INFORMATION SCIENCE

Efficient network immunization under limited knowledge

Yangyang Liu[1,†], Hillel Sanhedrai[2,†], GaoGao Dong[3,*], Louis M. Shekhtman[2,4,*], Fan Wang[2], Sergey V. Buldyrev[5] and Shlomo Havlin[2,*]

ABSTRACT

Targeted immunization of centralized nodes in large-scale networks has attracted significant attention. However, in real-world scenarios, knowledge and observations of the network may be limited, thereby precluding a full assessment of the optimal nodes to immunize (or quarantine) in order to avoid epidemic spreading such as that of the current coronavirus disease (COVID-19) epidemic. Here, we study a novel immunization strategy where only n nodes are observed at a time and the most central among these n nodes is immunized. This process can globally immunize a network. We find that even for small n (≈ 10) there is significant improvement in the immunization (quarantine), which is very close to the levels of immunization with full knowledge. We develop an analytical framework for our method and determine the critical percolation threshold p_c and the size of the giant component P_∞ for networks with arbitrary degree distributions $P(k)$. In the limit of $n \to \infty$ we recover prior work on targeted immunization, whereas for $n = 1$ we recover the known case of random immunization. Between these two extremes, we observe that, as n increases, p_c increases quickly towards its optimal value under targeted immunization with complete information. In particular, we find a new general scaling relationship between $|p_c(\infty) - p_c(n)|$ and n as $|p_c(\infty) - p_c(n)| \sim n^{-1} \exp(-\alpha n)$. For scale-free (SF) networks, where $P(k) \sim k^{-\gamma}$, $2 < \gamma < 3$, we find that p_c has a transition from zero to nonzero when n increases from $n = 1$ to $O(\log N)$ (where N is the size of the network). Thus, for SF networks, having knowledge of $\approx \log N$ nodes and immunizing the most optimal among them can dramatically reduce epidemic spreading. We also demonstrate our limited knowledge immunization strategy on several real-world networks and confirm that in these real networks, p_c increases significantly even for small n.

Keywords: percolation, complex networks, network immunization, critical phenomena

[1]Department of Systems Science, National University of Defense Technology, Changsha 410073, China; [2]Department of Physics, Bar-Ilan University, Ramat Gan 5290002, Israel; [3]Faculty of Science, Jiangsu University, Zhenjiang 212013, China; [4]Network Science Institute, Northeastern University, Boston, MA 02115, USA and [5]Department of Physics, Yeshiva University, New York, NY 10033, USA

*Corresponding authors. E-mails: dfocus.gao@gmail.com; lsheks@gmail.com; havlins@gmail.com
†Equally contributed to this work.

INTRODUCTION

Networks play a crucial role in many diverse systems [1–11]. Connectivity of components is critical for maintaining the functioning of infrastructures like the Internet [12] and transportation networks [13], as well as for developing efficient immunization against epidemics [14] and the spread of information in social systems [15]. Because of this importance, researchers have long focused on how a network can be optimally immunized or fragmented to prevent epidemics or maintain infrastructure resilience [16–20]. Many approaches have used percolation theory from statistical physics to prevent the spread of viruses or assess network resilience under the infection or failure of some fraction of nodes or links [21–32].

Early studies in networks found that immunizing real networks against an epidemic is highly challenging due to the existence of hubs that prevent eradication of the virus even if many nodes are immunized [33–35]. These largest degree nodes, are targeted, the network can quickly be separated, leading to immunity or failure [36]. However, previous models of targeted immunization have assumed full knowledge of the network structure that in most cases is not available. Research has shown that even those in control of a network often have knowledge of only a small portion of the whole network structure [37–41]. This has been further demonstrated with the current coronavirus disease 2019 (COVID-19) epidemic where the detailed social network of individuals is unknown.

In this paper, we study targeted immunization in networks with limited knowledge. Note that the methodology equally applies to efficient

attacks (where the targeted nodes are removed) on networks. We assume that at each stage, n nodes are observed and the node with highest degree is immunized and thus unable to continue spreading infection. This procedure is repeated until a $1 - p$ fraction of nodes are immunized. In particular, our model could apply to a situation where several co-operative teams are sent to immunize a network and each team has access to information on a small subset (n nodes) of the network. We develop a theoretical framework for this model of immunization with limited information using percolation theory for networks with arbitrary degree distribution. In the limit of $n = N$ we recover prior work on targeted immunization [36], whereas for $n = 1$ we recover the case of random immunization [34,35]. We observe excellent agreement between our theoretical framework and simulation results regarding the efficiency of the immunization as measured by both the critical threshold p_c and the size of the giant component P_∞ for $p > p_c$ as a function of n. We note that a smaller giant component implies improved immunization, whereas a larger critical threshold p_c is also more efficient since fewer quarantines/vaccines are needed to stop the epidemic. We find a scaling relationship between n and p_c for both Erdős-Rényi (ER) and scale-free (SF) networks, which we determine both analytically and through simulations. Surprisingly, we find that p_c quickly reaches a large value and a plateau even for relatively small n (of order 10), after which increasing n has negligible effect on p_c. This suggests that obtaining information on even a small number (n) of people can lead to significant mitigation of the epidemic.

