A new immunization algorithm based on spectral properties for complex networks

Ramin Zahedi

Department of Science and Engineering
Sharif University of Technology
Kish, Iran
zahedi@kish.sharif.edu

Mohammad Khansari

Faculty of New Sciences and Technologies
University of Tehran
Tehran, Iran
m.khansari@ut.ac.ir

Abstract— **Nowadays, we are facing epidemic spreading in many different areas; examples are infection propagation, rumor spreading and computer viruses in computer networks. Finding a strategy to control and mitigate the spread of these epidemics is gaining much interest in recent researches. Due to limitation of immunization resources, it is important to establish a strategy for selecting nodes which has the most effect in mitigating epidemics. In this paper, we propose a new algorithm that minimizes the worst expected growth of an epidemic by reducing the size of the largest connected component of the underlying contact network. The proposed algorithm is applicable to any level of available resources and, despite the greedy approaches of most immunization strategies, selects nodes simultaneously. In each iteration, the proposed method partitions the largest connected component into two groups. These are the best candidates for communities in that component, and the available resources are sufficient to separate them. Using Laplacian spectral partitioning, the proposed method performs community detection inference with a time complexity that rivals that of the best previous methods. Experiments show that our method outperforms targeted immunization approaches in real networks**

Keywords: Complex Networks; Immunization; Graph Spectra;

I. INTRODUCTION

Networks can be used to model many complex systems, such as human social interactions, protein interactions inside cells, and the Internet. In these networks, the actors within the system (i.e. people, proteins, or computers) are the network nodes, and the interactions between them constitute network edges [1]. The threat of harmful spreading processes (i.e. epidemics) such as computer viruses [1, 2], rumours, and infectious diseases [2, 4, 5] has led to the establishment of strategies and algorithms that immunize the complex systems. Resource limitations require immunization strategies to somehow prioritize nodes based on the impact of their removal on preventing the epidemic from spreading [5-8].

The problem of network immunization under a limited budget is equivalent to decreasing the network's vulnerability by removing a limited number of nodes. In immunization algorithms, the immunized nodes and all their edges are removed from the network, because they have no further role in epidemic transmission. This is a combinatorial optimization problem whose goal is to minimize the number of patients by immunizing a limited subset of nodes.

The subset of nodes from which all others are reachable denotes a network connected component. The aim of node removal is to partition the network into smaller components such that, if an epidemic starts from one network component, it would fail to spread to other components. In the case of an epidemic spreading, the size of the largest connected component (LCC) denotes the network vulnerability. The size of the LCC is an effective measure of the network vulnerability to the worst-case epidemic size.

Although mass immunization of the population can be an effective strategy [9], the high cost and lack of resources mean this is often impossible [10]. Random immunization, where a specific number of randomly selected individuals are immunized, is common in scenarios with strict budget constraints. The effectiveness of random immunization is very low, with around 80% of nodes requiring immunization to eliminate an epidemic in a population [3, 11, 12]. Because of the inefficiency of mass and random strategies, researchers are continually seeking solutions that focus on those nodes that play a key role in the rate and extent of epidemic transmission.

Most immunization algorithms first prioritize the nodes, and then select those with the highest priority [1, 5, 7, 13, 14]. Most approaches use node centralities (e.g. the degree or betweenness centrality) in the network to perform prioritization (i.e. targeted immunization). The degree centrality is measured by counting the number of edges connected to a node. Immunization of nodes with the highest degree centrality (HD) reduces the network density, and has a significant effect on the growth rate of an epidemic [15, 16]. The betweenness centrality is measured by calculating the proportion of times a node lies on the shortest paths in a network [16]. Immunization of nodes with the highest betweenness centrality (HB) cuts off the main transmission routes. Although these strategies are much better than random immunization, they are far from optimal [17].

