



Presentation of Fuzzy Model to Compute the Edge Betweenness Centrality in Social Networks

Noushin Saed

Department of computer engineering, College of engineering, Khouzeestan Science and Research Branch, Islamic Azad University, Ahvaz, Iran

Mehdi Sadeghzadeh and Mohammad Hussein Yektaie

Department of Computer engineering, Islamic Azad University, Khuzestan, Mahshahr, Iran

ABSTRACT

Nowadays, we live in network area. The area through which the formation of various social network, new communicative and informing methods are introduced to the widespread social communications. A social network is a social structure which is made out of individuals and meanwhile, by the pass of time, the analyzing these social network will gain increasing primacy. In this research, one of the parameters of social network analysis called edge betweenness centrality is introduced. Edge betweenness is an edge to compute the shortest paths between pair of no desin the network that passes through it most frequently. In this research, to detect the communities through edge betweenness centrality algorithm, a method is introduced in such a way that each edge by receiving one fuzzy membership degree in the interval $[1,0]$ the measure of its effect on the network will be different. One of the features of this algorithm that makes it distinguished from others is the application of fuzzy logic to detect the communities of social network. Then by introducing the density of each cluster the density measure of the communities graph is computed through considering the fuzzy detected structures. The finding of the implementation of algorithm indicated that introduced algorithm to compute the density of samples and to detect the number of mono-nodes while clustering has revealed more accuracy rather than the related works.

Keywords

Social Networks, Community Detecting, Community Clustering, Membership Degree, Edge Betweenness.

1. INTRODUCTION

The researches have revealed that mostly there are some common features among real networks, such as biological systems and cooperating scientific systems. Among these common features “community structure” has attracted the focus of attention. The community system is defined to explicate the clustering of social networks. A community structure of network can be simply divided into different community [1]. The connection between the different nodes in community is denser in proportion to the connection of these nodes with other nodes. Although



there is not any general definition of community yet; however some accepted measurable parameters are investigated to recognize these kinds of community structure and their importance in determined networks to improve the efficiency and the time of implementation [2]. But in this research study we believe that to realize the detection of communities in social networks the priority should not be given to the time of implementation. Of course, known that clustering is a NP HARD algorithm, a kind of algorithm should be introduced that when we have polynomial it can solve the problem. Therefore, giving priority to and much attention the time of implementation can reduce the accuracy in community detecting and will result in the fact that a number of communities that are really present in social network be lost due to the improvement of the time of implementation. Structure of the present paper is as follows: Next part is assigned to review of literature. In part 3, edge betweenness centrality algorithm and its characteristics have been described. Proposed method is presented in part 4, and description of Simulation is presented in part 5. Finally, part 6 includes conclusion and some future works.

2. RELATED WORKS

Clustering, put it in another term, the recognition of communities as one of interests to the science of data mining has been investigated and studied. Usually, a community in network is a group of nodes that their connection in intra-community arrangement is more than to the rest of the network [3]. This intuitive definition has been formalized in a number of competing ways, usually by way of a quality function, which quantifies the goodness of a given division of the network into communities. Some of these quality functions measures like modularity and normalized cuts are more common than other cases but none of them has gained public acceptance since one criterion in all situations is not workable. Algorithms for community discovery deal to problem as well as the features of efficiency from different perspectives. However, this will obviously improve a particular quality criterion. Spectral methods, Kernighan-Lin (KL) algorithm and flow-based postprocessing are examples of algorithm that attempt obviously to particular standard quality [4]. Hierarchical methods are one of the traditional methods of community clustering. In this hierarchical clustering method, a hierarchical structure in a tree model is dedicated to the final clusters according to the amount of their commonness. This hierarchical tree is called dendrogram. The methods of hierarchical clustering techniques are usually according to Greedy and Stepwise-optimal algorithms. The clustering methods according to productive hierarchical structure through them are usually divided into Bottom-Up (agglomerative) and Top-Down (Divisive) algorithms [5]. Agglomerative algorithms begin to function with a node as a community in the network and they merge similar communities in each stage. This repetitive process continues until an ideal number of communities emerge up or the other nodes for merging are dissimilar to one another. The divisive algorithms operate in regressive manner, they commence to function with a complete network as a community and in each stage determine one community then divided it into two



