An Adaptive Approximation Algorithm for Community Detection in Social Network

Kamal Sutaria Research Scholar, C.U.SHAH University, Wadhwan, Gujarat,
Dipesh Joshi Lecturer Computer Engineering Department V.V.P. Engineering College Rajkot,
Dr.C.K.Bhensdadiya Professor & Head, Dharamsin Desai University, Nadiad, Gujarat.
Kruti Khalpada Research Scholar, C.U.SHAH University, Wadhwan, Gujarat.

Abstract— Social network is one of the most important complex networks, which aims to describe the interactive relationship among a group of active actors that represent different kind of structure. Many systems in the real world such as human societies and different types of components can be modeled as social networks. We can represent such a network in terms of graphical community. Social Network Analysis provides inherent research for the success of social media sites and social content sharing facility. Social Network Analysis provides key terms to provide platform for industry to generate survey of product and facilitate to introduce new innovation ideas to public entity. Now a day, as increase the use of social media sites provide the entrepreneurs and user to define new concept of community creation that represents the relationship of users that might be interested in same kind of activity. To create such communities introduce new research area for researcher. This community detection is different from traditional clustering. In This paper, we propose new algorithm for community detection in social network to get some meaningful and important information.

Keywords— Social Network Mining, Community Detection, Social Network Analysis, Data Mining

INTRODUCTION

Data Mining is a technique that helps to extract important information from a large information resources. It provide a way to extract relevant information from these large volume using appropriate algorithms. Social Network Analysis is the way of study the social phenomena in particular social setting. The analysis is carried out based on some small community or social network, interviews, questionnaires and other methods. Study of social network basically focus on structural components of network, interaction between nodes and information supplied between nodes.

Basic issue that we considered during analysis of social network are preparation of data , data connection like follow friends, like and dislikes of node, request by nodes, interaction of nodes, extracting the nodes that have same interest, private and personal information of nodes. We are focusing on main concept of extracting nodes that have same interest for particular topic.

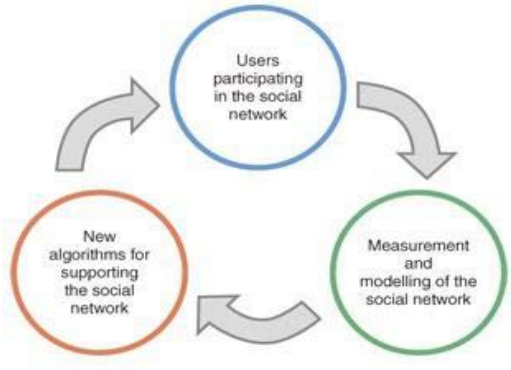

Figure 1: Life cycle of social network mining [7]

II. COMMUNITY DETECTION

Community detection is basic issue that encounter during social network mining. Community is structural component that represent interaction of nodes and for that initially we have link predicted between any nodes to generate the community. Community detection is the extraction of related nodes and put it in specific groups. Detected Community is generally used to analyses marketing strategy of organization, find the interaction of employee in organization, to get interest of customer for specific area. Community detection methods divide each node in a group that satisfies specific properties and these groups are disjoint set of social network. Hierarchical structure of such groups and networks represent complex community structure. Two main property is considered for detecting community: Betweenness and Modularity. Betweenness represent those edges or connection that are least central or most central between communities whereas modularity represent strong connectivity between nodes as well as other community.

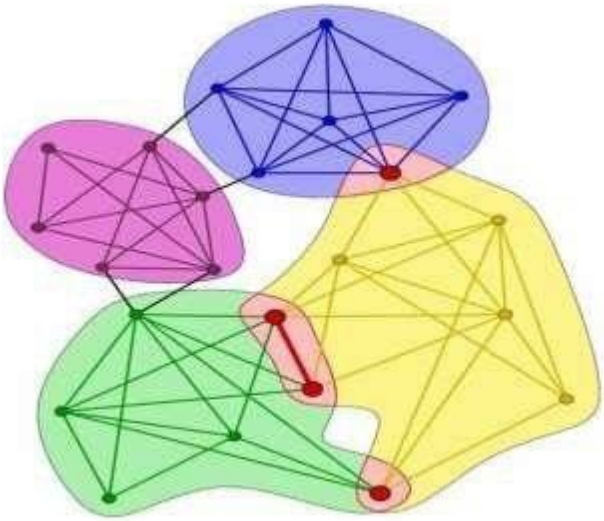

Figure 2: Community Structure

II. RELATED WORKS

Basic Algorithm for community detection is work on graph terminology. First, it divide the network in nonempty groups that represent the communities and each vertex belongs to one of the communities. Second, Division of nodes are performed based on certain properties that provide best partition. Girvan Newman algorithm provide better understanding for community detection and use efficient way to find the community based on below steps.

1. Calculate the betweenness for all edges in the network

2. Remove the edge with the highest betweenness.
3. Recalculate betweenness for all edges affected by the removal.
4. Repeat from step 2 until no edges remain. As this algorithm gives the better result using edge betweenness but it has many limitation such that it is unable to find overlapping community structure, no splitting technique, when to stop execution is not defined.

III. METHOD AND PROCEDURE

We proposed a community detection method that are performed based on vertex relation and their modularity. We make our method more precise based on dynamically Steps of Proposed Algorithm:

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 this proposed algorithm, we repeatedly checking for changes in node structure by deleting some node or by updating of node performed. Algorithm is adaptively checking the status of nodes and performed the community detection dynamically as well as calculation of communities of partition. **Case 1, Case 2, Case 3 are describe below.** Case 1: when both vertices are new then new event is executed and both vertices are placed in same community index and added in to partition list and visited vertex list.

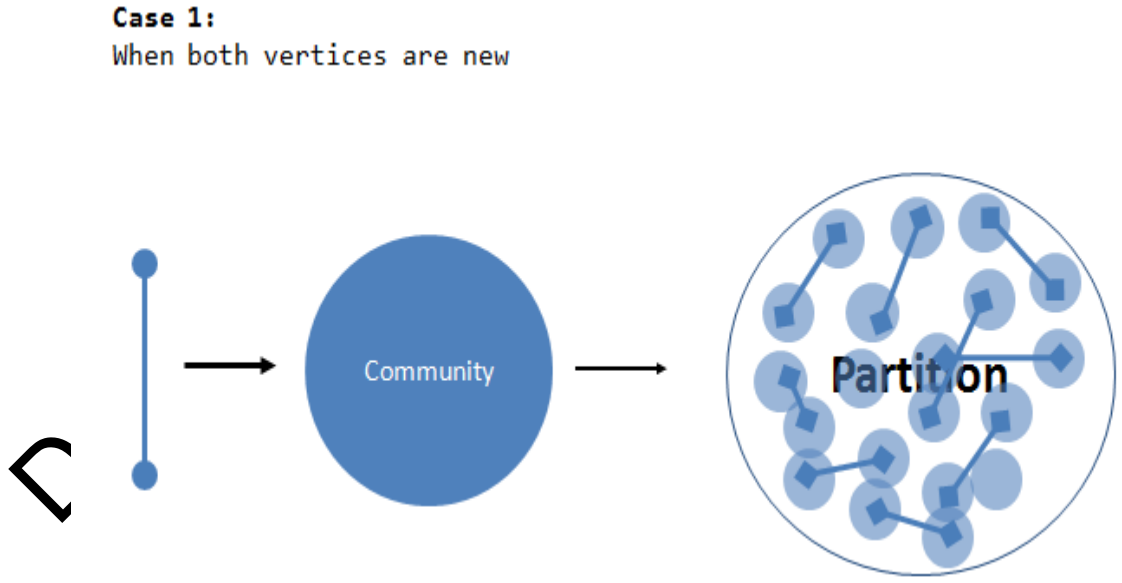

Figure 3: Case 1

Case 2 when one vertex is already visited and other one is new then either join or split event is performed based on condition represented in below figure. When vertex I is already visited but vertex j is new then need to calculate strongly connected property using Bernoulli distribution or betweenness. If we find vertex I is strongly connected then

vertex j move to the community of vertex I otherwise it will create new community.

