

A Review on Overlapping Community detection in Social Network

Richa¹

¹ M.Tech scholar, Department of Computer Science, Central University of Punjab, India, richarich05@gmail.com

-----***-----

Abstract - Community detection in a Social network is a trending issue in the study of network system and it helps to realize and study the overall network structure in detail. Communities are basically the partition of network nodes into some subgroups in which nodes within these subgroups are densely connected, but between the subgroups, the connections are sparser. In this paper different overlapping community detection algorithms are studied in order to detect the overlapping node in the network. Some modularity measures are also considered to measure the quality of communities detected by the algorithms.

Keywords: Complex network, Community Structure, Overlapping community.

1. INTRODUCTION

Networks are omnipresent on the Web. The most profound Web network is the Web itself comprising billions of pages as vertices and their hyperlinks to each other as edges. Moreover, collecting and processing the input of Web users (e.g. queries, clicks) results in other forms of networks, such as the query graph. Finally, the widespread use of Social Media applications, such as Delicious, Digg, Flickr and YouTube, is responsible for the creation of even more networks, ranging from folksonomy networks to rich media social networks. Since networks originating from Social Media data are of particular interest to this study, it shall be collectively refer as **Social Media networks**.

A social network can be defined as a set of people connected by a set of people. A social structure made of nodes (individuals and organizations) that are related to each other by various independencies like friendship, kinship etc. Social network analysis provides both visual and mathematical analysis of human relationship.

Community Structure assumes more significance with the increasing popularities of online social network (OSN) services like FACEBOOK, TWITTER, and MYSPACE etc.

2. CONCEPT OF COMMUNITY

A community in social network usually contains the users having similar characteristics that make them different from others. Community is actually a subset of “actors among whom there are relatively strong, direct, intense, frequent and positive ties”. Community is a set of actors interacting with each other frequently e.g. people attending the same dance classes.

A set of people without interaction is NOT in a community e.g. People waiting for bus at stop but doesn’t talk to each other. People form communities in social media.

NOTE: The term “Community”, first appeared in the book “GEMEINSCHAFT & GESELLSCHAFT”, published in year 1887.

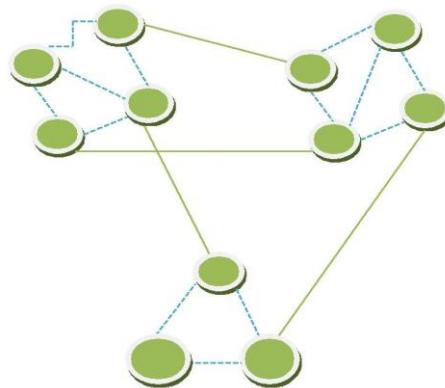


Fig- 1: A simple graph with 3 different communities.

Defining a community is quite a challenging task. The definitions vary from author to author and from algorithm to algorithm.

There are three levels to define a community in the most efficient way:

- Local definitions group.
- Global definitions group.
- Definitions based on vertex similarity.

2.1 Local definition group

In this group, communities consist of “parts of the graph with few ties with the rest of the system. In this the communities are studied from their inner structure independently of the remaining part of the graph. For example, cliques like K-clique, K-club etc.

2.2 Global definition group

In this, a global criterion associated with the graph is used to compute communities. The global criterion is dependent on the algorithm implemented to locate the communities. For this global definition, either a clustering or distance-based criterion may be introduced.

2.3 Vertex similarity based group

In this, communities are considered as groups of vertices similar to one another. In order to detect the similar vertices, the use of hierarchical clustering algorithm is done. The hierarchical clustering aims to find the communities made of similar vertices.

2.4 Community detection

Discovering the groups in a network where individuals “group memberships” are not explicitly given is the concept behind the community detection. Community detection attempts to solve the problem which is the identification of groups of vertices (nodes) that are more densely connected to each other than to the other remaining network. Detecting communities is of great importance in sociology, biology and computer science where systems are often represented as graphs.

Real networks are not random graphs, as they display big in homogeneities revealing a high order and organization.

The distribution of edges is not only globally, but also locally inhomogeneous, with high concentration of edges within special groups of vertices, and low concentrations between these groups. This feature of real network is called “community structure or clustering”.

