Immunizing complex networks with limited budget

Baharan Mirzasoleiman, Mahmoudreza Babaei and Mahdi Jalili (a)

Department of Computer Engineering, Sharif University of Technology - Tehran, Iran

received 9 January 2012; accepted in final form 3 April 2012
published online 15 May 2012

PACS 89.75.-k – Complex systems
PACS 89.75.Fb – Structures and organization in complex systems
PACS 89.75.Hc – Networks and genealogical trees

Abstract – In this letter we studied the epidemic spreading on scale-free networks assuming a limited budget for immunization. We proposed a general model in which the immunity of an individual against the disease depends on its immunized friends in the network. Furthermore, we considered the possibility that each individual might be eager to pay a price to buy the vaccine and become immune against the disease. Under these assumptions we proposed an algorithm for improving the performance of all previous immunization algorithms. We also introduced a heuristic extension of the algorithm, which works well in scale-free networks.

Copyright © EPLA, 2012

Introduction. – Infectious diseases have been a major cause of death, disability, social and economic disruption for millions of people worldwide. The dynamics of epidemics is strongly influenced by the underlying structure of the contact network. Many real networks show scale-free degree distributions \cite{1-3}, and it has been revealed that the spread of infections on scale-free networks does not have any epidemic threshold \cite{4}. Finding an efficient immunization method against the spread of infectious diseases in these networks can significantly reduce the cost and pains in case of an epidemic outbreak.

Models such as susceptible-infected-susceptible (SIS) and susceptible-infected-removed (SIR) are well-known models to consider infectivity \cite{5} that assume equal transmission probability over all links of the contact network. However, this is not a realistic assumption in real scale-free networks. In fact, the infection transmission probability over a link can be considered as a function of the contact time between an infected and a susceptible individual, the body resistance of the susceptible individual against disease and many other factors. In the network of possible contacts, these parameters can be modeled as the weight of the links between the nodes or individuals. On the other hand, as a disease cannot be transmitted from an immunized individual to healthy ones, the immunized neighbors of an individual can make him/her immune against the disease to some extent. In this paper, we propose a new model in which the probability for an individual to be exposed to the infection is a function of the immunized individuals in the network.

The effects of immunization strategies are often studied by ignoring the vaccination costs or assuming an infinite budget; however, this might not be the case in real situations. Here, we aim at a more natural objective by finding an effective immunization strategy considering a limited budget. We are interested in reducing the minimum budget required for globally immunizing a network against infections. To this end, models available for diffusion of innovations \cite{6,7} and influence maximization \cite{8-10} can be used. Here we aim at determining appropriate offers to those who are eager to pay a price for the vaccination. The lower price increases the probability of selling the vaccine thus decreasing the revenue that can be added to the budget in order to immunize a higher number of individuals in the network. The strategy should be able to make a trade-off between these issues. We introduce two approaches for determining applicable offer sequences in different networks. Computational experiments show that these methods can extensively increase the efficiency of the immunization strategies.

Model. – Consider a network of individuals in which each link represents a connection along which the infection can spread. Let us suppose that producing each vaccine piece has a unit cost and a limited budget is allocated for vaccination. In addition, each individual \( i \) might be eager to pay a price to become immunized.

Disease transmission and immunity model. The individuals who have been vaccinated against the disease do
not get infected, and thus, do not transmit the disease. Thus, immunization is of value from two perspectives. On one hand, it decreases the overall probability of infections by reducing the number of susceptible individuals in the network. On the other hand, as the immunized individuals cannot become infected, the disease cannot be transmitted through their links to the susceptible ones. This reduces the infections probability of the susceptible individuals in the network. Based on this intuition, we model the immunity of individual $i$, $q_i(.)$, as a function of the set of other individuals in the network who have already been immunized.

