Volume 8, No. 4, May 2017 (Special Issue)
International Journal of Advanced Research in Computer Science
REVIEW ARTICLE
Available Online at www.ijarcs.info
ISSN No. 0976-5697

Comparison of Clique Percolation , Fuzzy Detection and Line Graph & Link partitioning algorithms used in overlapping community detection

Er. Harpreet Arora[1]
Asst. Prof.: Dept. of C.S.A[1]
SBBSU, Punjab, India
er.harpreetarora@gmail.com[1]

Er. Rohit Kumar[2]
Lecturer: Dept. of C.S.E.[2]
SBBSU, Punjab, India
roh1986@gmail.com[2]

***Abstract*—The online social media is the great area of research which is explored by the researchers now days. There are number of areas in which overlapping community detection works. In this paper comparison of various algorithms used to detect the overlapping will be made. The disjoint clusters can be identified by the use of algorithm. The framework for evaluating various algorithms will be described which will help disclosing the person's membership in multiple clusters. The cluster will be collection of number of distinct users belonging to only one group. This paper evaluates all the algorithms which are used for overlapping community membership detection.**

***Keywords*—community overlapping, clique percolation, fuzzy detection.**

I. Introduction

Community overlapping or modular structure is one of the most widely used in real world social media as it will elaborate functioning of the system. The dataset will be required in order to detect the overlapping. The dataset is fetched from the Snap University. The data which is fetched is converted into the adjacency matrix [2].The adjacency matrix will indicate that nodes are adjacent or not. The nodes are said to be adjacent if they are connected by the single edge. If nodes are adjacent then there will be 1 in the corresponding adjacency matrix otherwise it will contain 0 [3].The overlapping community detection will be represented through the following diagram.

The overlapping community detection will help in identifying the interest of the community and hence presenting them the content which they most often like. Social networking sites are chosen for this purpose. Network communities represent basic structure for understanding the organization of the real world environment. a community will be a group of nodes which are attached by some logical links. In a social network it is well understood that people in a social network are naturally characterized by multiple community memberships. A person can have link to many active area including people, movies, news, etc. All of the above stated active things are group to which a user may belong.

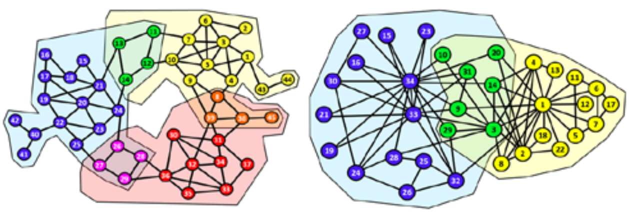


Figure 1: Showing the overlapping community structure

For this reason, there is growing interest in overlapping community detection algorithm that identifies a set of groups which are not disjoint. In the proposed paper comparison of the various techniques used for detection of the overlapping will be made.

A. *Preliminaries*

The basic definitions which are used throughout this paper are presented [1]. Consider a graph

G= {E,V} where E is the set of edges and V is the set of vertexes or nodes. For dense graphs m=O(n2) and for the sparse network m=O(n) [2].The adjacency matrix will be used in order to determine which nodes are reachable from the current nodes. If the elements of the adjacency matrix 1 then it means the node is reachable from the current node otherwise adjacency matrix will contains 0.

In the case of overlapping community detection, the set of clusters found is called covers which are denoted by C= {c1,c2,c3,……cn). [3]In the presented clusters a node can belong to more than one cluster. Each node i associate with a community according to a belonging factor. The association between the nodes and cluster will be indicated with the aic. This aic is a measure of the strength of association between node i and cluster c. There is following constraint to be satisfied in order for the node to be member of the cluster

$$0<=aic<=1 \text{ , } i€V, \quad c€C \qquad (1)$$

II. ALGORITHMS

In order to analyze the overlapping community detection we will use the following algorithms.

A. *Clique Percolation*

The Clique Percolation Method (CPM) is based on the assumption that a community consists of overlapping sets of fully connected sub graphs and detects communities by searching for adjacent cliques. It begins by identifying all cliques of size k in a network [4].Once these have been identified, a new graph is constructed such that each vertex represent those shares k-1 members. Connected graph will be used to detect the overlapping community [1].The relatively small values of k lie between 3 to 6 appear to give accurate results [5].A threshold value will be maintained. Only k cliques with intensity larger than a fixed threshold are included into the community [1]. CPM introduces a sub graph intensity threshold for weighted networks. Only k-cliques with intensity

