International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 09 | Dec-2015 www.irjet.net p-ISSN: 2395-0072

Overlapping Community detection Algorithms:-A Review

Mini Singh Ahuja[1], Jatinder Singh[2], Neha[3]

[1] *Research scholar, Department of Computer Science, Punjab Technical University, Punjab, India*
[2]*Professor, Department of Computer Science, KC Group of Institutes Nawashahr, Punjab, India*
[3]*Student (M.Tech), Department of Computer Science, Regional Campus Gurdaspur, Punjab, India*

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

Abstract - ***Community detection is an important task in the study of network system as it provides information about overall network structure in depth. Community is the division of network nodes into subgroups in such a way that nodes inside the subgroup have more connections as compared to connections between the subgroups. Node may belong to more than one group. In this paper several overlapping community detection algorithms are considered in order to detect these overlapping nodes. Several modularity measures are used to measure the quality of communities detected by such algorithms.***

Keywords: *Complex network, Community Structure, Overlapping community.*

1. INTRODUCTION

Network is a collection of entities that are interconnected with links. For example: people that are friends, computers that are connected, and web pages that point to each other. In graph theory, entities are vertices and links are edges. Large graph of real life are called complex network. Real-world complex systems are composed of interacting entities with nontrivial dynamical behavior and complicated interaction topology. Examples of Complex network are: Internet, WWW, Transport networks, Food webs, Social Networks etc [1].

Social networking is very important application because it has a unique ability to make social contact over Internet for geographically dispersed users. Social network is a finite set of nodes which represents single person or group and edges represents relations among them [2]. Social networking is an application that enables users to communicate with each other. For example Facebook, twitter etc.

2. COMMUNITY

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 [3].

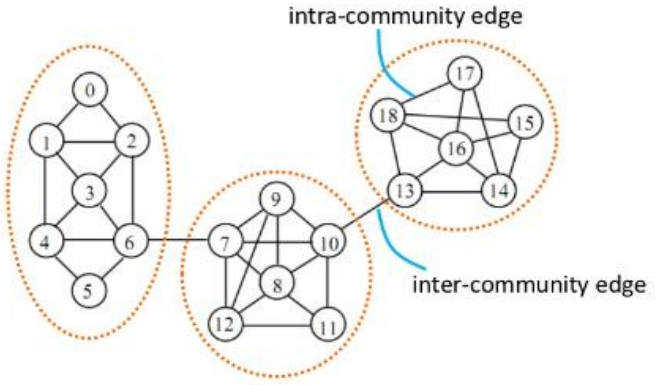

Fig- 1: Shows three different communities.

Definitions of community can be classified into the following three categories:

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

2.1 Local definition

The local definition focuses on the vertices of the sub network under investigation and on its immediate neighborhood. Local definitions of community can be further divided into self-referring ones and comparative ones. The examples of self referring definitions are clique, n-clique [4].

2.2 Global definition

Global definitions of community characterize a sub network with respect to the network as a

whole .These definitions usually starts from a null model. Null model is designed simply by introducing randomness in the distribution of edges among vertices. Then, the linking properties of sub networks of the initial network are compared with those of the corresponding sub networks in the null model. If there is a wide difference between them, the sub networks are regarded as communities. The most common null model is given by Newman Girvan [5].

2.3 Vertex similarity:

According to this definition, community is a group of vertices which are similar to each other. To find similar vertices we use hierarchical clustering algorithm. Hierarchical clustering is a way to find several layers of communities that are composed of vertices similar to each other [6].

2.4 Community structure

Community structure is also named as cover of community. Community structure is a set of communities present in network. It is represented as C= {c1, c2, c3, c4…ck}. Here C is the community structure and c1, c2, c3, ck are communities. For example there are two communities c1= {1, 2, 3} and c2= {3, 4, 5}. Thus community structure is C= {c1, c2}.

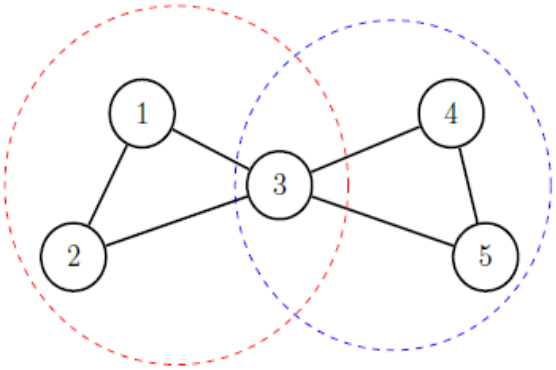

Fig-2: Shows community structure having two Communities c1 and c2.

