IMPACT OF STRUCTURAL CENTRALITY BASED ATTACKS IN COMPLEX NETWORKS

Anurag Singh

Department of Computer Science and Engineering
National Institute of Technology Delhi, Delhi, India

Rahul Kumar, Yatindra Nath Singh

Department of Electrical Engineering, IIT Kanpur, India

(Received November 7, 2014; revised version received January 26, 2015)

In this paper, we study a new strategy to find the influential nodes in the complex networks. This strategy is based on Structural Centrality (SC) of the node in the network. In this strategy, by using graph spectral analysis of the network, we find the hierarchy of the influential nodes in the form of central nodes in the network. The structural centrality of each node is ranked in the topology of complex networks which are modeled as the scale free networks. We have explored the structural centrality based targeted attack and compared our result with the degree based targeted attack. The robustness of the real world complex network has been measured efficiently against the degree, structural centrality based targeted attack and compared with the random attack and compared it. In the social networks, the mechanism to suppress the harmful rumors is of great importance. A rumor spreading model has been defined using the susceptible-infected-refractory (SIR) model to characterize the rumor propagation in the social networks. Inoculation strategy based on the structural centrality has been applied on the rumor spreading model for the heterogeneous networks. It is compared with the random and degree based targeted inoculations. The nodes with higher structural centrality are chosen for the inoculation in the proposed strategy. The structural centrality based targeted inoculation strategy is found to be more efficient in comparison to the random and degree based targeted inoculation strategies. One of the bottlenecks of this approach is the high complexity in computing the structural centrality of the nodes in the complex networks with very large number of nodes. Further, appearance of giant component has been studied in the network with random attacks, and degree and structural centrality based attacks. The proposed hypothesis has been verified using simulation results for e-mail network data and also for the generated scale free networks.

DOI:10.5506/APhysPolB.46.305
PACS numbers: 89.75.–k, 64.60.aq, 05.10.Gb, 89.75.Fb

1. Introduction

In today's world, Internet has become the most powerful medium to circulate the information. We use online social network sites to express our attitude, emotions and to communicate with friends, almost on daily basis. Twitter and Facebook have become the most important mechanisms for information broadcasting. A large number of users share information on them. Consequently, lot of research has been carried out to provide valuable insights in the information diffusion in social networks. It has been found that the topologies of many real world networks have three main properties: small world, scale free and high clustering.

If any information is circulated without officially publicized confirmation, it is called a rumor. In other words, rumors are unreliable information. The rumor spread phenomenon is similar to epidemic spread, in which all the informed nodes spread rumor by informing their neighboring nodes [1, 2]. Recent research in complex network theory has given a new direction to the epidemic spreading model [3, 4]. The susceptible-infected-refractory (SIR) model for dynamic process of epidemic spread is used to model the rumor spread in this paper. A susceptible node can be infected by an infected neighbor with some spreading rate and introduces a new refractory state in which nodes cannot be infected. The SIR model for rumor spreading, was first introduced by Daley and Kendal [5] and its variants by Maki–Thompson [6]. In Daley–Kendal (DK) model, homogeneous population is subdivided into three groups *viz.*, ignorants (I), spreaders (S) and stifler (R). The rumor is propagated throughout the population by pairwise contacts between spreaders and other individuals in the population. Any spreader involved in a pairwise meeting attempts to infect other individuals with the rumor. In Maki–Thompson (MK) model when a spreader contacts another spreader, only initiating spreader becomes a stifler. DK and MK models have an important shortcoming that these models do not take into account the topology of the underlying social interconnection networks along which the rumors spread. To consider the topology of network, the rumor spreading models on small world network and scale free (SF) networks [7, 8] have been defined. Some studies have been reported on how to stop the rumor spread [9–15] in small world and SF networks. These studies are more important since false and fatal rumors have negative impacts on the society especially during disasters.

In this paper, a strategy based on structural centrality is applied to find the most central nodes in the network. Using this method, the hierarchy of nodes can be made, according to their structural centrality, to find the most influential nodes. The robustness of real world complex networks has been studied using the attack based on structural centrality. The inoculation of the nodes using this method has been used in the rumor spreading models in this paper. The results are compared with degree and structural centrality based targeted and random inoculations [16]. The structural centrality based inoculation will not be useful for a complex network with very large number of nodes because the complexity will be high in finding the structural centrality of the nodes. In real world networks, the scale free properties have been found *e.g.*, e-mail networks, Internet networks, telephone call graphs, *etc.* [4]. Thus, in this work, for all the simulations of the complex networks, the scale free property has been considered with power law degree distribution.

In order to study the stability of network, some proper stability metric need to be defined. To measure the network's failure tolerance, Albert *et al.* [17] studied that removing a fraction of node may causes a change in diameter, largest component size and average component size. Due to the absence of deleted fraction of nodes, there is an increase in distance between the remaining nodes and hence in the diameter of network due to reduced system interconnectedness [18]. When the node is removed from the network, it gets detached from the large component. We can use giant component GC (largest connected component) size as the stability metric [19]. This metric is based on the topological properties of the network. We can calculate the stability metric by visualizing the network break-down point using *percolation threshold* [20]. If we remove more fractions of nodes than the *percolation threshold* from the network, it results in large number of disconnected components and the giant component disappears. Below that threshold, there exists a giant component which spans the macroscopic part of network. Molloy *et al.* [21] theoretically showed the existence of giant component by the ratio $\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle}$, where $\langle k \rangle$ and $\langle k^2 \rangle$ are the first and second moments of degree distribution respectively. With the help of κ, we can find out the presence of giant component; if the value of $\kappa \geq 2$, it indicates the presence of giant component in the network. Paul *et al.* [22] obtained the percolation threshold from the simulations by removing a fraction of nodes and then visualizing the value of κ. When the ratio is $\kappa < 2$, a fraction of nodes is removed and recorded. In this paper, the percolation threshold has been calculated by random, degree based and structural centrality based attacks. The size of giant component has been calculated after every fraction of attack and the phase transition of giant component have been studied. It helps to know the exact fraction of nodes needed to be attacked to break the network into small components with no giant component.

2. Centrality

The node through which most of the nodes are connected is the most active node in the network. Centrality in the network is basically defined as the measure criteria of the node's importance w.r.t. to other nodes. The central node is the most important point of stability w.r.t. to other nodes and it is likely to be the main stem for the network connectivity. There are different types of measure by which we can categorize the centrality. Here, we discuss a few which are of importance. These measures have been discussed in the networks to explain the idea.

