International Journal of Scientific Research & Engineering Trends
Volume 4, Issue 5, Sept-Oct-2018, ISSN (Online): 2395-566X

A Survey on Online Social Network Based User Community Identification Techniques and Features

Asst.Prof. Raju Sharma
Dept. of Computer Science
CIST Bhopal, India
sharma.raju696@gmail.com

***Abstract* – As the amount of web customers are extending each day. Presently a user connecting the social personality over the distinctive web-based social networking stages is of basic significance to business intelligence. In this paper, a survey of similarity based user community detection methodologies were discussed for finding the exceptional arrangement of users. Here different features of the social user profiles were identified as per there requirement. So current issues are summarized in the paper for the solution of the work.**

***Keywords*— User Identification, Cross-Media Analysis, Genetic Algorithm, Clustering**.

I. INTRODUCTION

Digital Social communities have moved toward becoming piece of our regular day to day existence. It is presently typical that individuals have accounts in numerous Social communities, sharing their considerations, advancing their work and most likely impacting a piece of the populace by means of them [1,7]. An assortment of functionalities are given by these administrations, for example, video and photograph transferring, posting, informing, republishing and so on,contrasting as indicated by the stage and its point. Persuaded by the need to check the legitimacy and reliability of data that shows up on Social organizations, the nearness of people in various systems can be demonstrated especially valuable

Open data from one system can be utilized to approve the wellspring of data in another system. To accomplish this objective, there is a requirement for user recognizable proof crosswise over Social communities. As same user available in many platform may with same name or not but there community is always be same in all networks as shown in fig.1.

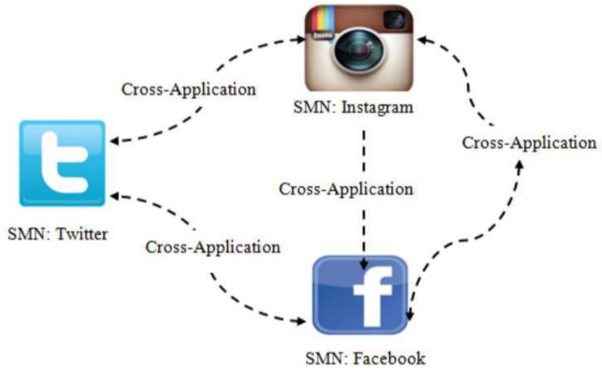

Fig. 1 Cross-application research to merge a variety of SMN’s.

Since there are many users sharing their opinions and experiences via social media, there is aggregation of personal wisdom and different viewpoints. Such aggregation has limitations as viewpoints are subject to change with time. In a sense the social media prediction problem is paralleled by prediction of financial time series based on past history, which has its uses in trading.

In general, if extracted and analyzed properly, the data on social media can lead to useful predictions of certain human related events. Such prediction has great benefits in many realms, such as finance, product marketing and politics, which has attracted increasing number of researchers to this subject. Study of social media also provides insights on social dynamics and public health. A survey provides us perspective and is helpful for carrying out further research.

II. COMMUNITY DETECTION

Community discovery process is related with fundamental concept of data clustering as it clusters set of vertices while latter is used for clustering data points. There are two required properties of community which must need to be satisfied by sub graph:

- Edge Density There must be maximum no. of edges shared within sub graph while less no. of edges outside of it.
- Connectedness Path between pair of vertices of sub graph must run only through vertices of sub graph. Any sub graph which satisfies above conditions is considered as community shown in Fig. 2. There are various types of community defined on the basis of the different parameters.

