CONTROLLING SPREAD OF RUMOR USING NEIGHBOR CENTRALITY

Anoop Mehta†, Bhaskar Mukhoty‡, Ruchir Gupta§

Department of Computer Science Engineering
Indian Institute of Information Technology, Jabalpur, India

(Received July 19, 2016; revised version received September 7, 2016)

Social networks are collaboration of individual entities where propagation of information, disease or ideas could take place because of the interaction between the entities. In such type of networks, it is often observed that propagation starts from a few unknown nodes and spreads through the whole network. However, propagation of a malicious information or a disease is often not desired and tried to be restricted to as few nodes as possible, using some external interventions. Inoculation is one such strategy, where few nodes are barred from further communication in order to restrict the spread. Such inoculations are costly, as they could require actions such as vaccination or disruption of normal operation of a node. Hence, we have minimization objective on both the parameters the final number of affected nodes and the number of inoculations. For such purpose, centrality measures, that rank the nodes according to some importance, are often used in identifying the nodes to be inoculated. Such measures mostly require the topology of the whole network in order to compute the centrality of the nodes. We propose a new centrality measure that requires information only on the neighboring nodes and can work in a distributed fashion depending upon the local information, but can also be used in centralized global inoculations. We empirically show that the centrality outperforms the existing ones in minimizing the spread in both the strategies.

DOI:10.5506/APhysPolB.47.2325

1. Introduction

Epidemic data dissemination in social networks and its control have been extensively studied in recent past [1–3]. Compartmental models have considered the network as a collection of heterogeneous classes of nodes depending

† anoopmehta007@gmail.com
‡ bhaskar.mukhoty@gmail.com
§ ruchir.in@gmail.com

upon the behavior towards the propagation of information. The dynamics of information propagation is studied under fraction of nodes, transiting from one class to another. Generally, a viral information or rumor spreading in a network may need to be restricted by external interventions like barring certain nodes from further propagation. Such process of stopping nodes from propagating the information is known as inoculation of the node. As inoculations are expensive and disrupt the natural functioning of a node, it is imperative to seek central nodes in the network, which would be most effective to inoculate in order to restrict the spread to a minimum [4–6]. We propose a control strategy under a twofold objective. The first step is to restrict the rumor size to a minimum number of nodes and the second, to minimize the number of inoculations.

Out of several models that are available for information dissipation [7], one of the most popular is SIR model introduced for rumor dynamics by Daley and Kendal (DK) [8] and its variation of by Maki–Thompson (MK) [9]. In DK model, nodes are divided into three categories known as ignorants, spreaders and stifler. A node that is yet to receive the information is ignorant (I), while a rumor spreading node is termed as spreader (S). A spreader node when stops propagating the rumor becomes a stifler (R). The nodes interact among themselves using some probabilistic rules, parameterized by the following three factors. The spreading rate, λ, denoting the probability with which a spreader node propagates the information to an adjacent ignorant node, which then becomes a spreader. The stifling rate, σ, being the probability with which a spreader node becomes a stifler in contact with an adjacent spreader or stifler node. The model requires that the stifling rate σ applies to both of the nodes if they are spreader, in contrast to the MK model when a spreader node comes into contact with an adjacent spreader node, then only initiating spreader becomes a stifler. Nekovee *et al.* [10] introduced the probability termed as recovery rate, δ, with which a spreader may become stifler independently, irrespective of its neighbors.