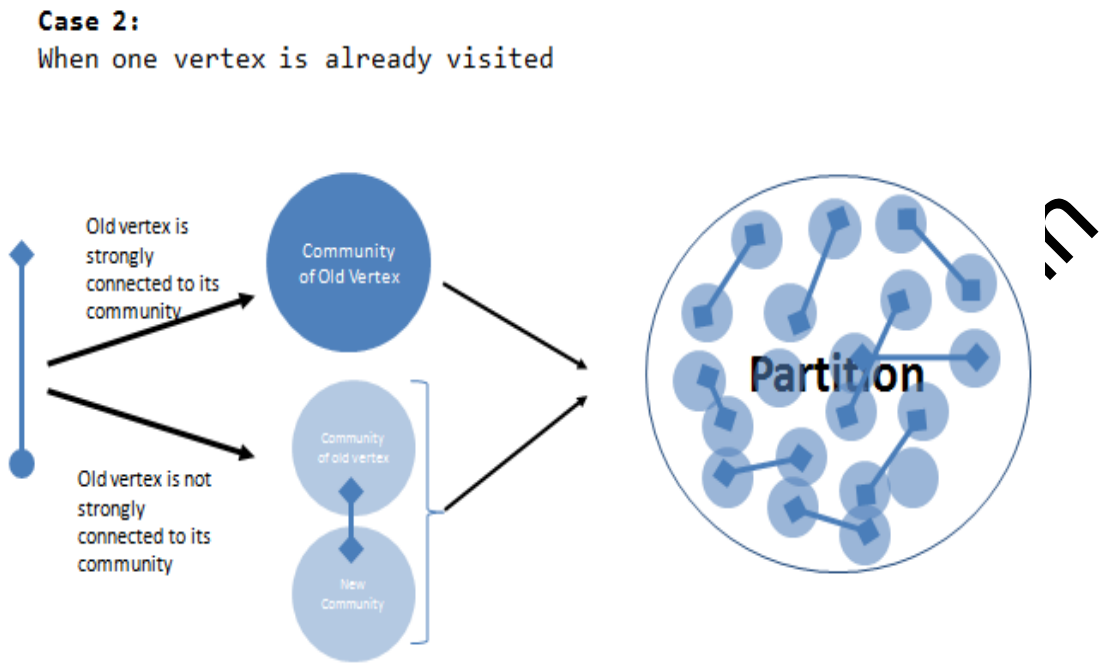

Figure: 4 Case 2

Case 3: When both vertex are already visited then calculation is made based on modularity and tightly coupled Node. Node is transferred to the community which contain maximum modularity because it is densely connected with friend node.

