Efficient local strategies for vaccination and network attack

P. Holme

Department of Physics, Umeå University - 901 87 Umeå, Sweden, and
Department of Physics, University of Michigan - Ann Arbor, MI 48109, USA

received 8 September 2004; accepted in final form 22 October 2004
published online 12 November 2004

PACS. 89.65.-s – Social and economic systems.
PACS. 89.75.Hc – Networks and genealogical trees.
PACS. 89.75.-k – Complex systems.

Abstract. – We study how a fraction of a population should be vaccinated to most efficiently stop epidemics. Our starting point is that only local information —about the neighborhood of specific vertices— is usable in practice. For this reason we consider only local vaccination strategies. The efficiency of the vaccination strategies is investigated with both static and dynamic measures. We test the strategies on real-world networks and model networks with real-world characteristics. Among other things, we find that the most efficient strategy for many real-world situations is to iteratively vaccinate the neighbor of the previous vaccinee that has the most links out of the neighborhood.

Introduction. – Diseases spread over networks. The spreading dynamics is closely related to the structure of networks. For this reason, network epidemiology has turned into one of the most vibrant subdisciplines of complex network studies [1–3]. A topic of great practical importance within network epidemiology is the vaccination problem —how shall a population be vaccinated to most efficiently prevent a disease to turn into an epidemic? For economic reasons it is often not possible to vaccinate the whole population. Some vaccines have severe side effects which is another reason for limiting the number of vaccinated individuals. So if cheap vaccines, free of side effects, do not exist, then having an efficient vaccination strategy is essential for saving both money and lives. If all ties within the population are known, then the target persons for vaccination can be identified using sophisticated global algorithms (cf. [4]); but this is hardly possible for large-scale vaccination campaigns. In ref. [5], the authors suggested a vaccination strategy that only requires a person to assess whom, he or she, gets close enough to for the disease to spread to —i.e., to name the “neighbors” in the network over which the disease spreads. For networks with a skewed distribution of degree (number of neighbors) the strategy to vaccinate a neighbor of a randomly chosen person is much more efficient than random vaccination. In this work we assume that each individual knows a little bit more about his, or her, neighborhood than just the names of the neighbors: We also assume that an individual can guess the degree of the neighbors and the ties from one neighbor to another. This assumption is not very far-fetched —people are believed to have a
Table I – Statistics of the networks. Note that the arxiv and seceder model networks are not connected —the largest connected components contains 48561 and 2162(1) nodes, respectively (the one s.d. error in parentheses).

| Network    | N   | M   | C       | r       |
|------------|-----|-----|---------|---------|
| arxiv      | 58342 | 294901 | 0.420 | +0.324 |
| karate club | 34  | 78  | 0.256 | -0.476 |
| seceder    | 2200 | 6600 | 0.266(1) | +0.012(2) |
| WS         | 2000 | 4000 | 0.4219(1) | -0.01267(2) |

good understanding of their social surroundings (this is, for example, part of the explanation for the “navigability” of social networks) [6].

Finding the optimal set of vaccinees is closely related to the attack vulnerability problem [4, 7] —how communication networks are affected by sinister attacks. The major difference is the dynamic system confined to the network —disease spreading for the vaccination problem and information flow for the attack vulnerability problem.

Preliminaries. – We will use two measures of network structure in our discussion: One is the clustering coefficient $C$, defined as the ratio of triangles to connected triples normalized to the interval $[0, 1]$. If $C = 1$, there is a maximal number of triangles (given the set of connected triples); if $C = 0$, the graph has no triangles. We also measure the correlation between the degrees at either side of an edge through the assortative mixing coefficient defined as:

$$ r = \frac{4\langle k_1 k_2 \rangle - \langle k_1 + k_2 \rangle^2}{2\langle k_1^2 + k_2^2 \rangle - \langle k_1 + k_2 \rangle^2}, $$

where $k_i$ is the degree of the $i$-th argument of an edge in a list of the edges, and $\langle \cdot \rangle$ denotes average over that edge-list. We let $N$ denote the number of vertices and $M$ the number of edges.

The networks. – We will test the vaccination strategies we propose on both real-world and model networks. The first real-world network is based on scientific coauthorships [10]. The vertices of this network are scientists who have uploaded manuscripts to the preprint repository arxiv.org. An edge between two authors means that they have coauthored a preprint. We also study a small real-world social network constructed from an observational study of friendships in a karate club [11]. The edges of this small network are, probably, more relevant for disease spreading than the arxiv network, but may suffer from finite-size effects.