2.1. Degree centrality

The simplest and perhaps the most common notion is that the node centrality is a function of degree of the node. The degree of a node N_i can simply be understood by the number of other nodes $N_j (i \neq j)$ which are neighbors of this node. The nodes having higher degree are more strongly connected with the other nodes and they have more options to share the resources or to communicate with the neighbor nodes. Because of the higher degree, a node can act as an important communication link between two different nodes.

In social network, from the communication point of view, the person with the highest number of neighbors is in direct contact with many others. This person can be seen as the major source through which information spreads. On the other hand, a person with the less number of neighbors makes the person aloof from the communication process. Using this basis, Niemen has introduced the general measure of calculating the degree centrality [23] as

$$C_d(N_k) = \sum_{i=1}^{n} e(N_i, N_k)\,, \tag{1}$$

where $e(N_i, N_k) = 1$ if the node is connected by an edge between i and k, otherwise 0. Now, from the value of C_d for the node N_k, we can predict that if the value is large then this node is more likely to be degree central. As the value decreases, the centrality of the node in the network decreases and if the value equals zero than the node is isolated from the network.

2.2. Betweenness centrality

The second view point of centrality is betweenness centrality. Betweenness implies the node which is common in most of the paths connecting all possible pairs of nodes. This centrality was first introduced by Freeman [23]. He defined that if any node which is present in between the path connecting two nodes it has the potential to control the connectivity. Thus, node having this property for the maximum possible paths is defined as betweenness central.

In a complex network, there is a situation when some nodes are present on the connecting path and sometimes not if more than one path are there between two nodes, and one of the paths is chosen randomly. In that case, some sort of partial betweenness is defined in terms of probabilities. The node which is more common in the paths is more central than the other nodes present in the network. Let us assume that total number of communication paths between two nodes N_i and N_j is represented by t_{ij}, and the probability of using one out of these is

$$\frac{1}{t_{ij}}\,. \tag{2}$$

The centrality of node N_k, which lies on the connecting paths between nodes N_i and N_j, is defined as the probability that N_k falls on the randomly selected path between these two nodes. If $t_{ij}(N_k)$ is the number of paths between the nodes i and j that contains N_k, then $b_{ij}(N_k) = \frac{1}{t_{ij}} \times t_{ij}(N_k)\,.$ The $b_{ij}(N_k)$ is the probability that node k occurs on the connecting path between node i and j. To determine the overall centrality of a node N_k, we have to take sum of probabilities for all the pairs of nodes which contain node k on the path between these with the constraints that $i \neq j \neq k$

$$C_b(N_k) = \sum^{N} \sum_{i<j}^{N} b_{ij(N_k)}\,. \tag{3}$$

$C_b(N_k)$ is the measure of partial centrality for the node N_k. If N_k is the only node through which all possible path p_{ij} passes, then the value is 1, otherwise the value is less than 1.

3. Complex network topology using graph spectra

The complex network topology can be understood by the graph structure $G = (V, E)$ [23, 24]. Here V is the set of vertices or nodes and E is the set of edges or links. In the graph structure form, a network can be represented by the symmetric adjacency matrix. $A = [a_{ij}]$ of $|N \times N|$ size in which $a_{ij} = 1$ if edge is present, and $a_{ij} = a_{ji}$. A diagonal matrix, $D = [d_i]_{N\times N}$, where $d_i = \sum_j a_{ij}$ is the degree of the i^{th} node. A Laplacian matrix L of a graph is given by $L = D - A$.

Spectral graph theory using eigenvalues and eigenvectors can be applied in the graphs to find out the structural centrality of the networks. If a matrix is square, symmetric and positive semidefinite [25] then eigenvectors and eigenvalues will exist for the matrix. Thus, eigenvectors and eigenvalues exist for A, since the adjacency matrix A of a graph is square, symmetric and positive semidefinite. The Laplacian matrix L is [25]:

— Symmetric, thus N real eigenvalues, and real eigenvectors form orthonormal basis.

— Positive semidefinite, thus eigenvalues are non-negative.

— Doubly centred , thus the centroid of the position vectors for the set of nodes lies at the origin of this n-dimensional space.

The Laplacian matrix L of the network is also square, symmetric and positive semidefinite, therefore, it has all eigenvalues, *i.e.* $\lambda_i \geq 0$, $\forall i$. These eigenvalues (λ_i) ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = 0^2$, with corresponding eigenvectors $\vec{z_i}$ such that $||\vec{z_i}||^2 = \vec{z_i}^T \vec{z_i} = 1$. The eigenvalues of L are positive with minimum eigenvalue is equal to 0. Hence, there is a trivial eigenvector $[1, 1, 1, \ldots, 1]$ with eigenvalue zero. In the undirected network, sum of row and column of Laplacian matrix, L, is zero. The set of eigenvectors of L, $\vec{Z} = [\vec{z_1}, \ldots \vec{z_i} \ldots, \vec{z_n}]$, will be orthonormal *i.e.*, $Z^T Z = I$. If Λ is a diagonal matrix, *i.e.* $\Lambda = [\lambda_{ii}]$, of eigenvalues, then L follows the eigen decomposition as $L = \vec{Z} \Lambda \vec{Z}^T$. The Laplacian matrix L is [25]:

— Symmetric → N real eigenvalues, and real eigenvectors form orthonormal basis.

— Positive semidefinite → non-negative eigenvalues.

— Doubly centred → the centroid of the position vectors for the set of nodes lies at the origin of this n-dimensional space.

4. Structural centrality

The Moore–Penrose pseudo inverse matrix L^+ that follows all the properties (square, symmetric, doublycentered, positive semidefinite) of L, can be defined, from the Laplace matrix. The eigen decomposition of L^+ will be $\vec{Z}^T \Lambda^{-1} \vec{Z}$. $\vec{Z}$ is an orthonormal matrix made of the eigenvectors of L^+. If Λ has an eigenvalue value, $\lambda_i = 0$ then corresponding eigenvalue λ^{-1} in Λ^{-1} will also be 0. As L^+ has the doubly centered (all rows and columns sum will be zero) property, therefore, centroid of the nodes (having position vectors) lies on the origin of the space [25]. The graph matrix maps into the new Euclidean space. We can represent each node by a unit vector $\vec{v}$ as

$$\begin{aligned} \vec{v_i} &= [0 - -- \underset{i}{1} - - - 0]^T\,, \\ \vec{v_j} &= [0 - -- \underset{j}{1} - - - 0]^T\,. \end{aligned}$$

Now, we can calculate the distance between node i and j in terms of number of hops required to reach j from i ($m(j|i)$) and *vice versa.* Average commute hop distance measure is

$$n(i,j) = m(j|i) + m(i|j)\,. \tag{4}$$

$n(i,j)$ will follow the following distance measures for any nodes i, j and k

1. $n(i,j) \geq 0$,
2. $n(i,j) = 0$ if $i = j$,
3. $n(i,j) = n(j,i)$, and
4. $n(i,j) \leq n(i,k) + n(k,j)$.

