Finding a Better Immunization Strategy

Yiping Chen,[1] Gerald Paul,[1] Shlomo Havlin,[2] Fredrik Liljeros,[3] and H. Eugene Stanley[1]
[1]*Center for Polymer Studies, Boston University, Boston, Massachusetts 02215, USA*
[2]*Minerva Center and Department of Physics, Bar-Ilan University, 52900 Ramat-Gan, Israel*
[3]*Department of Sociology, Stockholm University, S-106 91 Stockholm, Sweden*
(Received 13 February 2008; revised manuscript received 14 May 2008; published 31 July 2008)

The problem of finding the best strategy to immunize a population or a computer network with a minimal number of immunization doses is of current interest. It has been accepted that the targeted strategies on most central nodes are most efficient for model and real networks. We present a newly developed graph-partitioning strategy which requires 5% to 50% fewer immunization doses compared to the targeted strategy and achieves the same degree of immunization of the network. We explicitly demonstrate the effectiveness of our proposed strategy on several model networks and also on real networks.

DOI: 10.1103/PhysRevLett.101.058701 PACS numbers: 89.75.Hc, 64.60.ah, 87.23.Ge, 89.75.−k

There is much interest in the question of how to immunize a population, or a computer network such as the Internet, with a minimal number of immunization doses. This question is very important since in many cases the number of immunization doses is limited or very expensive. This question is mathematically equivalent to asking how to fragment a given network with a minimum number of node removals. To achieve this goal, many immunization strategies have been developed recently [1–8], ranging from local strategies, like acquaintance immunization [2] to global strategies like targeted immunization [3]. Many network structures have been studied, such as Erdős-Rényi (ER) networks [9,10] and random regular graphs [11], but most work focuses on networks with a broad degree distribution such as scale-free (SF) networks, which have features in common with many real networks ranging from the Internet to human contact patterns of importance for the transmission of contagious diseases [6,7,12–15].

It is widely accepted that the most efficient immunization strategies are based on targeted strategies [1–7]. The basic idea of targeted strategies is first to rank the importance of nodes and then remove the nodes from highest importance to lowest until the network becomes disconnected. The importance of nodes is often represented by node degree or betweenness centrality, which is the frequency of appearance of a node in the shortest paths between other nodes [16,17]. The breakdown of the network is quantified by the ratio F of the size of the largest cluster to the total cluster size [3]. See also a related definition of network breakdown used in social sciences [18,19]. To further improve the targeted strategy, it was suggested to apply this method adaptively by recalculating the importance of nodes after every step of node removal [3,4].

In this Letter we propose a novel "equal graph partitioning (EGP)" immunization strategy which we find to be significantly better than targeted methods, with 5% to 50% fewer immunization doses required (on the networks studied here). Our method is based on the heuristic optimal partitioning of graphs [20,21] and is motivated by Refs. [18,22]. The main idea of the EGP is to fragment the network into many connected subnetworks (clusters) of approximately *equal* size. This strategy leads to the need to immunize fewer nodes compared to the targeted strategies. This is since in targeted strategies a broad distribution of cluster sizes appears after fragmentation, including many very small clusters. Hence, one wastes many immunization doses to isolate these small clusters, which is unnecessary in the EGP method. We confirm the improved efficiency of our approach on ER and SF networks, random regular graphs, and on several real networks.

The EGP immunization strategy is based on the nested dissection (ND) algorithm, which was developed to solve sparse systems of linear equations efficiently [20]. The original ND algorithm can separate a network into two equal-size clusters with a minimum number of nodes removed. A network is first divided into three groups (Fig. 1): first cluster, second cluster, and the group of nodes separating the first and second clusters (separator group). To

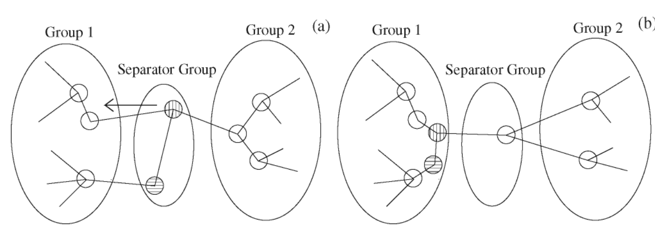

FIG. 1. Diagram showing one attempt of moving nodes in separator group. In (a) we try to move the top node in the separators to group 1 and the result is shown in (b), where the adjacent node of the moved node is dragged into separator group. The bottom node in the separators is dragged to group 1 because all its adjacent nodes are in group 1 and its existence in separators becomes meaningless. After this trial, the size of separator group is reduced from 2 to 1. Thus, this movement is permitted.

