COMMUNITY DETECTION IN SOCIAL NETWORKS

[1]Jyoti Shokeen, [2]Partibha Yadav, [3]Meenakshi
[1,2]Ph.D Scholar
[1]Department of Computer Science & Applications, M.D.U., Rohtak, India
[2]I.G.U., Meerpur, Rewari, Haryana, India
[3]Govt. Nehru College, Jhajjar, Haryana, India

Abstract- A social network is a social structure made up of individuals or organizations called nodes or actors which are related through edges. The actors are related to one another in various ways such as friendship, interests, dislike, kinship, prestige or financial exchange. Social networks play a vital role in spreading information. To mine and analyze the information of social networks has become an attracted research area. Such networks can be represented in the form of graphs and analyzed as a whole to detect a number of communities present in the network. Community detection is one of the vital areas in the field of social networks. The social interactions among the individuals can be classified on the basis of behavior, positions, desires and tastes, etc. A community structure in social networks consists of dense connections within the group and sparse connections to the rest of the network.

Keywords- Social networks, Community detection, overlapping communities.

I. INTRODUCTION

With the growing technology, the internet users not only consume information but they also produce information. The users interact with each other, exchange view, participate in online discussions and thus form social networks. The flow of vast amount of data referred to as big data has received significant attention and has given boom to an emerging field. Identifying the group of users that interact more frequently with each other enable us to understand how knowledge, influence, happiness and even interests owe to social network. Identifying the groups that share similar tastes and desires is called community detection in social network analysis. Nodes in a community are densely connected and are sparsely connected to other communities [1].

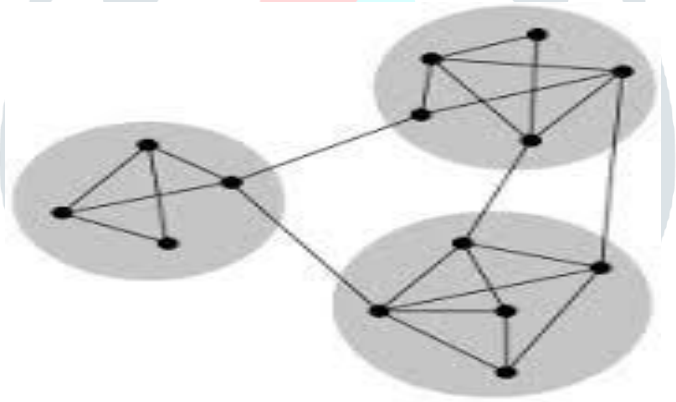

Fig 1. A small network with three communities

The interconnection of nodes in social networks can be represented in the form of graphs [2]. Community structure is an important property of social networks and detection of communities in the network is an issue. A number of approaches have been discovered to identify the hidden and overlapping communities. A network is overlapping when actors can be considered members of multiple communities. Social networks are often dynamic in nature and the social structure changes with the passage of time. New actors appear and form new relations while old actors may disappear and their corresponding relations fade away. The community structure of dynamic social network changes with the change in like, dislike, behavior, interests, etc.

The paper is organized as follows: Section II presents the literature review. In Section III, a comparison of different approaches for community detection is presented in tabular form. Section IV describes conclusion part.

II. LITERATURE REVIEW

Early research in the area of community detection [3] was mainly focused on the static networks but real- world networks are dynamic networks. As time changes, new nodes appear in the network and edges are added with the interactions among them and old nodes and edges disappear leading to the changes in network.

Elbarawy et al. [4] proposed a community detection method using DBSCAN algorithm. DBSCAN algorithm can detect the outliers thus making the dataset free from noise and making it efficient. Existence of communities can be concluded with interaction among the individuals in the network. In the approach, communities are detected on the idea of three types of nodes or members in the community: core, border and outlier members. Core nodes have high influence, border nodes have low influence and outlier nodes have no influence to other nodes in the network. Core members having more influence are primarily focused in the approach.

Ganjaliyev [5] proposed a method to identify communities in the network by clustering. In this method, a set of constraints that every data object is assigned to exactly one cluster and every cluster contains at least one object [8]. A certain number of clusters are selected and the total weight of the selected clusters is maximized and the similarity between the clusters is minimized.

