



# **A Review on Community Detection in Dynamic Social Networks**

Dr. G.T.Prabavathi<sup>1</sup>, T. Sreeshanmugapriya<sup>2</sup>

Associate Professor, Dept. of Computer Science, Gobi Arts & Science College, Gobichettipalayam, TamilNadu, India <sup>1</sup>

Research Scholar, Gobi Arts & Science College, Gobichettipalayam, TamilNadu, India. <sup>2</sup>

**ABSTRACT:** The identification of the community structure in the social network is an important issue in many domains. Community detection is an interesting area of research in the domain of complex networks. Due to the evolving nature and diverse interactions in complex networks, the topological changes in the network is quiet interesting area of study. Various algorithms have been proposed in static and dynamic social networks. Social dynamics has interesting applications in real life problems. This paper presents an overview of algorithms developed for detecting community detection in dynamic complex networks.

**KEYWORDS:** Community detection, Social Networks, Dynamic Networks, Overlapping Communities.

## **I. INTRODUCTION**

A social network (SN) is a social structure made of individuals (or organizations) called "nodes," which are tied (connected) by one or more specific types of interdependency. Social networks can be represented as a graph  $G = (V, E)$  where  $V$  is the number of nodes present in network and  $E$  is the edges used to connect these nodes. People present in different groups form different communities. Community is a group of nodes that have some common properties and have common role in organization. Group of nodes are more densely connected if they belongs to the same community and less likely to be connected if they are not the members of same community. A community in a network is a group of nodes within which the network connections are dense, but between which they are sparse [7]. Communities of social network can be friendship/family/colleague circles or groups of people sharing common interests or activities.

Social networks are categorized as static and dynamic social networks. Detecting communities in static networks is relatively a simple task. Many algorithms exist to accomplish this [4]. Detecting community structure in a dynamic social network has gained more importance recently. As the network grows in size, formation of new communities takes place and previous communities dissolve. Due to the evolving nature of social networks, communities may become denser and lead to the failure of existing static community detection methods. So, researchers are concentrating more on the dynamic aspect of social network.

Community can be disjoint community where a node belongs to single community; disjoint community is also known as crisp assignment, where binary relationship is being held between a node and a community. A node can belong to at-most one Community and at-least zero community (none) [3]. Overlapping communities are community where a node may belong to more than one community. Overlapping community is an interesting structure to analyze in the social network dynamics. In this paper an attempt has been made to review few algorithms that detects either static or overlapping communities in dynamic social network. Section 2 discusses about the necessity of community detection and Section 3 highlights the importance of dynamic nature of networks and in Section 4 few algorithms in dynamic social networks have been reviewed.

## **II. COMMUNITY DETECTION**

Many systems in nature can be modeled as networks. Examples include smart phone communications, online social networking sites, animal sightings, and molecular interactions. One phenomenon that has been frequently observed across different kinds of networks is the presence of latent community structures. Community structure is a set of communities present in network. It is represented as  $C = \{c1, c2, c3, c4...ck\}$  where  $C$  is the community structure



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

and  $c_1, c_2, c_3, \dots, c_k$  are communities. Community structures are quite common in real networks. Social networks include community groups based on common location, interests, occupation, etc. In social network extracting such community structures help to study the overall structure of network and the insights in the network.

Main goal of community detection algorithm is to find the group of nodes of interest in a network. There are numerous algorithms for detecting static communities from social networks. 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[4]. There are many complex networks which consist of billions or trillions of edges where node changes takes place dynamically. In many networks a node may belong to multiple communities simultaneously. Overlap is one of the characteristics of complex networks, in which a node may belong to more than one group. In particular, communities in social networks overlap with each other since many active users can possibly participate in multiple groups at the same time. There are many algorithms to detect overlapping communities such as Clique percolation[8], Agent and Dynamic based[9], Fuzzy based, Local expansion and Optimization, Line graph and Link partition etc.

## III. DYNAMIC SOCIAL NETWORKS

Social interactions are often modeled with networks. A key characteristic of social interactions is their continual change. Most past analyses of social networks are essentially static in that all information about the time that social interactions take place is discarded. Many researches in social network is connected with static investigation by considering independent graphs at different snapshots or one aggregated graph over the time period. In a dynamic network a node can be removed from the network or the node can newly join a network. The addition or deletion of nodes and links at different time stamps form a dynamic network. Since each and every minute new nodes and links are formed and deleted in online social networks, it is dynamic in nature. Hence community detection in complex networks, especially in social networks is an emerging area of interest.

Detecting the evolution of communities within dynamically changing network is crucial to understand the underlying dynamic processes driving complex system. In dynamic social networks, finding patterns of social interaction within a population has wide-ranging applications including: disease modeling, cultural and information transmission, and behavioral ecology. The next section discusses about the review of various researches done in dynamic social network.

