iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5152

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Community Overlapping Detection using Social Media Dataset and Redundant Node Elimination

Narinderpal Kaur[1], Meenakshi Sharma[2]
[1]*Research Scholar,* [2]*Professor,*
[1]*Department computer science, GIMET Amritsar, Punjab, India*
[2]*Department Computer Science, GIMET Amritsar, Punjab, India*

Abstract: ***In the proposed framework the primary goal is to distinguish community overlapping. The people that are identified with in excess of one community demonstrate overlapping and that gatherings are known as covered gathering. Since in the current work the sink hubs have been considered so the recognition of overlapping moderate. In our present work the association of sink hubs is to be considered through which overlapping has been computed. Clique detection mechanisms are used in order to detect any malicious node within the social networks. But techniques used in existing literature do not consider redundancy of nodes along with presence of sink node. The proposed mechanism considered both redundancy as well as sink nodes hence better result is obtained in terms of community detected and time consumed. Here in distinguishing covered hubs additional time is devoured. In proposed system the quantities of groups identified are more when contrasted with the current one.***

Keywords: community overlapping, redundancy handling, sink nodes, community

I. INTRODUCTION

The universal online social services improve individuals' social exercises with their families, companions and associates, applying an indispensable effect on individuals' life, changing their mindsets and acting. Social media sites including Face-book, Twitter, Wikipedia, Bloggers, MySpace are pulling in more clients than any other time in recent memory. In 2009, the worldwide time spent on social media sites expanded by 82%1 than the prior year. [1]Facebook, a standout amongst the most famous social media sites, has in excess of 500 million dynamic clients and the number is still increasing. The quick increment in social media populace proposes a dynamic social change and potential open doors for social showcasing organizations. In social media websites, clients are permitted to take an interest in social exercises, e.g., interfacing with other similarly invested individuals, refreshing their status, posting online journals, up-stacking photographs, bookmarks and labels etc.[2] Also, individuals can join express groups at various websites. For example, fans of games groups can join committed gatherings where they can impart their insights on group execution, remark on the most current data about player exchanges. Concentrate singular conduct is normally troublesome because of the great degree huge populace and in addition the mannerism of human conduct. [3]Considering insights at site level frequently neglect to get adequate detail. Gathering level examination can furnish valuable data with fluctuating granularity.
Community detection within social networks can be elaborated using the following figure