RESULTS

Model

Let $G(V, E)$ be a network where V and E are the sets of nodes and edges, respectively. The number of nodes in the network is $N = |V|$. We assume that an immunizing team has limited knowledge of the overall network structure and instead possesses only limited information on several nodes. Specifically, we randomly select n nodes for which the immunizing team is assumed to obtain information on their node degree. We then immunize the node with the highest degree among these n. This could be performed by many teams such that they collectively immunize a fraction of $1 - p$ nodes. The process leads to identical results if instead of immunizing the nodes they are put into quarantine. We note that the immunizing team only retains knowledge of the initial node degree and does not reduce the degree based on any neighbors that might have been immunized or quarantined. This is due to the fact that when asking

individuals about their neighbors, the individuals are unlikely to know that a neighbor has recently been immunized or quarantined.

In Fig. 1, the limited information immunization is illustrated and compared to the global targeted immunization on a network. Here a total of $n = 3$ nodes are observed. In panel (a), an individual with global information about the network structure chooses the highest degree node u to immunize. However, in panel (b), the individual knows only the degree of three nodes in the network at any time, i.e. v_1, v_2, v_3. Consequently, node v_3 with the highest degree $k = 4$ (marked in red) is immunized.

ER networks

We now study our limited knowledge immunization strategy, i.e. the general result, Equations (M8) and (M9) in the Methods section, on ER networks. First, we analyze the giant component P_∞. For the case $n = 1$, limited knowledge immunization reduces to the classical random immunization, while for $n \to \infty$ (meaning that the global network is observed) corresponds to targeted immunization [17,33,35,36]. Using Equation (M9) in the Methods section, the giant component P_∞ can be solved numerically for any given p. In Fig. 2(a), simulations and analytical results are shown for the giant component P_∞ as a function of $1 - p$ under limited information immunization for different n. As the knowledge index n increases from 1 to N, the limited knowledge immunization moves from being like random immunization to being like targeted immunization. The simulations are in excellent agreement with the theoretical results (lines).

Next, we focus on the critical threshold, p_c, of limited knowledge immunization. Overall, we find that one does not need much knowledge of the network to improve the immunization and $n \sim 10$ is enough to achieve similar results as for targeted immunization with complete information. This can be seen by observing the critical threshold p_c as a function of n in Fig. 2(b). In Fig. 2(c) we show the variation of p_c with $\langle k \rangle$ for several fixed n.

We can also derive the behavior of p_c in the limit of large n analytically. By examining Equation (M5) in the Methods section we note that, when $n \to \infty$, there are two distinct behaviors depending on whether k is small, $F(k) < p$; or k is large, $F(k) > p$. It can be shown (see Section II.C of the online supplementary material) that the leading term behaves as

$$F_p(k) \sim \begin{cases} \frac{F(k)}{p} - \frac{1}{n}e^{-\alpha_k n}, & F(k) < p, \\ 1 - \frac{1}{n}e^{-\alpha_k n}, & p < F(k) < 1, \\ 1, & F(k) = 1, \end{cases}$$
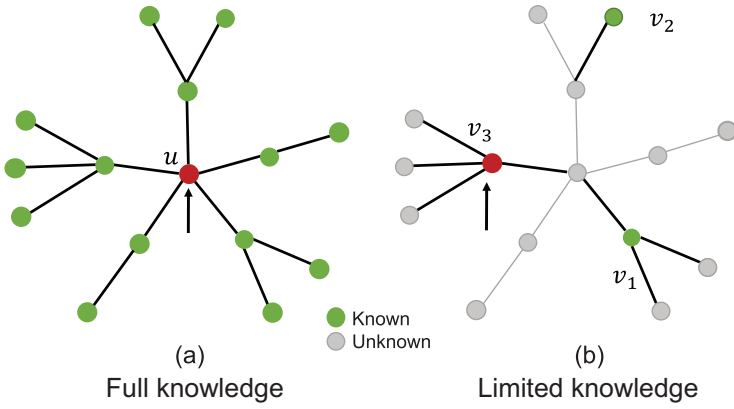
$$(1)$$