parts and this procedure will continue until culminating in clusters containing one member. Clauset and his associates [6] introduced a hierarchical method through that the nodes of graph are divided according to a greedy algorithm that the modularity resulted from this division reaches its maximum. Another method of clustering is partitioning method. The algorithm of partitioning clustering will gains one partition from data instead of the structure that dendrogram produces through a hierarchical technique. The partitioning method is useful as well as in total data collections this is one of the advantages of this method since prevents the dendrogram structure that involves a lot of computations. One of the problems relevant to the algorithms of partitioning is the selection of ideal clusters number and preliminary [7]. Most of the community discovery algorithms discussed in this section were designed with the implicit assumption that the underlying network is unchanging. This is in the case that in real social networks the relation among nodes changes by the pass of time and consequently their membership in different communities will as well as change [8].

3. EDGE BETWEENESS CENTRALITY ALGORITHM

In the traditional hierarchy methods, the structures with high degree joints are gradually created and develop in graph. These methods are prosperous in detecting communities only in limited cases. One of the main problems of these methods is that there may be nodes in the graph which connect to the other nodes only via an edge. These nodes by algorithm will have no place in any community. But it is clear that these nodes are belonged to the community through that mono-edge is connected to that the same community. To solve the problems of hierarchical methods Newman and Girvan [9] proposed a divisive algorithm for community discovery, using ideas of edge betweenness. In this method the border of community is delineated through the criterion of edge betweenness. The centrality of an edge is the number of the shortest path that exists among nodes that pass through this edge. In the other word, the centrality of edge means that what measure of edge exists across the paths between two collections of connected nodes. While conducting this method in each stage the edge having the most centrality will be deleted from the graph, then the centrality of other edges will again be computed. While conducting this method in each stage the edge having the most centrality will be deleted from the graph, then the centrality of other edges will again be computed. As a result of the repetition of this procedure, a collection of isolated categories will be created in graph that each of them is the representative of communities in graph. The reason for the recalculation step is as follows: if the edge betweenness are only calculated once and edges are then removed by the decreasing order of scores, these scores won't get updated and no longer reflect the new network structure after edge removals.

4. METHODOLOGY

This section deal with the introduction of a method applied for community clustering in the social network graphs. Of course the intended graph in our



suggested method is a kind of directed and weighted graph. The method introduced in this research study is based on a divisive algorithm. As well as, our suggested method is verbalized through the algorithm of edge betweenness centrality. Considering the issue that in social network edge is the representative of relationship between two individuals, it can be concluded that edges, the most frequently used to compute the shortest paths pass through them, are the bridges which much density is available on their two ends under the graph. The difference here creates in Newman -Girvan algorithm is that the value between two nodes will no longer be identical, but by accepting one fuzzy degree in the $[0,1]$ interval in proportion to other edges will be more valuable and consequently the proportion of this edge in the edge betweenness centrality will be different. It should be added that the degree of fuzzy membership is attributed to each edge on the premise of the measure of the relationship among individuals in social network. If the detecting of these edges is conducted by using the shortest paths under created graphs in the first stage is regressively continuing, almost the maximum of communities in graph can be detected. The stages of conducting the suggested method are as the following:

ALGORITHM 1: BREADTH FIRST SEARCH

1. The initial node called A, $d_a=0$ and gained the weight $W_a=1$.
2. Each node i in the neighborhood of A will gain the distance of $d=d_a+1=1$ and the weight of $W_i= W_a= 1$.
3. One of the three following choices will be performed for each node j in the neighborhood of node i :
 - If the node j has not receive distance to which the distance of $d_j= d_i+ 1= 1$ and the weight of $W_j= W_i$ will be dedicated.
 - If the node j already has received the distance and $d_j= d_i+ 1$, the weight of node will W_i be increased. It means that $W_j= W_j+ W_i$.
 - If the node j has already received and $d_j < d_i+1$ no operation will be performed.
4. From the stage (3) the algorithm will be repeated as long as no node is remained to investigate.

