RESEARCH REVIEW JOURNALS
Volume-04
Issue-05
May-2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com[UGC Listed Journal]

Analysis of Eigenvector Centrality Measurement Algorithm for Tracking Online Communities

[1]Deepak Adhikari & [2]Dr. Jitendra Sheetlani

[1]Research Scholar, Department of Computer Science , Sri Satya Sai University of Technology & Medical Sciences, Sehore, M.P.
[2]Research Guide, Department of Computer Science , Sri Satya Sai University of Technology & Medical Sciences, Sehore, M.P.

ARTICLE DETAILS

Article History
Published Online: 25 May 2019

Keywords
Social Network Analysis, Centrality, Communities, Eigenvector.

ABSTRACT

The centrality of the network is used to identify the most influential people in a network center or those well connected through a network. Tracking the single community in social networks is usually carried out using some of the key measures used in social networks. The SCAN (Social Cohesion Network Analysis) approach used the centrality measure to track social network communities. This paper introduces new alternative eigenvector centrality measures for group monitoring and is more suitable compared to the centrality measures algorithm for betweenness.

1. Introduction

Social network analysis as a social relationship for the theory of networks comprising of nodes and ties also called links, or connections, edges,). Nodes are the individual Communities within the networks and links are the relationships among the Communities. Centrality measures signify the importance of the communities of the network / units. There are different measurements of a vertex's centrality within a graph in theory of graphs and network analysis that determines the corresponding significance of the vertex in the graph.

The Internet has generate various types of systems for sharing information, including the Web. Online social networks have recently gained significant prominence and are now one of the web's most popular sites. Online social networks are structured between users, unlike the Internet, which is primarily organized around content. Participating users enter a network, post their profile, and (optionally) build connections to any other users they wish to associate with. The resultant social network provides a basis for social relationships, recognizing people of similar interests, and discovering information and expertise that other users have contributed or endorsed.

2. Related Work

The characterization of the social behavior and connectivity of nodes within networks has been characterized by numerous central measures such as betweenness, degree, closeness, information, eigenvector and dependence centrality. The rationale of using centrality metrics is that individuals involved in one or more subgroups would score higher in comparison to the related network's centrality ratings. Freeman [1977], Betweenness centrality, is mostly used to identify and quantify subgroup membership and community membership, while centrality of degree and closeness are used to classify leading members.

While measurements of network centrality can be easily calculated using computer programs such as Pajek, de Nooy et al.[2005] and UCINET, Brandes and Pich[2007], researchers did not have a majority on the most important centrality measure to be used to classify members of the subgroup. Computational efficiency can become a problem in extremely large social networks when choosing which centrality measure to use. After UCINET's analysis, a measure of betweenness centrality has a high measurement and time complexity compared with eigenvector centrality and degree centrality is the easiest to calculate. The Social Cohesion Analysis of Networks (SCAN) method, Chin[2009], was developed to automatically identify long-term cohesive subgroups of people in social networks. The SCAN approach should be used on the basis that from the experiences a social graph can be generated of the online community where the links are not typed(i.e., there are no associated semantics).

In the social graph, each link represents an interaction between two individuals where one individual has responded to the other's post in the online community. The SCAN approach is designed to identify coherent subgroups based on social networks that are derived from online interactions around common topics of interest and concentrate only on the betweenness centrality; other central measures can be helpful.

3. Existing betweenness centrality measurement algorithms for tracking online communities

➢ **Betweenness Centrality**

Newman [2005] defined betweenness is a vertex measure of centrality inside a graph. Vertices which appear on many of the shortest routes between other vertices are more betweenness than those that do not.

The betweenness CB(v) for vertex v is calculated as follows for a graph G: = (V, E) with n vertices:

1. Calculate the shortest paths between them for each pair of vertices (s, t).
2. Determine the fraction of the shortest paths that go through the vertex in question for each pair of vertices (s, t) (here, vertex v).
3. Sum this fraction across all vertices pairs (s, t).

