethz.ch
†irina@mech.kth.se
‡hans@ifb.baug.ethz.ch

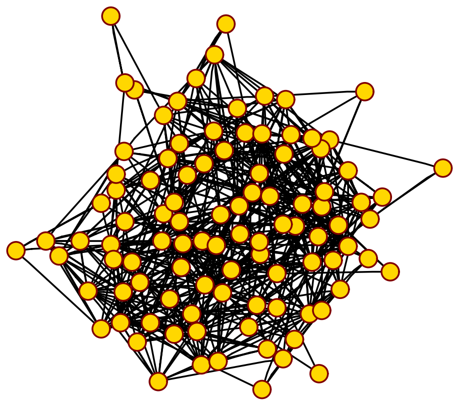

FIG. 1. (Color online) Friendship network of $N=100$ agents.

generalization like $f(n_{ij})=1-e^{-\eta n_{ij}}$, where η defines the number of times until the friendship reaches a value of $1-e^{-1}$ would be more appropriate. It turns out, however, that the results do not substantially differ from the ones obtained with Eq. (1). All further simulations were thus performed with Eq. (1).

Each relative probability for meeting already known contacts is calculated as $p_{ij}=f_{ij}/\Sigma_k f_{ik}$. The maximal acquaintance parameters λ_i are chosen to be normally distributed around one with a variance σ. λ_i represent the willingness of an agent i to make new connections. Not every agent is equally involved in this system (e.g., at a university) since it might possess some friends already, which are outside the system. We assume as a first approximation a normal distribution.

An agent can choose to meet either yet unknown agents or agents in the contact list. The probability for agent i to meet an agent in the contact list is given by

$$p = 1 - e^{-\lambda_i k_i}, \tag{2}$$

where k_i is the degree, i.e., the number of encounters with different other agents. Thus the probability to meet a yet unknown agent outside the contact list is $1-p$, which is decaying strongly with increasing size of the contact list.

Depending on the agent's λ_i, the contact list increases in time. The probability to meet outside agents decreases accordingly but never vanishes. However, in order to prevent an unbounded growth of the list, which would lead eventually to a fully connected network, a fixed threshold $\theta_\sigma = 1/\sqrt{4\pi}e^{-1/\sigma^2}$ was introduced, which is the largest number of contacts the agent can have. This threshold corresponds to the distance $1/\sqrt{\sigma}$ from the expectation value of the normal distribution. For $p>1-\theta_\sigma$, new contacts are not accepted any more. Thus the agent has built up its individual social community. In real life, this corresponds to the fact that the time to meet other students is limited and that therefore there is a natural limit on how many friends a particular person can have. Newly met agents are only admitted to the agent's list if they present an added value; then the new contact replaces an old one. This threshold prevents social isolation and accounts for beneficial random encounters since the probability to meet new agents does not drop below this threshold. As the contacts are chosen probabilistically, it may happen that an agent i tries to make an acquaintance with an agent j, which has already a full contact list. Agent j can then decide to reject the contact or conversely accept it in case that the possible friendship f_{ij} is larger than one of f_{jk}'s in the contact list of j. Then the old weak contact is dropped in favor of the new link with agent i.

The current choice of the functional form of Eqs. (1) and (2) is considered to be a first approximation since the exact relations are to the authors' knowledge not known. As will be shown, these approximations seem however to be surprisingly successful in describing the empirical data collected from social studies on friendship networks.

Intuitively, one finds that the number of encounters alone cannot be the only objective quantity, which defines friendship. Otherwise, the people we meet every day such as working colleagues, neighbors, schoolmates, newspaper agents, etc. would all be part of our best friends. While meeting people often naturally leads to a certain familiarity, friendships do not necessarily develop. Thus, we introduce a second characteristic—the affinity a_i of an agent i with $a_i \in [0,1]$—which summarizes the agent's fields of interest. The affinities can be distributed according to any kind of distribution P_a. An agent tends to optimize its friends in the contact list with respect to its own interest a_i while maintaining the maximal possible number of contacts.

