International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 07 | Jul 2023 www.irjet.net p-ISSN: 2395-0072

New Similarity Index for Finding Followers in Leaders Based Community Detection

Sunil Patel[1], Dr Kumar Gaurav[2]

[1]*M.tech Scholar, Dept. Of Electronics Engineering, HBTU, Uttar Pradesh, India*
[2]*Assistant Professor, Dept. Of Electronics Engineering, HBTU, Uttar Pradesh, India*

Abstract *- Currently, the problem of social research has emerged as a challenge, there have been various debate to understand and analyze human relationships in the network. Many researchers have found ways to search the community and select a leader individually, but the algorithm discussed in this paper is suitable for detection of community and community leaders. Generally speaking, people love to connect with people who share the same behaviour of interest, and to form communities based on shared ideas. A small number of residents in the community are in charge of spreading consciousness there, i.e. leader, they represent a group who have conveyed their thoughts and experiences.*

The existing algorithms are based on modularity optimization techniques. Leader based community detection method is not modularity optimization dependent techniques. The accuracy of this technique decreases as network size increases so this paper tells about a modified similarity indexing in leader based selection method to increase the accuracy of the method.

Keywords—Community Detection, Similarity Method, LFR Benchmark

1. INTRODUCTION

A network is defined as group or connection system between people or things. These include a number of devices such as computers, servers, peripherals, humans, etc. These things are known as nodes in the context of network science, and the connections between them are known as edges. The internet is one of the best examples of a network since it connects millions of individuals to a single platform and offers a variety of information and services, including the World Wide Web and e-mail. The people gather together in a place called social network. People tend to connect with people who share the same personality or interest. They form a group of people who have common interest and interact more with each other as compared to others, these group of people are called community. When people within a community interact more, this is referred to as an intercommunity link, and when those outside the community interact less, this is referred to as an intercommunity link.

Currently the goal of banks and business houses is to look for active groups in their network as well as in their customer network. In many communities, some nodes play an important role in innovation, with knowledge and ideas shared in the community. These nodes act as a catalyst to cause turmoil in community. Many researchers look for this catalyst node or most important person in the community [1].

Identifying useful nodes in the network is one of the biggest tasks. In biological system, identification of important node in communities is important, for example in cancer therapy, which require the identification and destruction of cancer cells from the blood and require the maintenance of the body. The second example is the September 11 attack, which involved a network of 62 nodes and 153 connections made up of 5 communities. To keep these communities in control the owners of these communities should be kept in check.

Finding the most influential node in a network depends on the network's structure and centrality. In order to determine a node's power within a network, centrality is utilized to determine which node is the most influential. There exist two types of centrality local centrality and global centrality [2]. Local centrality includes degree centrality and betweenness centrality. The total number of focal nodes connecting one node to another is used to determine degree centrality. The shortest path between each pair of nodes is used to determine betweenness centrality. Closeness centrality is a sort of global centrality that has an inverse connection with the sum of shortest distances between network nodes. This influential node acts as tools in community to share ideas and information making community powerful.

2. LITERATURE REVIEW

The research related to community detection has been started by Girvan and Newman [3]. They proposed GN algorithm in 2002 that detects the community by continuously removing the edges according to the betweenness centrality but this algorithm cannot be applied to large network. They also coined the term "modularity" to describe the strength of a community's

structure. In 2006 a new method was introduced by Pascal pons and Mattieu Latapy called as Walk Trap method [4]. This method works can be applied in directed network. In 2008 again Girvan and Newman improved the algorithm to increase the speed of community detection. This was called Leading Eigen Vector [5]. A new method called Louvain Algorithm was introduced in 2008 by Blondel *et al*. This is also called fast unfolding method; it works in two phases: Modularity Optimization and Community Aggregation [6]. This method increases the speed of detection along with its modularity. This method is unable to detect internally disconnected community.V.A Traag *et al*. discovered the problem in Louvain method and proposed a solution called Leiden Method [7]. It operates in three stages: (1) local node movement, (2) partition refinement, and (3) network aggregation based on refined particles. The above algorithm mentioned optimizes the modularity. Santo Fortunato *et al.* found that modularity optimization may not work if modules identified are of smaller size compared to the size of the network [8].In 2014 Sorn Jarukemratana *et al.* detected the community on the basis of centrality and node closeness [9]. In 2020 leader-based community detection was introduced by A. sarwani *et al*. which was divided in two stages. First we leader is found then according to the leader followers are found according to the similarity index.

3. PROBLEM STATEMENT

The existing algorithms in community detection are based on modularity optimization techniques. Leader based community detection does not include modularity, but similarity. The existing Leader based community detection fails for the network size above 2000 nodes. Here in this paper we shall discuss the method for the selection of the followers which helps in detecting community using LBSD for network size above 2000, maintaining its accuracy.

4. LEADER BASED COMMUNITY DETECTION

Leader-based community detection is a method used to identify communities or clusters within a network or graph structure. This approach relies on the concept of leaders, which are nodes that act as representatives or central points for their respective communities.

Leader-based community detection is a method used to identify and analyze communities within a network or graph. This approach involves the concept of "leaders" or "representatives" that act as central nodes within their respective communities. By identifying these leaders and examining their connections, it becomes possible to uncover cohesive groups of nodes that share similar attributes or exhibit strong interconnections. Leader Based Community Detection involves three steps. The initial step is to locate or pick the network's leader nodes. The next stage is to locate all of the followers who are present. After finding all the followers we will check if any isolated node is present or not if isolated node is found it is added list of followers. Now all the followers form a community for a particular leader and this community is removed from the network.

4.1 LEADER SELECTION

In each community there is most important nodes ,the most important nodes we consider as a leader of the community. To measure the importance of node we calculate the centrality of any each node.

$$\arg\max_{n \in Community(l)} Centrality(n) \quad (1)$$

There is various method available to measure the centrality of any network. To find the leader of the community we calculate the centrality of the network of each node. To calculate the centrality, we have used eigen vector centrality method. It measures the centrality of the node considering the neighbors of node. The eigenvector centrality for node i is the i_{th} element of the vector x defined by the equation

$$Ax = \lambda x \quad (2)$$

Where is A is the adjacency matrix of graph G and λ is eigen value. The node with maximum centrality value is selected as the leader node.

4.2 FIND THE FOLLOWER

To find the follower nodes of the particular leader node, similarity with the leader node is measured.

In this paper a new to method to find follower node has been proposed.

Graph G(u,v), A is the leader node of Graph ,B is node of Graph, X is the number of common neighbor of leader node and node B,Y is the Degree of B

$$\text{Follower Index} = \frac{X}{Y} \quad (3)$$

The Follower Index value define the how strength of connectivity between both node its

value varies from 1 to 0

4.3 FINDING THE COMMUNITY

In this step we find any isolated node which is part of the community but not detected by this method. to find such isolated node we calculate the connected component of the network. if any isolated nodes available in network we add that isolated network to the community. the communities have already been detected, if any isolated node has been detected it is merged with the community.

5. STEPS FOR COMMUNITY DETECTION

Step 1: Calculate the node centrality of the network.

Step 2: Consider the node with maximum centrality as the leader node.

Step 3: Calculate the follower index of every node using eqn (1)

Step 4: Arrange the nodes in ascending order according to the follower index.

Step 5: All the nodes whose follower index value is greater than threshold value is removed from the graph and added to the community.

Step 6: Check all the nodes, if any isolated node is found it is added to the community and removed from the graph.

Step 7: Check the graph, if any node is found go to STEP 1.

Step 8: If no node is found terminate the algorithm, all the communities are found.

6. SIMULATION STEPS

This section simulates Community detection algorithm to obtain a leader using similarity method. We have performed the simulation in Python Language in Anaconda Navigator. The code is executed on Jupyter Notebook in Anaconda Navigator on an Intel(R)Core (TM) i5-7200U CPU running at 2.50GHz with 16GB RAM on Windows 10 64-bits.

The synthetic network is created using LFR Model [10]. The performance of the community detection is measured using Normalized Mutual Information (NMI) and Adjacent Rank Index (ARI) [11].

6.1 EVALUATION METRICS

To measure the performances of community detection by comparing with ground truth.

Adjusted Rand index(ARI): The Adjusted Rand Index (ARI) is a measure of the similarity between ground truth and detected community, but it has the disadvantage of being sensitive to chance. ARI ranges from-1 to 1. Negative 1 means there is no agreement between ground truth and detected community and 1 means there is full agreement between ground truth and detected community. Let a,b,c and d denote the number of pairs of nodes that are respectively in the same community in G and R , in the same community in G but in different communities in R , in different communities in G but in same communities in R and in different communities in G and R then ARI is computed by the following formula [11]

$$\mathrm{ARI} = \frac{\binom{n}{2}(a+d)-[(a+b)(a+c)+(c+d)(b+d)]}{\binom{n}{2}^2-[(a+b)(a+c)+(c+d)(b+d)]} \qquad (4)$$

NORMALIZED MUTUAL INFORMATION (NMI): Normalized Mutual Information (NMI) is a measure used to evaluate network partitioning performed by community finding algorithms. To compute the NMI we use following formula [11]

$$\mathrm{NMI(G, R)} = \frac{2XI(G,R)}{H(G)+H(R)} \qquad (5)$$

Where I(G,R) mutual information of G and R H(G) and H(R) is entropy

Its value range from 0 to 1.

6.2 DATASET

The proposed method is run on synthetic network social network. there are limited number of real social network dataset are available. There are various model available to generate the social network but LFR model is network is so close to real world network and give better control to generate to network .

7. SIMULATION

A synthetic network is created and leader-based community detection method is applied to it. The various parameter considered to make synthetic network are:

TABLE 1: Parameters of LFR

PARAMETER	VALUE
No. Of node(n)	200-4900
Maximum number of communities	10 % of n
Mixing (μ)	0.3
Maximum Degree	10 % of n
μ1	3
μ2	1.5

7.1 STEPS OF SIMULATION

Step 1: Create a network.

Step 2: Define the threshold value.

Step 3: Apply the above algorithm on the network.

Step 4: If number of detected communities is less than The Number of ground truth community, update The threshold value and go to STEP 2.

Step 5: Calculate the NMI, ARI value.
Step 6: Store the results Step 7: Plot the result.

8. RESULTS

a) To find the accuracy of the algorithm, we plot the graph between number of network size and ARI value. This is shown in fig 1. The graph shows that ARI value will lie between 0.8 to 0.9 as network size increases from 2000 to 4000.

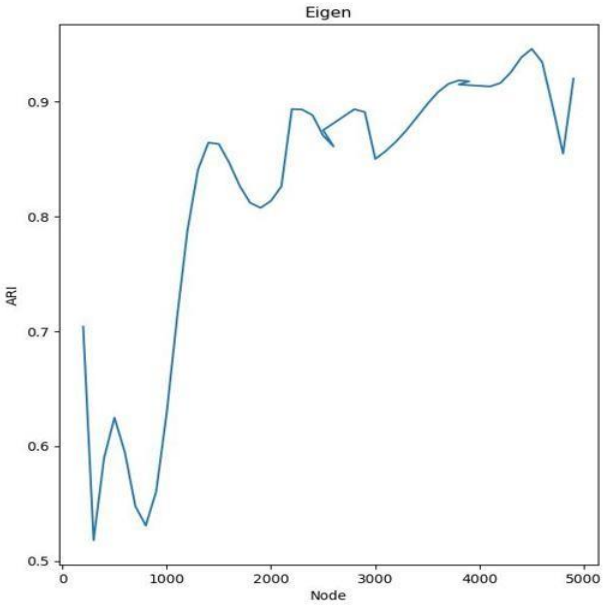

Fig 1: Plot between ARI and network size

b) To find the accuracy of the algorithm, we plot the graph between number of network size and NMI value. This is shown in fig 2. The graph shows that NMI value will lie between 0.85 to 0.95 as network size increases from 2000 to 4000.

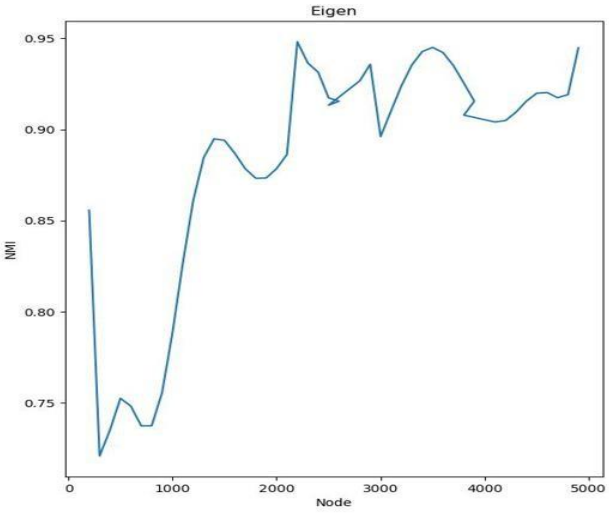

Fig 2: Plot between NMI and network size

c) Plot between number of detected community and number of ground truth community. This is shown in fig 3. This graph shows that number of communities found are equal in number to the number of communities formed by LFR

Benchmark also called as ground truth.

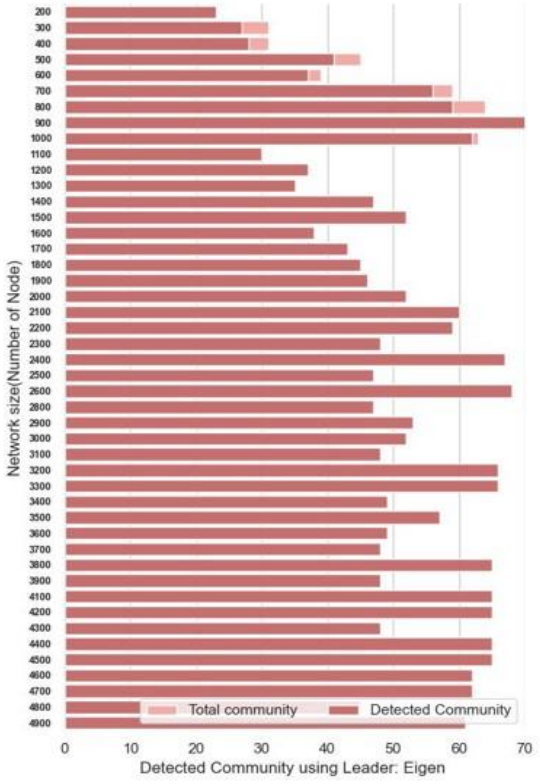

Fig 3: Plot between number of detected

d) Plot between community and network sizeexecution time and network size .

This is shown in fig 4. This graph shows that as the network size increases, the execution time also increases along it.

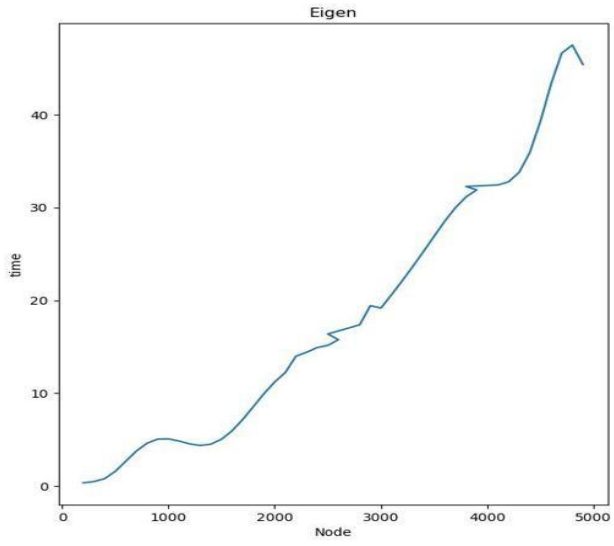

Fig 3: Plot between execution time and network size

9. CONCLUSION

The aim of this paper was to detect the community in the large network without using modularity optimization with high accuracy. The leader based community detection was used for this purpose as it is the technique which does not uses modularity optimization. The drawback of this method is that for network size greater than 2000, it fails to detect the community and the accuracy of this method decreases. To overcome this problem, we have used "New Similarity Index for Finding Followers in Leaders Based Community Detection"

The ARI value ranges from 0.8 to 0.9 for the network with node size between 2000-4000.The NMI value ranges from 0.85 to 0.95 for the network with node size between 2000-4000.These results shows that the proposed similarity method for the community detection is capable to detect the community in the network size range of 2000-4000 with high accuracy. The disadvantage this algorithm is that the time increases with the network size.

REFERENCES

[1] Wang, Y., Di, Z., Fan, Y., Identifying and characterizing nodes important to community structure using the spectrum of the graph. PLoS One 6 (11), p.e27418, (2018).

[2] Gao, S., Ma, J., Chen, Z., Wang, G., Xing, C., Ranking the spreading ability of nodes in complex networks based on local structure. Phys. Stat. Mech. Appl.403,130–147, (2014).

[3] Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks." Proceedings of the national academy of sciences 99.12 (2002): 7821-7826.

[4] Pons, P., Latapy, M.. "Computing Communities in Large Networks Using Random Walks". In: Yolum, p., Güngör, T., Gürgen, F., Özturan, C. (eds) Computer and Information Sciences - ISCIS 2005. ISCIS 2005. Lecture Notes in Computer Science, vol 3733. Springer, Berlin, Heidelberg

[5] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," Physical Review E, vol. 74, no. 3, Sep. 2006

[6] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E., "Fast unfolding of communities in large networks",Journal of Statistical Mechanics:

Theory and Experiment, vol. 2008, no. 10, p. 10008, 2008.

[7] Traag, V.A., Waltman, L. & van Eck, N.J. "From Louvain to Leiden: guaranteeing well-connected communities". Sci Rep **9**, 5233 ,2019

[8] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," Physical Review E, vol. 78, no. 4, Oct. 2008

[9] S. Jarukasemratana, T. Murata, and X. Liu, "Community Detection Algorithm based on Centrality and Node Closeness in Scale-Free Networks," Transactions of the Japanese Society for Artificial Intelligence, vol. 29, no. 2, pp. 234–244, 2014

[10] Avani Kesarwani, Ashutosh Kumar Singh, Kumar Gaurav, and Ashok Kumar Shankhwar, "Leader Similarity Based Community Detection Approach for Social Networks," Nov. 2020

[11] Liu, Xin, Hui-Min Cheng, and Zhong-Yuan Zhang. "Evaluation of community detection methods." IEEE Transactions on Knowledge and Data Engineering 32.9 (2019).

[12] Dao, V., Bothorel, C., & Lenca, P. (2020)."Community structure: A comparative evaluation of community detection methods." Network Science, 8(1), 1-41.

[13] Taha, Kamal. "Detecting disjoint communities in a social network based on the degrees of association between edges and influential nodes." IEEE Transactions on Knowledge and Data Engineering 33.3 (2019): 935-950