With the above probabilistic rules of propagation and some fraction of randomly distributed informed nodes, the process of propagation takes place to model the natural phenomenon of information spreading. A control strategy might intervene in the process at the start by inoculating some fraction of nodes selected using a centrality measure. The control strategy is assumed to have no knowledge of the informed nodes, but decides on the nodes to be inoculated by using topological structure of the network. There are different kinds of centrality measures *viz.* degree [11], betweenness [12], structural [13], Katz [14], eigenvector centrality [15] *etc.* that rank the nodes according to certain kind of importance. Most of these requires structural topology of the whole network and works in a centralized manner to identify the nodes to be inoculated. Such a strategy which requires the whole topological information of the network is called global inoculation strategy, whereas a strategy which can work in distributed manner with information about small part of the network is called a local strategy. We propose a new centrality measure, called Neighbor Centrality (NC), which can work in both global and local format. NC needs only the degree information of the neighboring nodes, so that it is locally computable, and can be implemented in a distributed manner, where each node in the network could carry out the process of inoculation. The global inoculation strategy is assumed to have the authority to inoculate a fraction of the nodes in the network in the initial round.

In the following section, we introduce some of the important centrality measures in use. In Section 3, we introduce the new centrality measure NC which turns out to be very effective for the purpose of inoculation. Section 4 presents the global and local inoculation method implemented with it. Experimental evaluations are conducted in Section 5 in order to measure the relative performance of the proposed centrality with respect to the existing ones. Section 6 concludes the paper.

2. Related work

In rumor dynamics, inoculation strategy is a way in which some nodes are barred from communication in order to control the spread of information in the network. Various inoculation strategies exist based on random and targeted inoculation. In random inoculation, a fraction of nodes are inoculated randomly, but in targeted inoculation, nodes may be selected using different centrality measures. Centrality measures rank the nodes in a network according to some kind of importance. As the nature of the application varies, the notion of importance also varies. A centrality measure suitable for one application might be inappropriate for the other. In the context of controlling the rumor spreading in a network, we consider a node to be central depending upon whether its inoculation restricts the final number of informed nodes or rumor size. Here, we discuss a few of the measures in this context.

Degree centrality of a node in an undirected graph is the number of its adjacent nodes [11]. The node with higher degree is considered to be central. In the rumor spreading model, a high degree node is important because it has the potential to spread information to many neighbors when it becomes a spreader. In the degree centrality based inoculation, the fraction of nodes with a higher degree will be inoculated to control information spreading. We will show by an example that it is not always optimal to remove the highest degree node in order to restrict the spread.

The betweenness centrality [12] considers a node to be central if many shortest paths go through the node. We take the ratios of the number shortest paths that pass through node i with the number of all possible shortest paths between the nodes s and t. We get the betweenness centrality of node i by adding up all such ratios obtained from the possible pairs of nodes s and t, other than i. If a particular node in the network is located such that most of the nodes pass information through it, then it has the potential to control the connectivity. Thus, in betweenness centrality a node which lies in between the larger number of shortest paths is considered as a central node. The main drawback of this centrality is its higher computational complexity incurred on tracing all pairs of shortest paths.

The structural centrality measures which node is close to the center according to the structure of the network in the Laplacian domain [13]. It is defined using the pseudo-inverse of the Laplacian matrix [16]. The reciprocal of the diagonal elements l_{ii}^{+} of the pseudo-inverse matrix L^{+} gives the centrality. The l_{ii}^{+} shows the squared distance of node i from the origin in the network in that domain. A node i with a smaller value of l_{ii}^{+} lies close to the center of the network and hence the value of the structural centrality of a node i is higher and *vice versa*. Therefore, for node i, $1/l_{ii}^{+}$ is considered as the structural centrality. This centrality is not viable for a large network due to the high computational complexity and requires global information which is not optimal for the large network.

The Katz centrality [14] of node i counts the number of distinct walks that can emerge from node i discounting importance of a walk as the path length increases. Suppose $A_{n\times n}$ is the adjacency matrix of a network with $A(i,j)=1$ if there is an edge from node i to j, otherwise $A(i,j)=0$. Then, $A^2(i,j)$ would represent the number of paths of length 2 from node i to j. The Katz centrality of node i sums the number of paths to all nodes j in the network, but discounting paths of length k, with α^k, where α is a positive quantity less than 1. Thus, the Katz centrality of a node i is given by

