Immunization and targeted destruction of networks using explosive percolation

Pau Clusella,[1] Peter Grassberger,[2, 1] Francisco J. Pérez-Reche,[1] and Antonio Politi[1]

[1]*Institute for Complex Systems and Mathematical Biology, SUPA, University of Aberdeen, Aberdeen, UK*
[2]*JSC, FZ Jülich, D-52425 Jülich, Germany*

(Dated: April 4, 2016)

The explosive percolation (EP) paradigm is used to propose a new method ('explosive immunization') for immunization and targeted destruction of networks. The ability of each node to block the spread of an infection (or to prevent the existence of a large cluster) is estimated by a score. The algorithm proceeds by first identifying *low score* nodes that are *not* vaccinated / destroyed, similarly as links that do *not* lead to large clusters are first established in EP. As in EP, this is done by selecting the worst node from a finite set of randomly choosen 'candidates'. Tests on several real-world and model networks suggest that the method is more efficient and faster than any existing immunization strategy. Due to the latter property it can deal with very large networks.

Network robustness is a major theme in complex-systems theory that has attracted much attention in recent years [1]. Two specific problems are immunization of networks against epidemic spreading (of infections diseases, computer viruses, or malicious rumors), and the destruction of networks by targeted attacks. At first sight these two look completely different, but they can actually be mapped onto each other. The key observation is that infections spreading in a population use the network of contacts between hosts for their spread. Accordingly, from the viewpoint of the infection, immunization corresponds to an attack that destroys the network on which it can spread. Vaccination of hosts (network nodes) is often the most effective way to prevent large epidemics. Other strategies include manipulating the network topology [2–4] or introducing heterogeneity in transmission of the infection [5–7].

The main task both in network immunization and in targeted attacks is to find those nodes ("blockers") whose removal is most efficient in destroying connectivity. Important blockers ("superblockers") are often assumed [8] to be equivalent to "superspreaders", i.e. the most efficient nodes in spreading information, supplies, marketing strategies, or technological innovations. Identifying superspreaders is the subject of a vast literature [1] but, as pointed out in, e.g., Ref. [9], identifying superblockers is not the same as finding superspreaders. Indeed, a node in a densely connected core will in general be a good spreader [10], but it will be in general a very poor blocker, since the infection can easily find ways to go around it.

Here we devise a strategy which identifies superblockers. Vaccinating such nodes provides an efficient way to fragment the network and reduce the possibility of large epidemic outbreaks. We focus on "static" immunization which aims at fragmenting the network before a possible outbreak occurs ("dynamical" immunization strategies where one tries to contain an ongoing epidemic were studied, for instance, in [11]). In our approach, the network consists of N nodes out of which qN are vaccinated; the rest are left susceptible to the infection. The size of an invasion will depend on the fraction of immunized nodes,

q, the type of epidemic (e.g. SIR or SIS [12]) and its virulence. However, the maximum fraction of nodes infected at any time will always be bounded by the relative size $S(q)$ of the largest *cluster of susceptible* nodes, $\mathcal{G}(q)$. Keeping $S(q)$ as small as possible will therefore ensure that epidemic outbreaks of any type are as small as they can be for a vaccination level q [8, 13]. For large networks, $N \to \infty$, the aim of immunization is to fragment them so that $S(q) = 0$ [8]. The immunization threshold q_c is defined as the smallest q-value at which $S(q) = 0$. Although q_c is not well defined for finite N, it can nevertheless be estimated reliably. Our algorithm deteriorates only when the network is too small (in this case, however, an extensive search of the optimal solution can be performed). In general, the smaller q_c, the more effective is the corresponding strategy, since the epidemic can be prevented by vaccinating a smaller set of nodes.

The identification of superspreaders is in general an NP-complete problem [14], and most likely this is also true for finding superblockers. Therefore, heuristic approaches have to be used. Typically one defines for each node a score based on local [15, 16] or global properties [8, 13, 17]. Here we use two different scoring schemes for $q > q_c$ and $q < q_c$, which both combine local and semi-local information. We refer to our approach as "explosive immunization" (EI), since it is based on the concept of *explosive percolation* (EP) proposed by Achlioptas *et al.* [18]. Although EP has been discussed in a large number of papers because of its very unusual threshold behavior, so far no application of EP had been proposed. To our knowledge, EI is the first problem where it is practically used.

