Epidemic spreading in a hierarchical social network

A. Grabowski[1,*] and R. A. Kosiński[1,2,†]
[1] *Central Institute for Labour Protection–National Research Institute, 00-701 Warsaw, Poland*
[2] *Faculty of Physics, Warsaw University of Technology, 00-662 Warsaw, Poland*
(Received 5 April 2004; published 21 September 2004)

A model of epidemic spreading in a population with a hierarchical structure of interpersonal interactions is described and investigated numerically. The structure of interpersonal connections is based on a scale-free network. Spatial localization of individuals belonging to different social groups, and the mobility of a contemporary community, as well as the effectiveness of different interpersonal interactions, are taken into account. Typical relations characterizing the spreading process, like a range of epidemic and epidemic curves, are discussed. The influence of preventive vaccinations on the spreading process is investigated. The critical value of preventively vaccinated individuals that is sufficient for the suppression of an epidemic is calculated. Our results are compared with solutions of the master equation for the spreading process and good agreement of the character of this process is found.

DOI: 10.1103/PhysRevE.70.031908 PACS number(s): 87.10.+e, 89.75.−k, 05.40.−a

I. INTRODUCTION

The structure and the dynamics of complex networks have been extensively investigated in recent years [1–11]. It was found that many real-world networks, like the web of human sexual contacts [1], e-mail networks [2], or the internet [3], have similar properties. They are called scale-free networks, because the probability that the number of k links connected to a node equals $P(k) \sim k^{-\gamma}$ [4]. Many authors have used this type of complex network to model a network of social contacts [9,12–15]. In particular, complex networks with a hierarchical structure, corresponding to the real structure of human communities, have been studied [7,16–18], e.g., an epidemic spreading in a population with a two-level structure of interpersonal interactions was analyzed in Ref. [19]. A small average shortest path between nodes (individuals) and a high value of the clustering coefficient [5,6], e.g., the probability that "a friend of my friend is my friend" in the community is high, are the most important properties of social networks. These properties are typical for the structure of a social network and they have a strong influence on dynamical phenomena in the population.

In our work we investigate an epidemic spreading in the human population, treated as a scale-free network, taking into account spatial localization of individuals, with a three-level hierarchical structure of interpersonal interactions.

We assume that each individual belongs to some social groups [7]: from the smallest one (e.g., family or friends), to a large one (e.g., a community of the whole city). Interpersonal interactions among individuals in the same group are stronger than interactions among individuals from different groups. The smaller the group, the stronger an individual's influence on the other individuals in that group. From the point of view of the spreading of an epidemic, most effective are social connections with the family, close friends, etc.; however, random contacts with unknown individuals are important too. Such a random contact is most probable for individuals, who live (or work) in the same place, e.g., in the same building. On the other hand, a contemporary community is very mobile; therefore there is a nonzero probability of contact between two arbitrarily chosen individuals from a population. Such contact can occur, e.g., commuting, in the cinema or in another public place, and can result in infection of a new individual. In our model we take into account this hierarchical structure of a social network, with interpersonal connections between neighbors and contacts between random individuals referring to the mobility of a community.

The spreading of epidemics has been investigated by many authors with different models of interpersonal interactions [20–27]. The hierarchical structure of interpersonal interactions described in the present paper seems to be more plausible for modeling real social networks.

II. THE MODEL

In our model each individual is in one of four permitted states: healthy and susceptible (H), infected (I), ill (S), healthy and unsusceptible or dead (R). The state of the individuals evolves in time and depends on their previous state and the connections or random contacts with other individuals. The probabilities of transitions between different states in one time step are described with the following parameters: $W_{H\to I}$, the probability that a susceptible individual will be infected by an ill individual (it also denotes how contagious the disease is); $W_{I\to S}$, the probability that the infected individual becomes ill (this value is connected with the average time of incubation); $W_{S\to R}$, the probability that an ill individual will recover, die, or be isolated from the rest of the population (e.g., in a hospital).