Choudhury et al. [6] focused on a new method using Newman-Girvan algorithm [7] to detect communities and sub-communities in a social network. The approach was implemented on real world networks such as Zachary Karate Club, Bottlenose Dolphin Network and College Football Network.

Rees and Galahar [9] proposed an approach for detecting overlapping communities using swarm intelligence. Swarm intelligence removes the need for central control mechanism. The approach follows some assumptions to detect communities: Homophiliy or common interest exists to bind the communities; edges representing same level of relationship strength and all edges have same amount of reciprocity [8]. The algorithm is based on the notion that a person interacting in a social group knows his/her friends and friends-of-friends. Firstly, the friendship groups are identified and assigned a unique identifier to each group. Then, non-propagating nodes in the group are identified followed by propagation of friendship group IDs if the ID for a group changes.

Conrad Lee and Padraig Cunningham [10] described the difference in evaluation of large datasets and small, hand-curated datasets. The algorithm which works well on smaller datasets may perform poorly on large datasets. They introduced a framework where social network datasets is annotated with meta-data. The framework is based on machine learning task. It is assumed that if a community detection algorithm is working well, then the classifier should be able to use the set of detected communities to infer missing valued of a node attribute that is closely related to community structure. The attribute suffers from two deficiencies: incompleteness and nesting. They also explained how mined data suffers from purpose-gathered data.

Feyza Altunbey and Bilal Alatas [11] proposed a parliamentary optimization algorithm to detect the overlapping communities. An overlapping community is one in which a person belongs to more than one group. The approach optimizes network modularity function and fitness function. In this approach, a population of people is created and partitioned into groups. Individuals with high fitness function are selected as the leaders and then intra-group competition begins followed by inter-group competition. After meeting the stopping criteria, a group is declared winner and its best member is the solution to solve the optimization problem.

Z. Ding et al. [12] proposed a new overlapping community detection method, NDOCD, based on network decomposition. The algorithm removes all links in derived link communities, thereby, splitting the network iteratively. Node clustering techniques has been used to identify the link communities. Decomposition of network has reduced the computational time and removal of noisy links has improved the quality of communities so obtained. Moreover, node clustering technique is used to find link communities. The algorithm has been tested on both real-world and synthetic networks.

Nasif Muslim [13] discussed four schemes for detection of community structure which use structural and attribute similarity of nodes. Each scheme provides different outputs which may be combined to find communities in social networks. The first scheme uses structural similarity between the nodes. The second scheme utilizes attribute similarity between the nodes. Both structural similarity and attribute similarity between the nodes are exploited in the third scheme. In the fourth scheme, attribute similarity between the nodes are used considering only the relevant attributes.

Wang et al. [14] proposed a method NEIWalk for community detection in dynamic content-based network. The paper proposes a transformation of content-based network into Node-Edge Interaction (NEI) network where node content, edge content and linkage structure are embedded seamlessly. At first, the content-based network is transformed into an NEI network, a multi-mode network, consisting of two types of nodes and three types of edges, termed as n-node and e-node respectively. The three types of edges respectively characterize the node content similarity, edge content similarity and structural similarity. As the content-based network evolves, a differential activity based approach is proposed to maintain the NEI network incrementally. Latent communities are discovered by applying heterogeneous random walk in the NEI network.

III. COMPARATIVE STUDY OF COMMUNITY DETECTION METHODS

A tabular representation is presented to provide the comparative study of different community detection methods with their advantages and limitations as follows:

Authors	Paper Title	Advatages	Limitations
Yomna M. ElBarawy, Ramadan F. Mohamedt and Neveen I. Ghali	Improving Social Network Community Detection Using DBSCAN Algorithm	DBSCAN algorithm is robust to outliers and eliminates noise.	Border points can be the part of one or more clusters.
F. Ganjaliyev	New Method for Community Detection in Social Networks Extracted from the Web	Similarity between clusters in minimized and maximized within the clusters.	Not suitable for all types of networks.
Deepjyoti Choudhury, Saprativa Bhattacharjee, Anirban Das	An Empirical Study of Community and Sub-Community Detection in Social Networks Applying Newman-Girvan Algorithm	A new concept is defined to detect sub-communities.	Computation cost is high.
Bradley S. Rees and Keith B. Gallagher	Overlapping community detection using a community optimized graph swarm	Remove the need of central control mechanism, good performance, scalable to large datasets.	Performance is limited to single Java Virtual Machine.