$$C_{\mathrm{Katz}}(i)=\sum_{k=1}^{\infty}\sum_{j=1}^{n}\alpha^k A^k(i,j)\,.$$

The eigenvector centrality of a node is defined on the principle that centrality of a node should be defined by the cumulative centralities of its neighbors [15]. It is an extension of the degree centrality. In the degree centrality, all the neighbor vertices are considered equivalent which is not the case because some of them are more connected to other nodes. In the eigenvector centrality, a node is important if it is linked to other important nodes in the network. This apparent cyclic definition can be resolved using eigenvector of the adjacency matrix [17] defined as

$$Ax=\lambda x\,,$$

where λ is the largest constant satisfying the criteria. The vector x then gives the centralities of the nodes. All entries of x are guaranteed to be non-negative in this case, λ is the largest such value, also known as the principle eigenvalue.

All the above centralities have been used to inoculate nodes, but as they are not defined for the specific purpose of inoculation, they could be outperformed by a centrality designed for it. In the next section, we propose the new centrality.

3. Proposed neighbor centrality

The proposed centrality method can be seen as a modification of the degree centrality. Suppose two nodes in the network have equal degree, with different average degree of the neighbors. In this case, the node with higher average degree of neighbors has a greater chance that its neighbors gets the rumor due to the higher number of connections they have in the network. As our purpose is to isolate the nodes, so that they could not get affected by the rumor, it would be more effective to inoculate the node with lower average degree neighbors. Hence, we take the centrality to be inversely proportional to the average degree of neighbors. Also as before, the centrality should be directly proportional to the degree of the node as a higher degree would signify more connections to spread the rumor. Hence, we propose the centrality as the ratio of the degree of a node and the average degree of its neighbors

$$C_{\mathrm{N}}(i) = \frac{d_i}{\left(\sum_{j \in N_i} d_j\right) / d_i} . \tag{1}$$

Here, d_i is the degree of node i, N_i is the set of the neighbors of node i and $C_{\mathrm{N}}(i)$ denotes the neighbor centrality score. A higher score would signify that it is more effective to inoculate the node in order to restrict the rumor spreading.

As per example, in Fig. 1, both nodes 2 and 10 have the degree 5. It would be more effective to inoculate node 10 rather than inoculating node 2, as node 10 would isolate all its neighbors from getting information, which is not the case for node 2, whose neighbors has further connections to get the rumor. Later, using correlation matrix we show empirically that NC differs significantly from the other centralities including the degree centrality.

If $A_{n\times n}$ is the adjacency matrix of a network and $e_{n\times 1}$ is a vector of all ones, then $(Ae)_{n\times 1}$ gives the degree of the nodes. $e_i^T(Ae)$ gives the degree of the i^{th} node, where $(e_i)_{n\times 1}$ is a vector with unity in the i^{th} position and zero in all others. Now, $(A^2e)_{n\times 1}$ is the vector with the sum of the degree of the neighboring nodes for each node and $e_i^T A^2 e$ is the same for i^{th} node. Thus,

$$C_{\text{N}}(i) = \frac{\left[e_i^T (Ae)\right]^2}{e_i^T A^2 e}\,. \tag{2}$$

This would help to compute the centralities globally using matrix operations.

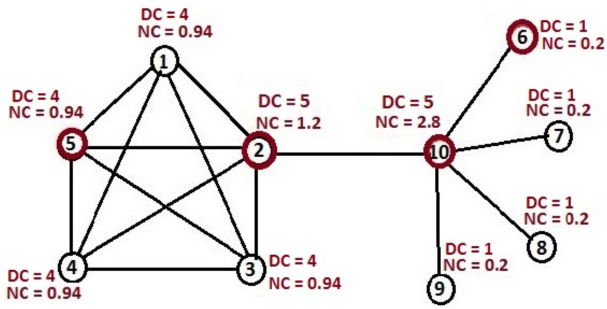

Fig. 1. Example showing effectiveness of NC.