The spreading process in a population can be treated as a nonstationary process, which is described by a master equation, and that approach was applied in a number of studies [14,23,28–32]. The results obtained in our model will be compared with the solutions of this equation in Sec. IV. For

*Electronic address: atomic@olimp.if.pw.edu.pl
†Electronic address: rokos@ciop.pl

the present case the changes in time of the probabilities $P_X(t)$ that an individual is in one of the possible states X (where $X=H$, I, S, or R) are described with the master equation

$$\frac{dP_H(t)}{dt} = -W_{H\to I}P_S(t)P_H(t),$$

$$\frac{dP_I(t)}{dt} = W_{H\to I}P_S(t)P_H(t) - W_{I\to S}P_I(t),$$

$$\frac{dP_S(t)}{dt} = W_{I\to S}P_I(t) - W_{S\to R}P_S(t),$$

$$\frac{dP_R(t)}{dt} = W_{S\to R}P_S(t). \tag{1}$$

This simple analytical model has one serious disadvantage—it does not take into account the structure of interpersonal interactions in the human population which is included in the present paper. In our model the population and its structure are described as follows. The population consists of $N=L\times L$ individuals S_{ij} localized by the indices i,j in a two-dimensional lattice. Connections and random contacts between individuals have a hierarchical structure. The connections of each individual with k neighbors is the first level of interpersonal interactions [see Fig. 1(a)]. All connections are symmetrical and have the same value. We have assumed that the network of social connections is scale-free, i.e., the distribution of connectivity of individuals has the form $P(k)\sim k^{-\gamma}$ ($\gamma=3$ was used in most of computations), with k generated from the range $(k_{\min},k_{\max})$. Initially all individuals are not connected. Next, connections between individuals are created with the probability $P(l)$, depending on the distance l between individuals S_{ij} and S_{nm}, where $n=i\pm l_1, m=j\pm l_2$ (l_1,l_2 are two independent random variables and the sign is generated with the probability 0.5):

$$P(l) \sim \frac{1}{1+\exp[(l-a)/b]} + 0.01\frac{L-l}{L}. \tag{2}$$

The second term in Eq. (2) causes $P(l)$ to reach zero slowly enough. The whole population is divided into local groups of $G=L_G\times L_G$ individuals, where the size of those groups is connected by the parameters $a=L_G$ and $b=L_G/4$ of the distribution (2). Thus, most connections are created between individuals located in the same local group. The structure of the network from the point of view of a certain individual is depicted in Fig. 1(a). Having created the connection between S_{ij} and S_{nm}, the connections between the individual S_{nm} and each neighbor of the individual S_{ij} are created with the probabilities p_c [Fig. 1(b)]. Similarly, new connections between S_{ij} and the neighbors of the individual S_{nm} are created, also with the probabilities p_c. However, each pair of individuals can be connected only once, and a new connection is added to each individual only when its actual number of connections is smaller than the value k_{ij} (where $i,j=1, 2, \ldots, L$) generated with the distribution $k^{-\gamma}$. In this way a desirable distribution of connectivity is obtained. It should be noted that in this procedure the value p_c influences the clustering coefficient C of the network [4,33]

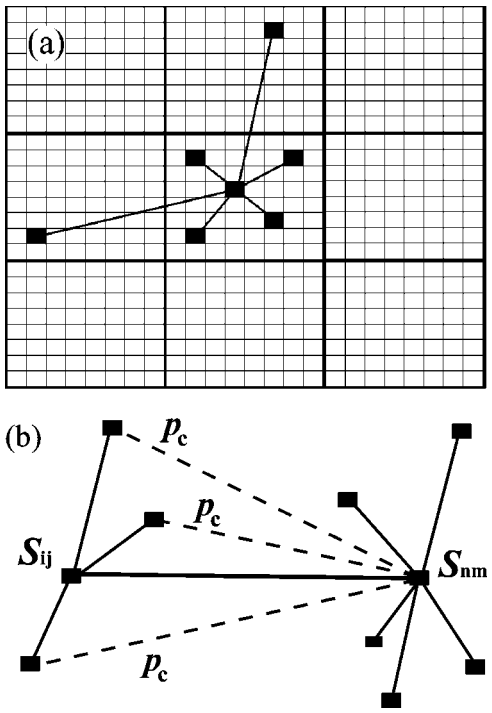

FIG. 1. An example of a network with $L=24$ and $L_G=8$ (nine local groups) from the point of view of the $S_{12,12}$ individual, who is connected with $k_{12,12}=6$ neighbors and four of whose connections are located in its local group (a). When a connection between two individuals S_{ij} and S_{nm} is created (solid line), the individual S_{nm} is connected with the neighbors of the individual S_{ij}. Next, the connections between the neighbors S_{nm} and the individual S_{ij} are created (this is not shown in the figure). Each new connection (dashed line) is created with the probability p_c (b).

$$C = \left\langle \frac{2E_{ij}}{k_{ij}(k_{ij}-1)} \right\rangle \tag{3}$$

where E_{ij} is the number of connections between neighbors of the ijth individual.