The two model networks are: 1) The networked seceder model, modeling social networks with a community structure and exponentially decaying degree distributions [12]. Briefly, it works by sequentially updating the vertices by, for each vertex $v$, rewiring all $v$’s edges to the neighborhood of a peripheral vertex. With a probability $r$ an edge of $v$ can be rewired to a random vertex (so $r$ controls the degree of community structure). We use the parameter values $M = 3N = 6600$, $r = 0.1$ and $10M$ iterations on an Erdős–Rényi network [13]. 2) The Watts-Strogatz (WS) model [14] generates networks with exponentially decaying degree distributions and tunable clustering. The WS model starts from a circular topology with edges between vertices separated by 1 to $k$ steps on the circle. Then one goes through the edges and rewire one side of them to randomly selected vertices with a probability $P$. We use $P = 0.05$ and $M = kN = 2N = 4000$. Sizes of the networks, along with $C$ and $r$ values, are displayed in table I.

The strategies. – Now we turn to the definition of the vaccination strategies. We assume a fraction $f$ of the population is to be vaccinated. As a reference, we consider random vaccination (Rnd, equivalent to site percolation). We use the neighbor vaccination (RNB) of ref. [5]
— to vaccinate the neighbor of randomly chosen vertices. We also consider the trivial improvement [15] if knowledge about the neighbors’ degrees is included: Pick a vertex at random and vaccinate one (randomly chosen) of its highest-degree neighbors (we call it Deg). To avoid overvaccination of a neighborhood, one can consider vaccinating neighbors of a vertex \( v \) with a maximal number of edges out of \( v \)'s neighborhood (Out). For all strategies except RND, we also consider “chained” versions where one, instead of vaccinating a neighbor of a randomly chosen vertex, vaccinates a neighbor of the vertex vaccinated in the previous time step (if all neighbors are vaccinated a neighbor of a random vertex is chosen instead). For the acronyms of the chained versions a suffix “C” is added. As it turns out, the Deg and DegC strategies are for all cases similar to, or worse than, the Out and OutC strategies, respectively. To simplify the figures and discussion, we therefore omit the Deg and DegC strategies.

Results and analysis. – The results of this letter are presented in three sections: 1) We study how the average size of the largest connected subgraph \( S_1 \) depends on the fraction \( f \) of vaccinated vertices (vertices deleted from the network). \( S_1 \) can be interpreted as the largest possible size of an outbreak. It is the main quantity of percolation-based epidemiology studies [16]. 2) We confirm the conclusions from the \( S_1 \) study by measuring the average outbreak size \( s \) in dynamic simulations. More common dynamic measures of a population’s susceptibility to disease are basic reproduction rate (the average number of people of an uninfected population an infected individual will pass the disease to) and the epidemic threshold (the critical transmission probability for an epidemic to occur) [1–3]. These measures are quantitatively more informative than \( S_1 \) and \( s \), but need a size scaling study to be properly established. However, the order of efficiency of the strategies as derived from \( S_1 \) and \( s \) (\( N - S_1 \) and \( N - s \), respectively) would be unlikely to change (this is confirmed in a preliminary study), so \( S_1 \) and \( s \) is, qualitatively, elucidative enough. 3) To interpret the results, we also investigate \( S_1 \) for a fixed \( f \) as a function of the clustering and assortative mixing coefficients.

Our main static measure \( S_1 \) is obtained by averaging the number of vertices in the largest connected subgraph over \( n_{\text{vac}} = 1000 \) runs of the vaccination procedures. The model networks are also averaged over \( n_{\text{net}} = 100 \) network realizations. (Smaller or larger values of \( n_{\text{vac}} \) and \( n_{\text{net}} \) do not make any qualitative difference.) In fig. 1 we display \( S_1 \) as a function of \( f \). For all networks, except the WS model network, the Out (chained and unchained versions) is most efficient. Within this group the order of efficiency varies: For the arxiv network the Out strategy is more than twice as efficient as any other in the interval \( 0.25 \lesssim f \lesssim 0.4 \). For the HK and seceder model networks, the chained strategies are considerably more efficient than the unchained ones. Another interesting observation is that, even if the degree distribution is narrow (such as for the seceder model), the more elaborate strategies were much more efficient than random vaccination. This is especially clear for higher \( f \) which suggests that the structural change of the network of susceptible vertices during the vaccination procedure is an important factor for the overall efficiency. For the WS model network, the chained algorithms performed poorer than random vaccination. This is in contrast to all other networks. We conclude that epidemiology-related results regarding the WS model networks should be cautiously generalized to real-world systems.