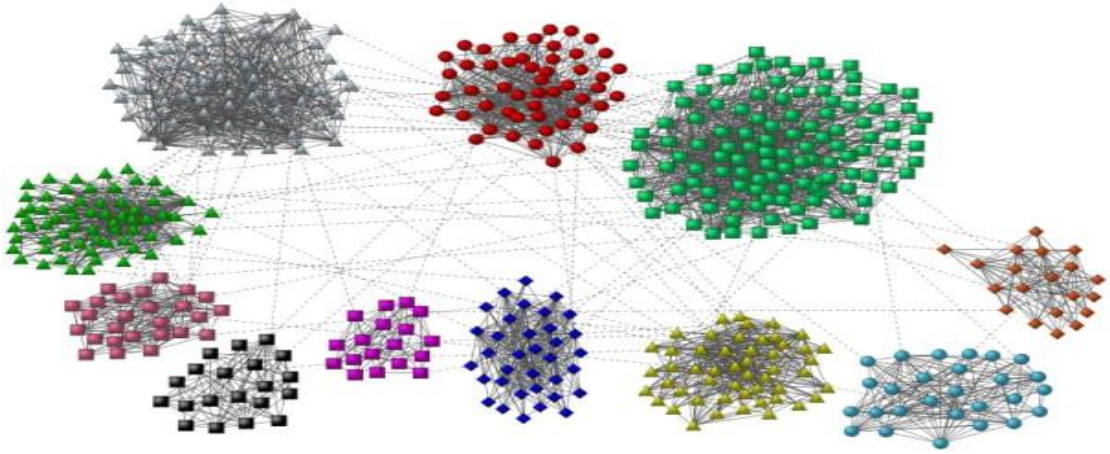


Figure 1: Community detection using clustering

A gathering (or community) can be considered as an arrangement of clients where every client associates more often with clients inside the gathering than with clients outside the gathering. [5]Some social media websites (e.g., Flickr, YouTube) give outside gatherings which enable clients to buy in or go along with them. However, some profoundly powerful sites (e.g., Twitter, Delicious) have no unmistakable gathering structure, which requires quality com-munity identification ways to deal with find them.[6] Community location approaches are normally in light of auxiliary highlights (e.g., joins). Since social media sites give metadata and in addition content data, such data can likewise characterize the performers' social positions.

Community overlapping[7] or measured structure is a standout amongst the most broadly considered points in certifiable social media as it expounds the working of the system. So as to identify overlapping groups from the Facebook arrange we have taken the dataset from Stanford University. System people group speak to fundamental structure for understanding the association of this present reality condition. [8]A community is a gathering of hubs which are associated by some intelligent connections. In a social system it is surely knew that individuals in a social system are normally described by numerous community participations. A man may have connections to numerous dynamic zone including individuals, motion pictures, news, and so on. The majority of the above expressed dynamic things are gathering to which a client may have a place. The structure of community detection is given as under

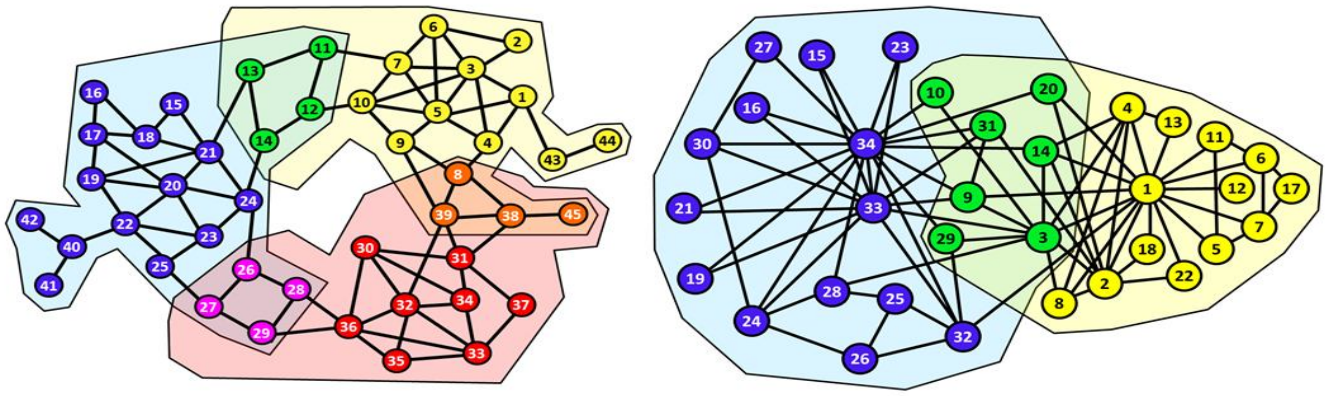


Figure 2: Structure of Community overlapping process

Consequently, there is developing enthusiasm for overlapping community location calculation that recognizes an arrangement of gatherings which are not disjoint. In this paper correlation of the different calculations for overlapping community location has done.

II. LITERATURE SURVEY

G. Palla et.al. proposed this method finds the overlapping cliques by determining strongness of connection between nodes. cliques with same liking are grouped together within common cluster. [10] Kumpula, j. m. et.al. proposed the size of the network is specified and critically used for detection process[11]. Steve Gregory et.al. proposed once identification process is complete next subgraph with k-1 nodes is found out. Connected graph will be utilized to detect the overlapping community. small values of overlapping factor k used to give good result[12].