Let us describe the effectiveness of first-level connections in the spreading of epidemics. The probability of an infection of an individual by one of k neighbors is

$$p_1 = W_{H\to I}\sqrt{\frac{k_S}{k}} \tag{4}$$

where k_S is the number of neighbors in the state S. This probability is a nonlinear function of the number of ill neighbors and it increases fast for a low value of k_S, because interpersonal connections are a more effective way of spreading the epidemic than random contacts.

The second level of interpersonal interactions is random contacts between individuals in the same local group of G individuals. They are most probable for individuals living (or

working) in the same place, e.g., in the same building. In our model the probability of infection resulting from such a random contact is

$$p_2 = W_{H\to I}\left(\frac{G_S}{G}\right)^2 \tag{5}$$

where G_S is the number of ill individuals in a local group.

The third level of interpersonal interactions is random contacts between pairs of individuals who do not know each other and are chosen arbitrarily from the whole population. The probability of infection caused by such a contact does not depend on the localization of the individuals:

$$p_3 = W_{H\to I}\left(\frac{N_S}{N}\right)^2 \tag{6}$$

where N_S is the number of ill individuals in the whole population. The nonlinear factors in Eqs. (5) and (6) cause the probabilities p_2 and p_3 to initially increase very slowly and become significant for a great number of ill individuals.

It can be seen that from the point of view of each individual its interpersonal interactions have a hierarchical structure and they can be divided into three levels: k neighbors, individuals from the same local group, and individuals from the rest of the population. Note that, as results from Eqs. (4)–(6) the probabilities p_1, p_2, and p_3 of an infection of each individual depend nonlinearly on the number of ill individuals and their localization in one of the above-mentioned levels. This is why the probability of an infection of a certain individual is greatest when an ill individual is one of its k neighbors, it is smaller when an ill individual belongs to the same local group, and it is smallest when an ill individual is located somewhere in the rest of the population. Other probabilities of a transition between states X, Y are described by the parameters $W_{X\to Y}$, as in the master equation [Eq. (1)].

III. RESULTS

Computations were performed for the initial conditions with one ill (S) and randomly located individual and the rest of the population healthly and susceptible (H). Synchronous dynamics and the values $L=200$ and 500 were used.

Figure 2 shows the influence of the localization of the source of infection in one of three levels of interpersonal interactions on the number of newly infected individuals as a function of time (epidemic curves). It can be seen that the number of newly infected individuals resulting from connections with k neighbors is approximately ten times greater for times $30<t<50$ than in the case of random contacts. In the first stage of the epidemic, new infections result from the interactions with k neighbors (probability p_1), whereas the possibility of infection resulting from random contacts (p_2 and p_3) becomes significant when the number of ill individuals is large enough.

In our model it is possible to investigate the influence of the value of the clustering coefficient C [Eq. (3)] on the spreading process by changing the value of p_c. This problem was also discussed in earlier papers [11,27]. The final number V of individuals in the state (R) when the epidemic dies out (i.e., the range of the epidemic) is slightly influenced by C, but the progress of the epidemic depends significantly on the value of C. As is shown in Fig. 3, the greater C, the greater the time $t_{\max}$ in which the number of ill individuals reaches its maximum value—$w_{\max}=N_S(t_{\max})/N$. Moreover, the value of this maximum decreases with increasing C. Higher values of C cause the number of individuals grouped in clusters of highly connected nodes to increase. When one ill individual appears in a cluster, first individuals from that

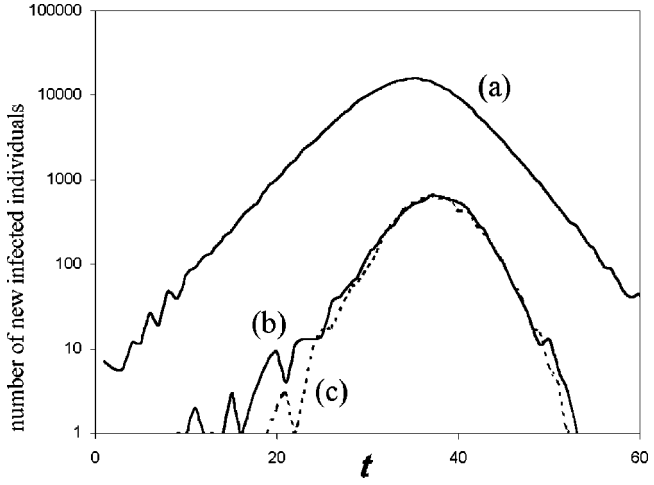

FIG. 2. Epidemic curves (the number of newly infected individuals per time step) as a function of time for different types of interpersonal interactions. Infections resulting from interpersonal connections, curve *a*; infections resulting from random contacts with the individuals from a local group, curve *b*; infections resulting from random contacts with individuals from the rest of the population, curve *c*. The values of the other parameters are $\gamma=3$, $L=500$, $L_G=20$, $W_{H\to I}=0.5$, $W_{I\to S}=0.5$, $W_{S\to R}=0.5$, $k_{\min}=8$, and $k_{\max}=24$.