2.5 Types of communities

Community can be of two types:

a) Disjoint community: In disjoint community 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 1 community and at-least 0 community (none) [7].

.

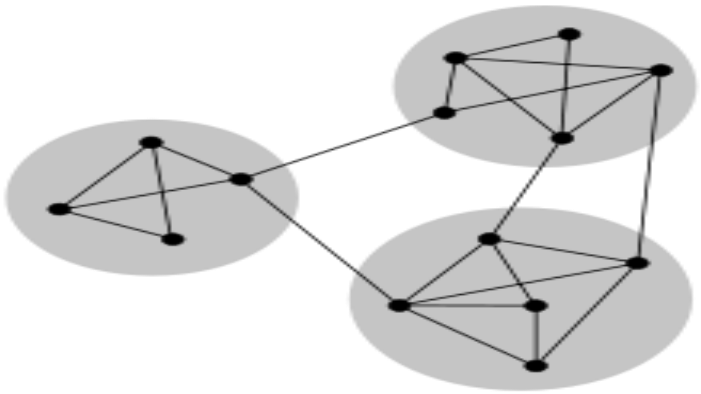

Fig-3: Disjoint communities.

b) Overlapping community: In overlapping community a node may belongs to more than one community [8]. This is known as fuzzy assignment of nodes where, a node may belong to more than one community.

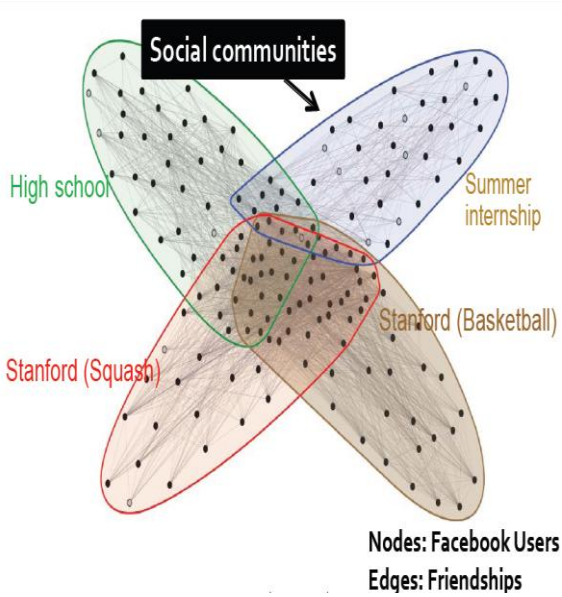

Fig.-4: Overlapping communities in Facebook network.

In social network extracting such community structure is very useful as it helps us to study the overall structure of network. It is a very challenging task. The concept of community discovery is similar to graph partitioning but there are some differences like in graph partitioning the number of groups and their size is already known to us but in case of community detection we don't know about the number of communities in network and the communities may not be of same size.

3. LITERATURE SURVEY

Community detection is a stimulating field of research. The purpose of this paper is to help you to understand the roles of overlapping communities in various fields. There are various algorithms for community detection but most of the algorithms fail in detection of overlapping communities they can only detect disjoint communities. Some of these algorithms are kl graph partitioning algorithms, multilevel graph partitioning algorithms, hierarchical clustering, divisive algorithms [9] etc.

Palla et al in 2005 [10] presented the first overlapping community detection algorithm. In this approach communities were identified based on the k-cliques. According to this algorithm node may belongs to many communities Clique is a subset of nodes where every node is adjacent to every other node. K-clique represents size of clique. for example 4-clique indicates a sub graph having 4 nodes.

Lancichinetti et al. in 2009 [11] presented a method which uncovers both overlapping community structure and hierarchical properties of complex network .This algorithm is based on local optimization of a fitness function. This algorithm detects overlapping communities by maximizing the fitness value. Community structure was exposed by peaks in the fitness histogram.

Shen et al. in 2009 [12] proposed another overlapping community detection approach. Overlapping communities are detected based on maximal cliques. An overlapping modularity measure was proposed here based on number of maximal cliques. In this way we can find overlapping communities by partitioning the maximal clique network by using any of the modularity optimization method.

Gregory in 2009 [7] presented a two phase method for detecting overlapping communities. During the first phase of this method a new network is formed from the existing network by splitting node using split betweeness concept. After this, in second phase disjoint community detection algorithm has been applied to this new network. In this way this approach has ability to convert disjoint community detection algorithm to overlapping community detection algorithm.

Ahn et al in 2010 [13] presented another overlapping community detection algorithm based on link partition. Using hierarchical clustering, links are partitioned to link dendogram. Overlapping communities are detected by cutting this dendogram at some threshold point. Here modularity measure is based on partition density.

