Improved Dynamic Community Detection based on Distance Dynamics in Real World Network

Komal Prajapati

PG Student, Department of Computer Engineering,

GEC Gandhinagar, Gujarat, India

Prof. M. B. Chaudhari

Professor, Department of Computer Engineering,

GEC Gandhinagar, Gujarat, India

Abstract

In today's era, real world networks have become quite prominent and a lot of research is getting done for making these networks valuable, by proposing helpful communities and similar interest communities to users. Moreover, due to freely available web space and interactions, there are different communities per user, making it extremely hard for highly iterative detection algorithms perform quickly and give important suggestions. Furthermore, real world networks are dynamic in nature. Along this lines, there is a requirement for dynamic community detection algorithm which can appropriately detect communities with time differs. Because it accepts the changes in network rather than static community detection algorithm. A Real world network is vital complicated network. There is a number of algorithms developed to detect communities. Dynamic Community Detection based on Distance Dynamics algorithm doesn't think about different cohesiveness of each neighbor node. In this way, because of the different degree of neighbor node, different attractive strength from the neighbor node to two end points. Here, we propose improved Dynamic Community Detection based on Distance Dynamics (iDC3D) algorithm which use neighbor cohesion for getting different cohesiveness of neighbor nodes. In this, we calculate distances between nodes. Then it uses neighbor cohesion. After that it uses two new interaction patterns for getting the influences between nodes. Then we will get new distances between nodes. We will get communities from the real world network. We demonstrate that the proposed improved Dynamic Community Detection based on Distance Dynamics (iDC3D) algorithm terminated on a decent community number and additionally has comparable detection accuracy with existing approach.

Keywords: *Communities, Dynamic community detection, Distance dynamics*

---

1. INTRODUCTION

Using data mining technique, we can extract the necessary information from massive resources of information. It provides extracting relevant information from the massive volume using suitable algorithms. Community is an important feature of real world networks. Real world network includes number of networks such as Social network, Biological network, Information network, Technological network, etc. Basically, networks are constituted by nodes which represent "entities" and edges which represent "interaction" between pairs of entities. For example, WWW (World Wide Web) is a real world network in which nodes represent a web pages and edges represent reference between web pages.

The aim of a community detection algorithm is to divide the nodes of a network into some number of groups. The connections between nodes inside a community are much denser while the connections between communities are sparser[1]. These groups known as communities of network. For example, take a group containing a set of people, who have strong mutual relationships and weak with people outside the group. Using the community detection algorithm in real world network, one can get some meaningful and important information.

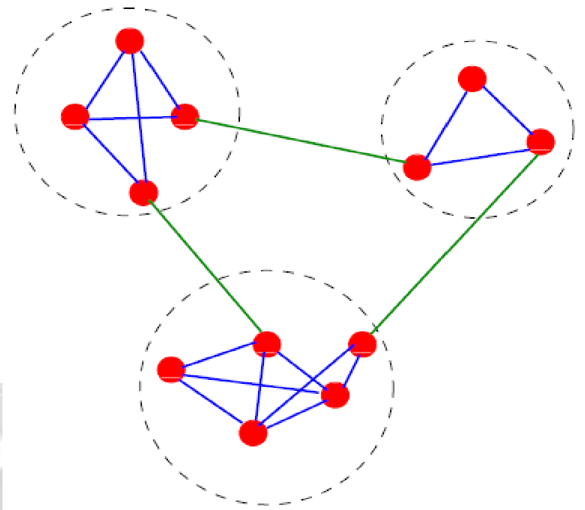

Fig 1. Three communities in a graph, enclosed by dashed circles[8]

2. RELATED WORK

In this section, we present the different kind of algorithm which exists for detecting dynamic community from real world network.