Figure 1. Schematic illustration of our limited knowledge immunization strategy. The immunizer is able to observe the degree of nodes that are colored green, while the gray nodes are unknown. (a) For the classical targeted immunization, one has complete information on the global structure of the network and chooses the highest degree node (u) to immunize. (b) Here the case of an individual with limited knowledge of the network is demonstrated. In this figure, we set $n = 3$ and only the degrees of nodes v_1, v_2 and v_3 are known. Given this limited information, the immunizer would choose to immunize v_3, being unaware that an unobserved higher degree node exists. At the next immunization, only nodes that have not been immunized yet will be observed.
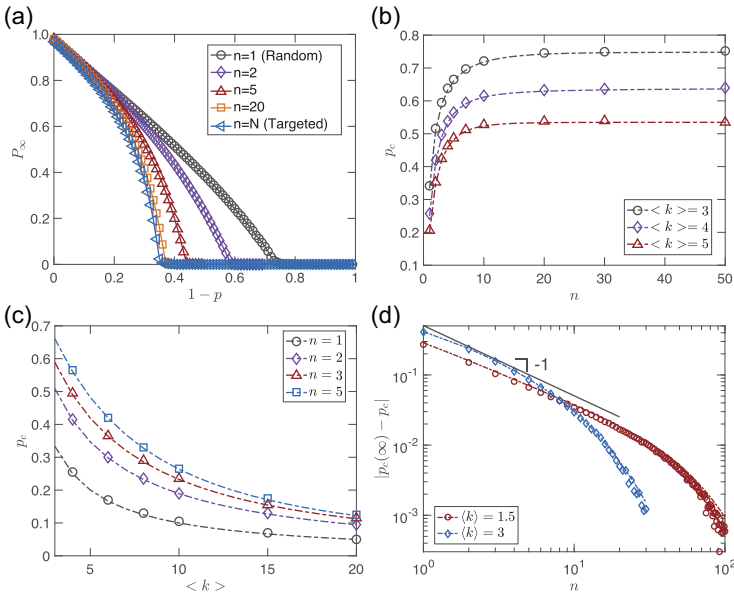

Figure 2. Results on ER networks. (a) The giant component P_∞ of an ER network with $\langle k \rangle = 4$ varies with the fraction of immunized nodes $1 - p$ under limited knowledge. As n is increased, the limited knowledge immunization tends to have the same immunization effect as targeted immunization. (b) The critical threshold p_c of limited knowledge immunization as a function of n on ER networks. The values of p_c were determined from the p of the maximal second largest clusters [28]. Note that already for small $n \sim 10$, p_c is high and close to targeted immunization (global knowledge, $n \sim N$). (c) Critical threshold p_c as a function of the mean degree $\langle k \rangle$ of ER networks for limited knowledge immunization. (d) The scaling of $|p_c(\infty) - p_c|$ with n on ER networks. Symbols are average results of simulations over 100 independent realizations of ER networks with 10^6 nodes. All simulation results (symbols) agree well with theoretical results of Equation (M10) in the Methods section (dashed lines).

where $\alpha_k = |\log[p/F(k)]|$. In the limit $n \to \infty$, we can get the expected result for targeted immunization, $F_p(k) = \min\{F(k)/p, 1\}$ [35,36].

Substituting Equation (1) into Equation (M10) in the Methods section and noting that, from a sum of exponentials decaying with n, only the lowest decay rate contributes to the leading term, we obtain (see Section II.C of the online supplementary material)

$$p_c(n) \sim p_c^\infty - A\frac{1}{n}e^{-\alpha n}, \qquad (2)$$

where $p_c^\infty = p_c(n \to \infty)$, and the decay rate α is now

$$\alpha = \min_k |\log(p_c^\infty/F(k))|.$$

The prefactor is $A = (2p_c^\infty k_{slow})/(k_> k_<)$, where $k_<$ is the largest degree such that $F(k) < p_c^\infty$, $k_> = k_< + 1$ and k_{slow} is the degree that gives the lowest rate α. (See the illustration in the online supplementary material).

It is clear that k_{slow} must be $k_<$ or $k_>$ because $F(k)$ is monotonic. If $F(k_{slow}) = F(k_>) = 1$ then $k_<$ should be taken as k_{slow}, and the corresponding α should be taken. It should also be noted that if k_{slow} is not unique, it would simply change the prefactor A in Equation (2). Another special case is where $F(k_{slow}) = p_c^\infty$; then $|p_c^\infty - p_c| \sim 1/n$ (see Section III of the online supplementary material).

In Fig. 2(d) we show $\Delta p_c = |p_c^\infty - p_c|$ as a function of n. As expected from the theory, we can see that $\Delta p_c \sim 1/n$ for small n and exponential decay for large n. When $p_c \to 1$, which occurs for the ER network when $\langle k \rangle \to 1$, the power law regime becomes much broader, as explained in Section II.C of the online supplementary material.

