ISSN: 2321-7782 (Online)
Volume 2, Issue 2, February 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

# Design and Development of Overlapping Community Detection Algorithm Using Multi-Label Propagation

**G.T.Prabavathi**[1]
Assistant Professor in Computer Science
Gobi Arts & Science College
Gobichettipalayam
Tamilnadu - India

**Dr. V. Thiagarasu**[2]
Syndicate Member, Bharathiar University, Coimbatore &
Associate Professor in Computer Science
Gobi Arts & Science College
Gobichettipalayam
Tamilnadu - India

***Abstract:*** ***Social network communities exhibit an overlapping community structure. Detecting overlapping communities and overlapping nodes in social network is an active field of research. Label Propagation Algorithms uses only the network structure to detect communities and does not require any external parameters. Propagating multiple labels between nodes of network based on probability occurrence of labels helps to detect the overlapping communities effectively. A modified label propagation algorithm that uses multiple labels for propagation has been designed and developed to detect overlapping communities from social networks.***

***Keywords:*** ***Social networks, community detection, overlapping communities, label propagation***

## I. INTRODUCTION

Complex networks allow studying a given system by representing its components and their relationships with nodes and links respectively. A network is a structure made up of nodes, representing entities, and links or edges, representing relationships of interactions between entities. Networks are natural representation for various kinds of complex systems such as biology, computer science, sociology etc. Various networks like biological networks, computer networks, social networks and communication networks exhibit the property of community structure. A community in a network is a group of nodes within which the network connections are dense, but between which they are sparse (Girvan and Newman, 2002). The goal of the community detection algorithms is to find group of nodes of similar interest in a given network. The size of large real-world networks has forced the research community to develop scalable approaches that can be applied to networks with several millions of nodes and billions of edges. Lot of algorithms have been proposed to deal with community structure discovery based on the principles such as hierarchical clustering, graph clustering, optimization methods, spectral partitioning of the network etc. Community detection is a challenging and interesting area of research in the domain of complex networks.

A social community is formed in web by people having common ideas about a subject, sharing hobbies, working to gather etc. When the communities they form are modeled as a structure, it is represented as a complex network called as social network. At present, social networks have emerged as one of the popular means of communications and it describes a structure of relationship between individuals and interpersonal influences. Communities of social network can be friendship/family/colleague circles or groups of people sharing common interests or activities. Community detection has been widely used in social network analysis to study the behavior and interaction patterns of people within the network. Overlap is one of the characteristics of social networks, in which a person may belong to more than one social group. For example, a person joins different hobby groups in the real world or on the internet, a researcher may belong to multiple research groups etc. The diversity of people's interest and social interactions suggest that the community structures overlap based on the necessity, personal preferences and interests. Detecting communities and their overlap, in social network has received an enormous

 

amount of attention in recent years. Therefore, detecting overlapping communities from social networks has become an active research direction in the last decade.

This paper is structured as follows: Chapter 2 discusses about the relevant work in community detection, Chapter 3 discusses about the importance of multi-label propagation, Chapter 4 lists the design and development of the proposed algorithm and Chapter 5 concludes the paper.

## II. BACKGROUND STUDY

There are number of community detection algorithms in the literature to detect both disjoint and overlapping communities. A review of algorithms can be found in (Xie et al., 2013; Prabavathi et al., 2013a and 2013b). Clique Percolation algorithm (CPM) was the first overlapping community detection algorithm (Palla et al., 2005). Since then, numerous algorithms have been proposed with great improvements in time and efficiency. Label Propagation is a community detection technique which uses only the network structure to guide its progress and requires no external parameter settings. LPA belongs to the family of agent based community detection algorithms. The main idea behind Label Propagation Algorithms (LPA) is to propagate the labels of node throughout the network and form communities (Raghavan et al., 2007). LPA uses unique identifiers of nodes as labels and propagates the labels based on an agreement with the majority of the neighbor nodes, with ties broken uniformly, randomly. As the process continuous, nodes connected densely acquire the same label to form communities. At the end of the propagation, nodes connected with the same label form a community. When a node belongs to a single community it is disjoint and if it belongs to multiple communities it is overlapping. The main advantage of LPA is its near linear-time complexity. It runs linearly in the number of edges, thus linearly also in the number of nodes for sparse network. Since LPA algorithms are considered as fast and efficient, recently much research work has been progressed with Label propagation algorithms.