In the following we test the performance of EI for both real-world and model networks. It gave in all cases by far the lowest values of q_c, compared to all other strategies we are aware of. It also gives the best values of $S(q)$ in most cases, although there might be values of q where other strategies perform better. In fact, following the mainstream in network studies, we focused on $S(q)$, which corresponds to outbreaks starting in $\mathcal{G}(q)$. However, outbreaks can also start in any other cluster:

an improved success measure $\bar{S}(q)$ can be defined which takes this possibility into account. If this is used, our algorithm turns out to be yet more efficient, and is optimal even when $S(q)$ might suggest that it is not (see Appendix A). In addition to being very efficient, our algorithm is also extremely fast, if implemented using the Newman-Ziff percolation algorithm [19].

Explosive immunization: The basic concept of EP [18] is as follows. We consider site percolation transition on a network on which all links are already present, but nodes are still to be inserted one by one (note that Achlioptas *et al.* [18] studied bond percolation, but the basic concept is the same for site percolation; the original, bond-based, version can be used directly for immunizing bonds instead of sites).

While nodes are chosen randomly in ordinary percolation, this is done in a two-step way in EP:

(a) One first selects m "candidate" nodes ($m = 2$ was used in [18]; we shall use typically $m \sim 10^3$), and determines for each candidate a score that should indicate its success in creating a big cluster. Various scores were used in [18] and in papers following it. The simplest one is the size of the cluster formed by adding the candidate, but we had best results with the more sophisticated scoring scheme discussed below.

(b) In a second step, one selects from the m candidates the one with the *worst* score, i.e. the one which would delay the formation of a large cluster as long as possible. Once all low-score nodes have been chosen, one has to use high-score candidates, and the cluster size may increase explosively.

When interpreted as an immunization strategy, this means that one first identifies the nodes that are poor blockers for infection – either because they have small degree or because they do not join already large clusters – and can remain unvaccinated. As long as these are few, and the rest are immunized, no big epidemic can occur. Only when nodes are identified as poor blockers but are not, one starts to form larger clusters of susceptible nodes that can be invaded by an epidemic. Thus we first assume that all nodes are vaccinated (i.e. we start from $q = 1$) and then "undo" the immunization of nodes identified by the EP algorithm.

A similar strategy was proposed in [13], but here we introduce several important modifications. In particular, we show that using different scores depending on the value of q leads to better results than following more traditional strategies based on a single score for all q. Also, *all* nodes were used as candidates in [13] and a slow algrithm for finding the worst was used, both of which make the algorithm unsuitable for large networks.

The first score, typically used for values of q where no giant cluster exists, is defined as

$$\sigma_i^{(1)} = \sum_{\mathcal{C} \subset \mathcal{N}_i} (\sqrt{|\mathcal{C}|} - 1) + k_i^{(\text{eff})} , \qquad (1)$$
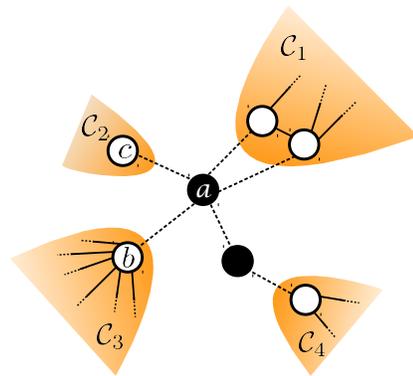

FIG. 1. Instance of possible configurations encountered during the "de-immunization" process. White and black circles correspond to susceptible and vaccinated nodes, respectively. Node a is a potential candidate for being de-immunized. Assuming that b is a hub and c is a leaf, then $M_a = 1$ and $L_a = 1$. From Eq. (2) it follows that $k_a^{(\text{eff})} = 3$.

where $\mathcal{N}_i$ is the set of all clusters linked to i , $|\mathcal{C}|$ is the size of cluster $\mathcal{C}$, and $k_i^{(\text{eff})}$ is an effective degree of i defined below. The square root in the first term on the r.h.s. is used in order not to give too large weight to large clusters, and subtracting 1 from it avoids counting leaves.

In the effective degree we want to discount both leaves (since they don't contribute much to spreading) and hubs (because they will presumably be vaccinated anyhow). A hub is defined as a node with $k_i^{(\text{eff})} \geq K$ with some arbitrarily chosen cutoff K. We will see later that best results typically are obtained with $K \approx 6$. Formally, $k_i^{(\text{eff})}$ is given by the self-consistent equation

$$k_i^{(\text{eff})} = k_i - L_i - M_i(\{k_j^{(\text{eff})}\}) , \qquad (2)$$

where k_i is the bare degree of node i, L_i is the number of leaves and M_i is the number of hubs attached to i. It is calculated at the start of the simulation iteratively with forward substitution, starting with $k_i^{(\text{eff})} = k_i$ for all i. An example of how $k_i^{(\text{eff})}$ is determined is given in Fig. 1, under the assumption that node b is a hub.