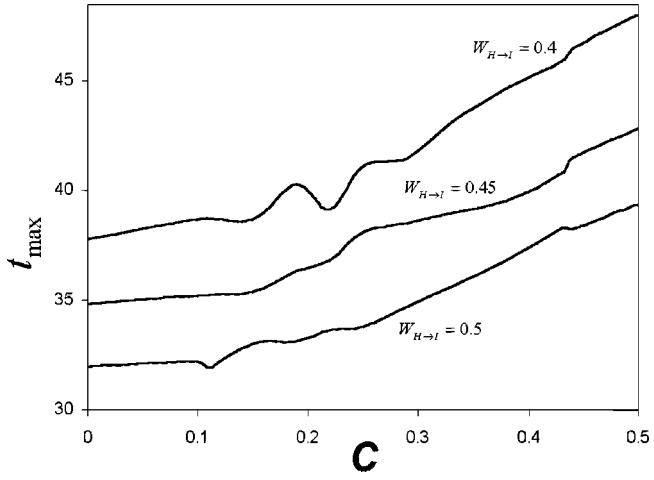

FIG. 3. The time $t_{\max}$ in which the number of ill individuals reaches a maximum value as a function of the clustering coefficient C for different values of $W_{H\to I}$. Results were averaged over 100 independent simulations. The values of other parameters are $\gamma=3$, $L=200, L_G=20$, $W_{I\to S}=0.5$, $W_{S\to R}=0.5, k_{\min}=8$, and $k_{\max}=24$.

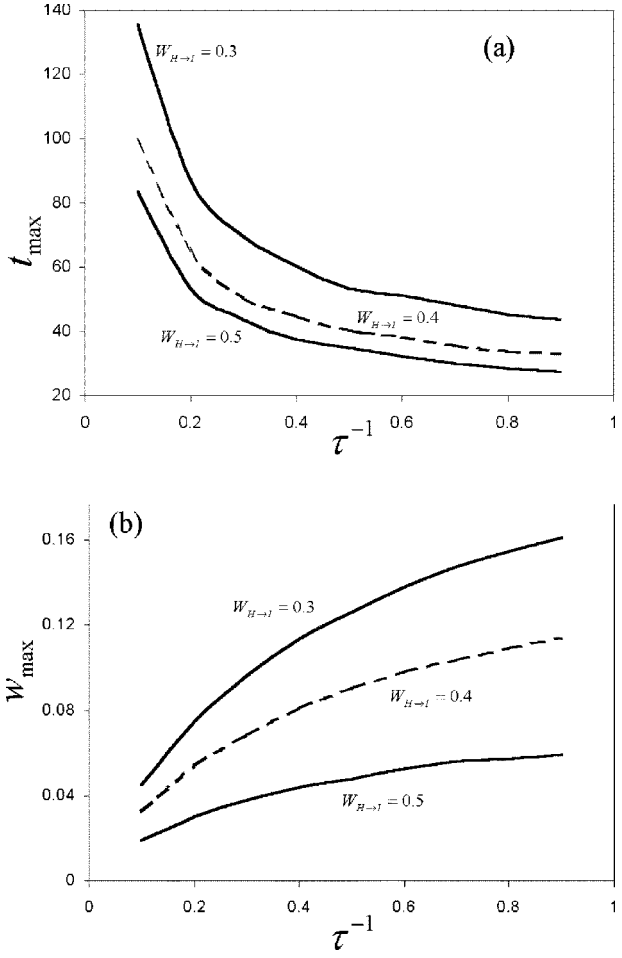

FIG. 4. The influence of the time of incubation on the time $t_{\max}$ (a) and the maximum value of the number of ill individuals $w_{\max}$ (b) for different values of $W_{H\to I}$ and $p_c=0.5$. Results were averaged over 100 independent simulations. The values of the other parameters are as in Fig. 3.

cluster are infected and next the infection spreads outside the cluster. This slows down the spreading process in the whole population.