SF networks

Next, we study SF networks with $P(k) = Ak^{-\gamma}$, $k = m, \cdots, K$, where $A = (\gamma-1)m^{\gamma-1}$ is the normalization factor, and m and K are the minimum and maximum degrees, respectively [33,34]. Similar to ER networks, the size of the giant component P_∞ can be obtained from Equation (M9) in the Methods section. In Fig. 3(a), we show P_∞ as a function of $1 - p$ for different n values. The results demonstrate that SF networks become more immunized as n increases. Compared with ER networks, we can observe that slightly higher values of n (more knowledge) are needed to reach the near-steady-state region of a fully targeted strategy. In Fig. 3(c) we show how p_c varies with γ for small n for the SF network where the minimum and maximum degrees of the network are $m = 2$ and $K = 1000$, respectively. In Fig. 3(d) we show $\Delta p_c = |p_c^\infty - p_c|$ as
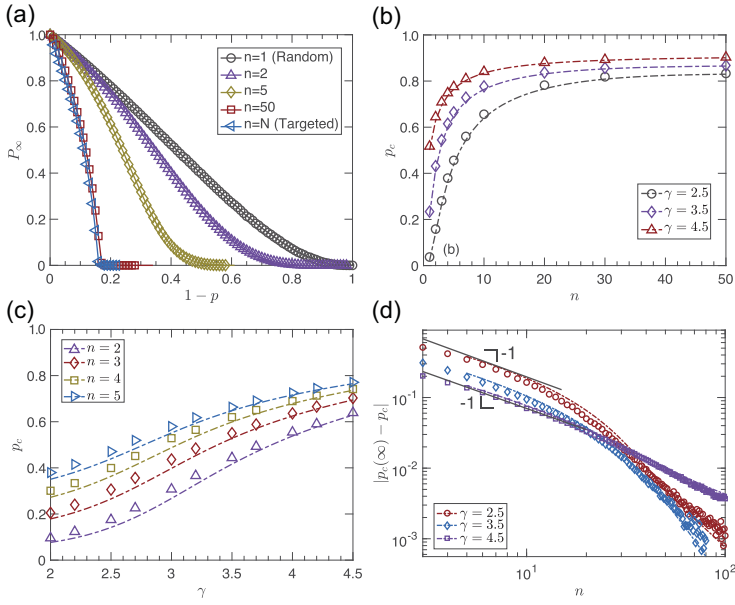
Figure 3. Results for SF networks. Comparison of theory (lines) and simulation (symbols) for limited knowledge immunization, n, for SF networks. (a) The size of the giant component versus n for the SF network with $\gamma = 2.5$. (b) Critical threshold p_c versus n on SF networks. (c) The critical threshold p_c as a function of γ for different values of n. (d) The log-log plot of $|p_c(\infty) - p_c|$ with n on SF networks showing both a power law and exponential behavior. Simulations are obtained for networks of 10^6 nodes and averages are taken over 100 independent realizations. The minimum and maximum degrees of the network are $m = 2$ and $K = 1000$, respectively. The theoretical results (dashed lines) are calculated from Equation (M10) in the Methods section and are in excellent agreement with the simulation results (symbols).

a function of n. As expected from the theory, Equation (2), the log-log plot captures the slope of -1 for small n while an exponential regime exists for large n.

For SF networks with $2 < \gamma < 3$, under random immunization ($n = 1$), it has been shown that $p_c = 0$ for an infinite system [34], while for high-degree immunization ($n \to \infty$), $p_c > 0$ [35,36]. Next we determine for which n, p_c becomes nonzero and how it depends on the system size N. To this end, we analyze Equations (M5) and (M10) in the Methods section for large k (high degrees govern the behavior in SF networks) and small n and p as follows (elaborated in Section II.D of the online supplementary material). It can be shown that, for large degrees,

$$F(k) \approx 1 - (k/m)^{1-\gamma}.$$

Substituting this into Equation (M5) in the Methods section and assuming that $(k/m)^{\gamma-1} \gg n$ for large degrees, it can be concluded that

$$P_p(k) \approx p^{n-1} P(k). \tag{3}$$

In addition, we note that $P_p(k)$ has a new natural cutoff, K_p, which depends on p and N as

(see Section II.D of the online supplementary material)

$$K_p \sim p^{n/(\gamma-1)} N^{1/(\gamma-1)}.$$

This helps us to evaluate the second moment of $P_p(k)$:

$$\langle k_p^2 \rangle \sim \int_m^{K_p} k^2 p^{n-1} A k^{-\gamma}$$

$$\sim p^{n-1} K_p^{3-\gamma} \sim p^{n-1+n\beta} N^\beta.$$

Here $\beta = (3 - \gamma)/(\gamma - 1)$.

Considering this, and substituting Equation (3) into Equation (M10) in the Methods section, keeping the leading terms in the limit of large N, we obtain (see Section II.D of the online supplementary material for further details)

$$p_c \sim C(n) N^{-\delta/n} \sim C(n) \exp\left[-\delta \frac{\log N}{n}\right], \tag{4}$$

where

$$\delta = \frac{\beta}{1+\beta} = \frac{3-\gamma}{2}$$

and

$$C(n) = \phi^{1/n}, \quad \phi = \left(\frac{3-\gamma}{\gamma-2} \frac{1}{m}\right)^{(\gamma-1)/2}.$$

From Equation (4), it is easy to see that if $n \ll \log N$ then $p_c \to 0$, while if $n \sim \log N$ then p_c is nonzero. Note that the prefactor $C(n)$ depends on n but not on N.

In Fig. 4(a) we show p_c versus γ. It is known that, for $2 < \gamma < 3$ and $n = 1$, if $N \to \infty$ then $p_c \to 0$ [34]. Also, for $n = 5$, we can see that system size matters and p_c decreases as N increases. In Fig. 4(b) we show that the scaling with $n/\log N$ of Equation (4) is valid. Furthermore, it can be seen from Fig. 4(b) that, when n is small or N is large, such that $n/\log N \ll 1$ (in Fig. 4(b) it is 0.07), p_c approaches 0. Figure 4(c) supports the exponential scaling of p_c versus $n^{-1}\log N$ obtained analytically in Equation (4).