Thus, in terms of networks, each node optimizes its interest under the constraint of a fixed degree distribution (nodewise optimization). We introduce, as a first approximation, a decaying interest match function that favors matching interests and penalizes differences in interests

$$f_m(a_i,a_j) = \frac{1}{1-e^{-1}}(e^{-|a_i-a_j|} - e^{-1}), \tag{3}$$

which is essentially a rescaled exponential decay, so that it becomes 1 for $|a_i-a_j|=0$ and 0 for $|a_i-a_j|=1$. By multiplying the friendship function with the match function

$$f_{ij} = f(n_{ij}) f_m(a_i,a_j), \tag{4}$$

we have introduced a friendship optimization. Hence, with time, every agent optimizes the contacts, which fit to its own taste, and local self-organized social communities of common interest naturally emerge.

III. RESULTS

A. Measurable network quantities

First we study the model independently on the interests a_i by using the friendship function of Eq. (1). Figure 2 shows the degree distribution for the averaged data of 84 schools in USA where questionnaires from 90 118 students were evaluated [11]. These data are compared to the model of Ref. [9] (thick lines) and the present model (thin line). The results for the present model have been averaged over 20 000 realizations for $N_a=1000$ and 500 realizations for $N_a=10\,000$ for $\sigma=2.8$. It can be seen that the calculated results fit the data much better than exponential or Poisson distributions and is in agreement with the calculations of Ref. [9]. However, the data of Ref. [9] fit the experiments only up to a degree of

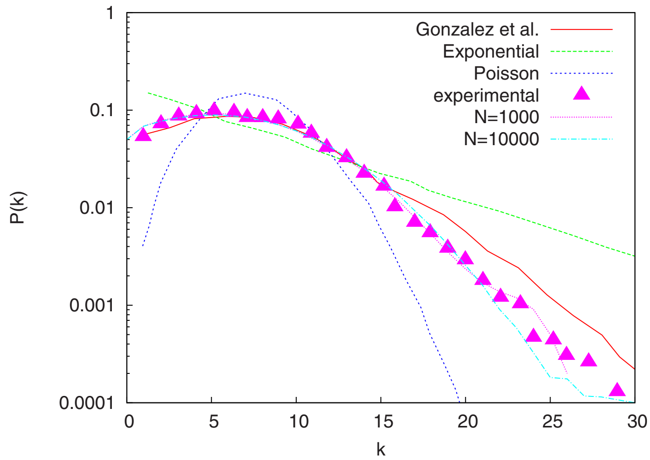

FIG. 2. (Color online) Degree distribution $P(k)$ vs k: experimental data (triangles), calculated data from the model of Ref. [9] (thick line), and the present model (fine lines) together with exponential and Poisson fits of the data. While the predictions of Ref. [9] fit the data well for $k<15$, a deviation is observed for larger k. The present model fits the data in the whole range accurately. The parameters used are $N=1000$ and 10 000 and $\sigma=2.8$.

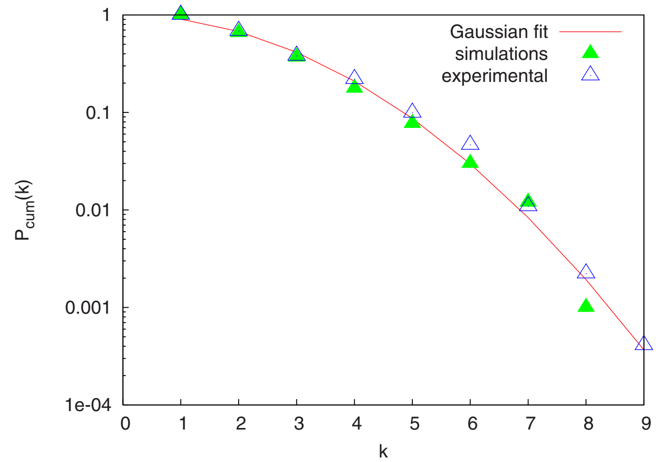

FIG. 4. (Color online) Friendship distribution normalized: connectivities for the friendship network of 417 high school students, where the number of links corresponds to the number of times a student is chosen by another student as one of his or her two best friends. The experimental data (triangles) were taken from Refs. [4,12]. The calculated data of the present model are superimposed. The distribution is Gaussian and is in excellent agreement with the experiments.

