

System of Mobile Agents to Model Social Networks

Marta C. González,^{1,2} Pedro G. Lind,^{1,3} and Hans J. Herrmann^{1,2}

¹*Institute for Computational Physics, Universität Stuttgart, Pfaffenwaldring 27, D-70569 Stuttgart, Germany*

²*Departamento de Física, Universidade Federal do Ceará, 60451-970 Fortaleza, Brazil*

³*Centro de Física Teórica e Computacional, Avenida Professor Gama Pinto 2, 1649-003 Lisbon, Portugal*
(Received 20 September 2005; revised manuscript received 23 December 2005; published 3 March 2006)

We propose a model of mobile agents to construct social networks, based on a system of moving particles by keeping track of the collisions during their permanence in the system. We reproduce not only the degree distribution, clustering coefficient, and shortest path length of a large database of empirical friendship networks recently collected, but also some features related with their community structure. The model is completely characterized by the collision rate, and above a critical collision rate we find the emergence of a giant cluster in the universality class of two-dimensional percolation. Moreover, we propose possible schemes to reproduce other networks of particular social contacts, namely, sexual contacts.

DOI: 10.1103/PhysRevLett.96.088702

PACS numbers: 89.65.Ef, 02.50.Le, 64.60.Ak, 89.75.Hc

Friendships among a group of people, actors working in the same movie or coauthors of the same paper, are all examples of systems represented as networks, and the study of such systems have imprinted to social networks an unquestionable place in the field of complex networks [1,2]. However, the topological features of networks of acquaintances fundamentally differ from other networked systems [2,3]. First, they are single-scale networks and present small-world effect [4]. Second, they are divided into groups or communities [2]. Additionally, their evolution process differs from standard growth models as those that govern, e.g., the World Wide Web. An interesting development in this area is given in [5], where a simple procedure of transitive linking to generate small-world networks is proposed. While each of the mentioned features can be reproduced with some previous model, there is still no single model that incorporates simultaneously dynamical evolution, clustering, and community structure.

In this Letter we show that all these characteristics can be reproduced in a very natural way, by using standard concepts and techniques from physical systems. Namely, we propose an approach to dynamical networks based on a system of mobile agents representing the nodes of the network. We will show that, due to this motion, it is possible to reproduce the main properties [1,2] of empirical social networks, namely, the degree distribution, the clustering coefficient (CC), and the shortest path length, by choosing the same average degree measured in the empirical networks and adjusting only one parameter, the density of the system. The community structure emerges naturally, without labeling *a priori* the community each agent belongs to, as in previous works [6]. Moreover, this approach gives some insight to further explain the structure of empirical networks, from a recently available large data set of friendship networks [7] concerning 90 118 students, divided among 84 schools from the USA, constructed from an in-school questionnaire. The acquaintance between pairs of students was rigorously defined. Each student

was given a paper-and-pencil questionnaire and a copy of a list with every student in the school. The student was asked to check if he/she participated in any of 5 activities with the friend: like going to his/her house in the last seven days, or meeting him/her after school to hang out or go somewhere in the last seven days, etc. Other studies [4] have used a slightly different definition of friendships and obtained the same kind of degree distribution, an indication of the robustness of the concept of friendship.

Our model comprehends N particles (agents) with radius r moving continuously in a square shaped cell of linear size L with periodic boundary conditions and low density $\rho \equiv N/L^2$. One link (acquaintance) is formed whenever two agents intercept. After each collision, each colliding agent moves in a random direction with an updated velocity, till it collides again acquiring a new random direction, and so forth. In this way, the resulting movement alternates between drift (between collisions) and diffusion (collisions). Similarly to human communities, agents arrive and depart after a certain time of residence, the total number of agents remaining fixed in time, which enables the system to reach a quasistationary state. Initially all agents are placed randomly, with the same velocity modulus v_0 and random directions. At each time step Δt , the position $\mathbf{x}_i$ of agent i is updated according to

$$\mathbf{x}_i(t + 1) = \mathbf{x}_i(t) + \mathbf{v}_i(t)\Delta t. \quad (1)$$

After collisions, the velocity modulus of each agent, say, i , is updated proportionally to its degree k_i , defined as the number of links connected to an agent i at time t :