Real networks

Our analytical framework is derived and well supported, as shown above, on an ensemble of networks generated by the random configuration model. Here we also test our analytical framework on six real-world networks ranging from computer to social networks. They include (a) autonomous computer networks from the Skitter project [42], (b) the reply network of the social news website Digg [43,44], (c) the network of autonomous systems of the Internet from the CAIDA project [42], (d) the social
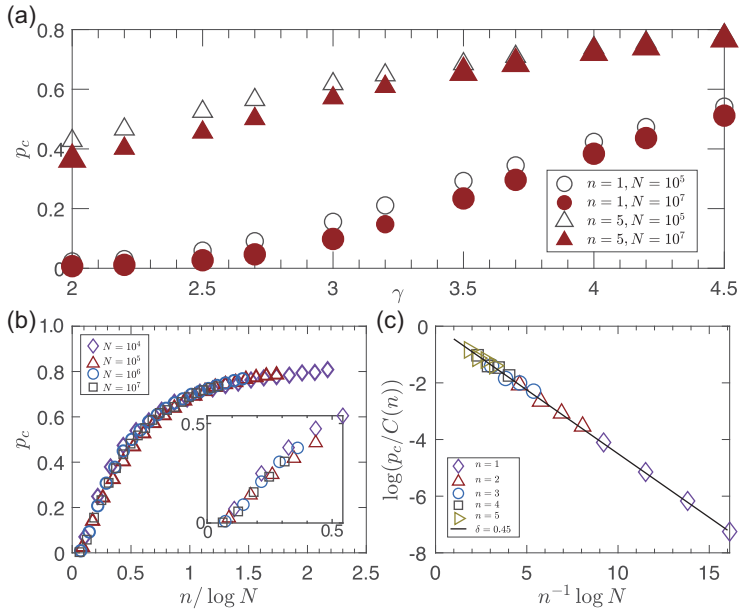
Figure 4. How p_c for SF networks depends on γ and the system size N. (a) The critical threshold p_c as a function of γ for different values of n and N; p_c decreases with increasing N. For $n = 1$ and $2 < \gamma < 3$, it is well known that p_c approaches zero for an infinite system. (b) The critical threshold p_c as a function of $n/\log N$ for the SF network with $\gamma = 2.1$. The inset shows magnifying the data for small $n = 1, \ldots, 5$; note that p_c approaches zero for $n/\log N \ll 1$. (c) The scaling of p_c with N and n for large N and small n. Here $C(n)$ is the prefactor. The minimum and maximum degrees of the nodes are $m = 2$ and $K = N^{1/(\gamma-1)}$, respectively. This confirms Equation (4) for $\gamma = 2.1$ and $\delta = (3 - \gamma)/2 = 0.45$.

friendship network from Douban [45], (e) tech-p2p, the eDonkey peer-to-peer network [43,46] and (f) sc-rel9, the scientific computing network [43,47]. Detailed information and statistical features of these networks are summarized in Section IV of the online supplementary material. To test our theory, we execute the limited knowledge immunization strategy on the above real-world networks and use the peak of the second giant component to determine the critical point p_c, as shown in Fig. 5. For six real-world networks, we plot the distribution of the critical point over 200 independent realizations, together with the theoretical results. As can be seen from Fig. 5, the critical thresholds p_c for the empirical data are generally consistent with our theoretical predictions. Moreover, we find that our limited knowledge immunization approach is very efficient at preventing outbreaks even for small n of order 10.

DISCUSSION

In summary, our results provide a framework for understanding and carrying out efficient immunization (or efficient attack) with limited knowledge. Especially in cases of global pandemics, e.g. the current COVID-19 epidemic, it is impossible to know the full interactions of all individuals and immunize the most central. We find that an effective
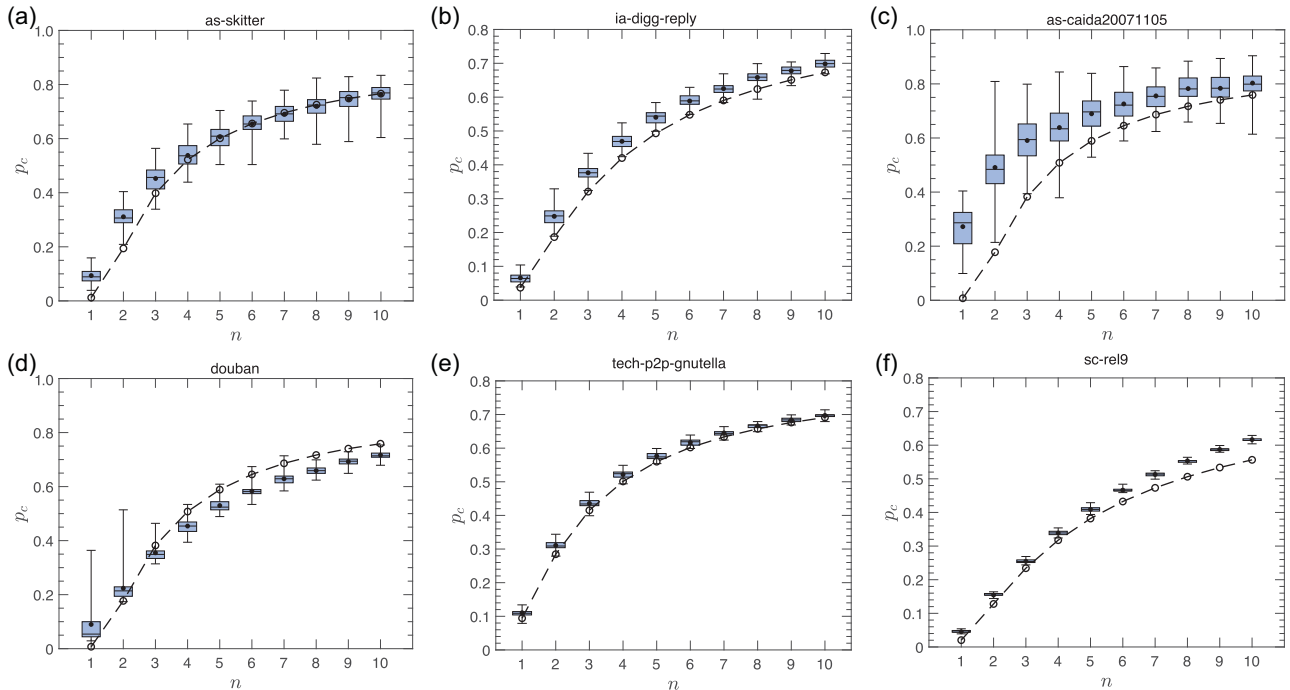