Notice that, according to Eq. (2), nodes surrounded by hubs may play a minor blocking role and can be left unvaccinated, as compared to nodes without hub neighbors. This is also verified numerically for Erdös-Rényi (ER) networks (Appendix B). This idea is similar to the score used in [15], but it is opposite of what is assumed e.g. in page rank [1] and in the "collective influence" defined in [8].

As we will see, using the score $\sigma_i^{(1)}$ yields excellent estimates of the critical point q_c, even though in finite networks, the onset of a giant cluster is not precisely defined, and we approximate it by a value q^* where $S(q^*) \approx 1/\sqrt{N}$.

Below q_c, the score defined in Eq. (1) leads to big jumps that arise from the joining of large clusters. For

very small q, finally, after the last two large clusters have joined, many single-node clusters survive that had made up the boundaries between them. The vaccination of these is very inefficient. In an optimal strategy we would like to immunize even for very small q clusters of reasonably large size, i.e. we would like to prevent intermediate size clusters from joining too early.

For $q < q^* \approx q_c$ we thus switch to a different score $\sigma_i^{(2)}$ defined as

$$\sigma_i^{(2)} = \begin{cases} \infty & \text{if} \quad \mathcal{G}(q) \not\subset \mathcal{N}_i, \\ |\mathcal{N}_i| & \text{else, if } \arg\min_i |\mathcal{N}_i| \text{ is unique,} \\ |\mathcal{N}_i| + \epsilon|\mathcal{C}_2| & \text{else.} \end{cases} \tag{3}$$

Here $|\mathcal{N}_i|$ is the number of clusters in the neighborhood of i, $\mathcal{C}_2$ is the second-largest cluster in $\mathcal{N}_i$, and ϵ is a small positive number. Thus we select only candidates which have the giant cluster in their neighborhood; among these we pick the candidate with the smallest number of neighboring clusters, and if this is not unique, we pick the candidate for which the second-largest neighboring cluster is the smallest. This preserves as much as possible intermediate-size clusters.

Two remarks are in order regarding the efficiency of our algorithm: (i) In [13] all nodes were considered as possible candidates in step (a) of EI. This makes the algorithm very slow and prevents its use for large networks. In our tests already $m = 10$ candidates gave very good results, and using $m = 1000$ candidates led to no noticeable degradation. (ii) When joining clusters, we used the very fast Newman-Ziff percolation algorithm [19] which has time complexity $O(N)$ for networks with bounded degrees. It also gives, at each moment of the evolution, the size of the largest cluster, whose determination would otherwise need most of the CPU time. As a result, we could analyze networks with 10^8 nodes within hours on normal work stations.

Numerical results: As a first test we studied ER networks with average degree $\langle k \rangle = 3.5$ (to compare with results from [8]). Overall, the best results are obtained by using the scores given in Eqs. (1) and (3) (Fig. 2, solid line in main plot). The dashed line is obtained by using $\sigma_i^{(1)}$ for all q (the big jumps, which were also seen in [13], correspond to joinings of big clusters). It is in general worse than the continuous curve, except close to the jumps (see, however, the Appendix). Finally we show in Fig. 2 also the results obtained with the recently proposed collective influence algorithm [8], which was hailed in as "perfect" [20]. They are significantly worse. Our estimate $q_c \leq 0.1838(1)$ is also smaller than the best estimate $0.192(9)$ obtained in [8] using extremal optimization [21], and used there as "gold standard" for small networks (it is too slow to be used for large networks).

