ISSN:0975-9646

Vinay et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3339-3341

A Review paper for Detection of Overlapping Communities in Complex Networks

Vinay[#1], Sumit[*2], Jai Parkash[#3]

[#1, 3](Research Scholars) University Institute of Engineering and Technology
Kurukshetra University Kurukshetra, India

[* 2](Research Scholars) Department of Computer Science and Application
Kurukshetra University, Kurukshetra, India

***Abstract*: This paper presents a comprehensive study about different methods which were used to detect the communities in networks. Most of the algorithms were failed to perform in all kind of networks i.e. some perform only in weighted networks, some in bi-partite networks and some in directed networks etc. Label propagation algorithm plays the most effective role in community detection as it takes much less time in comparison of many algorithms. But the main disadvantage of this algorithm is that it fails in finding overlapping communities. Despite of its disadvantages it works as a base method for different algorithms like rank based label propagation algorithm and multiple label propagation approach, which are enhancements in label propagation algorithm. So, there must be an algorithm which can be used in most of the networks and able to detect communities both disjoint and overlapping in less running time. We will try to overcome these problems in our research with the help of label sharing approach**.

***Keywords:* community, community detection, label propagation algorithm, label sharing approach, social networks.**

I. INTRODUCTION

A. *Social networks*

Social networks are the most used and common phenomenon for interaction among the peoples living in a same country or in different countries. A group of people interact with each other in different ways and in this way they form a social network (SN). 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.

B. *community*

Social network analysis is a technique which is used to describe the structure and characteristics of the networks by establishing the relationships among the individuals and similarities present among them. Based on the similarity among the individuals a network is divided in communities. There is no any proper definition for a community but many researchers try to define a community. Based on their definitions we can say that a community in a network is a group of nodes which are similar to each other and dissimilar to the rest of network. Despite there is no universally accepted definition of a community, most of the real world networks display community structure.

C. *Community Detection*

In recent years there has been a lot of work done for defining, detecting and identifying a community present in a network. In community detection algorithms, the main aim is to find the group of nodes of interest in a network. Communities in social networks can provide insights about common characteristics or beliefs among people that makes them different from other communities [1]. Community detection is similar to network portioning problems, in which a network is divided into n groups of approximately equal sizes and the number of edges between the groups is minimized. Community detection is an NP hard problem [1] and a number of heuristic approaches are developed for solving this problem existing among the networks. A simple community diagram is shown below:

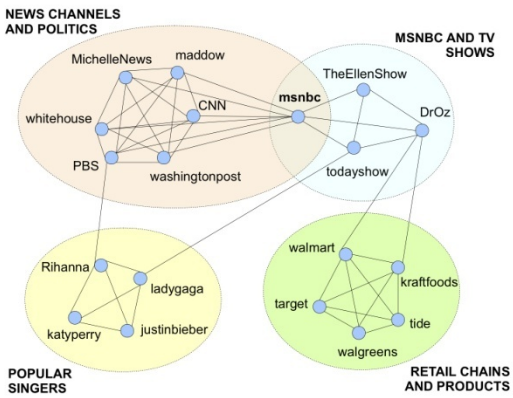

Fig.1 source: Google images [12].