Or, more succinctly:

$$C_B(v) = \sum_{s \neq v \neq t \in V}^{|Y|} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad \text{..............................(3.1)}$$

where σ_{st} is the number of shortest route, Kahng et al.[2003], from s to t and $\sigma_{st}(v)$ is the number of the shortest paths from s to t passing through the vertex v. Measuring the betweenness and closeness of all the vertices in a graph requires measuring the shortest paths on a graph between every pair of vertices. In determining the centrality of betweenness and closeness in a graph of all vertices, graphs are assumed to be undirected and linked to the loop and multiple edge allowance. When dealing specifically with network graphs, often graphs do not have loops or multiple edges to maintain simple relationships (where edges represent links between two persons or vertices). In this case, using Brande's algorithm, the final centrality scores will be divided by 2 to account for every twice-counted shortest path.

➢ **The Sequential Algorithm**

The sequential algorithm was developed by Brandes [2001], the basic idea is to identify shortest paths by breadth-first search (or BFS). Conducting a BFS from the beginning vertex s gives every other vertex the lengths of the shortest paths from s. With a bit of extra bookkeeping, this BFS can also calculate the shortest paths from s to each t and may even keep track of how some of the paths are passing through each other v. The query from s requires two sweeps across the graph: the first sweep is the BFS, which measures for each t and also reports information on the "predecessors" of each vertex reached in the hunt, the second sweep goes in opposite order via the BFS list, modifying each vertex's centrality scores using its predecessors. The searching from the vertex s determines the contribution to $C_B(v)$ from the internal sum of the equation (3.1) for each v. The algorithm performs the search from any initial vertex s, adding the contributions as measured.

The algorithm has a number of places to introduce parallelism. The easiest thing to do is to trace the outermost loop over s; that is, to simultaneously perform several breadths-first searches on various processors.

A second approach is to parallel the individual breadth-first searches, either by working on multiple nodes at a level simultaneously or by working on multiple neighboring single nodes. The fastest parallel benchmark codes used by BC use more than one parallel source.

In topology and related fields of mathematics, closeness is one of the fundamental concepts. Intuitively, if they are random close to each other, two sets are similar. The definition can be extended naturally in a metric space where the theory of distance between space elements is defined, but the topological spaces cannot be generalized where there is no specific way of measuring distance.

Naturally, the definition may be extended to an area mathematical space topological space} wherever the theory of distance between space components is outlined,

Closeness is among the basic principles in a topological space in topology and related fields in mathematics. Intuitively, if they are arbitrarily close to each other, two sets are close. Obviously, the definition can be applied to a metric space where the notion of distance between space elements is established, but it cannot be extended to topological spaces where there is no precise way to evaluate distances.

4. Proposed alternative eigenvector centrality measurement algorithm for tracking online communities

➢ **Eigenvector Centrality**

Ruhnau [2000] and Freeman [1978] defined eigenvector centrality as a proportion of the importance of a node in a network. It allocates relative scores to all the network nodes dependent on the rule that connections with high-scoring node contribute a bigger number of scores of the node being referred to than equivalent connections of low-scoring nodes. Google's PageRank is a variation of the Eigenvector centrality measure.

• **Using the adjacency matrix to determine eigenvector centrality**

let x_i a chance to indicate the score of the i^{th} hub. Let $A_{i,j}$ be the nearness grid of the system. Consequently $A_{i,j}$ = 1 if the ith hub is nearby the j^{th} hub and $A_{i,j}$ = 0 generally. All the more for the most part, the sections in A can be genuine numbers speaking to association qualities, as in a stochastic network.