Recent studies have shown that many real networks contain distinct communities [18-20]. These are groups of nodes that have dense sets of interactions among themselves and

relatively few interactions with other nodes. In a small network, these communities can be identified by eye, but in real networks consisting of thousands of nodes, the use of algorithms is required to perform community detection inference [21, 22]. It has been shown that the nodes connecting different communities (so-called bridge nodes) play an important role in the spread of epidemics through community-structured networks. Their immunization breaks down the transmission routes and prevents infection propagation. Hence, identifying and immunizing bridge nodes can reduce the cost of immunization and minimize the number of infected individuals with a limited budget. Some immunization strategies attempt to find these communities and cut off their between-community interactions [13, 23, 24]. As a result, the network will consist of separated communities, such that if an epidemic starts in one community, it cannot propagate to the others. The problem with such strategies is that they need a certain level of resources, and cannot be applied if the available resources do not meet this threshold.

In this paper, we introduce a new immunization algorithm that is applicable for any level of available resources and, unlike the greedy approaches of targeted strategies, selects nodes simultaneously. The aim of the proposed algorithm is to reduce the maximum expected growth of an epidemic by minimizing the size of the LCC. Our iterative method partitions the LCC into two groups in each iteration. These groups are the best candidates for communities in that component, and the available resources are sufficient to separate them. We use graph spectra to perform community detection inference with a time complexity that rivals that of the best previous methods.

II. PROPOSED ALGORITHM

In [25], it was shown that one of the statistical inference methods used for community detection could be mapped to some versions of the problem of minimum-cut graph partitioning. Therefore, to perform community detection inference we can use Laplacian spectral partitioning. In our algorithm, we employ the method presented in [25]. Hence, we briefly describe its execution.

First, we review the Fiedler method for partitioning a graph into two separate groups using Laplacian spectra. For a minimum-cut graph division into two separate groups of arbitrary size n_1 and n_2 ($n_1 + n_2$ is equal to the number of nodes), we calculate the eigenvector corresponding to the second smallest eigenvalue of the graph Laplacian matrix $\mathbf{L}$ (the graph Laplacian is a symmetric $n \times n$ matrix where $\mathbf{L} = \mathbf{D} - \mathbf{A}$, here $\mathbf{D}$ is an $n \times n$ diagonal matrix which D_{ii} equals to the degree of vertex i and $\mathbf{A}$ is the adjacency matrix). After calculating this eigenvector (known as the Fiedler vector), we can divide the graph into two groups by assigning the n_1 largest elements to group one, and the remainder to group two. This method enables a graph to be divided into two separate groups in $n - 1$ different ways: First, we put the node with the largest corresponding element in the Fiedler vector in one group and the rest in the other group. Then, one-by-one, we move the nodes with the largest corresponding elements in the Fiedler vector from the second group to the first, until only

one node remains in the second group. Although the Fiedler method only gives an approximation for a minimum-cut graph partitioning, practical experience suggests it yields a good estimation [26].

Reference [25] first calculates the Fiedler vector, then computes a likelihood for every $n - 1$ different partitions. The likelihood of a partition is proportional to the probability that such a partition is a good estimate for the communities in that graph, and is calculated as:

$$Q = m_{in} \ln\left(\frac{2m_{in}}{K_1^2 + K_2^2}\right) + m_{out} \ln\left(\frac{m_{out}}{K_1 K_2}\right),$$

Where m_{in} and m_{out} are the number of edges inside groups and between groups, respectively, and K_1 and K_2 denote the sum of the degrees of the nodes in groups one and two. Partitions with a higher likelihood are more likely to represent the communities of a network [25].

After calculating the likelihood of all partitions, we prioritize the partitions, and choose the partition with the highest priority that could be separated with the given level of resources. First, we determine the largest connected component of the network, and then use the method described above to find the best partition for dividing the LCC into two separate groups. Next, we calculate the minimum number of nodes required to separate these groups. If this number can be immunized with the available resources, we choose that partition. However, if immunizing this number would require more than the available resources, we consider the partition with the next highest likelihood. This process continues until we reach a partition for which the minimum number of nodes required for separation consumes less than or equal to the available resources (we will never reach a deadlock, because there always exist two partitions with one node in one group that can become separated by immunizing that one node). We immunize the relevant nodes, and remove them from the network (i.e. separate the two groups). After updating the level of available resources, we again determine the largest connected component of the network, and continue the algorithm until all resources have been exhausted.