The weight attributed to node i , in fact, verbalize the number of independent paths of initial node to node i . These weights are necessary to compute edge betweenness centrality, for two connected nodes i and j (that j is more far that i from the resource A) a fraction of the shortest paths between i and j nodes that passes through i is determined via W_i/ W_j . Now, the following stages will be conducted to compute edge betweenness through the all shortest paths from resource A.

ALGORITHM 2: THE COMPUTATION OF EDGE BETWEENNESS

1. All the leaves under the title node B should be searched.
2. The amount of W_i/ W_j will be dedicated to the edges between i and j .
3. Now, the movement is performed from the farthest edges from the source A, it means lowest level to up, towards node A. The amount of one plus the total



number of graph on the neighboring edges (present lower to this edge without mediator) is dedicated to the edge between the node i to j and the conclusion will be multiplied by amount W_i/W_j .

4. The algorithm will be repeated from the stage (3) until it reaches node A.

At present, this procedure, for the total number of node (n), will be repeated as initial node and the gained conclusion in each stage to each edge will be pulsed (aggregated) to compute total betweenness by time for all edges. All the computations are again performed for all other edges after each time an edge is deleted. The dependence of this algorithm on many computations will result in the fact that it be only employed for the networks with ten thousands nodes and in the wider networks their growth of time will be intractable. Now, after the clustering of social network graph, the internal density of each cluster will be computed by considering formula (1). Naturally, determining the dense time of a cluster is performed by computing the density of graph. The under graph density gained by the cluster is considered as intra-cluster density. In the following formula the u and v are two intended vertexes in cluster G . The intra-cluster density measure is computed via the following formula:

$$\delta_{int}(G) = \frac{|\{\{v,u\}\}v \in G, u \in G|}{|G|(|G|-1)} \quad (1)$$

5. SIMULATION OF PROPOSED MODEL

This social network is the friendship relation among 200 student of IT University that according to the amount of the relation among the members, values 0.2, 0.4, 0.6, and 0.8 is attributed to each edge. The graph of this social network, by dedicating numerical figures to the individuals, is designed as the following via NodeXL software.

