2nd International Conference on Recent Trends in Engineering Science and Management
International International Center, YMCA, New Delhi (ICRTESM-16)
20 February 2016, www.conferenceworld.in ISBN: 978-81-932074-3-7

A STUDY ON EFFICIENT COMMUNITY DETECTION TECHNIQUES IN SOCIAL NETWORKS

Arindam Roy[1], Priyanshu Thakuria[2], Gypsy Nandi[3]

Department of CSE &IT,School of Technology, Assam Don Bosco University, Azara Guwahati (India)

ABSTRACT

Presently, there are many algorithms and techniques used for detecting communities in social networks, but not all of them are well suited for various types of networks such as dynamic social networks. Detection of community in social networks is broadly classified into four types – node-based, hierarchy-based, network-centric and group-centric. Majority of these techniques use the concept of graph mining which is a technique enormously used for mining online social networks. In this paper, we study efficient community detection techniques by making a comparative study of few such algorithms that are often been used for this purpose.

Keywords: Divisive Hierarchy-Based Technique, Girvan-Newman, Network Centric, Node Similarity, Online Social Networks

I. INTRODUCTION

Social networking in the recent years has become an increasingly important area because of its unique ability to enable social contact over the internet for geographically backward areas. A social network can be represented as a graph, in which nodes represent users and links represent the connections among the users of the graph. Social communities have emerged as new networked systems and are being of current research interests. Such communities are formed based on working environment, family relationships, friendships, demographic regions, etc. In graph terms, a social community can also be considered as a clique, in which the community is a subset whose nodes or vertices are all adjacent to each other.

A community in a social network can also be defined as a collection of users considered as a group such that there are high interrelations or associations among members within the group. Communities help us understand the structure of given social networks and clarify the functions and properties of the networks. An important problem in the area of social networking is that of community detection. In case of community detection the main aim is to partition the network into dense regions of the graphs. A wide application of community detection is for generating recommendations based on how members within a community display similar tastes and preferences.

Fig. 1 below illustrates a simple example of a social network where there are mainly three communities formed based on similar tastes and/or preferences. Each community in the network is highlighted with a different color.

Detecting communities in a network also provides us meaningful insights to its internal structure as well as its organizational principle. Knowing the structure of network communities also provide helpful points of view to some

uncovered parts of the network thus helps in preventing potential networking diseases such as virus or worm propagation.

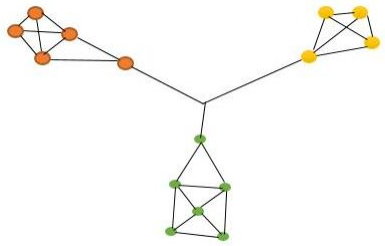

Fig 1: A simple social network showing communities (each community is colored differently)

The rest of the paper is organized as follows. Section II discusses about the standard community detection techniques that are currently often applied for such analysis. Section III discusses about the framework of the standard community detection techniques and algorithms. In section IV a brief conclusion of the paper is given along with scope of future works in this area of research.

II. DISCUSSION ON STANDARD COMMUNITY DETECTION TECHNIQUES

In [14], a survey on community detection techniques has been done in the context of online social networks (OSN). In OSN, nodes represent individuals and edges indicate relationships between them. In this survey various community detection methods for networks with static and dynamic nature are also discussed. Also in [15], a classification for community detection methods has been made based on the type and nature of social networks.

Overlapped community means that a node in the network may be affiliated to more than one community. After reviewing [16], it can be come to the conclusion that quality and scalability are the major issues in this problem area and also a measure that can be used for comparing and analyzing various such community detection techniques.

The standard community detection techniques can be broadly classified into four categories:

a. Node-centric Community Detection
b. Group-Centric Community Detection
c. Hierarchy-Centric Community Detection
d. Network-Centric Community Detection

In node-centric community detection, each of the nodes that are present in the graph has to satisfy certain properties. One of them is the complete mutuality cliques that have a complete subgraph in which all nodes are adjacent to each other [10].