Fortunato, S et.al. proposed a small value of k also known as threshold value provide base for detection of overlapping community. Complexity of the network greatly depends upon the factor that graph is strongly connected or node. Node strongly connected within the graph with k nodes will have k-1 path associated with it. The nodes with low complexity will be judged on the basis of degree of interrelationship. Nodes does not having any path from source to destination is known as sink nodes. [13] Evans et.al proposed the primary task of proposed approach is to eliminate sink nodes. [14] This makes the overall process of detection easier and less complex.

The nodes detected by the use of overlapping community detection mechanism is represented with the help of entropy. The entropy indicates degree of relationship between pixels.

Chuan Shi et.al. proposed group exposure is an important task in network analysis. A group (also referred to as a clump) is a set of cohesive vertices that have more connections inside the set than outside. In many social and information networks, these communities naturally overlap. The key idea of our algorithm is to find good seeds, and then greedily expand these seeds based on a group metric. Within this seed expansion method, we investigate the problem of how to determine good seed nodes in a graph. The

performance analysis indicates that sink node elimination produces better result overlapping group exposure methods in terms of producing cohesive clumps and identifying ground-truth communities. Social media is expanding day by day. With advancement in technology users can interact with each other through social media[16]. K. L. Wu et.al proposed companies sell their products by publicizing their product on social media. Users and vendors of social media is expanding with the rate of knots. Users are of varying intensions. Some users (consumers) and vendors are malicious in nature causing frauds. Fraud detection is objective of this literature [17].

W.Zhu et.al. proposed detecting frauds on social media is complex. The proposed literature uses modified k means clustering for detection of frauds on social media. Dataset derived from UCI is used for this purpose. Simulation is conducted in MATLAB. Detection results indicate improvement in prediction by 20% [18].

Jierui Xie et.al proposed routing in delay tolerant networks (DTN) adopts the store-carry-forward mode, and it requires nodes to forward data in a cooperative way. However, nodes may be not willing to help others in many applications and this behavior can be called as individual selfish. On the other hand, nodes often can be divided into different communities, and nodes in the same group often have some social ties. Due to these social ties, nodes are more willing to help the one in the same group, but not others.

This behavior can be called as social selfish [19].

Ahn, Y.-Y. et.al proposed social media is providing resources for the user in order to make user interact with each other. It provides mechanism by which users interact with each other without physically moving from one location to other locations. Lack of time is causing more and more users to participate in the social media take over.

The people now days do not have time. They communicate with each other by the use of internet. Internet provides number of mechanisms by which users can communicate with each other. Most common mechanisms which are used involve social media. Social media will help in establishing linkage between the different communities of users. The social media has allowed many users to share their views and also help the users around.

But with the advent of the technology problems also start to appear. The main problem which is caused with the social media is deception. The deception model is then created in order to detect the problems with the online user accounts. Some users can have multiple accounts or some wrong information is provided by them.

This paper describe that the deception is deliberate attempt to mislead the others. The deception will be such that the other user will not able to detect the falsifying information provided by the malicious users. The privacy of the users will be at stake if deception takes place[20].

III. METHOD OF EXPERIMENT

In existing paper to detect community overlapping Louvain method is used to analyze the network in a quick manner. This method is modularity based and performs action on isolated parts of your system. Firstly it finds the smallest community and then in second phase builds the network where each node is the community of first phase.

After that we got a belonging matrix as a outcome from first phase. So the existing work considers both disjoint and overlapping communities. In our proposed paper sink node method is used along with the existing work in order to find community overlapping. The sink node is the node which has no out degree in the graph or no neighboring nodes.

So these type of nodes will not be considered in formulating the community. In our proposed work we will eliminate these nodes in order to achieve optimal results.

