International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 6 (2020), pp. 1386-1390
© International Research Publication House. https://dx.doi.org/10.37624/IJERT/13.6.2020.1386-1390

Community Detection Methods and Tools for Various Complex Network

Dipesh Joshi[1], **Dr. Tejas Patalia**[2]

[1]*Research Scholar, Gujarat Technological University, Ahmedabad, India.*

[2]*Professor, Computer Engineering Department, V.V.P. Engineering College, Rajkot, India.*

[1]*ORCID: 0000-0002-8420-8591* [2]*ORCID: 0000-0001-6008-8380*

Abstract

The analysis of complex networks like social, biological network is a new era of research in information mining. The analysis of social networks has gained considerable attention in the current era, mainly due to the high growth of social networks and content exchange sites. A social network can be seen as a complex interconnection of social entities in terms of vertex and nodes. The detection of the community is the task of grouping social entities based on the linking of nodes and relationships. Most of the research has been done on the basis of grouping algorithms and extraction techniques. There are many problems associated with the task of researching social group and network analysis, such as the grouping of nodes, community mining, the generation of graphics, the prediction of links. The detection of the community in the mining of social networks is different from the traditional grouping methods. Community detection is a way to recognize the network nodes in a group or community within which the property of identified vertices is maximized in terms of similarity. Communities can have concrete applications.

Keywords: Data mining, clustering algorithm, link prediction, community detection, social network

I. INTRODUCTION

The analysis of social networks and mining become evident as an important research topic in the social sciences. Most of the previous works were performed on data collected from individual profiles in social environments, in order to investigate specific social entities. The research was generally taken as a study in small communities or groups, collecting data through a set of printed or written questions with a choice of answers, interviews and different methods.[6][14] Comprehensive study on SNA focus on structural component of social network, methods for social network mining, issue regarding social network mining and tools used for social network mining. Clustering nodes that have similar group interest and are location bases close to each other can enhance the performance of the methods provided on the internet, since each group of nodes can be managed by a dedicated resource.

The identification of groups of customers that have similar interests in the purchasing connections between customers and products of online retail shop allows for the establishment of influenced recommendation systems, which direct customers through the retailer's list of different items and improve opportunities of business. The detection of the community is important to identify network and their limits allow a classification of nodes, according to their structural behaviour in the network. The objective of the community detection in the graphs is to recognize the components and, perhaps, their hierarchical representation, using only the data present in the different level structure of the graph.

II. ELEMENTS OF COMMUNITY DETECTION

We can define elements of community detection in terms of graphs as it will represent structure in graph with interconnection of densely connected nodes.

A. Complexity:

Complexity is represented as the estimation of the amount of resources needed by the method to execute a single task. This implies both the number of calculation steps required and the number of memory units that must be assigned simultaneously to execute the calculation. In most cases, we obtain greater complexity due to the large number of nodes that must be processed, so we cannot use any traditional algorithm to find the structure, but we need to use the approximation algorithm to find the structure and reduce the complexity of the network.

B. Communities:

The community detection is a method to classify the nodes of the network in a group or community within which the attribute of the nodes of similarity is maximized. The community can be formed by people who share hobbies, who work together, who live together or who have a similar interest in the subject. [2][4][8][9] There is no clear definition for communities as it will always depend application by application. Most of the cases, community can be algorithmically defined as the end product.

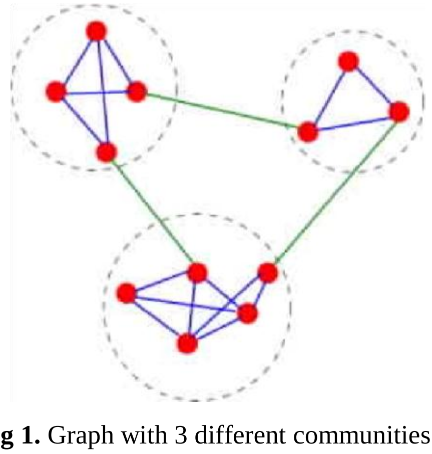

Fig 1. Graph with 3 different communities[12]

C. Nodes & Edges:

For graph G(V,E), we can define node as user for social network and edges as relationship between these users. When these user are densely connected with each other that means there are strong relation between them and sharing same interest between them. For different application, Nodes and edges can be defined differently and share different similarity. We can use quality parameter to identify strong relation between them.

