

A Study on Community Overlapping Detection Algorithms in Social Networks

Eaint Mon Win ¹⁺, May Aye Khine ²

University of Computer Studies, Yangon

Abstract. Community detection is one of the most important research area wherein invention and growth of social networks. Community is a set of members densely connected within a group and sparsely connected with the other groups. In social networks, the singular characteristic of communities is multi membership of a node resulting in overlapping communities. Another relevant feature of social networks is the possibility to evolve over time. In recent years, many researchers have worked on various methods that can efficiently unveil overlapped structure on dynamic network. This paper reviews the previous studies done on the problem of overlapping community detection algorithms. Moreover, some approaches for dynamic network that change from time to time are also described.

Keywords: overlapping community, community detection, dynamic, social network.

1. Introduction

With the development of information technology, people living in society start to use social media as a virtual environment to find online friends. Social media refers to the interaction of users by exchanging idea in virtual communities. Social Networks are designed as graphs consisting of nodes where each node represents individual user such as people, organizations, and edges or lines between these nodes represents their relationship like friendship or mutual interaction. One of the problems in the study of large complex networks is the detection of community structure i.e. the decomposition of a network into groups (clusters or modules) consisting set of nodes. The process of uncovering these groups is called community detection. It is a clustering approach and difficult task in social network analysis [1]. Community detection is used in various applications where group decisions are taken, e.g., delivering information within group or recommending products to group. It can be also used for Target Advertising, Criminology, Public health and Politics [2]. Moreover, community detection or clustering with increasing number of smart devices is applied on smart devices network to generate the device social relation without human intervention by clustering approach of device interactions [3].

Much effort in identifying efficient and effective methods for community detection has been lead on finding disjoint communities. However, communities in real networks may overlap, i.e. a node belongs to one or more communities because an individual (node) share many things in common (e.g. regions, topic, hobbies, and interest). Fig 1 shows the visualization of overlapped community structure with three groups. Detection of overlapped structure becomes a challenge because traditional algorithms that detect only disjoint communities are not suitable. Therefore, overlapping community detection algorithms are defined on static network. In recent year, many researchers analysed overlapping community detection algorithms and evolution of community in social networks [4] [5]. In this paper, a description of the recent developed overlapping community detection algorithm is given and also represents developed algorithms on dynamic networks due to traditional overlapping community detection algorithms fails the problem on dynamic nature of social network.

This paper is organized as follows. The next section describes static and dynamic communities. Section 3 considers categorised overlapping community detection methods and proposed methods by many researchers in recent year. Section 4 describes common use network dataset in literatures. Finally, conclusion and future research are represented.

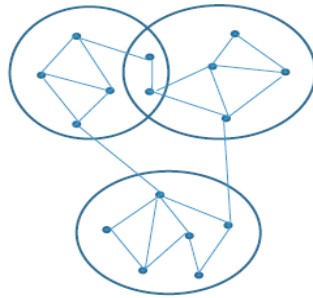


Fig. 1: Visualization of two overlapping communities with three groups

2. Static and Dynamic Communities

Since evolving communities are occurred as dynamic networks, a new research area concerned with dynamic structure has appeared. Communities having dynamic structure can change after a timestamp. For example, let's imagine a social network of American college football, and communities are identified on it. The static community represents all players in community at initial state t_0 . However as time goes on, players may change such as leave or enter into team and initial players will not be in the team; but, the corresponding community still exists. These dynamic network can be represented by a time-sequence of static networks called time frames (snapshots), each snapshot corresponding to interactions derived from data collection within a time period in Fig 2.

