# Measuring Nodal Contribution to Global Network Robustness

T. Feyessa and M. Bikdash
Department of Electrical and Computer Engineering
North Carolina A&T State University Greensboro, North Carolina 27411
tmfeyess@ncat.edu, bikdash@ncat.edu

***Abstract*—Network centrality indices are quantification of the fact that some nodes/edges are more central or more important in a network than others. Different centrality indices are suitable for different applications, but most of them have structural significance and require that the network be connected. Most centrality measures are only definitions and there hasn't been much work done in measuring the effectiveness of the measure in describing a network's performance, robustness and survivability. In this work, we have quantified the importance of a node that its centrality index implicates. We have conducted empirical analysis on different network robustness measures. The contribution of a node to these measures is studied as a function of its centrality index. We also propose a new type of betweenness centrality that is based on the flow circulating in a network. We compare circulation-based betweenness, eigenvector and shortest-path based betweenness centralities using network average clustering and shortest-path based network efficiency.** [1]

## I. Introduction

A network can be defined as an object composed of elements and interactions or connections between these elements. A graph, $G(V,E)$, made up of node set, $V$, and link set, $E$, is a natural means to model networks mathematically. Centrality measures the relative importance of a node or a link in terms of network efficiency and utilization of the network resources [1,2]. Node-centric measures are more convenient for computation and interpretation, and hence more common than edge-centric measures. Centrality is one of the most studied concepts in social network, and is used mainly to analyze social power and structural influence [3].

There is little work on comparing the performance of different centrality measures as an assessment and prediction tool. Most centrality measures are intuitive. For instance, the eigenvector centrality, which states that a center of centers is more central than the center of peripherals, is more intuitive than the simple degree centrality.

It is clear that from the definition of the centrality itself we can tell what is special about the most central node. But, how does this transform into describing the general importance of a node in the network? In this work, we quantify the impact of centrality measures by comparing the contribution of a node to the overall network robustness and its centrality index. We have used efficiency, connectivity and average clustering to quantify network robustness to failures of nodes or edges due to attacks or accidents.

The rest of this paper is organized as follows: In section two we revise recent works in network centrality. In section three we discuss ways of quantifying 'importance' based on network efficiency, connectivity and clustering. Then the concept of circulation is used to adopt a circulation-based betweenness centrality in section four. The circulation-based betweenness is similar with flow-betweenness in nature but much simpler to compute. Numerical experiment on scale-free and random networks to test our approach and the results obtained make the last section.

## II. Related Works

Centrality indices are quantifications of the fact some nodes/edges are more central or more important in a network than others. Different centrality indices are suitable for different applications, but most of them have structural significance and require that the network be connected [1]. Centrality indices that involve volume or length of a walk are usually referred as radial. Examples include degree-like and closeness-like measures. Indices based on number of paths passing through a node, such as the betweenness measure, are called medial [4].

A large amount of work in centrality comes from the social network studies [2,5,6]. Centrality in such networks is usually interpreted to measure power and social stratification. There are also many instances of centrality applications in biological networks [7,8], communication networks [9, 10], power networks [11] and transportation networks [12, 13]. In spatial networks, centrality measures are used in developing design requirement and studying vulnerability. In vulnerability analysis, centrality is used to identify critical locations and "vulnerability backbones" in the network and assess how well the network is distributed. Removing central nodes generally leads to increase in diameter, reduction of flow and lesser structural connectivity [14].

The simplest centrality measure is node degree, that is, the number of neighbors a node has. Therefore the degree centrality index is the degree vector of the graph. In contrast, spectral centrality, e.g., Bonacich centrality and $\alpha$-centrality, takes the importance of these neighbors into consideration [6]. Eigenvectors are important in spectral analysis of network and centrality measures [15]. The eigenvector centrality of a node

[1]This work is supported by Pennsylvania State University and The Defesne Threat Reduction Agency under contract DTRA01-03-D-0010/0020 and Subcontract S03-34.

$i$, $c_i^E$, can be defined [6] as a quantity satisfying

$$c_i^E = \frac{1}{\lambda}\sum_{j=1}^{n} A_{ij} c_j^E \qquad (1)$$

where $\lambda$ is a constant (an eigenvalue), $n$ is the number of nodes in the network and $A$ is the adjacency matrix.

