



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Overlapping Community Detection Algorithms in Dynamic Networks: An Overview

Ms. G.T.Prabavathi¹, Dr. V. Thiagarasu²

¹Assistant Professor in Computer Science, Gobi Arts & Science College, Gobichettipalayam, India

²Syndicate Member, Bharathiar University, Coimbatore & Associate Professor in Computer Science, Gobi Arts & Science College, Gobichettipalayam, India

Abstract: Many real-world networks, especially social networks, exhibit an overlapping community structure. Identifying overlapping communities and overlapping nodes is a recent and increasingly active field of research. Dynamic complex networks model the evolving relationships between communities. Significant attention is attracted to the literature of community detection in complex network. Detecting and tracking communities in a dynamic network where changes arrive as a stream is a challenging issue in social network. Many algorithms have been proposed to identify and detect the dynamic communities in complex networks. This paper provides an overview of overlapping community detection algorithms in dynamic networks.

Keywords: Overlapping communities, community detection, dynamic networks, social networks

I. INTRODUCTION

This paper provides an overview of various algorithms that detects overlapping communities in dynamic networks and also lists the algorithms that can be extended to dynamic networks. In complex networks, community structure indicates groups of vertices such that vertices within a group are much more connected to each other than to the rest of the network [Lancichinetti A. and Fortunato S 2009]. Identifying community structure is one of the key properties of social networks since it depicts the behavior and interaction patterns of people in the social network. This paper attempts to study communities from the social network point of view. Social communities are very close to the traditional definition of community, i.e. communities are group of nodes within which the connections are dense and between which the connections are sparse.

The networks for which communities to be identified can be either static networks or dynamic networks. In static network, the graph is derived from a snapshot of data at a particular time. Communities in static network provide meaningful groups. But, in complex networks, communities are not static; they can change or even disappear in time. The dynamic features of community are also important in the study of complex networks. Many dynamic algorithms find static communities at each time stamp and evolutionary characteristics are introduced to explain the difference between the community structures over time which results in high temporal variations. But, the goal of community detection in dynamic networks should be to track community evolution and to identify their dynamics in a unified framework.

Both, in static and dynamic networks, communities may or may not overlap depending upon the nature of the network. Taking a personal social network into consideration, the user belongs to several communities: for example family, friends, co-workers and so on. These communities can be identified by density of connections in between the people of these communities which may overlap with each other with respect to the level of interest between the users. So, to understand and analyze the structure of social network, overlapping communities should be detected. In the recent literature [Lancichinetti et al. 2009; Gregory 2010], the topic of overlapping community detection has attracted attention in the data mining area.

This paper reviews the algorithms that attempts to identify overlapping communities in complex network particularly in the perspective of dynamic social network. This paper is structured as follows. Chapter 2 discusses about the overlapping communities. The characteristics of dynamic communities are discussed in Chapter 3. Various algorithms used for identifying overlapping communities in dynamic network are listed in Chapter 4. Chapter 5 concludes the review.

II. Overlapping communities

Real networks like social networks, biological networks and semantic networks exhibit the overlapping property in communities, as they reveal the internal structure of large networks. Particularly, in social network explicit

formation of groups is based on user's interests and preferences. When a person is involved in several communities such as family members, friends, co-workers, then community overlapping occurs. Contact based social networks might change from day to day due to people's varying activities. An individual can belong to multiple communities with respect to the simultaneous interactions with multiple groups. This feature makes overlapping as significant characteristics of many complex networks particularly in social networks. The overlapping nodes play a key role as an interface between communities, and it is interesting to study the community formation of these nodes, because it reflects the dynamic behavior of individual user in a network.

Clique Percolation Method (CPM) which was proposed by Palla *et. al.* in 2005 was the first algorithm to detect overlapping communities in complex networks. This method is based on rolling k -clique (a complete sub graph of k nodes) over the network through other cliques with $k-1$ common nodes. Since one node can participate in more than one community, overlap naturally occurs. Many algorithms that detect overlapping communities do not take into account their evolution over time. An efficient static overlapping community detection algorithm detects both individual overlapping nodes and overlapping communities, whereas dynamic overlapping community detection algorithms should not only detect overlapping nodes and communities but also take into account their dynamic evolution.