In a community, nodes are connected with each other based on their human relationships like friendship, colleague etc. In computer science, communities can be regarded as sub-graphs of networks. The whole complex network can be generated as a graph, which is consisted of many sub-graphs.

Community detection attempts to solve the problem which is the identification of groups of vertices, that are ‘more densely connected’ to each other than to the rest of the network. Detecting and analyzing the community structure of network ranges to important findings in wide range of domains like biology to social sciences to web.

Community detection has increasing interest in applying on social media not only as a means of understanding the underlying phenomena taking place in such systems, but also to exploit its results in intelligent services and applications e.g. automatic event detection in social media content.

2.5 Types of communities

The communities are broadly divided into two types:

a) Disjoint community: In this community, node will only belong to a single community. There will be no overlapping between the nodes. This type of communities is called as CRISP assignment for the nodes in which only binary relationship is allowed between any node and community. A node will be in at most one community or none.

b) Overlapping community: In this community, a node may belong to more than one community at the same time. This is sometimes called as the FUZZY assignment of the nodes in which a node may present in different communities simultaneously.

3. WHY WE STUDY SOCIAL MEDIA NETWORKS?

Different types of social media networks are there and it is very important to study and at the same time analyze them. These social media networks are the significant source of the INTELLIGENCE, as they encode the online activities properly and efficiently and the inputs of the social media participants at the same time.

3.1 Advantages of studying and analyzing social media networks

- By analyzing such networks, we gain insights into the social phenomena and processes that take place in our world.
- Also we can extract actionable knowledge which is beneficial in several information management and the retrieval tasks, for example online content navigation, online recommendation etc.
- We know, as social networks starts gaining prominence, the first obvious question that comes to a researcher’s mind in observing these networks is, How to extract the meaningful knowledge from these data?
- The identification of these high order structures within the networks will yields insight into the functional organization, which contributes more knowledge while offering some possible actions like marketing plans, recommendations and user interface adaptations.

3.2 Network and Representation

Social networks can be represented by two ways:

- Graphical Representation:
Nodes = members.
Edges = relationships.

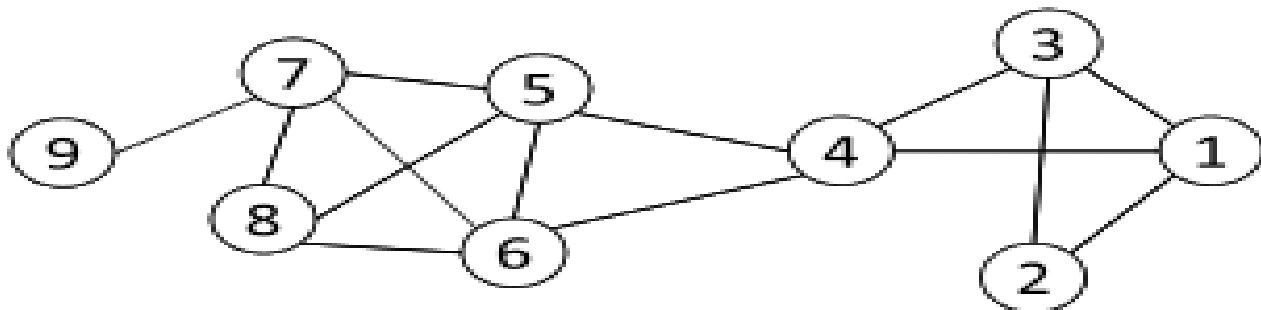


Fig-2: Graph Representation

- Matrix Representation:
Social networks can also be represented as matrix called **SOCIOMATRIX**.

Node	1	2	3	4	5	6	7	8	9
1	-	1	1	1	0	0	0	0	0
2	1	-	1	0	0	0	0	0	0
3	1	1	-	1	0	0	0	0	0
4	1	0	1	-	1	1	0	0	0
5	0	0	0	1	-	1	1	1	0
6	0	0	0	1	1	-	1	1	0
7	0	0	0	0	1	1	-	1	1
8	0	0	0	0	1	1	1	-	0
9	0	0	0	0	0	0	1	0	-

Fig-3: Matrix Representation

3.3 Why Communities in Social Media?

- Human beings are generally social in nature.
- The social media is easy to use and it allows people to extend their social life in unprecedented ways.
- Generally it is difficult to meet friends in physical (real) world, so it is much easier to find friend online with similar interests.
- The interactions between the different nodes may help in determining the communities.

Two types of groups in social media:

- **Explicit groups:** formed by user subscriptions.
- **Implicit groups:** implicitly formed by social interactions.

Network “interactions” provide rich information about the relationships between the users. Groups which are implicitly formed can complement other kinds of information e.g. user profile, help network visualization and navigation and provide basic information for other tasks, like recommendation and therefore each of the information can act as a research topic.

3.4 Taxonomy for the Community Criteria:

The criteria roughly depend on the tasks and community detection is divided into 4 categories:

- **Node-centric community:** Each node in a group satisfies certain properties.
- **Group-centric community:** Consider the connections within a group as a whole. Group has to satisfy certain properties without zooming into the node-level.
- **Network-centric community:** Partition the whole networks into several “disjoint sets”.
- **Hierarchy-centric community:** Construct a hierarchical structure of communities.

4. LITERATURE SURVEY:

Community detection is a stimulating field of research. The purpose of this paper is to make you familiar with the methods and the concepts of the community detection in online social networks. The online social networks (OSN) have become embedded in our everyday lives so much that we cannot ignore it. One specific interest in social networks is that of detecting overlapping communities, instead of considering online communities as autonomous islands acting independently, communities are more like sprawling cities bleeding into each other.

The assumption is that online communities behave more like complex networks and this creates new challenges especially in the area of size and complexity.

The algorithms for detecting these overlapping communities need to be fast and accurate. The detection of overlapping community structure in networks can give insight into the structures and functions of many complex systems.

Now, the study done by the different researchers in the field of community detection can be summarized as follows:

a) **“Detecting Community Structure in Networks”** by **M.E.J Newman**, Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor :

In this paper, the focus is given on reviewing the different algorithmic method for finding community of densely connected vertices in the network data. The discussion of some of the traditional approaches, such as **spectral graph partitioning** and **hierarchical clustering** is also been done, but further it was found number of shortcomings as far as the concern for the analysis of the large real-world network. There was also the description of the, methods based on Iterative removal of “Between community edges”, which also includes the “betweenness-based method” of Girvan & Newman and Monte Carlo resample variation ,proposed by Tyler and also the algorithm based on “counts of short loops”, which was proposed by Radicchi. The two recent algorithms which was good for their computational efficiency [1]:

- Modularity Maximization algorithm of Newman.
- Resistor Network algorithm of Wu and Huberman.

b) **“Identifying overlapping communities in networks using evolutionary method”** by **W.Zhan, J.Guan & H.Chen** of Journal Physica A:

In this paper, the presentation of an encoding scheme for an overlapping partition of a network is done. The two informativeness measure for a node is proposed and presents an evolutionary scheme between two segments over the population. This evolutionary method was for detecting overlapping community structure in the network. For the representation of the overlapping part of the network , there has been developed an encoding scheme composed of two segments, the first one represents a disjoint partition and the other one represents an extension of the partition which allows the multiple membership.

The work has been done to give two measures for the informativeness of a node and then present a co evolutionary scheme between two segments over the population which ultimately solve the overlapping partition of the network. Later, the experimental results reveal that this method can give a better result to the network and showed that a best overlapping partition of the network might not be rooted from a best disjoint partition. The variation of overlapping modularity is also presented and the experiments done on this study indicate that through optimizing the measure, the method can yield a better result. But, it could suffer from the problem of Resolution limit, when particularly applied to a large heterogeneous network.

When the method has to be applied to a large network, the focus has to be given to the efficiency of the evolutionary method. It appears if the method is combined with multi-level technique, then this problem can be resolved [2].

c) **“Detecting Overlapping communities by seed community in weighted complex networks”** by **J.Li, X.Wang & J.Eustace** of Journal Physica A:

