Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Engineering

Elixir Comp. Engg. 93 (2016) 39833-39836

ISSN: 2229-712X

A Comparison of Overlapping Community Detection Algorithms on Facebook Network

Mini Singh Ahuja, Jatinder Singh and Neha

[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.

ARTICLE INFO

Article history:
Received: 17 March 2016;
Received in revised form:
21 April 2016;
Accepted: 26 April 2016;

Keywords
Overlapping communities,
Clique percolation method,
Fuzzy detection,
Line graph and link partitioning.

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 we have done comparison of various algorithms in order to detect the overlapping communities in the Facebook network. The framework for evaluating various algorithms is described which helps in disclosing the person's membership in multiple clusters. The cluster is a collection of number of distinct users belonging to one or more groups. This paper evaluates the comparison between various algorithms which are used for overlapping community membership detection. Experimental results shows that CPM can give better results in less amount of time.

Introduction

Community overlapping or modular structure is one of the most widely studied topics in real world social media as it elaborates the functioning of the system. In order to detect overlapping communities from the Facebook network we have taken the dataset from Stanford University.

Network communities represent basic structure for understanding the organization of the real world environment [1]. A community is a group of nodes which are connected 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 may have links to many active area including people, movies, news, etc. All of the above stated active things are group to which a user may belong [2].

For this reason, there is growing interest in overlapping community detection algorithm that identifies a set of groups which are not disjoint. In this paper comparison of the various algorithms for overlapping community detection has done.

Algorithms

In order to analyze the overlapping community detection we have used the following algorithms:

Clique Percolation Method

The Clique Percolation Method (CPM) is a clique based algorithm introduced by Palla .et. al to detect overlapping communities [8]. Clique is a sub graph in which every two distinct vertices are adjacent to each other. This algorithm 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 [6].Once these have been identified, a new graph is constructed in which each vertex represents a k-clique. This graph is known as Clique Graph in which two nodes are connected only if they share k-1 members. Connected graph will be used to detect the overlapping community. The relatively small values of k lies between 3 to 6 appear to give accurate results [5].

CPM introduced a sub graph intensity threshold for weighted networks. Therefore a threshold value was maintained. Only k-cliques with intensity larger than this fixed threshold were included into the community.

Further kumpula et.al presented Sequence Clique Percolation method (SCP) which is faster than CPM as it finds cliques of fixed size instead of all values of k [9]. 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.

The result for k=4 will be describe as

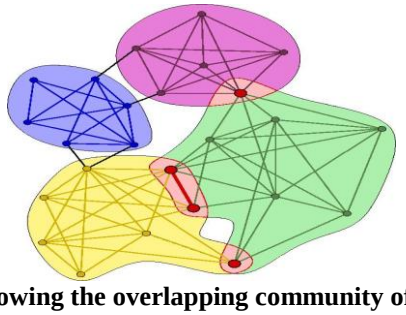


Fig. 1 showing the overlapping community of order 4

Tele:
E-mail address: nehachalotra10@gmail.com

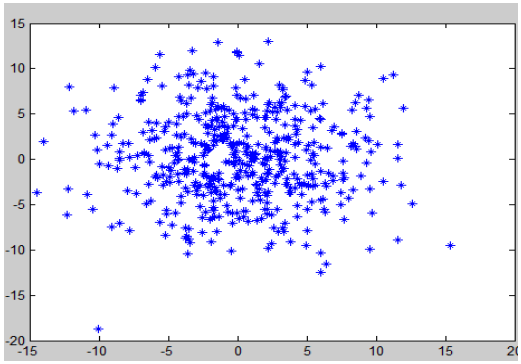


Fig. 2 showing k-clique mechanism

Fuzzy Detection

The fuzzy detection mechanism can be used to detect modular overlaps ie group of nodes shared by community. Fuzzy community model is used to quantify the strength of the associations that exists between nodes and edges. [3]For overlapping community detection generally c mean clustering is preferred. 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. In case of crisp overlapping every vertex belongs to one or more community with equal strength but in fuzzy overlapping each node belongs to one or more community but their strength of belongingness may vary. To compare these algorithms consistently, we propose using a common measure: the Fuzzy Rand Index.