III. Dynamic Networks

There is a high demand for an efficient community detection algorithm that is able to handle large amount of dynamic data on a daily basis due to rapid emergence of large-scale online social networks. There exists several community detection algorithms based on static and global view of network which uses local connections between nodes as structural information. But in complex networks, the temporal correlation between different snapshots of data should be considered to detect dynamic communities. In a dynamically evolving network, where changes arrive as a stream and new data come in continuously, detecting communities is a challenging issue. In the past few years, many dynamic community detection methods have been designed which fall within two categories: incremental community detection methods which is suitable for data which is evolving in real time and offline community detection where all the changes of the network evolution are known *a priori*. A good dynamic community detection algorithm must be able to detect robust communities dynamically and identify evolutionary relationships simultaneously. It should also estimate the number of communities in the dynamic network without human intervention, while handling node addition and node attrition naturally.

Few algorithms represent dynamic network through an evolution of static networks in time stamps. These algorithms follow a two-stage approach. In the first stage, communities are identified from the static network at each time stamp and then in the second stage evolutionary relationship between the existing communities at each time stamp is calculated. But, this approach is longitudinal in nature, rather than dynamic and also induces a significant loss of information. It also becomes inappropriate if some intricate temporal relationships occur, such as some nodes which were part of a community at some point are not anymore after some time, but underlying group remains the same. Moreover data in real world are often noisy in nature. So, a two-step approach results in unstable community structures. Instead of computing communities on each snapshot independently, algorithms that incrementally update communities automatically are effective in monitoring huge data streams like Internet traffic or social interactions. This approach concentrates on the transition or change of community memberships, which is the key to the analysis of dynamic social network.

The main phenomena that occur during the life time of a community are growth, reduction, fusion, split, birth and death [Palla *et al.* 2005]. When a new community emerges without predecessor, it is referred as birth and when a community disappears without successor, the community dies. When several communities join together to form a new community the process is called merging and a community can split into several new communities. A community grows when a community gains new member and Contracts when it loses members. There are very few algorithms that deal with overlapping communities as well as these phenomena's. The aim of this study is to examine the algorithms based on overlapping community detection in dynamic complex network particularly in social network.

IV. Overlapping Community Detection Algorithms in Dynamic network

Many algorithms have been proposed for finding both overlapping and non-overlapping communities on static networks. Complex networks have been shown to demonstrate periodic dynamics. Detecting communities on dynamic networks, both on overlapping and disjoint structures is an emerging area of interest in complex networks. This chapter gives an overview of overlapping community detection algorithms in dynamic network and non-overlapping algorithms that can be extended to detect overlapping communities.

Lin Y-R, Chi Y, Zhu S, Sundaram H, and Tseng B.L., proposed FacetNet, a unified framework for analyzing communities and their evolution in dynamic networks. In this framework, the community structure at a given timestamp t is determined both by the observed networked data at t and by the prior distribution given by historic community structures. Instead of assigning a hard community partition to an individual who usually participate in multiple communities at the same time, the author uses a soft community memberships which provides more details about the author's participation in multiple communities. A stochastic block model is used for generating communities and a probabilistic model based on the Dirichlet distribution is applied for capturing the community evolutions. Due to these probabilistic models, soft community memberships are assigned naturally to nodes and it overcomes the problems faced by the parameters which exist in many dynamic algorithms. Total complexity is $O(\ell m)$ where ℓ is the number of non-zero elements in the sparse timestamp graph and m being the number of communities.

[Cazabet et al. 2010] developed iLCD (Intrinsic Longitudinal Community Detection) algorithm, which detects both static and temporal communities. This algorithm updates the communities by adding a new node if it's number of second neighbors and robust second neighbors are greater than expected values. The similarity between two communities is measured as the ratio of nodes in common. If the similarity between two communities is high, then the communities are merged. A new community is created, if the minimum pattern is detected by the edges. iLCD is based on two parameters size of the minimal clique which effects the creation of new communities and the threshold for community merging. This algorithm is not sufficient when new nodes are added or removed, or when existing edges are removed from the dynamic network. The complexity of iLCD is $O(nk^2)$ in general which depends on community structures and its parameters.