ALGORITHM 1. THE PROPOSED ALGORITHM

Input: G: a network, k: number of immunization resources
Output: G′: immunized network
Repeat
  Determine the largest connected component of the network.
  Calculate the Fiedler vector for the LCC.
  Calculate the likelihood for all n-1 possible partitions.
  Choose the partition with the largest likelihood for which the minimum number of nodes required to separate the two groups would be less than or equal to k.
  Immunize the required nodes and separate two groups.
  Update the value of k.
Until k becomes zero.

To calculate the minimum number of nodes required to separate the two groups, we first find the connecting nodes, then count the number of edges from each connecting node to the other group. The groups are separated by removing the nodes with the most edges going to the other group. After removing a node, we update the number of edges going to the other group for the remaining connecting nodes, and repeat this process until the two groups become separated.

ALGORITHM 2. SEPARATING GROUPS BY REMOVING MINIMUM CONNECTING NODES

Input: a partition of graph, consisting of two groups
Output: two separated groups
Find the connecting nodes.
Count the number of edges from each connecting node to the other group.
Repeat
  Remove the node with the maximum going to another group edges.
  Update the number of going to another group edges for remaining connecting nodes.
Until two groups become separated.

III. TIME COMPLEXITY ANALYSIS

In this section, we calculate the time complexity of one iteration of the proposed algorithm, then discuss the overall complexity. In each iteration, we can determine the LCC of a network using a breadth first search or depth first search in a time proportional to the number of nodes and edges $O(m+n)$. The Fiedler vector can be calculated using the iterative Lanczos method, where each iteration is of $O(m)$. The Lanczos method requires relatively few iterations, but the exact number is unknown. The $n-1$ partitions considered in each iteration differ from the previous partition in just one node, and the movement of node i from one group to the other causes the likelihood equation to change as follows:

$$K_1 \rightarrow K_1 - k_i \qquad\qquad K_2 \rightarrow K_2 + k_i$$
$$m_{in} \rightarrow m_{in} - \Delta m \qquad\quad m_{out} \rightarrow m_{out} + \Delta m$$

Here, Δm is equal to the number of edges between the nodes in group one and node i minus the number of edges between the nodes in group two and node i. For the movement of one node, these changes and the changes in likelihood can be determined in a time proportional to the degree of that node. Thus, for all n nodes, these changes can be calculated in a time proportional to sum of all degrees ($2m$). Therefore, searching for the partition with maximum likelihood has a time complexity of $O(m)$. Finding the connecting nodes by searching all nodes and edges takes $O(m+n)$ time. We can conclude that our algorithm has a time complexity equal to $O(m+n)$ for one iteration.

The number of iterations depends on the level of resources and the network's community structure. If the level of resources is approximately equal to the number of nodes and the network's community structure is such that causes the algorithm to separate one node in each iteration, then the worst-case scenario would see the number of iterations equal the number of nodes. In this case, the time complexity of the proposed algorithm is $O(n*(m+n))$, but our experiments suggest that for the level of resources equal to the number of nodes, the number of iterations in real networks is proportional to $lg(n)$. In addition, the number of edges in the worst case is proportional to n^2, but we know that real networks are sparse and have an edge number m that is proportional to number of nodes n. Thus, the time complexity in real cases is $O(n(lg(n))$.

IV. SIMULATION AND RESULTS

We now compare the effectiveness of the proposed algorithm with targeted immunization by comparing different algorithms in real networks. The considered networks have been widely used in the literature as examples of real networks. The first one is a community of dolphins at Doubtful Sound, New Zealand [27]. The second is a network of books on US politics. The edges represent the frequency of co-purchasing by the same buyers; the network can be found on Krebs' website[1]. The third network considers American football games between some colleges in the fall of 2000 [20], and the fourth is a neural network of C. Elegans [28]. We test the effectiveness of the proposed algorithm in comparison with HD centrality and HB centrality in real networks by plotting the size of the LCC versus the number of immunized nodes (fig.1).

The proposed algorithm exhibits a clear advantage compared to the others. The size of the LCC given by our approach is smaller than HD and HB by *14%* and *10%*, respectively, in the dolphin network and by *8%* and *11%*, respectively, in the political book network. This advantage increases in networks three and four (to *16%* and *14%* for the college football network and to *13%* and *16%* for the neural network). As expected, the proposed algorithm yields better performance than centrality-based methods. Unlike the greedy approach of HD and HB approaches, which first prioritizes all the nodes and then, for any level of immunization resources selects the highest-priority nodes, the proposed algorithm chooses nodes simultaneously.