Figure 5. Limited knowledge efficient immunization on six real-word networks (a–f). Each panel is the result of 200 independent simulations. In each panel, filled circles, horizontal lines and dashed lines represent the mean, median value and analytical results of Equation (M10) in the Methods section, respectively. In the simulations, we use the peak of the second largest component to identify the critical point p_c.

way to limit spreading is to obtain information on a few (n of the order of 10) individuals and target (test or quarantine) the most central. For example, testers could stand at a supermarket and select a small number of people entering the store simultaneously. Information on the connections of these people, e.g. the number of people they live with or work closely with, where and how often they meet with other people etc., could be obtained quickly through interviews or through cell phone tracking and only the individual with the most connections in the group could be tested, quarantined or immunized. Our results demonstrate that even when this is done in small groups of people (low n on the order of 10), it is possible to obtain a significant decrease in global epidemics compared to randomly selecting individuals. In our model, this can be seen by the significantly reduced size of the giant component and the significantly high critical threshold p_c. Overall, these findings could help to develop efficient approaches for immunizing large networks and designing resilient infrastructures.

Finally, we note the following two possible variations of our model. In one approach, once we get the degree of a node, we could also obtain partial information (some links) of neighboring nodes and utilize that information to target the highest degree node. We find that adding this knowledge does not change the key results of the proposed model. For detailed results, we refer the reader to Section V of the online supplementary material. Additionally, our immunization approach assumes the existence of a link as binary. Another variant of the model that would be interesting is to incorporate difference frequencies of interactions between different nodes. This would suggest that some nodes are likely to infect some of their neighbors more than others. Such a setting could be implemented through a weighted network, which would be an interesting topic for future study.

METHODS

Suppose that the degree distribution of a network is given by $P(k)$, and $F(k) = \sum_{s=0}^{k} P(s)$ is the cumulative probability that the degree of a randomly chosen node is less than or equal to k. Furthermore, at an arbitrary time t during the iterative percolation process, assume that the distribution of the original degree (including the immunized neighbors) of the remaining nodes is $P(k, t)$. Then the degree distribution of the node that is immunized at time t is given by

$$P_r(k, t) = F(k, t)^n - F(k - 1, t)^n$$
$$\equiv \Delta[F(k, t)^n], \qquad (M1)$$

where $F(k, t)$ is the cumulative distribution of $P(k, t)$. This formula can be recognized as being derived from 'order statistics' giving the 'maximum' of several independent random variables [48] (see Section II.A of the online supplementary material). For $k = 0$, Equation (M1) becomes $P_r(0, t) = F(0, t)^n$. Hence, we define $F(k = 1, t) = 0$, and then Equation (M1) is valid for $k \geq 0$.

In a limited knowledge immunization, each node's immunization changes the degree distribution of the remaining nodes as

$$N(k, t + 1) = N(k, t) - P_r(k, t), \qquad (M2)$$

where $N(k, t)$ is the number of nodes with degree k at time t and $P_r(k, t)$ is the likelihood that a node immunized at time t has degree k.

Then, substituting Equation (M1) into Equation (M2) gives

$$N(k, t + 1) = N(k, t) - \Delta[F(k, t)^n],$$

which in the continuous limit becomes

$$\frac{\partial N(k, t)}{\partial t} = -\Delta[F(k, t)^n].$$

Substituting $N(k, t) = (N - t)P(k, t)$ yields

$$-P(k, t) + (N - t)\frac{\partial P(k, t)}{\partial t}$$
$$= -\Delta[F(k, t)^n],$$

and using $P(k, t) = \Delta F(k, t)$, we obtain

$$\Delta\left[-F(k, t) + (N - t)\frac{\partial F(k, t)}{\partial t} + F(k, t)^n \right] = 0.$$