As regards ER networks with other values of $\langle k \rangle$, we first looked at $\langle k \rangle = 4$, since this had been used in [13]. Our results are similar to those of [13], but significantly
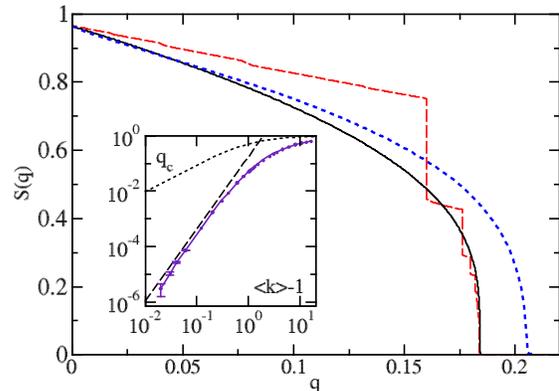

FIG. 2. Relative size $S(q)$ of the largest clusters against q, for ER networks with $N = 10^6$ and $\langle k \rangle = 3.5$. The dashed curve with jumps is obtained, if EI is used with score $\sigma^{(1)}$ for all q, 2000 candidates, and $K = 6$. The continuous curve is obtained with $\sigma^{(2)}$ for $q < q^*$, where $S(q^*) = 1/500$. The dotted line shows the result from [8]. The inset shows a log-log plot of q_c against $\langle k \rangle - 1$. The straight line indicates the power law $q_c \sim (\langle k \rangle - 1)^{2.6}$, while the dotted curve shows the result for random immunization.

better. Next we estimated q_c for a wide range of $\langle k \rangle$. By using networks with N up to 2^{24} we were able to obtain precise results even for $\langle k \rangle$ very close to the threshold $\langle k \rangle = 1$ for the existence of a giant cluster. The results, shown in the inset of Fig. 2, suggest that q_c satisfies for small $\langle k \rangle$ a power law

$$q_c \sim (\langle k \rangle - 1)^{2.6}, \tag{4}$$

where the error of the exponent is $\approx \pm 0.2$. This should be compared to random immunization [22], $q_c^{\text{rand}} = (\langle k \rangle - 1)/\langle k \rangle$ (dotted curve in the inset of Fig. 2). The difference in the exponents reflects the fact that a nearly critical cluster can be destroyed by removing a few "hot" nodes, whence targeted attacks become more efficient as $\langle k \rangle$ approaches the threshold.

Surprisingly, for all $\langle k \rangle$ except very close to 1, best results are obtained with $K = 6$. This suggests that most nodes with $k_i^{(\text{eff})} > 6$ are vaccinated at q_c, independently of the average degree. This was also verified directly: Although there is no strict relationship between effective degree and blocking power (some hubs were not vaccinated at q_c, while some nodes that were vaccinated are not strong hubs), there is a very strong correlation, stronger that between actual degree and blocking power (see Appendix B). On the other hand, very few weak nodes have to be vaccinated (about 1 per mille of the nodes with $k_i^{(\text{eff})} = 3$), in contrast to claims in [8] that weak nodes are often important blockers.

Scale-free networks. EI also gives excellent results for scale-free (SF) networks with node degree distribution
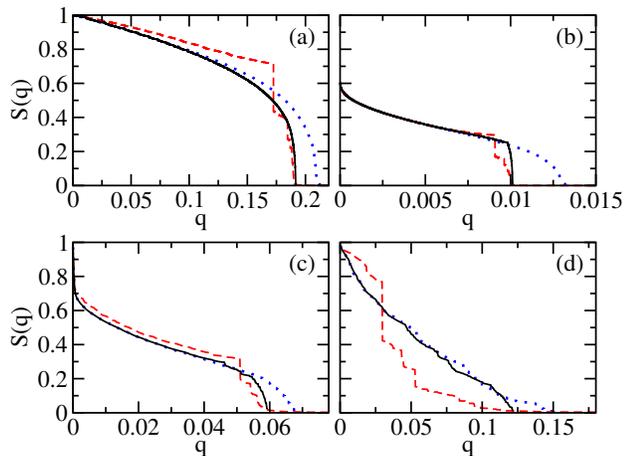
FIG. 3. Relative size $S(q)$ of the largest clusters in SF networks of size $N = 10^6$ obtained with (a) the Albert-Barábasi model ($\gamma = 3$) and (b) the configuration model with $\gamma = 2.5$. Panels (c) and (d) show results for the cattle and airport transportation networks, respectively. Different line types correspond to different algorithms: EI using scores $\sigma_i^{(1)}$ and $\sigma_i^{(2)}$ (continuous line) or only score $\sigma_i^{(1)}$ (dashed line) and the algorithm in [8] (dotted line).

$p_k \sim k^{-\gamma}$, built with both the Barabási-Albert method (fixed $\gamma = 3$) and the configuration model (γ can be tuned) [1, 23]. Our results are significantly better than those obtained with the method from [8] for both settings (Fig. 3(a) and (b)). Using a single score across the entire q-range gives again the best estimate for q_c, while the two-score strategy proves generally superior for $q < q_c$. The jumps obtained in the single-score strategy are less strong for the configuration model (and thus the two-score strategy seems there less preferable), but the superiority of the two-score strategy becomes again clear when using the improved $\bar{S}(q)$ discussed in Appendix A.