In 2016, Qian Guo , Lei Zhang , Bin Wu , Xuelin Zeng proposed a DC3D algorithm [1], which is based on distance dynamics. The drawback of this algorithm is that it doesn't consider the different cohesiveness of neighbor nodes. In 2015, Kamal Sutaria, Dipesh Joshi, Dr. C. K. Bhensdadiya, Kruti Khalpada [4] proposed an algorithm for detecting communities which based on vertex relation and their modularity. The drawback of that is it cannot contain many overlapping nodes in communities. GeoSim algorithm [11] is based on their geodesic location and similarity between their interests. Drawback is that it focus on both distance and similarity interest terms. So, for placing two nodes in one community, both have to close to each other and have similarity interest. In 2014, Zeineb Dhouioui, Jalel Akaichi proposed an Data Warehouse for Dynamic Communities(DDCom) algorithm [12], which is based on indicators where important events can occur. In 2016, Zhe Dong proposed a Hidden Markov Model for Dynamic Community detection(HMM_DC) algorithm [13], to detect communities in dynamic social network. Drawback is that it has lower performance in sparse graph and cannot get global optimization community structure without enough information. Dynamic Algorithm Based on Permanence(DABP) algorithm [14], based on incremental identification according to vertex based metric called permanence. The drawback of this algorithm is that it gets much lower modularity. In 2017, Nivin A. Helal, Rasha M. Ismail, Nagwa L. Badr, Mostafa G. M. Mostafa proposed a Leader Based Community Detectio(LBCD) algorithm [20], which is based on forming local communities around nodes with great influence. Drawback is that it doesn't include more information attached to nodes, so it decrease the accuracy.

3. BASIC TERMINOLOGY

3.1 Jaccard Distance :

It measures the similarity of neighbor set between nodes. The smaller distances between nodes shown there is more and more common neighbors between nodes. Jaccard distance defined as below [5]:

$$d(u,v) = 1 - \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$$

where, N(u) is the neighbors of node u and N(v) is the neighbors of node v.

3.2 Accumulated Increment(AI) :
Accumulated Increment[1] is used to measure accumulative changes in dynamic networks. It is introduced to estimate the error[1]. It is used to find the changes from every network snapshot. It give different result if any changes occurs in number vertex and/or edges. It is defined as:

$$AI(t)=AI(t\text{-}1)+\frac{|vertex^{+}(t)|+|vertex^{-}(t)|}{|V_{t-1}|}+\frac{|edge^{+}(t)|+|edge^{-}(t)|}{|E_{t-1}|}$$

where, $|V_{t-1}|$ is number of vertex at time t-1, $|E_{t-1}|$ is number of edge at time t-1.

4. PROPOSED METHODOLOGY

Given a real world network G (V, E), where V=set of nodes and E=set of edges. Our algorithm aims to find dynamic communities in real world network.

In a proposed algorithm, we have used jaccard distance to finding the distances between nodes in a network, which is a very well known and precise measure in a field of data mining.

Input: Undirected graph G (V, E)
Output: Set of communities

Explanation:

In this algorithm, we find distance between nodes in a network using jaccard distance method. After getting distances, we find cohesiveness of different neighbor nodes. Then after for getting influences between nodes, we use two new interaction patterns namely Common Neighbor Influence(CI) and Exclusive Neighbor Influence(EI). Then using these all, we get new distances. We will get communities from real world network.

Algorithm:

Steps of proposed algorithm are shown below:

Step-1: First it will calculate initial distances between nodes by Jaccard distance in a graph of the network.

Step-2: Then it will find neighbor cohesiveness of nodes.

Step-3: Then it use two new interaction patterns for getting influence between nodes.

Step-4: Then it add these influence result to initial distance for getting new distance.

Step-5: After that it calculate Accumulative Increment (AI) for every snapshot. Then using AI to determine whether need to recalculate whole graph.

Step-6: Then check whether vertex and/or edge can be added/deleted.

Step-7: Distance updated iteratively to achieve community detection.

Step-8: It will update with new interaction patterns detect small communities.

Step-9: Then display communities.

As shown in Fig. 2, we will get the final two communities of network. It shows two communities with two different color of nodes. We will take one community as nodes with orange color and other community as nodes with red color. So, that's the output of our proposed algorithm on karate club network.

