Exploring Agglomerative Spectral Clustering Technique Employed for Social Network Community Detection, A Survey

Shafia Bashir[1] and Manzoor Ahmad Chachoo[2]

[1]Department of Computer Science, University of Kashmir Srinagar, J&K, India.
[2]Department of Computer Science, University of Kashmir Srinagar, J&K, India.

Abstract - *Data mining from social networks by identifying actionable patterns has been an emerging and leading research. These patterns have been known as communities wherein a pattern is a densely interconnected group of the whole network. Various approaches and algorithms have been devised to identify such groups. This study introduced an agglomerative spectral clustering with edge weights and conductivity one of the famous clustering approach in recent years. This methodology agglomerates related nodes based on edge weights and eigenvector space. Moreover, to distinguish the densely interconnected groups during agglomeration. Conductance is utilized. This methodology proved to be efficient and showed better performance as compared to the state of the art approaches.*

Key words: Community structure, Probabilistic graphs, Spectral clustering

1. INTRODUCTION

In the social network domain community identification is an emerging and challenging issue. Community detection in social networks depends on analyzing the network structure and identifying the people who interact much with one another than with others. Humans have natural tendency to make groups, within family, in locality and work places and exchange their views and ideas in text form. However the technology has improved to such an extent that nowadays ideas can be exchanged through multimedia.

Social network analysts employ two kinds of mathematics tools to represent information among social actors on patterns of ties: graphs and matrices. Problems can be solved in almost any possible discipline using graphic designs. A graph G = (V, E) where V consists of vertices and E a set of unordered pairs of vertices called edges that that if an edge exists between the vertices V_i and V_j then the vertices V_i and V_j are related to each other and the edge between them is considered as $E_{ij.}$

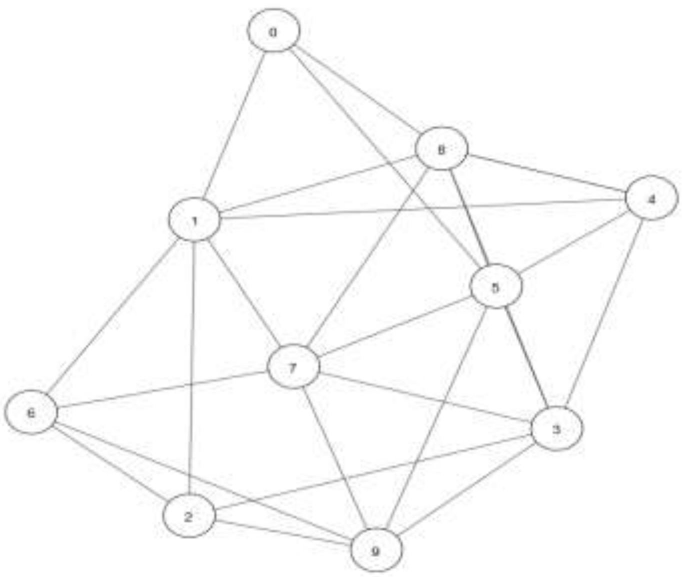

Fig-1: Showing a simple graph with 10 vertices connected by edges.

When modeling social networks as graphs, an entity is modeled as a node, and an edge is modeled on the relationship that links two entities. We can utilize graph models to show different people-to people relations.Everyperson is defined by a vertex in a given group of people.When two people know one another, an ed ge is used to bind these two people. However Matrix form is the most widely recognized form of data representation in social network analysis, because it is most suitable and convenient for many mathematical o perations. Matrices can be converted into graphs and vice versa. Using adjacency matrix and adjacency list are the most familiar ways to store a graph.

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

Fig -2: Showing adjacency matrix

An adjacency list shall be represented as an indexed list array. Every node has a list of unordered nodes that are neighbors of that node. Every edge is shown twice, one for source node and one for destination.

0	1 → 5 → 8
1	0 → 2 → 4 → 6 → 7 → 8
2	1 → 3 → 6 → 9
3	2 → 4 → 5 → 7 → 9
4	1 → 3 → 5 → 8
5	0 → 3 → 4 → 7 → 8 → 9
6	1 → 2 → 7 → 9
7	1 → 3 → 5 → 6 → 8 → 9
8	0 → 1 → 4 → 5 → 7
9	2 → 3 → 5 → 6 → 7

Fig -3: Showing Adjacency list

Social communities primarily consist of individuals from all fields of life. The community members are bonded together because of common interests like profession, hobby or location. Identifying these communities is a challenging task chiefly due to vast definitions of network community, intractability of methods, evaluation issues and the lacking of reliable and ground-truth standard.