Observe that the shape of $S(q)$ near $q = 0$ is concave/convex for large/small γ (compare panels (a) and (b) in Fig. 3). The convex shape for small γ is due to the presence of many hubs which lead to a drastic decrease of $S(q)$ when immunized at small q.

Real world networks: We have also studied the performance of EI on a number of real-world networks, starting from an example in which immunization plays an important role for food security [24, 25]: a network of Scottish cattle movements. The network consists of $N = 7228$ premises (nodes) connected by $E = 24784$ transportation events (edges) occurring between 2005 and 2007. The node distribution obeys a power-law with exponent $\gamma = 2.37 \pm 0.06$ (Maximum likelihood fit). The scenario is similar to that of SF networks with small γ (compare panels (c) and (b) in Fig. 3). Again, $S(q)$ decreases quite quickly because of the presence of many well connected nodes (e.g. markets and slaughterhouses), whose immunization leads to a drastic decrease of the largest cluster.

Once again we see that our strategy is superior to the previous approaches.

We have also studied several networks that were used as benchmark in previous works. This includes the high-energy physicist collaboration network [26] and the internet at autonomous system level [27]. In both cases our results are similar to, but slightly better than, in [13] (which were the best previous estimates). The results for these and soil networks [28] are shown in Appendix C.

A particularly problematic case is the airline network [29], also studied in [2]. This is a rather small network ($N = 3151$ and $E = 27158$) with a very broad degree distribution (power-law with $\gamma = 1.70 \pm 0.04$), for which q_c is not well defined. The results are shown in Fig. 3(b), where we see that using $\sigma_i^{(1)}$ alone provides the optimal response almost everywhere. In this case, the outcome of the score $\sigma_i^{(2)}$ strongly depends on where the critical point is selected. It is anyway clear that a suitable combination of them provides the optimal results.

Conclusions: In this paper, we extend the explosive percolation concept to propose a two-score strategy for attacking networks that proves superior to all previously proposed protocols. The comparison between the two scores suggests that an everywhere optimal strategy using a single score is unlikely to exist. This is to be traced back to the NP completeness of the problem. Since immunization of a network by immunizing (vaccinating) nodes can be regarded as a strategy for destroying the network on which an infection can propagate, this also gives a nearly optimal strategy for immunization. Our explosive immunization method seems superior, both as regards speed and minimal cost (as measured by the number of immunized nodes) to all previous strategies.

We have focused on immunization of nodes but the idea behind EI can also be applied to immunization of links. This would provide nearly optimal quarantine strategies which might significantly improve the typical brute-force implementation which cut all the links between two parts of a network. Targetted removal of links with high betweenness centrality is the basis for one of the most efficient algorithm for finding network communities [30]. We propose that explosive immunization of links should also provide a very efficient algorithms for community detection.

Appendix A: Improved success measure for network immunization

The standard measure for the success of an immunization strategy is the relative size of the largest connected cluster, $\mathcal{G}(q)$, after a fraction q of nodes have been vacci-

nated,

$$S(q) = |\mathcal{G}(q)|/N , \quad\quad (A1)$$

where $|\mathcal{G}(q)|$ is the size of the largest cluster. The motivation for this is that a strongly infective disease that hits a random node will in average infect a region of size $NS(q)^2$ in the largest cluster, where the first factor of $S(q)$ is for the probability that the largest cluster is hit at all, and the other factors give the number of infected sites, if it does so. This quantity neglects the effect of smaller clusters, following a widespread habit in network science. Often this is justified because their contribution is small and/or hard to estimate. But in the present case, the contribution of clusters other than the largest one can be substantial, and it can be taken into account easily. In order to incorporate the effect of epidemics starting in all clusters in our analysis, let us assume the clusters to be ordered by size, $|\mathcal{G}(q)| \equiv |\mathcal{C}_1(q)| \geq |\mathcal{C}_2(q)| \geq \dots$. The probability that a random outbreak starts in a cluster $\mathcal{C}_i$ is $S_i(q) = N^{-1}|\mathcal{C}_i|$ and its maximum size is $NS_i(q)$. Accordingly, the average number of infected sites in an random outbreak is

$$\langle n_{\text{infected}} \rangle = NS(q)^2 + N\sum_{i\geq 2} S_i(q)^2 = N\bar{S}(q)^2, \quad (A2)$$

where, perturbatively,

$$\bar{S}(q) = S(q)\left[1 + \frac{1}{2}\sum_{i\geq 2}\left(\frac{S_i(q)}{S(q)}\right)^2 + \dots\right] \quad (A3)$$