Another important parameter is the time of incubation, proportional in our model to $\tau=1/W_{I\to S}$. It was found that the range of the epidemic V is influenced by τ^{-1}, i.e., the higher the value of τ^{-1}, the greater the range V, especially when the value $W_{H\to I}$ is low and $W_{H\to I}<W_{S\to R}$. However, for high enough values of $W_{H\to I}$, the range of the epidemic does not depend on the time τ. On the other hand, the duration of the epidemic T and $t_{\max}$ decrease with decreasing time of incubation—the epidemic spreads more rapidly [Fig. 4(a)]. The maximal number of ill individuals $w_{\max}$ increases with increasing parameter τ^{-1}, as results from Fig. 4(b).

Changes in the parameter L_G have the strongest influence on the spatial character of the spreading process. For the lowest values of L_G there is a small number of long range connections, and infections of individuals spatially located near ill individuals are more likely. The spreading process is similar to the propagation of the wave front when secondary sources of epidemics are activated [21]. With increasing L_G, the average length of connections increases. Therefore the

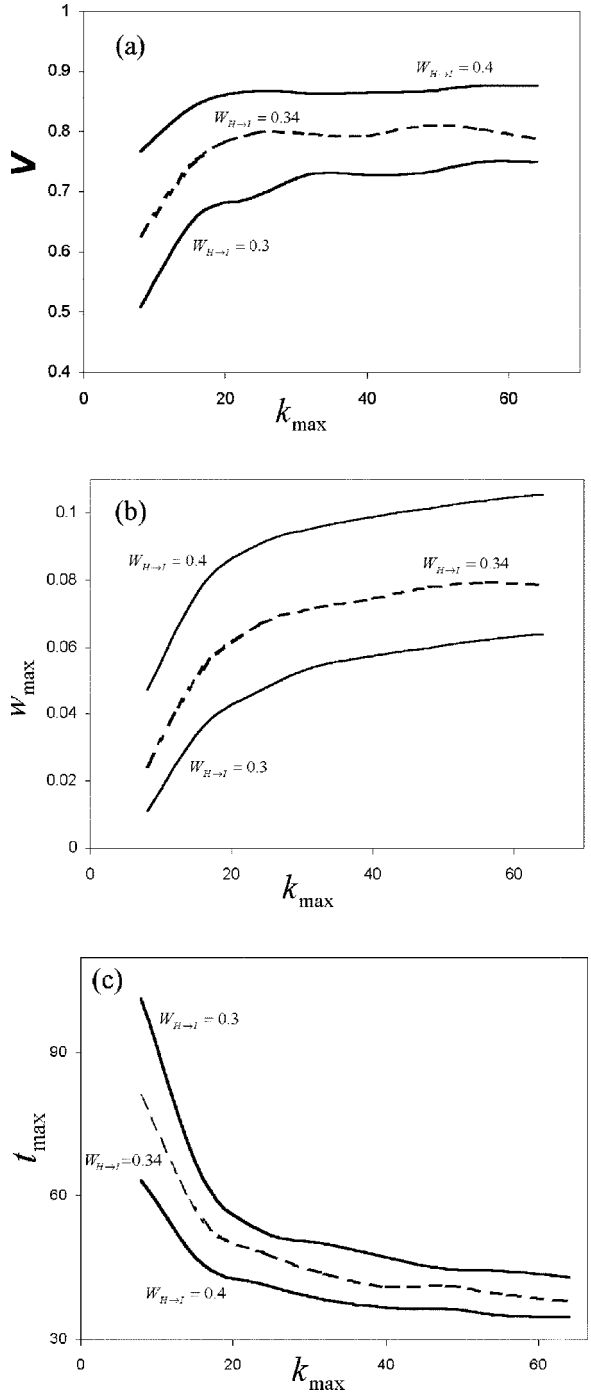

FIG. 5. The influence of the maximum value of the connectivity $k_{\max}$ on the range of the epidemic V (a), the maximum value of the number of ill individuals $w_{\max}$ (b), and the time $t_{\max}$ (c), for different values of $W_{H\to I}$ and $p_c=0.5$. Results were averaged over 100 independent simulations. The values of the other parameters are as in Fig. 3.

epidemic spreads slightly faster and the range of epidemic is slightly smaller, which results from weaker interactions between individuals in the same local group.

