Systematic Exploration on Centrality Measures for Diverse Networks

M.Poongothai,
Department of Information Technology,
Institute of Road and Transport Technology,
Erode, India.
poongothaiirtt@gmail.com

Abstract:
Many complex systems such as biological and social systems can be modelled as graph structure like Protein -Protein interaction Network, Social Network, Gene-Gene Interaction Network. In most of the above network the basic issue is to find the influential node. Centrality Measures are an important analysis mechanism to explore influential element of complex networks. For this reason, many research work have been done on centrality measures and vast number of Centrality Measures have been proposed over years. Identifying the influential node is an important task in finding the overlapping community structure. The selection of a appropriate set of centrality measures is vital to gather important functional properties of a network. Centrality measures has its own application in several field.. In this survey paper, we investigate ten centrality measures and its suitability for applications.

Keywords: Complex Networks, Centrality Measures, Community Detection

1. Introduction

Community Detection is one of the most popular research topics in network science. Communities or clusters refer to that, the nodes in the same community have dense connections and have sparse connections between communities[1,2]. Community Detection has its application in the field of the biological network such as protein-protein interaction network, [3,4] residue interaction [5] , and gene-gene interactions networks[6].

Complex network models are everywhere in the field of Computer Science. Several Complex systems are modelled as graph structure which helps the study of complex system much easier.[1,7,8,9] This motivates the development and application of several metrics for their analysis, understanding and improvement. Once such analysis and understanding method is community Detection. Most of these analysis are based on identifying , evaluating and ranking the influential or vital node based on their power, influence and relevance. This class of metric captures idea of centrality which is used in many applications.

On a given network , the centrality is a quantitative measures that uncover the vital node in a network. The node is influential if it is more centred. Centrality measure is formally defined as real valued function on the nodes of the graph. But in real world examples the interpretation of 'centrality' is heavily dependent on the context.

Most of the metric algorithms are polynomial in time, when applied to real world network becomes difficult because of the huge size and complexity. Hence selecting the right centrality measure for the application is itself needed a study.

This paper is structured as follows In Sect. 1,.Given the introduction on community detection and the importance of centrality measure in community detection algorithm for complex networks. In Sect. 2. We describe the application of centrality measure in various networks. In Sect. 3.We have given ten centrality measures with its category, applications and its limitations. In Sect. 4,Finally we have concluded the importance of centrality measure and how the selection of centrality measure influence the community detection algorithm.

2.Applications of Centrality Measure

As mentioned early, elements of network can be evaluated and ranked using centrality measures. This priority ranking is based on the structural features of studied network. For example to find opinion leaders in social network, to discover communities in networks [10] , to estimate network traffic and congestion [11] , in bioinformatics .The following are some of the applications where centrality measures play a major role. 1.To identify influential users in social network which in turn used for online virtual Marketing. 2.predicting essential proteins [4] 3.Detecting financial risks[12]. 4.Predicting career movement[13]. 6.predicting failure with developer networks.

3.Categories of Centrality Measures:

Each centrality measure is attempting to capture some intuition or some set of intuitions , so it's clear that each will be of limited use to address scientific questions where some other intuitions is needed. The choice of suitable set of centrality measures is crucial for inferring functional properties of a network. The categories is based on what it implies for an network element to be "influential " to the given network. The definition of centrality changes as per the network and its application . Koschutzki et al. (2005) attempted a classification of centrality measures based on Geo, information flow ,Vitality Measures, Random Walk, Feedback centralities[15]. Freeman (1979) reviewed a number of published measures and reduced them to three basic classifications.These were degree, closeness and betweenness[16].