There is no exact categorization of communities are possible according to above parameters still it’s significance affects performance of community detection process and network considered for

© 2018 IJSRET

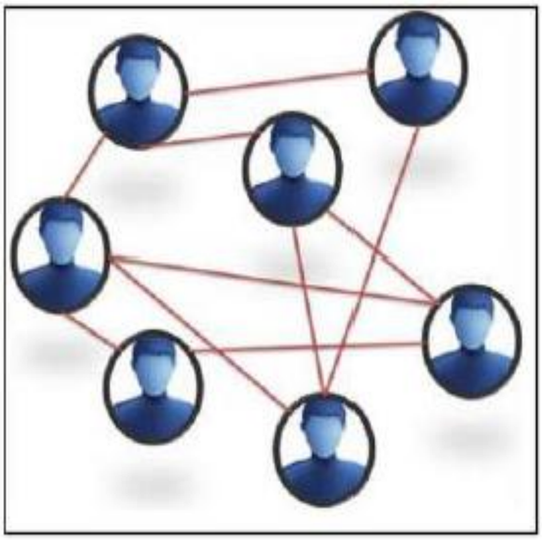

Fig. 2 Sample Network.

Community detection. Community criteria are defined to decide type of community which has been detected given as follows:

1. Community criteria Criterion parameters [2] [4] are defined for induced sub graph which decides whether sub graph is "Community" or not. Local and Global community relies on different set of criteria which are discussed in the following section. Provided set of parameters also helps to decide stopping condition for "Community Discovery" process.

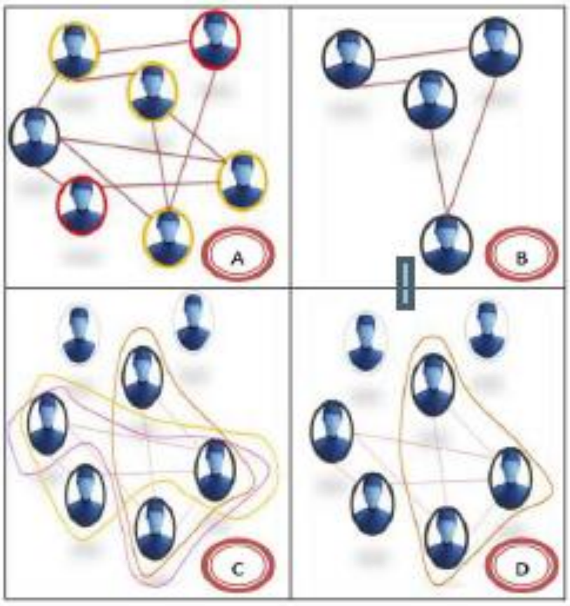

Fig .3 Different types of communities detected in Sample network using different approach: a) Vertex similarity b) Global community c) Local community (Overlapping) d)Local community (Non overlapping).

2. Local community

Local community can be derived from incomplete part of Whole graph of real world network as shown in Fig.3 Neighborhood vertices of initial source vertex plays important Role in community formation. Different degree distribution of vertices of graph leads to identify set of valid vertices from neighborhood region which satisfies required properties of community. For the same purpose following properties are devised:

3.Global community

Global community considers community structure as a property of the whole graph shown in Fig. 3. It is stated in previous section that it is necessary to have complete knowledge of whole graph. due to following reasons:

- In many real world networks, sub graph of original graph comprises basic structural properties of whole graph. Only such sub graph can be identified as "Global Community" which is basically prototype model of original whole graph.
- If complete graph is unavailable, it can lead to generate false positives for com- munity which does not satisfy basic properties

III. RELATED WORK

Palsetiay et al. in [3], researcher analyze the issue of element determination crosswise over Twitter and Instagram utilizing general methods. Our techniques fall into three classifications: profile, substance, and chart based. For the profile based strategies, researcher consider systems in view of rough string matching . For content-based strategies, researcher perform creator ID. At long last, for diagram based strategies, researcher apply novel cross domain group location techniques and produce neighborhood-based features. The three classifications of techniques are connected to a vast chart of users in Twitter and Instagram to comprehend challenges, decide execution, and comprehend combination of various strategies. Last outcomes exhibit an equivalent mistake rate under 1%.