In this paper, it is clearly presented that detection of the community structures in the weighted complex networks is really significant for understanding the structure of the network and the further analysis of the network properties. A unique algorithm is presented for detecting overlapping communities in the weighted complex networks along with the considerable accuracy. A new EM-BOAD algorithm has been proposed for detecting the overlapping community structures from the weighted networks, which is somewhat better than the previous algorithms. For the given weighted network, all the seed communities are first extracted with two characteristics and then the absorbing degree has to be computed between the seed community and its neighbors and then the seed community absorbs new members and the larger seed community goes on expanding. The algorithm successfully mined the common vertex and further tests are also performed on the different real-world networks and it shows that EM-BOAD algorithm performs better than the GN algorithm, GCE algorithm and hence it was supposed to be taken as a good algorithm for detecting overlapping communities in the weighted networks [3].

d) **“Structural & functional analytics for community detection in large scale complex networks”** by **P.Chopade** of Journal of Big Data:

In this paper, the discussion has been done on the community dynamics and the complex network structural parameters. The importance of the network centrality or the degree centrality and also network robustness for the community detection has been highlighted in the paper.

This centrality is correlated with the degree. The network or degree centrality based on the modified Laplacian, weighted micro community centrality also discussed. The introduction of the k-clique sub community for weighted modularity optimization and overlapping Community

Detection based on degree & weighted micro-community centrality has been done. This will help in the identification of the hidden level vulnerabilities. It has been stated that network centrality and robustness will help for supervised community detection in overlapping communities.

The problem still needed to be resolved that is selection of the parameter k. Further study is to be done how to determine an appropriate k for a given network and also for the functional dynamics of the complex networks by including network centrality and weighted clustering coefficient which will help to identify the micro-level communities and also their associated relationship [4].

e) **“Personalized recommendations based on time-weighted overlapping community detection”** by **H. Feng& J.Tian** of Journal Information and Management:

In this paper, a recommendation based approach called TOTAR (Temporal overlapping community detection using time weighted Association Rules) is proposed which is based on time weighted overlapping community detection and association rule mining. The different approaches have been incorporated to synchronize the time effects in the proposed algorithm to improve its performance and predict the user’s dynamic interests over the time. Different data sets has also been used from MOVIELENS and NETFLIX and performance is compared with other algorithms especially the accuracy and diversity [5].

f) **“Overlapping community detection in social network using Parliamentary Optimization Algorithm”** by **F. Altunbey** of International Journal of Computer Networks & Applications:

In this paper, a Novel overlapping community detection algorithm, which tries to optimize the network modularity using the parliamentary optimization algorithm with the fitness function has been proposed. The POA (Parliamentary Optimization Algorithm) has been first time proposed and further some modifications or additions are also performed to the algorithm and the results are promising. This designed POA can help to the analysis of the community structure and also detect the overlapping community.

Further, it was noted that only the modularity measure has been used as the fitness function for finding out the overlapping community of a network and the work has been done to enhance this measure. The POA also have to be generalized for the multi-objective purposes in the large network [6].

5. COMPARISON OF DIFFERENT LITERATURES AND THEIR FINDINGS & GAPS:

S.NO.	Literature Name &Authors	Work done by Authors	Findings & Research gaps in literature
1.	Survey articles by Fortunato&Castellano.	Concerned with methodological foundations of community detection.	It describes community detection in a generic context, mostly under the statistical physics perspective, thus lacking any association to Web mining and Social Media research.
2.	Study by Danon.	Presents a comparative discussion on the computational complexity of several community detection methods.	It disregards their memory requirements, & other scalability considerations, such as possibility for incremental computations & it further lacks a Social Media context.

3.	Study by Schaeffer.	Presents generic overview of graph clustering which can be considered as community detection.	It lacks the association with the social networking modules and the consideration of the nodes in the social networks.
4.	Article by Tang & Liu.	Studies several social networks analysis problem with emphasis on community detection relating them with social media domain.	Limited to the methodological principles of methods, discussing neither the performance attributes of methods nor the interpretation & exploitation of their results.
5.	Detecting community structure in network by M.E.J Newman.	Reviewed the different algorithmic methods of finding communities. And discusses traditional approaches like spectral graph partitioning & hierarchical clustering.	There were many shortcomings, if the concern was given to the large real-world networks. Due to this new methods are developed which are flexible enough to generate new structure.
6.	Identifying overlapping community in network using evolutionary method by W. Zhan & H. Chen.	The encoding scheme for the overlapping partition of the network has been proposed and then a evolutionary scheme is developed to partition the network.	It was found that this method gives better result to the network & also shows that the overlapping partition of the network might not be part of the disjoint partition. But it suffers the problem of the Resolution Limit, when applied to large heterogeneous networks.
7.	Efficient Overlapping community detection in huge real-world networks by Z. Wu, Y. Lin & S. Tian.	An efficient algorithm CONA has been proposed which is capable of detecting overlapping communities in large social networks.	The algorithm is based on deduced conditions of overlapping nodes & aimed at identifying overlapping nodes from boundaries and inner nodes. Algorithm is found to be easy for extending weighted network by replacing degree of the node to the sum of link weights. But, still there was no algorithm which can detect highly overlapping community in large networks. It fails to work in the real large world networks.
8.	Detecting overlapping community by seed community in weighted complex network by J. Li & J. Eustace.	Presented that detecting community structures in weighted network is significant. And for that a unique algorithm for this purpose is proposed i.e.	The algorithm works well in the weighted complex network and aimed to extract the seed communities and absorbing is also computed. But when applied to the different real world networks,