When the network is modeled by a graph, $q_i : 2^V \to R^+$ is a function of neighbors of node $i$ in the graph, i.e. $q_i(S) = f_i\left(\sum_{j \in S \cup \{i\}} w_{ij} / \sum_{k \in V} w_{ik}\right)$, where $V$ is the set of all individuals in the network, $S \subseteq V \setminus \{i\}$ is the set of all individuals who have already been immunized in the network and $w_{ij}$ is the weight of link $e_{ij}$. Since the exact information about the link weights cannot be obtained in real contact networks, we assume that we know the distribution from which the weights are drawn, i.e. the values of $w_{ij}$'s are drawn independently of a distribution $F_{ij}$ for all $j \in S$.

**Influence model.** In order to model the willingness of individuals to buy the vaccine and become immune, we use the influence extraction model [10]. Consider a seller and set $V$ of potential buyers; $v_i(S) : 2^V \to R^+$ is the value of the good for buyer $i$ if the set $S$ of buyers already own the item. Each buyer $i$ receives feedbacks from the set of his/her neighbors who have bought the same item. For high-quality items these feedbacks will be positive and increase the valuation of the buyer $i$ for the good. The valuation of the buyer $i$ can be modeled as a non-negative monotone concave function $f_i : R^+ \to R^+$ [10]. For all $i \in V \setminus S$, we have $v_i(S) = f_i\left(\sum_{j \in S \cup \{i\}} w_{ij} / \sum_{k \in V} w_{ik}\right)$, where the link weights represent the influences that individuals have on each other.

**Immunization strategies.**

**Random immunization.** Random or uniform strategy selects all the individuals within the population with the same probability. Recall that a network remains contagious if immunization does not destroy its connectivity. In other words, a network acquires global immunization if removal of the immunized individuals damages the global integrity of the network. It is well-established that scale-free networks possess high resiliency to random connection failures, i.e. almost all the nodes in a scale-free network need to be removed in order for a network to become disconnected [4,11].

**Acquaintance immunization.** Without having global information about the node degrees in a network, acquaintance immunization can significantly boost the effectiveness of the random strategy [12]. In this approach a random acquaintance of a randomly chosen node will become immunized. The nodes with higher degree are more likely to be chosen by the acquaintance method. An enhanced acquaintance immunization strategy was also proposed to improve its efficiency [13]. In this method, first, a random node is selected. Then, one of its neighbors with higher degree is randomly chosen and becomes immunized. Alternatively, one might choose the neighboring nodes if they have degrees larger than a threshold $k_{cut}$. If there are no such neighbors, no node is chosen.

**Targeted immunization.** Although scale-free networks are very robust against random failures, selective damages can strongly affect their integrity [11]. We can take advantage of this property to devise an efficient targeted immunization scheme that takes into account the heterogeneity of scale-free networks. The heavily connected nodes are highly exposed to the infection and are more likely to spread the disease. Targeting these nodes, therefore, hinders the infections from penetrating through other parts of the network and noticeably boosts the effectiveness of an immunization strategy. Several studies on model and real scale-free networks revealed the efficiency of the targeted immunization strategy, see for example [11,14], which makes it the simplest solution to the optimal immunization strategy in heterogeneous networks.

**Greedy hill-climbing.** Finding an optimal immunization strategy can be considered as a problem of finding the initial set of nodes $S$ to be immunized in order to maximize the global immunity $g(S)$ in the network. The greedy hill-climbing algorithm is a mathematical optimization technique for finding a local optimum out of all possible solutions. Starting with an empty set $S$, in any iteration, the element which maximizes the function value is added to the solution set $S$. The algorithm based on a hill-climbing approach is as follows.

1. Initialize set $S = \emptyset$.
2. Among buyers who accept the offered price choose buyer $i$ such that $i = \arg\max g(S \cup \{i\}) - g(S)$.
3. If $g(S \cup \{i\}) \leq g(S)$, output $S$.
4. $S = S \cup \{i\}$ and go to step 2).