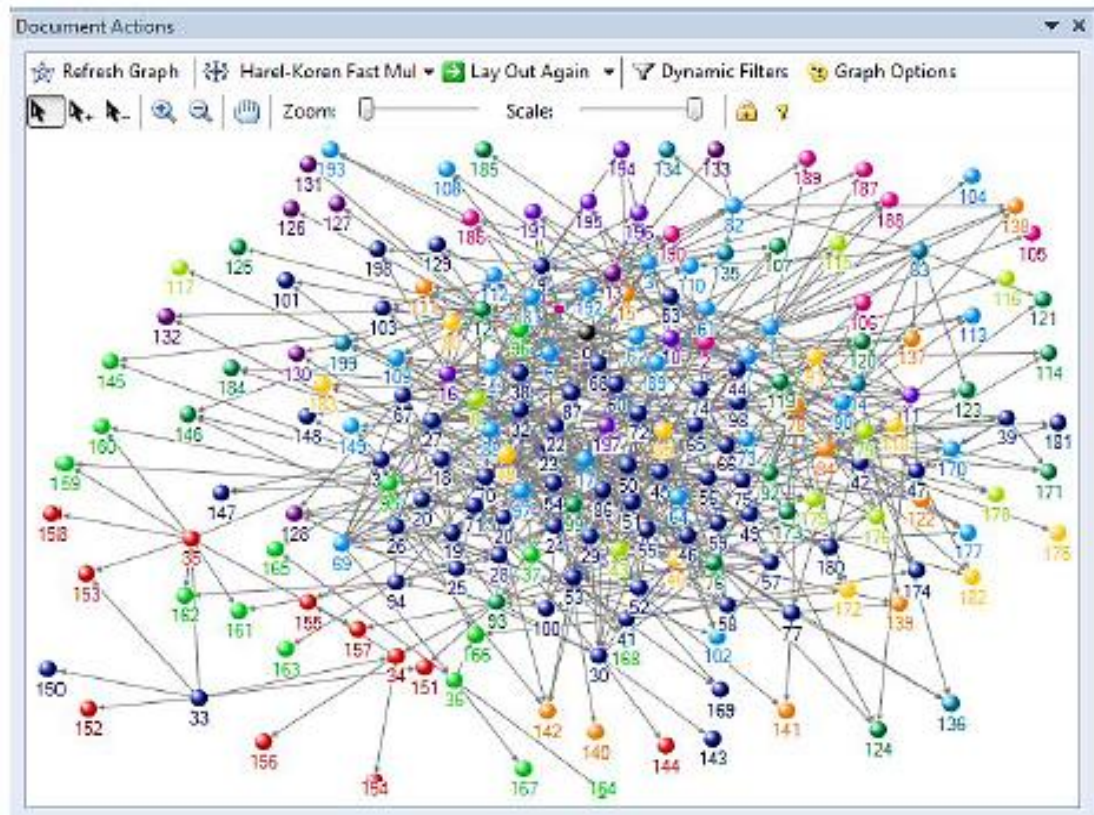


Figure. 1. The graph of communications in the network

The feature of this network is as following:

- The total number of the members of network is 200 n.
- The kind of edge (communications) is fuzzy directed and weighed.
- The number of communications among the members (the number of the edges in the network) is equal to 208.

After the designing the graph of collected data, the code relevant to execution is written in Visual Basic. Net language and are executed on the collection of data as the following:

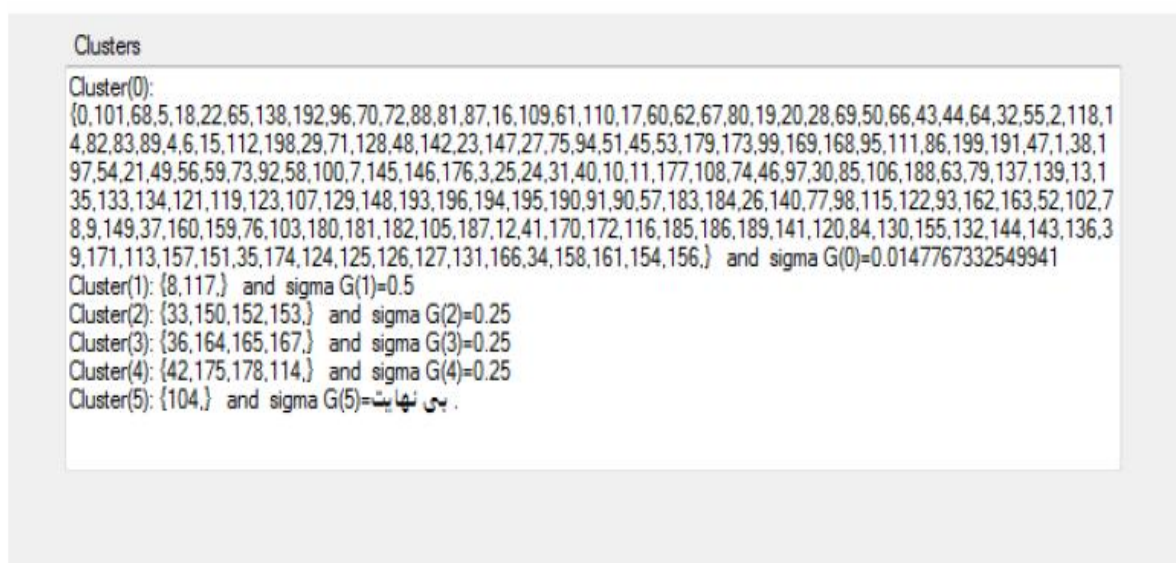


Figure 2. The results of the evaluation of gained clusters

Drawing attention to the execution of suggested method the nodes of the main graph are categorized into 6 grouping cluster.