4. Inoculation strategies

The inoculation strategy is a way in which nodes that going to be inoculated are chosen. The method may be random, where we randomly select a given fraction of nodes without any information of the network and inoculate them — it is called random inoculation [18, 19]. The random inoculation strategy performs very well in homogeneous network, because there is no large variations in the degree distribution of the network. The random inoculation however, performs badly in heterogeneous network due to large degree variation. It gives the same importance to all the nodes and does not choose the influential nodes in the network. The heterogeneous networks are strongly affected by targeted inoculation, where we find the influential nodes in the network using different centrality measures and inoculate them [20]. Below, there are the two main strategies in which targeted inoculation is done, based on the information about the influential nodes in the network.

4.1. Global strategy

A global strategy lists all nodes in the network according to a centrality measure, and inoculates a given fraction of nodes with top ranks. The inoculation takes place in the first round and the process stops when there is no spreader node left in the network. The stopping of the process is guaranteed due to the recovery rate, δ. With this probability, a spreader becomes a stifler at each round. Thus, after r rounds of propagation, the probability that a spreader still remains a spreader is $(1-\delta)^r$, which exponentially decreases to zero with r.

4.2. Local strategy

In contrast to the global strategy, a local strategy works in a distributed manner from individual nodes with co-operation of its neighbors. Not all centrality measures are locally computable. The neighbor centrality needs information only on the neighbors, so local strategies can be implemented with it. Local strategies have already been used for inoculation where a random node from neighbors is inoculated. Since our centrality is locally computable, we can use targeted inoculation instead of random acquaintances inoculation [21]. Below, we give a simple local strategy for inoculation.

Algorithm 1 *LocalStrategy*

Input: Degree of the neighbors of node i, probability p
  $\{rand$ is a random number in $[0,1]\}$
if $rand < p$ **then**
  for all nodes $j \in N_i$ **do**
    $C_{\mathrm{N}}(j) = \frac{d_j^2}{(\sum_{k \in N_j} d_k)}$
  end for
  select $s = \operatorname{argmax}_{j \in N_i}(C_{\mathrm{N}}(j))$
  inoculate s
end if

Assume each node in the network has the capability to inoculate its neighbors and individual nodes participate in the inoculation process with a probability p. Parameter p would control the fraction of inoculations that takes place in the network on an average. A node that is participating in the process computes the neighbor centrality of its neighboring nodes, and then removes the node with the highest score. This happens only in the first round, as it is the case with global strategies. The process stops when there is no more spreader node left in the network. Algorithm 1 lists the inoculation procedure as performed by the nodes of the network, individually. It requires the degree of its neighbors and the probability p as input.

As an example, in Fig. 1, we have shown a network of 10 nodes, with their corresponding NC score. The spreader nodes 2, 5, 6 and 10 are shown as thick/red circles. According to the local strategy, let randomly selected nodes with probability p be 4 and 6. Nodes 2 and 10 have respectively the highest NC score, as the neighbors of 4 and 6. Hence, nodes 2 and 10 would be inoculated in the first round. Here, it bears no significance if the node to be inoculated is a spreader or not.

4.3. Complexity analysis

It can be followed from Eq. (2) that the centrality of all the nodes can be computed in linear time once we have vectors $A(Ae)$ and Ae. Thus, the centrality could be found in $O(n^2)$ time as required by multiplication of a matrix to a vector. But if we perform the computation on every node and the purpose is to find the centralities of the neighbors instead of the whole network, then a node can compute the centrality of its neighbors in $O(n)$ time, thus greatly reducing the computation time.

5. Experimental results

5.1. Datasets

In the real world, most of the large scale networks follow the power law degree distribution. Scale free networks have many features of real world networks [22]. These types of networks have more nodes with low degree and less nodes with high degree. For experimental validation of the centrality, scale-free networks are generated using configuration model [23] with power law degree distribution $P(k) \propto k^{-\gamma}$, where $2 < \gamma \leq 3$. To keep diversity in the networks, we have taken $\gamma = \{2.5, 3\}$. 5 random networks are generated from two of the distributions. The final results are obtained by taking a mean on the set of 5 networks for each distribution. Every network has 5000 nodes. Below, we describe the SIR model of rumor dynamics taken for simulating the spread of rumor.

