

## Efficient Immunization Strategies for Computer Networks and Populations

Reuven Cohen,<sup>1,\*</sup> Shlomo Havlin,<sup>1</sup> and Daniel ben-Avraham<sup>2</sup>

<sup>1</sup>*Minerva Center and Department of Physics, Bar-Ilan University, Ramat-Gan, 52900, Israel*

<sup>2</sup>*Department of Physics, Clarkson University, Potsdam New York 13699-5820, USA*

(Received 24 April 2003; published 9 December 2003)

We present an effective immunization strategy for computer networks and populations with broad and, in particular, scale-free degree distributions. The proposed strategy, *acquaintance immunization*, calls for the immunization of random acquaintances of random nodes (individuals). The strategy requires no knowledge of the node degrees or any other global knowledge, as do targeted immunization strategies. We study analytically the critical threshold for complete immunization. We also study the strategy with respect to the susceptible-infected-removed epidemiological model. We show that the immunization threshold is dramatically reduced with the suggested strategy, for all studied cases.

DOI: 10.1103/PhysRevLett.91.247901

PACS numbers: 89.75.Hc, 64.60.Ak, 87.19.Xx

It is well established that random immunization requires immunizing a very large fraction of a computer network, or population, in order to arrest epidemics that spread upon contact between infected nodes (or individuals) [1–7]. Many diseases require 80%–100% immunization (for example, measles requires 95% of the population to be immunized [1]). The same is correct for the Internet, where stopping computer viruses requires almost 100% immunization [5–7]. On the other hand, targeted immunization of the most highly connected individuals [1,5,8–11], while effective, requires global information about the network in question, rendering it impractical in many cases. Here, we develop a mathematical model and propose an effective strategy, based on the immunization of a small fraction of *random acquaintances* of randomly selected nodes. In this way, the most highly connected nodes are immunized, and the process prevents epidemics with a small finite immunization threshold and without requiring specific knowledge of the network.

Social networks are known to possess a broad distribution of the number of links (contacts),  $k$ , emanating from a node (an individual) [12–14]. Examples are the web of sexual contacts [15], movie-actor networks, science citations, and cooperation networks [16,17], etc. Computer networks, both physical (such as the Internet [18]) and logical (such as the World Wide Web [19], and email [20] and trust networks [21]) are also known to possess wide, scale-free, distributions. Studies of percolation on broad-scale networks show that a large fraction  $f_c$  of the nodes need to be removed (immunized) before the integrity of the network is compromised. This is particularly true for scale-free networks,  $P(k) = ck^{-\lambda}$  ( $k \geq m$ ), where  $2 < \lambda < 3$ , the case of most known networks [12–14], where the percolation threshold  $f_c \rightarrow 1$ , and the network remains connected (contagious) even after removal of most of its nodes [6]. In other words, with a random immunization strategy almost all of the nodes need to be immunized before an epidemic is arrested (see Fig. 1).

When the most highly connected nodes are targeted first, removal of just a small fraction of the nodes results

in the network's disintegration [5,10,11]. This has led to the suggestion of targeted immunization of the HUBS (the most highly connected nodes in the network) [8,22]. However, this approach requires a complete, or at least fairly good knowledge of the degree of each node in the network. Such global information often proves hard to gather, and may not even be well defined (as in social networks, where the number of social relations depends on subjective judging). The acquaintance immunization strategy proposed herein works at low immunization rates,  $f$ , and obviates the need for global information.

In our approach, we choose a random fraction  $p$  of the  $N$  nodes and look for a random acquaintance with whom they are in contact (thus, the strategy is purely local,

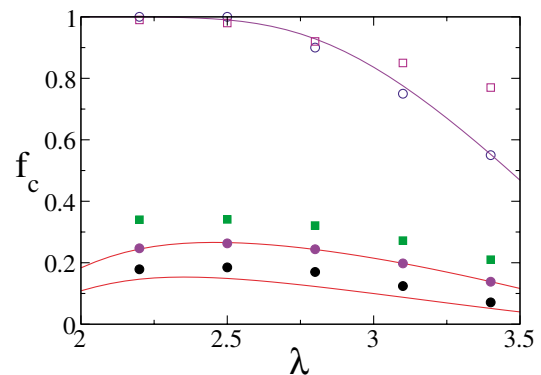


FIG. 1 (color online). Critical probability,  $f_c$ , as a function of  $\lambda$  in scale-free networks (with  $m = 1$ ), for the random immunization (top curve and open circles), acquaintance immunization (middle curve and top full circles), and double acquaintance immunization (bottom curve and bottom full circles) strategies. Curves represent analytical results (an approximate one for double acquaintance), while data points represent simulation data, for a population  $N = 10^6$ . (Because of the population's finite size,  $f_c < 1$  for random immunization even when  $\lambda < 3$ .) Squares are for random (open) and acquaintance immunization (full) of assortatively mixed networks [where links between sites of degree  $k_1$  and  $k_2 (> k_1)$  are rejected with probability  $0.7[1 - (k_1/k_2)]$ ].

requiring minimal information about randomly selected nodes and their immediate environs). The acquaintances, rather than the originally chosen nodes, are the ones immunized. The fraction  $p$  may be larger than 1 [23], for a node might be queried more than once, on average, while the fraction of nodes immunized  $f$  is always less than or equal to 1.

Suppose we apply the acquaintance strategy on a random fraction  $p$  of the network. The critical fractions,  $p_c$  and  $f_c$ , needed to stop the epidemic can be analytically calculated. In each event, the probability that a node with  $k$  contacts is selected for immunization is  $kP(k)/(N\langle k \rangle)$  [6,10], where  $\langle k \rangle = \sum_k kP(k)$  denotes the average degree of nodes in the network. This quantifies the known fact that randomly selected acquaintances possess more links than randomly selected nodes [24,25]. Suppose we follow some branch, starting from a random link of the spanning cluster. In some layer,  $l$ , we have  $n_l(k)$  nodes of degree  $k$ . In the next layer ( $l+1$ ) each of those nodes has  $k-1$  new neighbors (excluding the one through which we arrived). Let us denote the event that a node of degree  $k$  is susceptible to the disease (not immunized) by  $s_k$ . To find out the number of nodes,  $n_{l+1}(k')$ , of degree  $k'$  that are susceptible, we multiply the number of links going out of the  $l$ th layer by the probability of reaching a node of degree  $k'$  by following a link from a susceptible node,  $p(k'|k, s_k)$ . Then, we multiply by the probability that this node is also susceptible given both the node and the neighbor's degrees, and the fact that the neighbor is also susceptible,  $p(s_{k'}|k', k, s_k)$ . Since below and at the critical percolation threshold loops are irrelevant [6], one can ignore them. Therefore,

$$n_{l+1}(k') = \sum_k n_l(k)(k-1)p(k'|k, s_k)p(s_{k'}|k', k, s_k). \quad (1)$$

By using Bayes's rule:

$$p(k'|k, s_k) = \frac{p(s_k|k, k')p(k'|k)}{p(s_k|k)}. \quad (2)$$

Assuming that the network is uncorrelated (no degree-degree correlations), the probability of reaching a node with degree  $k'$  via a link,  $\phi(k') \equiv p(k'|k) = k'P(k')/\langle k \rangle$ , is independent of  $k$ .

A random site (of degree  $k$ ) is selected in each step with probability  $1/N$ . The probability of being redirected to a specific acquaintance is  $1/k$ . Thus, the probability that the acquaintance is *not* selected in one particular attempt, is  $(1 - 1/Nk)$ , and in all  $Np$  vaccination attempts, it is

$$\nu_p(k) \equiv \left(1 - \frac{1}{Nk}\right)^{Np} \approx e^{-p/k}. \quad (3)$$

If the neighbor's degree is not known, the probability is  $\nu_p \equiv \langle \nu_p(k) \rangle$ , where the average (and all averages henceforth) is taken with respect to the probability distribution  $\phi(k)$ . The probability that a node with degree  $k'$  is susceptible is  $p(s_{k'}|k') = \langle \exp(-p/k) \rangle^{k'}$ , if no other infor-

mation exists on its neighbors. If the degree of one neighbor is known to be  $k'$ :  $p(s_k|k, k') = e^{-p/k'} \times \langle e^{-p/k} \rangle^{k-1}$ . Since the fact that a neighbor with known degree is immunized does not provide any further information about a node's probability of immunization, it follows that  $p(s_k|k, k') = p(s_k|k, k', s_{k'})$ . Using the above equations one obtains

$$p(k'|k, s_k) = \frac{\phi(k')e^{-p/k'}}{\langle e^{-p/k} \rangle}. \quad (4)$$

Substituting these results in (1) yields

$$n_{l+1}(k') = \nu_p^{k'-2} \phi(k') e^{-p/k'} \sum_k n_l(k)(k-1)e^{-p/k}. \quad (5)$$