• **Use the adjacency matrix to determine eigenvector centrality**

Assumme x_i denotes the i^{th} node score. Assume $A_{i,j}$ be a network adjacency matrix. Therefore, if the i^{th} node is adjacent to the j^{th} node, $A_{i,j}$ =1 and $A_{i,j}$ = 0. More commonly, as in a stochastic matrix, the entries in A can be real numbers that show the intensity of the connection,

Let the centrality rating be equal to the sum of the scores of all nodes linked to it for the i^{th} node. Hence

$$X_i = \left(\frac{1}{\lambda}\right) \sum_{j \in N(i).J=1}^{N} (X_j) = \left(\frac{1}{\lambda}\right) \sum A_i, j(X_j) \quad \text{........................(3.2)}$$

where $M(i)$ is the set of nodes bound to the i^{th} node, N is the total number of nodes and λ is a constant. This can be rewritten in vector notation as

$X = \left(\frac{1}{\lambda}\right) X = \left(\frac{1}{\lambda}\right) AX$ or as the eigenvector equation

$AX = \lambda X$. In particular, there will be several separate eigenvalues λ for which there will be an eigenvector solution. However, the alternative option that all the entries in the eigenvector be positive (by the Perron–Frobenius theorem) means that the desired centrality measure results only in the highest eigenvalue. The i^{th} component of the associated eigenvector then provides the centrality score of the i^{th} node of the network. Power iteration is considered one of the many eigenvalue algorithms that can be used to determine this dominant eigenvector. Eigenvector centrality is a sort of recursive model of degree centrality. The general algorithm is as follows:

➢ Power Method

An efficient method for finding the largest eigenvalue/eigenvector pair is the power iteration or power method. The power approach is an iterative calculation that entails performing a matrix-vector multiply again and again until the change in the literate (vector) falls below the threshold provided by the user. Outline of the power method algorithm are stated below:

X[n] =1; //create initial vector and set to a vector of all ones.

While (convergence criteria not met)

{

for(i=0;i<=n-1;i++)

for(j=0;j<=n-1;j++)

*tmp[i]+=x[j]*A[i][j]; //metrix vector multiply for(k=0;k<=n-1;k++)*

*norm_fector +=tmp[j]*tmp[k]; //calculate Euclidian norm*

for(k=0;k<=n-1;k++)

tmp[i]/=norm_fector; //2109ormalize the vector

X=tmp;

}

In the initialization step, the starting iterates is set to a vector of all ones, allowing the algorithm to behave deterministically, a useful property for performance analysis. The algorithm joins a while loop after initialization consisting of a multiplication of the matrix-vector, followed by a Euclidean norm calculation and ultimately a vector normalization. Important graphs tend to be sparse, meaning that the adjacency matrix is generally sparse and can be defined efficiently in compressed row storage (CRS). Apart from being space-efficient, storing the matrix in CRS increases the performance of the multiplication matrix-vector in the loop. The pseudo-code for the sparse matrix-vector multiply is:

// sparse matrix vector multiply

for i=0:(n-1)

for j=0:(numNonZeroElements[i]-1)

colIdx = C(I,j);

tmp[i] += x[colIdx];

5. Analysis of centrality measurement algorithm

Assuming a Social network of communities and evaluating various centrality measures by UCINET simulator in following manner:

Step1: Suppose social network dataset contain node label and each row represents the relationship among the nodes.

L1 N1 N2 N3 N4
LI3
N1 LI1 N2 N3 N4 N5 U1
N2 L1 N1 N3 N4 U1
N3 L1 N1 N2 N4 N5 U1
N4LI1 N1 N2 N3 N5 U1
N5 N1 N3 N4 N7 U1
N6 N7 N8 N9
N7 N5 N6 N8 N9 U4
N8 N6 N7 N9 U4
N9 N6 N7 N8 U4
U1 N1 N2 N3 N4 N5
U2
U4 N7 N8 N9

Table 1.1: Comparatively analysis of Betweeness and Eigen Vector measures for Synthetic Dataset