V. CONCLUSION

In this paper, we have used the mapping between community detection inference and minimum-cut graph partitioning to develop a new community-based immunization algorithm that is applicable for any level of available resources. In each iteration, the proposed algorithm prioritizes candidate partitions in the LCC, and selects the highest priority partition for which the available resources are sufficient for its separation. Our approach has a time complexity of $O(n(lg(n))$ in real networks.

(a)

(b)

(c)

(d)
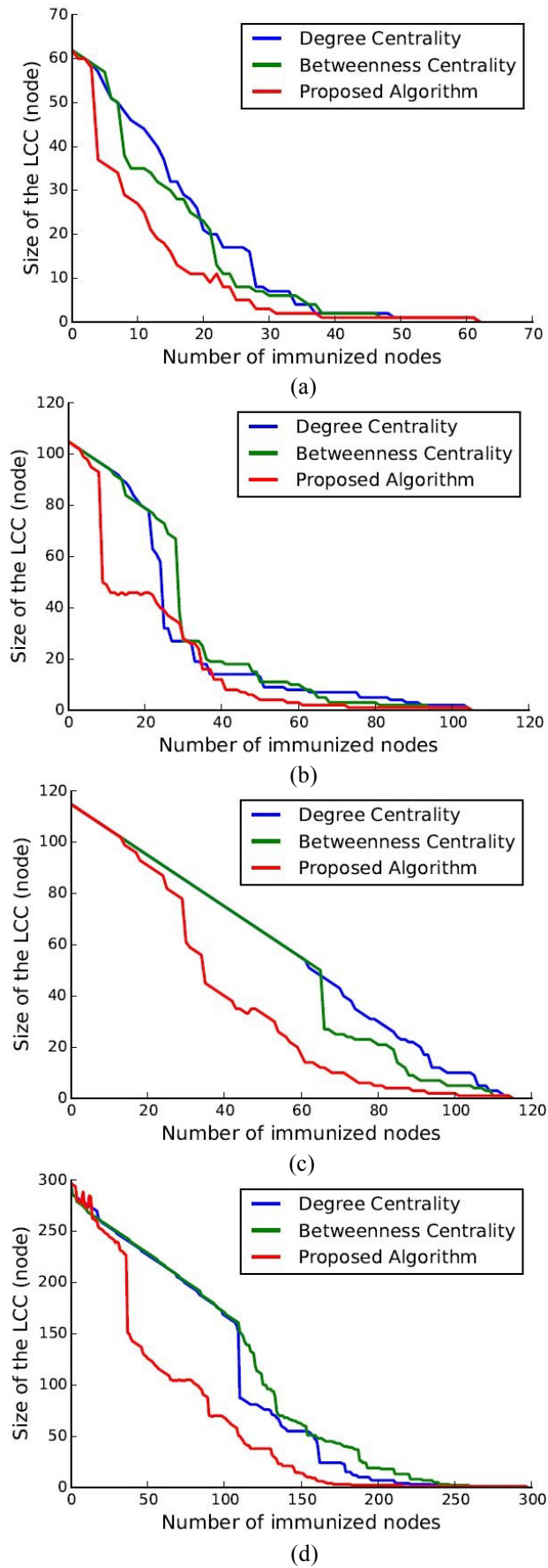
Fig.1)Size of the LCC versus the number of immunized nodes for a) dolphin network b) political book network c) college football network and d) Elegans neural network.

We measured the efficiency of the proposed algorithm by comparing it with targeted immunization on some real networks. The simulation results showed that our approach yields better performance.

VI.   REFERENCES

[1]   Gao, C., J. Liu, and N. Zhong, *Network immunization with distributed autonomy-oriented entities.* Parallel and Distributed Systems, IEEE Transactions on, 2011. **22**(7): p. 1222-1229.

[2]   Bishop, A.N. and I. Shames, *Link operations for slowing the spread of disease in complex networks.* EPL (Europhysics Letters), 2011. **95**(1): p. 18005.