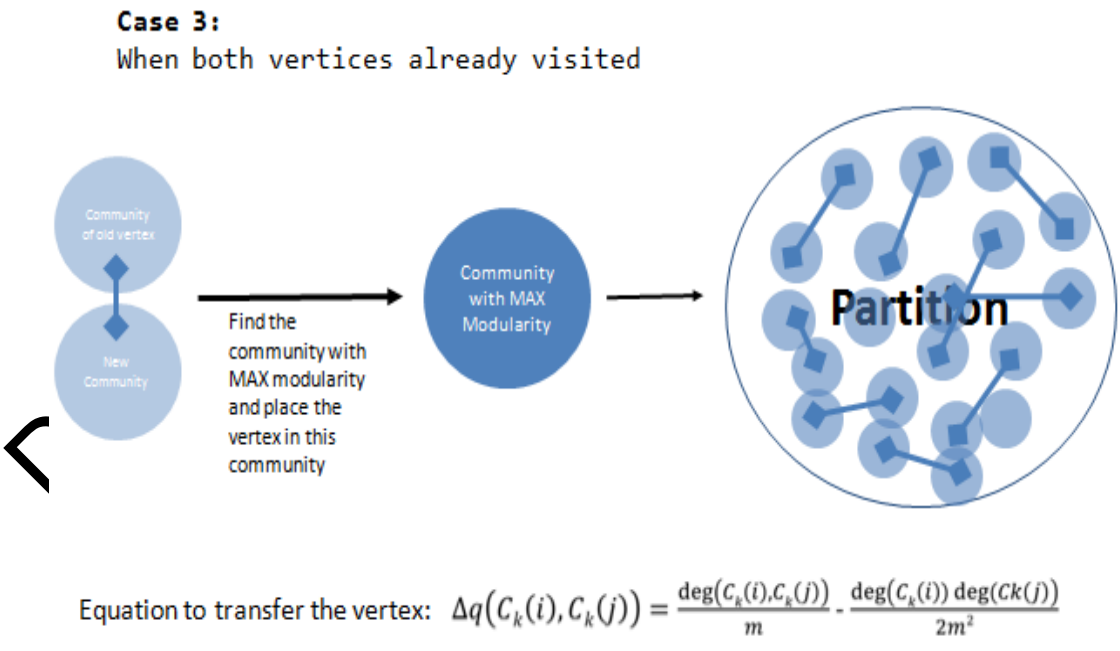

Figure 5: Case 3

We implemented this algorithm in web platform of .Net. We use C# and ASP.Net concept with .Net Framework 4.0.

IV. EXPERIMENTAL RESULTS

Dataset[8]

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

Results after Performing the algorithm.

Sr no	Dataset Name	No. Of Com-	Modularity	Time(In ond)
1	CA	1473	0.6020	0.23
2	CA Cond-	3089	0.6513	0.24
3	Web Stan-	41575	0.8683	7
4	Web Google	130865	0.8375	20

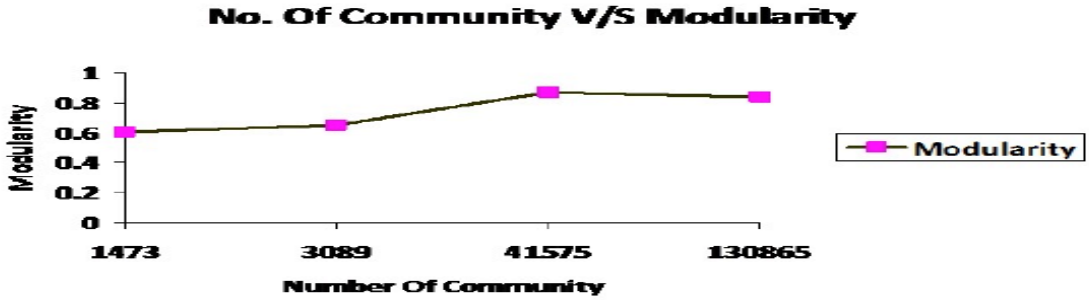

Figure 6: Graph represents the X-axis (No of Community) and Y axis (Modularity).

Modularity based comparison our method represents Our Proposed method whereas EIG represent eigenvector based method [Newman, 2006].

Sr no	Dataset Name	Modularity(Our Method)	Modularity(EIG)
1	CA	0.6020	0.571
2	CA Cond-	0.6513	0.251
3	Web Stan-	0.8683	0.050
4	Web Google	0.8375	0.034

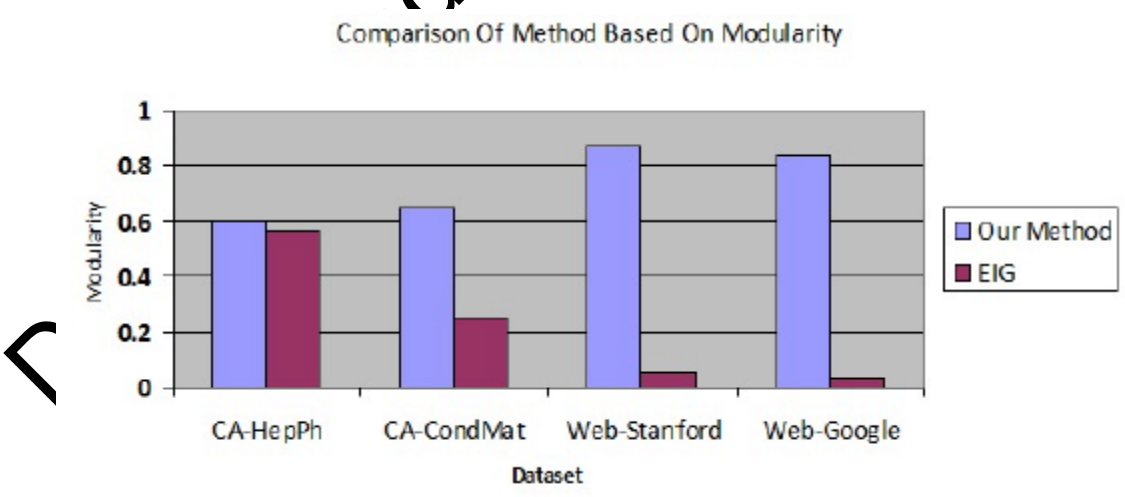

Figure 7 : Comparison of Our Method V/S EIG Based On Modularity

Time based comparison

Sr no	Dataset Name	Time in ond(Our Method)	Time in ond(EIG)
1	CA	0.23	2.7
2	CA Cond-Mat	0.24	1.5
3	Web Stan-ford	7	25
4	Web Google	20	48.3

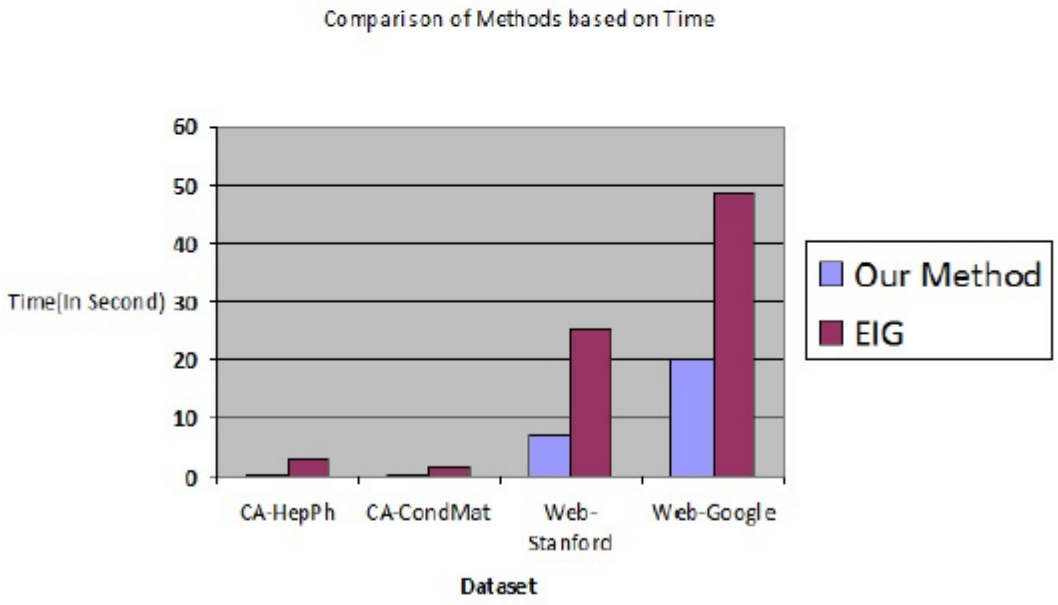

Figure 8: Comparison of Our Method V/S EIG Based On Time

CONCLUSION

As Proposed Algorithm is more precise then Eigen vector based algorithm based on modularity and computation time. Our aim to use this algorithm for complex and dense network as network contains many overlapping nodes and crossed edges. In future, we are looking to implement community detection algorithm for following properties: • Overlapping Community • Interest Based Community Detection

REFERENCES

[1] Vincenzo Nicosia , Dipartimento di Ingegneria Informatica e delle Telecomuni- cazioni Universit di Catania Italy,lecture notes on Modularity for community detection: history,perspectives and open issues

[2] Wangqun Lin, Xiangnan Kong, Philip S. Yu, Quanyuan Wu, Yan Jia, Chuan Li , Community Detection in Incomplete Information Networks,Presented in WWW 2011

[3] Yiannis Kompatsiaris ,(2013), Social Networks Mining for Innovative Applications and Users Well-Being FP7 ICT Work Programme 2013 Consultation Networked Media

[4] 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.

[5] Ellison, Nicole B. Social network sites: Denition, history, and scholarshipJournal of Computer-Mediated Communication 13.1 (2007): 210-230.

[6] Fast algorithm for detecting community structure in networks, M. E. J. Newman

[7] Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks" Proceedings of the National Academy of Sciences

[8] (Stanford Large Network Dataset Collection https://snap.stanford.edu/data/ Time