Static measures of vaccination efficiency are potential over-simplifications — there is a chance that the interplay between disease dynamics and the underlying network structure has a significant role. To motivate the use of \( S_1 \), we also investigate the SIS and SIR models [1] on vaccinated networks. In the SIS model a vertex goes from “susceptible” (S) to “infected” (I) and back to S. The SIR model is similar, except that an infected vertex goes to the “removed” (R) state and stays there. The probability of going from S to I (per contact) is zero for vaccinated vertices and \( \lambda = 0.05 \) for the rest. The I-state lasts \( \delta = 2.5 \) time steps. We use
synchronous updating and one randomly chosen initially infected individual. The dynamics is stopped when there are no I-vertices present, or (for SIS) an endemic state (defined in the simulations as when there are no susceptible vertices that have not had the disease at least once) is reached. We measure the outbreak size by $s$ — the number of individuals that
at least once have been infected during an outbreak averaged over \( n_{\text{dis}} = 100 \) outbreaks and \( n_{\text{vac}} = 1000 \) realizations of the vaccination schemes. In fig. 2(a) we plot \( s \) for the arxiv network. Other networks and simulation parameters give qualitatively similar results. Qualitatively, the large picture from the \( S_1 \) calculations remains — the chained and unchained OUT strategies are very efficient, and the chained versions are more efficient than the unchained ones. A difference is that the unchained RN\( \bar{b} \) also performs rather well. Quantitatively, the differences between the strategies are huge, this is a result of the threshold behaviors of the SIS and SIR models [5]. The conclusion of fig. 2 (and similar plots for other networks) is that the order of the strategies' efficiencies is largely the same as concluded from the \( S_1(f) \)-curves.

To understand how the network structures govern the relative efficiencies of the strategies, we measure \( S_1(f = 0.2) \) for varying assortative mixing and clustering coefficients. The results hold for other small \( f \) values. We keep the size and degree sequence constant to the values of the arxiv network. To perform this sampling we rewire pairs of edges \((v_1, v_2)\) and \((w_1, w_2)\) to \((v_1, w_2)\) and \((w_1, v_2)\) (unless this would introduce a self-edge or multiple edges). To ensure that the \( n_{\text{rew}} = 100 \) rewiring realizations are independent, we start with rewiring \( n_{\text{init}} = 3M \) pairs of edges. Then we go through pairs of edges randomly and execute only changes that make the current \( r \) or \( C \) closer to their target values. When the values of \( r \) or \( C \) are within 0.1% of the target value the iterations are braked. The results seen in fig. 3 show that, just as before, the OUT strategy, chained or unchained, is most efficient throughout the parameter space. The unchained versions are most efficient for \( r \gtrsim 0.3 \). An explanation is that, for high \( r \), the chained versions will effectively only vaccinate the high-connected vertices (that are grouped together for very high \( r \)) and leave chains of low-degree vertices unvaccinated. The \( C \)-dependence plotted in fig. 3(b) shows that the unchained versions outperform the chained versions for \( C \gtrsim 0.15 \). This is probably a result of chains getting stuck in one part of the network. It is not an effect of biased degree-degree correlations (since if the rewiring procedure is conditioned to a fixed \( r \), fig. 3(b) remains essentially unaltered). We note that the structure of the original arxiv network differs from the rewired networks. For example, at \( f = 0.2 \) of fig. 1(a) the OUT is 22\% more efficient than OUTC, but in fig. 3 the OUT and OUTC curves differ very little. For the RN\( \bar{b} \) strategy the chained version is better than the unchained one throughout the range of \( r \) and \( C \) values.
Summary and conclusions. – To summarize, we have investigated strategies for vaccination and network attack that are based on the knowledge of the neighborhood only—information that people arguably possess and utilize. Both static and dynamic measures of efficiency are studied. For most networks, regardless of the number of vaccinated vertices, the most efficient strategy is to choose a vertex v and vaccinate the neighbor of v with most links out of v’s neighborhood (Out). v can be picked either as the most recently vaccinated vertex (chained selection) or at random (unchained selection). For real-world networks, the chained versions tend to outperform the unchained ones, whereas this situation is reversed for the two types of model networks we study. We find that the unchained strategies are preferable for networks with a very high clustering or strong positive assortative mixing. In ref. [5], the authors propose the strategy to vaccinate a random neighbor of a randomly selected vertex. This strategy (RNb) requires less information of the neighborhood than Out do. Thus, the practical procedure gets simpler: One only has to ask a person “name a person you meet regularly” rather than “name the acquaintance of yours who meet most people you are not acquainted with regularly” (for Out). (“Meet with regularly” can be replaced with some phrase signifying a high risk of infection transfer for the pathogen in question.) On the other hand, if the information of the neighborhoods is incomplete, Out will, effectively, be reduced to RNb (and thus not perform worse than RNb).