For EI with both scores, there will (at least for large networks with a well defined q_c) never be more than one large cluster, since there is no large cluster for $q > q^*$, and for $q < q^*$ the growth of a second large cluster is suppressed. In this case $\bar{S}(q)$ is practically the same as $S(q)$, as we indeed checked for ER networks. This is not true, however, for $q < q^*$ if score $\sigma^{(1)}$ is used also there. In that case there are in general more than one large cluster, and $\bar{S}(q)$ is considerably larger than $S(q)$, see Fig. 4.

Thus even when it seems better to use score $\sigma^{(1)}$ for all q according to the success measure $S(q)$, using a more refined success measure might show that the strategy of using both scores $\sigma^{(1)}$ and $\sigma^{(2)}$ is superior.

Appendix B: Degree distributions of vaccinated nodes at $q = q_c$

Naively, one expects that it is the strongest hubs that should be vaccinated first, but the fact that network immunization is non-trivial shows that this is not exactly true. If the motivation that led us to define the effective degree $k^{(\text{eff})}$ is correct, we should expect the vaccinated nodes to be more strongly concentrated in the high-$k^{(\text{eff})}$
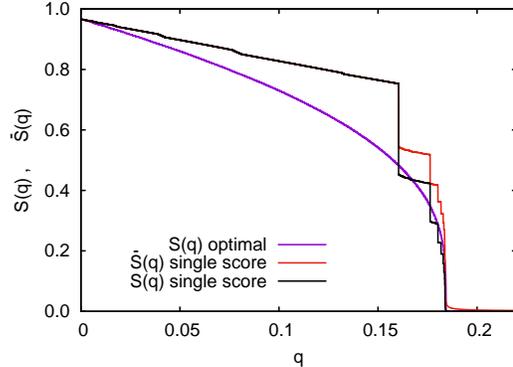

FIG. 4. Success measures $S(q)$ (lower (black) steplike curve) and $\bar{S}(q)$ (upper (red) steplike curve) for ER networks with $\langle k \rangle = 3.5$, if score $\sigma^{(1)}$ is used for all q. The smooth curve corresponds to the case where $\sigma^{(2)}$ is used for $q < q_c$. In that case, the curves for $S(q)$ and $\bar{S}(q)$ are indistinguishable except in a small region $q > q_c$.

region, than they are concentrated in the high-k region. Here we show data that indeed confirm this, although the difference is rather small. More precisely, we show in the left panel of Fig. 5 two histograms: The $k^{(\text{eff})}$-distribution of *all* nodes in an ER network with $\langle k \rangle = 3.5$ and the ditribution of those nodes that are vaccinated at $q = q_c$. In the right panel the corresponding two k-distributions are shown.

We see that in both cases nearly all nodes with degree > 7 are vaccinated, while nearly all nodes with degree < 4 are left unvaccinated. This agrees with our findings that $K = 6$ is optimal in this case. But a closer look shows that the $k^{(\text{eff})}$-distribution of vaccinated nodes has indeed a slightly sharper cut off than the k-distribution. For instance, while ≈ 25 % of nodes with $k = 6$ are not vaccinated, this is true for only ≈ 10 % of nodes with $k^{(\text{eff})} = 6$.

Appendix C: Real-world networks

We present the detailed results for three other real-world networks. In each of them we compare the two different scores of Explosive Immunization (EI) with the Collective Influence (CI) method proposed by Morone *et. al.* [8]. We also show an example of how the hub cut-off parameter K in the computation of the effective degree $k_i^{(\text{eff})}$ modifies the results. In all three plots we use only the success measure $S(q)$ (for more easy comparison with previous literature), but we should keep in mind that methods producing large steps would become worse when using $\bar{S}(q)$.
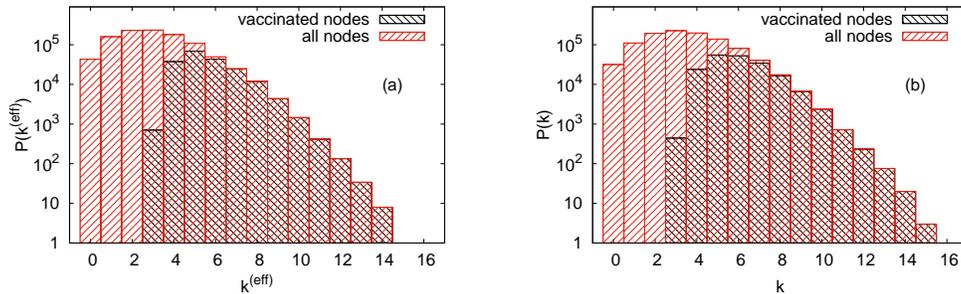
Soil network: We study a network of the structure