The spreading process and the range of the epidemic are strongly influenced by the parameter $k_{\max}$. An increase in the value of $k_{\max}$ (and, as a result, the total number of connec-

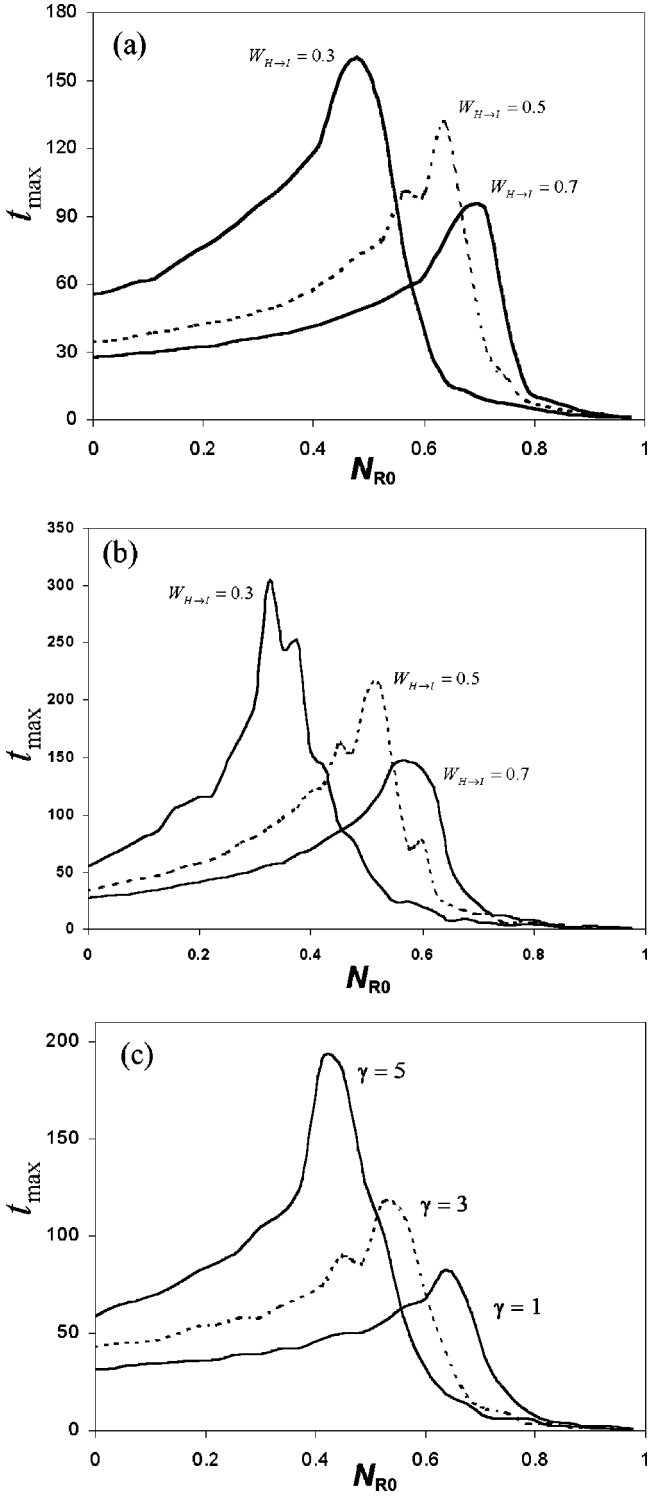

FIG. 6. The relation between time t_{max} and the number of preventively vaccinated individuals N_{R0} for different values of $W_{H\to I}$, k_{max}, and $p_c=0.5$ (a),(b). The vaccinated individuals were randomly chosen (a) or those with the highest value of k were vaccinated (b). The influence of the parameter γ on the $t_{max}(N_{R0})$ relation for $W_{H\to I}=0.4$, $k_{max}=64$, and $p_c=0.5$ is shown in (c). Results were averaged over 100 independent simulations. The values of the other parameters are as in Fig. 3.