D. Vertex Similarity

Communities are groups of nodes based on vertex similarity. We can use different graph theory concept to find similarity. Euclidean distance, adjacency matrix, cosine similarity, max flow, in cut theorem are used to find vertex similarity.

III. METHODS OF COMMUNITY DETECTION

There are many techniques and method available but almost use graph partitioning approach to analyse it.

A. Hierarchical Algorithms:

It is always different aspect to find where to place node in different clusters. It is also uncommon to know that graph partitioned in how many communities. Hierarchical algorithm is commonly used algorithm for community formation as it will reveal multilevel hierarchy of nodes. It gives hierarchical distribution of the nodes of the social network. Such hierarchical methods have traditionally been used in sociology. One property of hierarchical methods is that they give not only a flat partition of the network into different communities, but a hierarchy of communities and subcommunities structure.[12] It has two different categories of hierarchical clustering. One that represent agglomerative algorithms and other one that represent divisive algorithm. It is single linkage clustering and complete linkage clustering.

B. Modularity Maximization:

Higher values of modularity indicate best partitioning of graphs in community. There is different greedy technique available to maximize modularity. Basic divisive algorithm that is Girvan Newman to maximize modularity. Girvan and Newman proposed a measure to evaluate the quality of a partition of a network in communities and select the best community partition of a hierarchical decomposition in 2002. The parameter of measurement is called modularity and it is defined as the proposition of edges that fall within the communities minus the same proposition if the edges were randomized. [12]

C. Graph-Partitioning:

The fundamental problem that is tried to solve is the one to divide a great irregular graph in k parts. The partition is usually carried out in a way that satisfies certain restrictions and optimizes certain objectives. The most common restriction is to produce partitions of equal size, while the most common goal is to minimize the number of cut edges. [12] The problem of graph partitioning in dividing groups of predefined size, such that node that have most communication or similarity stay in same group. For partitioning any graph, first point that must analyze is where to provide cut. Edge betweenness play wise role for partitioning any graphs. There is different method available like minimum bisection, spectral bisection, level structure partitioning, geometric algorithm, max flow min cut algorithm.

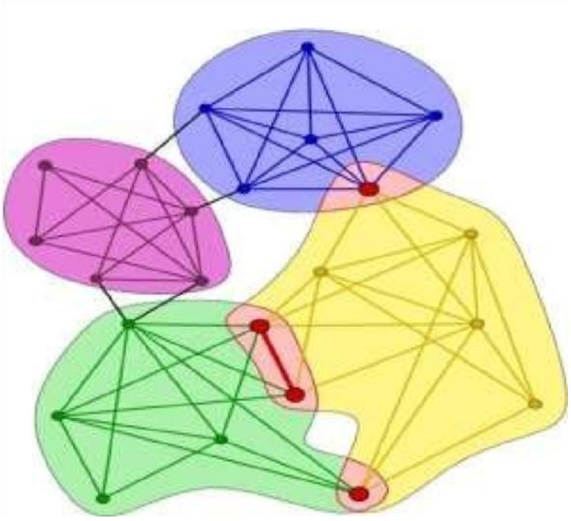

Fig 2. Graph with communities [12]

D. Spectral Clustering

The spectral grouping or clustering includes all the methods and techniques that divide the set into groups through the use of matrix vectors or eigen vector, such as S itself or other matrices derived from it. The spectral grouping consists of a transformation of the initial set of objects into a set of points in space, whose coordinates are elements of eigenvectors: the set of points is grouped through normal techniques, such as the kmeans partitioning. Some of the methods use eigenvectorbased method with adjacency matrix or Laplacian matrix.