3.1Degree Centrality

An important node is involved in large number of interactions. Degree Centrality is the simplest which is defined as the number of links incident upon a node. The degree can be interpreted in terms of the immediate risk of node for catching whatever is flowing through the network. It is used to infer how this node is used to give information to other node(In-Degree) and how this node depends on other node to get some information.(Out-Degree)[21]. Mathematically defined as for a given graph G=(V,E) with |V| vertices and |E| edges is defined as

C_d=deg(V). (1)

Degree centrality is simple to compute.It's purely local and not able to compare centrality between nodes with same degree. Degree centrality does not consider global structure.

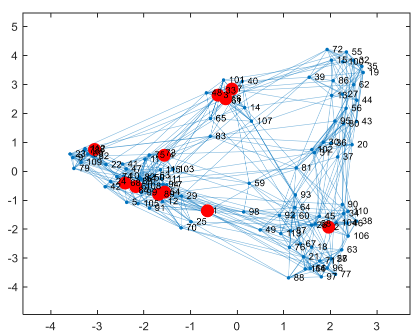

Fig. (1). American College Football Network with color indicates the node with high degree centrality measure.

3.2 Closeness Centrality

An important node is typically "close" to, and can communicate quickly with, the other

nodes in the network. It is calculated as the reciprocal of the sum of the length of the shortest path between the node and all other nodes in the graph[14,16].More central the node is the closer it is to all other nodes. Nodes with smallest closeness centrality has the best vision of information flow. This centrality infers how this node gets information faster to other node in the network. Mathematically defined by

$$C_c(x)=\frac{N}{\sum_y d(y,x)} \quad (2)$$

It indicates nodes as more central if they are closer to most of the nodes in the graph. It works good for dense connected graph. It can be applied to estimate level of efficiency and convenience, to capture how importance nodes for a given network. Nodes with smallest closeness centrality has best vision of information flow. Limitations Of Closeness Centrality is that numerical values is very small if the network is disconnected, all nodes have centrality zero. This measure is used to find service facility location of a network. Limitations is that this measure cannot be used if the network has disconnected components.

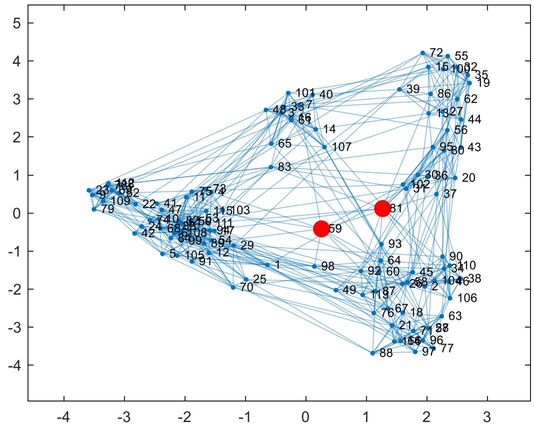

Fig. (2). American College Football Network with color indicates the node with high Closeness centrality measure.

3.3Eccentricity Centrality:

Eccentricity centrality is related to Closeness Centrality.

The diameter of a graph G is the maximum eccentric of a node that is the maximum distance between two nodes of G. The radius of Graph G is minimum eccentric of graph G. The eccentricity is the inverse of maximum geodesic distance between a vertex and any vertex of the network. [17]. A node is central if e(V)=r(G) and the centre of the G is the set of all central nodes. Eccentricity Centrality of vertex v_i is

$$Ce(V_i)=\frac{1}{max\{dis(v_i,v_j):j\in V\}} \qquad (3)$$

When it comes to distance-based measures eccentric is better than closeness. This measure is used to find the node that spreads the information fast. This is a global measure that considers the whole network. Limitations is that Eccentricity considers only maximum distance.

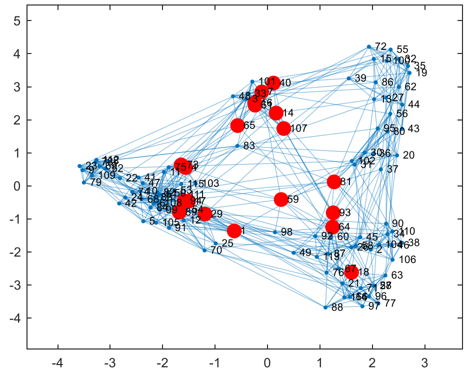

Fig. (1). American College Football Network with color indicates the node with high Eccentricity centrality measure.