		EM-BOAD (Extended Modularity Based On Absorbing Degree).	it was difficult to identify the starting seed to make the community grow as well as the absorbing degree calculation was also cumbersome.
9.	Overlapping community detection approach in online social network (OSN) by X. Zhang & W. Zhu.	For detecting overlapping community detection in OSN, a novel index CLA (Classification Accuracy) is proposed to evaluate the quality of detecting communities.	Traditional network & OSN datasets are used for checking algorithm and it was found the algorithm performs well than other algorithms. Algorithm performs well but slower than already existing CNM & COPRA, but the division accuracy of algorithm was not well and the time complexity of the algorithm was also higher.
10.	An ant colony based algorithm for overlapping community detection in complex network by X. Zhou & T. Liu.	An ant -colony based overlapping community detection was proposed and the heuristic formula for the calculation of similarity between the nodes was also redefined.	The algorithm performs better than the other algorithms & was inspired by behaviors of gregarious ant colonies. The experiments show that Ant CBO algorithm performs well and has high NMI value than CPM & COPRA. But the accuracy factor for uncovering the overlapping communities in the large network was not proper and the weighted and directed graphs were not considered.
11.	Overlapping community detection in social network using Parliamentary Optimization algorithm by F. Altunbey.	A novel overlapping community detection algorithm i.e. POA parliamentary optimization algorithm is proposed.	It has been first time when POA has been proposed and it facilitates in the analysis of the community structure and detects overlapping communities. But it only considers one measure of the network to calculate the overlapping communities but there are different indexes which affects the overlap between the communities. So considering only modularity measure as fitness function will not be sufficient to calculate overlap in the network.
12.	Finding Overlapping community in social networks, a rigorous	The framework has been proposed for rigorously thinking of communities	These techniques are based on the sampling and they infer the global structure of the vertices of the

	<p>approach by S. Arora & G. Schoenebeck.</p>	<p>and concepts of cliques, dense sub-graphs etc.</p> <p>And gap assumption, expected degree model was also proposed.</p>	<p>graph.</p> <p>The expected degree model is a static model, but the community evolution is dynamic process. So, static model could not be able to deal with the dynamic network evolution and clumps and hierarchies are not resolved with the model.</p>
13.	<p>Structural & functional analytics for community detection in large scale complex networks by P. Chopade.</p>	<p>The discussion has been done on the dynamics of the communities and the complex network structure parameters.</p> <p>The parameters of network centrality and network robustness are highlighted in the paper. And k-clique sub community & weighted micro-community detection has also proposed.</p>	<p>The centrality of the network or the degree is important for the detection of the overlapping communities, as centrality is related with the degree.</p> <p>But the selection of the parameter k needs to be resolved. The selection of k is time consuming and difficult, as functional dynamics of the network also not included at that point.</p>
14.	<p>Personalized recommendations based on time weighted overlapping community detection by H. Feng & J.Tian.</p>	<p>An algorithm based on the recommendation is proposed i.e. TOTAR (Temporal overlapping community detection using time weighted association rules).</p>	<p>TOTAR performs well in comparison to other algorithms & is based on the association rule mining.</p> <p>As the time effects is incorporated in the proposed algorithm, so to clear out the dynamics changes in user's interests over time is difficult. Algorithm uses some real time datasets and the dynamic interest's prediction needs to be improved.</p>
15.	<p>Community structure in social & biological networks by M. Girvan & M. Newman.</p>	<p>The method was proposed in the paper to detect the communities in the network, by analyzing information about the edge betweenness.</p>	<p>Previous methods find out strong connected cores of the communities, but this method finds out the edge betweenness and finds community's peripheries. It was found that it performs well on the computer-generated graphs & detects structures. But it was not able to use or handle the weighted & directed graphs and the speed of</p>