5.2. SIR model

For performance measure and comparison with existing methods, we have taken MK model [9] of rumor dynamics along with additions from Nekovee *et al.* [10]. Below, we show the interactions between the nodes in form of equations:

$$S + I \xrightarrow{\lambda} S + S\,, \tag{3}$$

$$S + S \xrightarrow{\sigma} R + S\,, \tag{4}$$

$$S + R \xrightarrow{\sigma} R + R\,, \tag{5}$$

$$S \xrightarrow{\delta} R\,. \tag{6}$$

The process starts by taking 0.1 fractions of the nodes being spreaders that are randomly scattered in the network and the remaining nodes are ignorant. In each round of discrete time, the following activities take place. A spreader node tries to send the information to each of its ignorant neighbors with the probability λ, as given in Eq. (3). Each of the spreader or stifler node tries to convert its neighbors that are spreader to a stifler with the probability σ, as shown in Eqs. (4) and (5). A spreader node also becomes a stifler independently with probability δ, given by Eq. (6).

The values of λ, σ and δ have been taken from previous research as 0.2, 0.1 and 0.1, respectively. The experiment is assumed to be completed when there is no spreader node left in the network.

5.3. Performance metric

In order to measure the performance of a centrality used in inoculation, the final number of informed node or the rumor size should act as the determining parameter. It is evaluated with different fraction of nodes being inoculated according to the centrality measure of nodes. Evidently, the smaller the final rumor size, the better is centrality used for inoculation.

We have taken three other performance metrics in order to comprehend the overall effect of the inoculation process on a network. They are namely the size of the giant component, a total number of components and a number of edges removed for a given fraction of inoculation. All these parameters are reported by taking the mean on a set of 5 networks.

The giant component is the largest connected component in an undirected graph [18]. If removal of a node reduces the size of the giant component significantly, then it should be understood that the node is important for robustness of the network. The same could be claimed when removal of a few nodes breaks the networks in more number of components. That is connectivity of the network breakdowns rapidly.

5.4. Results

Figures 2 and 3 show the performance of the proposed centrality with respect to other centralities as is given in Section 2. The results show that the final rumor size is best reduced when the inoculation is done using the neighbor centrality. This performance is followed by inoculation using the degree and betweeness centrality. It can be observed that relative performance of the centralities has not changed with the distribution parameter $\gamma = 2.5$ and 3, which gives greater confidence on the success of the proposed centrality in other distributions.

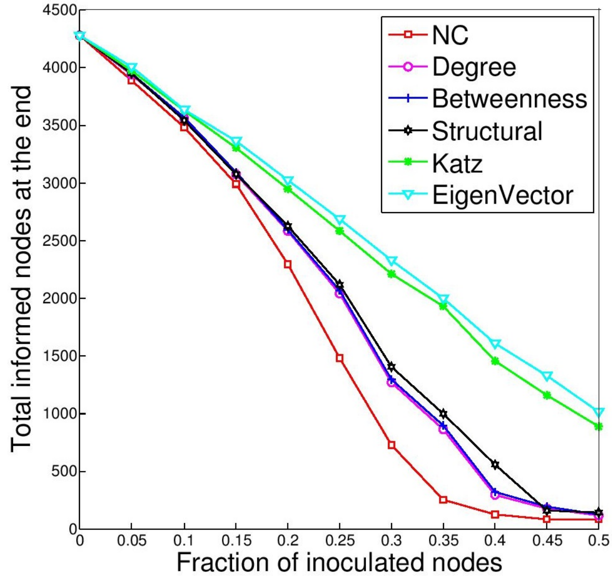

Fig. 2. Rumor spread with global inoculations in networks from $\gamma = 3$.

