A Comparative Study of Various Frameworks for Community Detection in Dynamic Social Network

Mr. Atul S Choudhary
Assistant Professor, Dept. of Information Technology
MIT Academy of Engineering, Alandi (D)
Pune, India

Mr. Sunil S Mhamane
Assistant Professor, Dept. of Information Technology
MIT Academy of Engineering, Alandi (D)
Pune, India

Abstract—**Social Networks are extensively used in today's realm. It has engrossed huge amount of people as it is describes the set of people and their relationship with each other. People are considered as entities in the expanse of social networks and the relationship between these entities is preserved by the interaction between them. Social network has its application in many spheres and business areas it has become one of the key areas of interest for the researchers. Community detection is one of the most significant and stimulating research areas in social network analysis. As the activities and interaction between the entities change over time, the speed with which the network is changing is phenomenal. Because of this reckless and frequent change it has become very important to detect the community in dynamic social network. Many methods have been proposed for the community detection in dynamic social network. This paper presents a detail study on various community detection methods and a comparison between them to show the effectiveness of any methods.**

Keywords—Community, Community Structure, Social Network

I. INTRODUCTION

Recently, the use of internet empowered devices has significantly increased. Huge volume of people now a day does depend on such devices for their businesses and personal work. Bulk quantity of data is shared amongst different nodes via internet in the form of social networking, which made the social networking issues very prevalent and has enthralled researchers for solving those issues. There are numerous of ways to define a social network. Barker defined social network as "Individuals or groups linked by some common bond, shared social status, similar or shared functions, or geographic or cultural connection. Social networks form and discontinue on an ad hoc basis depending on specific need and interest." [R. Barker, The social work dictionary. NASW press Washington, DC, 2003.]. Social networks are categorized as static social networks and dynamic social networks. Communities play a very vital role in social networks and detecting the communities is an imperative challenge for wide-ranging application predominantly allied to marketing and business. Social network communities can be perceived through various properties of the network such as a many people interacting extensively by forming a cluster and have slacker linking to the other members of the network [1]. More precisely, a community means a set of vertices in which there are more edges between vertices within the set than to vertices outside of it.

Community detection has evidenced to be a very vital problem in the area of social networking. The primary goal is to partition the network into dense regions of the graph. Such dense regions usually resemble to entities which are meticulously related, and can hence be said to belong to a community. The determination of such communities is expedient in the context of a variety of applications in social-network analysis, including customer dissection, endorsements, link inference, vertex classification and influence analysis. As a consequence, a significant amount of research has been devoted towards algorithms for solving this problem [2].

The social network is classified in various categories, Signed, Positive, Heterogeneous, Static, Dynamic, Directe [3]. The problem of community detection can occur in all the categories of social networks static but researchers are more interested now a days in dynamic content-based social network. There are numerous algorithms and methods proposed by researchers for community detection in a static social network like graph partitioning, hierarchical clustering, Bayesian models, Enf-Gibbs sampling algorithm and many more. The prompt change in the era of internet the data on social network is shared and transferred so recurrently that static methods are not able to detect the communities very proficiently. Due to latest internet technologies frequent change in the size of the social network attract huge amount of people to join social network, due to this the previous static community frameworks were not efficient enough to handle the dynamic nature of social network. As the network grows in size formation of new communities takes place and previous communities will become denser and will lead to the failure of previous methods used in community detection. So, researchers are concentrating more on the dynamic aspect of social network. Activities and interactions between the entities change over time. Also, a complex network system is dynamic and has a nontrivial topology and structure. Thus, to reveal imperative, essential structure and relationships underlying these networks has engrossed great interest from artificial intelligence and data mining communities.

www.ijert.org

So research in the area of dynamic social network has become very crucial order to solve the problem of community detection. Many methods have been presented with the conjecture of a static network structure leading to a lack of consideration of the evolution and dynamics of networks. The remaining paper will throw a light on different frameworks and algorithms that were proposed for community detection in a dynamic social network.

II. COMMUNITY STRUCTURE

In the study of complex networks a network is said to have community structure if the nodes of the network can be easily grouped into sets of nodes such that such that each set of node is densely connected internally. The more general definition is based on the principle that pairs of nodes are more likely to be connected if they are both members of the same community or many communities, and less likely to be connected if they do not share communities. Fig. 1 shows a sketch of a small network displaying community structure, with three groups of nodes with dense internal connections and sparser connections between groups.