## IV. RELATED WORK

Many community detection algorithms require prior information such as the size of the community before running the algorithm. The size is often not predictable beforehand in many complex networks, particularly, in social networks. Moreover it is computationally expensive to determine the size of communities before executing the algorithm. Label Propagation Algorithms (LPA)[10] are algorithms that are able to detect communities from complex networks without having any knowledge about the number or size of communities. LPA uses network structure alone as its guide to detect communities. The main idea behind LPA is to propagate the nodes throughout the network through some techniques and identify the communities from the propagation process itself. In this section, many label propagation based algorithms are reviewed.

Due to the label correlation problems; classification algorithms have difficulty in dealing with multi class/multi label problems. One vs. all strategy is the common approach used to address the multi-class/ multi-label problem, which also leads to difficult classes that are not properly utilized. Hence, the authors [1] proposed a semi-supervised multi-class/multi-label classification algorithm called Dynamic Label Propagation (DLP). It simultaneously deals with multi class and multi label problem by incorporating the label correlations and instance similarities dynamic and achieved significant improvement on benchmark data set. The DLP algorithms dynamically update the similarity measuring by fusing multi-label/multi-class information. The main idea of DLP is to have an improved transition matrix, by fusing information of both data features and data labels in each iteration.

DLP converges much faster than Label Propagation (LP) [10]. Usually, LP needs, more iteration to converge, while DLP only needs very few iterations due to diffusion process which manifolds into a KNN structure where only local similarities are preserved. For the step of kernel fusion addition of two matrices is performed. DLP has a find complexity of  $O(n^2)$  by decomposing it to observe whether only local neighbors are used to propagate the similarities.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

By keeping a record of the KNN matrix and performing the diffusion process repeatedly, its speed is improved. Wang et al. [1] extracted the fixed local structure from the KNN structure and performed multiplication  $K$  times for each pair of data points. Therefore, this can update the transition kernel in time  $Kn^2 + Kn$ . To summarize, the overall time complexity of DLP is  $O(Kn^2)$  where  $k \ll n$ .

LabelRank [15] updates all of its nodes at the same time, which is inefficient. LabelRankT[12] solves this updation problem by using local operations, working incrementally and only updating nodes that have changed in consecutive timestamps. LabelRankT is a generation of LabelRank for dynamic networks. LabelRankT detects a community adapting the network structure changes, LabelRankT uses inflation operator, cutoff operator, conditional update operator, and an extra conditional update rule which accept the new distribution only for nodes, involved in changes. The nodes are updated when they are changed between two consecutive snapshots. The added or removed edges are captured by LabelRankT. By comparing neighbors of a node  $I$  at two consecutive steps,  $t-1$  and  $t$ , i.e.,  $Nb^{t-1}(i)$  and  $Nb^t(i)$  are not equal. The node  $i$  is called as changed node and reinitialization of label distribution is performed only for changed node. The updation is performed until the stop criteria occur as specified in LabelRank algorithm. LabelRankT is more efficient than LabelRank. Since only the changed node and their neighbors are involved, the time complexity for the LabelRankT is  $O(m)$  in general.

A Density Based Structural Clustering Algorithm (DBSCAN) proposed by Ester et al. [2] forms communities from individual nodes (core point) which satisfies a user-defined number of neighbors within a given radius. Core points are connected to nearby core point to form communities. A distance function is used to relate the closeness between two nodes. Structural Clustering Algorithm (SCAN) [11] derived from DBSCAN detects communities in static networks using a similarity function that finds a ratio between the number of neighbors they both share in common to the number of neighbors that they each have. In SCAN a node  $v$  is a hub node if it does not belong to any community, but it does have more than two neighbors belonging to different communities. A node is an outlier if it does not belong to any community, and it does not contain more than one neighboring node that belongs to different communities.

Nathan et al. [6] improved SCAN by allowing it to form communities without the user-defined threshold for updating a dynamic network of timestamps. In Dynamic Scan (DSCAN), SCAN is performed on the first timestamp, for the next consecutive timestamps, the difference in edges between the two timestamps are obtained. When there are edge changes in the network, it is updated. During updation of edges a node can become a core, or a node can no longer be a core node. An existing cluster id or a new cluster id is propagated through all structurally connected nodes to form a new community. Through propagation, a node can either become a hub or an outlier. DSCAN performs better than SCAN by removing the need for parameter tuning.

The authors [6] also proposed Genetic Algorithm for dynamic networks namely Genetic Algorithm Dynamic (GAD). In GAD, the genes represent all individual nodes and their corresponding community. To find a gene that satisfies a good community structure of a network, the gene must maximize a given fitness function, such as modularity. A user-defined number of iteration is performed over the population which mutates each gene into a cross between two or more genes that produce the highest values due to the fitness function. At the end of the iteration, the gene with the highest fitness function value is used as the final community structure. GAD uses the previously found "best" gene of the previous timestamp as a gene in the population of the next. This results in the next timestamp starting with a gene in the population with a relatively good fitness.