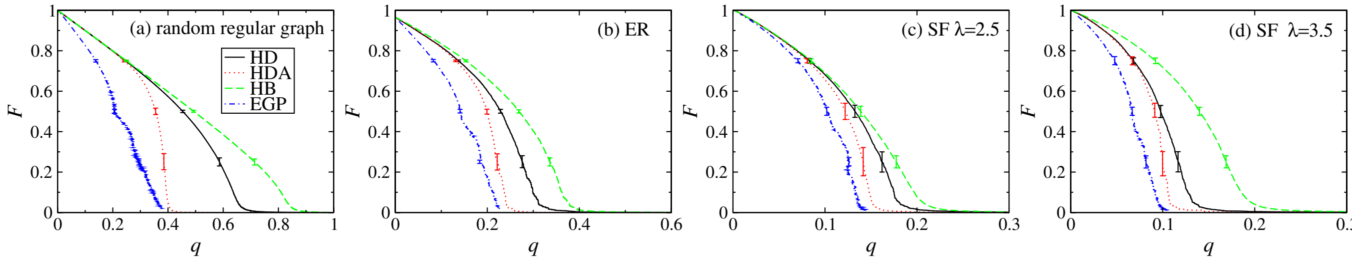

FIG. 2 (color online). The fraction F of the size of the largest cluster that can be infected versus the fraction q of the immunized nodes for HD, HDA, HB, and EGP strategies for (a) random regular graph with $N = 10^4$ and $k = 4$, (b) ER network with $N = 10^4$, $\langle k \rangle = 3.5$, (c) SF network with $N = 10^4$, $\lambda = 2.5$,and $\langle k \rangle = 4.68$, (d) SF network with $N = 10^4$, $\lambda = 3.5$, and $\langle k \rangle = 2.89$. We show also the error bars in F, which are derived from simulating realizations. The error bars for EGP are so small that they are barely seen.

minimize the size of the separator group, the nodes in separator group are attempted to be moved into the first or second cluster. As a result, in each of these trials, the adjacent nodes of the moved node may be dragged in or out of the separator group. The movement is kept if the size of the separator group after the movement is smaller than before. This algorithm is applied iteratively until no further optimization can be gained.

In our implementation of ND algorithm, we separate a network into two clusters with arbitrary size ratio instead of only equal sizes, with the number of separators minimized. In this way, networks can be partitioned into any number of same size clusters by applying the ND algorithm on the network recursively. For example, to partition a network into three equal-size clusters, first we partition the network into two clusters with size ratio 2:1, then apply another partition on the larger cluster with size ratio 1:1. Thus, in order to immunize a network of size N nodes so that only a small fraction F can be infected, one separates the system into $n \approx 1/F$ equal-size clusters. In contrast to the targeted strategies which are based on local decisions, the EGP immunization strategy is closer to global optimization [23].

We test the effectiveness of the EGP strategy by plotting F versus q, the removal fraction of nodes, for ER, SF, and random regular networks (Fig. 2) as well as for several real networks (Fig. 3). We also compare (in Figs. 2 and 3) between the efficiency of the EGP strategy to high degree targeted (HD), high degree adaptive (HDA), and high betweenness (HB) targeted strategies for immunization. In the HDA strategy we apply the high degree targeted method adaptively by recalculating the importance (the node degree) of every node after each node removal. Two values of the degree exponent λ, 2.5 and 3.5, for SF networks are shown. In all four network models tested, our EGP strategy exhibits clear advantage of less nodes to be immunized compared to all targeted strategies (Fig. 2). The curves of EGP strategy are far below curves of HDA, the most effective known targeted strategy. Regarding the threshold point q_c, where F approaches 0, the EGP strategy shows 30% to 50% improvement than nonadaptive and 5% to 10% than adaptive targeted strategies [24].

