International Journal of Advanced Engineering, Management and Science (IJAEMS) *[Vol-2, Issue-6, June- 2016]*
Infogain Publication (Infogainpublication.com) *ISSN : 2454-1311*

A Modified Weight Balanced Algorithm for Influential Users Community Detection in Online social Network (OSNs)

Bhupinder Singh[1], Manju Bala[2] , Manoj Kumar[3]

[1]Department of Computer Science, CTIEMT, IKG PTU, Jalandhar, Punjab, India
[2]Professor, CTIEMT, Jalandhar, Punjab, India
[3]Principal, Daviet, Jalandhar, Punjab, India

Abstract**—*In the modern era online users are increasing day by day. Different users are using various social networks in different forms. The behavior and attitude of the users of social networking sites varies U2U (User to User). In online social networking users join many groups and communities as per interests and according to the groups'/Communities' influential user. This paper consist of 7 sections , first section emphasis on introduction to the community evelotion and community. Second section signify movement between communities ,third section involve related work about the research.. Fourth section includes Problem Definition and fifth section involve Methodology (Proposed Algorithm Process ,Get Community Matrix, Community detetcion).Sixth section involve Implementation. Furthermore implementation include Datasets ,Quantitative performance, Graphical Results, Enhancement in the existing work..Last section include Conclusion and then references. In this paper,we are implementing and proposing the community detection in social media .In the proposed we have deployed a Longest Chain Subsequence metric for finding the number of connections to the kernel community.*

Keywords—SocialMedia, OSN, Community, U2U, Evolution, Weight Balanced, greedy approach.

I. INTRODUCTION

In online social network community analysis and detection is a vast area of research. Many researchers are working to find the behavior and influence behavior of communities in OSNs'. Classification, detection and evolution of communities in a computational sense, seeks to assign labels to data. Given a set of features for community, a classifier attempts to assign a label to that community. The classifier does this by drawing upon knowledge derived from examples of how other communities have been labeled. These examples, referred to as training data, serve as a source of prior knowledge that the classifier uses to make decisions about previously unseen communities. Categorization is a specialization of classification.

It deals with assigning a category to community. Other classification algorithms may simply make a yes/no decision based upon inputs, such as a fraud detector that indicates whether a credit card transaction is fraudulent. Categorization algorithms place an object in one of a small set of categories, such as categorizing cars as coupes, sedans, SUVs, or vans. Many of the concepts we discuss in this chapter pertain to classification as a whole, whereas others are more closely related to categorization. You'll see some terms used interchangeably. Community categorization in the sense is to find the community and to determine its desired level of surviving.

This new type of data mandates new computational data analysis approaches that can combine social theories with statistical and data mining methods[14]. The pressing demand for new techniques ushers in and entails a new Inter disciplinary field – social media mining. Thus we have studied the various techniques for community evolution in the online social media and we are discussing existing methods in this paper.

II. MOVEMENT BETWEEN COMMUNITIES

Having analyzed the membership and growth of communities, we have a tendency to currently intercommunicate the question of however folks and topics move between communities. A elementary question here is that the degree to which individuals bring topics with them from one community to a different, versus the degree to that topics arise in a very community and afterward attract folks from alternative communities. In alternative words, given a collection of overlapping communities, do topics tend to follow folks, or do folks tend to follow topics? we have a tendency to conjointly investigate a connected question: once folks get into a community square measure they additional or less doubtless than alternative members of the community to be participants in current and future "hot topics" of dialogue therein community?. While these queries square measure intuitively terribly natural, it's a challenge to outline sufficiently precise versions of them that we are able to build quantitative observations.

III. RELATED WORK

A substantial quantity of hardwork has been dedicated to the task of distinctive and evaluating closely knit communities in giant social networks, most of that is predicated on the premise that it's a matter of common expertise that communities exist in these networks. Akrati Saxena et al. [1] analysed the impact of the topology of a social network, specifically its meso scale properties- community structure and core-periphery structure, on a acculturation traversing over it. we have a tendency to propose a acculturation propagation model for artificial scale free graphs .David Burth et al. [2] represented however process researches have approached this subject and therefore the strategies accustomed analyse such systems. supported on a good tho' non-exaustive review of the literature, a taxonomy is projected to classify and describe totally different classes of analysis.