The group-centric criterion requires the whole group to satisfy a certain condition. It considers connection inside a group as a whole. It is acceptable to have some nodes in the group with some low connectivity as long as the group overall satisfies certain requirements.[10]

Another line of community detection research is to build a hierarchical structure of communities based on network topology. There are divisive algorithms that can be applied for detection of the communities. One is the Newman-Girvan algorithm [4] where the betweenness for all edges is calculated in the network and the edge with the highest betweenness is removed.

The Network-centric community detection has to consider the global topology of a network. It aims to partition nodes of a network into a number of disjoint sets. Typically, network-centric-community detection aims to optimize a criterion defined over a network partition rather than over one group.[10]

There are also some other models and concepts of community detection techniques that have been proposed by several authors. Bayesian models [2] are used for detecting latent communities from a social network. For instance, two users engaging in conversation related to politics are more likely to be members of a community on politics. Quick Community Adaptation [3] can be an adaptive modularity-based method for identifying and tracing community structure of dynamic online social networks.

Again, a node with a neighborhood that is too small is labeled a noise point, unless it falls within the neighborhood of a core point, and then it is labeled a border point. Core points connect to other nearby core points to form the center of a community [7]. The structure of social networks is an important research area and has attracted much scientific interest. Algorithm for detecting communities can be found which considers influential nodes based on random walk. Random walk can explore the underlying network structure and this concept has also been employed to reveal communities structure in social networks [8].

A new approach called "mutual accessibility" is proposed [12] to discover communities in a social network environment. Experimental results on a real world social network [6] show that a simple thresholding method with edge weights based on node attributes is sufficiently to identify a strong network community.

III. FRAMEWORK OF THE STANDARD COMMUNITY DETECTION TECHNIQUES

In this section, we will discuss in detail two main standard community detection techniques used for social networks by considering the several advantages and simplicity of these techniques. The first one is the Girvan-Newman algorithm that detects communities by progressively removing edges from the original network and the second is the clustering based on the node similarity technique.

a). Girvan-Newman Community Detection Technique

In Girvan-Newman algorithm, weak edges are progressively removed from the network and then the connected components of the remaining network form the communities. Instead of trying to construct a measure that tells us which edges are the most central to communities, the Girvan–Newman algorithm focuses on edges that are most likely "between" communities.

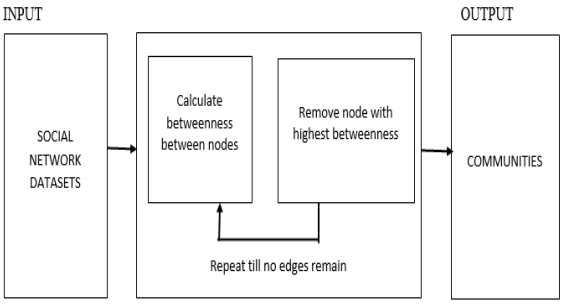

Fig 2: The standard process of Girvan-Newman Technique for community detection

As shown in Fig. 2 above, in case of Girvan-Newman algorithm, we can consider social network datasets as the input and then we calculate between ness between nodes and then remove node with the highest betweenness. This process is repeated until no edges remain. Finally, we get several clusters or communities of nodes as the output. Algorithm 1 below discusses the Girvan–Newman algorithm.

Algorithm 1: Girvan-Newman [4]

---

1. The betweenness of all existing edges in the network is calculated first.
2. The edge with the highest betweenness is removed.
3. The betweenness of all edges affected by the removal is recalculated.
4. Steps 2 and 3 are repeated until no edges remain.

In the first step of Algorithm 1, the betweenness of all existing edges in the network is calculated. Edge betweenness of an edge is defined by the number of shortest paths that pass along the edge. Then in the second step the edge with the highest betweenness is removed. In step 3 the betweenness of all the edges affected by the removal is recalculated. Step 2 and step 3 are repeated until all edges are considered and no edges remain.

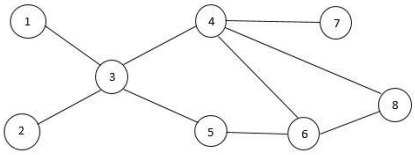

Fig 3: A graph with 8 nodes

For the above graph displayed in Fig. 3, the betweenness of edge (3, 4) is 3, as the shortest path from node 1 to 6,7 and 8 passes through the edge (3,4).

The edge with the highest betweenness generally acts as the bridge between 2 communities. Thus, when we remove the edges with highest betweenness, we get different communities.

b). Clustering-based Community Detection Technique

In network-centric community detection, we consider connections as features and then apply Jaccard similarity to compute vertex similarity. The Jaccard similarity coefficient of two vertices is the number of common neighbors divided by the number of vertices that are neighbors of at least one of the two vertices being considered. For instance, for the graph given below we can compute the Jaccard similarity as follows:

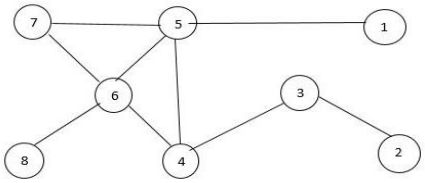

Fig 4: A graph with 8 nodes

For the above graph displayed in Fig. 4 the Jaccard similarity of nodes 3 and 5 would be:

$$\textbf{Jaccard Similarity (3,5)} = \frac{|\{4\}|}{|\{1,2,4,6,7\}|} = 0.2$$

Then, after applying Jaccard similarity, we use classical K-Means clustering technique to cluster the communities. This is explained in Algorithm 2 given below.

Algorithm 2: Clustering-based on node similarity [10]

---

1. Consider connections as features.
2. Apply Cosine or Jaccard similarity to compute vertex similarity
3. Apply Classical K-Means clustering algorithm

In Step 1 of Algorithm 2, all the connections or edges are considered as features. In the next step, we apply Cosine or Jaccard similarity to compute the vertex similarity. In step 3, we apply classical K-means clustering algorithm.

As shown in Fig. 5, for clustering-based community detection technique, we can take as input any social network dataset. Then we apply Jaccard similarity to compute vertex similarity. And then, we apply classical k-means technique. After all the steps are completed, we get the required communities.

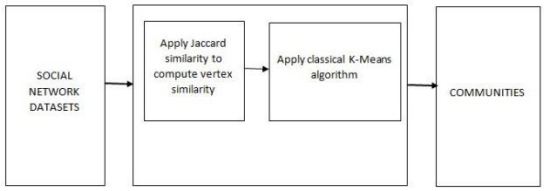

Fig 5: The standard process of clustering based on node similarity technique for community detection

c). Node-centric Community Detection Technique

The next technique that we will talk about is the node-centric community detection technique. In this technique the node has to satisfy certain properties. One of the properties is complete mutuality which is based on the concept of cliques. Clique is a maximum complete subgraph in which all nodes are adjacent to each other. For example in Fig. 6 below, nodes below, nodes 1,2,3 and 4 form a clique.

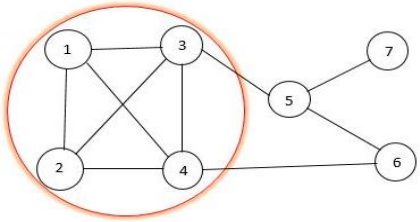

Fig 6: A graph with 7 nodes

Now if we are required to find the maximum clique, then in a clique of size k, each node maintains degree>=k-1. Nodes with degree <k-1 will not be included in the maximum clique.

Algorithm 3 below discusses the Maximum Clique technique used for community detection among nodes.

Algorithm 3: Maximum Clique [10]

---

1. We first sample a sub-network from the network.
2. Then we find a clique in the sub-network, by a greedy approach.
3. Say the clique above is size k, in order to find out a larger clique, all nodes with degree <=k-1 should be removed.
4. We repeat until the network is small enough.

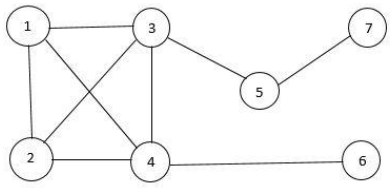

Fig 7: A graph with 7 nodes

For Fig. 7 above, if we want to find the maximum clique, then we can first find a clique {1,2,3}of size 3. In order to find a clique>3, we have to remove all nodes with degree <=3-1=2. Hence, we can remove nodes 5, 6 and 7. The resultant clique is shown in Fig. 8.

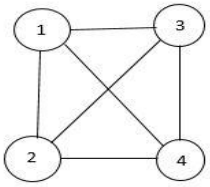

Fig 8: A clique consisting of four nodes

TABLE 1 below illustrates about the strengths and weaknesses of the three techniques discussed above.

TABLE 1: Strength and Weaknesses of Three Standard Community Detection Techniques

Community Detection Technique	Strength	Weakness
Girvan-Newman	Removes the nodes with highest betweenness	Is computationally costly
Clustering Based on Node	Takes common node attached to a	Difficult for large networks

Similarity	node	
Node-Centric Technique	Many nodes will be pruned	It will repeat until the network is small enough

IV. CONCLUSIONS AND FUTURE WORK

Social networking in the recent years has become an increasingly important application because of its unique ability to enable social contact over the internet for geographically backward areas. This paper covers the fundamental concepts of community detection and gives an overview of three standard community detection techniques that are available. In the future, new community detection technique can be developed that takes into consideration the limitations mentioned for these three community detection techniques.

REFERENCES

[1]. Guo-Jun Qi, Charu C. Aggarwal, Thomas Huang: Community Detection with edge content in social media networks. Data Engineering (ICDE), 2012 IEEE 28th International Conference.

[2]. Mrinmaya Sachan, Danish Contractor, Tanveer A. Faruquie, L. Venkata Subramaniam: Using Content and Interactions for Discovering Communities in Social Networks, Proceedings of the 21st international conference on World Wide Web, 331-340.

[3]. Nam P. Nguyen, Thang N. Dinh, Ying Xuan, My T. Thai: Adaptive Algorithms for Detecting Community Structure in Dynamic Social Networks, in Proc. IEEE INFOCOM, Shangai,Apr.2011,pp.2282-2290.

[4]. Deepjyoti Choudhury, Arnab Paul: Community detection in social network: An Overview, Volume: 02 Special Issue: 02, Dec-2013, Available @ http://www.ijret.org

[5]. Jaewon Yang, Julian McAuley, JureLeskovec: Community Detection in Networks with Node attributes, pp. 1151-1156,2013.

[6]. KarstenSteinhaeuser and Nitesh V. Chawla: Community Detection in a Large Real-World Social Network In: Liu, H., Salerno, J.J., Young, M.J.(eds.)Social Computing, Behavioral Modeling and Prediction, pp. 168-175. Springer US(2008).

[7]. Nathan Aston, Wei Hu: Community Detection in Dynamic Social Networks, Date of Conference 10-15 April 2011, Conference Location: Shangai

[8]. Chang Su, Yukun Wang, Lan Zhang: Community Detection in Social Networks Based on Influential Nodes, Journal of Software, Vol. 9, No. 9, September 2014

[9]. Lei Tang and Huan Liu ,Chapter 1: Social Media and Social Computing

[10]. Lei Tang and Huan Liu. Community Detection and Mining in Social Media, Morgan & Claypool Publishers, 2010.

[11]. Tsuyoshi Murata, Handbook of Social Network Technologies and Applications, Chapter 12 : Detecting Communities in Social Networks

[12]. The Stanford University By Jure Leslovec: Retrieved fromhttps://snap.stanford.edu/data/egonetsFacebook.html

[13]. The Dolphins Dataset Retrieved from http://www.casos.cs.cmu.edu/computational_tools/datasets/external/dolphins/index.php

[14]. AmitDhumal and PravinKamde. Survey on Community Detection in Online Social Networks, International Jounal of Computer Aplications (0975-8887) Volume 121-No.9, July 2015

[15]. Pourkazemi,M.,Keyvanpour,M. A survey on community detection methods based on the nature of social networks. Published in Computer and knowledge Engineering (ICCKE), 2013 3th International eConference , Oct. 31 2013-Nov.1 2013

[16]. Chintalapudi, S.R, Krishna Prasad, M.H.M.A survey on community detection algorithms in large scale real world networks. Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference. 11-13 March 2015