A REVIEW ON DYNAMIC COMMUNITY IDENTIFICATION IN INTERACTION NETWORKS

Gauri R. Pagare [1], S. S. Sane [2]

[1]M.E. Student [2] HOD & Vice Principal
[1,2]Department of Computer Engineering
[1, 2] K. K. W. I. E. E. R., Nashik
Savitribai Phule Pune University, Maharashtra, India
[1]pagare.gauri@gmail.com

ABSTRACT

Network consists of community structure, if nodes of the network can be easily clustered into group of nodes such that each set of nodes are heavily connected internally. In social media users interact in diverse ways. One of the types of network is interaction network; where it acknowledges not only social web, but also precise time that nodes take in interaction. A central problem is to study and identify core communities i.e. set of entities among which interactions are prevalent and uniform. In social groups membership often changes, it makes very hard to distinguish structure of communities. Existing method focuses on issue of identifying interaction networks in concrete dynamic communities, where edge corresponds with a time-stamp, and for the same set of nodes multiple edges may fall .In addition to the basic characteristics of node this method considers interaction time for community analysis. In some specific time instances groups of individuals who interact with each other are represented in such communities. Proposed work focuses on social network analysis using additional information such as frequency, hash tags, vocabulary for discovering the community which may result in finding community in an efficient way.

Keyword: - *Community detection, dynamic graphs, graph mining, interaction networks, social-network analysis, time-evolving networks.*

1. INTRODUCTION

To reveal and understand hidden structures and complexities, computer science contributes significantly. Fundamental approach for analyzing structures and their dependencies and also interactions between entities is a graph or network. Community is the important structural feature of graph.

In compound networks, a graph or network is said to have community structure if nodes can be clustered into sets of nodes such that each set of nodes is densely connected within. Networks are growing tremendously, so there is a need to find out the interaction between the communities. We can consider one type of network such as social network; where we can represent community as nodes and their communication as edges. Communication can occur between nodes at multiple times and at different locations. In graph mining finding communities in network is most researched problem. Many methods have been employed such as diverse set of algorithmic tools, spectral methods; min cut formulations, random walks etc.

People are connected to each other via small chains, and as the network increases it is most probable that a huge number of connections can get outside of the community. For example employees working in industries are interconnected. They not only work in industry but they also have other activities outside. And many people may have same hobbies in common, so the boundary between the club and the industry community disappears. Because of social interaction, integration of both communities can be possible. [1] The main goal of this example is that communities mentioned here are not isolated.

Members interact within community but they also communicate with other people outside their communities. If only the static topology is considered then the communities will be unrevealed and it will be difficult to discover them. In above example interactions occur but within short intervals.

It is assumed that all the interaction events [2] between nodes are familiar. Illustrations of such interaction networks include call graphs of telecommunications, email communication networks etc. The issue of discovering dense communities in an interaction network is studied in this paper. Proposed algorithm has proved that the problem defined is NP-hard.

2 .RELATED WORK

The preliminary study in this area mainly focused on the static assets of the specified networks, evading the statement that maximum real world communication systems exist dynamic in nature.

In practical life, several of the stated networks continuously develop over time with the addition and removal of edges and nodes demonstrating deviations in the interactions amongst the displayed objects. Finding the portions of the network which are varying, describing the form of transformation, guessing forthcoming events and emerging generic models for evolving networks are tasks that need to be addressed. For example, the fast growth of online communities has dictated the necessity for evaluating huge quantities of historical data to disclose community organization, dynamics and also development.

Social webs are partitioned as static or dynamic. Detecting communities in static networks is relatively simple task. Survey has been done on the case of static graphs and typical strategy is to divide graph into disjoint clusters. Static network can be optimization based algorithms, this algorithm solves problem by trying to find optimal solution with respect to predefined objective function example: cut criteria by spectral methods. Heuristic based algorithm solves problem using some intuitive assumptions or heuristic rules. Real world networks grows in space and size over time, they are not always static. So real world networks lend themselves to dynamic networks where resources and controls are updated frequently.