Computational experiments on large networks showed that this approximation algorithm significantly outperforms the targeted strategy and results in a better influential set [9]. Although the problem of selecting the most effective nodes for immunization is NP-hard, the greedy hill-climbing provides an approximation guarantee arbitrarily close to $(1 - 1/e)$—slightly better than 63%—for any non-negative submodular function as the immunity function of each node in a network [8]. Function $f(.)$ is submodular if it satisfies the intuitive “diminishing returns” property, i.e. adding an element to a small set increases the function value more than adding an element to a large set.
Finding the optimal pricing strategy. – Each individual in the network might be eager to voluntarily pay a price to buy the vaccine and become immune. In order to improve the result of an immunization strategy, we offer the vaccine to the individuals chosen by the algorithm with a discounted price in a way that most of them accept the offer. These values can be added to the initial budget to make a higher number of people immune against the disease.

In the sequel we introduce two strategies to determine an appropriate offer sequence.

Pricing based on average degree. Here we introduce a strategy that uses only the average degree of a network (μ) to determine an appropriate offer sequence in any network. We then prove that this approach improves the result of any immunization strategy.

As the individuals in the network are influenced by those who have already been vaccinated, all the nodes have zero value before the immunization starts. Therefore, we should give away the vaccine for free to the first nodes chosen by an immunization algorithm to increase the exerted influence on the remaining individuals. Let us suppose that giving away the item to k1 nodes who are added to the set S, the exerted influence on any nodes i ∈ V \ S becomes at least d1/μ, where d1 is degree of node i. From this point on, all the nodes chosen by the algorithm buy the vaccine with a price at least equal to f(1/μ). Therefore, we can offer the vaccine to the next k2 individuals chosen by the algorithm with the price of f(1/μ) and be assured about the acceptance of the offers. We continue this process until the exerted influence on any nodes i ∈ V \ S becomes at least 2d1/μ, and then, offer the vaccine to the next k3 individuals with the price of f(2/μ). Indeed, in step j, where j = 1, . . . , n, n = [μ] 1, we offer the item with the price of f(j−1/μ) to the kj individuals chosen by the algorithm. Adding these nodes to the set S the values of the remaining individuals in V \ S become at least equal to f(j/μ).

Lemma 1. In a network of size N, for any set S, |S| ≥ kN渥 and all i ∈ V \ S, the expected value of v1 is at least f(kN渥).

Proof. Consider real-valued random variable X(j), j ∈ S as follows:

\[ X(j) = \begin{cases} 1, & \text{if } w_{ij} > 0, \\ 0, & \text{if } w_{ij} = 0, \end{cases} \quad (1) \]

where i is an arbitrary node. The expected influence on node i is

\[ E \left( \sum_{j \in S \cup \{i\}} [X(j)] \right) = \sum_{j \in S \cup \{i\}} E([X(j)]) \approx \sum_{j \in S \cup \{i\}} \frac{d_i \cdot E_{ij}}{N \cdot F_{ij}} = \frac{k d_i}{\mu}. \quad (2) \]

where \( F_{ij} \) is the expected value of the distribution from which the link weights are derived.

Thus, the value of buyer j is

\[ f_j \left( \frac{\sum_{i \in S \cup \{i\}} w_{ij}}{\sum_{k \in V} w_{ik}} \right) \approx f_i \left( \frac{k \cdot d_i \cdot F_{ij}}{\mu \cdot d_i \cdot F_{ij}} \right) = f_i \left( \frac{k}{\mu} \right). \quad (3) \]

Lemma 2. Using the pricing strategy based on μ with any immunization algorithm will increase the global immunity in the network.

Proof. Considering Lemma 1, for any immunization algorithms, we guarantee that any node chosen by the algorithm can buy the item with the offered price. As we do not give away the item for free to all nodes in set S some value will be added to the initial budget which can be utilized to immunize more people and improve the global immunity in the network.