Fig. 1. A Graph showing a Community Structure [3]

Detecting communities within an arbitrary network can be a computationally difficult task. Despite of so many complexities several methods for community finding have been developed and employed with varying level of success. Hierarchical clustering is one of the methods for detecting community structures in networks. In this method one defines a similarity measure quantifying some type of similarity amongst node pairs. Some common measures include the cosine similarity, the Jaccard index, and the Hamming distance between rows of the adjacency matrix. Another commonly used algorithm for community detection is Girvan-Newman algorithm, this algorithm identifies edges in a network that lie between communities and then removes them, leaving behind just the communities themselves. The identification is performed by employing the graph-theoretic measure betweenness, which assigns a number to each edge which is large if the edge lies "between" many pairs of nodes.

Modularity maximization and clique based methods are some other algorithms for detecting the community structure.

III. COMMUNITY DETECTION

Networks in general and social networks in particular often contain some form of group structure known as communities (other common terms used are partitions, modules, and clusters). In the context of data clustering, each node inside the community is in some sense similar to its neighbors. For example, we can often find friends, family, and colleagues in the social network of a typical person. These groups are mostly quite isolated and not many friendships exist between the groups and those groups are the communities of the network. They are also similar in regard with their position and social roles in the network. Therefore, it is in general possible to use the obtained community structure for identifying social roles, hierarchies and hidden groups within the network data material [4].
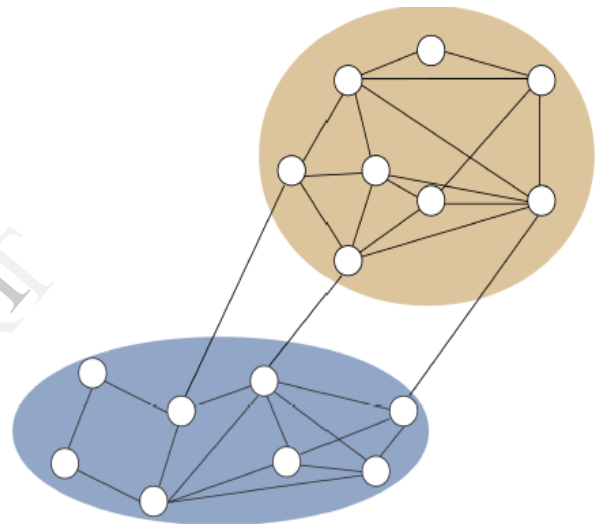

Fig. 2. Two communities in a simple network, the number of edges in each community is much higher than between the two communities.[4]

Communities are a vague and fragile concept found in some networks. It is difficult to find communities and verify their existence and uniqueness but some methods do exist. The most promising is to evaluate the robustness of the clustering. The definition of a community commonly used in social network analysis is that a community is a subset of nodes which have more internal connections than external.

IV. LITERATURE SURVEY

Early research in this area [5] has primarily focused on the static properties of the stated networks, avoiding the fact that most real-world communication networks are dynamic in nature. In practical life, many of the stated networks constantly evolve over time, with the addition and deletion of edges and nodes representing changes in the interactions among the modeled entities. Identifying the portions of the network which are changing, characterizing the type of transformation, predicting future events (e.g., link prediction), and developing generic models for evolving networks are challenges that need to be addressed. For example, the speedy growth of online communities has

dictated the need for analyzing large amounts of temporal data to reveal community structure, dynamics and evolution [5].

The structure and dynamics of social networks are of vital to many social phenomena, ranging from organizational efficiency to the spread of knowledge and disease. The study of such networks has spawned numerous research conferences and even a series of rapidly growing start-up companies providing software tools for social networking. Academic research in social networks has an abundance of interesting and important questions, but has been faced with a paucity of data rich enough to answer many of these questions[13].

Many researchers are working in this domain to solve challenges like Growing size of the network, Dynamic evolution of a network, Ambiguity in social network and various performance issues. In dynamic social network the frequent changes in the nodes and their weight have opened so many research problems for the research scholars. A lot of research has already being done on dynamic nature of the social network and many issues were worked on.