We emphasize that the structure of real-world contact networks is debated and not fully understood. Furthermore, different networks are relevant for different diseases. Sexually transmitted diseases are known to have a small fraction of short circuits [17] and highly skewed degree distributions [18, 19]. Other diseases predominantly spread over acquaintance networks, that have a high clustering [20] but where the degree distribution is yet obscure [21]. We justify our conclusions above by the fact that they are the same for the wide variety of networks we test. They are also consistent for simulations with Barabási-Albert [22] and Holme-Kim [23] networks, an empirical acquaintance network of prisoners [24]. The only real exception is that the chained strategies work poorly for the WS model networks.

To epitomize, choosing the people to vaccinate in the right way will save a tremendous amount of vaccine and side-effect cases. The best strategy can only be selected by considering both the structure of the network the pathogen spreads over, and the disease dynamics. If such detailed information is lacking we recommend OutC—a strategy that often is the best, and never much worse than the best.

***

The author is grateful for comments from M. Rosvall and acknowledges support from the Swedish Research Council through contract no. 2002-4135.

REFERENCES

[1] Giesecke J., *Modern Infectious Disease Epidemiology* (Arnold, London) 2002.
[2] Liljeros F., Edling C. R. and Amaral L. A. N., *Microbes Infect.*, 5 (2003) 189.
[3] Newman M. E. J., *SIAM Rev.*, 45 (2003) 167.
[4] Holme P., Kim B. J., Yoon C. N. and Han S. K., *Phys. Rev. E*, 65 (2002) 066109.
[5] Cohen R., Havlin S. and Ben Avraham D., *Phys. Rev. Lett.*, 91 (2003) 247901.
[6] Watts D. J., Dodds P. S. and Newman M. E. J., *Science*, 296 (2002) 1302.
[7] Albert R., Jeong H. and Barabási A.-L., *Nature*, 406 (2000) 378.
[8] Barrat A. and Weigt M., *Eur. Phys. J. B*, 13 (2000) 547.
[9] Newman M. E. J., *Phys. Rev. Lett.*, 89 (2002) 208701.
[10] Newman M. E. J., Proc. Natl. Acad. Sci. USA, 98 (2001) 404.
[11] Zachary W., J. Antropol. Res., 33 (1977) 452.
[12] Grönlund A. and Holme P., Phys. Rev. E, 70 (2004) 036108.
[13] Erdős P. and Rényi A., Publ. Math. Debrecen, 6 (1959) 290.
[14] Watts D. J. and Strogatz S. H., Nature, 393 (1998) 440.
[15] Kim B. J., Yoon C. N., Han S. K. and Jeong H., Phys. Rev. E, 65 (2002) 027103.
[16] Moore C. and Newman M. E. J., Phys. Rev. E, 61 (2000) 5678.
[17] Bearman P., Moody J. and Stovel K., Am. J. Soc., 110 (2004) 44.
[18] Liljeros F., Edling C. R., Amaral L. A. N., Stanley H. E. and Aberg Y., Nature, 411 (2001) 907.
[19] Schneebberger A., Mercer C. H., Gregson S. A. J., Ferguson N. M., Nyamukapa C. A., Anderson R. M., Johnson A. M. and Garnett G. P., Sex. Transm. Dis., 31 (2004) 380.
[20] Holland P. W. and Leinhardt S., Am. J. Soc., 72 (1972) 1205.
[21] Amaral L. A. N., Scala A., Barthélémy M. and Stanley H. E., Proc. Natl. Acad. Sci. USA, 97 (2000) 11149.
[22] Barabási A.-L. and Albert R., Science, 286 (1999) 509.
[23] Holme P. and Kim B. J., Phys. Rev. E, 65 (2002) 026107.
[24] MacRae J., Sociometry, 23 (1960) 360.