Equation (1) can be rewritten as $\lambda C^E = AC^E$, where $C^E$ is vector of centrality indices. Both $\lambda$ and elements of $C^E$ must be nonnegative. The Perron-Frobenius theorem and its extensions [16] state that, a symmetric irreducible matrix with nonnegative entries has a simple real maximum eigenvalue $\lambda_{\max}$ and the entries of the corresponding eigenvector $\upsilon_{\max}$ are all positive. This result is directly applicable to the adjacency matrix of a connected graph. Similar, but weaker, results are available for digraphs. A matrix is irreducible if it couldn't be made block diagonal by row and column permutations. In short, $\upsilon_{\max}$ can be interpreted as a centrality vector.

Closeness-based centrality finds the distance center or the median of a graph. It has application in facility location [5], package delivery [2] and similar operations research problems. It is computed by summing up the distances from the current node to all remaining nodes. Let $d_{ij}$ be length of shortest path between nodes $i$ and $j$, then closeness centrality $c_i^C$ is

$$c_i^C = \sum_{j=1}^{n} d_{ij} \qquad (2)$$

The lesser the sum is, the more central is the node.

Betweenness is one of the most prominent centrality measures. It measures the influence/brokerage of a node over the connection of other nodes by summing up the fraction of shortest paths between the other nodes that pass through it [17,18]. This definition qualifies betweenness as a medial measure [4]. Given any two nodes $i$ and $j$, one can compute the number of shortest paths, $P_{ij}$, between $i$ and $j$ that satisfy a given criterion such as a threshold on path length. Out of all the possible shortest paths, some will pass through the node $h$ whose centrality is considered. Let $P_{ij}^h$ denote the number of these paths, the shortest path betweenness centrality of node $h$ is defined as

$$c_h^B = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{P_{ij}^h}{P_{ij}} \text{ for } i,j \neq h \qquad (3)$$

There are also many more centrality measures but they are usually variants of the indices defined above. Variants of shortest path betweenness can be obtained by including influence of endpoints, sources and targets, bounding or scaling the shortest distance, considering edge betweenness and group betweenness [17]. Sometimes betweenness is defined as a communicability betweenness. Communicability betweenness takes all paths into account rather than just the shortest paths [18]. Delta centrality, proposed in [19], is defined as the decrease in generic cohesiveness measure of the graph when the node is deactivated or removed from the network.

Network connectivity, flow, and shortest path are deeply connected with the notion of centrality. A network flow problem is: if nodes $i$, $j \in V$ in a directed graph $G(V,E)$ are connected by an edge $(i,j) \in E$, with a nonnegative capacity $c_{ij}$, a nonnegative quantity called flow $f_{ij}$ is assigned to the edge. The set of the flows $\{f_{ij}\}$ is called the flow pattern. A feasible flow in the network is a pattern that satisfies the constraint and conservation equations.

A graph $G$ is k-connected if every vertex cut has at least k number of vertices. Connectivity of $G$, $k(G)$, is the maximum $k$ such that $G$ is k-connected [20]. Connectivity is often used to define fault-tolerance of a network, a $k-node/link$ $connected$ graph can tolerate $k$ number of node/link failures [21]. Centrality of a node can be defined as the ratio of the connectivity after and before the node is deleted from the network [22].

## III. Measuring node importance

The influence of a node on the network can be computed by comparing certain network robustness measures before and after the removal of a node. The claim of centrality is to locate the most influential node, hence it is reasonable to expect a reasonable amount of degradation in the network robustness when the most central node is removed. Efficiency and connectivity are among the most desired features of a network in most applications. Here, we define importance/significance, $S$, of a node as the nodes contribution to these features.

### A. Network Efficiency

In a weighted graph, global efficiency can be measured as a function of inverse sum of shortest distances [23,24]. Let $d_{ij}$ be the shortest distance between nodes $i$ and $j$, and $n$ be the number of nodes in the graph, then shortest path based efficiency is,

$$\eta = \frac{1}{n(n-1)} \sum_{i \neq j \in V} \frac{1}{d_{ij}} \qquad (4)$$

If every arc in the network has a unity length, a complete graph will have an efficiency of 1.

Therefore the importance of a node $i$ as a function of network efficiency, $S^\eta(i)$, is defined as the decrease of the network's efficiency upon removal of the node, i.e.,

$$S^\eta(i) = 1 - \frac{\eta^i}{\eta} \qquad (5)$$

where, $\eta^i$ is the efficiency of the network without node $i$.

