International Conference On Emanations in Modern Technology and Engineering (ICEMTE-2017) ISSN: 2321-8169
Volume: 5 Issue: 3 86 - 90

A Survey paper on Techniques used for Community Detection in Social Networks

Apeksha P. Naik[1], SachinBojewar[2]
[1]*PG Student, Department of Computer Engineering*
[2]*Associate Professor, Dept of Information Technology*
Vidyalankar Institute of Technology, Mumbai.

Abstract:- Community detection is the most popular and growing area of interest in the field of Social and Real-Time Network applications.Several community detection algorithms have been introduced in recent years. This paper presents community detection techniques, which have already been proposed, and also discussing the type of social networks on which those proposed techniques are applicable.It also talks about some of the traditional algorithms for overlapping or disjoint community detection on large-scale real-world networks to identify the best community in real- time networks. This paper can play a significant role in the analysis and evaluation of community detection approaches in different application domains.

Keywords: *Community detection, Social Networks, Real-Time Networks, Community Structure*

*****

1. INTRODUCTION

In the world of internet,social networking has become an extremely important application in the past few years, because of its capability to enable social contact over the internet for geographically dispersed users. The web users interact with each other, participate in online discussions, and exchange different views forming social networks. A social network can be represented as a graph, in which nodes represent users (e.g. peoples, organizations, etc.) and links or edges represent the connections between the users. Detecting clusters or communities in large real-world graphs such as large social networks is a problem of considerable interest. Figure 1 shows a simple graph with three communities, enclosed by the dashed circles.

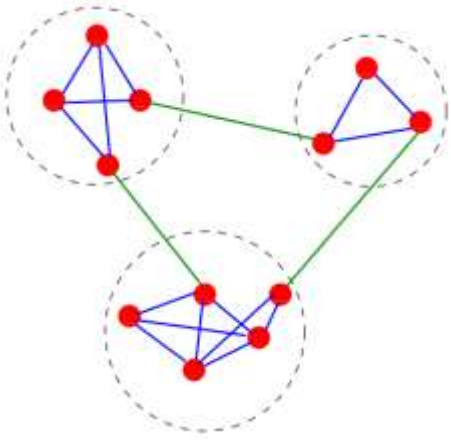

Figure 1: A simple graph with three communities

In social or real-time networks, finding a community means finding a group of users who interact on different entities like photos, comments, tags, stories or any other posts. Community Detection is of great importance in Biology, Sociology and Computer Science disciplines where the system is often represented as graphs.A network is said to have community structure if the nodes of the network can be easily grouped into sets of nodes such that each set of nodes is densely connected internally.Detecting communities in such a complex network is a difficult task.This survey reviews about the different community detection algorithms and methods for finding the best community in a network.A brief study of the community detection algorithms and approaches are presented in section 2. Results and discussions are presented in section 3. The concluding remarks are given in section 4.

2. DETECTION OF COMMUNITIES

In social networks, community detection is done by looking out for the nodes which are similar to each other and keeping those nodes in same community. When the nodes of a network, belonging to the same community, can be arranged to form a group, then that network is said to have a community structure. Community Structure is quite common in real-time networks. The community detection problem has many widespread applications and hence proven to be very important.The main advantage of community detection is accessing the information from diverse sources and clusters. A community structure consists of members with similar interests. Detection of communities makes exchanging or offering information easier because members of same community often have similar tastes.

2.1 Proposed Approaches for Community detection

Louvain Modularity: The Louvain Method for community detection is a method to extract communities from large networks.The method is a greedy optimization method that

attempts to optimize the "modularity" of a partition of the network. The optimization is performed in two steps.First, it looks for "small" communities by optimizing modularity in a local way. Second, it aggregates nodes of the same community and builds a new network whose nodes are the communities. Although the exact computational complexity of the method is not known, the method seems to run in time O(nlogn) with most of the computational effort spent on the optimization at the first level.

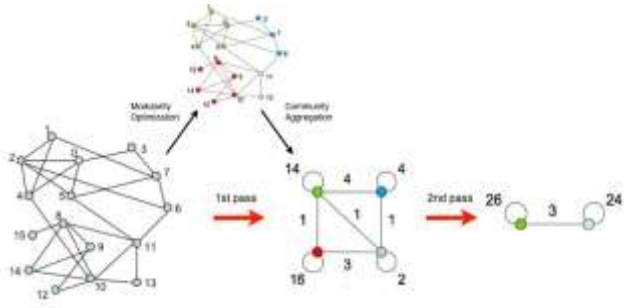

Figure 2: Louvain method for community detection

The method was first published in: Fast unfolding of communities in large networks, Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000.This is obviously an approximate method and nothing ensures that the global maximum of modularity is attained, but several tests have confirmed that the algorithm has an excellent accuracy[10].

Girvan and Newman Algorithm: Another commonly used algorithm for finding communities is the Girvan-Newman algorithm. The method is important because it marked the beginning of a new era in community detection field. This algorithm identifies edges in a network thatlie between communities and then removes them, leaving behind just the communities themselves. The identification is performed by employing the graph-theoretic measure betweenness centrality, which assigns a number to each edge which is large if the edge lies "between" many pairs of nodes.

The steps of the algorithm are:

1. Computation of the centrality for all edges;
2. Removal of edge with largest centrality: in caseof ties with other edges, one of them is picked atrandom;
3. Recalculation of centralities on the running graph;
4. Iteration of the cycle from step 2.

The Girvan–Newman algorithm returns results of reasonable quality and is popular because it has been implemented in a number of standard software packages. But it also runs slowly, taking time $O(m^2n)$ on a network of n vertices and m edges, making it impractical for networks of more than a few thousand nodes.[12]

Infomap Algorithm: The core of the algorithm follows closely the Louvain method, neighboring nodes are joined into modules, which subsequently are joined into super modules and so on. First, each node is assigned to its own module. Then, in random sequential order, each node is moved to the neighboring module that results in the largest decrease of the map equation. If no move results in a decrease of the map equation, the node stays in its original module. This procedure is repeated, each time in a new random sequential order, until no move generates a decrease of the map equation. Now the network is rebuilt, with the modules of the last level forming the nodes at this level, and, exactly as at the previous level, the nodes are joined into modules. [7]

This hierarchical rebuilding of the network is repeated until the map equation cannot be reduced further. With this algorithm, a fairly good clustering of the network can be found in a very short time. Let us call this the core algorithm and see how it can be improved. The nodes assigned to the same module are forced to move jointly when the network is rebuilt. As a result, what was an optimal move early in the algorithm might have the opposite effect later in the algorithm. Because two or more modules that merge together and form one single module when the network is rebuilt can never be separated again in this algorithm, the accuracy can be improved by breaking the modules of the final state of the core algorithm in either of the two following ways:

Submodule movements. First, each cluster is treated as a network on its own and the main algorithm is applied to this network. This procedure generates one or more submodules for each module. Then all submodules are moved back to their respective modules of the previous step. At this stage, with the same partition as in the previous step but with each submodule being freely movable between the modules, the main algorithm is re-applied on the submodules.

i. *Single-node movements.* First, each node is re-assigned to be the solemember of its own module, in order toallow for single-node movements. Then all nodes are moved back to their respective modules of the previous step. At this stage, with the same partition as in the previous step but with each single node being freely movable between the modules, the main algorithm is re-applied on the single nodes.

In practice, we repeat the two extensions to the core algorithm in sequence and as long as the clustering is improved. Moreover, we apply the submodule movements recursively. That is, to find the submodules to be moved, the algorithm first splits the submodules into subsubmodules, subsubsubmodules, and so on until no further splits are possible. Finally, because the algorithm is stochastic and fast, we can restart the algorithm from scratch every time the clustering cannot be improved further and the algorithm stops. The implementation is straightforward and, by repeating the search more than once, 100 times or more if possible, the final partition is less likely to correspond to a local minimum. For each iteration, we record the clustering if the description length is shorter than the previously shortest description length.[8]

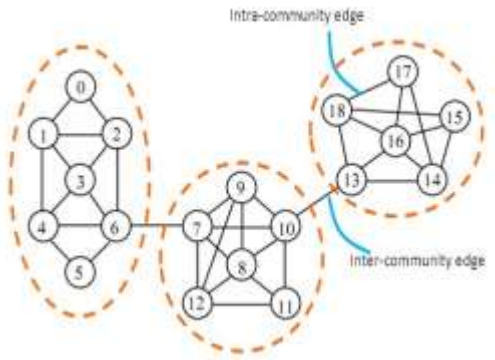

Figure 3: Community Structure in a graph showing intra-community edges and inter-community edges

Clique guidedcommunity detection: A new approach developed for fast and efficient community detection. Clique guidedcommunity detection consists of two phases. In the first phase, the framework finds disjoint cliques. In thesecond phase, the cliques from the first phase are used to guide the merging of individual vertices until a good quality solution is obtained. For the first phase, we develop an algorithm named MACH (Maximum Clique Heuristic), which is a newapproach to compute disjoint cliques using a heuristic-based branch-and-bound technique. The experimental resultsare provided to demonstrate the efficiency of the new algorithm and compare the approach with other previously proposed algorithms.As the framework is adopted the community merge step for the proposed paper takes O(k) time, where k is the number of communities. If the merging is very unbalanced in this phase, it could perform O(n) merges, taking up to O(n2) time for this phase.[14]

Graph Partitionmethod: A graph partition method based on min-max clustering principle was proposed by Ding and Zha et al. The principle states that the similarity or association between two sub graphs is minimized, while the similarity or association within each sub graph is maximized. Luo and Wang et al. proposed a framework to identify modules within a biological network. Networks are divided into subnetworks and the identification of modules is based on their topology. For this, the concept of edgebetweenness was used. Edge-betweenness is the number of shortest path between all pairs of vertices that run through the edge. Edges between modules tend to have shortest paths through them than do edges inside modules and thus have higher betweenness values. The deletion of edges with high betweenness can separate the network, while keeping the modules structure in the network intact. Sun and Castro et al. proposed a framework, MetaFac that extracts community structures from social media networks. Mehler and Skicna presented a general method for network community expansion from seed set of members. It is achieved by assigning a score to all entities in the network and selecting the highest scoring outside vertex to join the community. Some of the scoring criteria in order to rank the selection are neighbor count, juxta position count, neighbor ratio, juxta position ratio, binomial probability. The essential function of the community expansion method is to identify the most promising next member to be added to the community. Some representative community detection methods such as latent space models, block model approximation, spectral clustering and modularity maximization.

Community Detection method Using DBSCAN Algorithm: A community detection methods using DBSCAN algorithm was proposed, which is the most effective unsupervised clustering algorithm. The DBSCAN algorithm can identify clusters in large spatial data sets by looking at the local density of database elements, using only one input parameter. Furthermore, the user gets a suggestion on which parameter value that would be suitable.The DBSCAN can also determine what information should be classified as noise or outliers. In spite of this, its working process is quick and scales very well with the size of the database-almost linearly. From the graphical representation structure of social network, the interactions or connection between individuals or entities, or nodes can be viewed, from which the existence of communities can be concluded. In this approach detection of communities was done on the basis of three types of members in the community, which are core, border and outlier members, and which are of high, low and no influence respectively. The outliers were eliminated from the dataset because they are noises which are free to deal with it.The DBSCAN algorithm is robust to outliers, and by deleting the outliers the dataset will be noise free to deal with. The outliers can be detected by changing the radius of the cluster. In the analysis of social networks it mostly focuses on cores as they have influence to other members and the eliminated outlier member's leads to an accurate

clustering result that helps with the community detection issue in the social network analysis field.[4]

3.RESULTS AND DISCUSSIONS

Community detection is an emerging and rapidly growing area of data mining that is focusedon finding patterns in data by exploiting and explicitly modelling the links among the data instances. Several models have been surveyed for community detection that implements both popularity and productivity of link in community detection.Community detection proves to be an important role for analysis of a large community graph. This paper discusses various techniques on particular community detection and its retrieval in social community graph.

The tabular view of different methods is as shown:

Algorithm	Description	Runtime complexity
Louvain Modularity Method	Greedy optimization method	O(nlogn)
Girvan and Newman Algorithm	Hierarchical method for detecting communities	$O(m^2n)$
Clique GuidedCommunity Detection Method	Modularity based technique	O(k) for balanced merging
DBSCAN Algorithm	Data clustering algorithm	O(n log n)

From the survey it can be inferred that community detection earlier was complex and tedious task. Many methods have been proposed to extract community structures from different types of networks. However these methods may not be full proof but they guarantee correct results. For analyzing large networks such as social networks where the users are constantly changing detecting community is of utmost importance. Different techniques for community detection vary differently in case of accuracy, efficiency and their complexity.Based on simple network properties and the aforementioned results, this paper provide guidelines that may help to choose the most adequate community detection technique for a given network. Table 1 shows the details and the runtime complexity of various methods described in the paper.

Thus a range of network community detection methods have been explored and compared to understand their relative performance. Another important aspect to be considered for large networks is that, one should first choose algorithms which are able to detect the organization of nodes in a reasonable time. This paper explores all the different methodologies and concepts used to identify different kinds of community on the basis of certain pattern or properties, structure and trends in their linkage to extract remarkable knowledge from World Wide Web.Based on the previous results, and taking into account both factors, accuracy and computing time, it is possible to suggest under which situations to use each algorithm depending sorely on topological properties of the network under study.In the end, the paper highlights that detecting the community structure of networks is an important issue in network science. However, based on the results discussed, existing community detection algorithms still needs to be improved to better uncover the ground truth of networks.

4. CONCLUSION

Community detection play a significant role in large social network and also helps in understanding the structure of social networks.Community detection algorithms are widely used to study the structural and topological properties of real-world networks. In this paper, we have compared some of the community detection techniques for social and real-time networks.Identifying and implementing the best community detection technique among the networkis a big challengeThis paper, describes different types of techniques, related to social or real-time networkcommunity detection. Different type of approaches with their run time complexity has been discussed properly and in detail.The proposed techniques were based on the mining algorithms and some of the graph mining approaches for the detection of communities in the social network. All the proposed approaches discussed in the paper will be useful for the other researcher's, doing research in the social or real-time network community detection.

References

[1] Deepjyoti Choudhury, Saprativa Bhattacharjee, Anirban Das, "*An Empirical Study of Community and Sub-Community Detection in Social Networks Applying Newman-Girvan Algorithm*", Emerging Trends and Applications in Computer Science (ICETACS), 2013 1st International Conference.

[2] Dhanya Sudhakaran, Shini Renjith," *Survey of Community Detection Algorithms to Identify the Best Community in Real-Time Networks*", International Journal of Scientific Engineering and

Applied Science (IJSEAS) – Volume-2, Issue-1, January 2016.

[3] Gulab R. Shaikh, Digambar M. Padulkar, "*ASurvey on Template Based Abstractive Summarization of Twitter Topic Using Ensemble SVM with Speech Act*",International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 ,Vol. 2 Issue 11, November – 2013.

[4] Mehjabin Khatoon, W. Aisha Banu, *"A Survey on Community Detection Methods in Social Networks",*I.J. Education and Management Engineering, 2015, 1, 8-18.

[5] Nikita Jain, Vishal Srivastava, "*DATA MINING TECHNIQUES: A SURVEY PAPER*",International Journal of Research in Engineering and Technology(IJRET), Volume: 02 Issue: 11, Nov-2013.

[6] Soumi Dutta, Sujata Ghatak, Moumita Roy, Saptarshi Ghosh and Asit Kumar Das, "*A Graph Based Clustering Technique for Tweet Summarization*",978-1-4673-7231-2/15/$31.00 ©2015 IEEE.

[7] *"Infomap-community-detection"* http://www.mapequation.org/code.html.

[8] LOuvian method http://arxiv.org/abs/*0803*.0476

[9] Yomna M. ElBarawy, Ramadan F. Mohamedt and Neveen I. Ghali, "*Improving Social Network Community Detection Using DBSCAN Algorithm*", Computer Applications & Research (WSCAR), 2014 World Symposium, 2014 IEEE.

[10] L. F. Rau, P. S. Jacobs, and U. Zernik, *"Information extraction and text summarization using linguistic knowledge acquisition"*, Information Processing Management, vol. 25, no. 4, pp. 419 - 428, 1989. \

[11] L. Vanderwende, H. Suzuki, C. Brockett, and A. Nenkova, *"Beyond sumbasic: Task-focused summarization with sentence simpliffcation and lexical expansion,"* Inf. Process. Manage., vol. 43, no. 6, pp. 16061618, Nov-2007.

[12] Sudip Misra, Romil Barthwal, Mohammad S. Obaidat, "*Community Detection in an Integrated Internet of Things and Social Network Architecture*", Global Communications Conference. (GLOBECOM), 2012 IEEE.

[13] "*Community detection and grph partitioning*" https://arxiv.org/abs/1305.4974.

[14] "*Girvan-Newman Method*" https://arxiv.org/pdf/0906.0612v2.pdf