A. THE COMPARISON OF CLUSTERING ALGORITHM

In this section the suggested method is compared with the previous method in the area of community detecting, especially the standard Newman and Girvan and the algorithm of Clauset and his associates which are hierarchical methods, and the conclusions of the research are gathered in the following table. The results of the two algorithms became simulated using NodeXL software (version 2012).



Table 1. The comparison of algorithms

The kind of algorithms	The number of gained clusters	The average density of the whole graph	The number of mono-node clusters
Neman-Girvan	47	0.12	23
Clauset-Newman-Moore	6	0.07	2
Fuzzy Edge Betweenness	6	0.21	1

B. ADVANTAGES AND DISADVANTAGES OF PROPOSED METHOD

- One of the advantages of this suggested method is the fact that in comparison with the previous methods, it has much reduced the number of the mono-node cluster, since the unnecessary clusters prevent from accurate detecting of the communities resulted from clustering algorithms.
- In this method the centrality of the whole graph is improved in proportion to the other methods.
- This method is very time- consuming because of long statistical computations in the complicated graphs.

6. CONCLUSION

In this article through modeling the social networks to a graph, in which the nodes are the same individuals or groups and the edges are the same communications between the individuals and groups, a huge graph containing so many numbers of nodes and edges is created. We could explicate a new algorithm to perform the clustering of the network and explore the structure of community containing nodes and edges. Also, drawing attention to the point that the nearer the density of the intra- clustering is to 1, the higher the quality of clustering. The findings of simulation and the evaluation of density gained from this algorithm signify the matter that the suggested algorithm is more efficient in smaller graphs, since the smaller the graph, the more it is liable that the detected clusters in the graph have a fewer node number and consequently, the intra-cluster density even with fewer number of edges will quickly desire to 1. We



could also reduce to much extent the number of mono-node clusters which have undesirable effects on creating earned clusters and the average density.

FUTURE WORKS: To detect the nodes located at high degree centrality or the nodes located to the high closeness centrality rather than the other nodes in social network and introduce them as the centers of clusters in the algorithm of edge centrality in Newman's algorithm. Also, the amount of the influence of a node, called Eigenvector centrality in the graph of a social network can be introduced and determined as the core of clusters and the leader node in the graph.

REFERENCES

- [1] U. Brands, D. Delling, M. Gaertler, R. Gorki, M. Hoofers, Z. Mikulski, D. Wagner 2008. On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering*. 20, 2, pp. 172-188.
- [2] R. Xu 2005. Survey of Clustering Algorithms. In: *IEEE Transactions on Neural Network*, vol. 16, no. 3:645-678.
- [3] M. Girvan, M.E.J. Newman 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences USA*. 99, 12, pp. 7821-7826.
- [4] L. Freeman 1979. Centrality in social networks: Conceptual clarification. *Social Networks*. Elsevier Sequoia S.A., Lausanne - Printed In the Netherlands. 2, 2, pp. 215-239.
- [5] M.E.J. Newman 2010. *Networks: An Introduction*. Oxford University Press; 1 edition. pp. 1-34.
- [6] A. Clauset, M.E.J. Newman, C. Moore 2004. Finding community structure in very large networks. *Physical Review E*, 70, 6, pp. 66-111.
- [7] P. Zhao, C. Zhang, S. Chatterjee 2011. A new clustering method and its application in social networks, *Pattern Recognition Letters*, Elsevier, 32, 2109-2118.
- [8] O. Green, D. Bader 2013. Faster betweenness centrality based on data structure experimentation. *International Conference on Computational Science, ICCS*. Available online at www.sciencedirect.com. Volume 18, pp. 399-408.
- [9] M.E.J. Newman, M. Girvan 2004. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):026113-026120.

This paper may be cited as:

Saed, N., Sadeghzadeh, M. and Yektaie, M. H., 2014. Presentation of Fuzzy Model to Compute the Edge Betweenness Centrality in Social Networks. *International Journal of Computer Science and Business Informatics*, Vol. 14, No. 2, pp. 130-138.