$$|\mathbf{v}_i(t)| = v_0 + \bar{v}k_i(t), \quad (2)$$

where $\bar{v}$ is a constant having unit of velocity and v_0 is the initial velocity of the agents, corresponding to a characteristic time $\tau_0 \equiv 1/(2\sqrt{2}\pi r\rho v_0)$ between collisions. We assume that “age” $\mathcal{A}_i$ is the only intrinsic property of each agent i , initially randomly and homogeneously

chosen from an interval $[0, T_l]$, and updated as

$$\mathcal{A}_i(t+1) = \mathcal{A}_i(t) + \Delta t. \quad (3)$$

When $\mathcal{A}_i = T_l$, agent i leaves the system, all its links are removed, and a new agent replaces its position with the initial conditions stated above, namely, velocity modulus v_0 and an age randomly distributed in the range $[0, T_l]$. Therefore the time of permanence of an agent in the system is given by $T_l - \mathcal{A}_i(0)$.

After a certain transient the system reaches a quasistationary (QS) state. Thus, the degree distribution, degree correlations, and community structure depend only on two parameters, namely, ρ and T_l/τ_0 . Figure 1(a) illustrates the convergence towards the QS state for the average degree $\bar{k}(t)$ per agent.

In Fig. 1(b) we show the degree per agent $\langle k \rangle$ vs T_l/τ_0 . For each value of T_l/τ_0 the average degree was averaged over different snapshots in the QS regime, yielding a non-linear function of T_l/τ_0 , which depends on the chosen density. An approximate analytical treatment of this dependence can be made and will be presented elsewhere. Further, the average degree is a function of the average number λ of collisions during the average residence time $T_l - \langle \mathcal{A} \rangle$, and is defined as

$$\lambda \equiv \frac{1}{v_0 \tau_0} \langle v \rangle (T_l - \langle \mathcal{A} \rangle). \quad (4)$$

As illustrated in the inset of Fig. 1(b), we find $\langle k \rangle = \lambda/2$ (solid line), independently of the density.

In the presented model, we find a critical value $\lambda_c = 2.04$, beyond which a giant cluster of connected nodes emerges. Table I shows the values obtained numerically with the standard method of finite size scaling for systems of $N = 2^{10}, \dots, 2^{16}$; the results are compared with exponents for mean field and two-dimensional (2D) percolation. Since the agents move on a 2D plane and have only a finite life time, they can establish connections only within a

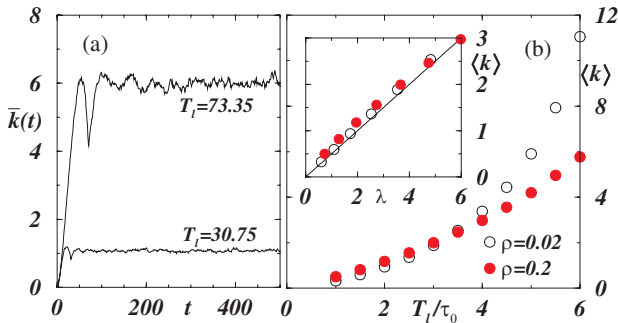


FIG. 1 (color online). (a) Average degree $\bar{k}$ per agent as a function of time t , illustrating the convergence towards a QS state ($N = 4096$). (b) Average degree $\langle k \rangle$ vs T_l/τ_0 for $N = 10^4$, averaged over 100 realizations. Inset: linear dependence between $\langle k \rangle$ and λ (see text); the solid line indicates $\langle k \rangle = \lambda/2$. In all cases, $v_0 = \sqrt{2}$ and $\bar{v} = 1$.

restricted vicinity. This effect corresponds to a connectivity which is short ranged at each snapshot of the system. So, although our clusters are not quenched in time, the underlying problem corresponds to short range 2D percolation. We have also explicitly calculated the correlation length as the linear size of clusters, and confirm that near the critical point this quantity diverges with precisely the same exponent ν obtained from the finite size scaling.