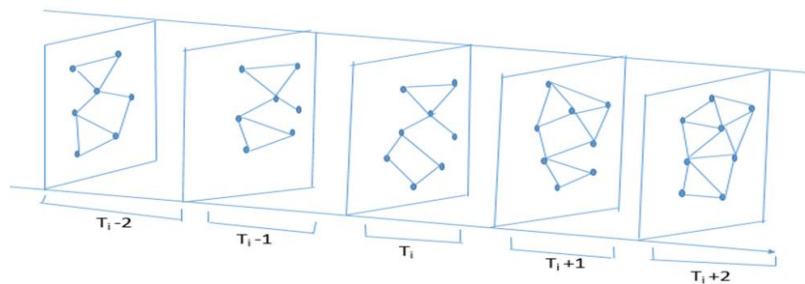


Fig. 2: A temporal network changes within five time frames

2.1. Events of Dynamic Communities

The changes of dynamic network are appearance and disappearance of nodes and edges in network over time. These operations or events are quite simple, but some operations are complex due to changes of entire or partly community. Therefore, implementation on all operation are not easy [6]. These operations on communities are (1) Growth: community's size grows when new nodes come into a community. (2) Contraction: community's size decreases when loses some nodes from community. (3) Merging: one or more communities are merged into one to form a new community. (4) Splitting: a community is decomposed into two or more communities. (5) Birth: a community emerges as first appearance at a given time. (6) Continue: community remain unchanged. (7) Death: a community disappears when loses its all members.

3. Overlapping Community Detection Algorithms

There are many algorithms for detecting community structures. However, most of the traditional algorithms are focused on identifying disjoint communities. Therefore many researchers tend to research the algorithms for detecting overlapping communities. The algorithms are categorized as Clique Percolation, Link Partition, Local Expansion, Label Propagation and Dynamic Networks.

3.1. Clique Percolation

The basic idea of Clique Percolation method to detect overlapping communities is creating a set of many cliques (completely connected subgraph) as communities. Firstly, construct vertices of the k clique graph and then construct edges of the k clique graph (percolate if two vertices in k -clique graph have strong connections). In the new graph, each connect component of the clique graph is a community [7].

Greedy Clique Expansion (GCE) [8] is proposed to detect accurate overlapped communities and get good performance on synthetic data. It begins finding core nodes by detecting maximum cliques and then core nodes are used to obtain community C by inserting node until added node obtain lower fitness. Extended Clique Percolation Method (ECPM) [9] is developed to address problem that cannot discover complete network of CPM. This method finds initial communities and lists left out nodes which are not included in any initial community and calculate their belonging coefficient. Communities are expanded by adding left out into corresponding community according to their belonging coefficient. Finally much similarity communities are merged into single one using Jaccard Similarity. But most algorithms based on CPM is high complexity because it finds many small cliques to get community.

3.2. Link Partition

Link partition method partitions the links in the original graph into several communities. It does not need to know prior knowledge and extracts the link graph using edges instead of nodes. After that graph partition algorithm is applied to cluster or some link community similarity functions are used to directly cluster.

A link clustering based novel algorithm [10] is proposed for loosely connected network that cannot be solved by CPM and for inaccurate overlapping community detection in weak-tie membership. In this paper, algorithm LinkSCAN based on link-space transformation (transforms original graph into a link-space graph). Overlapping Communities are detected using disjoint community detection algorithm on line graph and link similarity on original graph. By using this framework LinkSCAN* that enhances the efficiency of LinkSCAN by sampling. [11] developed a link based clustering algorithm called meme link which optimizes modularity density function to find densely connected links among communities by using weighted graph via similarity function. An efficient algorithm LEPSO [12] is proposed for solving problem of traditional algorithm based PSO that generate superfluous small communities, The algorithm is based on line graph theory, ensemble learning and particle swarm optimization.

3.3. Local Expansion

Local seed expansion method selects seeds and expand selected seed by using various fitness functions and then generates overlapping communities by merging intermediate communities into global communities. The problem to deal with this algorithm is to find good seed sets. In Local Expansion, each implementation differs a lot by depending various fitness function.

Therefore, an algorithm represented in [13] is seed set expansion algorithm to detect overlap communities. The basic concept of algorithm is to identify good seeds, and then expands these seed sets using the personalized PageRank clustering procedure. This algorithm consists of four phases namely filtering (use graph partitioning), seeding (use Graclus Center and Spread Hub), seed set expansion (use PageRank) and propagation to find overlap community. In 2016, [14] detected the minimal cluster using density function to find the nodes that are closely connected with the initial nodes and then finds the local community extended from the minimal cluster. In 2018, proposed an algorithm, OCDNW to find a good seed. The algorithm consists of three parts: initial seed selecting, local community expansion and community optimization. For initial seed selecting, choose seed node with highest weighting of each node by sum score of all edges which connected to the neighbors. Then expand seed to form local community using node fitness function. Finally, merge two communities into larger ones to improve the quality of community if there are too many overlapped vertices between two communities [15].

