Contact tracing and epidemics control in social networks

Ramon Huerta[1,2] and Lev S. Tsimring[1]
[1]*Institute for Nonlinear Science, University of California, San Diego, La Jolla, California 92093-0402*
[2] *GNB, E.T.S. de Ingeniería Informática, Universidad Autónoma de Madrid, 28049 Madrid, Spain*
(Received 9 April 2002; published 19 November 2002)

A generalization of the standard susceptible-infectious-removed stochastic model for epidemics in sparse random networks is introduced which incorporates contact tracing in addition to random screening. We propose a deterministic mean-field description that yields quantitative agreement with stochastic simulations on random graphs. Both the stochastic simulations and the mean-field equations show secondary epidemics if the contact tracing is not performed with sufficient strength. We also analyze the role of contact tracing in epidemics control in small-world networks and show that its effectiveness grows as the rewiring probability is reduced.

DOI: 10.1103/PhysRevE.66.056115 PACS number(s): 89.75.Hc, 87.23.Ge, 87.19.Xx

Properties of complex networks recently attracted much attention in physical community [1]. Although perhaps it was prompted by the advent of the Internet and World-Wide Web, the importance of this subject goes far beyond computer networks. Indeed, daily commute, power and goods traffic, wired and wireless communication, infection spreading, etc. occur within certain physical or social networks. The theory of infection spreading, which is known as mathematical epidemiology, has a long and rich history (see, e.g., Ref. [2]). However, until recently the epidemiological studies have been mostly concerned with the so-called mean-field description of epidemics, in which it is assumed that at any time the probability to get infected is the same for all individuals. In some other works, spreading of an infection on relatively simple lattices of individuals has been studied within the so-called ''forest-fire'' models [3]. Only recently, the studies that elucidate the role of underlying network structure in the infection spreading began to appear in the literature [4,5,7,8].

Most of the epidemiological models are based on several simple assumptions regarding infection contracting and cure. In particular, the most common mechanism of infection is through contact with an infected individual, and the mechanism of recovery is either deterministic or purely stochastic with a certain typical time of recovery. In the simplest susceptible-infectious-susceptible model, a recovered individual immediately becomes susceptible again, while in a more complicated susceptible-infectious-removed (SIR) model, cured individuals become immune and effectively excluded from further dynamics.

While these models give a good description of the evolution of many common infectious diseases, they usually neglect the role of intelligent strategies to stop nascent epidemics. Few epidemiological models take into account prevention strategies such as, for example, mass and ring vaccination [9]. In practice, one of the main counterepidemics measures is *contact tracing*, when individuals that have been in contact with an infected (and identified) individual, are found and thoroughly checked. It applies, among others, to the treatment of sexually transmitted diseases, tactics of law-enforcement organizations trying to uncover criminal or terrorist networks, cleaning of computer virus infection, etc. We are only aware of one theoretical paper [10] where a model of this kind has been studied. The model [10] is based on the assumption that infection is a slow branching process, while contact tracing occurs at a much shorter time scale. This leads to a familiar SIR-type model with rescaled parameters and similar dynamics. In this paper we consider a more realistic model in which infection and contact tracing occur concurrently, and their interplay determines the dynamics of the system.

Stochastic model. We assume that the population consists of N hosts whose connections to one another form a fixed graph. The hosts are enumerated with index $n=1,\ldots,N$. A node n is said to have a degree $k(n)$ if it is connected to k other hosts. In case of random graphs the degree distribution is Poissonian with a certain mean degree $K=\langle k(n)\rangle$.

For simplicity, we assume that there in no spontaneous recovery, an infected individual can only be disinfected externally through screening. Immediately upon disinfecting, the individual becomes *traced* (T) for a certain period of time during which its neighbors are checked for possible infection. After that time, the individual spontaneously becomes removed, and its neighbors are no longer traced.

Infection $S\rightarrow I$. Initially, the whole population except for one host is assumed to be susceptible to infection. The probability of host infection depends on the state of its nearest neighbors. The infection dynamics is modeled as a simple contact process: if a susceptible node n has $k_i(n)$ infectious neighbors, the probability that it becomes infectious during a small Δt time interval is $\alpha k_i(n)\Delta t$.

Tracing $I\rightarrow T$. The process of infection elimination consists in finding infectious hosts and then curing them. Hosts are being checked with certain probability β that depends on the state of their neighbors. We postulate that if an infectious host is checked, it is immediately cured, eliminated or at least isolated so it cannot infect other hosts. We introduce two nonexclusive strategies of checking for infectious hosts: random checking and contact tracing. Random checking means choosing an arbitrary host with probability $\beta_r\Delta t$, while contact tracing of host n is done with probability $\beta_t k_t(n)\Delta t$, where $k_t(n)$ is the number of neighbors of n which are in the *traced* state T. The random checking process is equivalent to the removal process of general epidemics [2].

Removal $T\rightarrow R$. With certain probability $\gamma\Delta t$, traced hosts are transformed into the *removed* state, in which they

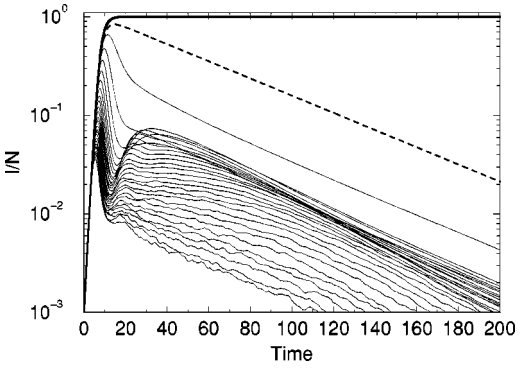

FIG. 1. Infected population in a random graph of 1000 nodes and $K=10$ for $\alpha=0.1, \gamma=0.5$. The solid thick line corresponds to $\beta_r=0, \beta_t=0$, and the dashed line to $\beta_r=0.02, \beta_t=0$. Thin lines are for $\beta_r=0.02$ and $\beta_t=0.1, 0.2, \ldots, 2.5$.

also cannot be infected, but they are no longer under observation, so they do not initiate contact tracing.

Stochastic simulations of the described process were performed using the Gillespie algorithm [11]. It is significantly superior over synchronous and asynchronous update schemes, both in terms of accuracy and computational speed. In this event-driven scheme we select the time lapsed between two consecutive events from a Poisson distribution with a combined probability of all events (infection, tracing, recovery, etc.), then choose a node (each node has its own probability to be chosen, depending on its state and the states of its neighbors), and apply the transition from one state to another according to the ratio of individual transition probabilities.

In our simulations with random graph based networks, we typically built networks with average degree $K=10$ and 1000 nodes. For every random graph we ran 100 simulations starting every time from a single (but different in each run) infected host. Then we averaged the results for 50 random graphs. The time bin was $\Delta t=10^{-6}$ for most simulations. In all simulations we varied the tracing parameters β_r and β_t, while the infection constant was set at $\alpha=0.1$, and the transition rate from T to R, $\gamma=0.5$. The latter parameter is important for the effectiveness of the targeting elimination, because the longer a node remains in the traced state, the more probable it is to trace its neighboring infectious nodes. However, since tracing presumably bears a significant cost, an optimal choice of the tracing parameters (β_r, β_t, and γ) for a given epidemic is an important issue.

In Fig. 1 we present the "prevalence" of epidemics (the fraction of infectious nodes in the whole population $i=I/N$) as a function of time for several values of β_r and β_t. When $\beta_r=\beta_t=0.0$ we have a simple SI process, and all the nodes eventually get infected (thick solid line in Fig. 1). Other curves show the fraction of infectious nodes as a function of time for $\beta_r=0.02$ and different values of β_t. The ratio α/β_r is chosen to be above the epidemics threshold [2]. For $\beta_t=0$ we obtain the classical SIR process with randomly removed infectives (dashed line in Fig. 1). The epidemic eventually saturates, and the fraction of infectious nodes decays exponentially. The lower lines display the evolution of the infection fraction for values of β_t ranging from 0 to 2.5 with a step value of 0.1. The initial (exponential) phase of the epidemics growth is nearly independent of β_t, because the contact tracing process is intrinsically nonlinear (it requires the presence of I-T connected pairs and, therefore, only begins after the first infected node is randomly screened). As expected, the tracing process significantly reduces the magnitude of the epidemics (maximal value of i), but at large times the infection decays with the same exponential rate as for $\beta_t=0.0$ (again, because we return to the linear regime at small i). The most interesting feature of the process at large $\beta_t>0.35$ is the presence of a second maximum of i which indicates a second epidemic. Due to this second epidemic, the percentage of the infectious population at large times $t>40$ may actually increase with increase of β_t. It means that the range of β_t values from 0.4 to 0.9 are not better to control the epidemics than values smaller than 0.4.