Conrad Lee and Padraig Cunningham	Community detection: effective evaluation on large social networks	Missing value can be inferred from the detected communities, can measure recall.	Can only measure recall, not precision. May not discover more complicated relationships. Computationally expensive to train a classifier. Classifier's performance evaluation may introduce complexity.
Feyza Altunbey and Bilal Alatas	Overlapping Community Detection in Social Networks Using Parliamentary Optimization Algorithm	Can analyze community structure, optimize network modularity.	Only modularity measure is used as the fitness function to detect overlapping communities.
Zhuanlian Ding, Xingyi Zhang, Dendgi Sun and Bin Luo	Overlapping Community Detection based on Network Decomposition	Good accuracy, efficient computation and improved quality of obtained communities.	Network topology is not taken into consideration.
Chang-Dong Wang, Jian-Huang Lai and Philip S. Yu	NEIWalk: Community Discovery in Dynamic Content-Based Networks	Transformation of content-based network into a Node-Edge Interaction (NEI) network where linkage structure, node content and edge content are embedded seamlessly.	Bounded accuracy loss due to the random walk sampling

IV. CONCLUSION

Communities play a vital role in understanding the structure of large social networks. In this paper, we have described various methods used in the detection of communities. A comparison of various community detection approaches has also been stated in tabular representation. Each approach has its own advantages and weaknesses. The aim is to develop algorithms that can effectively detect the communities in different application domains.

REFERENCES

[1] Clauset, A., Newman, M.E.J., and Moore, C. "Finding community structure in very large networks", Physical Review E, 70:066111, 2004.

[2] S. Wasserman and K. Faust, "Social Network Analysis", Cambridge University Press, Cambridge, 1994.

[3] S. Fortunato, "Community Detection in Graphs", Physics Reports, 486, 75-174. http://dx.doi.org/10.1016/j.physrep.2009.11.002.

[4] Yomna M. ElBarawy, Ramadan F. Mohamedt and Neveen I. Ghali, "Improving Social Network Community Detection Using DBSCAN Algorithm", Computer Applications & Research (WSCAR), 2014 World Symposium, IEEE, 2014.

[5] Ganjaliyev. F, "New Method for Community Detection in Social Networks Extracted from the Web", Problems of Cybernetics and Informatics (PCI), International Conference, IEEE, 2012.

[6] Deepjyoti Choudhury, Saprativa Bhattacharjee and Anirban Das, "An Empirical Study of Community and Sub-Community Detection in Social Networks Applying Newman-Girvan Algorithm", Emerging Trends and Applications in Computer Science (ICETACS), 2013 1st International Conference.

[7] M. E. J. Newman, "Fast algorithm for detecting community structure in networks", Phys. Rev. E 69, 2003.

[8] Mehjabin Khatoon and W. Aisha Banu, "A Survey on Community Detection Methods in Social Networks", MECS, 2015.

[9] Bradley S. Rees and Keith B. Gallagher, "Overlapping community detection using a community optimized graph swarm", Social Network Analysis and Mining, Springer, pg. 405-417, 2012.

[10] Conrad Lee and Padraig Cunningham, "Community detection: effective evaluation on large social networks", Journal of Complex Networks, pg. 19-37, 2014.

[11] Feyza Altunbey and Bilal Alatas, "Overlapping Community Detection in Social Networks Using Parliamentary Optimization Algorithm", International Journal of Computer Networks and Applications, Vol. 2, Issue 1, 2015.

[12] Zhuanlian Ding, Xingyi Zhang, Dendgi Sun and Bin Luo, "Overlapping Community Detection based on Network Decomposition", Scientific Reports 6, 24115, 2016.

[13] Nasif Muslim, " A combination Approach to Community Detection in Social Networks by Utilizing Structural and Attribute Data", Social Networking, pg. 11-15, 2016.

[14] Chang-Dong, W., Jian-Huang, L., Philip S. Y., "NEIWalk: Community Discovery in Dynamic Content-Based Networks", IEEE transactions on knowledge and data engineering, Vol. 26, No. 7, July 2014.