A. Algorithm of proposed work

1) Input Dataset
2) Analysis the adjacency matrix
3) If a[i,j]==0 then
4) No community overlapping detected
5) Else
6) Set count =count+1 End if
7) Check for similarity by removing the redundancy
8) Redundancy =redundancy +1 if similar node detect End of if
9) Check for cycle within the graph if exist community overlapping detected

The flowchart for the above listed methodology is given as under

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

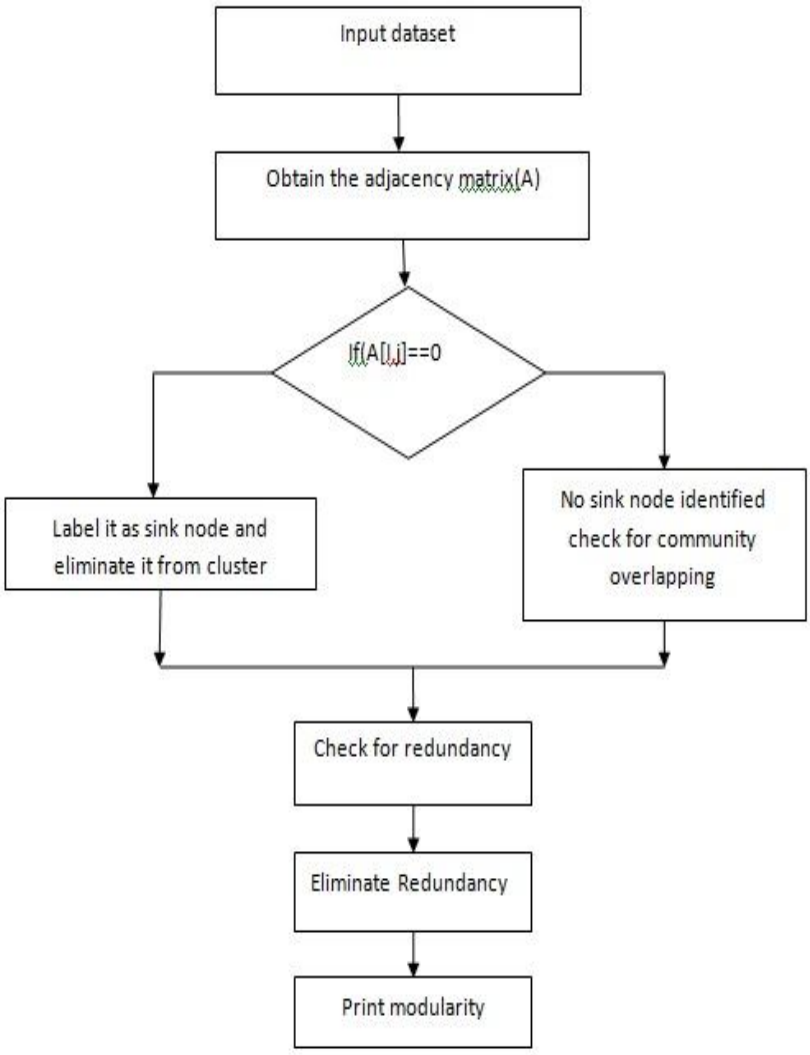


Figure 3: flow diagram of proposed work

B. Performance Evaluation

1) Entropy: Entropy is the relationship of degree between the nodes. In the below table we will consider the entropy of existing work and proposed work.

Table 1: Nodes with base and proposed entropy

No. of Node	Base Entropy	Proposed Entropy
5500	0.5	0.75
7000	0.65	0.86
8500	0.67	0.82
8800	0.73	0.79
9500	0.79	0.83