Therefore, using L^+ matrix and graph volume, $V_{\mathrm{G}}(= \sum_{k=1}^{n} d_{kk})$, $n(i,j)$ can be expressed as [25]

$$n(i,j) = V_{\mathrm{G}} \left(l^+_{ii} + l^+_{jj} - 2l^+_{ij} \right)\,. \tag{5}$$

The node vector $\vec{v_i}$ can be mapped into the new Euclidean space by using the following transformations

$$\vec{v_i} \;=\; \vec{Z}\vec{y_i}\,, \tag{6}$$

$$\vec{y_i}' \;=\; \Lambda_i^{-1/2}\vec{y_i}\,m\,, \tag{7}$$

where $\vec{y_i}$ is the transformation node vector. Then Eq. (5) can be decomposed as

$$\bar{n}(i,j) = V_{\mathrm{G}} \left(\vec{y_i}' - \vec{y_j}' \right)^T \left(\vec{y_i}' - \vec{y_j}' \right)\,. \tag{8}$$

L^+ contains the inner product of transformed vector $\vec{y}'_i$:

$$\begin{aligned} \vec{y}'^{T}_{i}\vec{y}'_{j} &= \left(\wedge_i^{-1/2}\vec{y_i} \right)^T \wedge_j^{-1/2}\vec{y_j} = \vec{y}_i^T \wedge^{-1} \vec{y_j} \\ &= \vec{v}_i^T \vec{Z} \wedge^{-1} \vec{Z}^T \vec{v_j} = \vec{v}_i^T L^+ \vec{v_j} = l^+_{ij}\,. \end{aligned}$$

The vectors $\vec{y}'_i$ are centred

$$\sum_{i=1}^{n} \vec{y}'_i = \wedge^{-1/2} \sum_{i=1}^{n} \vec{y}_i = \wedge^{-1/2} \vec{Z}^T \sum_{i=1}^{n} \vec{v}_i = \wedge^{-1/2} \vec{Z}^T \vec{v}$$

$$\text{using} \quad \wedge^{-1} = \vec{Z}^T L^+ \vec{Z} \quad \text{and} \quad \wedge^{-1/2}\, \vec{Z}^T = \wedge^{1/2} \vec{Z}^T L^+$$

$$\sum_{i=1}^{n} \vec{y}'_i = \left(\wedge^{1/2} \vec{Z}^T L^+\right) \vec{v} = 0 \quad \text{since} \quad L^+ \vec{v} = 0\,.$$

If $\vec{Y}'$ denotes data matrix containing $\vec{Y}' = [\vec{y}'_1, \vec{y}'_2, \vec{y}'_3, \vec{y}'_4, \vec{y}'_5, .., \vec{y}'_n]$ then we have $L^+ = \vec{Y}'(\vec{Y}')^T$ with $l_{ij} = \vec{y}'^T_i \vec{y}'_j$. Thus, taking into account the above fact and having $\vec{Z}$ as orthonormal basis, one can conclude that matrix $\vec{Y}'$ represents an embedding of network in a new n-dimensional Euclidean space. Hence, in the new Euclidean space, the node vector $\vec{y_i}$ and $\vec{y_j}$ are separated by average commute euclidean distance measure $(\bar{n}(i,j))$.

Therefore, the Euclidean distance measure for the node i from the origin can be found as the diagonal entry of the L^+

$$\left|\left|\vec{y'_i}\right|\right|_2^2 = l^+_{ii}\,. \tag{9}$$

Definition If L_e is the Laplacian of the graph on n vertices consisting of just the edge e and $\vec{w} \in \Re^n$ then

$$\vec{w}^T L \vec{w} = \sum_{e \in E} \vec{w}^T L_e \vec{w} = \sum_{(i,j)\in E} \left(\vec{w}_i - \vec{w}_j\right)^2\,. \tag{10}$$

Definition Structural centrality is able to make the hierarchy from the most influential nodes to least influential nodes.

The structural centrality of the node i for the graph G is

$$\mathrm{SC}(i) = \frac{1}{l^+_{ii}}\,. \tag{11}$$

From Eq. (11), for the lower value of l^+_{ii} the structural centrality (SC) will be high and *vice versa*. Therefore, the value of l^+_{ii} determines the influential nodes.

If a node i is closer to the origin in the n-dimensional space, then it will have lower value of l^+_{ii}, *i.e.*, more centrally located in the network. Therefore, the value of l^+_{ii} in the pseudo inverse matrix L^+ can be defined as

$$l^+_{ii} = \sum_{k=1}^{n-1} \frac{\vec{z}^{\,2}_k}{\lambda_k}\,. \tag{12}$$

It has been observed from Eq. (12) that structural centrality of a node is defined by the eigenvectors and eigenvalues of the Laplace matrix, L of the graph.

It has been defined that L^+_{ii}, the squared distance of each node i to the origin in the L^+ geometric embedding, provides a strong measure of the structural centrality of a node i in the network. In fact, the larger is L^+_{ii} (*i.e.* farther away it is from the origin of the embedding), the less structurally central is the node i in the network. In other words, node i with larger L^+_{ii} lies closer to the periphery of the network, whereas the node i with smaller L^+_{ii} lies closer to the center of the network. Therefore, for node i, $1/L^+_{ii}$ is referred as the structural centrality measure as in Eq. (11) (thus the larger is $1/L^+_{ii}$, the more structurally central is node i), whereas L^+_{ii} will be referred to as the random eccentricity metric. Intuitively, the structural centrality metric, $1/L^+_{ii}$, implies that if the network is attacked (inoculated) and divided in two parts, nodes i with larger structural centrality metrics are more likely to lie in the larger part of the remaining network after the attack.