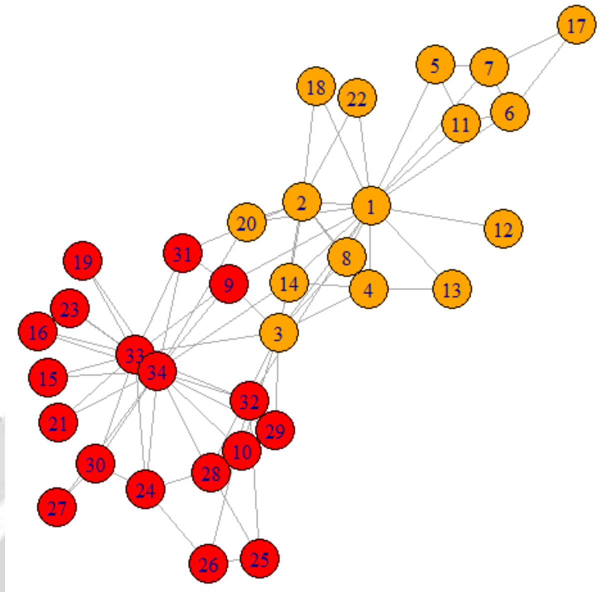

Fig. 2 Two communities of Zachary's karate club network

5. EXPERIMENT

5.1 Evaluating Measures:

1. Modularity :

It is used to measure the strength of division of a network into communities. High modularity of network have dense connections between the nodes within communities rather than in different communities. It defined as a difference between actual interactions of nodes and expected number of connections.For a network, it has divided into number of communities and its modularity can be defined as[8]:

$$Q = \sum_{c=1}^{n_c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right].$$

where, n_c is a number of communities, l_c is a number of edges inside the community, d_c is a total degree of nodes inside a community, m is a total number of edges.

2. Normalized Mutual Information(NMI) :

It is a popular criterion for evaluating the accuracy of community detection. It measures the difference or similarity between the true partition and detected partition. If X = $\{x_1, x_2,\ldots,x_a\}$, Y = $\{y_1, y_2,\ldots,y_b\}$ are two partitions, the formulas of NMI are given below:

$$NMI(X,Y) = \frac{I(X,Y)}{\sqrt{H(X)H(Y)}}.$$

H(X) is the entropy of partition X.

$$H(X) = \sum_{x \in X} p(x) \log p(x)$$

Mutual information of two random nodes is a measure of the mutual dependence between them. It shows the shared information between two partition.

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right),$$

5.2 Experimental results

Karate club

Karate club is a classic data set used to validate the social network analysis community detection algorithm [12]. The dataset is obtained by sociologists Zachary in the early 1970s, with two years watched the social relations between members of a university in karate club in the US. It includes 34 nodes and 78 edges.

Table 1: Results of Modularity and NMI on karate club

|V|	|E|	Modularity(Q)		NMI	
		DC3D[1]	iDC3D	DC3D[1]	iDC3D
28	61	0.3737	0.4635	0.5583	0.6846
33	74	0.3279	0.4904	0.5383	0.6741
34	78	0.4634	0.4933	0.4997	0.6686
35	83	0.4595	0.4926	0.604	0.7406
40	87	0.4463	0.4875	0.5463	0.6965

Now, we have taken a karate network with different number of nodes and edges for showing its dynamic nature. We have calculated modularity and NMI value of our proposed algorithm and DC3D algorithm for karate network with different number of nodes and edges. In a Table 1, we can see that the comparison of our proposed algorithm and DC3D algorithm in terms of modularity and NMI value. Modularity and NMI of improved DC3D algorithm is good than DC3D algorithm. Benefit of improved DC3D algorithm is that it gives accurate result because it uses neighbor cohesion. Existing algorithm doesn't consider this.

In a fig. 3, it shows graph of the value of modularity of our proposed algorithm and DC3D algorithm. In a graph, x-axis shows the number of nodes and y-axis shows the value of modularity. Here, iDC3D has higher value of modularity than DC3D.