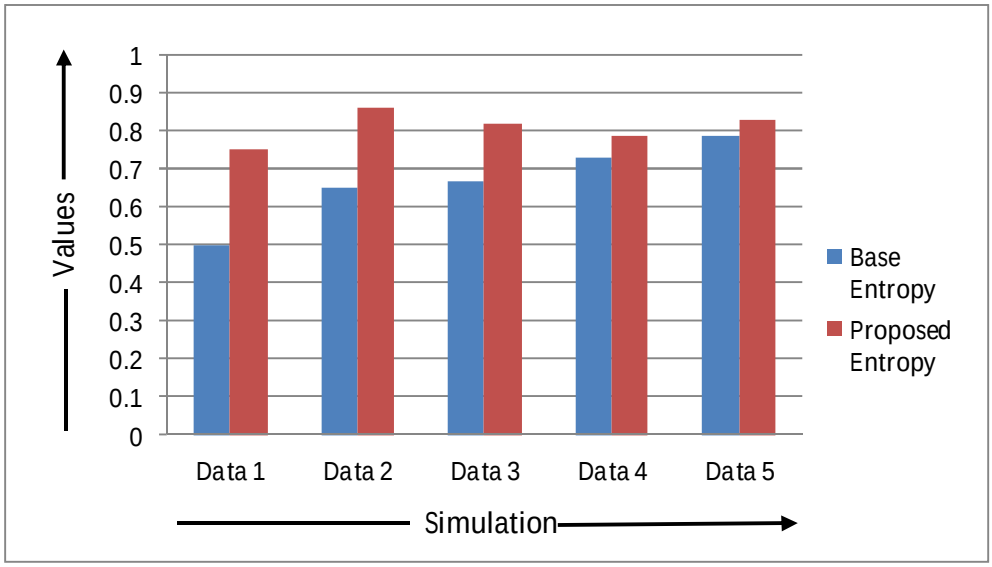


Chart 1. Bar chart of existing and proposed entropy

2) *Modularity:* Modularity defines the number of nodes derived from an existing node. The below table describes the modularity in the existing and proposed paper.

Table 2: Base and proposed modularity

No. of nodes analysed	Base Modularity	Proposed Modularity
5500	0.52	0.93
7000	0.65	0.94
8500	0.67	0.97
8800	0.75	0.91
9500	0.9	0.98