[3]   Gao, C., J. Liu, and N. Zhong, *Network immunization and virus propagation in email networks: experimental evaluation and analysis.* Knowledge and information systems, 2011. **27**(2): p. 253-279.

[4]   Hartvigsen, G., et al., *Network structure, and vaccination strategy and effort interact to affect the dynamics of influenza epidemics.* Journal of theoretical biology, 2007. **246**(2): p. 205-213.

[5]   Schneider, C.M., T. Mihaljev, and H.J. Herrmann, *Inverse targeting—An effective immunization strategy.* EPL (Europhysics Letters), 2012. **98**(4): p. 46002.

[6]   Chen, Y., et al., *Finding a better immunization strategy.* Physical review letters, 2008. **101**(5): p. 058701.

[7]   Schneider, C.M., et al., *Suppressing epidemics with a limited amount of immunization units.* Physical Review E, 2011. **84**(6): p. 061911.

[8]   Ventresca, M. and D. Aleman, *Evaluation of strategies to mitigate contagion spread using social network characteristics.* Social Networks, 2013. **35**(1): p. 75-88.

[9]   Kaplan, E.H., D.L. Craft, and L.M. Wein, *Emergency response to a smallpox attack: the case for mass vaccination.* Proceedings of the National Academy of Sciences, 2002. **99**(16): p. 10935-10940.

[10]   Gallos, L.K., et al., *Improving immunization strategies.* Physical Review E, 2007. **75**(4): p. 045104.

[11]   Dezső, Z. and A.-L. Barabási, *Halting viruses in scale-free networks.* Physical Review E, 2002. **65**(5): p. 055103.

[12]   Pastor-Satorras, R. and A. Vespignani, *Immunization of complex networks.* Physical Review E, 2002. **65**(3): p. 036104.

[13]   Masuda, N., *Immunization of networks with community structure.* New Journal of Physics, 2009. **11**(12): p. 123018.

[14]   Hébert-Dufresne, L., et al., *Global efficiency of local immunization on complex networks.* Scientific reports, 2013. **3**.

[15] Hadidjojo, J. and S.A. Cheong, *Equal graph partitioning on estimated infection network as an effective epidemic mitigation measure.* PloS one, 2011. **6**(7): p. e22124.

[16] Borgatti, S.P., *Centrality and network flow.* Social networks, 2005. **27**(1): p. 55-71.

[17] Borgatti, S.P., *Identifying sets of key players in a social network.* Computational & Mathematical Organization Theory, 2006. **12**(1): p. 21-34.

[18] Palla, G., et al., *Uncovering the overlapping community structure of complex networks in nature and society.* Nature, 2005. **435**(7043): p. 814-818.

[19] Newman, M.E. and M. Girvan, *Finding and evaluating community structure in networks.* Physical review E, 2004. **69**(2): p. 026113.

[20] Girvan, M. and M.E. Newman, *Community structure in social and biological networks.* Proceedings of the National Academy of Sciences, 2002. **99**(12): p. 7821-7826.

[21] Fortunato, S., *Community detection in graphs.* Physics Reports, 2010. **486**(3): p. 75-174.

[22] Newman, M.E., *Modularity and community structure in networks.* Proceedings of the National Academy of Sciences, 2006. **103**(23): p. 8577-8582.

[23] Yoshida, T. and Y. Yamada, *Community structure based node scores for network immunization*, in *PRICAI 2012: Trends in Artificial Intelligence.* 2012, Springer. p. 899-902.

[24] Yamada, Y. and T. Yoshida, *A comparative study of community structure based node scores for network immunization*, in *Active Media Technology.* 2012, Springer. p. 328-337.

[25] Newman, M., *Community detection and graph partitioning.* EPL (Europhysics Letters), 2013. **103**(2): p. 28003.

[26] Pothen, A., H.D. Simon, and K.-P. Liou, *Partitioning sparse matrices with eigenvectors of graphs.* SIAM Journal on Matrix Analysis and Applications, 1990. **11**(3): p. 430-452.

[27] Lusseau, D., et al., *The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations.* Behavioral Ecology and Sociobiology, 2003. **54**(4): p. 396-405.

[28] Watts, D.J. and S.H. Strogatz, *Collective dynamics of 'small-world' networks.* nature, 1998. **393**(6684): p. 440-442.