Chang Su et al. [3] projected a community detection rule supported influential nodes. First, we tend to introduce a way to notice cogent nodes supported stochastic process. Then we tend to mix the rule with order statistics theory to seek out community structure. we tend to apply our rule in 3 classical information sets and compare to alternative algorithms. Anna Zygmunt et al. [4] conferred a proper model of social structure with evolving groups. a brand new rule decisive a life cycle of social groups and their cores is conferred. supported the knowledge derived from them, one will - victimization Social Network Analysis strategies - build a network of connections between users, that then may be analysed to seek out influential users and groups that square measure shaped around them.

Weishu Hu et al. [5] outlined and study a completely unique community detection drawback that's to find the hidden community structure in giant social networks supported their common interests. we tend to observe that the users usually pay a lot of attention to those users who share similar interests, that change how to partition the users into totally different communities per their common interests. Liaoruo Wang et al. [6] define and explore a novel problem called community kernel detection in order to uncover the hidden community structure in large social networks. They discover that influential users pay closer attention to those who are moresimilar to them.

Beiming Sun et al. [7] described that abundant endeavor has been conducted to investigate info from social networks, as well as finding the influential users. during this paper, we tend to propose a graph model to represent the relationships between on-line posts of 1 topic, so as to spot the influential users . Michael Trusov et al. [8] mentioned that the success of web social networking sites depends on the amount and activity levels of their user members. though users generally have varied connections to alternative web site members (i.e., “friends”), solely a fraction of these alleged friends may very well influence a member’s web site usage. Emilio Ferrara [9] involved with the analysis of aggregation patterns and social dynamics occurring among users of the most important OSN because the date: Facebook. In detail, we tend to discuss the mesoscopic options of the community structure of this network, considering the angle of the communities, that has not nevertheless been studied on such an oversized scale.

Suhani Grabowicz P et al. [10] mentioned that in trendy days social network services are popularly utilized in electronic commerce systems. so as to market the business, it's fascinating to explore hidden relationships among users developed on the social network. during this paper, we tend to outline and study a completely unique community detection downside that's to get the hidden community structure in massive social networks supported their common interests. David R Hunter et al. [11] present a scientific examination of a true network information set victimization most chance estimation for exponential random graph models moreover as new procedures to judge however well the models match the determined networks.

Salvatore Catanese et al. [12] given access to information regarding Facebook users and their friendly relationship relations. For this they noninheritable the mandatory info directly from the side of the web site, so as to reconstruct a sub graph representing anonymous interconnections among a big set of users. Author delineate impromptu, privacy-compliant crawler for Facebook information extraction. Pasquale American state Meo et al. [13] explained to work out whether or not 2 distinct users may be thought of similar. during this the authors planned AN approach so as to estimate the similarity of 2 users supported the information of social ties (i.e., common friends and teams of users) existing among users, and therefore the analysis of activities (i.e., social events) during which users area unit concerned.

IV. PROBLEM DEFINITION

The basic structure of social area network in community detection is categorized in two tasks. Initial step begins with the identification and extraction of influential kernel members. The second step is to identify the corresponding structure of the community kernels. The problem arises in detection of the exact structure among the influential kernel members and auxiliary communities. Most of the conventional algorithms ignore the link among the auxiliary communities and community kernels, resulting in failure to distinguish among them. The overlapping of the communities drops the actual structure of the community. Although, in order to distinguish the kernel members, the Page rank and conductance based algorithm are designed that utilizes the link information. However, the link information among the auxiliary and kernel members are ignored finding the community kernels. Most importantly, the growing vertices of the network must be addressed developing robust algorithm with large scalability.

The problem of community detection has many smart applications, yet as representative user finding, friend recommendation, network visual image, and marketing. However, this drawback is non-trivial and poses a group of challenges. First, it's hard to identify the extremely potent users.