Since the sum in (5) does not depend on  $k'$ , it leads to the stable distribution of degree in a layer  $l$ :  $n_l(k) = a_l \nu_p^{k-2} \phi(k) e^{-p/k}$ , for some  $a_l$ . Substituting this into (5) yields

$$n_{l+1}(k') = n_l(k') \sum_k \phi(k)(k-1) \nu_p^{k-2} e^{-2p/k}. \quad (6)$$

Therefore, if the sum is larger than 1 the branching process will continue forever (the percolating phase), while if it is smaller than 1 immunization is subcritical and the epidemic is arrested. Thus, we obtain a relation for  $p_c$ :

$$\sum_k \frac{P(k)k(k-1)}{\langle k \rangle} \nu_p^{k-2} e^{-2p_c/k} = 1. \quad (7)$$

The fraction of immunized nodes is easily obtained from the fraction of nodes which are not susceptible,

$$f_c = 1 - \sum_k P(k)p(s_k|k) = 1 - \sum_k P(k)\nu_{p_c}^k, \quad (8)$$

where  $P(k)$  is the regular distribution, and  $p_c$  is found numerically using Eq. (7).

A related immunization strategy calls for the immunization of acquaintances referred to by at least  $n$  nodes. (Above, we specialized to  $n=1$ .) The threshold is lower the larger  $n$  is, and may justify, under certain circumstances, this somewhat more involved protocol.

The acquaintance immunization strategy is effective for any broad-scale distributed network. Here we give examples for scale-free and bimodal distributions, which are common in many natural networks. We also give an example of an assortatively mixed network (where high degree nodes tend to connect to other high degree nodes [26]). We also discuss the effectiveness of the strategy in conjunction with the susceptible-infected-removed (SIR) epidemiological model.

In Fig. 1, we show the immunization threshold  $f_c$  needed to stop an epidemic in networks with  $2 < \lambda < 3.5$  (this covers all known cases). Plotted are curves for the (inefficient) random strategy, and the strategy advanced here, for the cases of  $n=1$  and 2. Note that while  $f_c = 1$  for networks with  $2 < \lambda < 3$  (e.g., the Internet) it decreases dramatically to values  $f_c \approx 0.25$  with the

suggested strategy. The figure also shows the strategy's effectiveness in the case of assortatively mixed networks [26], i.e., in cases where  $p(k'|k)$  does depend on  $k$ , and high degree nodes tend to connect to other high degree nodes, which is the case for many real networks.

Figure 2 gives similar results for a bimodal distribution (consisting of two Gaussians, where high degree nodes are rare compared to low degree ones). This distribution is also believed to exist for some social networks, in particular, for some networks of sexual contacts. In Fig. 3 geographical effects, where nodes tend to connect to geographically adjacent ones [28], are also taken into account. The improvement gained by the use of the acquaintance immunization strategy is evident in both cases, as seen in Figs. 2 and 3.

The above considerations hold if full immunization is required. That is, given a static network structure, one wishes to stop any epidemic or virus propagation. However, most real viruses have a finite infection rate, and, therefore, a finite probability of infecting a neighbor of an infected node. The SIR model, widely studied by epidemiologists [29–31], assumes that nodes can be susceptible, infected, or removed (i.e., recovered and immunized against further infection or otherwise removed from the network). This epidemiological model can be mapped to a bond percolation model, where the concentration of bonds,  $q = 1 - e^{-r\tau}$ , where  $r$  is the transmissibility of the virus (infection rate over a link) and  $\tau$  is the infection time. To find the effect of the strategy, given this finite infection probability, the right hand side of Eq. (1) should be multiplied by  $q$ , giving:

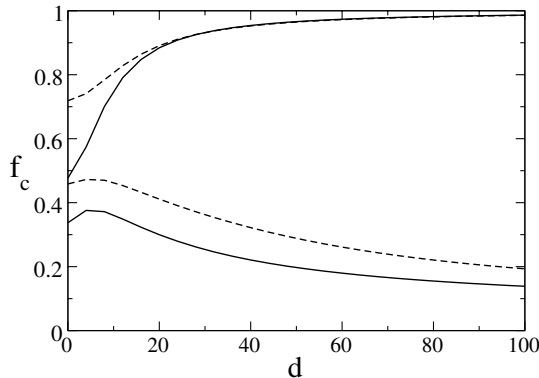


FIG. 2. Critical concentration,  $f_c$ , for the bimodal distribution (of two Gaussians) as a function of  $d$ , the distance between the modes. The first Gaussian is centered at  $k = 3$  and the second one at  $k = d + 3$  with height 5% of the first. Both have variance 2 (solid lines) or 8 (dashed lines). Top two lines are for random immunization. The bottom two lines are for acquaintance immunization. All curves are analytically derived from Eqs. (7) and (8). Very similar results have been obtained for bimodal distributions of two Poissonians. Note that also for the case  $d = 0$ , i.e., a single Gaussian, the value of  $f_c$  reduces considerably due to the acquaintance immunization strategy. Thus the strategy gives improved performance even for relatively narrow distributions [27].