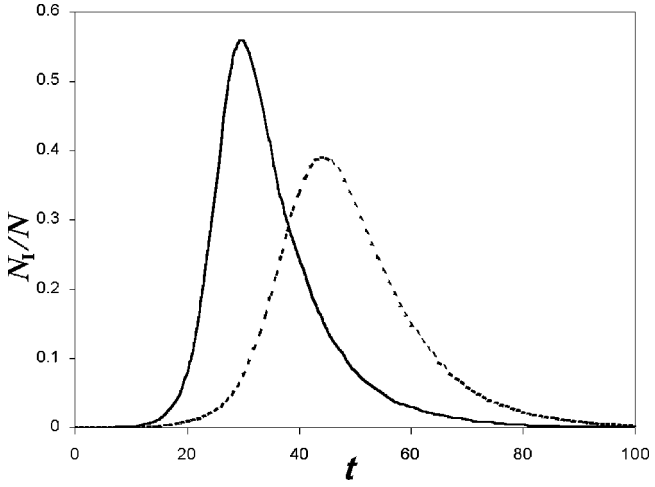

FIG. 7. Comparison of the relation $N_I(t)/N$ obtained in the present model (solid line) with the solution of the master equation (dashed line) for $W_{H\to I}=0.5, W_{S\to R}=0.1$, and $p_c=0.5$. The values of the other parameters are as in Fig. 3.

tions in the network) accelerates the spreading process and increases the range of epidemic V in the population [Fig. 5(a)]. For higher values of k_{max} the maximal number of ill individuals w_{max} has a higher value and occurs earlier [Figs. 5(a)–5(c)]. As results from Fig. 5 significant changes in V, w_{max}, and t_{max} are observed only for low values of k_{max}; then the curves saturate. It is also interesting to discuss the influence of the parameter $W_{H\to I}$ on the aforementioned relations. It was found that the higher the values of $W_{H\to I}$, the smaller the influence of the parameter k_{max} on the evolution of the epidemic, i.e., the saturation of the curves $V(k_{max})$, $w_{max}(k_{max})$, and $t_{max}(k_{max})$ occurs for lower values of k_{max}.

The connectivity of the individual—the initial source of the epidemic k_{start} — is another important parameter which determines the time evolution of the epidemic. This parameter has a similar influence on the behavior of V, w_{max}, and t_{max} as the parameter k_{max}. An earlier work discusses the influence of the localization of the initial source of epidemic in the population on the spreading process [20].

It is important to investigate the influence of preventive vaccination (the number N_{R0} of the individuals in the state R at time $t=0$) on the spreading process. In Fig. 6(a) the time t_{max} as a function of the number of preventively vaccinated (and randomly chosen) individuals for different values of $W_{H\to I}$ is shown. For low values of N_{R0} the time of duration of the epidemic T increases, because the epidemic cannot spread freely. For a certain value of N_{R0}, denoted N_{RC}, the time t_{max} reaches a maximum. The abrupt decrease of the times t_{max} and T observed for bigger values of N_{R0} indicates that a phase transition occurs at N_{RC}. This is proved by the significant increase of the transient times (i.e., the time before the system reaches the point attractor) for N_{R0} slightly smaller than N_{RC}, which is typical behavior for a phase transition. Such a phase transition in a regular lattice is the percolation phase transition [34]. When the value of $W_{H\to I}$ in-

creases, the part of the population that should be preventively vaccinated in order to suppress an epidemic also increases.

When preventively vaccinated individuals are not chosen randomly, but individuals with the greatest k are chosen, the results are similar [Fig. 6(b)]. However, the phase transition occurs for a lower value of N_{R0}, which means that a smaller number of preventively vaccinated individuals is needed to suppress the epidemic. Note that, in this case, for values of N_{R0} slightly smaller than the critical value N_{RC}, the times $t_{\max}$ and T increase quickly with N_{R0}. This means that the rate of the spreading of the epidemic is much smaller, because only individuals with small k can be infected [cf. Fig. 5(b)]. In Fig. 6(c) the influence of the parameter γ is shown. It can be seen that with decreasing values of γ, the times $t_{\max}$ and T decrease and the phase transition occurs for a higher value of N_{R0}, because there are more connections between individuals in the population.

IV. COMPARISON WITH MASTER EQUATION

In the master equation it is assumed that each individual interacts with all other individuals in the population and interactions with all individuals are treated in the same way. In contemporary and large communities this is not true, because people interact strongly only with a small (in comparison to the size of the whole population) number of other individuals. In Fig. 7 results obtained from the analytical solutions of the master equation (1) and from the present model are compared. The two curves are similar but in the case of our model the number of ill individuals increases faster and the maximum appears for lower values of time than in the case of the solutions of the master equation. When only one individual is ill at $t=0$, the number of infected individuals N_I resulting from the master equation increases very slowly, because P_S is very small. In our model, however, strong interactions with nearest neighbors are taken into account; as a consequence the epidemic spreads faster, which explains the discrepancy between the locations of the two curves.