Note that $F(k = -1, t) = 0$, and thus the entire term inside the Δ is 0 for $k = -1$. Similarly, this implies that, for $k = 0$ and likewise for any $k \geq 0$, this term is also 0. Thus, we obtain the simple ordinary differential equation

$$(N - t)\frac{\partial}{\partial t} F(k, t) = F(k, t) - F(k, t)^n \qquad (M3)$$

with the initial condition $F(k, t = 0) = F(k)$. It can be shown (see Section II.B of the online supplementary material) that the solution of Equation (M3) is

$$F(k, t) = (1 + (F(k)^{1-n} - 1)$$
$$\times e^{(n-1)\log[(N-t)/N]})^{-1/(n-1)}, \qquad (M4)$$

or, equivalently,

$$F_p(k) = \left(1 + (F(k)^{1-n} - 1)p^{n-1}\right)^{-1/(n-1)}, \tag{M5}$$

where $F_p(k)$ is the cumulative distribution of the degree after immunizing a $1 - p$ fraction of nodes. For $n = 1$, the solution of Equation (M3) is $F_p(k) = F(k)$, as expected. Also, Equation (M5) converges to $F(k)$ in the limit $n \to 1$.

We can now obtain the degree distribution of the occupied nodes after a $1 - p$ fraction of nodes are immunized, which is given by

$$P_p(k) = \Delta F_p(k) = F_p(k) - F_p(k-1). \tag{M6}$$

The equation for v, the probability of a randomly chosen link to lead to a node not in the giant component, is

$$1 - v = \sum_{k=0}^{\infty} \frac{kP(k)}{\langle k \rangle} P(\Theta|k)(1 - v^{k-1}), \tag{M7}$$

where $P(\Theta|k)$ is the probability of a node to be occupied given its degree is k. This self-consistent equation can be recognized as being derived based on the generating function method [3]. Using Bayes theorem, we note that $P(k)P(\Theta|k) = P(\Theta)P(k|\Theta) = pP_p(k)$. Hence, Equation (M7) becomes

$$1 - v = \frac{p}{\langle k \rangle} \sum_{k=0}^{\infty} kP_p(k)(1 - v^{k-1}). \tag{M8}$$

The giant component is given by

$$\begin{aligned}
P_\infty &= \sum_{k=0}^{\infty} P(k)P(\Theta|k)(1 - v^k) \\
&= p \sum_{k=0}^{\infty} P_p(k)(1 - v^k),
\end{aligned} \tag{M9}$$

where v is found from Equation (M8).

At criticality, we take the derivative of both sides of Equation (M8) and substitute $v = 1$, representing the location where the first solution with $v < 1$ exists, as opposed to only the $v = 1$ solution. Thus, the critical condition is

$$1 = \frac{p_c}{\langle k \rangle} \sum_{k=0}^{\infty} k(k-1)P_{p_c}(k). \tag{M10}$$

SUPPLEMENTARY DATA

Supplementary data are available at *NSR* online.

AUTHOR CONTRIBUTIONS

Y.L., G.D., L.S. and S.H. designed the research, Y.L., H.S. and S.B. analyzed the model, Y.L., H.S., G.D. and F.W. performed the experiments, and all authors contributed to the writing of the manuscript.

Conflict of interest statement. None declared.

REFERENCES

1. Newman MEJ. The structure and function of complex networks. *SIAM Rev* 2003; **45**: 167–256.
2. Albert R and Barabási AL. Statistical mechanics of complex networks. *Rev Mod Phys* 2002; **74**: 47.
3. Cohen R and Havlin S. *Complex Networks: Structure, Robustness and Function.* Cambridge: Cambridge University Press, 2010.
4. Cohen R, Havlin S and Ben-Avraham D. Efficient immunization strategies for computer networks and populations. *Phys Rev Lett* 2003; **91**: 247901.
5. Brockmann D and Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *Science* 2013; **342**: 1337–42.
6. Shi D, Lü L and Chen G *et al.* Totally homogeneous networks. *Natl Sci Rev* 2019; **6**: 962–9.
7. Liu YY, Csóka E and Zhou HJ *et al.* Core percolation on complex networks. *Phys Rev Lett* 2012; **109**: 205703.
8. Liu YY, Slotine JJ and Barabási AL. Controllability of complex networks. *Nature* 2011; **473**: 167–73.
9. Zhang Z and Zhang S. Towards understanding residual and dilated dense neural networks via convolutional sparse coding. *Natl Sci Rev* 2020; doi: 10.1093/nsr/nwaa159.
10. Watts DJ. A simple model of global cascades on random networks. *Proc Natl Acad Sci USA* 2002; **99**: 5766–71.
11. Tang D, Du W and Shekhtman LM *et al.* Predictability of real temporal networks. *Natl Sci Rev* 2020; **7**: 929–37.
12. Albert R, Jeong H and Barabási AL. Internet: diameter of the world-wide web. *Nature* 1999; **401**: 130–1.
13. Toroczkai Z and Bassler KE. Network dynamics: jamming is limited in scale-free systems. *Nature* 2004; **428**: 716.
14. Pastor-Satorras R and Vespignani A. Epidemic spreading in scale-free networks. *Phys Rev Lett* 2001; **86**: 3200–3.
15. Watts DJ and Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* 1998; **393**: 440–2.
16. Gallos LK, Cohen R and Argyrakis P *et al.* Stability and topology of scale-free networks under attack and defense strategies. *Phys Rev Lett* 2005; **94**: 188701.