3.4. Label Propagation

Label propagation algorithm refers to labels propagate between nodes. It begins assigning each vertex to unique label and these labels propagate through the network. In updating labels, ties are broken randomly if the node receive multiple labels. But LPA can detect only disjoint community.

COPRA [16] is a multi-label propagation algorithm to detect overlapping communities. It begins by giving unique label with belonging coefficient setting 1 to every vertex. Each vertex updates its labels by

summing and normalizing the belonging coefficients of vertices in the neighbor set. After propagation, each vertex has multiple labels. Another extension of label propagation technique, SLPA [17] is developed to avoid detecting only disjoint community. It can detect overlapping communities because receives multiple label. In SLPA, each node is initialized unique label and then one node is selected as listener. Each neighbor of the selected node randomly selects a label and sends the selected label to the listener. Listener accepts one of the propagated labels according to listener rule. Finally, threshold is used to generate overlapping communities as post processing. In 2018, [18] developed LPANNI (Label Propagation Algorithm with Neighbor Node Influence) to overcome weakness such as low accuracy, instability (some algorithms based on LPA require to set parameter as priori). It detects overlapping community structures by adopting fixed label propagation sequence based on the ascending order of calculated node importance and label update strategy based on neighbor node influence and historical label preferred strategy (i.e. using idea of COPRA for detecting overlap and DLPA for reducing complexity).

3.5. Dynamic Networks

Most real world social networks are inherently dynamic, growth rapidly in term of social interaction. With community structures change from time to time in evolving network, this research area is receiving more interesting from researchers. The more recent methods aiming to find out dynamic communities are described.

SLPA Dynamic (SLPAD) is an algorithm based on SLPA and incorporates the ability for this algorithm to handle dynamic networks. It involves running SLPA on communities that change from one timestamp to the next. But SLPAD considers only updates based on edge changes not node changes [19]. In 2018, an overlapping community's detection method is proposed using agent that observe the network and consequently update their communities on dynamic networks. For detecting communities each node is iteratively reassigned to the community that yield the highest positive gain by considering similarity attributes in initial partition. It allows all operation on communities: birth, death, growth, and contraction, merge and split [20]. OLCPM is based on clique percolation and label propagation methods. The paper introduced online CPM (OLCPM) by building upon CPM to identify the core nodes of communities in real time and it aimed on analysing the dynamic behaviours of the network which may appear from inserting or removing. As post-processing based label propagation, OLCPM is set out on the generated communities of OCPM to discover the peripheral nodes [21].

4. Standard Dataset for Testing Algorithms

The most commonly use dataset in literatures for testing algorithm are divided into two types: real world network dataset and synthetic dataset. Mostly use real world dataset and available sites are shown on Table 1. For Synthetic dataset, suitable dataset for overlapping community detection algorithms in literature is LFR benchmark (an algorithm generate benchmark dataset is proposed by Lancichinetti) [22].

Table 1: Available network dataset for testing algorithms

Type	Network	Number of Node	Number of Edge	Available site
Real World Network	Zachary Karate Club	34	78	http://www-personal.umich.edu/~mejn/netdata/
	Bottlenose Dolphin	62	159	
	College Football Club	115	613	
	US Politics books	105	441	
	Ego Facebook	2888	2981	https://snap.stanford.edu/data/ego-Facebook.html
Synthetic Network	LFR benchmark			https://sites.google.com/site/santofortunato/inthepress2

5. Conclusion

Community detection is more and more attention in social network with the rapidly growth of social data. Till present, related research area is still popular due to challenges of community detection such as dynamic nature, overlapping nature of social network and algorithm instability. This paper reviews algorithms for detection of overlapped communities and algorithms on dynamic networks. As future work, modified

algorithms are developing for instability of algorithm and leads to purpose for defining efficient and effective methods on dynamic network. As data can get large and distribute on web, researchers are comparing accuracy and performance of algorithms on Sparse and Hadoop architecture.