Chen et al in 2010 [14] proposed another algorithm for overlapping community detection in weighted network. It detects overlapping communities using a local algorithm which works by expanding a partial community which is started from a special single node.

Lazar et al. in 2010 [15] proposed an overlapping modularity measure for overlapping communities based on difference between inward and outward edges. In this approach author deals with a non-fuzzy measure which has been designed to rank the partitions of a network's nodes into overlapping communities. Such a measure can be useful for both quantifying clusters detected by various methods and during finding the overlapping community structure by optimization methods.

Coscia et al in 2012 [16] presented local first approach to detect overlapping communities in the complex network. In it each node votes for the communities it sees surrounding it using a label propagation algorithm and finally, the local communities are merged into a global collection. This method has limited time complexity so that it can be used on large scale networks.

Junqiu Li et al. in 2013 [17] said that in weighted complex networks, community detection was considerable to understand the structure and properties of network. He proposed a unique

algorithm to discover overlapping communities in the weighted networks. In the first step of his algorithm he detected all the seed communities. After that more community members are immersed by absorbing degree function. This algorithm successfully detects the nodes that belong to more than one community.

Reza Badie et al. in 2013 [18] said that Communities are groups of nodes forming strongly connected units in networks. Various nodes can be shared among different communities of a Network. He proposed a novel algorithm that is capable to find both types of structures in Complex networks such as overlapping and without overlapping community structure. This algorithm is based on the concept of nodes 'closeness and improve the result of label propagation algorithm.

M. Yaozu Cui et al. in 2014 [25] presented an algorithm for detecting overlapping communities in complex network. In this paper, author used various types of theories i.e. clustering coefficient and maximal sub-graph which are rolling in between two neighboring communities. Firstly all the maximal sub-graphs are eliminated from the original networks and then they merged them on the basis of clustering coefficient of two neighboring maximal sub-graphs. In this paper, an additional feature is also anticipated to enhance the algorithm i.e. a new extended modularity. Overlapping vertex can be discovered with the help of this algorithm. Then at last author compared the results of his algorithm with other correlated algorithms. But this paper leads to unsatisfactory results.

4. ALGORITHMS FOR OVERLAPPING COMMUNITIES

Community detection is NP-hard problem and there are many approaches for solving these problems like hierarchical clustering, Kernighan-Lin (KL) Graph Partitioning, divisive algorithm, multilevel graph partitioning algorithms but most of the algorithms are failed in the detection of overlapping communities. There are many algorithms to detect overlapping communities. Some algorithms are explained below:

Clique percolation method (CPM): Palla et al. [10] presented clique percolation approach which is based on k-cliques to discover overlapping communities in network. Clique is a subset of nodes where every node is adjacent to every other node. K-clique represents size of clique for example 4-clique indicates a sub graph having 4 nodes. In the following figure there are six 3-cliques that are a (A,B,C); b (A,B,H); c (B,D,E); d (B,D,F); e (B,E,D);f (D,E,F) and one 4-clique (B,D,E,F).

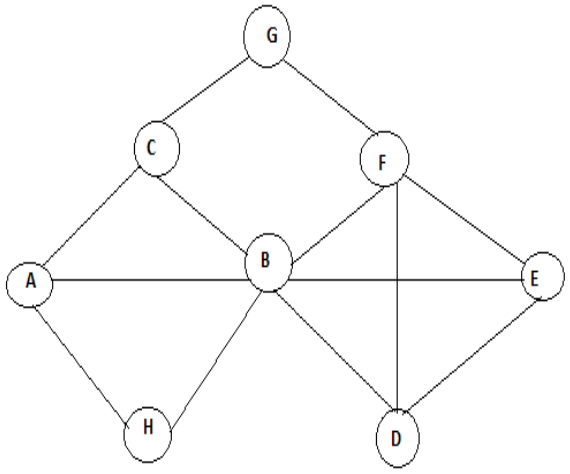

Fig-5: shows network having 3-clique and 4-clique.

Steps of the algorithm are given below:

Input: Network N, clique size – k

Output: Community structure C

Step1: In the first step, find all k-cliques in the network N.

Step2: Now construct a Clique graph Gc where each node represents clique which is identified in first step of algorithm. In clique graph two cliques are connected with an edge only if they share k-1 members.

Step3: In clique graph each connected component represents community.