$$\sum_k \frac{P(k)k(k-1)}{\langle k \rangle} \nu_{p_c}^{k-2} e^{-2p_c/k} = q^{-1} \quad (9)$$

instead of Eq. (7). Results for different infection rates and scale-free networks with  $\lambda = 2.5$  and  $\lambda = 3.5$  are shown in Fig. 4. As can be seen in the figure, in the limit  $\tau r \rightarrow \infty$  this model leads to the full immunization case of Fig. 1. For lower values of  $r$ , the proposed strategy still gives similar, or even larger, decrease in the immunization threshold.

Various immunization strategies have been proposed, mainly for the case of an already spread disease, and are based on tracing the chain of infection towards the super-spreaders of the disease [2]. This approach is different from our proposed approach, since it is mainly aimed at stopping an epidemic after the outbreak began. It is also applicable for cases where no immunization exists and only treatment for already infected individuals is possible. Our approach, on the other hand, can be used even before the epidemic starts spreading, since it does not require any knowledge of the chain of infection.

In practice, any population immunization strategy must take into account issues of attempted manipulation. We would expect the suggested strategy to be less sensitive to manipulations than targeted immunization strategies. This is due to its dependence on acquaintance reports, rather than on *self-estimates* of number of contacts. Since a node's reported contacts pose a direct threat to the node (and relations), we anticipate that manipulations would be less frequent. Furthermore, we would suggest adding some randomness to the process: for example, reported acquaintances are not immunized, with

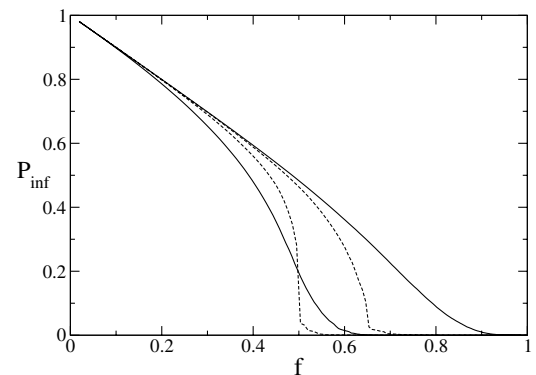


FIG. 3. The fraction of the population infected in the endemic state,  $P_{\text{inf}}$ , as a function of  $f$ , the immunized fraction of the population, for a scale-free network embedded in a  $2d$  geographical space (see [28]). The solid lines are for random (top) and acquaintance (bottom) immunization for a network with  $\lambda = 2.8$  and the dashed lines are for the same cases with  $\lambda = 4$ . In all cases  $N = 10^6$  and  $m = 4$ . In both cases ( $\lambda = 2.8$  and 4) the acquaintance immunization strategy provides a considerable improvement over random immunization. The high values for  $f_c$  stem from the fact that the network is very well connected with  $m = 4$ , which was taken in order to approach a regular square lattice at  $\lambda \rightarrow \infty$ .

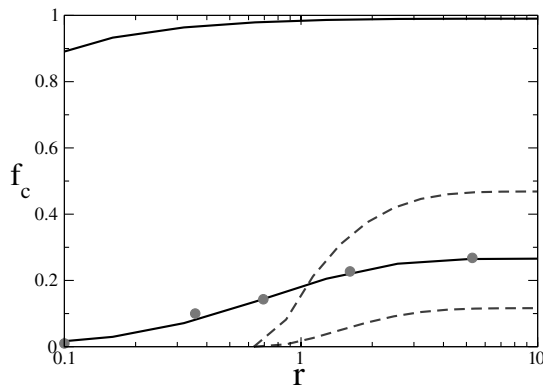


FIG. 4. Critical concentration,  $f_c$ , vs  $r$ , the infection rate, for the SIR model with  $\tau = 1$ . The solid lines are for random (top) and acquaintance immunization (bottom) for scale-free networks with  $\lambda = 2.5$ . The dashed lines are for  $\lambda = 3.5$  [random (top) and acquaintance (bottom) immunization]. The circles represent simulation results for acquaintance immunization for scale-free networks with  $\lambda = 2.5$ .

some small probability (smaller than the random epidemic threshold), while randomly selected individuals are immunized directly, with some low probability. This will have a small impact on the efficiency, while enhancing privacy and rendering manipulations less practical.