The concept of structural centrality can be understood with the help of an example given in Fig. 1. There are seven nodes in the graph and the hierarchy of their degrees is given in the center of the nodes. Hence, node 5 is most influential in the case of targeted inoculation based on nodal degree as shown in Fig. 1 (a). After defining the adjacency matrix A and degree matrix D of the given graph, we can calculate the Laplace matrix $L = D - A$, as

$$L = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 3 & -1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 & 4 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & 2 \end{pmatrix}.$$

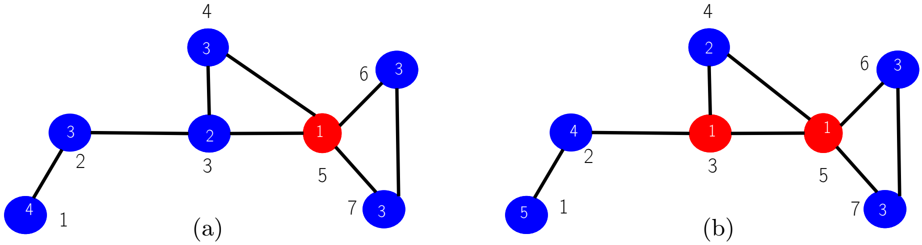


Fig. 1. The node ranks in graph with (a) degree centralities (b) structural centralities mentioned inside the nodes.

Laplace matrix, L, holds the desirable properties to calculate the structural centrality. It is:

1. Symmetric: $a_{ij} = a_{ji}$, in the given Laplacian, L.
2. Square: The size of given Laplacian matrix, L, is 7×7.
3. Doubly centered: Summation of all rows and columns in the given Laplacian matrix, L, are 0.
4. Positive semidefinite: Let $\vec{w}$ be any vector, *i.e.*, $\vec{w} = \begin{bmatrix} -0.8507 \\ -0.5257 \end{bmatrix}$, then $\vec{w}^T = \begin{bmatrix} -0.8507 & -0.5257 \end{bmatrix}$, for edge between node 1 and 2, $L_{12} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$, and

$$\vec{w}^T L_{12} \vec{w} = 0.3820\,. \tag{13}$$

Therefore, L will be positive semidefinite.

By given symmetric, square, doubly centred and positive semidefinite, Laplacian matrix, L, pseudo inverse matrix L^+ can be generated as

$$L^+ = \begin{pmatrix} 1.4626 & 0.6054 & -0.1088 & -0.3469 & -0.4422 & -0.5850 & -0.5850 \\ 0.6054 & 0.7483 & 0.0340 & -0.2041 & -0.2993 & -0.4422 & -0.4422 \\ -0.1088 & 0.0340 & 0.3197 & 0.0816 & -0.0136 & -0.1565 & -0.1565 \\ -0.3469 & -0.2041 & 0.0816 & 0.5102 & 0.0816 & -0.0612 & -0.0612 \\ -0.4422 & -0.2993 & -0.0136 & 0.0816 & 0.3197 & 0.1769 & 0.1769 \\ -0.5850 & -0.4422 & -0.1565 & -0.0612 & 0.1769 & 0.7007 & 0.3673 \\ -0.5850 & -0.4422 & -0.1565 & -0.0612 & 0.1769 & 0.3673 & 0.7007 \end{pmatrix}.$$

From the above matrix, diagonal value l_{ii}^+ is defined for i^{th} node. Thus, the vector for l_{ii}^+ $\forall i$ is

$$l_{ii}^+ = \begin{bmatrix} 1.462 & 0.7483 & 0.3197 & 0.5102 & 0.3197 & 0.7007 & 0.7007 \end{bmatrix}.$$

After observing the above values of l_{ii}^+, it has been found that nodes 3 and 5 with minimum l_{ii}^+, have the maximum structural centrality in the network. Therefore, node 3 can also be most influential like node 5 (*i.e.* most influential in degree centrality).

5. Structural centrality inoculations

The diagonal elements l_{ii}^+ can be sorted from low to high with their corresponding node ids. Now, we will be able to get the list of the nodes sorted according to their structural centralities from high to low from Eq. (11). Then, we can select fraction g of inoculated nodes from the sorted array. Consequently, we will be able to inoculate most structurally central nodes first. It is also considered as structural centrality based targeted attack.

6. Rumor spreading model

In this work, the modified SIR model for rumor spreading has been used, proposed by us in [8]. The mean field equations for complex networks has been used while considering non-linearly varying number of informed neighbor nodes by a spreader in each time step (not all neighbors of the node). It means that at a single time step a node may or may not inform their all neighbors. It depends on a parameter α, $0 \leq \alpha \leq 1$. This scenario can also be found in real life, where a person can share information only to some of his neighbors, not to all neighbors. $P(k) \propto k^{-\gamma}$ is the degree distribution of SF network and $\Phi(k) = k^{\alpha}$ is the non-linear rumor spreading function with $0 \leq \alpha \leq 1$. $P(l|k)$ is the degree–degree correlation function that a randomly chosen edge is emanating from a node of degree k leads to a node of degree l, and $P(l|k) = lP(l)/\langle k \rangle$, for uncorrelated networks, where $\langle k \rangle$ is the average degree of the network. Let $I(k,t), S(k,t), R(k,t)$ be the fraction of ignorants, spreaders and stifler nodes, respectively belonging to connectivity class k at time t. The rate equations for rumor diffusion model are

$$\frac{dI(k,t)}{dt} = -\frac{k\lambda I(k,t)}{\langle k \rangle} \sum_j j^{\alpha} P(j) S(j,t)\,, \tag{14}$$

$$\begin{aligned}\frac{dS(k,t)}{dt} &= \frac{k\lambda I(k,t)}{\langle k \rangle} \sum_j j^{\alpha} P(j) S(j,t) - \frac{k\sigma S(k,t)}{\langle k \rangle} \sum_j [S(j,t) + \\ &\quad R(j,t)] j^{\alpha} P(j) - \delta S(k,t)\,,\end{aligned} \tag{15}$$

$$\frac{dR(k,t)}{dt} = \frac{k\sigma S(k,t)}{\langle k \rangle} \sum_j [S(j,t) + R(j,t)] j^{\alpha} P(j) + \delta S(k,t)\,, \tag{16}$$

where, λ, σ and δ are the rumor spreading, stifling and forgetting rates respectively. After solving Eqs. (14)–(16) for $\delta = 1$, the rumor threshold (below this spreading rate rumor will not spread in the network) is $\lambda_c = \frac{\langle k \rangle}{\langle k^{\alpha+1} \rangle}$

7. Random inoculations

In random inoculation strategy, randomly selected node from the network are inoculated. This approach inoculates a fraction of nodes randomly, without any information about the network. Here, variable g $(0 \leq g \leq 1)$ defines the fraction of inoculated nodes. In the presence of random inoculation, rumor spreading rate λ is reduced by a factor $(1 - g)$.