Link Clustering:

This algorithm detects overlapping communities. In this algorithms set of links are partitioned instead of the set of nodes. So line graph L(G) is used. The line graph L(G) of an undirected graph G is the graph L(G) in which every node represent an edge of G. Two nodes in L(G) are adjacent only if their corresponding edges share a common endpoint in G. In this way the line graph represents the adjacency between edges of G. The main advantage of this approach is that it produces an overlapping graph of

division of the original graph thus nodes to take part in more than one communities [23].

Cluster Overlap Newman Girvan Algorithm (CONGA):

Gregory [7] presented CONGO algorithm which is based on Girvan Newman community detection algorithm but extends to detect overlapping communities. This algorithm based on the concept of split betweeness. In this a vertex is split up into two parts in such a way so that each part keeps some original egdes. Thus split betweeness is defined as egde betweeness of imaginary edge between two new vertices.

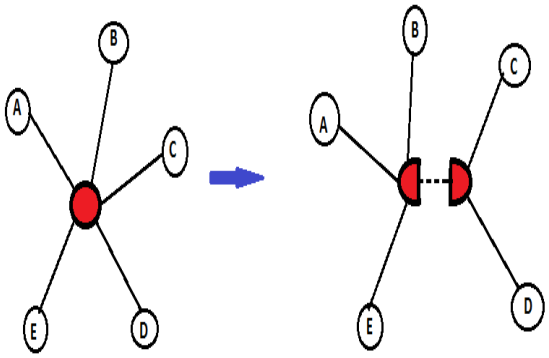

Fig- 7: Defines split betweeness.

Steps for algorithm are:

Step1. Firstly calculate edge betweeness of edges and split betweeness scores of all vertices.

Step2. a) Split the vertex at the optimal split if the maximum split betweenness is greater than maximum edge betweeness.

b) Otherwise delete the edge with maximum edge betweeness.

Step3. Recalculate all edges and split betweeness scores.

Step4. Repeat from step 2 until all edges have been removed.

Label propagation algorithm:

Raghavan et al [22] proposed this algorithm for community detection. Label propagation algorithm (LPA) is an extremely fast community detection method and is widely used in large scale networks Label propagation algorithm can be used for both disjoint as well as overlapping community detection. The main idea is that the nodes adopt the label that most of its neighbor nodes have and thus communities are formed on the basis of their label.

Steps for algorithms are:

Step1 Firstly a unique label is assigned to each vertex.
Step2 At each iteration of algorithm each node adopts the label that the maximum of its neighbor nodes have, ties are broken at random. If more than one label is contained by the same maximum number of its neighbors, then randomly select one from them.
Step3 All vertices carrying the same label are identified with the same community.
The COPRA algorithm by Gregory et al. [6] extends this approach to find communities. In this algorithm a vertex has a list of labels with corresponding belonging factors between 0 and 1. In the update step each vertex averages the belonging factors of its neighboring vertices and drops labels whose belonging factor is below some threshold.

Eagle algorithm:

Eagle algorithm [19] is an agglomerative hierarchical clustering method for overlapping community detection. Eagle algorithm basically deals with set of maximal cliques instead of set of nodes. This algorithm has two stages. In the first stage, dendogram is created and in the second stage we select appropriate cut method to cut this dendogram into communities.
Steps for this method are:

Step1 Discover all the maximal cliques in the network. The clique which is not subset of ant other cliques is known as maximal clique.

Step 2 Identify the pair of communities having maximum similarity and merge them to obtain new community. Then calculate similarities between this new community and other communities.

Step3 Repeat the above step until only one community left.

In this algorithm the similarity between two communities is defined as:

$$M = \frac{1}{2m}\sum_{v\in c1, w\in c2, v=w}\left[A_{vw} - \frac{k_v k_w}{2m}\right]$$

A_{vw} Represent element of adjacency matrix. If there is an edge between vertex v and w then it takes value 1 otherwise 0.

5. APPLICATIONS

1. Community detection can be used for information recommendation as the members of community have some similar interests and preferences. [20]

2. Communities will also help us to understand the structure of social network. Communities clarify the properties and functions of network. [20]

3. We detect communities to understand behavior of large scale social network as it will clarify the information sharing and information diffusion processes.

4. Community detection methods are of great advantages in social-aware routing in MANETs and worm containment on social networks.

5. In biological network communities helps us to understand basic mechanisms which control normal cellular processes.

6. In the network community of customers with similar interests can be used to make recommender system to enhance the business [21].

7. Complex graphs can be easily visualized with the help of community detection.

8. In World Wide Web link farms can be easily detected with community detection. A link farm is a group of websites which hyperlink to every other site in the group.