We also test our EGP immunization strategy on four real networks from different fields. Figure 3 shows simulation results for the workplace network [25], the autonomous system (AS) Internet network [26,27], the high energy particle (HEP) physics citation network [28], and the metabolic network [29]. The workplace network is extracted from a data set obtained from Statistics Sweden [25] and consists of all geographical workplaces in Sweden that can be linked with each other by having at least one employee from each workplace sharing the same household. Household is defined as a married couple or a couple having children together that are living in the same flat or house [30]. This kind of network has been shown to be important for the spreading of influenza [31] and is also likely to be important for spreading of information and rumors in society. The network has 310 136 nodes and

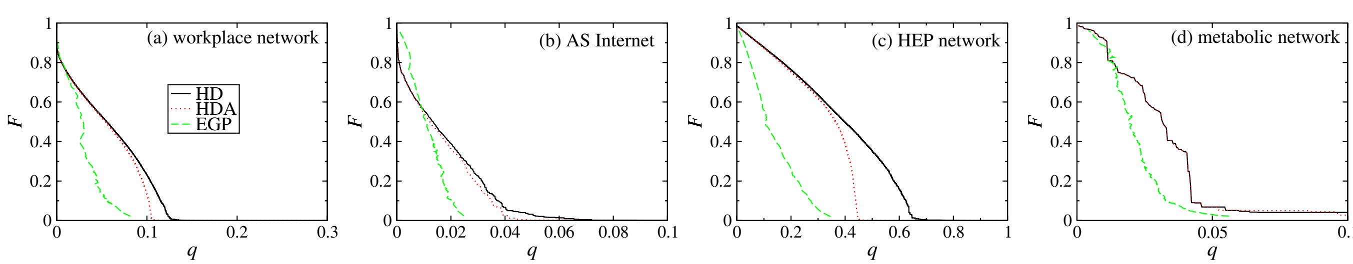

FIG. 3 (color online). The fraction F of the size of the largest cluster that can be infected versus the fraction of immunized nodes q for HD, HDA, and EGP strategies for (a) the workplace network, (b) the AS Internet network, (c) high energy physics HEP network, (d) metabolic network.

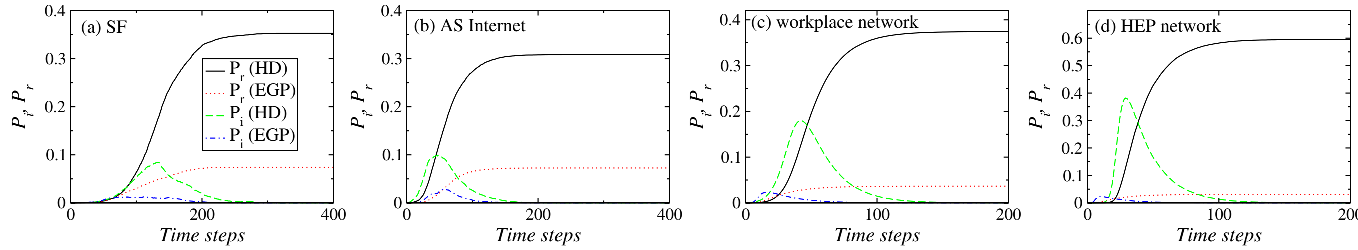

FIG. 4 (color online). Infected network fraction P_i and recovered fraction P_r versus time for the SIR model [17]. Comparison between HD and EGP strategies for (a) immunizing a fraction $q = 0.12$ of the nodes in a SF network with $N = 10^4$, $\lambda = 2.5$, (b) immunizing a fraction $q = 0.02$ in AS Internet network, (c) immunizing a fraction $q = 0.065$ of the nodes in the workplace network, (d) immunizing a fraction $q = 0.31$ of the nodes in the HEP network.