This study focussed on the community detection in social networks by employing agglomerative spectral clustering method which is simple and transparent in forming separable clusters from non-separable ones. Moreover the outline of the study is organized as: Section 2 addresses the findings and recent research in social networks related to detecting communities. Section 3 covers the agglomerative spectral clustering methodology. Lastly section 4 presents the discussion and conclusion of the study.

2. RELATED WORK

The identification of social network community structures is a challenging task. Various types of methods have been implemented for revealing such community structures. The most popular approach to detect communities is clustering as it is applicable on different graph types like multilayer [2] and complex probabilistic graphs [3]. For large complex networks hierarchical clustering is utilized [4]. It automates the subgroup identification process by clustering those nodes which are similar. Then successive fusion of clusters takes place as long as all the nodes get fused into single leftover cluster. The goal is to identify cluster of vertices having high similarity. The result is a new similarity matrix [5]. Hierarchical clustering comprises of agglomerative stage [6, 7] and divisive stage [8]. The agglomerative stage follows a bottom-up approach where the clusters with high similarity are agglomerated iteratively into a single cluster. Various algorithms that utilized this approach are community detection based on neighbourhood overlap [9], Girvan-Newman [10] etc. However the divisive stage follows a top-down approach where a network initially is deemed to be single cluster and after few iterations the cluster gets divided into may clusters by removing the edges that connect most dissimilar vertices. Various algorithms that implemented this approach are label propagation algorithm [11], Louvain algorithm [12].

Identifying the closest people in the network is typically the key goal in social network research, which is generally done by the use of a visualization technique within a social network. Thus clustering can lead to the emergence of a potential technique for discovering more groups and clusters in complex social networks. In fact, it may provide more comprehensive information than simulation, including a group's closeness, extensive membership details in a group, and the social network relationship between groups.

Though hierarchical clustering does not need the prior knowledge about the size and number of clusters yet it requires a well-explained similarity function hence decreases the community detection accuracy. Moreover the clustering may be unsatisfactory in case all the vertices are similar to one another. Therefore, community detection needs to be checked with some advanced formulation.

3. AGGLOMERATIVE SPECTRAL CLUSTERING

This section presents the agglomerative spectral clustering with conductance method in detail. The eigenvector space is used to define the similarity of the nodes and agglomerate the more related nodes to create a new agglomerated node (which becomes a clustered community after some iteration)This new clustered community is then added with the graph and the modified graph is iterated unless and until the termination condition is satisfied. The termination condition reaches once the clustered community is lightly connected to outside and is more densely connected to inside.

The agglomerative spectral clustering is a 3 stage process wherein the first stage original points are projected into the eigenvector feature space. The eigenvector feature space together with the no. of connecting edges between nodes is used for the evaluation of similarity within the nodes. In the final stage conductivity in between the node and its candidate is identified. The nodes are agglomerated to form clustered communities if and only if the conductivity improves. The conductivity is employed as the termination criteria and the edge weight is used to evaluate the most accurate similarity. The entire 3 stage procedure is iterated until no further agglomeration of nodes is required.