V. METHODOLGY

Proposed Algorithm Process

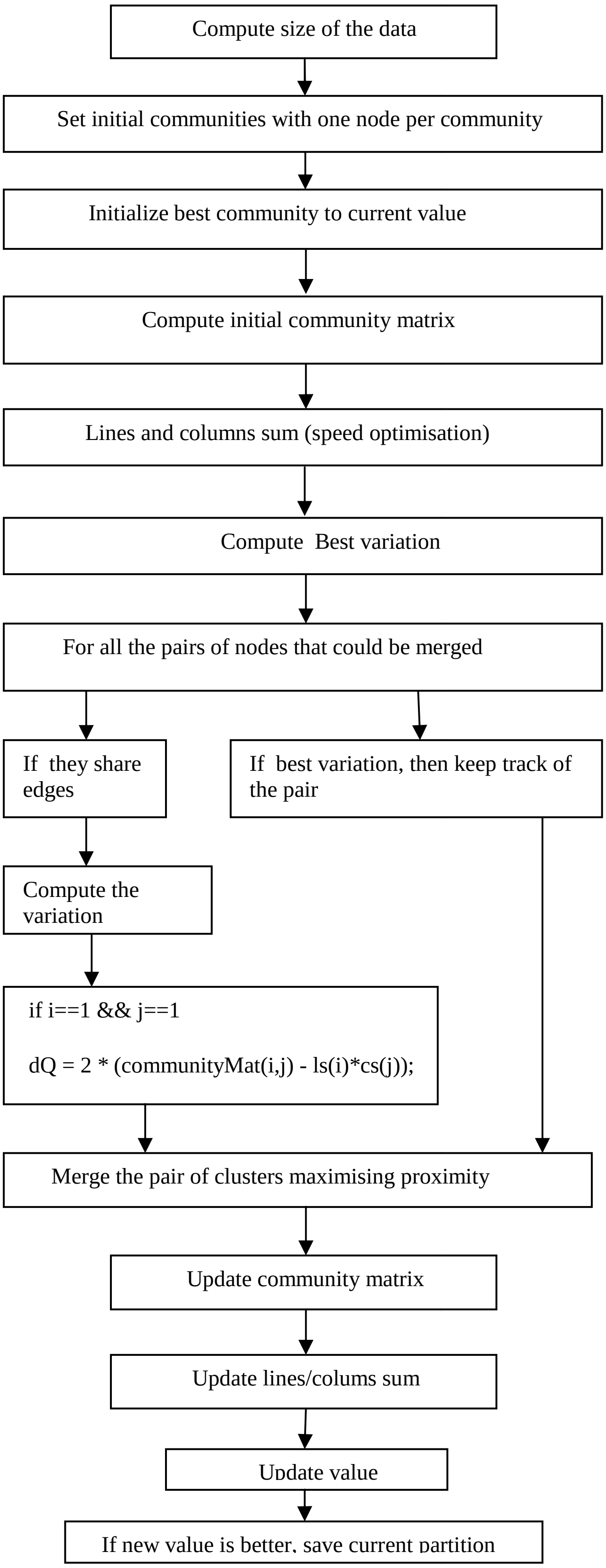

Fig 1: shows flowchart of Proposed Algorithm Process

Get Community Matrix

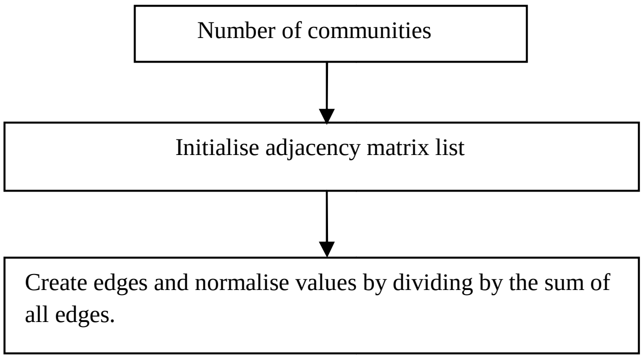

Fig 2: shows flowchart of Adjacent Matrix

Community Detection

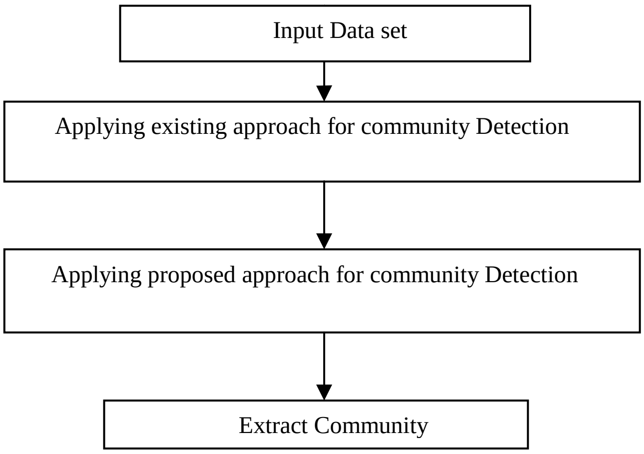

Fig 3: shows flowchart of community detection

VI. IMPLEMENTATION

Datasets[15] :- Our experiments are conducted on two facebook datasets. Facebook data was collected from survey participants using third party facebook app such as social circles. The dataset includes node features (profiles), circles, and ego networks.

Quantitative performance. We use Precision, Recall, and F1-score to evaluate existing techniques with our proposed Algorithm Technique.

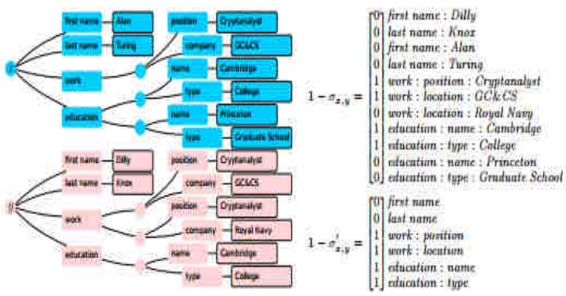

Fig 4: Feature construction[6]

We need to compute the size of data. Set initial communities with one node per community. Initialise best community to current value. Compute initial community

matrix. Lines and columns sum (speed optimisation) . Initialise best known and current Q values.
curQVal=trace(communityMat)-
sum(sum(communityMat^2));
Loop until no more aggregation is possible.Compute Best Q variation. For all the pairs of nodes that could be merged. If they share edges,then Compute the variation in Q. If best variation, then keep track of the pair. Then Merge the pair of clusters maximising Q. Update community matrix. Update lines/colums sum. Further Update the value. If new Q is better, save current partition.

Graphical Results:-

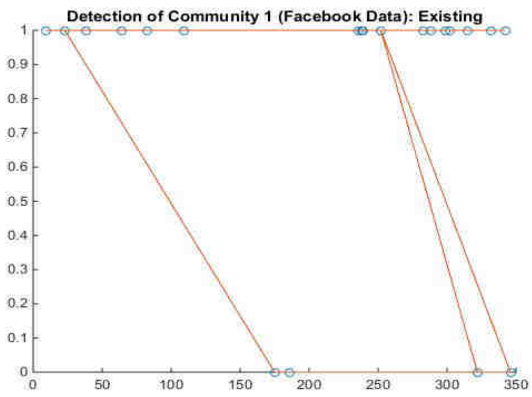

Fig. 5: Detection of Community 1 (Facebook Data) : Existing

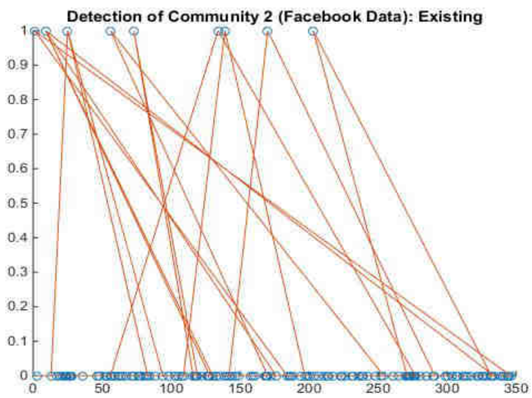

Fig. 5: Detection of Community 2 (Facebook Data) : Existing

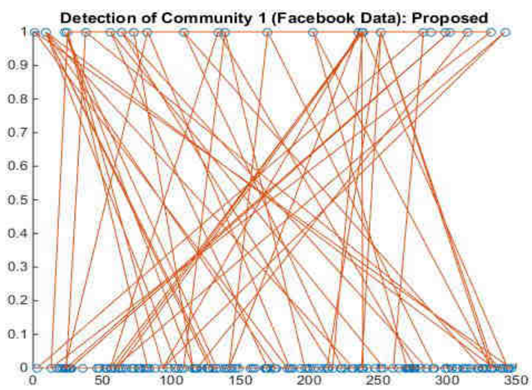

Fig. 6: Detection of Community (Facebook Data) : Proposed

Enhancements in Existing Work

The existing method works on two algorithms Greedy Algorithm and Weight-Balanced Algorithm (W_EBA). The greedy algorithm works on finding the maximum distance from a given set of vertices to other connections. The process is repeated for every vertex from the vertex set in the graph. The vertex having maximum number of degree is selected as community head. However, the greedy algorithm does not consider the link among auxiliary members and kernel members. In the W_EBA algorithm, a weight measure is associated with each kernel community followed by optimization for finding the exact kernel structure and auxiliary kernels. However, in such algorithm, the small communities' having less number of connections has a lower priority value and never gets identified.

In the proposed we have deployed a Longest Chain Subsequence metric for finding the number of connections to the kernel community. The method performs grouping the communities based on LCSS metric. Despite of considering only the maximum value, the algorithm considers the new connections as well as the overlapped connections for a particular community. This results in better identification of the kernel communities as well as the auxiliary communities intact.

The proposed method performs the distance measure with the Longest Chain Sub Sequences (LCSS) metric defined as.

$$d_{ij} = LCS(X_i, Y_j) = \begin{cases} \phi & if i = 0 and j = 0 \\ LCS(X_{i-1}, Y_{j-1}) & if x_i = y_j \\ \max(LCS(X_i, Y_{j-1}), LCS(X_{i-1}, Y_j)) & if x_i \neq y_j \end{cases}$$

Fig. 7: show Longest Chain Sub Sequences (LCSS) metric

VI. CONCLUSION

Now it s clear that communities in OSN are playing a vital role and the users of communities are changing their role as per community behavior. If the community is demising then its directly meaning that its users are leaving that particular community and if the users are joining the community then it is clear that particular community will grow in future. Various methods are user to find the community structure and the analytical behavior. In future work can be extended to find the spatial information of the user as well as to the community in social network. For future work, we would like to explore the dynamic behavior of community kernels and their auxiliary communities.We are interested in how community kernels take shape and evolve over time. In addition, we would like to combine link and content information in our problem definition and algorithm design for practical applications, such as query dependent community detection. Inspite of assuming the utmost price, the algorithmic program considers the new connections yet because the overlapped connections for a selected community. This ends up in higher identification of the kernel communities yet because the auxiliary communities intact.

REFERENCES

[1] Akrati Saxena,S.R.S.Iyengar,YayatiGupta, "Understanding Spreading Patterns on Social Networks Based on Network Topology", arXiv:1505.00457v1 [cs.SI] , May 2015

[2] David Burth Kurka, Alan Godoy, Fernando J. Von Zuben, "Online Social Network Analysis: A Survey of Research Applications in Computer Science", arXiv:1504.05655v1 [cs.SI] , Apr 2015

[3] Chang Su, Yukun Wang , Lan Zhang "Community Detection in Social Networks Based on Influential Nodes ", Journal of Software, VOL. 9, NO. 9, September 2014

[4] Anna Zygmunt, Jarosław Ko´zlak, Edward Nawarecki, Adam Mika, "Determining life cycles of evolving groups", AGH University of Science and Technology, Al. Mickiewicza VOL.35, PP.1102-1111 Krak´ow, Poland August ,2014

[5] Weishu Hu, Zhiguo Gong, Leong Hou U and Jingzhi Guo, "Identifying influential user communities on the social network ",Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macao, China, 2013 Taylor & Francis.

[6] Liaoruo Wang, Tiancheng Lou, Jie Tang and John E. Hopcroft, "Detecting Community Kernels in Large Social Networks", ICDM '11 Proceedings of the 2011 IEEE 11th International Conference on Data Mining PP. 784-793 ISBN: 978-0-7695-4408-3

[7] Beiming Sun, Vincent TY Ng, "Identifying Influential Users by Their Postings in Social Networks", MSM'12, June 25, 2012, Milwaukee, Wisconsin, USA. Copyright 2012 ACM 978-1-4503-1402-2/12/06

[8] Michael Trusov, Anand V.Bodapati, and Randolph E.Bulkin, "Determining Influential Users in Internet Social Networks", Journal of Marketing Research VOL. XLVII (August 2010), 643–658

[9] Emilio Ferrara, "A large-scale community structure analysis in Facebook",Ferrara EPJ Data Science 2012, 1:9

[10] Suhani Grabowicz P, Ramasco J, José J. Ramasco ,Esteban Moro,Josep M. Pujol, Victor M. Eguiluz, "Social features of online networks: the strength of intermediary ties in online social media", PLoS ONE 7(1): e29358.doi:10.1371/journal.pone.0029358,2012

[11] Hunter D, Goodreau S, Handcock M, "Goodness of fit of social network models", J Am Stat Assoc 103(481):248-258, 2008.

[12] Salvatore A. Catanese, Pasquale De Meo , Emilio Ferrara, Giacomo Fiumara, Alessandro Provetti,"Crawling Facebook for Social Network Analysis Purposes", WIMS'11, May 25-27, 2011 Sogndal, Norway

[13] Pasquale De Meo a, Emilio Ferrara b,a, Giacomo Fiumara a, Alessandro Provetti a,c, "Mixing local and global information for community detection in large networks,", Journal of Computer and System Sciences VOL. 80 (1), PP.72-87

[14] Manju Bala, Manoj Kumar, Bhupinder Singh, "Socia Media Evolution and Detection of influential Community Users in online Social Networks (OSNs)", ICACSET 2016 - Singapore on April 14 - 15, 2016

[15] https://snap.stanford.edu/data/