906 260 links. In the example of Fig. 3(a), the advantage of EGP (the value of q for which $F \approx 0$) is about 30% better compared to the nonadaptive targeted strategy and 15% better than the adaptive targeted strategy. The AS network is obtained from the DIMES project [26], which determines the Internet network at the autonomous system or interdomain level. This network has 20 556 nodes and 62 920 links and is very important in the study of computer virus spreading. For the AS network, the EGP method shows a larger improvement, about 50%, against both adaptive and nonadaptive targeted strategies [Fig. 3(b)]. The HEP network contains high energy particle physics citations from the hep-th section of arxiv.org [28]. In this network, a node represents a published paper and a link represents a citation between two papers. Simulation is performed regardless of the direction of citations. This network has 27 770 nodes and 352 807 links, which leads to a high average degree of 12.7. For the HEP network, the EGP algorithm shows an advantage of about 23% compared to adaptive and 46% compared to nonadaptive targeted strategies [Fig. 3(c)]. The metabolic network describes the interactions between the metabolites of *E. coli* in the course of the metabolic cycle, and has 2363 nodes and 5960 links [29]. From Fig. 3(d), this network is special because, unlike most networks, the HD and HDA curves of the metabolic network are almost the same, suggesting that degree recalculation is not necessary during the targeting process. However, the EGP method still gives an advantage of about 20%. It should be noted that the performance efficiency of the EGP is better in real networks compared to model networks. This is probably due to the structural behavior of the real networks that represent communities making it easier for the EGP to separate the networks into clusters compared to model networks.

To further estimate the immunization effectiveness using the EGP strategy as compared to HD strategy, we also studied the disease spreading behavior using the susceptible-infectious-recovery (SIR) epidemic spreading model [8,32] on SF model networks, the AS Internet network, the workplace network, and the HEP network. The SIR model is an epidemiological model widely used to simulate the spreading of epidemics, i.e., number of people infected with a contagious illness, in a closed population as a function of time [33]. At every time step, the model assumes a transition rate α for a susceptible person to become infected, if an infected neighbor is present, and a rate β for an infected person to become recovered or die. The recovered person will never be infected again. The simulation results are shown in Fig. 4. In the simulations we use $\alpha = 0.2$ and $\beta = 0.05$. For all networks studied here, the infected fraction is significantly (5 to 10 times) lower when using the EGP strategy compared to the HD strategy with the same fraction of immunization doses.

Next, we study, for ER and SF networks, the dependence of the threshold q_c for different immunization strategies as a function of the network parameters (Fig. 5). The threshold q_c is defined as the fraction of nodes immunized or

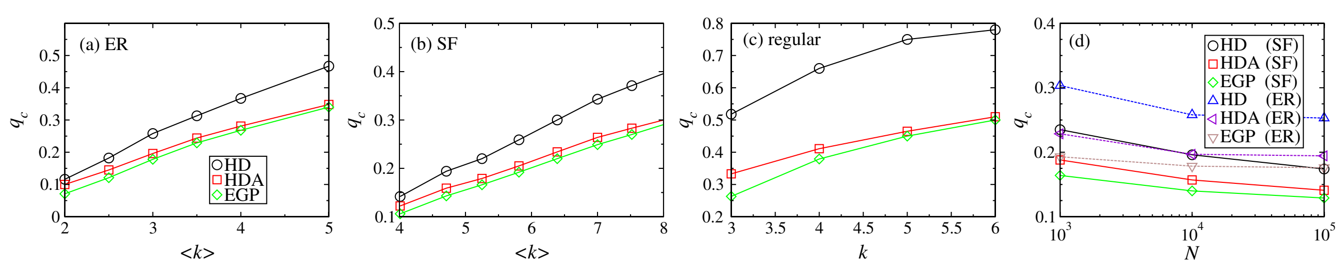

FIG. 5 (color online). The threshold q_c versus average mode degree $\langle k \rangle$ for HD, HDA, and the EGP strategies for (a) ER networks with $N = 10^4$, (b) SF networks with $N = 10^4$ and $\lambda = 2.5$, (c) random regular graph with $N = 10^4$, (d) the critical threshold q_c versus system size N for SF networks with $\lambda = 2.5$ and ER networks with $\langle k \rangle = 3$.

removed for which $F \approx 0$. It is expected that q_c increases with increasing the average degree $\langle k \rangle$, since it is harder to fragment the network apart when more links between nodes exist. The EGP method compared to adaptive high degree strategy is more significant for lower values of $\langle k \rangle$. This is due to the fact that when $\langle k \rangle$ increases, the relation between potential clusters becomes closer. More internal nodes in potential clusters will have external links connecting to other clusters. Thus, it is harder to reduce the size of the separator group. However, as noted above, this effect seems to be less pronounced in real structured networks. We also study the dependence of q_c on the system size N. Figure 5(d) shows q_c versus system size N for SF networks with $\lambda = 2.5$ and for ER networks with $\langle k \rangle = 3$. The value of q_c is affected by the finite size of the system. When N increases, q_c decreases and approaches its asymptotic value. However, the ratio between q_c of EGP and targeted strategies does not seem to be affected by N.