Nathan et al. [5] proposed an algorithm SLPAD (Speaker-Listener Label Propagation Algorithm Dynamic) which adapts the label propagation technique followed by SLPA to propagate labels in the network. SLPAD algorithm runs SLPA on communities that changes from one time to another. Initially for each group SLPAD detects the edges that have changed between two consecutive timestamps  $\Delta E$  and finds all the nodes incident to the edges  $\Delta V$ . The detected set of nodes is expanded to include all nodes belonging to communities of these nodes. With  $T$  as number of iterations and  $r$  as the cutoff threshold, the algorithm SLPA [14] is run on these nodes. In this algorithm, the selected listener nodes receive the propagation only from the speakers within the node set that is being updated. Speaker node



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

can also send their highest frequency labels to the listener. Due to this condition, SLPAD algorithm converges quickly and produce more stable results than SLPA and LabelRankT. SLPAD has a time complexity of  $O(Tx)$  where  $x < n$ .

Before running the SLPA, two cleanup techniques is performed by the authors. First, the labels that disappeared from the previous timestamp to the current timestamp are removed from all the nodes in the network. Due to this cleaning process, some labels can no longer propagate through the network and it is removed. Second, the labels of the nodes that are to be updated are reset. By checking each node's neighbours and moving that node to the community with the highest neighbour count, the nodes are pushed to the same community that the majority of its neighbours belong to. By performing this task, community repair function is established by SLPAD algorithm and produces better modularity than LabelRankT. The drawback of SLPAD is finding the right overlap for a node. Since only updates are done based on edge changes, problem arises when a community involved in the edge change being updated, which can complicate finding an overlapping node between two communities if only one community of two in a possible overlap is updated.

Nathan et al. modified DSCAN to detect overlapping communities using an algorithm named DSCAN-O [5]. The first phase of DSCAN-O is detecting the communities using DSCAN. Then, it checks every node to extract their maximum similarity of all incident edges using a threshold value. The overlapping threshold is user specified similarity value. Any node whose edge is incident to another community with a similarity greater than the node's maximum similarity minus the similarity decrease becomes part of the other community. It shows the overlap between the two communities.

## V. CONCLUSION

Identification of communities in static and dynamic networks are important research issues in the field of complex networks. Numerous algorithms have been proposed to detect static and dynamic communities that may overlap with each other Label propagation algorithms are simple and near linear time algorithms which are adapted my many researchers for identifying communities. This paper presents an overview on community detection algorithms that detect disjoint and overlapping communities in dynamic social networks.

## REFERENCES

1. Bo Wang ., and Zhuowen Tu., and John K. Tsotsos, "Dynamic Label Propagation for Semi-supervised Multi-class Multi-label Classification".
2. Ester M., and Kriegel H P., and Sander J., and Xu X, "A Density-Based Algorithm for Discovering Communities in Large Spatial with Noise", In Proceeding of 2<sup>nd</sup> International Conference on Knowledge Discovery and DataMining, pp.226-231,1996 .
3. Gregory S, "Finding Overlapping Communities Using Disjoint Community Detection Algorithms", In ComplexNetworks, Springer, pp. 47–61, 2009.
4. Lancichinetti A., and Fortunato S., and Kertesz J, "Detecting the Overlapping and Hierarchical Community Structure in Complex Networks", New Journal of Physics, Vol.11, 2009.
5. Nathan Aston., and Jacob Hertzler., and Wei Hu, "Overlapping Community Detection in Dynamic Networks (DSCAN)", pp.872- 882, 2014.
6. Nathan Aston., and Wei Hu, "Community Detection in Dynamic Social Networks", 2014, Vol.6, pp.124-136.
7. Newman M.E.J., and Girvan M, "Finding and evaluating community structure in networks", Physical Review E, 69(026113), pp.1–16, 2004.
8. Palla G., Derenyi I., Farkas I., Viscsek T., "Uncovering the overlapping community structure of complex Networks", Nature and Society, 435:814-8, 2005.
9. Prabavathi GT., and Thiagarasu V, "Design and development of overlapping community detection algorithm using Multi-Label Propagation", International Journal of Advance Research in Computer Science and Management Studies, issue 2, Vol.2, pp.195-199, 2014.
10. Raghavan U.N., and Albert R., and Kumara S, "Near linear time algorithm to detect community structures in large-scale networks", Physics Review E, Vol. 76(3), pp. 036-106, 2007.
11. Xiaowei Xu., and Nurcan Yuruk, Zhidan Feng., and Thomas A. J. Schweiger, "SCAN: A Structural Clustering Algorithm for Networks", In Proceeding of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining , pp. 824-833, 2007.
12. Xie J., and Chen M., and Szymanski BK, "LabelRankT: Incremental community detection in dynamic networks via label propagation" arXiv: 1305.2006 v2 [cs.SI], pp.25-32, May 2013,.
13. Xie J., and Kelley S., and Szymanski B. K, "Overlapping Community Detection In Networks: The State Of The Art And Comparative Study", Arxiv Preprint Arxiv: 1110.5813, 2011.
14. Xie J., and Szymanski B. K., and Liu X, "SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process", In Proceeding of IEEE ICDM Workshop on DMCCI, pp. 344–349, 2011.
15. Xie J., and Szymanski BK, "LabelRank: A stabilized label propagation algorithm for community detection in networks", IEEE Network Science Workshop, West Point, NY, pp. 138–143, 2013.