Nguyen N.P., Dinh T.N., Dung T., and Thai M.T. [2011] proposed DOCA (Detecting Overlapping Community Algorithm) which aims to classify network nodes into different groups by first identifying all possible densely connected parts of the input network. At second stage it tries to merge highly overlapped communities. In the next step, to ensure that all nodes are properly labeled, the algorithm revisits unassigned nodes to either classify them as outliers or to group it into a community. These three procedures are executed sequentially to discover network overlapping. The algorithm does not depend on the knowledge of the whole network as it works locally and scales well with the network's size and topology.

Many research articles have been published regarding Label Propagation Algorithms (LPA) which has been an active research area in community detection. LPA which was first proposed by Raghavan et. al. in 2007, is a simple opinion spreading model with many competing opinions among nodes. LPA algorithms use the labels (an artificial unique id) for each node and the labels are propagated based on an agreement with the majority of its neighbors. At the end of the process, nodes connected with the same label form a community. LPA algorithms are considered as fast and efficient. But, there are very few algorithms that can detect overlapping communities using LPA.

Leung [Leung et al 2009] initialized the node with a unique label and progressed by allowing each node to adopt the label most popular among its neighbors which produces densely connected groups of vertices from the current consensus. In order to avoid the monster community structure, the author proposed a hop attenuation parameter by initializing the label of each node by a score of 1.0 which declines by a given value in every propagation step. When the label score is smaller than a given threshold, it will stop propagating. This technique will prevent monster community, but it may damage some reasonable large communities because hop attenuation parameter limits the radius of community structure.

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]. This method allows each node to obtain multiple community labels. Each node updates its belonging coefficients by averaging the coefficients from all its neighbors in a synchronous fashion. For low overlapping density networks, COPRA offers better performance. But, it produces a number of small size communities in some networks. Since it adopts a global vertex-independent parameter (v), if there are some vertices with various numbers of community memberships in a network, it will be hard to choose a suitable value of v . In a network where most vertices are non-overlapping and a small part of vertices are overlapping with large community memberships, then if v is small, the algorithm will make it hard to identify overlapping vertices. If v value is large, then some non-overlapping vertices are identified as overlapping vertices. The time complexity is $O(vm \log(vm/n))$ per iteration, where parameter v controls the maximum number of communities with which a node can associate, m and n are the number of edges and number of nodes respectively.

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. In SLPA, each node can be a speaker or a listener depending whether it serves as a information provider or consumer. 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 significant characteristics of SLPA are (i) a node accumulates knowledge of repeatedly observed labels instead of erasing all but one of them (ii) no knowledge about the number of communities is required. The algorithm shows an excellent performance in identifying both node and community level overlapping structures. It is suitable for weighted, un-weighted, directed and undirected networks, but due to random tie breaking strategy, it produces different partitions in different runs, which is not desirable in applications like tracking the evolution of communities in a dynamic network. The time complexity is $O(tm)$, linear in the number of edges m , where t is a predefined maximum number of iterations.

To overcome the random tie-breaking problem, the author [Xie et. al. 2013] stabilized LPA and extended Markov Cluster Algorithm (MCL) approach that resulted in LabelRank algorithm. It stores, propagates and ranks labels in each node and relies on four operators namely propagation, inflation, cutoff and conditional update to stabilize the propagation dynamics. 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. Initialization of nodes takes $O(m)$, adding self-loop takes $O(n)$, each of the four operators takes $O(m)$ on average making the running time of LabelRank as $O(m)$, linear with the number of edges m .