8. Targeted inoculations

Scale free networks permit efficient strategies which depend upon the hierarchy of the degrees of nodes (or structural centrality). The SF networks are strongly affected by targeted inoculation of nodes [8]. In targeted inoculation, the high degree nodes (high degree centrality) are inoculated as they are more likely to spread the information. The robustness of SF networks decreases even with a tiny fraction of inoculated individuals.

Let us assume that the fraction g_k of nodes with degree k are successfully inoculated. An upper threshold of degree (structural centrality) is k_t , so that all nodes with degree (structural centrality) $k > k_t$ get inoculated ($g_k = 1$), fraction g_k of nodes with the degree (structural centrality) k are successfully inoculated. The fraction of inoculated nodes is given by

$$g_k = \begin{cases} 1, & k > k_t \,, \\ f, & k = k_t \,, \\ 0, & k < k_t \,, \end{cases} \tag{17}$$

where $0 < f \leq 1$.

9. Simulations and results

The results of simulations are shown for the simulated scale free network as well as for real world network data *e.g.* the e-mail communication network at the University Rovira i Virgili [26]. In the e-mail network, there are 1133 nodes and 10902 edges with maximum degree of 71. Here, all attack strategies are performed and compared to the robustness for e-mail network. The SF networks are generated according to the power law, $P(k) = k^{-\gamma}$, where $2 < \gamma \leq 3$ for $N = 5000$ and $\gamma = 2.3$ (Fig. 2 (a)). E-mail network has also been verified to be complex (Fig. 2 (b)). The random attack is implemented by selecting gN nodes randomly in the network. The targeted attack can

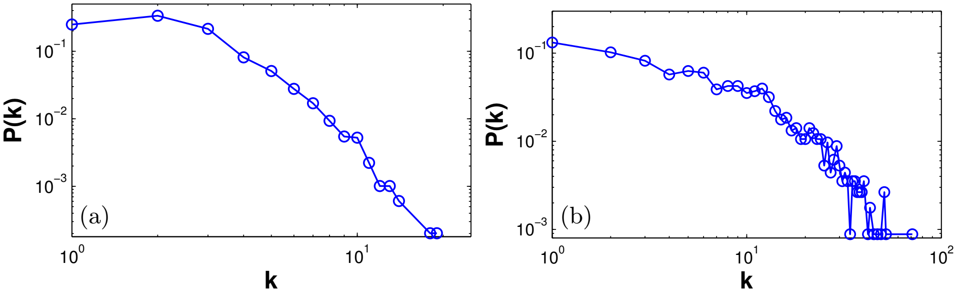


Fig. 2. The degree distributions of (a) generated SF network, (b) e-mail network.

be done after selection of the fraction of higher degree (structural centrality) nodes. The structural centrality inoculation can be done by getting the diagonal values l_{ii}^{+} of the pseudo inverse matrix L^{+}, for the corresponding node i. Using l_{ii}^{+}, we can sort the values in an array from low to high and inoculate fraction of the sorted nodes in the array. In structural centrality (SC) attack, we use the diagonal entries of matrix L^{+} which is l_{ii}^{+}. We delete the node in the hierarchy, in the order of node centrality (increasing order of algebraic value present at diagonal entries of matrix L^{+}).

In Fig. 3, the variation of the size of the giant component has been shown against the targeted based on degree and structural centrality and random attack. In the Fig. 3, the size of giant component is affected more by structural centrality than targeted. There is a range over which SC attack is showing much better results than the degree centrality based targeted attack. By deleting even a small fraction of more central connecting nodes, the size of giant component is reduced further. At higher fraction of deletion, structural centrality proves to be better than degree based targeted attack, because after deleting a large fraction, there remains a large number of small degree nodes as we can see in Fig. 9. Therefore, degree based targeted attack will delete all the nodes with the same probability. On the other hand, in the case of SC attack, the nodes which are more centrally placed in the network within the same degree of nodes are deleted. Hence, attack through SC chooses the nodes more efficiently even when less nodes of the same degree are present.

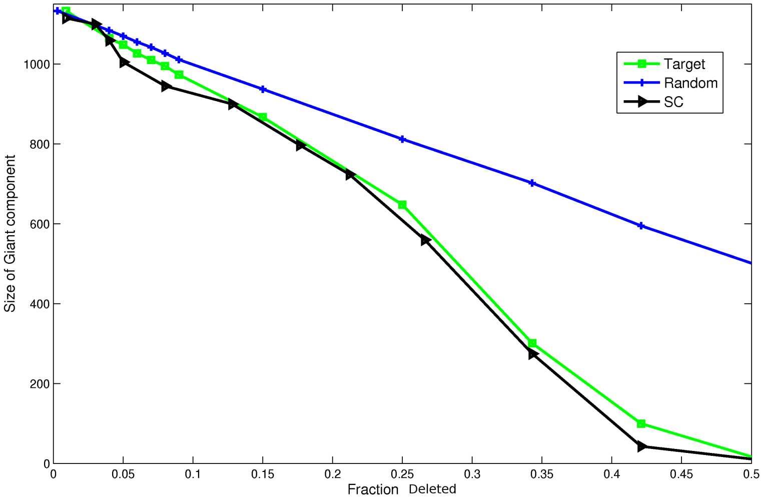


Fig. 3. Giant component against fraction of deleted nodes.

In Fig. 4, a variation of deleted edges against the fraction of deleted nodes is shown. It has been discussed above under what range structure centrality (SC) attack is showing better result. The same result has been observed in Fig. 4, in the case of edges deletion. At small fraction more edges are deleted, this is due to the presence of the large connected component getting attacked. Whereas, when the fraction is bigger, there is not much variation in the deleted edges. This is because at this stage only less connected components with smaller size having nodes with less degree are remaining in the network.

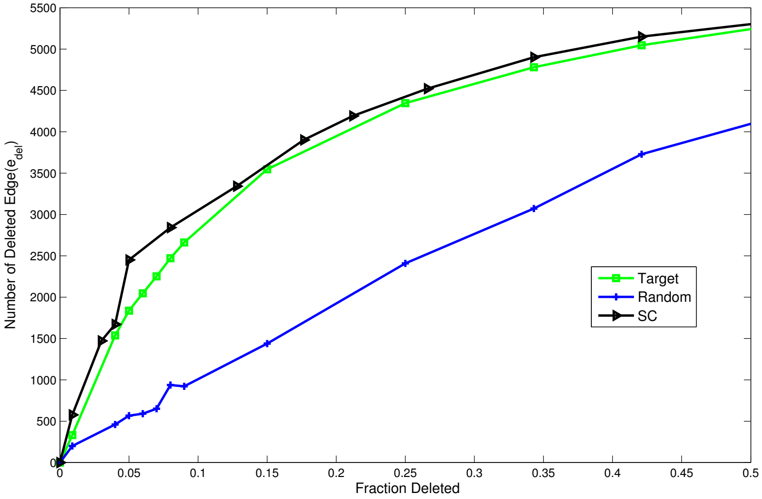


Fig. 4. Deleted edges against fraction of deleted nodes.