Chen et. al [4] proposed agglomerative clustering algorithm along with Max-min modularity quality measure. Proposed algorithm considers both topology of network and provided domain knowledge e.g. unrelated pair of vertices while community detection. Defined community measure considers absent links within vertices of same community which helps to identify sub graphs with number of edges within sub graphs and unrelated pairs between sub graphs having higher value than expected one.Vertex similarity approach is proposed in Dang et al. [5] to partition graph into network considering structural and attribute equivalence. Modularity optimization scheme is used to define cluster where modularity measure consider both link strength and attribute equivalence factor. K-nearest neighbor based approach is used to agglomerate vertex into community.

Karsten Steinhae user et. Al [6] proposes methodology which considers structural and attribute equivalence while Calculating vertex similarity. Structural equivalence evaluated on the basis of degree of node, participation in triangle community and total no. of common neighbourhoods while attribute equivalence is based on value of attributes.

Evaluated score are assigned as edge weight. Comprising use of attribute information of node for deciding vertex similarity allows efficient detection of community with even available little information about topology. Predefined threshold value is used to define eligibility of edge in community.

Xiao-Li Li et. al [7] proposes technique to detect community of people for future events based on available knowledge of previously attended events and topology of social network. Weighted graph is generated with inclusion of virtual links to depict relationship between people with similar interest but unknown to the each other. Vertex similarity approach is used to cluster event and to identify set of candidate people for new event.

Shuai Zhao et. al. in [8] researcher propose an overlapping community detection method, namely, latent Dirichlet allocation (LDA)-based link partition (LBLP), which uses a graphical model and considers network structure and content information.

Two feature integration strategies are proposed to combine the influence of network structure and content information on the network generation process. Experimental results on synthetic and real-world networks show that the LBLP method is effective, and content information is beneficial in mining community structure.

Table 1. As given Technique and limitations of work.

Paper	Techniques	Advantages	Limitations
In [12] Mohsin Hasan Hussein et. al	Friend of a friend (FOAF) to identify the communities of friends using PAM (Partitioning Around Medoid) clustering algorithm	The outlier nodes have been reduced after they had constituted a high percentage of nodes before using the proposed method	Only single social network was used for allotting a user to a community
In [13] Faliang Huang et. al	Meta-heuristic approach that combines together line graph theory, ensemble learning and particle swarm optimization techniques for overlapping communities detection	Efficiently detect overlapping community by PSO	Detection and elimation of fake user was not done in preprocessing of dataset
In [14] Longqi Yang et. al.	Proposed a method called CoNMTF (Coregularized Nonnegative Matrix Tri-Factorization	Large number of communities were identified by using less features	Proposed work was complex to implement with large dataset as execution time increases with nodes.
In[15]	Neighbour hood Of nodes	This study focuses on intelligent choice of initial centers of the clusters and has introduced an algorithm without parameters to detect graph clusters.	Work do not consider user inherent features for the classification

IV. COMMUNITY DETECTION METHODS

1. Categories of community detection methods

community detection algorithms are decided on the basis of following factors:

- Vertex similarity Methods are focuses on clustering vertices which have more similar feature.
- Edge density Methods are focuses on identifying clusters which improves intra cluster edge density rather than inter cluster edge density.

- Distance between vertices Methods are focuses on clustering vertices which are nearer to each other. Basic principle of clustering data points using data clustering algorithms is considered as base for same purpose.

2.Clique guided community detection- A new approach developed for fast and efficient community detection. Clique guidedcommunity detection consists of two phases. In the first phase, the framework finds disjoint cliques. In the second phase, the cliques from the first phase are used to guide the merging of individual vertices until a good quality solution is obtained. For the first phase, we develop an algorithm named

3.MACH (Maximum Clique Heuristic)- which is a new approach to compute disjoint cliques using a heuristic-based branch-and-bound technique. The experimental results are provided to demonstrate the efficiency of the new algorithm and compare the approach with other previously proposed algorithms. As the framework is adopted the community merge step for the proposed paper takes O(k) time, where k is the number of communities. If the merging is very unbalanced in this phase, it could perform O(n) merges, taking up to O(n2) time for this phase.[14]

4.**Graph Partition method-** A graph partition method based on min-max clustering principle was proposed by Ding and Zha et al. The principle states that the similarity or association between two sub graphs is minimized, while the similarity or association within each sub graph is maximized. Luo and Wang et al. proposed a framework to identify modules within a biological network. Networks are divided into sub networks and the identification of modules is based on their topology.