Deepayan Chakrabarti, Ravi Kumar and Andrew Tomkins in 2006 [6] solved the problem of clustering data over time. They proposed an evolutionary clustering framework. This framework requires that the clustering at any point in time should be of high quality while ensuring that the clustering does not change dramatically from one time step to the next. They presented two instantiations of this framework: k-means and agglomerative hierarchical clustering. The experiments on Flickr tags showed that these algorithms have the desired properties — obtaining a solution that balances both the current and historical behavior of data.

Chang-Dong Wang, Jian-Huang Lai AND Philip S. Yu had proposed a method named NEIWalk [7] which supports community discovery in dynamic content-based network. This paper proposes a novel transformation of content-based network into a Node-Edge Interaction (NEI) network where linkage structure, node content and edge content are embedded seamlessly. The content-based network is first transformed into the NEI network, which is a multi-mode network comprising two types of nodes and three types of edges. In the NEI network, the two types of nodes correspond to the nodes and the edges of the original contentbased network, which are respectively termed n-node and enode. On the other hand, the three types of edges respectively characterize the structural similarity, node content similarity and edge content similarity. A differential activity based approach is proposed to incrementally maintain the NEI network as the content-based network evolves. Then heterogeneous random walk is applied in the NEI network to discover latent communities. At the initialization stage, the communities are initialized by clustering n-nodes based on the distance measure computed from the hitting time between n-nodes. At the following timestamps, the evolution of communities is captured by performing heterogeneous random walks in the updated NEI network. Each random walk consists of structural hops, node content hops and edge content hops. Based on this, a random walk based approach termed NEIWalk (Node-Edge Interaction network based random Walk) is devised.

Jimeng Sun, Spiros Papadimitriou, Philip S. Yu, Christos Faloutsos in 2007 proposed a method GraphScope [8]. This framework is based on one form of the Minimum Description Length (MDL) principle and employs a lossless encoding scheme for a graph stream. The goal of GraphScope is to find the appropriate number and position of change points, and the number and membership of source and destination partitions so that the cost of (6) is minimized. Exhaustive enumeration is prohibitive, and thus we resort to alternating minimization. This method is rigorous and automatic, with no need for user-defined parameters. Instead, it uses the Minimum Description Language (MDL) principle, to decide how to form communities, and when to modify them. It is fast and scalable, carefully designed to work in a streaming setting. It is effective, discovering meaningful communities and meaningful transition points. The experiments were performed on several real data sets in which GraphScope gave excellent results.

Gergely Palla, Albert-László Barabási and Tamás Vicsek proposed an algorithm in 2007 [9] which is based on clique percolation that allows, for the first time, to investigate the time dependence of overlapping communities on a large scale and as such, to uncover basic relationships characterizing community evolution. Their results showed that the condition for stability of large communities is continuous changes allowing that after some time practically all members are exchanged.

Sitaram Asur, Srinivasan Parthasarathy, and Duygu Ucar in the 2009 have proposed a framework which was based on the events for the characterization of evolutionary behavior of interaction graphs [5]. The framework is based on the use of certain critical events that facilitate our ability to compute and reason about novel behavior-oriented measures, which can offer new and interesting insights for the characterization of dynamic behavior of such interaction graphs. The author has demonstrated how measures for Sociability, Stability, Influence and Popularity can be compiled. This framework can be used to drive other types of custom behavioral measures which are extremely useful in the context of social information management. The method does not depend on the clustering algorithm used for forming the different clusters of communities which increases the efficiency of the system.

The authors has demonstrated their research on three different datasets—DBLP, Wikipedia and a clinical trials dataset. The application of the behavioral patterns obtained to a cluster link prediction scenario provided favorable results, with the Sociability Index producing a large number of accurate predictions. Their experiments demonstrated that temporal change information can be extremely informative and can be well utilized for predictions in a broad range of applications. Apart from its general efficiency, the framework is also scalable and we have incorporated it in a visual toolkit for dynamic graphs.