$k=15$; for higher degrees, a substantial deviation is observed. For the present model with $N_a=1000$ and $N_a=10\,000$, the whole range of experimental data is predicted accurately. The experimental data in Fig. 2 are averaged over all 84 high schools. In Fig. 3 the degree correlation $K_{nn}(k)$ is shown. It can be seen that the calculated results match well to the experimental ones.

In Fig. 4 the normalized cumulative friendship distribution is plotted against the total number of being chosen to be a friend. The experimental data of the friendship network of 417 Madison Junior High School students were taken from Refs. [4,12]. The connectivity distribution shows no power-law regime but can be fitted well with a Gaussian distribution showing the single-scale character of the network. The number of links in this network corresponds to the number of times a student was chosen by another student as one of his or her two best friends. The simulations with the present model indicate a Gaussian distribution as well. The simulated results are in accordance with the experimental data.

In Fig. 5 the clustering coefficient of the experimental data of the 84 high schools is plotted as a function of the average degree. The thick solid line indicates the results obtained with the proposed model using the same range of values of $\langle k\rangle$ averaged over 50 realizations. The parameters used are $N_a=800$ and $\sigma\in[1,4]$. For comparison the data obtained by González *et al.* [9] using the mobile agent approach are plotted as well. It can be seen clearly that both models reproduce well the clustering coefficient within the error bars for the same average degree.

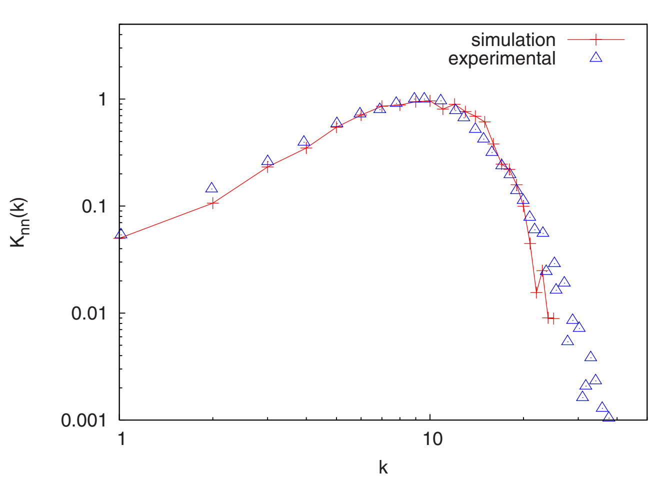

FIG. 3. (Color online) Degree correlation $K_{nn}(k)=\Sigma_{k'}P(k|k')k'$ from experimental data and simulations with the present model. The parameters used are $N=1000$ and $\sigma=2.8$.

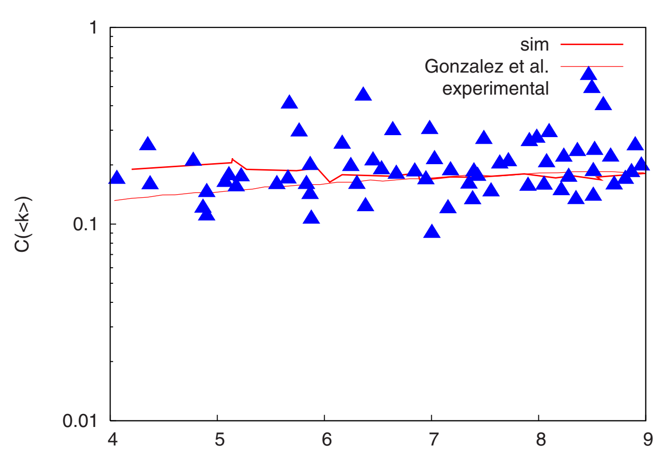

FIG. 5. (Color online) Clustering coefficient of the 84 high schools as a function of the average degree. Superimposed are the results of González *et al.* [9] and the results of the present model. The clustering coefficient is well predicted by the model.

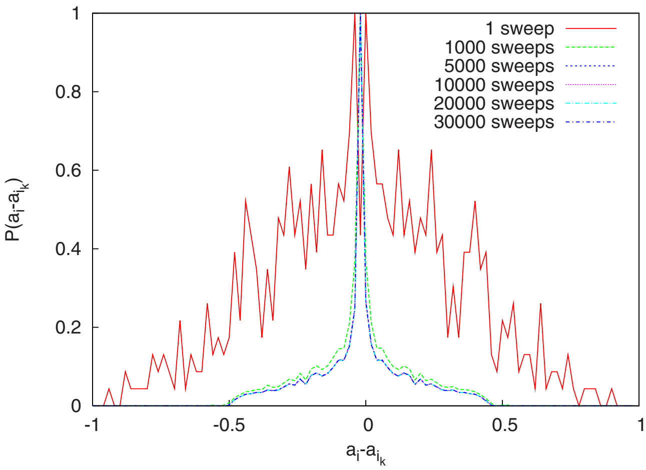

FIG. 6. (Color online) Probability distribution of $a_i - a_{i_k}$ for all agents i: the friendships are optimized according to Eq. (4).

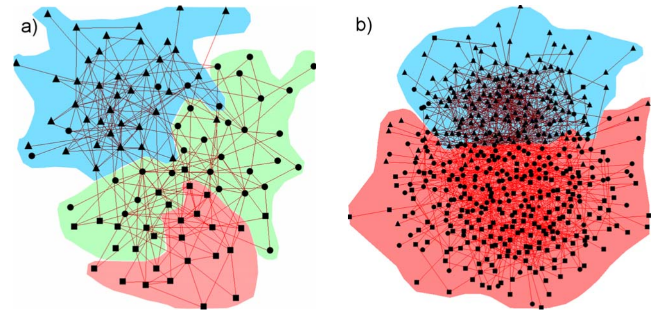

FIG. 7. (Color online) Network after 30 000 sweeps for (a) $N_a = 100$ and (b) $N_a = 500$. The symbols correspond to the interest $a_i \in [0,1]$, where $a_{i,\blacktriangle} \in [0, \frac{1}{3})$, $a_{i,\bullet} \in [\frac{1}{3}, \frac{2}{3})$, and $a_{i,\blacksquare} \in [\frac{2}{3}, 1]$. The layout has been calculated with the Kamada-Kawai algorithm. It can be seen that the nodes arrange in communities of common interest. Superimposed are the results of the community structure detection of the Girvan-Newman algorithm [14]. The communities of interest match well with the ones found by the Girvan-Newman algorithm.

B. Interest—network optimization

The model so far reproduces all the statistically measurable quantities, which can be determined from a purely statistical approach. Now, we study the influence of the affinities in the friendship by using the friendship function defined in Eq. (4).

For the sake of simplicity, we have assumed here a uniform distribution of $a_i \in [0,1]$; however, the results seem to hold for arbitrary distributions P_a. In Fig. 6 the friendship optimization of Eq. (4) is shown for different numbers of sweeps for the parameters $\sigma = 2.8$ and $N_a = 1000$. The decay of the function f_m of Eq. (3) determines the time scale at which the agents optimize their friendships. Any monotonically decreasing function will eventually optimize. Figure 6 also shows that the initial distribution is close to triangular as expected for friendship independent of a_i, which indicates that the optimization sets in only once some of the agents have a full contact list. In the beginning of the simulation, it is more favorable to meet many new contacts than to replace existing contacts. Once all agents have a fully occupied contact list, optimization is the only way to change the network. Thus, two regimes can be distinguished: (i) creation of the network, where all agents meet new agents mostly independently on their mutual interest, since every agent still has free contact capacities and (ii) in the subsequent regime most or all agents have a filled contact list, with not necessarily favorable contacts. In the search of optimizing the individual community, the agents start replacing existing contacts with more valuable ones until eventually an optimum network structure is found. In comparison with filling up contact lists by meeting arbitrary agents, the optimization process is much slower. Thus, as in real life, meeting many people is easy and is usually a quick process. However, sieving through the contacts to identify and cultivate new friendships takes much more time.