The degree distribution $P(k)$ is a direct consequence of the collision rule; i.e., it depends on $\bar{v}$ in Eq. (2). For $\bar{v} = 0$, the degree distribution is well fitted by a Poisson distribution, $P_p(k) = (\langle k \rangle^k / k!) \exp(-\langle k \rangle)$. The degree distribution obtained for $\bar{v} = 1$, resembles an exponential of the form $P_e(k) = (\langle k \rangle - 1)^{-1} \exp[-(k - 1)/(\langle k \rangle - 1)]$. However, while for small $\langle k \rangle$ the degree distribution of the giant cluster is exponential of the form of $P_e(k)$, for larger $\langle k \rangle$ it deviates from this shape. The same deviation as $\langle k \rangle$ increases is in fact found in empirical data, e.g., the friendship networks of the 84 schools. For each of the schools, Fig. 2(a) shows the average shortest path length l (circles) and the CC (triangles). The solid lines indicate the results obtained for the agent model using the same range of values of $\langle k \rangle$, averaged over 100 realizations with $N = 2209$ and $\rho = 0.1$. Since l depends on the network size, it is divided by the shortest path length l_0 of a random graph with the same average degree and size. Clearly, the agent model predicts accurately both the CC and the shortest path length for the same average degree.

By computing the average degree $\langle k \rangle$ of each school, one is able to obtain the value of T_l/τ_0 for which the agent model reproduces properly the empirical data, as illustrated in Fig. 2(b). Here the solid line indicates the prediction curve for the agent model, while triangles indicate the values of T_l/τ_0 chosen to reproduce the social network of the schools with the resulting value of $\langle k \rangle$. Moreover, the second moment $\langle k^2 \rangle_{\text{ag}}$ obtained with the simulations of the agent model is a rescaling of the same quantity $\langle k^2 \rangle_{\text{Sch}}$ measured for the empirical school networks, as shown in Fig. 2(c).

Figure 3(a) shows the degree distribution averaged over all the schools, compared with the average of the ones obtained from the agent model simulations using the chosen values of T_l according to the relation sketched in Fig. 2(b). As one clearly sees, the degree distribution obtained with the agent model fits much better the empiri-

TABLE I. Critical exponents related to the emergence of the giant cluster for the network of mobile agents, compared to the ones of mean field and 2D percolation.

	Mean field	2D percolation	Mobile agents
ν	0.5	$4/3 \sim 1.33$	1.3 ± 0.1
γ	1	$43/18 \sim 2.39$	2.4 ± 0.1
β	1	$5/36 \sim 0.139$	0.13 ± 0.01
σ	0.5	$36/91 \sim 0.397$	0.40 ± 0.01

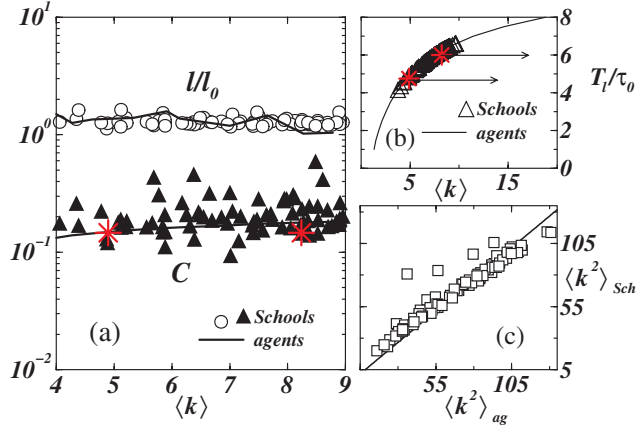


FIG. 2 (color online). (a) Average shortest path length l and clustering coefficient C as functions of the average degree $\langle k \rangle$. Empirical data (symbols) compared to simulations (solid lines). (b) Plot of T_l/τ_0 as a function of $\langle k \rangle$ for the agents models (solid line). Stars illustrate two particular schools for Figs. 3 and 5, having $T_l/\tau_0 = 4.75$ (school 1) and 6.0 (school 2), respectively. (c) Second moment $\langle k^2 \rangle$ for each school vs the second moment of the corresponding simulation with the agent model (solid line has slope 1).

cal data than the exponential (dotted line) or Poisson (dashed line) distributions for a given $\langle k \rangle$. The inset in the Fig. 3(a) shows the comparison of the network of one particular school (school 1 in Fig. 2), and the average over 20 realizations of its corresponding model (with $T_l/\tau_0 = 4.75$).