3.4 Betweenness Centrality:

An important node will lie on a high proportion of paths between other nodes in the network. Betweenness Centrality is calculated as ratio of shortest path in which a node lies and the total number of shortest paths in a graph [18]. Mathematically defined as

$$C_b(v)=\sum_{s\neq t\neq v\in V}\frac{P_{svt}}{P_{st}} \qquad (3)$$

P_{svt}=number of shortest paths in which node v lies from s to t.

Pst=number of shortest paths in a graph from s to t.

This centrality is used to identify the node that has control over the flow of network. The node with the largest betweeness centrality has the strongest control over the information flow. Suitable for the network where flow of information is important. Betweenness is conventionally thought to measure the volume of traffic moving from each node to every other node that would pass through a given node (Borgatti, 1995). Limitations is that the computation cost is very expensive and is not suitable for the very large network.

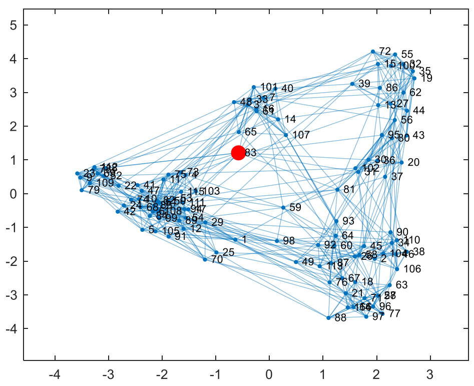

Fig. (4). American College Football Network with color indicates the node with high Betweenness centrality measure.

3.5 Eigen Vector Centrality:

The main principle of Eigen Vector centrality is that links from important nodes (as measured by degree centrality) are worth more than links from unimportant nodes. All nodes start off equal, but as the computation progresses, nodes with more edges start gaining importance. Their importance propagates out to the nodes to which they are connected. After re-computing many times, the values stabilize, resulting in the final values for eigenvector centrality [19-21].

An important node is connected to important neighbours. If a node has high degree then the node will have high Eigen Vector Centrality. Eigen vector centrality tries to generalize degree centrality by adding in the importance of neighbours.

Eigen vector Centrality of node x is Mathematically defined as

$$\text{Ceig}(v)=\frac{1}{\lambda}\sum_{y\to v} C_{eig}(y) \qquad (5)$$

Where Ceig converges to the dominant eigenvector of adjacency matrix A. λ converges to the dominant Eigen value of adjacency matrix A. This centrality infers how well this node is connected to all other well-connected nodes. Eigen Vector Centrality works well for strongly connected network.

Subject to the network topology, most of the weights of the eigenvector focus in a few nodes, like hubs. Most of the nodes in this case will present centrality close to zero and, hence the importance of nodes is not well quantified.

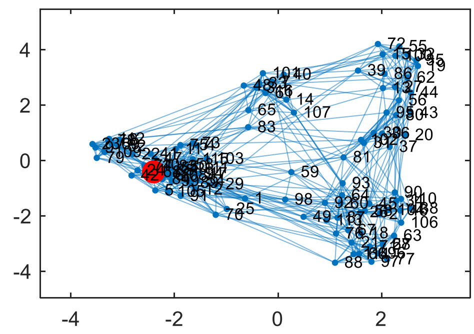

Fig. (5). American College Football Network with color indicates the node with high Eigen Vector centrality measure.

3.6 Katz's Centrality:

Katz centrality generalizes the concept of Eigen Vector Centrality to directed networks that are not strongly connected.