Mean-field equations. At first sight, it seems that the mean-field approach cannot be applied to contact tracing, since it does not take into account the nonuniform distribution of infection in the population. Nevertheless, a more sophisticated mean-field approach that operates not only with the mean densities of states, but also with the densities of links connecting nodes with different states, can be applied. In the context of nonequilibrium kinetics on lattices this approach is known as multisite or cluster mean-field theory (see, for example, Ref. [6]). In epidemiological context, this approach was pioneered by Rand [7] (see also Ref. [8]). Let us introduce the number of nodes A, the number of connected pairs $[AB]$, and triples $[ABC]$ of nodes, where A, B, and C stand for any of the types S, I, T, R. For example, the number of connected pairs of infectious and traced nodes is denoted $[IT]$. Note that $[AB]=[BA]$ and each pair in $[AA]$ is counted twice. For large N, the ratios A/N, $[AB]/N$, $[ABC]/N$ approach deterministic limits which we label $a, [ab], [abc]$, respectively.

The dynamics of the model is described by the following set of rate equations:

$$\dot{s}=-\alpha[si], \tag{1}$$

$$\dot{i}=\alpha[si]-\beta_r i-\beta_t[i\tau], \tag{2}$$

$$[\dot{ss}=-2\alpha[ssi], \tag{3}$$

$$[\dot{si}]=\alpha([ssi]-[isi]-[si])-\beta_r[si]-\beta_t[si\tau], \tag{4}$$

$$[\dot{ii}]=2\alpha([isi]+[si])-2\beta_r[ii]-2\beta_t[ii\tau], \tag{5}$$

$$[\dot{s\tau}]=-\alpha[ist]+\beta_r[si]+\beta_t[si\tau]-\gamma[s\tau], \tag{6}$$

$$[\dot{i\tau}]=\alpha[ist]+\beta_r[ii]+\beta_t([ii\tau]-[\tau i\tau]-[i\tau])-\gamma[i\tau] -\beta_r[i\tau]. \tag{7}$$

Here we used the notation $\tau=[T]/N$ to avoid confusion between the density of traced nodes and time t. Note that we omit here the equations for τ, $[\tau\tau]$, as well as any combina-

tions involving the removed state, since they do not affect the dynamics of the infectious population.

The meaning of these equations is rather straightforward. For example, the terms in the right-hand side of the last equation can be explained as follows. A (p,q) pair becomes $[i\tau]$ through infection of a susceptible node p, random screening of the infectious node q in an $[ii]$ pair, or through contact tracing of node q from node r in a (p,q,r) triple in the state $[ii\tau]$. However, we can lose an $[i\tau]$ pair by contact tracing of the p node in a $[\tau i\tau]$ triple (r,p,q), by direct tracing of p by q, by removing of q, or by random screening of p. Other equations can be obtained from similar arguments.

This set of equation is not closed, as the equations for the pair densities contain triple densities. We need to introduce a closure rule. Similarly to Refs. [7,8], we can use the approximation $[abc]=[ab][bc]/b$, which follows from the condition that the influence of a node on the state of its second neighbor in a triple is negligible [12].

Using this closure rule, we arrive at the following set of equations:

$$\dot{s}=-\alpha s\hat{i}, \tag{8}$$

$$\dot{i}=\alpha s\hat{i}-\beta_r i-\beta_t i\hat{\tau}, \tag{9}$$

$$\dot{\hat{i}}=(\alpha K s-\alpha-\beta_r)\hat{i}-\beta_t\hat{i}\hat{\tau}, \tag{10}$$

$$\dot{\hat{\tau}}=\alpha\frac{s}{i}\hat{i}\hat{w}+\left(\beta_t\hat{\eta}-\beta_t-\gamma-\alpha\frac{s}{i}\hat{i}\right)\hat{\tau}+\beta_r\hat{\eta}, \tag{11}$$

$$\dot{\hat{w}}=\beta_r\hat{i}+\beta_t\hat{i}\hat{\tau}-\gamma\hat{w}, \tag{12}$$

$$\dot{\hat{\eta}}=\alpha\frac{s}{i}(2\hat{i}+2-\hat{\eta})\hat{i}-(\beta_r+\beta_t\hat{\tau})\hat{\eta}. \tag{13}$$

where $\hat{i}=[is]/s$ is the mean number of infectious neighbors per susceptible node, $\hat{\tau}=[i\tau]/i$ is the mean number of traced neighbors per infectious node, $\hat{\eta}=[ii]/i$ is the mean number of infectious neighbors of an infectious node, and $\hat{w}=[st]/s$ is the mean number of traced neighbors per susceptible node. Notice that the equation for $[ss]$ dropped out as $[ss]=Ks^2$ satisfies the equations at all times. We used the initial conditions $s(0)=1-i_0, i(0)=i_0, \hat{i}(0)=(K-1)i_0, \hat{\tau}(0)=\hat{w}(0)=\hat{\eta}(0)=0$ which correspond to a small set of disconnected infectious nodes.