Degree correlations can be quantified by computing $K_{nn}(k)$, the average degree of the nearest neighbors of a

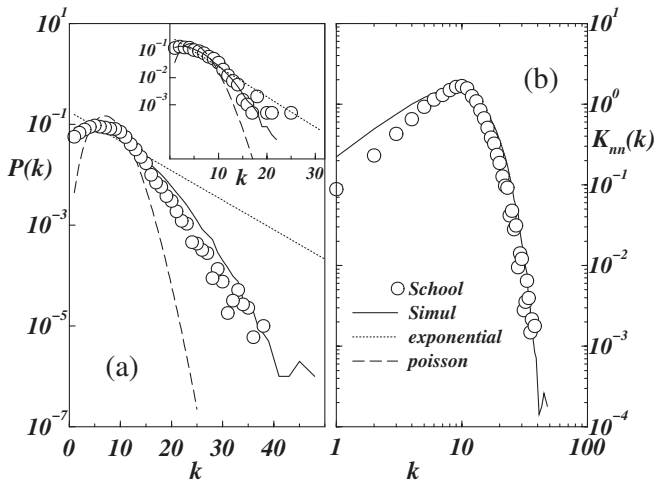


FIG. 3. (a) Degree distribution $P(k)$ averaged over all the schools (symbols) compared to $P(k)$ of the simulations (solid line). The inset shows the results for a particular school (school 1). (b) Average degree K_{nn} of the nearest neighbors as a function of k . Dashed and dotted lines indicate the Poisson and exponential distributions, respectively, for the same average degree $\langle k \rangle$.

vertex of degree k [3]. Figure 3(b) shows good agreement of this value between real data and the model for the same networks of Fig. 3(a). Similar to other social networks the mixing is assortative [2]; i.e., K_{nn} increases with k , but in contrast to networks with scale free degree distribution (i.e., collaboration networks), $K_{nn}(k)$ for friendship networks present a cutoff due to the rapid decay in the degree distribution.

Further, the typical community structure found in social networks is also reproduced with the agent model. Here, we use a precise definition of network community recently proposed [8] based on the concept of k -clique community. In Fig. 4 we plot the system of mobile agents, drawing only the trajectories of the agents which belong to two 3-clique communities, having 4 and 10 agents and sketched in Figs. 4(b) and 4(c) respectively. Agents that form a community share a region in space, and agents with larger trajectories are responsible for building up the community. It should be pointed out that the agent motion in the system has not the straightforward meaning of human motion in physical space, but may be better related with affinities among individuals.

Figure 5(a) shows the size distribution of 3-clique communities in a particular school (school 2) compared with the simulation for the suitable value of T_l/τ_0 (see Fig. 2), while in Fig. 5(b) the average over all schools is compared with the average over 10 realization of the corresponding model for each school. In both cases, the agent model reproduces the distribution of community size observed for the empirical data, particularly the feature related with the existence of a big community having a large fraction of the population, namely, $s \sim 10^3$ agents.

In the particular case of sexual contacts, it has been reported that the degree distribution presents a power law [9]. Figure 6 shows with triangles the cumulative degree distribution of a sexual contact network extracted from a tracing study for HIV tests in Colorado Springs (USA) with 250 individuals [10]. The dashed line indicates the degree

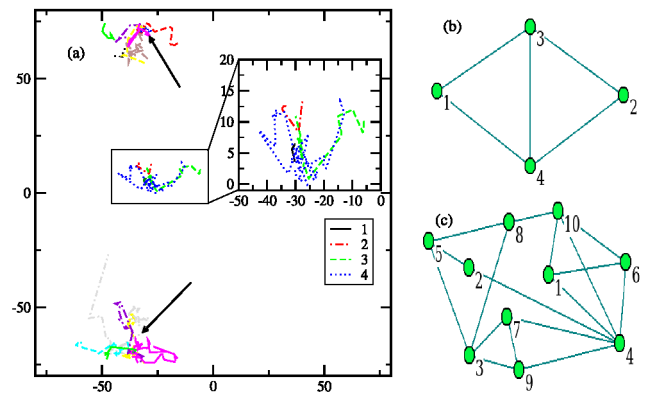


FIG. 4 (color online). (a) Example of trajectories of 4 agents (enclosed in a box and enlarged in the inset) and 10 agents (showed by arrows) forming a 3 clique, sketched in (b) and (c).

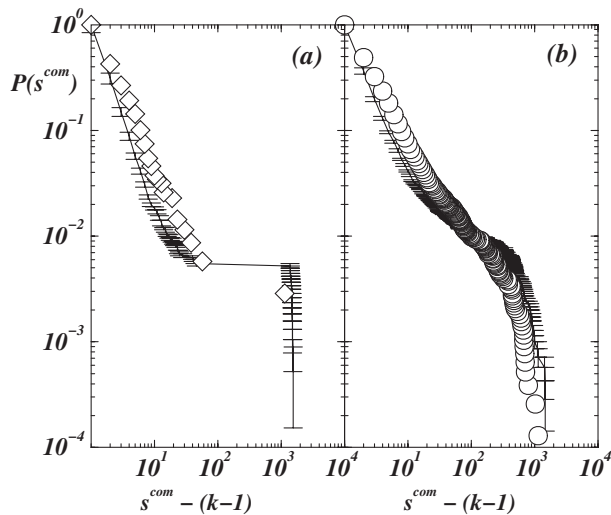


FIG. 5. (a) Distribution of community size s of 3-clique communities for one particular school (school 2); (b) the corresponding average over the 84 schools of the data set. Empirical data (symbols) compared to simulations (solid lines with error bars).

distribution of a social contact network simulated with the agent model, while the solid line is the degree distribution of a subset of contacts from the social network. The contacts in the subset are chosen by assigning to each agent an intrinsic property which enables one to select from all the social contacts the ones which are sexual. Namely, when two agents form a link, as stated before, this link is now marked as a “sexual contact” if the sum of the property values of the two agents is greater than a given threshold. These property values are assigned to the agents with an exponential distribution and the conditional threshold is $\ln N/2$, following the scheme of intrinsic fitness proposed in another context by Caldarelli *et al.* [11]. Interestingly, one is able to extract from the typical distributions of social contacts shown throughout this Letter, power-law distributions in QS states which resemble much the ones observed in real networks of sexual contacts.

In conclusion, we presented a novel approach to construct contact networks, based on a system of mobile agents. For a suitable collision rule and aging scheme, we have shown that one is able to produce quasistationary states which reproduce accurately the main statistical and topological features observed in recent empirical social networks. The QS state of the agent model is fully characterized by one single parameter and yields a phase transition belonging to the universality class of two-dimensional percolation. Moreover, we showed that, by introducing an additional property labeling the ability to select a particular type of social contact, e.g., sexual contacts, the degree distributions reduce to power-law distri-

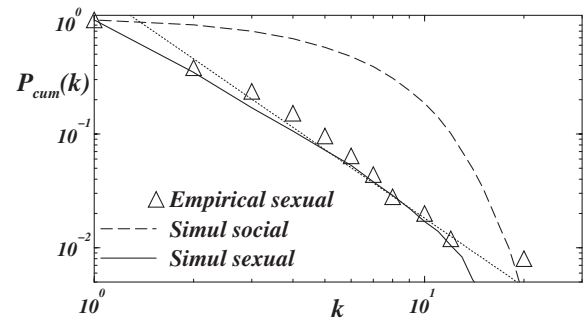


FIG. 6. Cumulative degree distribution of the number k of sexual partners in a real empirical network of sexual contacts (triangles) with 250 individuals, compared with the simulation of the agent model (solid line); the dotted line is a guide to the eye with slope 2. Here $N = 4096$, $T_1/\tau_0 = 5.5$, and $\langle k \rangle = 7.32$ and the average size of the resulting sexual network is 220.

butions as observed in real sexual networks. Summarizing, we gave evidence that the motion of the nodes is a fundamental feature to reproduce social networks, and therefore the above model could be important to improve the study and may serve as a novel approach to model empirical contact networks.

The authors would like to thank J. Kertész, J.S. Andrade, and M. Barthélémy for useful discussions. M.C.G. thanks DAAD (Germany), and P.G.L. thanks FCT (Portugal) for financial support.

-
- [1] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [2] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
 - [3] M. Boguñá, R. Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas, *Phys. Rev. E* **70**, 056122 (2004).
 - [4] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11 149 (2000).
 - [5] J. Davidsen, H. Ebel, and S. Bornholdt, *Phys. Rev. Lett.* **88**, 128701 (2002).
 - [6] D. J. Watts, P. S. Dodds, and M. E. J. Newman, *Science* **296**, 1302 (2002).
 - [7] This research uses data from 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).
 - [8] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, *Nature (London)* **435**, 814 (2005).
 - [9] F. Liljeros, C. R. Edling, L. A. N. Amaral, and H. E. Stanley, *Nature (London)* **411**, 907 (2001).
 - [10] J. J. Potterat *et al.*, *Sex. Transm. Inf.* **78**, Suppl. 1, I159 (2002).
 - [11] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).