Katz centrality computes the centrality for a node based on the centrality of its neighbours. It is a generalization of the eigenvector centrality[22].The Katz centrality for node i is Mathematically defined as

$$Ck(v)=\alpha \sum_j A_{ij}x_j + \beta , \qquad (6)$$

where A is the adjacency matrix of the graph G with Eigen Values λ .The parameter β controls the initial centrality and $\alpha < \frac{1}{\lambda_{MAX}}$.Katz centrality computes the relative influence of a node within a network by measuring the number of the immediate neighbours (first degree nodes) and also all other nodes in the network that connect to the node under consideration through these immediate neighbours. Extra weight can be provided to immediate neighbours through the parameter β.Connections made with distant neighbours are, however, penalized by an attenuation factor α which should be strictly less than the inverse largest eigenvalue of the adjacency matrix in order for the Katz centrality to be computed correctly.

One possible problem with Katz centrality is that a vertex v with high centrality and high out-degree will cause a large number of vertices to have high centrality as well. It is less desirable since not everyone known by a well known person is well known. Here in Katz centrality high centrality node pointing to large number of vertices gives all high centrality which is a drawback.

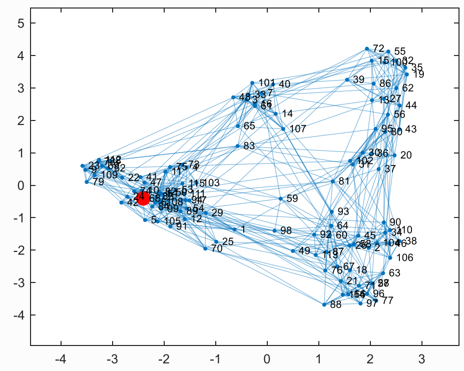

Fig. (6). American College Football Network with color indicates the node with high Katz centrality measure.

3.7. Leverage Centrality:

Joyce et al. introduce the concept of Leverage centrality measure to find highly essential vertices in the brain network. Leverage centrality considers the extents of connectivity of a vertex in relation to the connectivity's of its neighbours. This measure determines the extent to which the neighbours of a vertex rely on that vertex to access the network resources. By considering neighbours degrees, Leverage centrality gives different information about connectivity of a vertex in contrast with simple Degree centrality[23]. In fact, a vertex with high Degree centrality has not high Leverage centrality if its neighbours have also high degrees.

Leverage centrality is a measure of the relationship between the degree of a given vertex k_i and the degree of each of its neighbours k_j averaged over all neighbours N_i Leverage centrality for each vertex v_i in an undirected graph G(V, E) is Mathematically defined as

$$L(v_i)= \frac{1}{k_i}\sum_{N_i}\frac{k_i-k_j}{k_i+k_j} \quad (7)$$

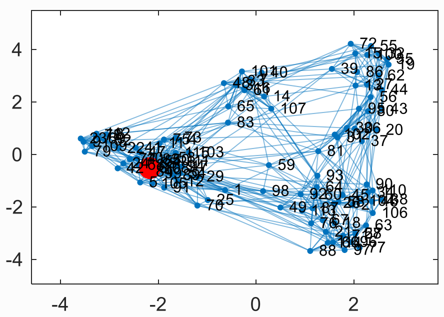

Fig. (9). American College Football Network with color indicates the node with high Leverage centrality measure

3.8. Information Centrality:

The Information Centrality is sensitive to how much an individual is close to the others and also to how much he stands between the others. The measure is based on the concept of efficient propagation of information over the network (Latora and Marchiori, 2001). The information centrality of an individual is defined as the relative drop in the network efficiency caused by the removal of the individual from the network. That is how the communication over the network is affected by the deactivation of the individual[24,25].The efficiency ϵ_{ij} in the communication between two points i and j is equal to the inverse of the shortest path length d_{ij} . The efficiency of G is the average of ϵ_{ij} .