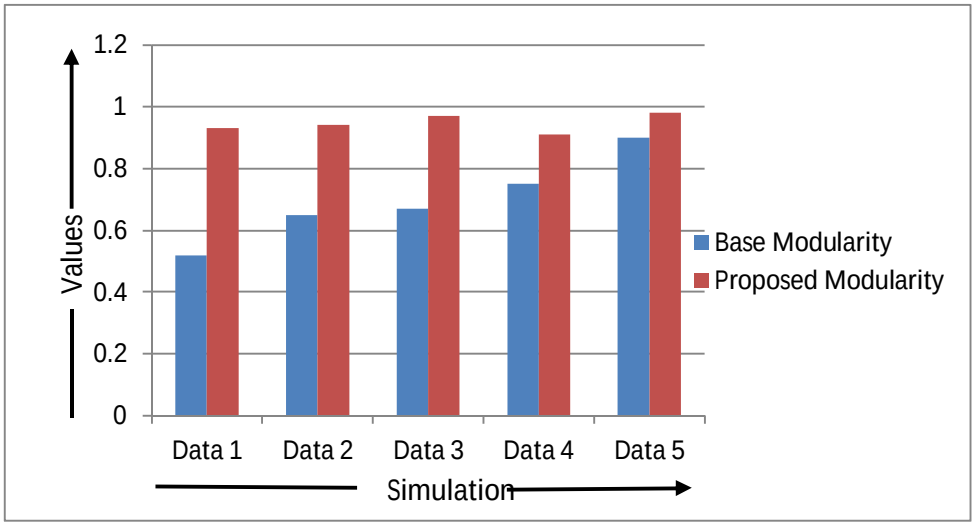


Chart 2.bar chart of base and proposed modularity

IV. DISCUSSION

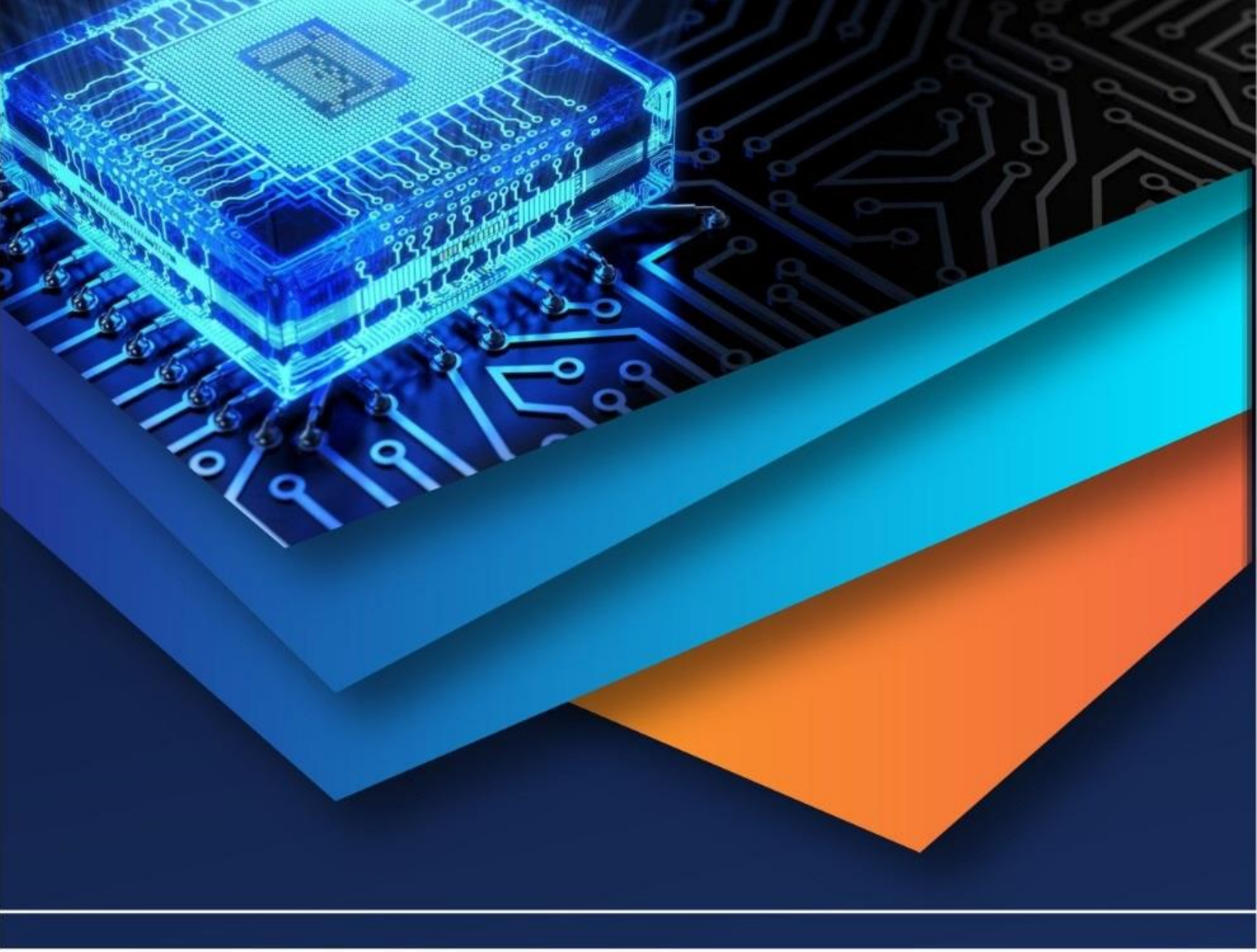
From the above comparison it is clear that the K-Clique algorithm generates faster results as compared to other algorithms. The K-Clique even goes through the nodes that are not required even then it takes less time as compared to other algorithms. It detects 6 communities of 4 clique size in only 10ms while the fuzzy detection algorithm takes 21 ms to detect such groups. Hierarchical algorithm detects only 4 communities in 26 ms. Modularity is also higher in clique algorithm.