II. RELATED WORK

Bing Kong *et al.* [2] elucidated a dynamic algorithm based on modularity for the community detection in social networks. The algorithm uses modularity as the criterion functions, based on a certain initial division, move the nodes inside the community through iteration, find the division creating the maximum modularity, and automatically detect the community number included in the network. The proposed algorithm is applied to the real network data. The results produced by this algorithm show that the algorithm can detect the communities dynamically in the network and the detected communities fits better with real communities. In a time growing world with the number of communities and nodes, the computational complexity of the algorithm should be lowered. Nikos Salamanos *et al.* [3] investigated the correlation between the social network communities defined by the community detection algorithms and the Facebook pages explicated as likes by its users. The authors firstly examined the relation between the underlined social dynamic, expressed indirectly by a community structure and the users, characteristics represented by likes and secondly, evaluated the outcome of the community detection algorithm. The experimental results demonstrated that in the case of users' Likes, the correlation ranges from small to medium between communities and the whole population, while it is even smaller between communities. Moreover there is a high correlation in terms of likes' categories between different communities and whole population. This research showed that Likes constitute a criterion of distinction among the communities and verifies the intuition. Some robust community detection methods should be applied to find the correlation among the communities and the annotated pages. Sudip Misra *et al.* [4] expounded a community detection scheme in an integrated Internet of Things (IoT) and Social Network (SN) architecture. Authors used a graph mining approach for solving the problem in complex network of IoT and SN. In this algorithm nodes in complex networks are divided into basic nodes and IoT nodes and after that execute the community detection algorithm. Authors consider two nodes to be in a community, only if the nodes are at most one hop apart and have at least two mutual friends. Main advantages of this approach are as follows: - a node in a SN can be a part of multiple communities, this approach can be used for suggesting friends and this algorithm works well for weighted networks also. Main drawback of this algorithm is that this algorithm is not generalized to all networks as it fails to give results for directed networks. Ye Conghuan [5] discussed the idea about the generalization of dense sub-graph problem by an additional distance restriction to the nodes of the dense sub-graph. Author proposed this method because of local core players, which are not always appeared in cliques. The author proposed a new quasi-clique detection algorithm based on the definition of dense sub-graph and derived some optimization techniques based on the idea of synchronization, which can detach the unpromising and redundant alien from the dense sub-graph. The method described in this paper could discover quasi-cliques and core players that were not shown in cliques. This algorithm detects most of the dense sub-graphs but not all. Hubs and outliers should be detected. Sercan Sadi *et al.* [6] illustrated an efficient community detection method using parallel clique-finding ants. High computational costs and non-scalability on large scale social networks are the biggest drawbacks of popular density detection methods. The main concept behind this paper is to reduce the original network graph to a maintainable size so that computational costs decrease without loss of solution quality, thus increasing scalability on such networks. In this paper, authors focus on Ant Colony Optimization techniques to find quasi-cliques in the network and assign these quasi-cliques as nodes in a reduced graph to use with community detection algorithms. The result produced by this method shows that the execution times of the community detection methods are decreased while overall quality of the solution is preserved. Some work on parallelized framework can improve the quality and execution time. Bonan Hou *et al.* [7] presented COMMPAR, a community-based model partitioning approach, which utilizes the community information of social networks for performance tuning. The authors proposed a two-phased network model partitioning as follows: first, a community detection algorithm is applied to discover community structure Residing in large-scale social networks; second, these communities are further equally partitioned to achieve an appropriate configuration of simulation execution, and facilitates mapping of the communities onto multiple computer processors. The results produced by this algorithm shows that this algorithm can effectively reduce the overhead of inter-processor communications. The primary focus of this paper was large-scale social network simulation; the approach proposed is applicable to many other parallel systems with community structure and social network modeling for performance tuning. If some work is done on this algorithm, this algorithm can also be used to determine the configuration of computing resources and performance prediction for dynamics simulation on specific social networks by using COMMPAR. Tetsuya Yoshida [8] proposed an approach for finding hidden communities based on user profiles in social networks. The author presented both observable connectivity relation technique and profile graph together. Instead of hierarchical approach, author proposed an embedding approach which utilizes the regularization via the profile graph. The results produced by this method are really encouraging, especially in terms of the gain obtained via the profile graph. The problem of detecting communities in networks is expounded by Bowen Yan *et al.* [9] using merging cliques. Many algorithms had been proposed for detecting disjoint communities in networks. One of the most popular techniques is to optimize modularity, a measure of the quality of a partition in terms of the number of intra-community and intercommunity edges. Greedy approximate algorithms for maximizing modularity can be very fast and effective. The authors proposed a new algorithm that starts by detecting disjoint cliques and then merges these to optimize modularity. Results produced by this algorithm showed the better performance than other similar algorithms in terms of both modularity and execution speed. The drawbacks of this paper are: first, speed of this algorithm is less than other similar algorithms. Second, this algorithm does not allow overlapping communities. Zhewen Shi *et al.* [10] elucidated PSO based community detection scheme in social networks. In this paper authors proposed a method based on particle swarm optimization which detects community structures by optimizing network modularity. At the beginning, an improved spectral method is used to transform community detection into a cluster problem and the weighted distance which combine eigen values and eigenvectors is advanced to measure the dissimilarity of two nodes. Then, PSO is employed for cluster analysis. Two definitive features are present in this algorithm: first, the number of communities is determined automatically; second, the particle has low dimensional structure. The results produced by this algorithm shows that the algorithm obtains higher modularity over other methods and achieves good partition results. Bin Wu *et al.* [11] narrated the problem of community detection in social networks using resume mining. In this we studied the three important aspects of resume mining: the characterization of community, the discrimination among communities and the community evolution mining. Unlike other similar algorithms, solutions provides by authors fully consider the inner topology of a community together with the attributes of nodes. Two cases were studies in this: first was about mobile call graphs and the second was about co-authorship networks. Results produced by these methods represent state and history of a community. These methods work effectively to make the state of community clear. The drawbacks of this paper are: first, the methods discussed in this paper do not employ the attributes of edges and second, results produced by these methods highly depend on the precision of community partition, which may not be accurate.