9. Human social networks have the property to show strong community structure. A network with strong community structure consist communities and these communities have multiple connections within them and less connections between communities. Community structure not only affects the spread of infectious disease in the community but also protects the network from large scale epidemics [23].

6. CONCLUSION

Overlapping community detection is a difficult task. In this paper we have studied various overlapping community detection algorithms for example clique percolation method based on k-cliques to detect the overlapping communities. Another algorithm is CONGO which is based on idea of split betweeness to detect overlapping communities. Link clustering and label propagation algorithms are also used to detect such communities. Eagle algorithm detects hierarchical and overlapping community structure in network. This topic of research is very useful in several fields like biology, physics, social science etc. With the community detection large complex graphs can be easily visualized.

REFERENCES

[1] M.E.J Newman, "The structure and function of complex network" Volume 2, pp, 167-256, 2003.
[2] L. Tang, H. Liu, "Community Detection and Mining in Social Media" Morgan & Claypool (2010).
[3] Karsten Steinhaeuser and Nitesh v. Chawla "Community detection in large real world networks".
[4] Symeon Papadopoulos · Yiannis Kompatsiaris · Athena Vakali and Ploutarchos Spyridonos "Community detection in Social MediaPerformance and application consideration" 2011.
[5] Newman, M.E.J., Girvan, M., "*Finding and evaluating community structure in networks*, Physical Review E", 69(026113), 1–16, 2004
[6] Borko Furht Florida Atlantic University,"Handbook of Social Network Technologies and Applications".
[7] S. Gregory, "Finding overlapping communities using disjoint community detection algorithms, in Complex Networks", pp. 47–61, Springer, 2009.
[8] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: the state of the art and comparative study", arXiv preprint arXiv: 1110.5813, 2011.
[9] Charu C. Aggarwal "Social Network Data Analytics".
[10] G. Palla, I. Der´enyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society",Nature, vol. 435, no. 7043, pp . 814– 818, 2005.
[11] A. Lancichinetti, S. Fortunato, and J. Kert´esz, "Detecting the overlapping and hierarchical community structure in complex networks," New Journal of Physics, vol. 11, no. 3, p. 033015, 2009.
[12] H.-W. Shen, X.- Q. Cheng , and J.-F. Guo, "Quantifying and identifying the overlapping community structure in networks, Journal of Statistical Mechanics: Theory and Experiment, vol. 2009, no. 07, p. P07042, 2009.

[13] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks,Nature", vol. 466, no. 7307, pp. 761–764, 2010.
[14] D. Chen, M. Shang, Z. Lv, and Y. Fu, "Detecting overlapping communities of weighted networks via a local algorithm," Physica A: Statistical Mechanics and its Applications, vol. 389, no. 19, pp. 4177–4187, 2010.
[15] A. L´az´ar, D. ´ Abel, and T. Vicsek, "Modularity measure of networks with overlapping communities, EPL (Europhysics Letters)", vol. 90, no. 1, p. 18001,2010.
[16] M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi, "Demon: a local-first discovery method for overlapping communities",in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 615–623, ACM, 2012.
[17] J. Li, X. Wang and J. Eustace, "Detecting overlapping communities by seed community in weighted complex networks", (2013) December 1.
[18] R.Badie, A. Aleahmad, M. Asadpour, and M. Rahgozar, "An efficient agent-based algorithm for overlapping community detection using nodes" closeness, vol. 392, Issue 20, (2013), pp. 5231-5247.
[19] Huawei Shen, Xueqi Cheng, Kai Cai, and Mao-Bin Hu , "Detect overlapping and hierarchical community structure in networks" November 2008.
[20]Mini Singh ahuja and Jatinder singh , "Future prospects in community detection". Vol. 4, Issue 5, Oct 2014, 37-48.
[21] P. De Meo, A. Nocera, G. Terracina, and D. Ursino, "Recommendation of similar users, resources and social networks in a social internetworking scenario", Information Sciences, vol. 181, no. 7, pp. 1285{1305, 2011.
[22] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks", Physical Review E, vol. 76, no. 3, Article ID 036106, 2007. View at Publisher ·View at Google Scholar · View at Scopus
[23] Alessia Amelio and Clara Pizzuti, "Overlapping Community Discovery Methods: A Survey, 2014".
[24] Marcel Salathe , James H. Jones , "Dynamics and control of disease in network with community structure", PLoS Comput Biol 6(4): e1000736. doi:10.1371/journal.pcbi.1000736, 2010.
[25] Y. Cui, X. Wang and J. Li, "Detecting overlapping communities in networks using the maximal subgraph and the clustering coefficient", Physica A, vol. 405, (2014), pp. 85–91