V. CONCLUSION AND FUTURE SCOPE

In this paper we have implemented three basic community detection algorithms to detect overlapping communities in the Facebook network. The K-Clique method is one of the simplest methods for the detection of the overlapping community detection. Here fuzzy detection method is also presented which determine the community overlapping detection graphically. The line partitioning method represents the overlapping communities by the use of dendograms. In this paper we have highlighted the methods and also described which method is useful in detecting the cliques.

REFERENCES

[1] M.E.J Newman, The structure and function of complex networks. Volume 2, pp, 167-256, 2003
[2] L. Tang, H. Liu. Community Detection and Mining in Social Media, Morgan & Claypool (2010).
[3] S.P.Chatzis, “A fuzzy c-means-type algorithms for clustering of data with mixed numeric and categorical attributes employing a probabilistic dissimilarity functional,” Expert syst. Appl.., vol. 38, no.7, pp. 8684-8689,2011
[4] Meenakshi sharma and Dr. Himanshu Aggarwal," Development and Implementation challenges in clinical-decision-support system"
[5] S. Ghosh and S. Dubey ,“Comparative analysis of K-means and Fuzzy C-Means Algorithms,” Ijacsa, vol.4, no.4, pp. 35-39,201
[6] W.L. Hung , M.-S yang and D.H Chen, “Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation ,” Pattern Recognit. Lett, vol .27,no 5, pp. 424-438, 2006
[7] Z.ji, Y. Xia, Q. Sun, and G. Cao, “Interval-valued possibilistic fuzzy C-means clustering algorithm,” Fuzzy sets syst, vol. 253, pp. 138-156, 2014
[8] S.A Mingoti and J.O. Lima, “Comparing SOM neural network with Fuzzy C-means, K-means and traditional hierarchical clustering algorithms,” Eur. J.Oper.Res, Vol. 174, pp. 1742-1759, 2006
[9] Meenakshi sharma and Dr. Himanshu Aggarwal," HL-7 Based middleware Standard for Healthcare Information System: FHIR”, ICCCN_2018,NITTTR Chandigarh.
[10] G. Palla, I. Der´enyi, I. Farkas, and T. Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society”,Nature, vol. 435, no. 7043, pp . 814– 818, 2005.
[11] Kumpula, j. m., Kivela, m., Kaski, k., and Saramaki, j. “Sequential algorithm for fast clique percolation.” Phys. Rev. E 78, 2008
[12] Steve Gregory, “Fuzzy overlapping communities in networks”.
[13] Fortunato, S. 2010, “Community detection in graphs.” Phys. Rep. 486, 75–174.
[14] Evans and Lambiotte 2009] Evans, T. S. and Lambiotte, R. “Line graphs, link partitions, and overlapping communities”, Physical Review E, 80 (2009).
[15] Meenakshi Sharma ,Dr. Himanshu Aggarwal," Evaluation factors for testing and validation of Clinical Reporting System", International Journal of Computer Science and Engineering,vol.6,issue-2,e- ISSN:2347-2693 on 2018.
[16] Chuan Shi □, Yanan Cai, Di Fu, Yuxiao Dong, Bin Wu “A link clustering based overlapping community detection algorithm” 2013.
[17] K. L. Wu and M.S. Yang, “Alternative c-means clustering algorithms,” Pattern Recognit.., vol. 35, no. 10, pp. 2267-2278, 2002
[18] W.Zhu, J. Jiang, C.song and L.Bao, “Clustering algorithm based on Fuzzy C-means and Artificial Fish Swarm,” Procedia Eng.., vol. 29, no. 0, pp. 3307-3311, 2012
[19] Jierui Xie, Stephen Kelley, Boleslaw k. Szymanski, “Overlapping Community Detection in Networks: The State-of-the-Art and Comparative Study”.
[20] Ahn, Y.-Y., Bagrow, J. P., And Lehmann, S. 2010, “Link communities reveal multiscale complexity in networks”.Nature 466, 761–764.
[21] Newman, M. E. J., Modularity and community structure in networks, Proceedings of the National Academy of Sciences (PNAS) , 103(23), 8577–8582, 2006.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY