COMMUNITY DETECTION IN NETWORKS USING PAGE RANK VECTORS

V.Greeshma[1] and K.Suvarna Vani[2]

Department of Computer Science & Engineering V R Siddhartha Engineering College, Vijayawada, A.P. INDIA

ABSTRACT

Nodes in the real world networks organize in the form of network communities. A community (also referred to as module or cluster)is defined as where the links are denser inside the nodes and sparser outside the nodes in the network. Communities in the networks also overlap because the nodes may belong to different clusters at once. The task of detecting communities in networks becomes an open problem because of lack of reliable algorithms. In practice all the existing community detection methods work good for non-overlapping communities and fail to detect communities with dense overlaps. We developed a novel method for detecting communities by considering a single seed node. This method successfully captures the overlapping networks ranging from social to information and from biological to citation networks. We believe that the proposed system works well for the overlapping communities.

KEYWORDS: *Network Communities, Overlapping communities, seed node, random walk, page rank nibble.*

1. INTRODUCTION

The networks in the real world such as social, biological and information networks are organized in the form of graphs. A graph consists of nodes and edges. Nodes in the networks are organized into densely linked groups and are referred as network communities [1].For example, in biological networks communities are functional modules of interacting proteins, and in social networks, communities correspond to groups of friends that come from the same college or from same home town. A community (also referred to as cluster or module) is defined as where the links are denser inside the nodes and sparser outside the nodes of a network. Detecting communities is nothing but clustering a set of nodes in which the nodes may belong to the multiple communities at once, because the nodes in communities may share common attributes or properties that is because they may have relationship between them [2].In Fig 1: the schematic example of a graph with communities is shown.

Communities in networks are in the form of groups in which the users in the network share a common functional property and the aim of the community

DOI :10.5121/ijbb.2015.5401

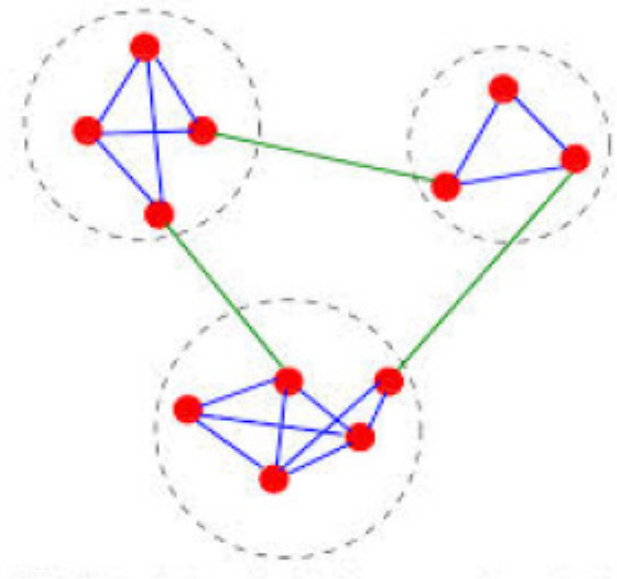

Fig. 1. A simple graph with three communities enclosed by dashed circles

detection is to find the sets of functionally related node from the undirected network. The model of network communities has evolved over time. The researchers started working on densely linked networks. Graph partitioning [3] , modularity [4] and betweeness centrality [1] are such algorithms for detecting communities for non-overlapping networks. These methods do not work well for overlapping communities.

There are two sources of data for clustering task. The first is the data about the nodes and their attributes. For example, if we take the social networking sites such as Facebook,Orkut the users profiles in it tell us which objects are similar and to which communities they belong to. The second source of data is that it comes from the network and the set of connections between the nodes. The users in the social networking sites form friendships with other users and in biological networks proteins interact with each other. The clustering method focuses only on these two sources of data. When considering attributes, clustering algorithms focuses only on the set of objects whose attributes are similar and leaves the connections between them. The community detection algorithms aim to detect communities based on the network structure and the links between the nodes.

In many networks the nodes may belong to multiple communities simultaneously this leads to the overlapping of nodes [5] . The overlapping nature of communities is that they have more external connections than internal connections. Overlapping cliques, articulation points and `The present overlapping community detection methods assume that the overlapping nodes are less connected than the non-overlapping parts. But this assumption went unnoticed.
This is because of the lack of reliable evaluation of ground-truth [6].

The ground-truth communities are nothing but the nodes in the network that share a common functional property. The ground truth examines the network communities whether they belong to

the real functional groups or not. We examine the sensitivity, quality and robustness of the communities. This is by using
the scoring function which scores a set of nodes based on their connectivity. If the scoring function is high then a set of nodes closely resembles the connectivity of the communities.

This paper is organized as follows: Section 2 summarizes the available research on the community detection. Section 3 describes the proposed system. Section 4 discusses Methodology. In Section 5 experiment results, Finally in Section 6 conclusion arrives.

2.RELATED WORK

Recently, several community detection algorithms have been proposed. These algorithms are based on clustering, spectral and modularity. Girvan et al. [1] proposed a method called edge betweenness. It is nothing but the number of shortest paths passing through an edge. This method gained popularity for divisive clustering methods. This clustering is based on the measure of similarity between different nodes. An agglomerative hierarchical clustering algorithms use bottom-up approach, as it starts vertices with separate clusters and ends up with a single unique cluster whereas in divisive algorithms it starts with top-down approach in which it starts with all the vertices as a unique cluster and separate them into a single cluster.

Andersen et al. [7] proposed a method called local graph partitioning using page rank vectors. In this method it finds a cut near a specified vertex defined, with a running time, which depends on the size of the cut, rather than the whole graph.

Newman, Mark EJ [4] proposed a method that is optimization of the quality function called as modularity. If this method is directly applied then the simulated annealing is costly. So, this method uses Eigen vectors a new matrix for the network and this is called as modularity matrix. This gives a spectral algorithm for community detection.

Radicchi et al. [8] proposed an improved version on Girvan and Newman algorithm. In this method edge betweenness is calculated and the edge is removed that is with the highest score. If the graph is partitioned then draw the corresponding dendrogram. Continue this process until no edges remained in the network.

Wang and Yuqin [9] proposed a method to find the community structure in complex networks by using k-means. In this method randomly select k nodes and these k nodes are considered separately as a community. The selected nodes are removed. Then select any node j from N, j and k-node and compute the first element from (d (G, in, j)). Find the minimum value and delete it from the set N. Repeat this process until N is empty.

S.Fortunato [10], the most popular method known as clique percolation method for detecting overlapping communities. This method is based on the concept of internal edges of a community and is used to form cliques due to their higher density. Cliques are the sub-graphs in which every node is connected to every other node.

Jure Leskovec et al. [5]proposed a model based community detection method known as Community-A_liation Graph Model that builds on bipartite-node community affiliation networks.

In this method by considering an undirected network, an affiliation graph is formed. Based on graph the clusters can be detected.

3.PROPOSED METHOD

The existing community detection methods, detect communities where the entire network is partitioned into clusters. The proposed method finds the communities near a specified vertex. This method works well for overlapping community detection. This method is based on the community-centric view where we want to discover all the members of the community of the node s. The main goal of this method is to find the communities at every point or node.

4 METHODOLOGY

The proposed algorithm finds the set of well-connected nodes around the nodes. This is achieved by using local partitioning method based on random walks [7]starting from the node s. Here we use the Page Rank-Nibble similar to that of Nibble which looks for a cluster containing the vertices that are closer to the node. The Page Rank-Nibble random walk method computes the personalized page rank vector with error< in time [11]. The nodes with highest page rank scores correspond to well-connected nodes around s. Now sweep technique is used in order to partition clusters from the vector. The nodes are arranged in the decreasing order based on the degree of the node. Then compute the community scoring function f (i.e., conductance) of the first j nodes, where j ranges from 1 to the number of non-zero entries in the vector. Conductance measures the fraction of outgoing edges to the set of nodes in the graph.

$$f(s) = \frac{|\,\partial(s)\,|}{min(vol(s), vol(\bar{s}))} \qquad (1)$$

If the value of the conductance is lower it forms a better cluster. After finding the conductance of all the vertices then detect the cluster with low conductance from all the sweep sets. In order to reduce the time for calculating page rank vector and perform sweep we can use approximation of it. The approximation algorithm computes an -approximate Page Rank vector [7] by calculating a random walk that only consider nodes that have more than d(v) probability in them, where d(v) is the degree of the vertex v. The resulting Page Rank vector has few non-zero entries. We can also consider sweep sets of certain size in order to find the cluster of specific size. This method is a parameter-free community detection method and this method is more reliable for detecting communities in dense overlaps.