### B. Connectivity

Node/Edge connectivity is defined as the minimum number of nodes/edges whose removal disconnects the network into more than one component [20]. Two nodes in a network can be separated by removing all the disjoint paths connecting them. Hence, network connectivity can be computed by evaluating the minimum number of such paths for all pairs of nodes. To make comparison between different networks easier we have normalized this quantity. The normalization is based on connectivity, $\psi$, of a complete graph (an arc exist between every pair of nodes) is 1. In a complete graph, between any pairs of nodes $i$ and $j$ there are $n-1$ disjoint paths, where $n$

is the number of nodes in the network. Therefore, if $N_{dp}$ is the minimum of the number of disjoint paths for all pairs of nodes, the network connectivity is

$$\psi = \left[\frac{N_{dp}}{(n-1)}\right]^{\alpha} \tag{6}$$

In (6) $\alpha$ is used to increase the linearity relationship between probability of connection and measure of connectivity. The number of numerical simulations we conducted indicate the optimal value is around 0.6. Similar to efficiency, a node's importance is quantified using the drop in connectivity, i.e.,

$$S^{\psi}(i) = 1 - \frac{\psi^{i}}{\psi} \tag{7}$$

### C. Clustering

The average clustering coefficient of a network, which is the arithmetic mean of the clustering coefficient of the nodes, indicates the robustness of the network by measuring the tendency of the network to form tightly connected neighborhoods. The clustering coefficient of a node is defined as the ratio of number of triangles of which the node is a member of to the number of triangles the node could possible participate in [25,26]. Let $\varsigma \in [0\ 1]$ be the average clustering of a network, then nodes importance measured using clustering is defined as

$$S^{\varsigma}(i) = 1 - \frac{\varsigma^{i}}{\varsigma} \tag{8}$$

## IV. Circulation Betweenness

In well connected flow networks, the maximum circulation can be used to describe the network's performance. Circulation in a network is defined as the flow in a network when there is no source or sink. For a circulation to exist, there needs to be at least one cycle in the network. Existence of a cycle in a graph can be verified by looking at the power matrix of the adjacency matrix. Mathematically maximizing circulation in a network can be described with the following linear program, which can be computed with the typical efficiency of the simplex method

$$F = \max \sum_{(i,j)\in E} f_{ij} \quad \text{subject to} \tag{9}$$

$$0 = \sum_{i=1}^{n} f_{ij} - \sum_{i=1}^{n} f_{ji} \quad \text{for} \quad j = 1, 2, ..., n \tag{10}$$

$$0 \leq f_{ij} \leq c_{ij} \tag{11}$$

where $f_{ij}$ is the flow in link $(i,j)$ and $c_{ij}$ is the capacity of the link and $E$ is the link set.

Flow betweenness is defined in the same way as the shortest path betweenness, except that in the former we use flow instead of path passing through the node of interest. Both types of betweenness measures can be computationally expensive, especially when a larger network is considered. Therefore we propose a simpler yet effective, as shown in figure 1, betweenness measure. We call it circulation-based betweenness centrality index. A circulation-based betweenness of a node is computed as the ratio of the circulation passing through the node to the total circulation.

$$c_i^{BC} = \frac{1}{2F} \sum_{j=1}^{n} f_{ij}$$

where $f_{ij}$ is the flow on a link $(i,j)$ and $F$ is the total circulation. Figure 1 shows circulation-based and shortest path based

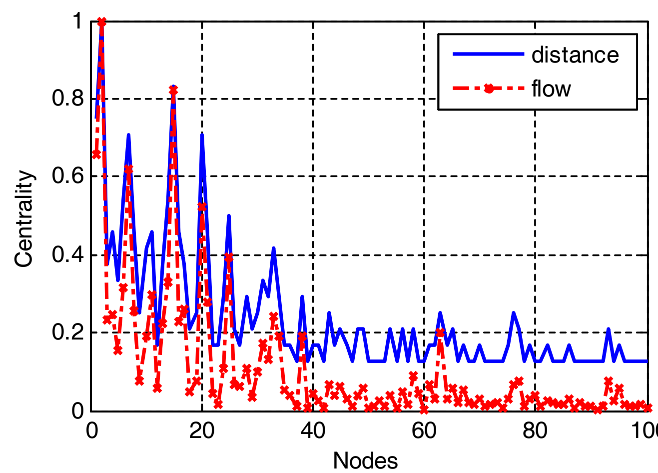


Fig. 1. Shortest path based betweenness (distance) and circulation based betweenness (flow) comparison

betweenness have similar pattern. Circulation betweenness can also be viewed as disjoint path betweenness when all the arcs in the network have a unity capacity. Circulation betweenness of a node in a symmetric network (there is exactly one return arc for every incoming arc) is equal to the in/out degree of the node.