Note that this method does not use any information about the structural properties of the network or the immunization algorithm. This is particularly interesting for the case of random strategy, since the exact structure of the contact networks cannot be usually extracted, in many cases and the random immunization is the only applicable strategy.

Greedy pricing approach. In many networks we can extract complete or at least partial information about the connectivity. The targeted strategy can appropriately use this information to significantly improve its efficiency in scale-free networks. Immunization of the heavily connected nodes has two major effects. On the one hand, many connecting paths between different parts of the network become disconnected and the giant component of the network is partitioned into many smaller isolated components. This prevents the disease from propagating throughout the network. On the other hand, highly connected nodes increase the willingness of many other nodes to voluntarily pay a price and become immunized. In this section we introduced a strategy which uses network structural properties to further improve the result of the targeted strategy in scale-free networks.

Similar to the previous approach, we determine the appropriate offer sequence in n step, n = [μ] 1. In step j, the goal is to offer the vaccine to the potential buyers at a discounted price of f(j−1/μ) in a way that they all accept the offer and buy the vaccine. From this point on, we can offer the vaccine at a higher price of f(j/μ) and still guarantee the acceptance of the offer. As the targeted strategy immunizes the nodes in decreasing order of their connectivity, in the earlier steps a lower number of nodes are needed to guarantee the acceptance of the offers in the following steps compared to the previous strategy.

Consider a non-negative, monotone concave function f as the value function of the nodes in a network. Let us suppose that adding kj nodes to set S in step j, the value of other potential buyers become at least f(j/μ). We divide the area under v1() into r = [μ] regions, and in each step, add the nodes in decreasing order of their degree to
the set $S$ until the number of nodes in interval $j$ becomes maximized. The goal is to maximize the number of nodes, especially influential ones, in some interval and offer them an appropriate discount for the item. This way we guarantee the acceptance of the offer by almost all the candidate buyers chosen by the targeted strategy. The normalized influence on an arbitrary buyer $i$ can be in the interval $[0, 1]$, i.e., $0 \leq v_i(S) = \sum_{j \in S, i \in (t)} w_{ij} / \sum_{k \in V} w_{ik} \leq 1$. Therefore, we divide the area under the concave value function $v_i(S)$ into $r$ influence regions each with width $1/r$ as shown in fig. 1.

Our greedy algorithm for determining the discount sequence is as follows.

1) Sort the nodes of a network in decreasing order of their degrees in array $D$.
2) Initialize $k_0 = 0$, $i = 1$.
3) For $j = 1, \ldots, n$, repeat the following steps.
4) If $j \leq \left\lfloor \frac{\mu}{r} \right\rfloor$, then $i = i + 1$.
5) $Z = \left\lfloor \frac{\mu}{r} \right\rfloor$.
6) Find $k_m = \arg \max_k \left( \sum_{S = D(1,k)} i \in V/S \right)$ where
   
   $$X(i) = \begin{cases} 1, & \frac{i}{r} \leq \sum_{j \in S, i \in (t)} w_{ij} / \sum_{k \in V} w_{ik}, \\ 0, & \text{otherwise,} \end{cases}$$

   for all $i \in V \backslash S$.
7) $k_j = k_m - k_{m-1}$.
8) Offer the vaccine with the price of $f(j - 1/\mu)$ to the nodes chosen by the targeted strategy until $k_j$ nodes accept the offer and buy the vaccine.

In order to find an appropriate value for $k_j$, we should examine the normalized influence on an appropriate set of potential buyers. In the early steps, the nodes with small degree are not likely to be chosen by the targeted strategy. Thus, we should ignore these nodes in the early steps. As the number of nodes chosen by the targeted strategy increases, the probability that the low-degree nodes are chosen increases. In scale-free networks, the probability for a node to have $k$ connections has inverse relation with $k$. Considering this fact, in order to study the normalized influence on an approximately equal number of nodes in each step, we examined the nodes whose degree lies in the same logarithmic interval in each step. In the first step, we consider nodes which have a degree larger than $\mu/2$ in the first $n/2$ steps. Then, we consider nodes which have a degree larger than $\mu/4$ in the next $n/4$ steps and so on.

**Experiments.**

Immunity and influence models. We choose a non-negative, monotone concave function (generalized Pareto cumulative distribution function) for $q_i$ as follows:

$$F_{(\xi,\mu,\alpha)}(y) = \begin{cases} 1 - \left( 1 + \frac{y - \mu}{\sigma} \right)^{-1/\xi}, & \text{for } \xi \neq 0, \\ 1 - \exp \left( - \frac{y - \mu}{\sigma} \right), & \text{for } \xi = 0. \end{cases}$$ (4)

Choosing $\xi = 0$, $\mu = -0.1$, we simply have an exponential cumulative distribution function which is shifted along the $x$-axis. Furthermore, we considered $\sigma = 2$, $y = 6x$, $x \in [0, 1]$, to have a function that takes values in the interval $[0.04, 0.95]$. By choosing $\mu = -0.1$ the possibility for “spontaneous” (or “automatic”) infection is also considered [15]. Here, we assume 0.04 as the possibility for spontaneous infection and a probability of 0.05 for each individual to be intrinsically immunized against the disease [15]. Finally, $\sigma = 2$ allowed us to have a function with low concavity. This makes our results comparable to those of standard SIR and SIS models. Recall that in these models the probability for a node to become infected is a linear function of the number of its sick neighbors [16] (fig. 2(b)).

As our influence model ($v_i$) we considered an exponential cumulative distribution function: $F(y/\mu) = 1 - \exp \left( - \frac{y}{\sigma \mu} \right)$ with $\sigma = 2$, $y = 6x$, $x \in [0, 1]$ whose values vary in the interval $[0, 0.95]$. We also chose $\sigma = 2$ in order to decrease the concavity of the exponential function. Under this assumption, the valuation of the individuals in set $V \backslash S$ does not rapidly grow as more of their friends buy

---

**Fig. 1:** (Colour on-line) $v_i(\cdot)$ as a function of normalized influence for a network with $4 \leq \mu < 5$. The area under $v_i(\cdot)$ is divided into $r = 4$ influence regions. In each step, the number of nodes in the blue interval should be maximized. Panels (a)-(d) show the intervals that should be maximized in steps 1–4.

**Fig. 2:** (Colour on-line) $v_i(\cdot)$ and $q_i(\cdot)$ as a function of “$y = 6 \times$ normalized influence on each individual”. The dashed line (in red) in (a) shows the probability of being exposed to the disease.
the vaccine (fig. 2(a)). Certainly, assuming a function with higher concavity or a function with values greater than the cost of producing each vaccine unit considerably improves the results of the immunization. In this work, we considered a more conservative approach by considering the smallest convexity.

**Algorithms and implementation.** Suppose that we want to immunize a population with a specific immunization strategy. During the immunization process, we offer the vaccine with a predetermined price to each person chosen by the immunization method. The offer sequence can be determined from the proposed pricing strategies before the immunization starts. If the chosen person accepts the offer, the value she pays for the vaccine will be added to the initial budget in order to produce more vaccine units and continue vaccinating more individuals with the immunization method. Otherwise, the immunization method chooses the next candidate and offers the vaccine with the specified price to him/her. We can determine the global immunity of the network as a function of the immunized individuals, $q(.)$.

We investigated the performance of the random, acquaintance, enhanced acquaintance and targeted immunization strategies using the two pricing methods introduced in the previous section. We also compared our results with those of the greedy hill-climbing algorithm. Despite of its high computational complexity, the hill-climbing strategy gives a $\left(1 - \frac{1}{e}\right)$-approximation of the optimal solution. It should be noted that, when the individuals pay a price to buy the vaccine, the problem of maximizing the immunity in the network becomes non-monotonic submodular, and thus, we cannot use the hill-climbing strategy. We assumed that the link weights are derived independently of the distribution function $F_{ij}$ that is considered as uniform cumulative distribution on the interval $[0, 2]$.