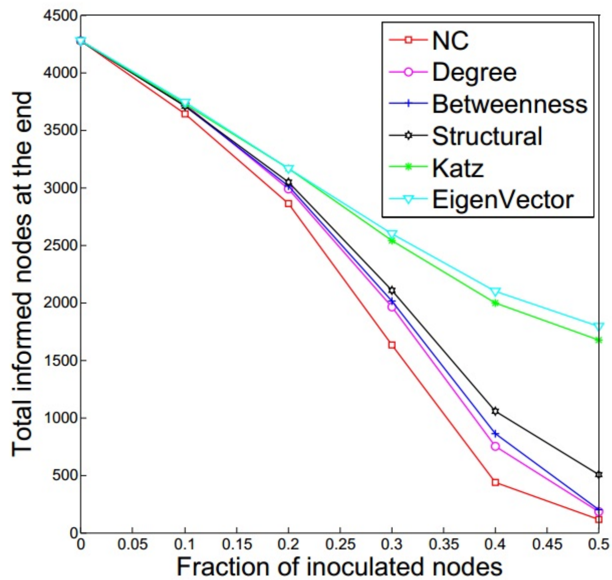

Fig. 3. Rumor spread with global inoculations in networks from $\gamma = 2.5$.

Figure 4 shows the performance of the neighbor centrality when used in the local strategy. In a local strategy, we cannot remove the top fraction of the nodes due to the required global comparison. Instead, a node participates in inoculation with some probability, which determines the number of inoculated nodes on average, as each node participating in the inoculation removes only one of its neighbors that has the highest NC score. We performed the experiment with probability varying from 0 to 0.5. As centralities that require the global information cannot be used in the local inoculation method, the comparison could only be done with the random and degree centrality. The figure depicts that the proposed centrality outperforms others in $\gamma = 3$.

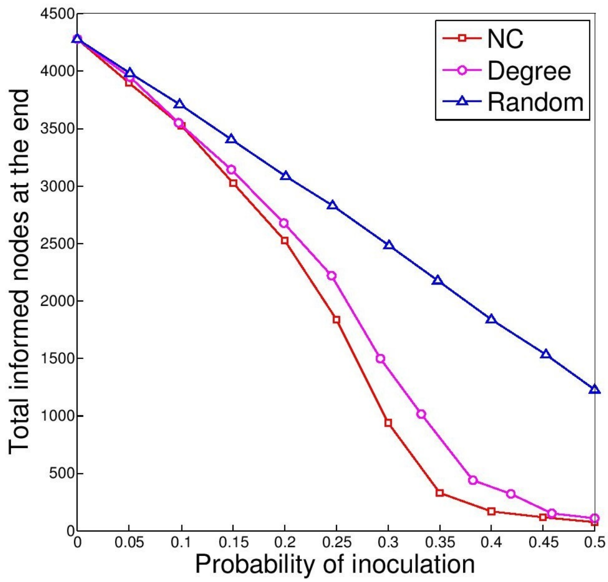

Fig. 4. Rumor spread with local inoculations in networks from $\gamma = 3$.

In Fig. 5, we plot the size of the giant component along with the fraction of inoculation. It shows that the size of the giant component stays the same for all centralities till 30 percent inoculation and decreases rapidly when the removal of nodes is done using neighbor centrality. Figure 6 shows that till 30 percent inoculation, number of components remains comparable across centralities and after that the network breaks quickly in smaller components when NC is used for inoculation.

Next, in Fig. 7, we show the number of edges that get removed as an effect of the inoculation performed on the network. It can be seen that NC removes the maximum number of edges for a given percentage of inoculation, followed by degree and betweenness.