## V. Numerical Simulation

### A. Networks

We have conducted simulation on scale-free networks. Scale-free networks exhibit power-law degree distribution [7,27]. The probability that a node has $k$ links, for a tunable scaling exponent $\gamma$, is $P(k)\tilde{}k^{-\gamma}$ [28]. In this work, the long-tailed distribution is achieved using Barabasi-Albert Algorithm (BA distribution) and random networks are created using Erdos-Renyi model. The algorithm grows a network by adding nodes progressively to a network with a linking probability between the new node and existing nodes favoring nodes that are well-connected. Another type of network growing method used in these simulations is random process. A directed-arc between pairs of nodes is assigned based on a predefined probability for a network.

Normalized centrality indices in random network generally tend to have smaller variance and higher value. In scale-free networks most nodes have low centrality and few have very high as shown in figure 2. This is a consequence of the power-law degree distribution in scale-free networks as opposed to uniform distribution in random networks.

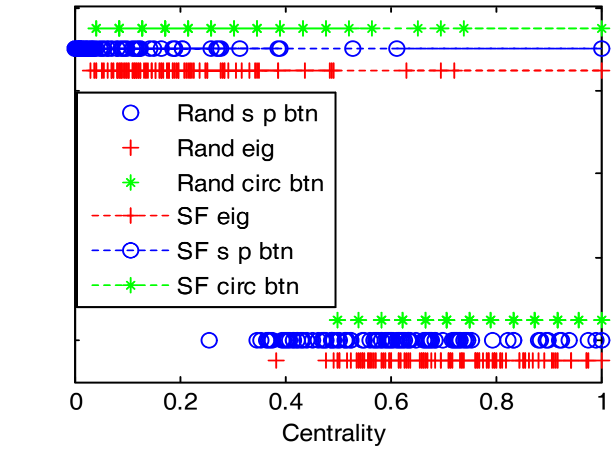


Fig. 2. Node normalized centrality distribution for scale free and random networks. The labels, Rand s-p btn, Rand eig and Rand circ btn indicate shortest path betweenness, eigenvector and circulation betweenness centralities of a random network. The labels SF eig, SF s-p btn, SF circ btn indicate eigenvector, shortest path betweenness and circulation betweenness centrality in a scale free network.

## *B. Results*

We experimented on the above two types of networks with three centrality measures; eigenvector, shortest path betweenness and circulation betweenness. Node importance is quantified using shortest path based efficiency and average clustering. Figures 3 and 4 show existence of a direct relation between the importance measures and centrality in random networks, especially when betweenness measures are used to find node centrality indices.

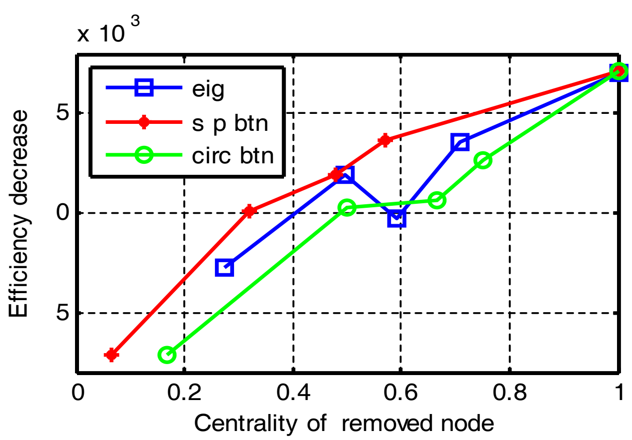


Fig. 3. Efficiency decrease in a random network when a single node is removed. The eig line shows nodes are selected based on eigenvector centrality, s-p btn and circ-btn stand for shortest path betweenness and circulation betweenness.