YU-RU LIN, YUN CHI and SHENGHUO ZHU, HARI SUNDARAM and BELLE L. TSENG in 2009 has analyzed the communities and their evolution in dynamic social network [10]. In this work they discover communities from social network data and analyze the community evolution. These communities are inherent characteristics of human interaction in online social networks, as well as paper citation

networks. They have termed their work as FaceNet which is a systematic framework for analyzing communities and their evolutions in dynamic networks.

In this proposed method they introduced FacetNet framework to analyze communities and their evolutions in a unified process. In FaceNet the community structure at a given timestep t is determined both by the observed networked data at t and by the prior distribution given by historic community structures. It is the first probabilistic generative model for analyzing communities and their evolution. The proposed model solves the evolutionary clustering problem from a probabilistic (Bayesian) perspective. Empirically, the discovered communities and their evolutions are more robust to noise and more reasonable.

This approach also adopt a stochastic block model for generating communities and a probabilistic model based on the Dirichlet distribution for capturing the community evolutions. These probabilistic models naturally assign soft community memberships to nodes and these models do not suffer from the problem of non-identifiability of parameters which exists in most existing models. Based on the probability distributions computed by the models the concept of Community Net and Evolution Net were evolved to interpret the community-level interaction and transition.

The results obtained from proposed model allowed assigning soft community memberships to individual nodes, to analyze the strength of ties among various communities, to study how the affiliations of an individual to different communities change over time, as well as to reveal how communities evolve over time. However, this approach has some issues like the factorization depends on the number of communities and it requires some domain knowledge to determine a range of candidate for m, otherwise evaluating the system over all communities can be expensive.

Tianbao Yang , Yun Chi , Shenghuo Zhu, Yihong Gong, Rong Jin in 2010 proposed a Bayesian approach [11] for community detection and community evolution in social network. This research presents a probabilistic framework for analyzing dynamic communities in social networks. The authors has developed dynamic stochastic block model for modeling communities and their evolution in a unified probabilistic framework. The framework has two versions the online inference version that progressively updates the probabilistic model over time, and the offline inference version that learns the probabilistic model with network data obtained at all time steps in a retrospective way. This is in contrast to most existing studies of social network analysis that only focus on the online inference approaches. The research also presents a Bayesian treatment for parameter estimation in the proposed framework.

Unlike most existing approaches for social network analysis that only compute the most likely values for the unknown parameters, the Bayesian treatment estimates the posterior distributions for unknown parameters, which is utilized to predict community memberships as well as to derive important characteristics of communities, such as community structures, community evolution, etc.

The proposed framework was implemented using an algorithm which works in an incremental fashion to minimize the computational cost. In addition, the algorithm is designed to fully take advantage of the sparseness of data. Results show that for each iteration the algorithm has a time complexity linear in the size of a social network provided the network is sparse.

Liang-Cheng Huang, Tso-Jung Yen in 2011 proposed a random walk approach for community detection in social network [12]. The main approach focuses on exploring the idea of random walk in formulating modularity functions for community detection. Under this approach, a modularity function is defined as the difference between the probability of a Markov chain induced by a community and the probability of a null model that assumes no detectable community structure exists in the network. The idea of random walk provides an intuitive connection between information flow and community detection. Ideally a community should have more observed within community links than one can expect from a random guess. If one drops a piece of information into the community, the information should circle within the community more likely than within a random graph having the same numbers of nodes and links, given that each link has the same capability of information flow.

V. COMPARISON

On the basis of the study carried out on various frameworks for community detection in dynamic social network a comparison can be done as shown in Table I. It compares the methods on the center of their performance and limitations.

Table I: Comparative Study of Frameworks

Method	Pros	Cons
Event based Framework [5]	Using efficient incremental algorithms for the identification of critical events.	Generalizing the proposed framework for reasoning about events over time. Improving the efficiency of semantic similarity by extending the ontology used in the paper.
Evolutionary Clustering []	Consistency. Noise removal. Smoothing. CI uster correspondence.	Evaluation of clustering algorithms based on trees in order to build weighted and non-binary trees.
GraphScope AJgoritlun	Does not require parameters as input. Is rigorous and automatic. It's an Incremental, fast and scalable algorithm, designed to work in a streaming setting.	Expanding it to produce hierarchical grouping
Palla Algoritlun	It is a local method. It is Based on link density. It allows overlap between nodes	After extracting communities in different time steps, A series of communities must be matched with each other in the next time step, This raises the problem of finding preimage of the given community at the time t .
FaceNet Algorithm	Allows the participation of individuals in multiple communities at the same time and with different participation levels. Introduces new concepts such as community net, evolution net and soft modularity/	Only link information is considered, While the information of the content is necessary in some applications. The model is only used for explaining the observed data, While it is possible to extend the model to predict the future behavior of the individuals of network.