5.Community Detection method Using DBSCAN Algorithm:

A community detection methods using DBSCAN algorithm was proposed, which is the most effective unsupervised clustering algorithm. The DBSCAN algorithm can identify clusters in large spatial data sets by looking at the local density of database elements, using only one input parameter. Furthermore, the user gets a suggestion on which parameter value that would be suitable. The DBSCAN can also determine what information should be classified as noise or outliers. In spite of this, its working process is quick and scales very well with the size of the database-almost linearly.

From the graphical representation structure of social network, the interactions or connection between individuals or entities, or nodes can be viewed, from which the existence of communities can be concluded. In this approach detection of communities was done on the basis of three types of members in the community, which are core, border and outlier members, and which are of high, low and no influence respectively.

V. FEATURES PROBLEM IDENTIFICATION OF COMMUNITY

Some of the important features [1] of the communities are considered to be as follows:

- Overlapping: communities can overlap in which users share the same interests and have the same edges in common between two or more communities.
- Directed: edges within a community can be directed or undirected. In terms of social networks, we can consider all of the edges to be directed.
- Weighted: edges in the communities can be weighted to denote that various users have different affiliations and interaction rate with the community. The more influence a node brings up to a community, the greater is its edge weight connecting the node to the community.
- Multi−dimensional: interactions within a community can be multi-dimensional, meaning that people can use various methods to interact with each other by posting, sharing, liking, commenting, tagging, etc.
- Incremental: communities and community detection algorithms are expected to be incremental in which adding a new node and assigning a community to it, would just need a local search for the node in its neighborhood. We definitely dont want to run the whole algorithm from the beginning just for finding a community for a newcomer node.
- Dynamic: communities can be dynamic and evolvethrough the time. Because most of giant social networks are dynamic, many of researchers have proposed the idea of streaming graph partitioning [15] which can be done using distributed computations and are mostly known as one-pass algorithms. In one-pass algorithms each node is assigned to a partition upon arrival in a greedy manner such that the objective function of the partition in graph is maximized.

Community detection approaches in the literature mostly rely solely on the link analysis and ignore the available information in the modern social networks. As an example, Twitter entails lots of metadata like user location, age, gender, geotags interests all of which can be used in clustering.

1. Problem Identification

In [8] only link feature was used by the work, where other inherent feature like social activity was ignored which tent to reduce the accuracy of the prediction community. Here single social network involved in the work which can be improved by introducing similar other kind of network. Learning of vast use network need some dynamic approach for predicting the class of the user, as variety of user is present in network and there change in community is also random.

VI. EVALUATION PARAMETERS

As various techniques evolve different steps of working for classifying users into appropriate community. So it is highly required that proposed techniques or existing work need to be compare on same dataset. So following are some of the evaluation formula shown in equation number which help to judge the community techniques. Precision = (True_positive (False_positive+ True positive)) Recall = (True_positive (False negative+ True positive)) F-Measure = (2xPrecisionxRecall/ (Recall + Precision))

This work adopts NDCG [6, 12] as the performance evaluation measure. The NDCG measure is computed as

$$NDCG \quad @ \; P = Z_P \sum_{i=1}^{P} \frac{2^{l(i)} - 1}{\log(\, i + 1)}$$

Where *P* is the considered depth, *l(i)* is the relevance level of the *i*-th image and *ZP* is a normalization constant that is chosen to let the optimal ranking's NDCG score to be 1.