(i) 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.

(ii)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.

(iii) 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

(iv) 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, but 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 [10].

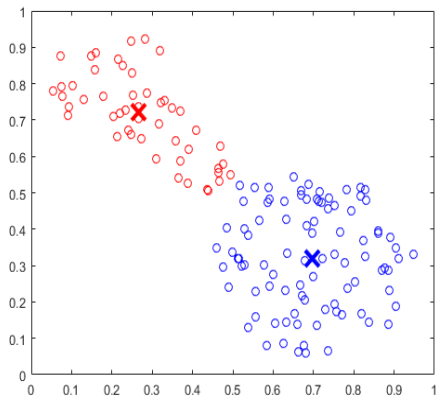


Fig. 3 Shows Fuzzy Clustering Mechanism

Line Graph and Link Partitioning

Evans and Lambiotte presented the Line graph algorithm which follows the partitioning of a link rather than nodes to analyze the cluster [12]. Links are partitioned via hierarchical clustering of edge similarity 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 [16]. In this algorithm the edges are analyzed. The edges belonging to more than one region define overlapping communities.

Chuan Shi proposed Link partitioning for overlapping community detection [13]. This algorithm provides higher quality results and it did not follow node based detection. This algorithm finds out the group of links that have similar characteristics. The concepts define within the line graph and link partitioning is ambiguous. Link based extended modularity is also purposed by author in this case. The modularity will decrease the complexity associated with the system [11].

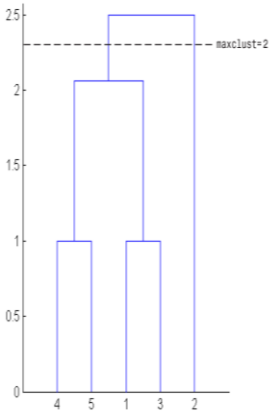


Fig. 4 Shows the dendogram describing overlapping community detection

Comparison of Various Techniques

There are following comparisons associated with the clustering algorithm.

Table 1. shows comparison among overlapping community detection algorithms.

Clique Percolation	Fuzzy Detection	Line Graph and Link partitioning
1. The clique of specified size is detected using this technique	1.The approximation about the clusters will be used	1. The clusters are specified in terms of the links or edges.
2. Clustering is detected by the use of the nodes	2. Clustering is detected in terms of the nodes	2. Clustering is detected in terms of the edges.
3. It is faster in nature.	3. It is slower as compared to clique percolation method.	3. It is slower as compared to both the methods specified.
4. The clique will be detected by discovering only those nodes which have same degree as the value of k	4. The clusters are detected using k-means and c-means techniques	4. The graph partitioning methods are used in order to detect the overlapping communities.
5. The problems associated with the undirected graph can be solved	5. The problems associated with undirected and directed graphs can be solved	5. Problems associated with directed and undirected graph can be solved.

Result Comparisons

For our demonstration we have consider the clique size to be 4 (which means we detect nodes that are connected to 4 communities). 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. When we applied Fuzzy detection method then result is produced in terms of nodes. In case of hierarchical clustering the cliques are detected in terms of dendograms.

The number of cliques detected will decide the complexity of the system within the network. There are many methods for partitioning a network into communities but sometimes we need to know which partition establish a real community structure. Therefore we need a quality function to find out how good a partition is. The most popular quality function is modularity of Newman and Girivan. [18]

In order to calculate the modularity we have used the following formula:

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(C_i, C_j)$$

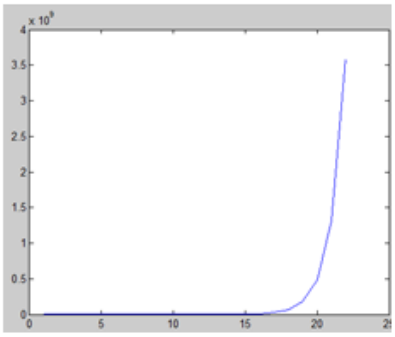
Where A is adjacency matrix, k_i is the degree of vertex of i , m is the total number of edges in the network. The element of A_{ij} of adjacency matrix is 1 if vertices i and j are connected.