Sl No	Node	Degree	Closeness	1	2
				Betweenness	Eigenvector
1	L1	30.769	23.636	0.000	43.368
2	L3	0.000	0.000	0.000	-0.000
3	U1	38.462	26.531	1.923	52.043
4	U2	0.000	---	0.000	0.000
5	U4	23.077	23.636	0.000	4.070
6	N1	46.154	27.083	4.808	58.960
7	N2	38.462	24.074	0.321	51.669
8	N3	46.154	27.083	4.808	58.960
9	N4	46.154	27.083	4.808	58.960
10	N5	38.462	28.889	38.462	45.718
11	N6	23.077	23.636	0.000	4.070
12	N7	38.462	27.660	36.325	12.011
13	N8	30.769	24.074	0.427	4.719
14	N9	30.769	24.074	24.074	4.719

Step 2: After examining the dataset L1, L3, U1, U4, N1, N2, N3, N4, N5, N6, N7, N8, N9 and track community, the comparatively analysis of both (eigenvector & Betweenness) measures and algorithm are given in Table 1.1.

Step 3: Graph 1 is generated by UCINET Simulator shows the highly influential nodes for selecting communities or tracking communities and also indicates that the eigenvector centrality measure is more effective as compared to betweenness centrality measure.

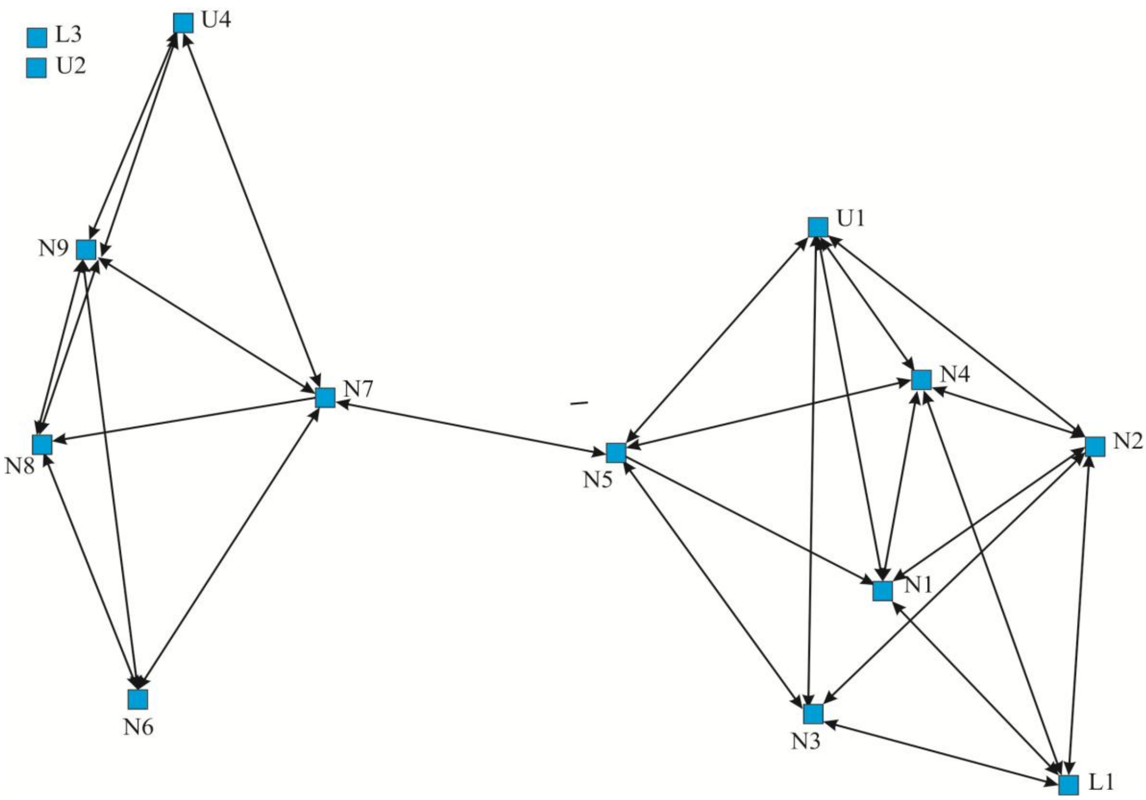

Fig 1.1: Graph for centrality measures work on various communities using UCINET simulator

Simulation is performed using UCINET Simulator for comparing two separate centrality measures based on the value produced by community information is shown in fig 1.1. Consequentially resulting graph show higher influential nodes. N1, N3, N4 are highly influential nodes in case of eigenvector centrality and N5 and N7 in case of betweenness centrality.