III.I TOOLS FOR NETWORK ANALYSIS

There are many tools used to implement social network mining. Some tools given below that are used to analyse it:

A. Gephi :

Gephi is an interactive visualization platform for all types of networks and complex graph structure, dynamism of graph and hierarchical graph topology. It is a tool for people who have to explore and understand graphics. The user interacts with the interface; Manipulate structures, shape and color to identify hidden properties. It uses a 3D rendering engine to show large networks in real time and to speed up exploration. A flexible, multi-tasking architecture offers new ways for working with complex data sets and producing valuable visual results.[18][21][19]

B. Graphviz:

Graphviz is an open source graph visualization platform. It has different graph design programs suitable for viewing social networks in interactive mode. [19][22]

C. SONDY:

SONDY is a tool for analyzing trends and dynamics in online social network data. SONDY helps end users, such as media analysts or journalists, to understand the interests and activity of social network users by providing emerging topics and event detection, as well as network analysis functionalities. [16][20]

D. NEO4J:

Neo4j is a graph database. It is an integrated, disk-based, fully transactional persistence engine that stores structured data in graph instead of tables. [19][23]

E. SocNetV [25]

Social Network Visualizer (SocNetV) is a multiplatform, user interactive interface friendly open source application for social network analysis and visualization of structure. We can design social network with some clicks and it use different file format like Pajek, GraphMP, UCINET etc..

F. Cytoscape[26]

Cytoscape is open source platform for visualizing large and complex network in few clicks. It is integrating complex networks with any data attributes. It will be use to analyse social network, bioinformatics and semantic web.

G. NodeXL[27]

NodeXL provides powerful feature to analyse social network including influencing node, brand evaluation based on product performance, content analysis, campaign analysis. It provides automation for different kind of analysis on social network. NodeXL makes it easy to explore, analyse and visualize network graphs

H. NetMiner[28]

NetMiner is not open source tools but it is premium tools to analyses social network features based on input data. It allows user to explore network data visually and interactively.

Many other tools are available for graph mining as well as community detection and topic detection.

III.II OUR METHODOLOGIES

1. Initialize Visited Vertex, Edge List and Partition List to empty
2. Read the Dataset text file containing edge list with Origin vertex and Destination vertex 3. for each Edge from Dataset File
4. If both vertices are new
5. Then Case - 1 is executed
6. If Any One Vertex is already visited
7. Then Case - 2 is executed
8. If both vertices are already visited
9. Then Case - 3 is executed
10. Result Writer Write vertex and corresponding community index in the file
11. Modularity of community is calculated
12. Modularity of Partition is calculated
13. Result Writer Write modularity of partition in the file
14. Exit

In above scenario, If both vertices are new then they will create single new community. If Anyone vertex is already visited then new vertex is added to old vertex's community. If both vertex are already visited then maximum modularity are calculated to move vertex with highest modularity community.

IV. RESULT

Table 1. Dataset List

Sr. no	Dataset Name	No. Of Nodes	No. Of Edges
1	CA HepPh[33]	12008	237010
2	CA Cond Mat[33]	23133	186936
3	Web Stanford[33]	281903	2312497
4	Web Google[33]	875713	5105039

Table 2. Average result after 10 times execution of algorithm

Sr. No.	Dataset Name	No. Of Community	Modularity	Time(In Second)
1	CA HepPh[33]	1473	0.6020	0.23
2	CA Cond Mat[33]	3089	0.6513	0.24
3	Web Stanford[33]	41575	0.8683	7
4	Web Google[33]	130865	0.8375	20

V. CONCLUSION

This paper presents analytical study of community detection in social network. here we give basic understanding of community detection method and tools. We also represent survey of community detection in social network. In order to collect data and process data which step to be conducted and which basic algorithm are available to implement. There are many tool, software and framework available for social network analysis. As per the analysis, Hierarchical clustering algorithm are best suited for large and complex social network. By using different tools, we can implement our algorithms well as dataset to analyze current trend. Our result set are better in terms of modularity. We will conduct overlapping node analysis in next step.

VI. ACKNOWLEDGEMENT

We thank our institute(V.V.P. Engineering College) for their support in research work progress. We greatly appreciate the reviewers and editor for their valuable suggestions and comments to improve this work.

REFERENCES

[1] Van der Aalst, Wil MP, and Minseok Song. ‘Mining social networks: Uncovering interaction patterns in business processes.’ International conference on business process management. Springer, Berlin, Heidelberg, 2004.