During the early stage of an epidemic the contact tracing can be neglected $(\tau=0)$, and Eqs. (8)–(13) are reduced to a set of three equations for $s,i,\hat{i}$ which coincide with the model that has been studied in Refs. [7,8]. Independently of β_t, the initial epidemics growth is characterized by the basic reproduction number $K\alpha/(\alpha+\beta_r)$. However, as the number of traced individuals grows, the growth rate is reduced and the epidemic is saturated. In Fig. 2 the dynamics of the epidemics calculated from Eqs. (8)–(13) are shown for different values of β_t. As seen from the figure, the maximum number of infectious nodes is drastically reduced with increase of

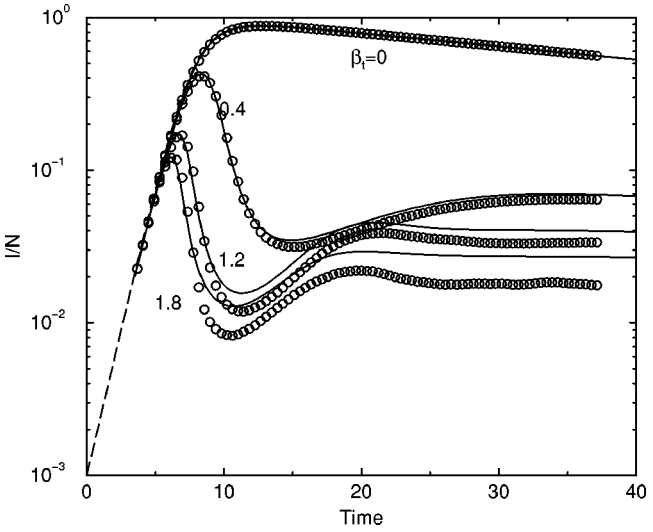

FIG. 2. Evolution of the infection prevalence for different β_t, mean-field model (lines) and stochastic simulations (symbols), $\alpha=0.1$, $K=10$, $\beta_r=0.02$, $\gamma=0.5$.

β_t. In the same figure we show the results of direct stochastic simulations for several β_t [13]. Both numerical simulations and the model exhibit the emergence of the secondary epidemic outbreak after the first one is nearly suppressed. The mechanism of the secondary epidemic is related to the interplay between the prevalence of invectious and traced nodes. After the first outbreak, the large number of traced nodes prevents the infection from spreading (the ''instantaneous'' basic reproduction number $K\alpha/(\alpha+\beta_r+\beta_t\hat{\tau})<1$), however, as this number diminishes with the rate γ below a certain threshold, the infection is able to spread again.

The most important question is whether contact tracing is capable of arresting the exponential growth of the epidemic before it engulfs a finite portion of the total population.

To answer this question, we consider the limit of small epidemics in a large population $(i,\hat{i}\ll 1)$, then we can set $s(t)=1$ and drop Eq. (9). We also observe that $\hat{i}/i=K-1$ and drop Eq. (9) as well. A simple calculation shows that the critical value of β_t at which the exponential growth of infection is arrested,

$$\beta_{cr}=\frac{[\alpha(K-1)+\gamma][\alpha(K-1)-\beta_r]}{\beta_r}. \tag{14}$$

For $\beta>\beta_{cr}$, epidemic remains small at all times, and so there is no major outbreak of the epidemic. For the parameter values of our stochastic simulations, $\alpha=0.1, K=10, \beta_r=0.02, \gamma=0.5$, we obtain $\beta_{cr}=61.6$. In Fig. 3 we show the relative size of the epidemic $i_{tot}=1-s(t=\infty)$ as a function of β_t at different initial epidemic sizes i_0. In agreement with the above argument, for $\beta_t<\beta_{cr}$ the value of $1-s_\infty$ is weakly independent of i_0, and it drops to zero as β_t approaches β_{cr}.

The mean-field equations proposed here are valid for networks with a low clustering coefficient [1]. Sparse random networks studied above represent a particular class of net-

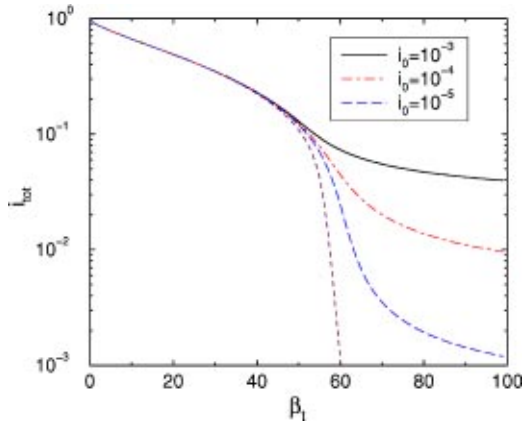

FIG. 3. Total number of infectious nodes vs β_t for different i_0. Parameters are the same as in Fig. 2.