4.1.Algorithm

Input: Graph G(V;E), starting vertex s, scoring function f.
Output: Clusters formation.

- Firstly we have to calculate the random walk scores from starting vertex in its neighborhood using page-rank nibble.
- From step1 a personalized page rank vector is calculated.
- The vertices are ordered in the decreasing value based on the degree of each vertex.

- Calculate the scoring function f (i.e., conductance) of the first j vertices,where j ranges from 1 to the number of non-zero entries in the vector.
- Identify the cluster with low conductance among all the sweep sets.

5.EXPERIMENTAL RESULTS

The experiments are conducted on a given dataset below.We proposed a method to detect communities from a given single vertex. We detect all the members of the communities that a single vertex belongs to. We find the conductance for all the sets. Then find the scoring function for all the nodes. If the value of conductance is low then it represents a good cluster. So detect the cluster among all sweep sets that has low conductance. This works well for both overlapping and non-overlapping communities.

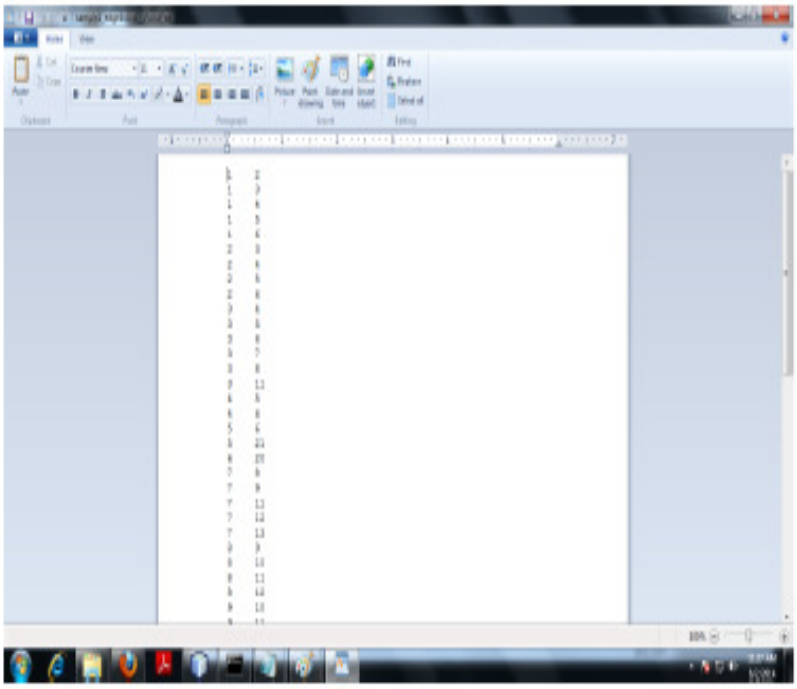

Fig. 2.