The Betweenness Centrality calculation informs how a node "in-between" is around a full graph by calculating the number of all the shortest paths passing through that node. The best-known algorithm for that takes O (VE +2VlogV) time and O(N+V) memory. Eigenvector centrality calculates the significance of a node by measuring its connectivity to other "significant" nodes in the graph. The process of selecting it is identical to the spread of the belief, and the algorithm is iterative to the time complexity of the O(k.E), Where k is the number of iterations required before convergence, V is the number of vertices and E is the number of edges of graphs. The number of edges E in dense graphs tends to approach O (2V). The calculation becomes super-linearly expensive with the increase in the value of E and V. To measure the Betweenness centrality and the eigenvector centrality for a graph, for instance is with V=656 and E=25000 took 142.5 seconds and 5.75 seconds respectively. As an example suppose the graph name is “Facebook” and betweenness centrality and eigenvector centrality for a graph “Facebook” V=2,625 and E=99,999 took 570 s and 23 s respectively where overall throughput can be achieved by eigenvector centrality. The observation reveal that resultant eigenvector centrality measure is more suitable for “select” step of SCAN method because eigenvector centrality can find or select high influential node in less time and in an easier manner.

6. Conclusion

This chapter evaluates eigenvector and betweenness centrality measurement algorithms for tracking online community in social network, accordingly the following conclusion are drawn. Ones, eigenvector centrality measures for tracking community are more suitable against betweenness centrality measures algorithm. Two, the betweenness centrality measure has high calculation complexity as compare to eigenvector measure algorithms. Three, eigenvector measure algorithms provide better time complexity against betweenness centrality measure algorithms. Four, eigenvector centrality is more desirable measure for SCAN method for selecting high influential nodes or communities.

References

1. Abelson, R.P. [1968] A Chapter in *Theories of cognitive consistency*: *A sourcebook* (Edited) (pp. 112-139). Chicago: Rand McNally.
2. Alba, R.D.A. [2003] Graph-theoretic definition of a sociometric clique. Journal of Mathematical Sociology 3, 113–126.
3. Alberto, F., Gómez S. [2008] Solving Non-uniqueness in Agglomerative Hierarchical Clustering Using Multi dendrograms. Journal of Classification 25: 43–65
4. Aronson, E., Wilson, T.D., Akert, A.M. [2005] *Social Psychology* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
5. Baumes, J., Goldberg, M., Magdon-Ismail, M. [2005] Efficient identification of overlapping communities.

Intelligence and Security Informatics (ISI) 27– 36.
6. Borgatti, S.P. [2002] NetDraw: Graph Visualization Software. Harvard: Analytic Technologies.
7. Estivill-Castro, V., Yang, J. [2000] A Fast and robust general purpose clustering algorithm. Pacific Rim International Conference on Artificial Intelligence, pp. 208-218.
8. Etzioni, A., Etzioni, O. [1999] Face-to-face and computer-mediated communities; a comparative analysis. Inform Soc 15(4):241–248.
9. Falkowski, T., Bartelheimer, J., Spiliopoulou, M. [2006] Community dynamics mining. In Proceedings of 14th European Conference on Information Systems (ECIS 2006) (Gteborg, Sweden, 2006).
10. Greene, D., Doyle, D., Cunningham, P. [2010] Tracking the evolution of communities in dynamic Social networks. In Proc. International Conference on Advances in Social Networks Analysis and Mining(ASONAM'10),IEEE. Jaccard,P.(1912).Thedistribut.
11. Gregson, A.M.R. [1975] *Psychometrics of similarity*. Academic, NY, USA
12. Huang, Z. [1998] Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery, 2:283–304.