COPRA is the first algorithm proposed to detect overlapping community using label propagation which is very fast and allows each vertex to belong to multiple communities at a time (Gregory S 2010). Xie (Xie et al., 2011) proposed a fast Speaker-listener Label Propagation Algorithm (SLPA), which spreads labels according to dynamic interaction rules and maintains label distributions in the memory of each node. Due to random tie breaking strategy, it produces different partitions in different runs. To stabilize the propagation dynamics and to produce deterministic output, Xie et al. (2013) proposed LabelRank algorithm that relies on four operators namely propagation, inflation, cutoff and conditional update to detect disjoint communities. Each node keeps multiple labels received from its neighbors eliminating the need of tie breaking. A community is formed of nodes with same highest probability. Due to cutoff and inflation operators, the number of labels in each node monotonically decreases and drops to a small constant within few steps. Introduction of the new operator, conditional rule and stopping criteria preserves the speed of the LPA based algorithms. The output is deterministic because there is no randomness in the simulation. In this paper, a label propagation technique that uses multiple labels to detect overlapping communities is proposed.

## III. MULTI-LABEL PROPAGATION

In social network, a user can spread, accept or reject an opinion received from its neighbors. Labels of each node can be treated as opinions in LPA and propagating the labels is similar to opinion spreading in social network. The agreement between the nodes to spread the labels forms the basis for Label propagation algorithms. Opinions are chosen based on the opinions of the neighbor list, which in turn, form their opinions from their neighbors and so on. This resembles the characteristics of social network where a user adds the friend's friend into their friends list. The proposed algorithm uses multiple labels for propagation and accepts multiple labels from the neighbors and decides to retain the labels for updating based on certain rules.

The aim of this paper is to design and develop an algorithm to identify overlapping communities in the social network using multi-label propagation to improve the speed and quality of detected communities. In SLPA, each node can be a speaker or a listener depending on whether it serves as an information provider or a receiver. A node stores as many labels as it likes,

 

depending on the propagation experience in the stochastic processes which is driven by the underlying network structure. The algorithm shows an excellent performance in identifying both node and community level overlapping structures and is suitable for weighted, un-weighted, directed and undirected networks.

The proposed algorithm combines the features of SLPA and LabelRank to detect overlapping communities from static social networks. In Speaker-listener Label Propagation algorithm, a speaker sends only one label to the listener. In the proposed algorithm, multiple labels are sent by the speaker during the propagation process. A listener accepts only one label from the set of labels received from the speakers based on probability of occurrence of labels. In the proposed algorithm, the listener accepts one or more labels and the probabilities of accepted labels are updated based on updating criterion. The node that provides information follows certain speaking rules to decide which labels to be send to its neighboring nodes and the listening node decides whether to accept the label or not based on listening rule. Based on the speaker and listening rule, the dynamics of output in the algorithm is determined. As propagation progresses, the probability of specific labels increase depending upon the frequency of occurrence. The decision made by a listening node to accept, reject or update the label is based on the probability occurrences of that label received in the past as well as the current probability. Taking into account both past and current probability of labels, accounts for the improvement of the algorithm. The propagation process continues until stop criterion is satisfied or until maximum iteration is reached. At the end of the entire propagation, post processing is done on the labels based on a threshold value. After the post processing, when a node contains multiple labels, overlapping communities occur. The main focus of this paper is on the development of an algorithm to detect overlapping communities by propagating multiple labels between nodes of social network.

## IV. Algorithm Design And Development

The overall design of the modified multi-label propagation algorithm is given below:

**Input: Network *G = (V, E)***

**Output: Community structure *C = {C1, C2…}***

*t*: user defined maximum iteration

*r*: threshold for post processing

1. Self loop is added to the adjacency matrix
2. Label distribution is initialized for all nodes, such that equal probabilities are assigned for each neighbor
3. Repeat *t* times or until stop criterion satisfied
    - An unvisited node is taken as a listener in a specific order
    - Each neighbor of the selected node sends a label or multiple labels based on the probability of occurrence (frequency) of labels
    - If the listener satisfy the update condition, it updates the probability of one or more labels and the listener is marked as visited
4. Post process the labels in the memory based on the threshold *r* and form the communities. If a node contains multiple labels, overlapping communities occur.