FIG. 5. (color online) (a) Log-linear plot of $P(k^{(\mathrm{eff})})$ for ER networks with $\langle k \rangle = 1.75$. The left histogram is for all nodes, the right one is for those nodes that are not declared as "harmless" at $q = q_c$ and which therefore must be vaccinated in order to immunize the network. Panel (b) shows the analogous distributions for the actual degrees. Notice that for the vaccinated nodes, the distribution of $k^{(\mathrm{eff})}$ has a slightly sharper cut-off than that of k, indicating that $k^{(\mathrm{eff})}$ is a better indicator for nodes that must be vaccinated than k. The same was found also for all other values of $\langle k \rangle$.

of soil pores presented in [28]. This network has a large clustering coefficient and a limited degree distribution with $\langle k \rangle = 2.8$. In figure 6(a) we plot the results of EI and CI using $K = 8$ and $S(q^*) = 0.06$. In this case, EI produces better results than CI everywhere. In particular, using the score $\sigma_i^{(1)}$ for all q is optimal except when q is very small (see, however, the above caveat about using $\bar{S}(q)$ instead of $S(q)$). In the inset panel we show different values of K effect the outcome of $\sigma_i^{(1)}$ in this network.

Internet: In figure 6b we show the results for a network representing the Internet at the level of autonomous system [27]. We set the parameters $K = 6$ and $S(q^*) = 0.02$. In this case, different values of K do not change significantly the outcome. In general we observe a behavior similar to the cattle network in which an early vaccination of nodes produces a strong decrease of $S(q)$. Again the EI method gives better results than CI.

High-Energy Physicists: Finally, in figure 6c we use the high-energy physicist collaboration network [26] also used in [13]. We plot the results using $K = 6$ and $S(q^*) = 0.01$. The proposed EI algorithm achieves a better value of q_c than the one obtained by Schneider *et. al.*. Both are better than the one obtained with CI. When the giant component is grown for small q (significantly $< q_c$), the CI method is similar but slightly better than EI. This is the only case that we have found where EI is not optimal everywhere.

[1] M. E. J. Newman, *Networks: an introduction* (Oxford University Press, Oxford, 2010).

[2] C. M. Schneider, T. Mihaljev, S. Havlin, and H. J. Herrmann, Phys. Rev. E **84**, 061911 (2011).

[3] C. M. Schneider, N. Yazdani, N. A. M. Araújo, S. Havlin, and H. J. Herrmann, Sci. Rep. **3**, 1969 (2013).

[4] A. Zeng and W. Liu, Phys. Rev. E **85**, 066130 (2012).

[5] F. J. Pérez-Reche, S. N. Taraskin, L. d. F. Costa, F. M. Neri, and C. A. Gilligan, J. R. Soc. Interface **7**, 1083 (2010).

[6] F. M. Neri, F. J. Pérez-Reche, S. N. Taraskin, and C. a. Gilligan, J. R. Soc. Interface **8**, 201 (2011).

[7] F. M. Neri, A. Bates, W. S. Fuchtbauer, F. J. Pérez-Reche, S. N. Taraskin, W. Otten, D. J. Bailey, and C. A. Gilligan, PLoS Comput. Biol. **7**, e1002174 (2011).

[8] F. Morone and H. A. Makse, Nature **524**, 65 (2015).

[9] Habiba, Y. Yu, T. Y. Berger-Wolf, and J. Saia, in *Adv. Soc. Netw. Min. Anal.* (Springer, Berlin Heidelberg, 2010) pp. 55–76.

[10] S. Pei and H. A. Makse, J. Stat. Mech. Theory Exp. **2013**, P12002 (2013).

[11] Q. Wu, X. Fu, Z. Jin, and M. Small, Physica A **419**, 566 (2015).