To estimate information centrality, first define the matrix $C = (L + J)^{-1}$, where L is the Laplacian of A and J is a N × N matrix with all elements equal to one. Information centrality is then defined as

$$IC(v_i)=\left(C_{ii} + \frac{\sum_j C_{jj} - 2*\sum_j C_{ij}}{N}\right)^{-1} \quad (8)$$

This centrality is used when information flows along the network is considered.

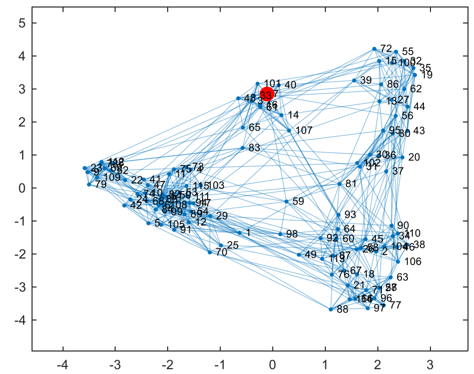

Fig. (8). American College Football Network with color indicates the node with information centrality measure.

3.9.Subgraph Centrality:

The subgraph centrality of the vertex i is defined as the "sum" of closed walks of different lengths in the network starting and ending at vertex i . As this sum includes both trivial and nontrivial closed walks it is to consider all subgraphs, i.e., acyclic and cyclic, respectively. The contribution of these closed walks decreases as the length of the walks increases[26].

A is the adjacency matrix of a network, the numbers of closed walks of length k are the diagonal entries of the matrix Ak , and the subgraph centrality of a node vi is defined as

$$CS(vi)=\sum_{k\geq 0}\frac{(A^k)_{ii}}{k!} \quad (9)$$

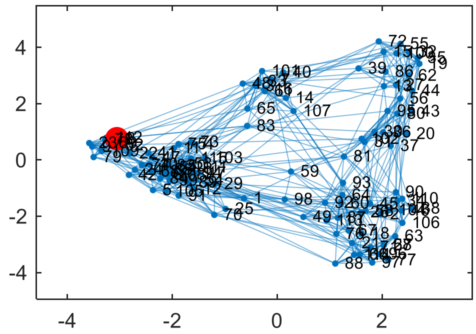

Fig. (9). American College Football Network with color indicates the node with high subgraph centrality measure.

3.10.Page Rank centrality:

A node is important if it linked from other important nodes or if it is highly linked. PageRank is an algorithm that measures the transitive influence or connectivity of nodes[27,28]. Page Rank is the Mathematically defined as

$$Ck(v)=\alpha \sum_j A_{ij}\frac{x_j}{k_j^{out}} + \beta \quad (10)$$

One limitation of the PageRank measure lies in the definition of connections among all pair of nodes. PageRank fails to calculate the importance of the node , that is it favors old page.

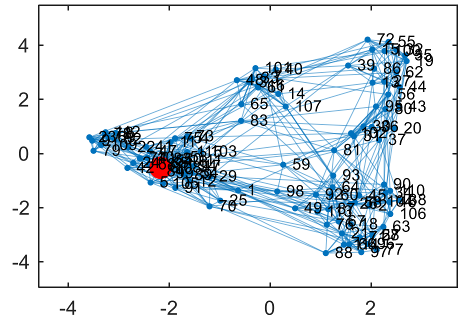

Fig. (1). American College Football Network with color indicates the node with Page Rank centrality measure.

4.Conclusions:

We have explored and compared different centrality measures. We visually exemplified more influential vertices for each centrality measures with a American College Football Network. We also briefly discussed the suitability, applications and limitations of the centrality measures. The survey concludes that different measures by using different beliefs and guidelines categorized different vertices as central and important. Each centrality measure has its unique property and hence to make difference and apply them in proper area is still an open issue.