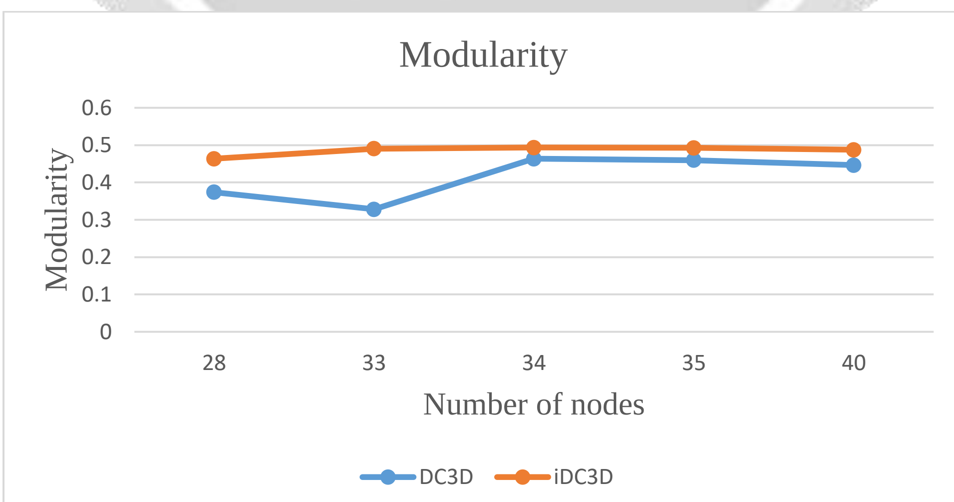

Fig. 3 Modularity of different algorithms in Karate network

In a Fig. 4, it shows graph containing NMI values of our proposed algorithm and DC3D algorithm. In a graph, x-axis shows the number of nodes and y-axis shows the value of NMI. Here, iDC3D has higher value of NMI than DC3D.

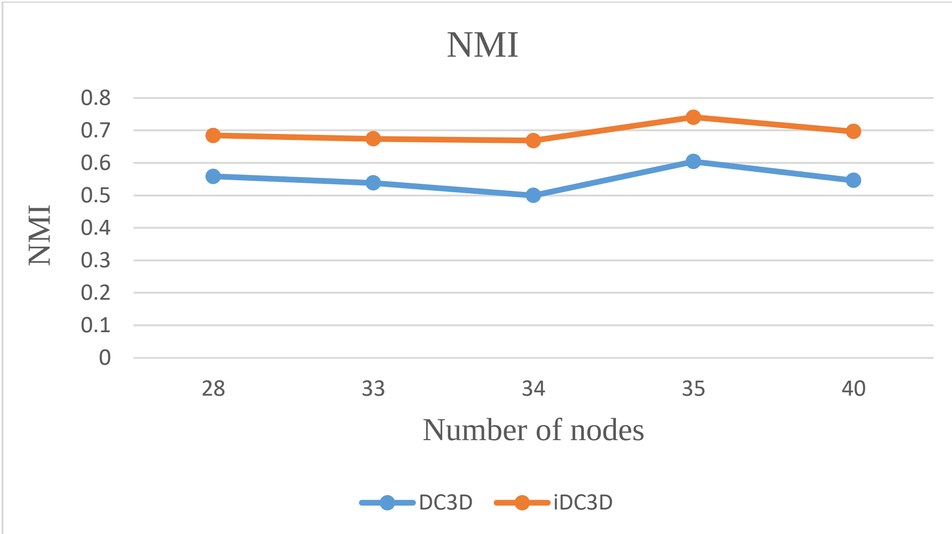

Fig. 4 NMI value of different algorithms in Karate network

So, that is the comparison of our proposed approach and DC3D algorithm. Experiments done on karate network dataset. These both graph shows that improved DC3D algorithm gives better result than existing one.

6. CONCLUSION

Here, we proposed an improved dynamic community detection algorithm based on distance dynamics in real world network. In the algorithm, we uses jaccard distance, neighbor cohesion and distance dynamics for discover community. Using jaccard distance method, we will get the initial distances between nodes in a network. After, we uses neighbor cohesion for getting different cohesiveness of neighbor nodes. Then it uses two new interaction patterns for getting influence between nodes. After that, we get the new distances between nodes and will discover the communities. Here, benefit of this proposed algorithm is that it use the neighbor cohesion, so we will get more accurate result. DC3D algorithm doesn't consider the different cohesiveness of neighbor nodes. For calculate the accuracy of proposed methodology, we did experiment on zachary's karate club data set. Then we compared the results with DC3D algorithm. The comparison of result shows that our proposed methodology obtains a greater modularity and NMI value. In future, we will do experiment on large dataset using proposed methodology and also extend it to detect overlapping community.