The tabular representation showing the numerical computations will be as follows:

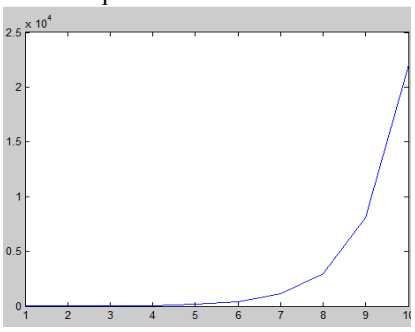
Table 2. Showing the difference in various Parameters of different algorithm

	Clique Peroration	Fuzzy Detection	Hierarchal Clustering
Number of Nodes	30	30	30
Clique Size	4	4	4
Nodes Compared	10	21	26
Cliques found	6	6	4
Time Consumed	10ms	21ms	26ms
Modularity	0.309	0.235	0.299

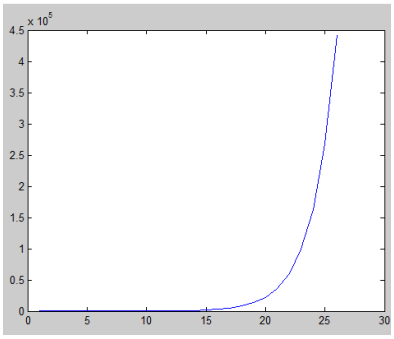
The time comparison of different algorithm will be listed as follows:



1. Clique Peroration Time 10ms



2. Fuzzy Detection Time 21ms



3. Line Graph and Link Partitioning time=26ms

Discussion

From the above comparison it is clear that the K-Clique algorithm generates faster results as compared to other algorithms. The K-Clique even goes through the nodes that are not required even then it takes less time as compared to other algorithms. It detects 6 communities of 4 clique size in only 10ms while the fuzzy detection algorithm takes 21 ms to detect such groups. Hierarchical algorithm detects only 4 communities in 26 ms. Modularity is also higher in clique algorithm.

Conclusion

In this paper we have implemented three basic community detection algorithms to detect overlapping communities in the Facebook network. The K-Clique method is one of the simplest methods for the detection of the overlapping community detection. Here fuzzy detection method is also presented which determine the community overlapping detection graphically. The line partitioning method represents the overlapping communities by the use of dendograms. In this paper we have highlighted the methods and also described which method is useful in detecting the cliques.

References

[1] M.E.J Newman, The structure and function of complex networks. Volume 2, pp, 167-256, 2003

[2] L. Tang, H. Liu. Community Detection and Mining in Social Media, Morgan & Claypool (2010).

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Kumpula, j. m., Kivela, m., Kaski, k., and Saramaki, j. "Sequential algorithm for fast clique percolation." Phys. Rev. E 78, 2008.

[10] Steve Gregory, "Fuzzy overlapping communities in networks".
[11] Fortunato, S. 2010, "Community detection in graphs." Phys. Rep. 486, 75–174.
[12] Evans and Lambiotte 2009] Evans, T. S. and Lambiotte, R. "Line graphs, link partitions, and overlapping communities", Physical Review E, 80 (2009).
[13] Chuan Shi □ , Yanan Cai, Di Fu, Yuxiao Dong, Bin Wu "A link clustering based overlapping community detection algorithm" 2013.
[14] K. L. Wu and M.S. Yang, "Alternative c-means clustering algorithms," Pattern Recognit.., vol. 35, no. 10, pp. 2267-2278, 2002.
[15] 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.
[16]. Jierui Xie, Stephen Kelley, Boleslaw k. Szymanski, "Overlapping Community Detection in Networks: The State-of-the-Art and Comparative Study".
[17] Ahn, Y.-Y., Bagrow, J. P., And Lehmann, S. 2010, "Link communities reveal multiscale complexity in networks".*Nature 466*, 761–764.
[18] Newman, M. E. J., Modularity and community structure in networks, Proceedings of the National Academy of Sciences (PNAS) , 103(23), 8577–8582, 2006.