17. Huang X, Gao J and Buldyrev SV *et al.* Robustness of interdependent networks under targeted attack. *Phys Rev E* 2011; **83**: 065101.

18. Neumayer S, Zussman G and Cohen R *et al.* Assessing the vulnerability of the fiber infrastructure to disasters. *IEEE/ACM TON* 2011; **19**: 1610–23.

19. Helbing D. Globally networked risks and how to respond. *Nature* 2013; **497**: 51–9.

20. Eubank S, Guclu H and Kumar VSA *et al.* Modelling disease outbreaks in realistic urban social networks. *Nature* 2004; **429**: 180–4.

21. Stauffer D and Aharony A. *Introduction to Percolation Theory*, revised 2nd edn. Boca Raton, FL: CRC Press, 2014.

22. Stanley HE. *Phase Transitions and Critical Phenomena*. Oxford: Clarendon Press, 1971.

23. Buldyrev SV, Parshani R and Paul G *et al.* Catastrophic cascade of failures in interdependent networks. *Nature* 2010; **464**: 1025–8.

24. Newman MEJ, Strogatz SH and Watts DJ. Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 2001; **64**: 026118.

25. Gao J, Buldyrev SV and Stanley HE *et al.* Networks formed from interdependent networks. *Nat Phys* 2012; **8**: 40–8.

26. Dong G, Gao J and Du R *et al.* Robustness of network of networks under targeted attack. *Phys Rev E* 2013; **87**: 052804.

27. Shekhtman LM, Danziger MM and Havlin S. Recent advances on failure and recovery in networks of networks. *Chaos Solitons Fractals* 2016; **90**: 28–36.

28. Bunde A and Havlin S. *Fractals and Disordered Systems*. Berlin: Springer, 2012.

29. Dong G, Fan J and Shekhtman LM *et al.* Resilience of networks with community structure behaves as if under an external field. *Proc Natl Acad Sci USA* 2018; **115**: 6911–5.

30. Coniglio A, Nappi CR and Peruggi F *et al.* Percolation points and critical point in the ising model. *J Phys A Math Gen* 1977; **10**: 205–18.

31. Morone F and Makse HA. Influence maximization in complex networks through optimal percolation. *Nature* 2015; **524**: 65–8.

32. Li M, Lü L and Deng Y *et al.* History-dependent percolation on multiplex networks. *Natl Sci Rev* 2020; **7**: 1296–305.

33. Albert R, Jeong H and Barabási AL. Error and attack tolerance of complex networks. *Nature* 2000; **406**: 378–82.

34. Cohen R, Erez K and Ben-Avraham D *et al.* Resilience of the internet to random breakdowns. *Phys Rev Lett* 2000; **85**: 4626–8.

35. Callaway DS, Newman MEJ and Strogatz SH *et al.* Network robustness and fragility: percolation on random graphs. *Phys Rev Lett* 2000; **85**: 5468–71.

36. Cohen R, Erez K and Ben-Avraham D *et al.* Breakdown of the internet under intentional attack. *Phys Rev Lett* 2001; **86**: 3682–5.

37. Yan G, Tsekenis G and Barzel B *et al.* Spectrum of controlling and observing complex networks. *Nat Phys* 2015; **11**: 779–86.

38. Yang Y, Wang J and Motter AE. Network observability transitions. *Phys Rev Lett* 2012; **109**: 258701.

39. Liu YY, Slotine JJ and Barabási AL. Observability of complex systems. *Proc Natl Acad Sci USA* 2013; **110**: 2460–5.

40. Wu J, Deng HZ and Tan YJ *et al.* Vulnerability of complex networks under intentional attack with incomplete information. *J Phys A Math Theor* 2007; **40**: 2665–71.

41. Wu J, Tan YJ and Deng HZ *et al.* A robustness model of complex networks with tunable attack information parameter. *Chin Phys Lett* 2007; **24**: 2138–41.

42. Leskovec J, Kleinberg J and Faloutsos C. Graph evolution: densification and shrinking diameters. *ACM Trans Knowl Discov Data* 2007; **1**: 2.

43. Rossi RA and Ahmed NK. The network data repository with interactive graph analytics and visualization. In: *Twenty-Ninth AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press, 2015, 4292–3.

44. Choudhury MD, Sundaram H and John A *et al.* Social synchrony: predicting mimicry of user actions in online social media. In: *2009 International Conference on Computational Science and Engineering*. New York: IEEE, 2009, 151–8.

45. Zafarani R and Liu H. Social computing data repository at ASU, 2009.

46. Opsahl T and Panzarasa P. Clustering in weighted networks. *Soc Networks* 2009; **31**: 155–63.

47. Bader DA, Meyerhenke H and Sanders P *et al. Graph Partitioning and Graph Clustering, Contemporary Mathematics*, Vol. 588. Providence, RI: American Mathematical Society, 2013.

48. Hoel PG. *Introduction to Mathematical Statistics*, 2nd edn. New York: Wiley, 1954.