VI. CONCLUSION

Social Networking now a days is considered as the most important feature as so many critical; activities are depended on it. In this survey, the basic concept of social networking and various terminologies related to social network are discussed. The study focuses on the concept of social network and community structure which is considered as one of the most important features of social network and also the importance of detecting these communities in a dynamically changing social network. Besides this various frameworks for detecting the communities in dynamic social network are discussed and a comparative study is shown which shows the advantages and limitations of the various methods. . In conclusion, providing a classification for the sustainability of the proposed approaches in terms of the nature and use of social networks as an important achievement of this study is noteworthy.

REFERENCES

[1] Subramani, K.; Velkov, A.; Ntoutsi, I.; Kroger, P.; Kriegel, H.-P., "Density-based community detection in social networks," Internet Multimedia Systems Architecture and Application (IMSAA), 2011 IEEE 5th International Conference on , vol., no., pp.1,8, 12-13 Dec. 2011 doi: 10.1109/IMSAA.2011.6156357

[2] Qi, G.-J.; Aggarwal, C.C.; Huang, T., "Community Detection with Edge Content in Social Media Networks," Data Engineering (ICDE), 2012 IEEE 28th International Conference on , vol., no., pp.534,545, 1-5 April 2012 doi: 10.1109/ICDE.2012.77

[3] Pourkazemi, M.; Keyvanpour, M., "A survey on community detection methods based on the nature of social networks," Computer and Knowledge Engineering (ICCKE), 2013 3th International eConference on , vol., no., pp.114,120, Oct. 31 2013-Nov. 1 2013 doi: 10.1109/ICCKE.2013.6682855

[4] Dahlin, J.; Svenson, P., "A Method for Community Detection in Uncertain Networks," Intelligence and Security Informatics Conference (EISIC), 2011 European , vol., no., pp.155,162, 12-14 Sept. 2011 doi: 10.1109/EISIC.2011.58

[5] S. Asur, S. Parthasarathy and D. Ucar, "An event-based framework for characterizing the evolutionary behavior of interaction graphs," Trans. Knowl. Discov. Data, ACM. USA, vol. 3, pp. 16, November 2009.

[6] D. Chakrabarti, R. Kumar and A. Tomkins, "Evolutionary Clustering," ACM, pp. 554-560, [Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining]

[7] Wang, C.; Lai, J.; Yu, P., "NEIWalk: Community Discovery in Dynamic Content-based Networks," Knowledge and Data Engineering, IEEE Transactions on , vol.PP, no.99, pp.1,1, 0 doi: 10.1109/TKDE.2013.153

[8] Sun, Ph. Yu, S. Papadimitriou and Ch. Faloutsos, "GraphScope :Parameter-free mining of large time-evolving graphs," ACM. USA, , pp. 687-696, August 2007 [In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining]

[9] G. Palla, A. Barabasi, T. Vicsek, T, "Quantifying social group evolution," Nature, vol. 446, pp. 664-667, April 2007.

[10] Yu-Ru Lin, Yun Chi, Shenghuo Zhu, Hari Sundaram, and Belle L. Tseng., "Facetnet: a framework for analyzing communities and their evolutions in dynamic networks" *page 685-694. ACM,* (*2008*)

[11] T. Yang, Y. Chi, Sh. Zhu, Y. Gong and R. Jin, "Detecting communities and their evolutions in dynamic social networks-a Bayesian approach," JMLR. USA, vol. 82, pp. 157-189,2010.

[12] Liang-Cheng Huang; Tso-Jung Yen; Chou, S.-C.T., "Community Detection in Dynamic Social Networks: A Random Walk Approach," *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on* , vol., no., pp.110,117, 25-27 July 2011 doi: 10.1109/ASONAM.2011.77

[13] http://www.cs.washington.edu/ai/socialnetworks/