In Figs. 5–7, the upper figure presents the variation in size of giant component against deleted fraction of nodes using random, structural centrality and targeted attacks, and lower figure presents the value of κ with the same deleted fraction. This result (Fig. 5–7) shows that at fraction 0.5 the value of $\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} < 2$. If $\kappa < 2$, then giant component disappears from the network [20]. Hence, we have concluded that the e-mail network is more robust against the attack as compared to our simulated network with structural centrality based attack. The e-mail network looses robustness once 40% of the nodes have been attacked (Fig. 7), whereas with targeted attack, it happens at 50% (Fig. 6), and with random attacks, robustness is lost at 90% (Fig. 5).

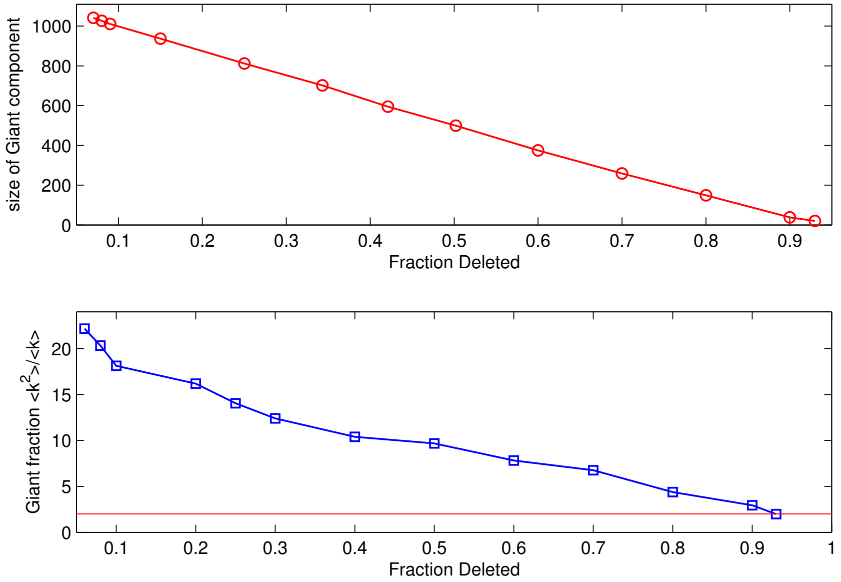


Fig. 5. Giant component in comparison with stability measure for e-mail network against random attack.

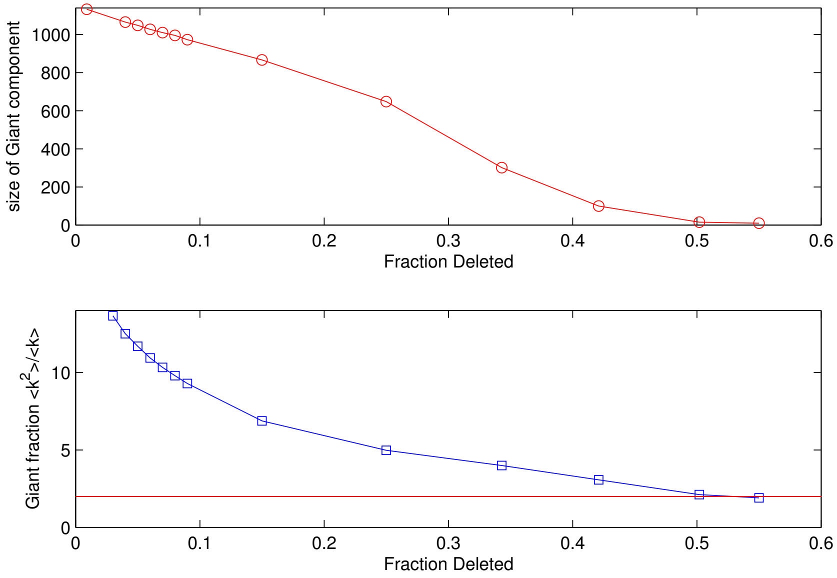


Fig. 6. Giant component in comparison with stability measure for e-mail network against targeted attack.