VII. CONCLUSIONS

Connecting user identities across social media sites is not a straightforward task. The primary obstacle is that connectivity among user identities across different sites is often unavailable. The existing approaches are illustrated with the main focus input parameters which are used while performing community detection. Type of network used as input for community detection affects the use of resulting community outputs e.g. target advertising, detecting information flow in social network etc.

VIII.FUTURE WORK

In future one can opt other feature combination with encryption for data security as well. Here feature of there schooling, locality, professional can be utilize. Most of work done in previous work focus on graphical based prediction but one can consider social activity between user as well like sharing messages, images, etc. Since social network is never be stable so some dynamic algorithm should be suggest which will absorb all these random changes.

REFERENCES

[1]. Reza Zafarani and Huan Liu." Connecting Users across Social Media Sites: A Behavioral-Modeling Approach" KDD'13, Chicago, Illinois, USA. Copyright ACM 978-1-4503-2174-7/13/08 August 11–14, 2013.

[2]. Denzil Correa et al. Top: interaction based topic centric community discovery on twitter,PIKM '12 Pages 51- 58,ACM 2012

[3]. Palsetiay et al., "User-interest based community extraction in social networks" ,SNAKDD'12, ACM 2012.

[4]. Jiyang Chen et al., Detecting Communities in Social Networks using Max-Min Modularity SDM, page 978- 989. SIAM 20.

[5]. Xiao-Li Li et. al, ECODE: Event-Based Community Detection from Social Networks DASFAA'11, Springer p22-37

[6]. The Anh Dang and Emmanuel Viennet, Community Detection based on Structural and Attribute Similarities ICDS'12, IARIA

[7].Karsten Steinhaeuser and Nitesh V. Chawla, Community Detection in a Large Real-World Social Network, Social Computing, Behavioral Modeling, and Prediction Springer 2008, pp 168-175

[8].Shuai Zhao1, Le Yu 2, and Bo Cheng. "Probabilistic Community using Link and Content for Social Networks". IEEE., Digital Object Identifier 10.1109/ACCESS.2017

[9].Hyun-Kyo Oh, Sang-Wook Kim, Member, IEEE, Sunju Park, And Ming Zhou." Can You Trust Online Ratings? A Mutual Reinforcement Model For Trustworthy Online Rating Systems.Ieee Transactions On Systems, Man, And Cybernetics: Systems, Vol. 45, No. 12, December 2015

[10].Jiawei Zhang and S Yu Philip. Multiple anonymized social networks alignment. Network, 3(3):6, 2015.

[11].Yutao Zhang, Jie Tang, Zhilin Yang, Jian Pei, and Philip S Yu. Cosnet: Connecting heterogeneous social networks with local and global consistency. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages, 1485–1494. ACM, 2015.

[12].Mohsin Hasan Hussein, Huda Naji Nawaf , Wesam S Bhaya. "Exploiting the Shared Neighborhood to Improve the Quality of Social CommunityDetection". Annual Conference on New Trends in Information & Communications Technology Applications-(NTICT'2017). 7 - 9 March 2017 978-1-5386-2962-8/17/$31.00 ©2017 IEEE

[13].Faliang Huang, Xuelong Li, Fellow, IEEE, Shichao Zhang, Senior Member, IEEE, Jilian Zhang, Jinhui Chen and Zhinian Zhai. Overlapping Community Detection for Multimedia Social Networks". 1520-9210 (c) 2016 IEEE.

[14].Longqi Yang, Liangliang Zhang, Zhisong Pan, Guyu Hu, Yanyan Zhang. "Community Detection Based on Co-regularized Nonnegative Matrix Tri-Factorization in Multi-view Social Networks".2018 IEEE International Conference on Big Data and Smart Computing.

[15].Binazir Balegh, Saeed Farzi. "Community Detection in Social Network Using a Novel Algorithm Without Parameter". IEEE 4th International Conference on Knowledge-Based Engineering and Innovation [6] (KBEI) Dec. 22, 2017