			the algorithm was also high i.e. $O(n^3)$ which is impractical for large networks.
16.	Clustering Online Social network community using Genetic Algorithm by M. Hajeer & S. Snayal.	In this, Genetic Algorithm is proposed and data mining method is used to find communities from the OSN.	<p>The term NOA is introduced which is important and represents node which attracts the other nodes the most.</p> <p>But the solution space is very large, so when this GA is used and the clustering of all the network nodes into groups is done; at this same time the track of NOA's also have to be recorded. This is pretty difficult, so this needs to be properly synchronized.</p>

6. APPLICATIONS:

1. Community detection is very helpful in the information recommendation because in this the members of community have few common preferences.
2. We detect communities to understand behavior of large scale social network as it will clarify the information sharing and information diffusion processes.
3. In biological network communities helps us to understand basic mechanisms which control normal cellular processes.
4. In the network community of customers with similar interests can be used to make recommender system to enhance the business.
5. In the field of online content navigation, it is necessary to identify the communities in the network and that's why the online social network is analyzed.
6. Till date, recommender systems make recommendations by getting data from the data warehouses. But due to dynamic nature of recommender system, data has to be distributed. So recommender system has to work parallel to provide recommendation to the user and to support different interfaces. In future, analysis of recommendation system in social networks quantitatively is done.

7. CONCLUSION:

Identifying meaningful community structure in social networks is a hard problem, and extreme network size or sparseness of the network compound the difficulty of the task. With a proliferation of real-world network datasets there has been an increasing demand for algorithms that work effectively and efficiently. Existing methods are limited by their computational requirements and rely heavily on the network topology, which fails in scale-free networks. Yet, in addition to the network connectivity, many datasets also include attributes of individual nodes, but current methods are unable to incorporate this data. Cognizant of these requirements, there is a need to propose a simple approach that steers away from complex algorithms, focusing instead on the edge weights; more specifically, leverage the node attributes to compute better weights.

Overlapping community detection is still a challenge. Though there are several proposed methods, but most of them take a huge amount of processing time. So emphasis should be given to effective algorithms which will be able to detect communities in a huge social network in allowable time. The weighted and directed networks are needed to be considered for community detection. Nowadays almost all social networks are dynamic i.e. some members are joining and some are leaving every moment. So it will be great if communities can be detected in dynamic networks.

8. ACKNOWLEDGEMENT:

I would like to thank all the respective authors of the papers and the guide for continuously providing me support and courage to do my work. And also to the academic sources and the environment provided to the students in the Central University of Punjab.

9. REFERENCES:

- [1] M. E. J. Newman, "Detecting community structure in networks," *cenetr for study of complex systems*, pp. 50-55, 2008.
- [2] W. Zhan, "Identifying overlapping communities in networks using evolutionary method," *Physica A*, pp. 182-192, 2016.
- [3] J. Li, "Detecting overlapping communities by seed community in weighted complex networks," *Physica A*, pp. 6125-6134, 2013.
- [4] P. Chopade, "Structural and functional analytics for community detection in large-scale complex networks," *Journal of Big Data*, pp. 1-28, 2015.
- [5] H. Feng, "Information & Management Personalized recommendations based on time-weighted overlapping community detection," *Information & Management*, pp. 789-800, 2015.
- [6] F. Altunbey, "Overlapping Community Detection in Social Networks Using Parliamentary Optimization Algorithm," *IJCN*, pp. 12-19, 2015.
- [8] A. R. Hajibagheri, "Social network community detection using the shapley value," *Shiraz university*, pp. 51-65, 2013.

10. BIOGRAPHIES:



Ms. Richa received her B.Tech degree from IMS ENGINEERING COLLEGE, Ghaziabad and currently pursuing her M.Tech from Central University of Punjab, Bathinda.