The Algorithm steps are listed:

**[A,n]=loadthenetwork()**

**Step 1 : Pre-Processing**

*For* i=1:n

 

$A_{ii=}1$

$K_{i=}$Neighbors(i)

*End for*

**Step 2 : Initializing Probability of nodes**

*For* i=1:n

*For* j=1:n $P_{ij=}1/K_i$ *End for*

*End for*

**Step 3 : Propagation**

*Repeat*

*For* i=1:n

Listener=i

Speakers=K(Listener)

*For* j=1:Speakers.length

ReceivedLabels(j)= SpeakerRule(Speakers(j))

*End for*

[SelectedLabels,Prob]=Listener.ListenerRule(ReceivedLabels)

ConditionalUpdateRule(Listener,SelectedLabels)

*End for*

*Until* stop criterion=*True* or *t* times

**Step 4 : Post processing**

*For* i=1:n

*For* j=1:n

*if* $P_{ij}$<r then $P_{ij}$=0

*endif*

*End for*

*End for*

The updating of labels is done asynchronously. In asynchronous updating, a node at $t^{th}$ iteration updates its label based on the labels of its neighbors taken from iteration *t* as well as *t-1*. At the beginning of the algorithm, there are n different labels and as iteration progresses the number of labels reduces. Community detection algorithms are normally tested with benchmark graphs or real-world networks. The proposed algorithm should be tested with synthetic and real-world networks and evaluated with overlapping quality measure to assess the performance of the detected qualities.

## V. CONCLUSION

Detection of overlapping communities in social networks is necessary for realistic social analysis. Label propagation algorithms are fast and efficient which runs linearly with the number of edges in the network and requires only local information. In this paper, a modified label propagation algorithm which propagates multiple labels between speaker and listener has been proposed and designed to detect overlapping communities in social networks. The algorithm should be tested both in benchmark networks and real-world to assess the quality of the output communities.

## References

1. Girvan M., and Newman M.E.J, Community structure in social and biological networks, Proceedings of the Natural Academy of Sciences, Vol. 99, 2002, pp. 7821–7826.

 

2. Gregory S., Finding overlapping communities in networks by label propagation, New Journal of Physics, 2010, Vol. 12:103018.

3. Lancichinetti A., and Fortunato S., Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities, Physics Review E., 2009, Vol.80:016118.

4. Palla G., Derenyi I., Farkas I., and Vicsek T., Uncovering the overlapping community structure of complex networks in nature and society, Nature, Vol. 435(1), 2005, pp. 814–818.

5. Prabavathi G.T., and Thiagarasu V., A Review on Overlapping Community Detection Algorithms, International Journal of Applied Research and Studies, Vol. 2(3), 2013a, pp.347-352.

6. Prabavathi G.T., and Thiagarasu V., Overlapping Community Detection Algorithms in Dynamic Networks: An Overview, International Journal of Emerging Technologies in Computational and Applied Sciences, Vol. 6(4), 2013b, pp. 299-303.

7. Raghavan U.N., Albert R., and Kumara S., Near linear time algorithm to detect community structures in large-scale networks, Physics Review E, 2007, Vol. 76:036106.

8. Xie J., Szymanski B.K., and Liu X., SLPA: Uncovering Overlapping Communities in Social Networks via A Speaker-listener Interaction Dynamic Process, Proceedings of IEEE ICDM Workshop on Data Mining Technologies for Computational Collective Intelligence, 2011, pp. 344–349.

9. Xie J., Kelly S., and Szymanski, B. K., Overlapping Community Detection in Networks: the State of the Art and Comparative Study, ACM Computing Surveys, Vol. 45(4), 2013.

10. Xie J., and Szymanski B. K., LabelRank: A stabilized Label Propagation Algorithm for Community Detection in Networks, Proceedings of IEEE Network Science Workshop, 2013, pp.138-143.

## AUTHOR(S) PROFILE

**G.T.Prabavathi** received her Master Degree in Computer Science from Gobi Arts & Science College, Gobichettipalayam in 1993, M.Phil in Computer Science from Bharathiar University, Coimbatore in 2002. She is presently working as Assistant Professor in Computer Science at Gobi Arts & Science college since 2000. She has completed two UGC sponsored research projects. Her current areas of interest are Social Network Analysis, Multi-Agent Systems.

**V.Thiagarasu** received his Master Degree in Mathematics from Gobi Arts & Science College,Gobichettipalayam in 1985, M.Phil and Ph.D in Computer Science from Bharathiar University, Coimbatore in 1996 and 2010. He is presently working as an Associate Professor in Computer Science, Gobi Arts & Science College since 1989. He has completed two UGC sponsored research projects during 2004 and 2012. His current research areas are: Networking, Scheduling, Multi-Agent Systems, and Data Mining. He has published 39 research papers in National and International Journals.

 