III. Problem Formulated

As defined by many authors that community detection is NP-hard problem and there are many heuristic approaches for solving these problems. The main problem in community detection is that most of the algorithms are failed in the detection of overlapping communities. In this review paper we study different methods which were used in the detection of communities in complex and social networks. These algorithms have its advantages in some aspects but have disadvantages also.

IV. Discussion And Conclusion

As we have discussed earlier that there are many algorithms which were used for community detection in different networks. Some of the algorithms are used in weighted networks, some are in bipartite networks and some for directed networks but most of them are failed in detection of communities in all type of networks. Some algorithms are failed in terms of modularity, which is the main factor in determining the performance of an algorithm. An algorithm based on label propagation approach plays an important role in the field of community detection. Despite it has disadvantage as it fails in detection of overlapping communities, but it works as a base method for different algorithms such as rank based label propagation and multiple label propagation algorithms. So, we need an algorithm which is able in finding both overlapping and disjoint communities in a less running time with most effective outputs such as in terms of modularity and can be used in most of the networks whether it is a weighted network or bipartite network or a directed network. We will try to overcome these problems by using label sharing approach.

References

[1] Usha Nandini Raghavan, Reka Albert and Soundar Kumara, "*Near linear time algorithm to detect community structures in large-scale networks*", Physical Review E 76, Vol. 3, 2007

[2] Bing Kong, Hongmei Chen, Weiyi Liu, Lihua Zhou "*A Dynamic Algorithm for Community Detection in Social Networks**", in proceedings of the 10th world congress on intelligent control and automation, pp. 350-354, 2012

[3] Nikos Salamanos, Elli Voudigari, Theodore Papageorgiou and Michalia Vazirgiannis "*Discovering Correlation between Communities and Likes in Facebook*", in proceedings of IEEE INTERNATIONAL Conference on Green Computing and Communications, Conference on Internet of Things, and Conference on Cyber, Physical and Social Computing, pp. 368-371, 2012

[4] Sudip Misra, Romil Barthwal and Mohammad S. Obaidat, "*Community Detection in an Integrated Internet of Things and Social Network Architecture*", in Globecom Communications QoS, Reliability and Modeling Symposium, pp. 1647-1652, 2012

[5] Ye Conghuan "*Dense Subgroup Identifying in Social Network*", in proceedings of international Conference on Advances in Social Networks Analysis and Mining, pp. 555-556, 2011

[6] Sercan Sadi, Sule Oguducu, A. Sima Uyar, "*An Efficient Community Detection Method using Parallel Clique- Finding Ants*", in proceedings of IEEE Congress on Evolutionary Computation, pp. 1-7, 2010

[7] Bonan Hou, Yiping Yao, "*COMMPAR: A Community based Model Partitioning Approach for large scale Networked Social Dynamics Simulation*", in proceedings of 14th IEEE/ACM Symposium on Distributed Simulation and Real-Time Applications, pp. 7-13, 2010

[8] Tetsuya Yoshida, "*Toward Finding Hidden Communities based on User Profile*", in proceedings of IEEE International Conference on Data Mining Workshops, pp. 380-387, 2010

[9] Bowen Yan and Steve Gregory, "*Detecting Communities in Networks by Merging Cliques*", in proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems, pp. 832-836, 2009

[10] Zhewen Shi, Yu Liu, Jingjing Liang, "*PSO-based Community Detection in Complex Networks*", in proceedings of Second international Symposium on knowledge Acquisition and Modeling, pp. 114-119, 2009

[11] Bin Wu, Xin Pei, JianBin Tan, Yi Wang, "*Resume Mining of Communities in Social Network*", in proceedings of Seventh IEEE International Conference on Data Mining – Workshops, pp. 435-440, 2007

[12]https://www.google.co.in/search?q=community+structure+diagram+in+social+networks&source=lnms&tbm=isch&sa=X&ei=2HJjU7z3DcL08QWbvoLQCQ&ved=0CAYQ_AUoAQ&biw=1280&bih=699#facrc=_&imgdii=_&imgrc=wKkiEGaRCkGPaM%253A%3B8HZh3K7Y6ODWkM%3Bhttp%253A%252F%252Fcucis.ece.northwestern.edu%252Fprojects%252FSocial%252Ffigures%252Fclique_based_communities.jpg%3Bhttp%253A%252F%252Fcucis.ece.northwestern.edu%252Fprojects%252FSocial%253F%3B1651%3B1275