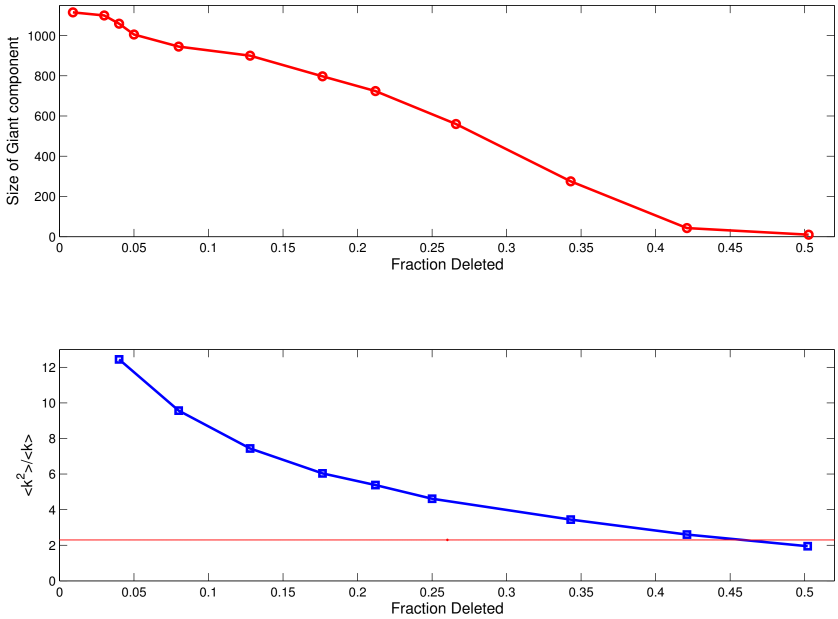
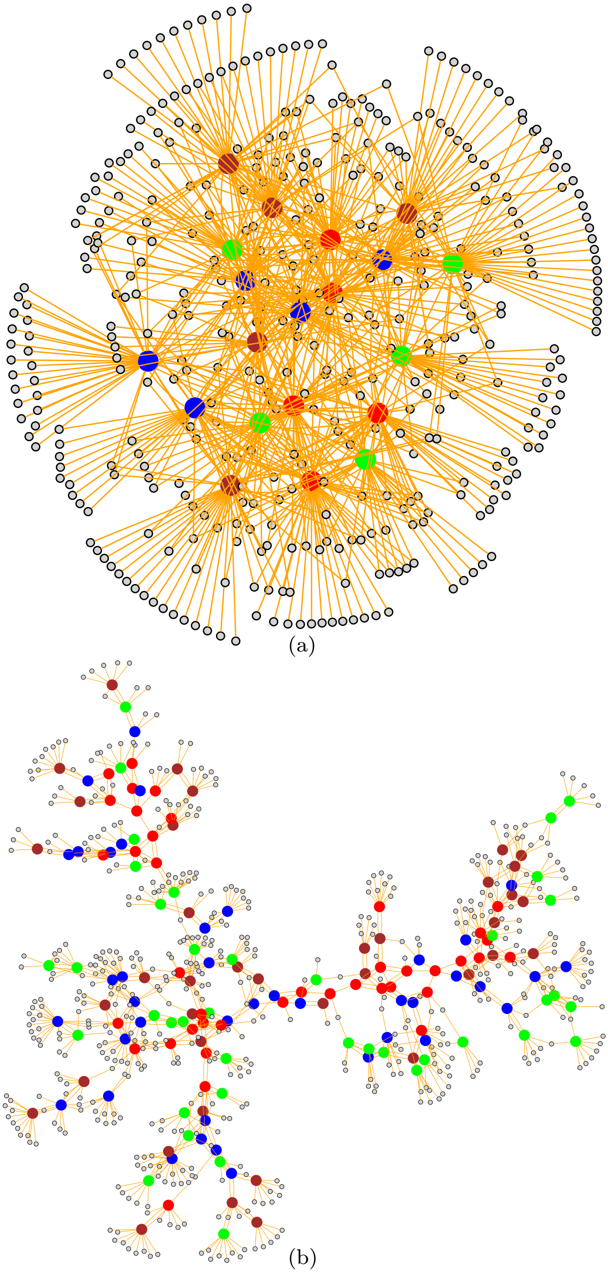


Fig. 7. Giant component in comparison with stability measure for e-mail network against Structural Centrality attack.

The structures of the e-mail and generated SF network are constructed for some nodes along with the structural centrality (Fig. 8). In the degree distribution of e-mail network more number of very high degree nodes are found as compared with the generated SF network, as shown in Fig. 2. Figure 8 (a) shows a lot of edges around more number of higher degree nodes as compared to the generated SF networks shown in Fig. 8 (b). For high structurally central node, less number of hops are required to reach the other nodes, even at lesser degree. The most structurally centered node provides the well connected path between the two dense nodes shown as a sub-graph.

In Fig. 9 (a) degree centrality has been mentioned for all the nodes in the decreasing order of degrees and corresponding node's structural centrality has also been shown in Fig. 9 (b) for e-mail networks. It is observed that even with lesser degree, for some nodes, the structural centrality is high, and can affect the network in the case of rumor spreading in comparison to the high degree nodes. Therefore, we observe influential nodes in the structural centrality. Hence, it is required to inoculate these nodes to suppress the rumor in the network.



(a)

(b)

Fig. 8. The structure of (a) e-mail network, (b) generated SF network with the diffrent ranking of structural centralities.

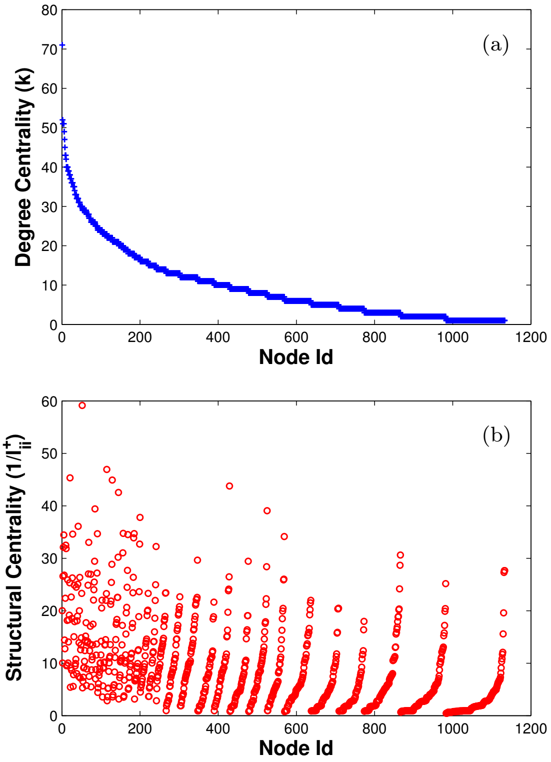
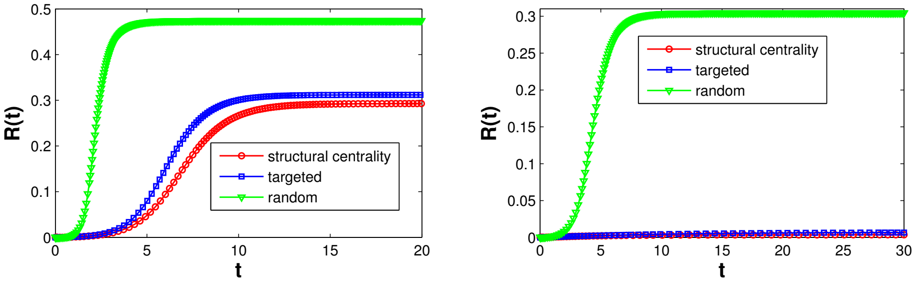


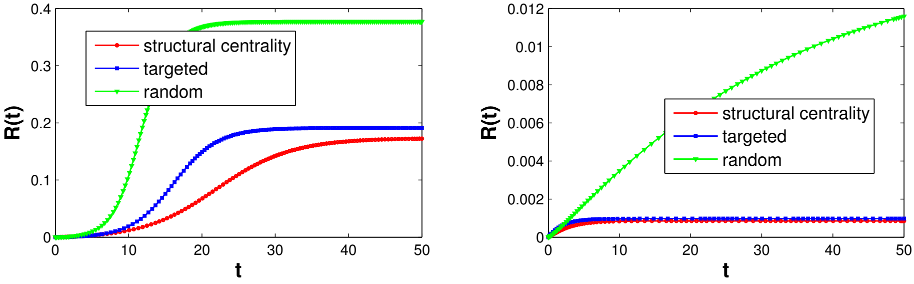
Fig. 9. Distributions of (a) degree centralities, (b) structural centralities with the node ids in e-mail network.