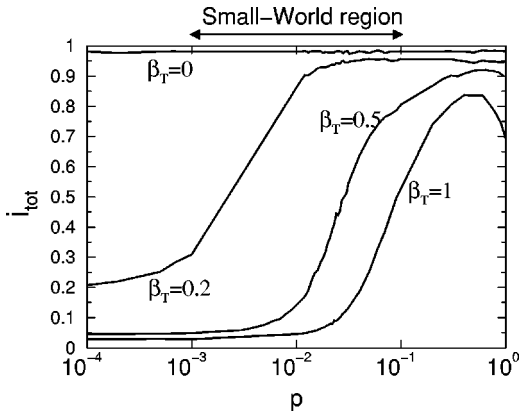

FIG. 4. The epidemic size as a function of the rewiring probability p. Effectiveness of the contact tracing becomes very significant in the small-world regime.

works with a short average minimal path and a small clustering coefficient. Many social networks are characterized by a relatively large clustering coefficient while keeping the average minimal path low. We studied numerically the effect of the network structure on the contact tracing of epidemics within the small-world model [14]. Changing the rewiring probability p allows us to scan the range of networks from regular ($p=0$) to random ($p\to 1$) through the small-world range $0.001<p<0.1$, which exhibits a short average minimal path and a large clustering coefficient typical for many social networks. We used the same number of nodes and edges as for the random graph simulations, and fixed the parameter values at $\alpha=0.1$, $\beta_r=0.02$, and $\gamma=0.5$. Figure 4 shows the dependence of the epidemic size i_{tot} on p for several different β_t. As we can see, i_{tot} changes mostly within the small-world range ($0.001<p<0.1$) where the clustering coefficient and the average path undergo large variations.

In this paper we studied the role of contact tracing as a part of the epidemics control strategy in complex networks. We demonstrated that by applying this strategy, a major outbreak can be significantly reduced or even eliminated at a small additional cost. Based on the pair correlation approach given by Rand [7], we developed the mean-field model of contact tracing for the case of random graphs. We also studied the influence of network topology on contact tracing using the small-world model with variable rewiring probability p, and found that its effectiveness grows as the reviring probability is reduced. The main change occurs within the small-world regime at $p\sim 10^{-2}$.

This work was supported by the U.S. DOE under Grant No. DE-FG03-96ER14592, ARO MURI-ARO under Grant No. DAAG55-98-0269, MCyT under Grant No. BFI2000-157, and NATO under Collaborative Linkage Grant No. PST.CLG.978512.

[1] M. E. J. Newman, J. Stat. Phys. **101**, 819 (2000); S. H. Strogatz, Nature (London) **410**, 268 (2001); R. Albert and A. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
[2] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases*, 2nd ed. (Griffin, London, 1975).
[3] A. Johansen, Physica D **78**, 186 (1994).
[4] C. Moore and M. E. J. Newman, Phys. Rev. E **61**, 5678 (2000).
[5] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[6] D. ben-Avraham and J. Köhler, Phys. Rev. A **45**, 8358 (1992).
[7] D. Rand, in *Advanced Ecological Theory: Principles and Applications*, edited by J. McGlade (Blackwell Science, Oxford, 1999), pp. 100–142.
[8] H. Andersson and T. Britton, *Epidemics: Stochastic Models and Their Statistical Analysis*, Springer Lecture Notes in Statistics, Vol. 151 (Springer-Verlag, New York, 2000).
[9] J. Müller, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **59**, 222 (1998); J. Math. Biol. **41**, 143 (2000).
[10] J. Müller, M. Kretzschmar, and K. Dietz, Math. Biosci. **164**, 39 (2000).
[11] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[12] This condition becomes asymptotically correct for $N\to\infty$, as the probability of a triple to form a connected triangle is $O(N^{-1})$. However, for other types of networks, such as lattices or small-world networks, this condition clearly is not satisfied.
[13] The discrepancy between theory and numerics for larger values of β_t is caused by the finite size effects: as the number of infectious and traced nodes reduces, the agreement between the theory and simulations becomes only qualitative.
[14] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).