[12] L. Hébert-Dufresne, A. Allard, J.-G. Young, and L. J. Dubé, Sci. Rep. **3**, 2171 (2013).

[13] C. M. Schneider, T. Mihaljev, and H. J. Herrmann, Europhys. Lett. **98**, 46002 (2012).

[14] D. Kempe, J. Kleinberg, and É. Tardos, in *Proc. ninth ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '03* (ACM Press, New York, New York, USA, 2003) p. 137.

[15] Y. Liu, Y. Deng, and B. Wei, Chaos **26**, 013106 (2016).

[16] P. Holme, Europhys. Lett. **68**, 908 (2004).

[17] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, Phys. Rev. E **65**, 56109 (2002).

[18] D. Achlioptas, R. M. D'Souza, and J. Spencer, Science **323**, 1453 (2009).

[19] M. E. J. Newman and R. M. Ziff, Phys. Rev. E **64**, 016706 (2001).

[20] I. A. Kovács and A.-L. Barabási, Nature **524**, 38 (2015).

[21] S. Boettcher and A. G. Percus, Phys. Rev. Lett. **86**, 5211 (2001).

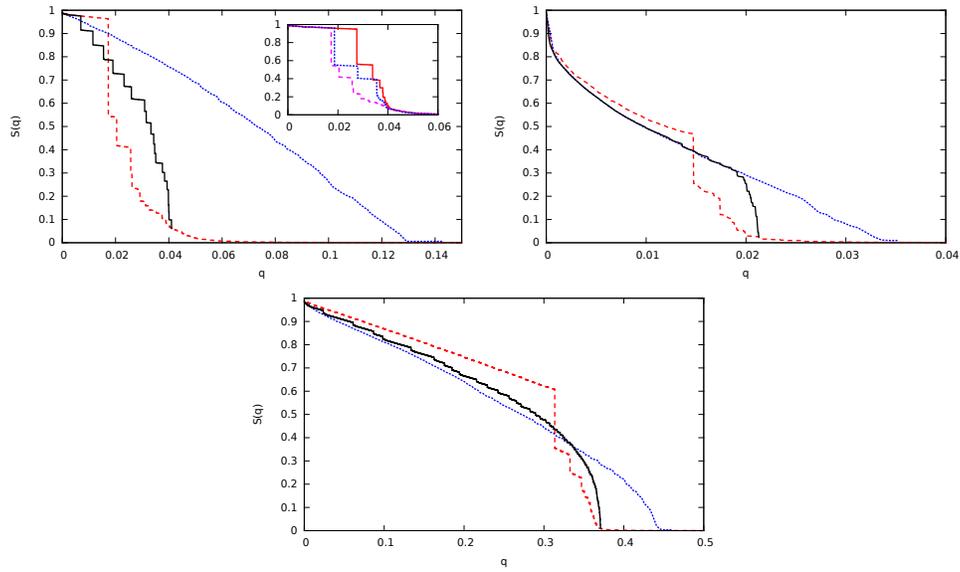[22] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin,

FIG. 6. Results of the different real-world networks. The red dashed, black solid and blue dotted lines corresponds to the $\sigma_i^{(1)}$ score, the $\sigma_i^{(1)}$ & $\sigma_i^{(2)}$ scores, and to the CI method, respectively. Panels a, b, and c correspond to the soil, internet, and high-energy physicist networks discussed in the text. In the small inset of panel (a) we show different results of $\sigma_i^{(1)}$ corresponding to $K = 6$ (red solid), 8 (magenta dashed) and 10 (blue dotted).

Phys. Rev. Lett. **86**, 3682 (2001).

[23] A.-L. Barabási and R. Albert, Science (80-.). **286**, 11 (1999).

[24] M. J. Keeling, M. E. J. Woolhouse, R. M. May, G. Davies, and B. T. Grenfell, Nature **421**, 136 (2003).

[25] R. R. Kao, D. M. Green, J. Johnson, and I. Z. Kiss, J. R. Soc. Interface **4**, 907 (2007).

[26] Http://vlado.fmf.uni-lj.sl/pub/networks/data/hep-th/hep-th.htm.

[27] Http://www.netdimes.org.

[28] F. J. Pérez-Reche, S. N. Taraskin, W. Otten, M. P. Viana, L. D. F. Costa, and C. a. Gilligan, Phys. Rev. Lett. **109**, 098102 (2012).

[29] Http://openflights.org/data.html.

[30] M. Girvan and M. E. Newman, Proc. Natl. Acad. Sci. **99**, 7821 (2002).