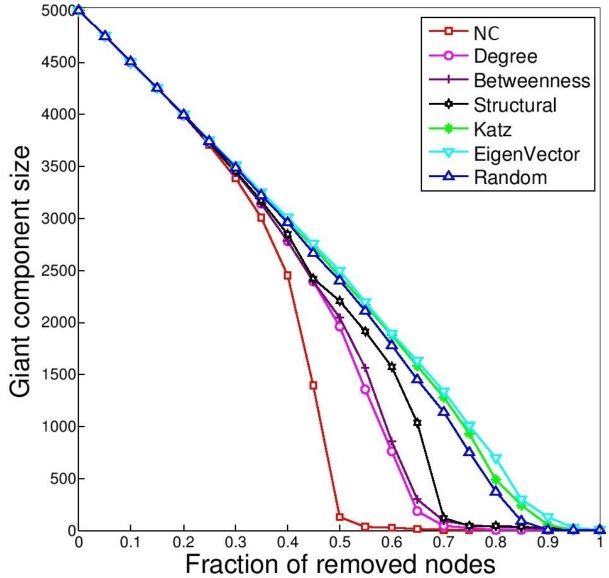

Fig. 5. Size of giant component *vs.* fraction of removed nodes, $\gamma = 3$.

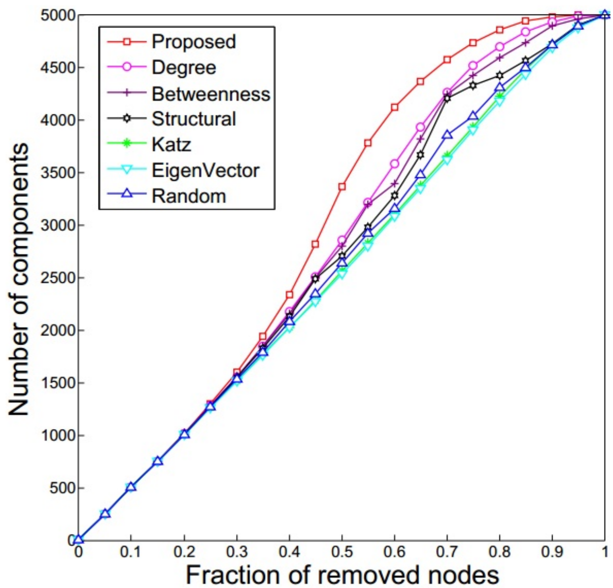

Fig. 6. Number of components *vs.* fraction of removed nodes, $\gamma = 3$.

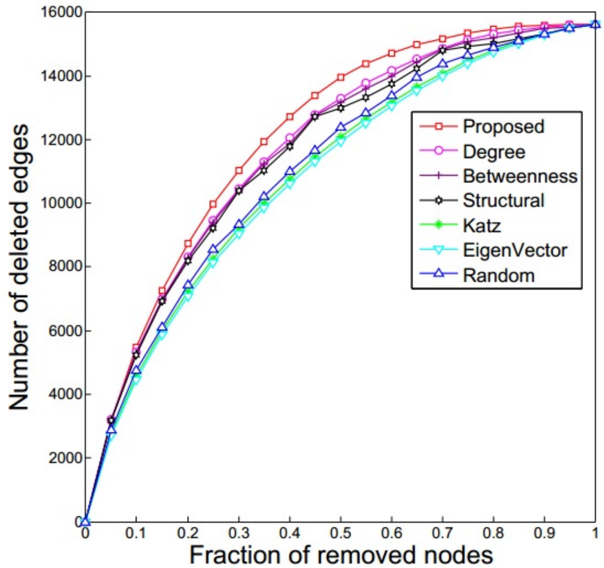

Fig. 7. Number of deleted edges *vs.* fraction of removed nodes, $\gamma = 3$.

Finally, Table I gives the Pearson correlation [24] matrix of the centrality measures obtained on these networks. It shows that neighbor centrality significantly differs from other centralities. The degree and betweenness centrality have the highest correlation, which explains their relatively the same performance in controlling rumor spreading. Similarly, the high correlation of eigenvector and Katz centrality explains their close performance.

TABLE I

Correlation of the centralities.