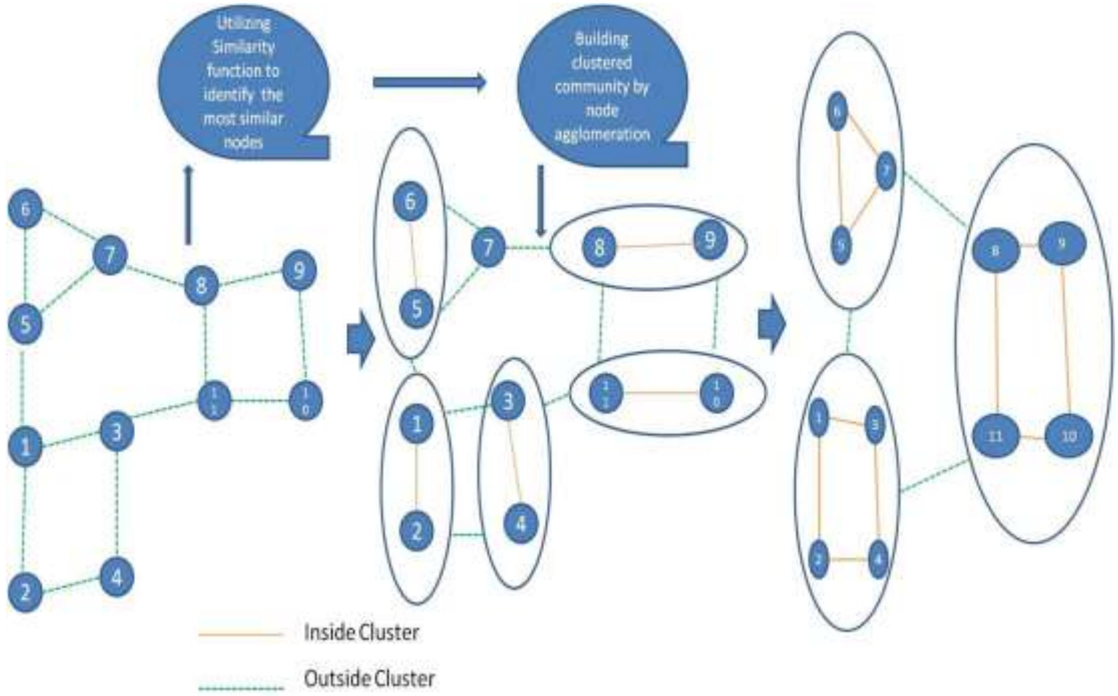

Fig - 4: A framework showing node agglomeration

The method improves both the accuracy and time complexity. Unlike most of the spectral clustering approaches which were applicable to synthetic networks only, agglomerative spectral clustering is well suited for real life social networks. Furthermore this method finds out vast application into the growing social network domain.

4. DISCUSSION and CONCLUSION

From the study of traditional clustering methods it can be inferred that eigenvector space was not efficient to convey the similarity within the agglomerated nodes as the interconnected nodes were projected to similar locations in a feature space. Thus was difficult to differentiate similar projections. Further, there was no termination criterion for satisfied clusters. Thus agglomerative spectral clustering aimed to resolve both these issues wherein similar projections are differentiated by unequal edge connections among the nodes. Since the agglomerated nodes may have more than one edge connecting with one another (fig -4). Therefore, the more the nodes are tightly connected the more is the similarity. Similarity score is obtained as a mean of the edges within two nodes. Based on this similarity nodes are agglomerated iteratively which results in formation of cluster after some iteration and are termed as clustered community. When these agglomerated nodes (clustered community) are connected more densely inside compared to outside termination condition reaches. Thus a cluster is fairly treated as a good community.

Though the agglomerative spectral clustering has better time complexity compared to other related methods yet the calculation of eigenvector space at every iteration slows down the computation time. The results discussed here present a comprehensive method for community detection however; further research is needed in this domain to achieve a mature stage.

REFERENCES

[1] U. U. Narantsatsralt, and S. Kang. (2017), "Social network community detection using agglomerative spectral clustering," http://doi.org/10.1155/2017/3719428.
[2] X. Dong, P. Frossard, and N. Nefedov, "Clustering on multi-layer graphs via Subspace analysis on grassmann manifolds," IEEE Transactions on Signal Processing, vol. 62, no. 4, 2014.
[3] G. Kollios, M. Potamias, and E. Terzi, "Clustering large probabilistic graphs," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 2, 2013.
[4] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," Nature, vol. 453, no.7191, pp. 98-101, 2008.

[5] S. Fortunato, “Community detection in graphs," in Complex Networks and Systems Lagrange Laboratory, ISI Foundation, 2010.
[6] M. C. Ramos, “Divisive and hierarchical clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean region,” Atmospheric Research, vol. 57, no. 2, pp. 123-138, 2001.
[7] L. Lin, T. Luo, J. Fu, Z. Ji, and D. Xiao, “A new community detection based on agglomeration mechanism'” in Proceedings of the IEEE 2nd International Conference on Computing, Control and Industrial Engineering, PP. 352-355, August 2011.
[8] M. Leng, J. Wang, P. Wang, and X. Chen, “Hierarchical agglomeration community detection algorithm via community similarity measures,” TELKOMNIKA Indonesian Journal of Electrical Engineering, vol. 10, no. 6, pp. 1510-1518, 2012.
[9] N. Meghanathan, “A Greedy algorithm for neighborhood overlap-based community detection,” Algorithms, vol. 9, p. 8, 2016.
[10] M. Girvan, and M. E. Newman, “Community structure in social and biological networks,” in Proceedings of the National Academy of Sciences of the United States of America, vol. 99, no. 12, pp. 7821–7826, 2002.
[11] 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, 2007.
[12] V.D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” Journal of Statistical Mechanics: Theory and Experiment, 2008.