To summarize, we have developed and applied a new EGP algorithm as an efficient network immunization strategy that partitions a network into clusters of approximately *equal* size. This strategy saves a large amount of immunization doses since immunizing small clusters is not needed. We find that our method performs significantly superiorly compared to previously known effective targeted strategies.

The study of the workplace network was approved by the Regional Ethical Review board in Stockholm, record 2004/2:9. We thank ONR, Israel Complexity Center, and the Israel Science Foundation for support.

[1] P. Holme, Europhys. Lett. **68**, 908 (2004).
[2] R. Cohen, S. Havlin, and D. ben-Avraham, Phys. Rev. Lett. **91**, 247901 (2003).
[3] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, Phys. Rev. E **65**, 056109 (2002).
[4] D. Magoni, IEEE J. Sel. Areas Commun. **21**, 949 (2003).
[5] L. K. Gallos *et al.*, Phys. Rev. Lett. **94**, 188701 (2005).
[6] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 036104 (2002).
[7] J. Gómez-Gardeñes, P. Echenique, and Y. Moreno, Eur. Phys. J. B **49**, 259 (2006).
[8] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).
[9] P. Erdős and A. Rényi, Publ. Math. **6**, 290 (1959); Publ. Math. **5**, 17 (1960).
[10] B. Bollobas, *Random Graphs* (Cambridge University Press, Cambridge, England, 2001).
[11] A. Bekessy, P. Bekessy, and J. Komlos, Stud. Sci. Math. Hung. **7**, 343 (1972).
[12] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[13] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos, IEEE/ACM Trans. Netw. **11**, 514 (2003).
[14] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).
[15] SF networks possess a power law degree distribution, $p(k) \sim k^{-\lambda}$, where k is the degree of a node.
[16] L. C. Freeman, Sociometry **40**, 35 (1977).
[17] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[18] S. P. Borgatti, Comput. Math. Org. Theory **12**, 21 (2006).
[19] Y. Chen, G. Paul, R. Cohen, S. Havlin, S.P. Borgatti, F. Liljeros, and H. E. Stanley, Phys. Rev. E **75**, 046107 (2007).
[20] R. J. Lipton, D. J. Rose, and R. E. Tarjan, SIAM J. Numer. Anal. **16**, 346 (1979).
[21] G. Karypis and V. Kumar, SIAM J. Sci. Comput. **20**, 359 (1998).
[22] G. Paul, R. Cohen, S. Sreenivasan, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **99**, 115701 (2007).
[23] To further improve the performance, the multilevel method is employed in our simulations [21]. A network is first coarsened into smaller networks via several levels by combining nodes and links using a heavy edge matching algorithm [21]. On each coarsening step, a number of nodes is randomly selected and collapsed with one of their connected neighbors. The weights of the resulting nodes are equal to the number of original nodes each of them contains. Links may also need to collapse as the result of the node collapse, and the weights of new links are the total number of original links they contain. The coarsened network is partitioned into the desired number of clusters using the ND algorithm by keeping the total weights of nodes in each cluster equal and minimizing the total weights of links between different clusters. Finally the network is uncoarsened back to the original network with ND refinement at each step.
[24] The computational complexity of EGP strategy is also investigated using simulations and found to be close to linear in N. The CPU time needed for EGP strategy is slightly higher compared to HD and much less than HDA and HB for both ER and SF networks.
[25] http://www.scb.se.
[26] http://netdimes.org.
[27] Y. Shavitt and E. Shir, Computer Communication Review **35**, 71 (2005).
[28] http://vlado.fmf.uni-lj.si/pub/networks/data/hep-th/hep-th.htm.
[29] http://www.nd.edu/~networks/resources/metabolic/index.html.
[30] Unmarried couples without kids and other individuals sharing a household are not registered in the data set as households.
[31] C. Viboud, O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell, Science **312**, 447 (2006).
[32] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[33] W. O. Kermack and A. G. McKendrick, Proc. R. Soc. A **115**, 700 (1927).