C. Community structure

In Figs. 7(a) and 7(b) a network with $N_a = 100$ and 500 agents is shown, respectively. The layout has been calculated by the Kamada-Kawai algorithm [13], which connects the agents by springs, whose interaction force is proportional to the shortest path in the network. The positions of the individual agents are calculated by finding the minimum-energy configuration of the spring system. The affinities a_i are plotted with different symbols. For better visibility only three categories of interest have been chosen: $a_i \in [0, \frac{1}{3})$, $[\frac{1}{3}, \frac{2}{3})$, and $[\frac{2}{3}, 1]$. As can be seen the communities that formed after 30 000 sweeps separate the triangle, circle, and square symbol agents rather well. Superimposed are the results of a conventional community detection algorithm (Girvan-Newman [14]). The detection led to only three respective two major communities, which separate the groups of triangles, circles, and squares in Fig. 7(a) and the groups of triangles and circles-squares in Fig. 7(b). The community boundaries are in accordance with the communities formed by interest; however, the latter creates a much finer community detection and separation, in particular with a finer binning of agents.

IV. DISCUSSION AND CONCLUSIONS

We introduce a model based on nonmoving agents, who build up connections based on preferential attachments in the beginning, and—later when the contact lists are filled up—on the emerging social community stored in the contact list. This model seems to build a bridge between the results obtained by the mobile agent calculations of Ref. [9] and the more traditional network structure models, in that it reproduces all the experimental results and is inherently single scale. Yet it is composed of standard elements and tools commonly used in the field of social networks. It is especially worth noting that the later stage development of the network is essentially a densely interconnected set of BA networks which, as a whole, show static converged distributions. In this sense it can be considered as a network of interconnected networks, which reproduces accurately experimental data.

The present model remedies some inconveniences of the mobile agent model. In particular no spatial topology is imposed and the notion of friendship is clearly defined by Eq. (4) and distinguished from simple encounters. Our model is capable of reproducing experimental data and is in excellent agreement with measured degree distributions, clustering coefficients, and friendship distributions. The choice of the functional form of Eqs. (1) and (2) was successful in describing empirical results. Future studies will concentrate on the robustness of the results on variations in these choices to establish ultimately the broad validity of this friendship model.

Friendship emerges from a local optimization process, which takes place under a constant degree distribution. The emerging communities based on the interest are in accordance with conventional community detection algorithms.

The simulations show that the buildup of acquaintances is a much faster process than finding friends: one possible network is created which fills up all the possible contact lists. Optimization of the friends takes place only after this initial stage. Thus, finding good friends is much harder than acquiring acquaintances. Agents are forced to find friends in the environment they are placed in. If the selection is too limited, no friends or rather in absolute values of a particular agent only "bad friends" are found. The total number of good friends is directly related to the number of acquaintances an agent can have since it increases the probability to meet other agents. In particular the number of encounters provides the statistical framework for the comparison with experimental data, whereas the mutual interest optimizes the network to form matching community structures.

[1] E. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[2] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

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

[4] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).

[5] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).

[6] B. Söderberg, Phys. Rev. E **66**, 066121 (2002).

[7] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).

[8] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).

[9] M. C. González, P. G. Lind, and H. J. Herrmann, Phys. Rev. Lett. **96**, 088702 (2006).

[10] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002).

[11] Data Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris and funded by a grant from the National Institute of Child Health and Human Development (P01-HD31921).

[12] T. J. Fararo and M. H. Sunshine, *A Study of a Biased Friendship Net* (Syracuse University Press, Syracuse, NY, 1964).

[13] T. Kamada and S. Kawai, Inf. Process. Lett. **31**, 7 (1989).

[14] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).