7. REFERENCES

1. Qian Guo , Lei Zhang , Bin Wu , Xuelin Zeng. "Dynamic Community Detection based on Distance Dynamics" International Conference on Advances in Social Networks
Analysis and Mining (ASONAM), IEEE/ACM, 2016

2. M. Girvan, M. E. J. Newman. "Community Structure in social and biological networks"
Proceedings of the National Academy of Sciences of the United States of America (PNAS), 2002, 7821-7826.

3. J. Shao, Z. Han, Q. Yang and T. Zhou, "Community Detection based on Distance Dynamics," in Proceedings of the 21st ACM SIGKDD international conference on Knowledge discovery and data mining,

ACM, 2015, PP. 1075-1084.

4. Kamal Sutaria, Dipesh Joshi, Dr. C. K. Bhensdadiya, Kruti Khalpada. "An Adaptive Approximation Algorithm for Community Detection in Social Network" 2015 International Conference on Computational Intelligence & Communication Technology, IEEE, 2015

5. Tao Meng, Lijun Cai, Tinqin He, Lei Chen, Ziyun Deng. "An improved community detection algorithm based on the distance dynamics" International Conference on

   Intelligent Networking and Collaboration Systems, IEEE, 2016.

6. C.D. Manning, P. Raghavan and H. Schutze, "Introduction to InformationRetrieval," Cambridge: Cambridge University Press, 1 edition, 2008.

7. Di Zhuang. "Modularity-based Dynamic Community Detection" JOURNAL OFLATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015.

8. Santo Fortunato, "Community detection in graphs," Physics Reports, vol. 486, no. 3-5, pp. 75-174,2010.

9. Danielle S. Bassett, Mason A. Porter, Nicholas F. Wymbs, Scott T. Grafton, Jean M. Carlson, Peter J. Mucha. "Robust Detection of Dynamic Community Structure in Networks" 013142 arXiv:1206.4358 [physics.data-an], June 2012.

10. Jiawei Han and Mickeline Kamber, Data Mining: Concepts and Techniques, 2ndEdition, Morgan Kaufmann Publishers, 2011.

11. Mohammed Ba Hutair, Zaher Al Aghbari, Ibrahim Kamel. "Social Community Detection based on Node Distance and Interest" 3rd International Conference on Big Data Computing, Applications and Technologies, IEEE/ACM, 2016.

12. Zeineb Dhouioui, Jalel Akaichi. "Tracking Dynamic Community Evolution in Social Network" International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM, 2014.

13. Zhe Dong. "Dynamic Community Detection Algorithm based on Hidden Markov Model" International Symposium on Advances in Electrical, Electronics and Computer Engineering (ISAEECE), 2016.

14. Xiaoming Li, Bin Wu, Qian Guo, Xuelin Zeng, Chuan Shi. "Dynamic Community Detection Algorithm Based on Incremental Identification" 15th International Conference on Data Mining Workshops, IEEE, 2015.

15. https://en.wikipedia.org/wiki/Social_network

16. https://scholar.google.co.in/

17. https://www.google.co.in/search?q=social+network&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjg1Zqm9ufXAhURTY8KHa51DkYQ_AUICigB&biw=1349&bih=635#img_rc=_

18. https://www.google.co.in/search?q=scientific+collaboration+network&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjmtMCr9ufXAhVEvY8KHf_kBIkQ_AUICygC&biw=1349&bih=635#imgrc=_

19. https://en.wikipedia.org/wiki/R_(programming_language)

20. Nivin A. Helal, Rasha M. Ismail, Nagwa L. Badr, Mostafa G. M. Mostafa. "Leader-based community

detection algorithm for social networks" WIREs Data Mining Knowl Discov 2017, e1213. doi: 10.1002/widm.1213

21. Margaret H. Dunham, S. Sridhar. "Data Mining: Introductory and Advanced Topics" 1st Impression, Pearson Education, 2006

22. https://en.wikipedia.org/wiki/Mutual_information

23. Strehl A, Ghosh J. "Cluster ensembles---a knowledge reuse framework for combining multiple partitions[J]". The Journal of Machine Learning Research, 2003, 3: 583-617.

24. Komal Prajapati, M. B. Chaudhary "A Survey on Dynamic Community Detection based on Distance Dynamics in Real World Network". International Journal of Innovative Research in Technology, February-2018, Volume 4 Issue 9 ISSN: 2349-6002.