**Results.** The results in fig. 3 correspond to applying the $\mu$-based pricing strategy on the random, acquaintance and enhanced acquaintance immunization strategies. The global immunity (the fraction of immunized nodes) as a function of initial budget for the immunization is shown for the preferential attachment Barabasi-Albert (BA) [17] and Watts-Strogatz (WS) [18] networks. It is seen that this simple approach already offers a significant improvement over the original immunization methods.

Figure 4 shows the global immunity as a function of initial budget for the immunization for BA [17] and Forest-Fire (FF) [18] network models, respectively. For each network, the results of using the pricing approach based on $\mu$ are shown for the random and high-degree targeted strategy (HD). Furthermore, the global immunity that resulted by using the greedy discount approach with the targeted strategy is compared to the global immunity of the greedy hill-climbing algorithm. These results revealed that the greedy hill-climbing strategy outperformed the basic targeted algorithms. However, the proposed average degree-based and the greedy pricing approaches outperformed the hill-climbing approach in the BA network, while they have a close performance in the other network types. Although the hill-climbing algorithm can guarantee the $\left(1 - \frac{1}{e}\right)$-approximation of the optimal solution for non-negative monotone submodular functions, its high computational complexity makes it inapplicable on the large networks. Therefore, it is interesting to find out that the targeted strategy reached a similar performance with a considerably lower complexity level.

Another interesting observation was that the initial budget required for stopping the epidemic spread significantly decreased by using the two pricing strategies. As it is seen, the proposed pricing-based random immunization can be more influential than the targeted immunization.

Figure 5 shows the result of the experiments on a number of real networks including the Facebook-like
Fig. 5: (Colour on-line) Global immunity of the network as a function of initial budget for immunization, for (a) the Facebook-like social network with $\mu = 14.61$, (b) the yeast protein interaction network with $\mu = 6.14$, (c) the high-energy physics theory citation networks with $\mu = 25.69$, (d) the CAIDA AS relationship network with $\mu = 4.03$, (e) the Enron email network with $\mu = 10.73$, and (f) the corporations inter-relationships with $\mu = 2.07$ (EVA) network. Other designations are as in fig. 4.

social network [19], the yeast protein interaction network whose social behavior has been discussed in [20], the high-energy physics theory citation networks [18], the CAIDA AS relationship network [18], the Enron email network [21] and the corporations inter-relationships (EVA) network [22].

In all these networks, similar qualitative behavior was observed for the discounting strategies. The average degree-based pricing strategy extensively improved the global immunity of the random strategy. Furthermore, this approach globally immunized the network against infections with considerably smaller budget than the traditional targeted strategy. This strategy can be simply used in networks in which no structural information is in hand. Another interesting observation was that using the greedy pricing approach with the targeted strategy, the network can be immunized with a considerably smaller budget compared to the hill-climbing strategy. In many cases, the greedy pricing approach achieved a higher immunity than the hill-climbing algorithm for small initial budgets.

Conclusion. – We proposed a general model in which the immunity of an individual against the disease depends on the set of his/her immunized friends in the network. Furthermore, we considered the possibility that each individual might be eager to pay a price to buy the vaccine and become immune against the disease. We proposed pricing-based immunization algorithms in order to enhance immunization strategies. We also proposed an algorithm based on the greedy hill-climbing strategy. Extensive simulations on model networks as well as several real networks showed that our strategies can make the result of the targeted strategy comparable to the results of the greedy hill-climbing algorithm. Furthermore, applying the pricing strategies on the random, acquaintance and targeted strategies, these algorithms can stop the spreading of the epidemic with a significantly smaller budget compared to the hill-climbing one.

REFERENCES