References:

1. Girvan M, Newman MEJ: Community structure in social and biological

networks. Proc Nat Acad Sci USA 2002, 99(12):7821–7826.
2. Newman M: Mixing patterns in networks. Phys Rev E 2003, 67(2):026126.
3. Maslov S, Sneppen K: Specificity and stability in topology of protein networks. Science (New York, NY) 2002, 296(5569):910–913.
4. Han JDJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, Vidal M: Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 2004, 430(6995):88–93.
5. Hu, Guang & Yan, Wenying & Zhou, Jianhong & Shen, Bairong. (2014). Residue Interaction Network Analysis of Dronpa and a DNA clamp.. Journal of theoretical biology. 348. 10.1016/j.jtbi.2014.01.023.
6. Arzucan Özgür, Thuy Vu, Güneş Erkan, Dragomir R. Radev, Identifying gene-disease associations using centrality on a literature mined gene-interaction etwork, *Bioinformatics*, Volume 24, Issue 13, 1 July 2008, Pages i277-i285,
7. Bullmore, E. & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 64(2), p. 10-12
8. Mason, O.; Verwoerd, M.: 'Graph theory and networks in Biology', IET Systems Biology, 2007, 1, (2), p. 89-119, DOI: 10.1049/iet-syb:20060038
9. Lesne, A. Lett Math Phys (2006) 78: 235. https://doi.org/10.1007/s11005-006-0123-1
10. Z. Yakoubi, R. Kanawati, "LICOD: A Leader-driven algorithm for community detection in complex networks", *Vietnam. J. Comput. Sci.*, vol. 1, no. 4, pp. 241-256, 2014.
11. Holme P. Congestion and centrality in traffic flow on complex networks. Adv Complex Syst 2003; 6(02): 163-76.
12. arXiv:1907.07908v2 [q-fin.MF]
13. Roebken, H. High Educ (2007) 54: 99. https://doi.org/10.1007/s10734-006-9048-y
14. Freeman, L.C., 1979. Centrality in networks: I. conceptual clarification. Social Networks 1, 2–239.
15. Koschützki, D., Lehmann, K.A., Peeters, L., Richter, S., Tenfelde-Podehl, D., Zlotowski, O.: Centrality indices. In: Brandes, U., Erlebach, T. (eds.) Network Analysis. LNCS, vol. 3418, Springer, Heidelberg (2005)
16. Latora V, Marchiori M. Efficient behavior of small-world networks. Phys Rev Lett. 2001;87(19):198701.
17. Hage P, Harary F (1995) Eccentricity and centrality in networks. Soc Netw 17:57–63
18. Linton C. Freeman. 1977. A set of measures of centrality based on betweenness. Sociometry 40, 1, 35–41 (March 1977). DOI: 10.2307/3033543
19. Bonacich P (1972) Factoring and weighing approaches to status scores and clique identification. J Math Sociol 2(1):113–120
20. Bonacich P, Lloyd P (2001) Eigenvector-like measures of centrality for asymmetric relations. Soc Netw 23(3):191–201
21. Bonacich P (2007) Some unique properties of eigenvector centrality. Soc Netw 29:555–564
22. Katz L (1953) A new status index derived from sociometric analysis. Psychometrika 18(1):39–43
23. Joyce KE, Laurienti PJ, Burdette JH, Hayasaka S. A new measure of centrality for brain networks. PloS One 2010; 5(8): e12200.
24. Fortunato S, Latora V, Marchiori M. Method to find community structures based on information centrality. Phys Rev E Stat Nonlin Soft Matter Phys 2004; 70(5 Pt 2): 056104.
25. Stephenson K, Zelen M. Rethinking centrality: methods and examples. Soc Networks. 1989;11(1):1–37.

26. Estrada E, Rodriguez-Velazquez JA (2005) Subgraph centrality in complex networks. Phys Rev 71:056103
27. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: bringing order to the web. 1999.
28. Xing W, Ghorbani A, editors. Weighted pagerank algorithm.Communication Networks and Services Research, roceedings Second Annual Conference on; 2004: IEEE.