In 2013 Xie J. and Szymanski B.K. extended LabelRank to LabelRankT with one extra conditional update rule. The main idea is to update only the nodes that are changed between two consecutive snapshots. The previous snapshot is used for inferring the dynamics in the current time step. It maintains the previous partitioning of communities and dynamically updates nodes that are involved in changes. Due to this new conditional update rule, this algorithm suits best to dynamic networks in which changes arrive as a stream. When a new edge is added in the stream, LabelRankT updates the nodes that are attached to this edge only. It gives the same performance of other static algorithms like MCL and Infomap, but with lower computational cost. It is faster than other dynamic detection algorithms like facetNet and iLCD. It is suitable for applications like wireless sensor networks and mobile ad hoc networks. The overall complexity for detecting evolving communities between two consecutive snapshots is $O(Tm)$, implying $O(m)$ in general where T is the number of iterations. LabelRankT can be extended to detect overlapping communities by integrating it with SLPA.

An overview of various algorithms that detects overlapping communities in static network and which can be extended to dynamic networks is listed in this chapter. Label Propagation algorithms are gaining more importance in the field of detecting communities from complex networks.

V. Conclusion

The detection of communities within complex networks is a challenging problem. When nodes in network are allowed to exist in multiple communities, which are more likely in social network, then detecting overlapping communities and overlapping nodes is an interesting area of research. Many algorithms are developed to detect both overlapping and non-overlapping communities in static network, but there are few such techniques in dynamic networks. Most of the work to detect communities in dynamic network use snapshots of data from static network and compare them. But, these algorithms do not take into account the temporal correlations between snapshots. An efficient overlap detection algorithm should not only detect overlapping nodes but take into account their dynamic behaviors. This paper gives an overview of such algorithms and lists the incremental community detection algorithms that can be extended to dynamic network to detect overlapping communities.

VI. References

- [1] Cazabet R., Amblard F., and Hanachi C, "Detection of overlapping communities in dynamical social networks", In Proceedings of SOCIALCOM, 2010, pp. 309–314.
- [2] Gregory S , "Finding overlapping communities in networks by label propagation", New Journal of Physics, vol. 12(10), 2010, pp.103018.
- [3] Lancichinetta A., Fortunato S, "Community detection algorithms: A comparative analysis", Physics Review E, Vol. 80, 2009, 056117.
- [4] Leung I.X.Y, Hui P., Li P., and Crowcroft J., "Towards real-time community detection in large networks", Physics Journal Rev., Vol. 79, pp. 066107, 2009.

- [5] Lin Y-R, Chi Y, Zhu S., Sundaram H, and Tseng B.L., “Analyzing communities and their evolutions in dynamic social networks,” ACM Trans. On KDD, vol. 3, no. 2, pp. 8:1–8:31, Apr. 2009.
- [6] Nguyen N.P., Dinh T.N., Dung T., and Thai M.T., “Overlapping community structures and their detection on social networks”, SOCIALCOM, 2011.
- [7] Palla G, Derenyi I, Farkas I, Vicsek T, “Uncovering the overlapping community structure of complex networks in nature and society”, Nature, Vol. 435(7043), 2005, pp. 814–818.
- [8] Prabavathi G.T., Thiagarasu V., “A Review on Overlapping Community Detection Algorithms”, iJARS, vol. 2(3), 2013, pp.347.
- [9] Raghavan U.N, Albert R., and Kumara S, “Near linear time algorithm to detect community structures in large-scale networks, Physics Review E”, Vol. 76(3),2007, pp. 036106.
- [10] Xie, J., Szymanski, B. K., and Liu X, “SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process”, In Proceedings of IEEE ICDM Workshop on DMCCI, 2011, pp. 344–349.
- [11] Xie J., Kelly S., and Szymanski, B. K.,” Overlapping Community Detection in Networks: the State of the Art and Comparative Study”, arXiv: 1110.5813v4[cs.SI], 2012.
- [12] Xie J. and Szymanski B.K., “ Labelrank: A stabilized label propagation algorithm for community detection in networks”, In Proceedings of IEEE Network Science Workshop, West Point, NY, pages 138–143, April 2013.
- [13] Xie J., and Szymanski B. K., “LabelRankT: A decentralized online algorithm for detection of evolving communities”. arXiv:1305.2006v2 [cs.SI], May 2013.