In conclusion, we have proposed a novel efficient strategy for immunization, requiring no knowledge of the nodes' degrees or any other global information. This strategy is efficient for networks of any broad-degree distribution and allows for a low threshold of immunization, even where random immunization requires the entire population to be immunized. We have presented analytical results for the critical immunization fraction in both a static model and the kinetic SIR model.

As a final remark, we note that our approach may be relevant to other networks, such as ecological networks of predator prey [32,33], metabolic networks [34], networks of cellular proteins [35], and terrorist networks. For terrorist networks, our findings suggest that an efficient way to disintegrate the network is to focus more on removing individuals whose name is obtained from another member of the network.

We are grateful to the Israel Science Foundation and to NSF Grant No. PHY-0140094 (DbA) for partial support of this research.

\*Email address: cohenr@shoshi.ph.biu.ac.il

- [1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, U.K., 1992).
- [2] H.W. Hethcote and J. A. Yorke, *Gonorrhea Transmission Dynamics and Control*, Lecture Notes in Biomathematics Vol. 56 (Springer-Verlag, Berlin, 1984).
- [3] R. M. May and R. M. Anderson, *Math. Biosci.* **72**, 83 (1984).

- [4] H.W. Hethcote and J.W. Van-Ark, *Math. Biosci.* **84**, 85 (1987).
- [5] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **406**, 378 (2000).
- [6] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).
- [7] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [8] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **65**, 036104 (2002).
- [9] A. L. Lloyd and R. M. May, *Science* **292**, 1316 (2001).
- [10] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
- [11] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001).
- [12] S. H. Strogatz, *Nature (London)* **410**, 268 (2000).
- [13] R. Albert and A. L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [14] S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002).
- [15] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, *Nature (London)* **411**, 907 (2001).
- [16] A.-L. Barabási, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, and A. Schubert, *Physica (Amsterdam)* **311A**, 590 (2002).
- [17] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2566 (2002).
- [18] S.-H. Yook, H. Jeong, and A.-L. Barabási, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13 382 (2002).
- [19] A. L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [20] H. Ebel, L.-I. Mielsch, and S. Bornholdt, *Phys. Rev. E* **66**, 035103(R) (2002).
- [21] X. Guardiola, R. Guimera, A. Arenas, A. Diaz-Guilera, D. Streib, and L. A. N. Amaral, *cond-mat/0206240*.
- [22] Z. Dezsó and A.-L. Barabási, *Phys. Rev. E* **65**, 055103(R) (2002).
- [23] For most cases  $p_c$  is well below 1. For the scale-free distribution for  $\lambda = 2.5, 3$  and  $3.5$ ,  $p_c \approx 0.81, 0.52$  and  $0.28$  respectively. Querying is, however, not expensive, and asking each individual for more than one neighbor will decrease these numbers significantly, to approx. 0.4, 0.3 and 0.2, respectively.
- [24] S. L. Feld, *Am. J. Soc.* **96**, 1464 (1991).
- [25] M. E. J. Newman, *Soc. Netw.* **25**, 83 (2003).
- [26] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [27] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11 149 (2000).
- [28] A. F. Rozenfeld, R. Cohen, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **89**, 218701 (2002).
- [29] P. Grassberger, *Math. Biosci.* **63**, 157 (1983).
- [30] C. P. Warren, L. M. Sander, and I. M. Sokolov, *Math. Biosci.* **180**, 293 (2002).
- [31] M. E. J. Newman, *Phys. Rev. E* **66**, 016128 (2002).
- [32] R. V. Solé and J. M. Montoya, *Proc. R. Soc. London Sect. B* **268**, 2039 (2001).
- [33] J. Camacho, R. Guimera, and L. A. N. Amaral, *Phys. Rev. Lett.* **88**, 228102 (2002).
- [34] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, *Nature (London)* **407**, 651 (2000).
- [35] H. Jeong, S. Mason, A.-L. Barabási and Z. N. Oltvai, *Nature (London)* **411**, 41 (2001).