Input Dataset

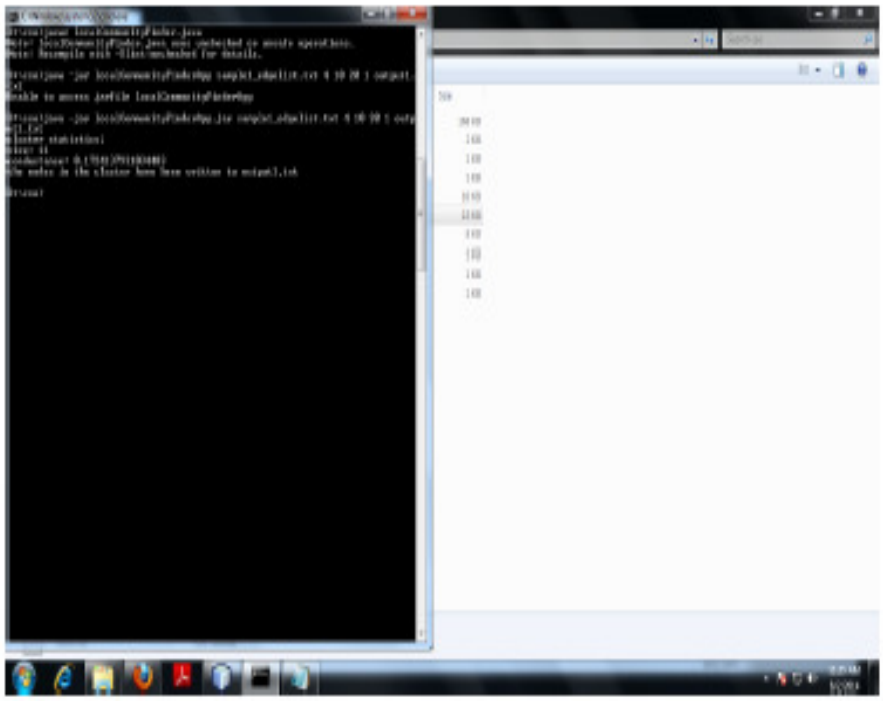

Fig.3.Executable Window

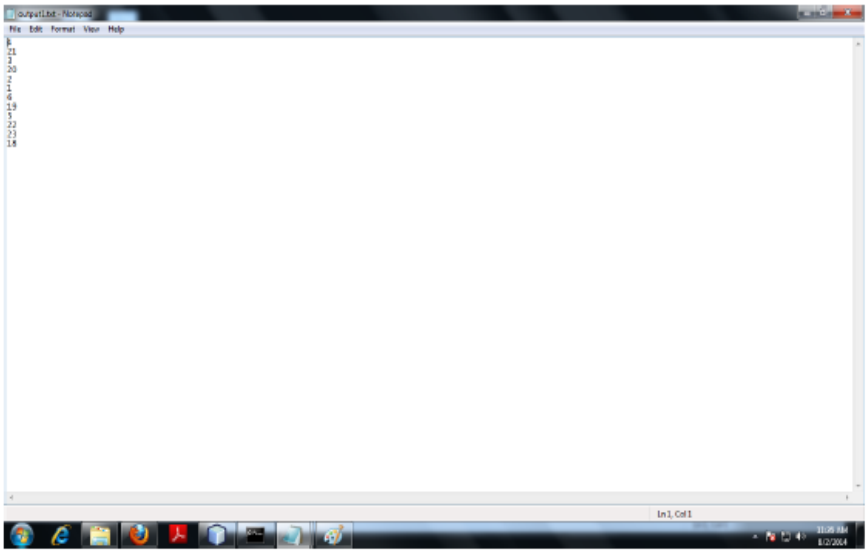

Fig.4.Cluster Detection

6.CONCLUSION

In this paper, community detection using page rank vectors has been presented. Here in this paper simple and effective algorithm for detection of local communities is proposed. The proposed method reliably detects the local communities in a network. If the good cluster exists in a network, it will be found by our method by looking at the right neighborhood. Our results show that by locally partitioning a network we can find communities whose vertices are functionally related to each other, and less related to other vertices in the network. This algorithm also works well for weighted networks i.e., the weights should be defined.

REFERENCES

1] Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks." Proceedings of the National Academy of Sciences 99.12 (2002):7821-7826.

2] Yang, J., McAuley, J., Leskovec, J. (2013, December). Community detection in networks with node attributes. In Data Mining (ICDM), 2013 IEEE 13th Interna- tional Conference on (pp. 1151-1156). IEEE.

3] Schaeer, Satu Elisa. "Graph clustering." Computer Science Review 1.1 (2007): 27-64.

4] Newman, Mark EJ. "Modularity and community structure in networks."Proceedings of the National Academy of Sciences 103.23 (2006): 8577-8582.

5] Jure Leskovec,Yang and Jaewon . "Community affiliation graph model for overlapping network community detection." Data Mining (ICDM), 2012 IEEE 12th International Conference on. IEEE, 2012.

6] Jaewon,Yang,and Jure Leskovec. "Defining and evaluating network communities based on ground-truth." Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. ACM, 2012.

7] Andersen, Reid, Fan Chung, and Kevin Lang. "Local graph partitioning using pagerank vectors." Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on. IEEE, 2006.

8] Radicchi, Filippo, et al. "Defining and identifying communities in networks." Proceedings of the National Academy of Sciences of the United States of America 101.9 (2004): 2658-2663.

9] Wang, Yuqin. "An Improved Complex Network Community Detection Algorithm Based on K-Means." Advances in Future Computer and Control Systems. Springer Berlin Heidelberg, 2012. 243-248.

10] Fortunato, Santo. "Community detection in graphs." Physics Reports 486.3 (2010):75-174.

11] Reid, Andersen and Kevin J. Lang. "Communities from seed sets."Proceedings of the 15th international conference on World Wide Web. ACM, 2006.