Scale-free networks behave in a different way than random networks, as seen in figures 5 and 6. This difference comes from the fact that the centrality indices are not uniformly distributed between the given range.

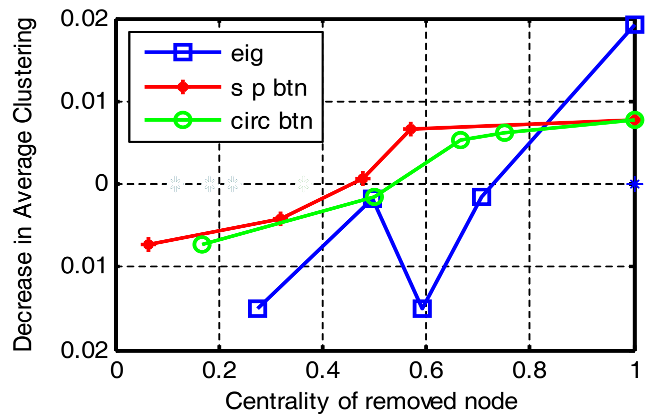


Fig. 4. Average clustering decrease in a random network

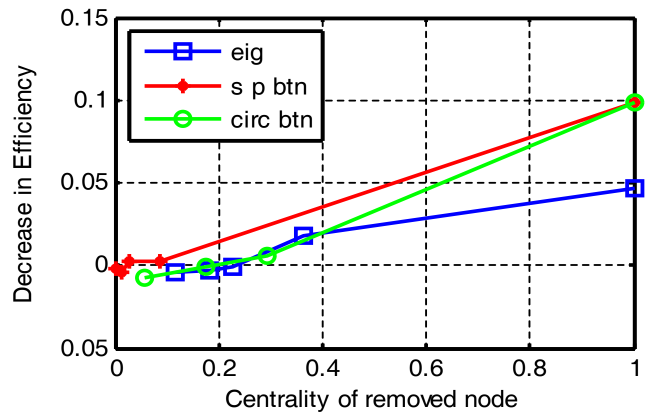


Fig. 5. Efficiency decrease in a scale free network

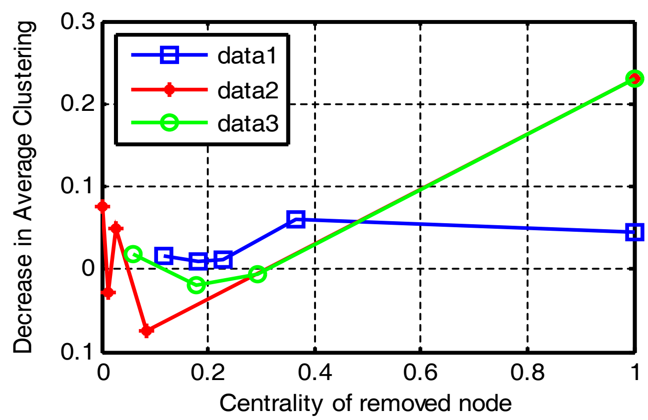


Fig. 6. Decrease in average clustering of a scale-free network

## VI. Conclusions

We have demonstrated both theoretically and with simulation that node centrality can be used to assess the impact of nodes failure in the whole network robustness. Centrality measure in scale-free network makes sense only in picking the top few central nodes due to the huge disparity in degree distribution created by the long tail. In random networks, betweenness measure give an insight about importance of the whole node set. Flow betweenness and shortest path betweenness centrality are also closely related function both by definition and empirical data. Another interesting observation from the simulation results is the robustness measures in random networks are less monotonic with the centrality value of the failing node when eigenvector centrality is used as opposed to shortest path and circulation betweenness centralities. As expected from its definition, shortest path efficiency is correlated with shortest path betweenness. The maximum loss of robustness in scale-free networks is much higher than random networks indicating scale-free networks are susceptible to targeted attacks.

## References


[1] Koschutzki, D.; Lehmann, K.A.; Peeters, L., Richter, S.; Renfelde-Podehl, D.; Zlotowski, O.; , " "Centrality Indices", Network Analysis: Methodological Foundations, pp.16-61, Springer-Verlag, 2005

[2] Borgatti, S. P.; , "Centrality and network flow", Social Networks, vol. 27, pp.55-71, 2005

[3] Marsden, P.V.; "Egocentric and sociocentric measures of network centrality", Social Networks, vol. 24, pp.407-422, 2002

[4] Borgatti, S.P.; Everett, M.G.; , "A Graph-Theoretic Perspective on Centrality", Social Networks, vol. 28, Is. 4, pp.466-484, 2006.