CONFERENCE PAPERS
National Conference on Emerging Trends on Engineering & Technology (ETET-2017)
On 21st April 2017
University Inst. of Engg. & Tech. & University Inst. of Computer, SBBS University, Punjab (India)

larger than a fixed threshold are included into a community. Instead of processing all values of k, SCP finds clique communities of a given size. In the first phase, SCP detects k-cliques by checking all the $(k-2)$-cliques in the common neighbors of two endpoints when links are inserted to the network sequentially in order of decreasing weights. In the second phase, the k-community is detected by finding the connected components in the $(k-1)$-clique

Projection of the bipartite representation, in which one type of node represents a k clique and the other denotes a $(k-1)$ -clique. Since each k-clique is processed exactly twice, the running time grows linearly as a function of the number of cliques. SCP allows multiple weight thresholds in a single run and is faster than CPM. Despite their conceptual simplicity, one may argue that CPM-like algorithms are more like pattern matching rather than finding communities since they aim to find specific, localized structure in a network.

The result for k=4 will be describe as:

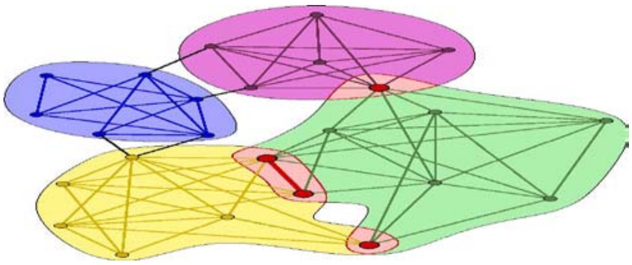


Figure 2 showing the overlapping community of order 4

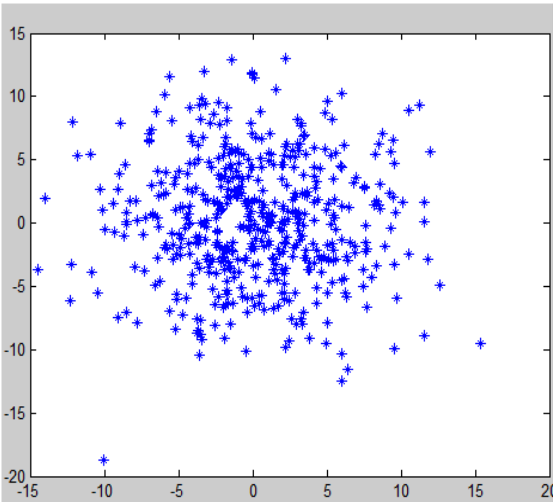


Figure 3 showing k-clique mechanism

B. *Fuzzy detection*

The fuzzy detection mechanism can be used in order to detect the overlapping of community. Fuzzy community model is used to quantify the strength of the associations that exists between nodes and edges. [1]For overlapping community detection generally c mean clustering is preferred. A soft membership function will be defined for this purpose. Algorithms to detect overlapping communities are either "crisp" or "fuzzy" by design: they produce crisp or fuzzy partitions regardless of the type of overlapping in the network. To compare these algorithms consistently, we propose using a common measure: the Fuzzy Rand Index.

a) To evaluate a fuzzy algorithm on a fuzzy network, we compare the fuzzy partition used to construct the network with the one produced by the algorithm.

b) To evaluate a crisp algorithm on a fuzzy network, we first convert the partition found by the algorithm to a fuzzy form by adding equal belonging coefficients for each community.

That is, if vertex v belongs to K communities in the crisp partition, its belonging coefficient is 1/K in those communities and zero in other communities, in the fuzzy partition.[1] One would expect this trivial fuzzy partition to be worse than one found by a good fuzzy algorithm, because it contains no information about the belonging coefficients.

c) To evaluate a fuzzy algorithm on a crisp network, we convert the crisp partition used to construct the network to a fuzzy form in the same way, and compare it with the fuzzy partition found by the algorithm.

d) If both the network and the algorithm are crisp, we convert both partitions (the original one and that found by the algorithm) to fuzzy form and compare them using the Fuzzy Rand Index. In this special case, the partitions could instead be compared by the Omega Index, these two measures are very similar.

so we use the Fuzzy Rand Index for consistency. Finally, we describe a simple procedure for obtaining a non-trivial fuzzy partition from a crisp one. For each occurrence of vertex i in community c, we add a belonging coefficient αic which equals the number of i's neighbors that occur in c divided by the size of c, normalized in the usual way. This technique, which we call Make Fuzzy, can be used to convert any crisp algorithm to a fuzzy one, which may produce better solutions than the crisp algorithm; we test this hypothesis in our experiments.

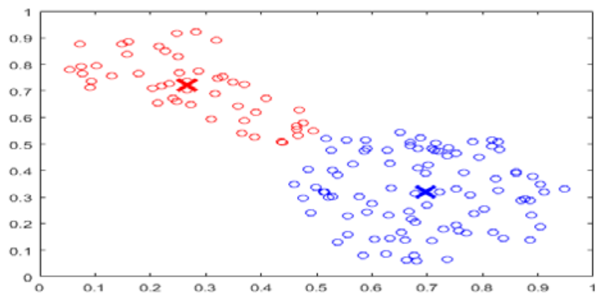


Figure 4 Showing Fuzzy Clustering Mechanism

C. *Line graph and link partitioning*

This technique will follow the partitioning of a link rather than nodes to analyze the cluster.[2] A node is said to be overlapped if the link connected to it is a part of many clusters. Links are generally clusters using the hierarchical clustering techniques. Single linkage hierarchical clustering is then represented by the dendogram. If these dendograms are cut at certain threshold value then overlapping communities are generated [3].The edges are analyzed in this case. The edges belonging to more than one region define community overlapping. Line graph is also extended to clique graph.

Link partitioning for overlapping community is also proposed. Link partitioning will help in identifying overlapping community [2]. Higher quality detection is

CONFERENCE PAPERS
National Conference on Emerging Trends on Engineering & Technology (ETET-2017)
On 21st April 2017
University Inst. of Engg. & Tech. & University Inst. of Computer, SBBS University, Punjab (India)

provided in this case. The node based detection is not used in this case. The concepts define within the line graph and link partitioning is ambiguous. Link based extended modularity is also purposed in this case. The modularity will decrease the complexity associated with the system. Link between the nodes i and j [11] .

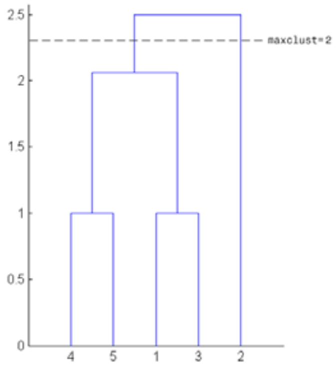


Figure 5 shows the dendogram describing overlapping community detection

III. COMPARISION OF VARIOUS TECHNIQUES

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

There are following comparisons associated with the clustering algorithm

Clique Percolation	**Fuzzy Detection**	**Line Graph and Link partitioning**
The clique of specified size is detected using this technique	The approximation about the clusters will be made	The clusters are specified in terms of the links or edges.
Clustering is detected by the use of the nodes	Clustering is detected in terms of the nodes	Clustering is detected in terms of the edges.
It is faster in nature.	It is slower as compared to clique percolation method.	It is slower as compared to both the methods specified.
The clique will be detected by discovering only those nodes which have same degree as the value of k	The clusters are plotted using k-means and c means techniques	The graph partitioning methods are used in order to detect the overlapping
The problems associated with the undirected graph can be solved	The problems associated with undirected and directed graphs can be solved	Problems associated with directed and undirected graph can be solved.

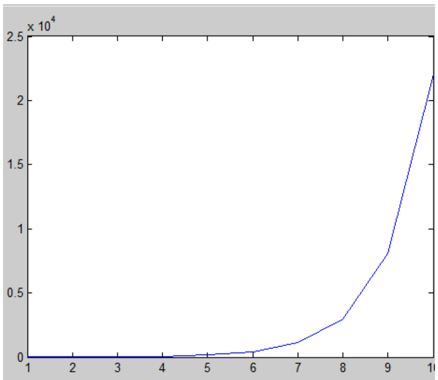
IV. RESULT COMPARISONS

For our demonstration we will consider the clique size to be 4. Number of nodes which are considered are 30. The Clique percolation method detects the cliques and result is listed in the form of cliques having size as specified. Fuzzy detection method when applied. Result is produced in terms of nodes. The cliques are listed in terms of clusters. In case of hierarchical clustering the cliques are detected in terms of dendograms. The tabular representation showing the numerical computations will be as follows:

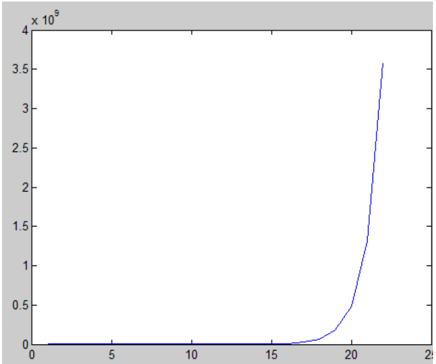
	Clique Percolation	Fuzzy Detection	Hierarchal Clustering
Number of Nodes	30	30	30
Clique Size	4	4	4
Nodes Compared	10	21	26
Cliques	6	6	4
Time Consumed	10ms	21ms	26ms

Table 1: Showing the difference in various Parameters of different algorithms

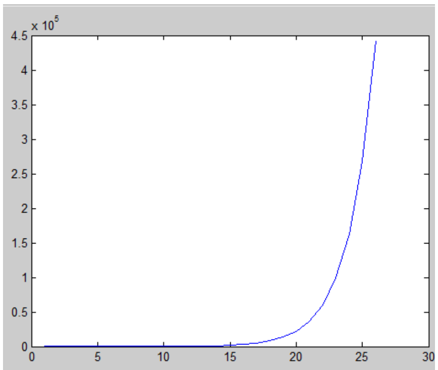
The time comparison of different algorithm will be listed as follows :



Clique Percolation Time 10ms



Fuzzy Detection Time 21ms



Line Graph and Link Partitioning Time 26ms

V. CONCLUSION

From the above comparison it is clear that the K-Clique algorithm will generate faster results as compared to previous algorithms. The K-Clique will go through the nodes even though they may not require hence consuming more time than normal. The fuzzy detection method is also present which will determine the community overlapping detection graphically. The line partitioning method will detect the overlapping community detection by the use of dendograms. The mechanisms which are available for detecting the overlapping are many. We have graphical user interface as well as command user interface for detecting the cliques. The users may be participants in the multiple communities which can be overlapping. Detection mechanisms are many. We highlighted the methods and also described which method is useful in detecting the cliques. The K-Clique method is one of the simplest methods for the detection of the overlapping community detection.

REFERENCES

[1] S.P.Chatzis, "A fuzzy c-means-type algorithms for clustering of data with mixed numeric and categorical attributes employing a probabilistic dissimilarity functional," Expert syst. Appl.., vol. 38, no.7, pp. 8684-8689,2011.

[2] S. Ghosh and S. Dubey ,"Comparative analysis of K-means and Fuzzy C-Means Algorithms," Ijacsa, vol.4, no.4, pp. 35-39,2013

[3] W.L. Hung , M.-S yang and D.H Chen, "Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation ," Pattern Recognit. Lett, vol .27,no 5, pp. 424-438, 2006.

[4] Z.ji, Y. Xia, Q. Sun, and G. Cao, "Interval-valued possibilistic fuzzy C-means clustering algorithm," Fuzzy sets syst, vol. 253, pp. 138-156, 2014.

[5] S.A Mingoti and J.O. Lima, "Comparing SOM neural network with Fuzzy C-means, K-means and traditional hierarchical clustering algorithms," Eur. J.Oper.Res.., Vol. 174, pp. 1742-1759,2006.

[6] Harpreer Arora, "A Novel Approach towards k-Mean Clustering Algorithm with PSO" in International Journal of Computer Science & Information Technologies Volume 5 issue 4 pp. 5978-5986. August 2014 .

[7] X. Wang, Y. Wang, and L. Wang, "Improving fuzzy c-means clustering based on feature-weight learning ," Pattern Recognitt. Lett, vol. 25, no. 10, pp. 1123-1132, 2004.

[8] K.L. Wu, "Analysis of parameter selections for fuzzy c-means," Pattern Recognitt.., vol. 45, no. 1, pp. 407-415, 2012.

[9] K. L. Wu and M.S. Yang, "Alternative c-means clustering algorithms," Pattern Recognit.., vol. 35, no. 10, pp. 2267-2278, 2002.

[10] W.Zhu, J. Jiang, C.song and L.Bao, "Clustering algorithm based on Fuzzy C-means and Artificial Fish Swarm," Procedia Eng.., vol. 29, no. 0, pp. 3307-3311, 2012.

[11] Jierui Xie, Stephen Kelley, Boleslaw k. Szymanski, "Overlapping Community Detection in Networks: The State-of-the-Art and Comparative Study".

[12] Harpreet Arora, "K-Mean Clustering and PSO: A Review" in International Journal of Engineering and Advanced Technology Volume/ Issue:3-5 Publication: June 30, 2014.

[13] W. Wang and Y. Zhang, "On fuzzy cluster validity indices," Fuzzy sets syst.., vol. 158, no.19, pp. 2095-2117,2007.