6. References

- [1] L. Tang, H. Liu, Community detection and mining in social media, Synthesis lectures on data mining and knowledge discovery . pp:1-137, 2010
- [2] A. Karatas, S. Sahin, Application Areas of Community Detection: A Review. 2018 IBIGDELFT
- [3] K. Dong-Oh, C. Jang-Ho, J. Joon-Young, and B. Changseok, "Clustering Approach of Device Interactions for Automatic Generation of Device Social Relation," International Journal of Machine Learning and Computing, 5, no. 2, pp. 132-136,2015
- [4] J. Xie, S. Kelley, B.K. Szymanski, Overlapping community detection in networks: The state-of-the-art and comparative study, 2013, ACM
- [5] N. Dakiche, FBS. Tayeb, Y. Slimani, and K. Benatchba, Tracking community evolution in social networks: A survey Information Processing & Management. pp:1084-1102,2019, Elsevier
- [6] G. Rossetti, R. Cazabet, Community discovery in dynamic networks: a survey. ACM Computing Surveys (CSUR), 2018
- [7] S. Fortunato, Community detection in graph. pp:75-174, 2010 ,Elsevier
- [8] C. Lee, F. Reid, A. McDaid, N. Hurley, Detecting highly overlapping community structure by greedy clique expansion. arXiv preprint arXiv:1002.1827,2010
- [9] S. Maity, S K. Rath, Extended Clique percolation method to detect overlapping community structure. International Conference on Advances in Computing, Communications and Informatics (ICACCI). pp: 31-37. 2014.IEEE
- [10] S. Lim, S. Ryu, S. Kwon, K. Jung, J-G. Lee, LinkSCAN*: Overlapping community detection using the link-space transformation, 2014 IEEE 30th International Conference on Data Engineering. pp: 292-303. 2014.
- [11] M. Li, J. Liu, A link clustering based memetic algorithm for overlapping community detection. Physica A , 2018
- [12] F. Huang, X. Li, S. Zhang, J. Zhang, J. Chen, Z. Zhai, Overlapping community detection for multimedia social networks. IEEE Transactions on multimedia. pp: 1881-1893. 2017
- [13] JJ. Whang, DF. Gleich, IS. Dhillon, Overlapping community detection using seed set expansion. Proceedings of the 22nd ACM international conference on Information & Knowledge Management. pp: 2099-2108. 2013
- [14] Y. Zhou, G. Sun, Y. Xing, R. Zhou, Z. Wang, Local community detection algorithm based on minimal cluster. Applied Computational Intelligence and Soft Computing.2016. Hindawi
- [15] X. Chen, J. Li, Overlapping Community Detection by Node-Weighting. ICCDA 2018, DeKalb, IL, USA. © 2018 Association for Computing Machinery
- [16] G. Steve, Finding overlapping communities in networks by label propagation. New Journal of Physics12, 2010
- [17] J. Xie, B K. Szymanski, X. Liu, Slpa: Uncovering overlapping communities in social networks via a speaker listener interaction dynamic process. IEEE, pp: 344-349. 2011.
- [18] M. Lu, Z. Zhang, Z. Qu, Y. Kang, LPANNI: Overlapping community detection using label propagation in large-scale complex networks. IEEE Transactions on Knowledge and Data Engineering 2018
- [19] Aston, Nathan; Hertzler, Jacob; Hu, Wei; overlapping community detection in dynamic networks. Journal of Software Engineering and Applications. 2014
- [20] A. Mahfoudh, H. Zardi, M A. Haddar, Detection of dynamic and overlapping communities in social networks. Int. J. Appl. Eng. Res. pp: 9109-9122. 2018
- [21] S. Boudebza, R. Cazabet, F. Azouaou, O Nouali, OLCPM: An online framework for detecting overlapping communities in dynamic social networks. pp: 36-51. 2018, Elsevier
- [22] A. Lancichinetti, S. Fortunato; Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities PHYSICAL REVIEW E80, 016118 , 2009.