[2] Nicosia, Vincenzo. ‘Modularity for community detection: history, perspectives and open issues.’ Workshop at the University of Catania. Catania. Recuperado em. Vol. 12. 2008.

[3] Lin, Wangqun, et al. ‘Community detection in incomplete information networks.’ Proceedings of the 21st international conference on World Wide Web. ACM, 2012.

[4] Sadi, Sercan, Sima Etaner-Uyar, and Sule GündüzÖğüdücü. ‘Community detection using ant colony optimization techniques.’ 15th International Conference on Soft Computing. 2009.

[5] Creamer, Germán, and Sal Stolfo. ‘A link mining algorithm for earnings forecast and trading.’ Data mining and knowledge discovery 18.3 (2009): 419-445.

[6] Yiannis Kompatsiaris , ‘Social Networks Mining for

Innovative Applications and Users Well-Being’, FP7 ICT Work Programme 2013 Consultation Networked Media

[7] Guy, Ido, et al. ‘Mining expertise and interests from social media.’ Proceedings of the 22nd international conference on World Wide Web. ACM, 2013.

[8] Griechisch, Erika, and András Pluhár. ‘Community detection by using the extended modularity.’ Acta cybernetica 20.1 (2011): 69-85.

[9] ‘Kai-Yang Chiang,Department of Computer

Science,University of Texas at Austin, Lecture notes on community detection’

[10] Leskovec, Jure, Kevin J. Lang, and Michael Mahoney. ‘Empirical comparison of algorithms for network community detection.’ Proceedings of the 19th international conference on World wide web. ACM, 2010.

[11] Bonchi, Francesco, et al. ‘Social network analysis and mining for business applications.’ ACM Transactions on Intelligent Systems and Technology (TIST)2.3 (2011): 22.

[12] Fortunato, Santo. ‘Community detection in graphs.’ Physics Reports 486.3 (2010): 75-174.

[13] Blondel, Vincent D., et al. ‘Fast unfolding of communities in large networks.’ Journal of Statistical Mechanics: Theory and Experiment 2008.10 (2008): P10008.

[14] Ellison, Nicole B. ‘Social network sites: Definition, history, and scholarship.’ Journal of ComputerMediated Communication 13.1 (2007): 210-230.

[15] Lin, C. Y., Wu, L., Wen, Z., Tong, H., Griffiths-Fisher, V., Shi, L., & Lubensky, D. (2012). ‘Social network analysis in enterprise.’ Proceedings of the IEEE,100(9), 2759-2776.

[16] Guille, Adrien, et al. ‘Sondy: An open source platform for social dynamics mining and analysis.’ Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data. 2013.

[17] Fire, Michael, Rami Puzis, and Yuval Elovici. ‘Organization Mining Using Online Social

Networks.’ arXiv preprint arXiv:1303.3741 (2013).

[18] Bastian M., Heymann S., Jacomy M. (2009). ‘Gephi: an open source software for exploring and manipulating networks.’ International AAAI Conference on Weblogs and Social Media.

[19] http://en.wikipedia.org/wiki/Social_network_analysis_s oftware

[20] http://mediamining.univlyon2.fr/people/guille/software.php

[21] https://gephi.org/

[22] www.graphviz.org/

[23] www.neo4j.org/

[24] ‘Albert Ching-man Au Yeung and Tomoharu Iwata, Research on Social Network Mining and Its Future

Development,NTT Technical Review’

[25] https://socnetv.org

[26] https://cytoscape.org

[27] https://nodexl.com

[28] Rezvanian, A., Moradabadi, B., Ghavipour, M., Khomami, M. M. D., & Meybodi, M. R. (2019). Social Community Detection. In Learning Automata Approach for Social Networks (pp. 151-168). Springer, Cham.

[29] Rashmi, C., and Mallikarjun M. Kodabagi. "A review on overlapping community detection methodologies." 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). IEEE, 2017.

[30] http://konect.uni-koblenz.de/networks/ucidata-zachary - Zachary karate club

[31] https://insightdatascience.com – Grapg Based Approach

[32] https://sites.google.com/site/ucinetsoftware/datasets - UCI Dataset

[33] https://snap.stanford.edu/index.html - Stanford Network Analysis Project