Centrality	NC	DC	BC	SC	KC	EC
Neighbor	1	0.8168	0.7865	0.7680	0.3982	0.3582
Degree	0.8168	1	0.9978	0.9964	0.7856	0.7455
Betweenness	0.7865	0.9978	1	0.9671	0.7984	0.7934
Structural	0.7680	0.9964	0.9671	1	0.8174	0.7785
Katz	0.3982	0.7856	0.7984	0.8174	1	0.9976
Eigenvector	0.3582	0.7455	0.7934	0.7785	0.9976	1

6. Conclusion

We have proposed a new centrality measure which can work with both centralized and distributed inoculation strategies and shown its effectiveness in restricting the final rumor size in both strategies. Comparisons have been made with the other centrality measures to demonstrate the effectiveness of the method and it can be seen that the proposed method outperforms the existing methods. As the centrality is locally computable, it can also be used in the distributed inoculation, where each node has the permission to carry out the inoculation process on its neighboring nodes. This work can be extended for directed or wighted network. We have also considered rumor spreading rate to be constant — this can be modeled as a variable entity which depends upon time.

REFERENCES

[1] Y. Moreno, M. Nekovee, A.F. Pacheco, *Phys. Rev. E* **69**, 066130 (2004).

[2] H.-F. Zhang, K.-Z. Li, X.-C. Fu, B.-H. Wang, *Chin. Phys. Lett.* **26**, 068901 (2009).

[3] Y. Moreno, R. Pastor-Satorras, A. Vespignani, *Eur. Phys. J. B* **26**, 521 (2002).

[4] A. Singh, Y.N. Singh, Rumor dynamics with inoculations for correlated scale free networks in: Proceedings of 2013 National Conference on Communications (NCC), pp. 1–5, DOI: 10.1109/NCC.2013.6488047.

[5] A. Singh, Y.N. Singh, in: *Complex Networks and their Applications*, Ed. H. Cherifi, Cambridge Scholars Publishing, 2014, pp. 132–173.

[6] M. Kitsak *et al.*, *Nature Phys.* **6**, 888 (2010).

[7] D.J. Daley, J. Gani, J.M. Gani, *Epidemic Modelling: an Introduction*, volume 15, Cambridge University Press, 2001.

[8] D.J. Daley, D.G. Kendall, *IMA J. Appl. Math.* **1**, 42 (1965).

[9] D.P. Maki, M. Thompson, *Mathematical Models and Applications, with Emphasis on the Social, Life and Management Sciences*, Englewood Cliffs, N.J.: Prentice-Hall, 1973.

[10] M. Nekovee, Y. Moreno, G. Bianconi, M. Marsili, *Physica A* **374**, 457 (2007).

[11] L.C. Freeman, *Social Networks* **1**, 215 (1978).

[12] L.C. Freeman, *Sociometry* **40**, 35 (1977).

[13] A. Singh, R. Kumar, Y.N. Singh, *Acta Phys. Pol. B* **46**, 305 (2015).

[14] L. Katz, *Psychometrika* **18**, 39 (1953).

[15] P. Bonacich, *Social networks* **29**, 555 (2007).

[16] J.C.A. Barata, M.S. Hussein, *Braz. J. Phys.* **42**, 146 (2012).

[17] P.D. Straffin, *Math. Mag.* **53**, 269 (1980).

[18] R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).

[19] R. Pastor-Satorras, A. Vespignani, *Handbook of Graphs and Networks: from the Genome to the Internet*, pp. 111–130, 2005.

[20] R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001).

[21] R. Cohen, S. Havlin, D. Ben-Avraham, *Phys. Rev. Lett.* **91**, 247901 (2003).

[22] S. Bornholdt, H.G. Schuster, *Handbook of Graphs and Networks*, Wiley Online Library, 2002.

[23] M.E.J. Newman, *SIAM Rev.* **45**, 167 (2003).

[24] K. Pearson, *Proc. R. Soc. Lond.* **58**, 240 (1895).