Using the rumor model from Eqs. (14)–(16), rumor dynamics is studied for random inoculation, and targeted inoculation on the basis of nodal degree and structural centrality. In Fig. 10, evolution of size of rumor is plotted against time for e-mail network. Final size of rumor is lesser with inoculation based on the structural centrality then the targeted inoculation based on degree for 10% inoculation of nodes (Fig. 10 (a)). Similar pattern for rumor evolution with time has been found for 30% inoculations (Fig. 10). The rumor is almost suppressed in this case. Thus, the structural centrality based inoculation suppresses the rumor in the networks more effectively. Random inoculation is not much effective for both e-mail network and generated SF network to suppress the rumor. In the case of generated SF networks, for very small fraction of time, rumor size has been found to be higher in structural centrality based inoculations initially for 10% as well as 30% inoculations of nodes, as shown in Fig. 11. But later, rumor size decreases in the structural centrality based inoculation in comparison with degree based targeted inoculations. The reason for that is that there are very few nodes in the network with highest degree, but in overall number of nodes with high degree are more. Therefore, initially the degree centrality plays an important role but later the structural centrality becomes more important.



(a) (b)

Fig. 10. Rumor evolution with the time for (a) 10% inoculations, (b) 30% inoculations for the e-mail network.



(a) (b)

Fig. 11. Rumor evolution during with the time for (a) 10% inoculations, (b) 30% inoculations for the generated SF network.

The numerical simulations have been done to observe the complete dynamical process with both the inoculation strategies with spreading ($\lambda = 0.5$), stifling ($\sigma = 0.2$) and spontaneous forgetting ($\delta = 1$) rates. Nodes interact with each other for rumor passing in each time step. After N nodes update their states according to the proposed rumor model, time step is incremented. To reduce the complexity, $\alpha = 1$ is considered. The random inoculation is implemented by selecting gN nodes randomly in the network. The targeted inoculation can be done after selection of the fraction of higher degree of nodes or structurally central nodes.

10. Conclusions

The robustness of the real world complex network has been measured efficiently with the help of the structural centrality base targeted attack in compare with the degree based targeted attack. The result shows that robustness of the network is lost only against a much higher fraction of deleted nodes for random attack. When we have full knowledge about the network structure, we can use an even better strategy than the targeted attack based on degree, *i.e.* based structural centrality. In SF network, the structural centralities of nodes have been derived in the complex networks and ranked with the help of l_{ii}^{+} values. A node with the high structural centrality needs lesser number of hops to reach the other node, even with small degree. The attacks, based on structural centralities of nodes were applied on SF networks and we found that the network is less robust as compared to targeted based on degree and random attacks. We have also inoculated nodes according to the rank of structural centrality. After this, we observed less rumor spreading then the degree centrality based targeted and random inoculations. It is also observed that there are a lot of nodes, having low degrees but high structural centralities and *vice versa*. It can be concluded that giant component will disappear after removal of fewer nodes in the structural centrality based targeted attack than for the degree based targeted and random attacks.

REFERENCES

[1] Y. Moreno, R. Pastor-Satorras, A. Vespignani, *Eur. Phys. J.* **B26**, 521 (2002).

[2] J. Zhou *et al.*, *Phys. Rev.* **E85**, 036107 (2012).

[3] A.-L. Barabasi, R. Albert, *Science* **286**, 509 (1999).

[4] M.E.J. Newman, *Siam Rev.* **45**, 167 (2003).

[5] D.J. Daley, J. Gani, J.M. Gani, *Epidemic Modelling: An Introduction*, Cambridge University Press, Cambridge, UK, 2001.

[6] D.P. Maki, M. Thompson, *Mathematical Models and Applications: with Emphasis on the Social, Life, and Management Sciences*, Prentice-Hall, NJ 1973.

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

[8] A. Singh, Y.N. Singh, *Acta Phys. Pol. B* **44**, 5 (2013).

[9] In *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion ACM 2012, p. 675.

[10] A. Singh, Y.N. Singh, *Rumor Dynamics and Inoculation of the Nodes in Complex Networks*, Cambridge Scholars Publishing, 2014.

[11] A. Singh, R. Kumar, Y.N. Singh, in: Eighth International Conference on Signal Image Technology and Internet Based Systems (SITIS), Nov. 2012, p. 798.

[12] A. Singh, Y.N. Singh, in: 2013 International Conference on Signal-Image Technology Internet-Based Systems (SITIS), Dec 2013, p. 514.

[13] A. Singh, Y.N. Singh, in: 2013 National Conference on Communications (NCC), 2013, p. 1.

[14] A. Singh, Y.N. Singh, Journal of Complex Networks, 2014, doi 10.1093/comnet/cnu047.

[15] R. Pastor-Satorras, A.Vespignani, in: *Handbook of Graphs and Networks: From the Genome to the Internet*, Eds S. Bornholdt, H.G. Schuster, Wiley-VCH, Berlin 2002, p. 113.

[16] A. Singh, R. Kumar, Y.N. Singh, in: *Computing and Combinatorics*, Springer 2013, p. 831.

[17] R. Albert, H. Jeong, A.-L. Barabasi, *Nature* **406**, 378 (2000).

[18] E. Lopez *et al.*, *Phys. Rev. Lett.* **99**, 188701 (2007).

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

[20] D.S. Callaway, M.E.J. Newman, S.H. Strogatz, D.J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).

[21] M. Molloy, B. Reed, *Comb. Probab. Comput.* **7**, 295 (1998).

[22] G. Paul, S. Sreenivasan, H.E. Stanley, *Phys. Rev.* **E72**, 056130 (2005) [`arXiv:cond-mat/0507202 [cond-mat.stat-mech]`].

[23] L. Freeman, *Soc. Networks* **1**, 215 (1979).

[24] D.A. Spielman, in: FOCS '07, 48th Annual IEEE Symposium on Foundations of Computer Science, Oct. 2007, p. 29.

[25] F. Fouss, A. Pirotte, J.-M. Renders, M. Saerens, *IEEE T. Knowl. Data En.* **19**, 355 (2007).

[26] J. Kunegis, *Konect — the Koblenz Network Collection*, `konect.uni-koblenz.de`, 2013.