V. CONCLUSIONS

A model of the spreading of an epidemic in a population with hierarchical structure of interpersonal interactions has been described and investigated numerically. The structure of interpersonal connections is based on a scale-free network. Spatial localization of individuals belonging to different social groups and the mobility of the contemporary community are taken into account. It was found that the type of interpersonal interaction has an essential influence on the spreading process. In particular, connections with the nearest neighbors (i.e., family or friends) are more important than random contacts between strange individuals.

As our calculations show, the epidemic spreads more slowly in a population with a higher value of the clustering coefficient C. This process depends also on the incubation time τ. With increasing values of τ the duration time T of the epidemic increases. On the other hand, an increase of the maximal number of connections in the population, $k_{\max}$, causes an increase of the range of the epidemic and accelerates the spreading process.

In our model the influence of preventive vaccinations on the spreading of the epidemic was investigated. We found a critical value of preventively vaccinated individuals sufficient for the suppression of the epidemic.

From all the results obtained a general conclusion emerges that an increase of the probability $W_{H\to I}$ decreases the influence of all the parameters characterizing the social network (i.e., $k_{\max}$, L_G, or clustering coefficient) on the dynamics and range of the epidemic. This observation shows how dangerous are most contagious diseases.

Our results were compared with the solutions of the master equation. The character of the two solutions is similar; however, there are discrepancies between the locations of the maxima of the relations of the number of ill individuals and time. It is caused by the assumptions in our model which take into account the hierarchical structure of interpersonal interactions in a more plausible way than in the case of the master equation.

[1] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).
[2] H. Ebel, L. I. Mielsch, and S. Bornholdt, Phys. Rev. E **66**, 035103(R) (2002).
[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[4] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[5] J. Travers and S. Miligram, Sociometry **32**, 425 (1969).
[6] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[7] A. E. Motter, T. Nishikawa, and Y. C. Lai, Phys. Rev. E **68**, 036105 (2003).
[8] M. E. J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003).
[9] H. Ebel, J. Davidsen, and S. Bornholdt, Complexity **8**, 24 (2002).
[10] C. Herrmann, M. Barthelemy, and P. Provero, Phys. Rev. E **68**, 026128 (2003).
[11] M. E. J. Newman, Phys. Rev. E **68**, 026121 (2003).
[12] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[13] K. Klemm, V. M. Eguiluz, R. Toral, and M. S. Miguel, Phys. Rev. E **67**, 026120 (2003).
[14] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).
[15] M. A. Janssen and W. Jansen, Artif. Life **9**, 343 (2003).
[16] E. Ravasz and A. L. Barabasi, Phys. Rev. E **67**, 026112 (2003).
[17] J. D. Noh, Phys. Rev. E **67**, 045103(R) (2003).
[18] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2003).
[19] F. Ball and P. Neal, Ann. Prob. **32**, 1168 (2004).

[20] R. A. Kosiński and L. Adamowski, Int. J. Mod. Phys. C (to be published).
[21] C. F. Moukarzel, Phys. Rev. E **60**, R6263 (1999).
[22] S. M. Dammer and H. Hinrichsen, Phys. Rev. E **68**, 016114 (2003).
[23] M. Boguna and R. Pastor-Satorras, Phys. Rev. E **66**, 047104 (2002).
[24] L. M. Sander, C. P. Warren, I. M. Sokolov, C. Simon, and J. Koopman, Math. Biosci. **180**, 293 (2002).
[25] C. Moore and M. E. J. Newman, Phys. Rev. E **61**, 5678 (2000).
[26] P. M. Tarwater and C. F. Martin, Complexity **6**, 29 (2001).
[27] M. J. Keeling, Proc. R. Soc. London, Ser. B **266**, 859 (1999).
[28] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases* (Springer, Berlin, 1993).
[29] M. J. Keeling, Proc. R. Soc. London, Ser. B **266**, 953 (1999).
[30] M. A. Aguirre, G. Abramson, A. R. Bishop, and V. M. Kenkre, Phys. Rev. E **66**, 041908 (2002).
[31] J. A. N. Filipe and C. A. Gilligan, Phys. Rev. E **67**, 021906 (2003).
[32] H. C. Tuckwell, L. Toubiana, and J. F. Vibert, Phys. Rev. E **57**, 2163 (1998).
[33] S. Wasserman and K. Faust, *Social Networks Analysis: Methods and Applications* (Cambridge University Press, Cambridge, England, 1994).
[34] M. E. J. Newman and D. J. Watts, Phys. Rev. E **60**, 7332 (2000).