[5] Hage, P.; Harrary, F.; , "Eccentricity and centrality in networks", Social Networks, vol. 17, pp.57-63, 1995

[6] Bonacich, P.; "Some unique properties of eigenvector centrality", Social Networks, vol. 29, pp.555-564, 2007

[7] Wuchty, S.; Stadler, P. F.; "Centers of complex networks", Journal of Theoretical Biology, vol. 223, pp.45-53, 2003

[8] Estrada, E.; Hatano, N.; , "Statisitcal-mechanical approach to subgraph centrality in complex networks", Chemical Physics Letters, vol. 439, pp. 247-251, 2007

[9] Guan, Z. H.; Chen, L.; Qian, T. H.; , "Routing in scale-free networks based on expanding betweenness centrality", Physica A, 2010

[10] Mishkovski, I.; Biey, M.; Kocarev L.; , "Vulnerability of complex networks", Communications in Nonlinear Science and Numerical Simulation, vol. 16, pp.341-349, 2011

[11] Zio, E.; Piccinelli, R.; , "Randomized flow model and centrality measure for electrical power transmission network anlaysis", Reliability engineering and System Safety, vol. 95, pp.379-385, 2010

[12] Crucitti, P.; Latora, V.; Porta, S.; , "Centrality Measures in Spatial Networks of Urban Streets", Physical Reveiew E, vol 73, 2006

[13] Strano, E.; Cardillo, A.; Lacoviello, V.; Latora, V.; Messora, R.; Porta, S.; Scellato, S.; , "Street Centrality vs. Commerce and Service Locations in Cities: a Kernel Density Correlation case in Bologna, Italy", ArXiv Physics e-prints, 2007

[14] Demsar, U.; Spatenkova, O.; Virrantaus, K.; , "Identifying Critical Locations in a Spatial Network with Graph Theory", Transactions in GIS, vol.12, pp.61-82, 2008

[15] Fiedler, M.; "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory", Czechoslovak Mathemtaical Journal, vol. 25, pp.619-633, 1975

[16] Berman, A.; Plemmons, R.J.; , Nonnegative Matrices in the Mathematical Sciences, SIAM, 1994

[17] Brandes, U.; , "On variants of shortest-path betweenness centrality and their generic computation", Social Networks, vol. 30. pp.136-145, 2008

[18] Estrada, E.; Hingham, d. J.; Hatano, N.; , "Communicability betwneenness in complex networks", Physica A, vol. 388, pp.764-774, 2009

[19] Latora, V.; Marchiori, M.; , "A measure of Centrality based on network efficiency", New Journal of Physics, vol. 9, 2007

[20] Wai-Kan Chen, Theory of Nets: Flows in Networks, John Wiley & Sons, Inc., 1990.

[21] Sen, A.; Shen, B. H.; Zhou, L.; Hao, B; , "Fault-Tolerance in Sensor networks: A New Evaluation Metric", 2006

[22] Kirkland, S.; , "Algebaic connectivity for vertex-deleted subgraphs, and a notion of vertex centrality", Discrete Mathemtaics. vol. 310, pp.911-921, 2010

[23] Latora, V.; Machiori, M.; "The Architecture of Complex Systems", Physical Review Letters, vol. 87, 2001

[24] Crucitti, P. Latora, V.; Marchiori, M.; Rapisarda, A.; , "Error and attack tolerance of complex networks", Pyhsica A, vol.340, pp.388-394, 2004

[25] Fagiolo, G.; , "Clustering in complex directed networks", Phys. Rev. E., vol. 76, no. 2, 2007

[26] Barmpoutis, D.; Murray, R.M.; "Networks with the Smallest Average Distance and the Largest Average Clustering", arXiv:1007.4031v1, 2010

[27] Amaral, L.A.N; Scala, A.; Barthelemy, M.; Stanley, H.E.; "Classes of small-world networks", Proceedigns of the National Academy of Sciences of the United States of America, vol. 97, no. 21, pp.11149-11152, 2000

[28] Barabasi, A.L; Albert, R.; , "Emergence of Scaling in Random Networks", Science, vol. 286 no.5439, pp509-512, 1999