In order to enhance performance analysis, dynamic interaction network should be considered in the graph network. Dynamic graph refers to the model where edges are inserted or removed. In this approach once an edge is added in the graph it stays awake until the existing time or until it is removed. For instance this approach may be utilized to model friendship connections in a network. Analysis has been done by the researchers that how network grows with respect to emergence of new nodes and edges. Kumar [3] has studied development of repeated structures in the networks like singleton, small isolated communities and giant connected components. In time evolving graphs for detecting communities a well-known approach is to take into account different graph snapshots, search communities in every snapshot one by one and then form correlation between the communities in successive snapshots. It helps in studying that how communities arrive, fade, partition, combine or develop. Such methods first detect the communities in static graphs induced by one/each time stamp, and then adjust solutions on consequent timestamps to provide consistency. An illustrative example is evolving clustering approach.

Dynamic graph studies are dedicated to event prediction problem. One common proposal of event detection is identification of graph substructure. Majority of work focuses on in what way to compare graph snapshots, and objects to identify those snapshots where the graph organization varies significantly. Network evolves fast and is dynamic in nature in that case snapshot model leads to significant loss of information. Interaction based model works with stream of communication and which is more informative. The problem of mining heavy subgraph in interaction networks is solved by Bogdanov [4].The approach proposed by this author is still centered on snapshots method, and it is complex to borderline; and their concept of dense sub graphs is based on edge weights, and their detecting problem relates to Steiner tree problem[5].

Problem of Steiner tree needs a tree of least weight which consists of all leaf nodes. In graph this problem can be divided into two prominent problems such as shortest path and minimum spanning tree [6].Steiner tree problem reduces for finding shortest route if it has exactly two terminals. And if all the vertices' are leaf nodes then this problem for graph is similar to minimum spanning tree.

Comparative study of various systems for community detection in dynamic social networks has been done. Abdelghani [7] central idea was to evaluate social interactions as they occur over time. Algorithm is constructed on the normal phenomena of bird flocking. However this algorithm had some drawbacks such as no filtering of observations and it also included observations with low entropy. Jingyong Li [8] approach is based on the fact that communities tend to evolve gradually over time, and will not suddenly appear or disappear. Performance decreased as the dataset size increased and thus this approach had this drawback.

Chang-Dong Wang had proposed a method named NEIWalk [9] which supports community detection in dynamic content based network. This method proposes a novel transformation of content-based network into a Node-Edge Interaction (NEI) network where linkage structure, node content and edge content are set in 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. This method gets a bounded accuracy loss due to the random walk sampling.

Yu-Ru Lin [10] proposed an approach called Facetnet which is a structure for examining communities and their evolutions in dynamic networks in this approach multiple communities can participate at the similar time with distinct contribution levels. This method comprises only the link information while content information is necessary in some applications. Future performance of the individuals of network line cannot be predicted the model is only used for explanation of observed data.

Bing Kong proposed a dynamic algorithm for community detection in social networks which was NG Modularity based dynamic algorithm. But it had drawbacks in this method networks with directed and weighted edges were not considered and its running time grows exponentially as the number of communities increases.

In contrast to these methods another approach in area of dynamic graphs is point-wise interactions among the network nodes. Discovering dynamic communities in interaction networks which consists of time stamped information regarding all the interaction among nodes is a base problem in interaction networks. Algorithm is designed for the problem of discovering communities in dense sub graph whose edges occur in short time intervals.

3. OVERVIEW

Fig.1 is the general block diagram for the system. Dataset is input to the system, which consists of interaction network G = (V, E) that is set of nodes and time stamped interactions between pair of nodes, where E = (u_i ,v_i, t_i). Here we have considered interactions that are undirected. For nodes of varying time stamped interaction may take place more than once. And similarly multiple interactions can occur between same time and dissimilar nodes. To handle this situation π (E) is defined which will consists of at least one interaction in that instance of time.

Figure 1 represents proposed system architecture. Processing of proposed work is takes places as following way:

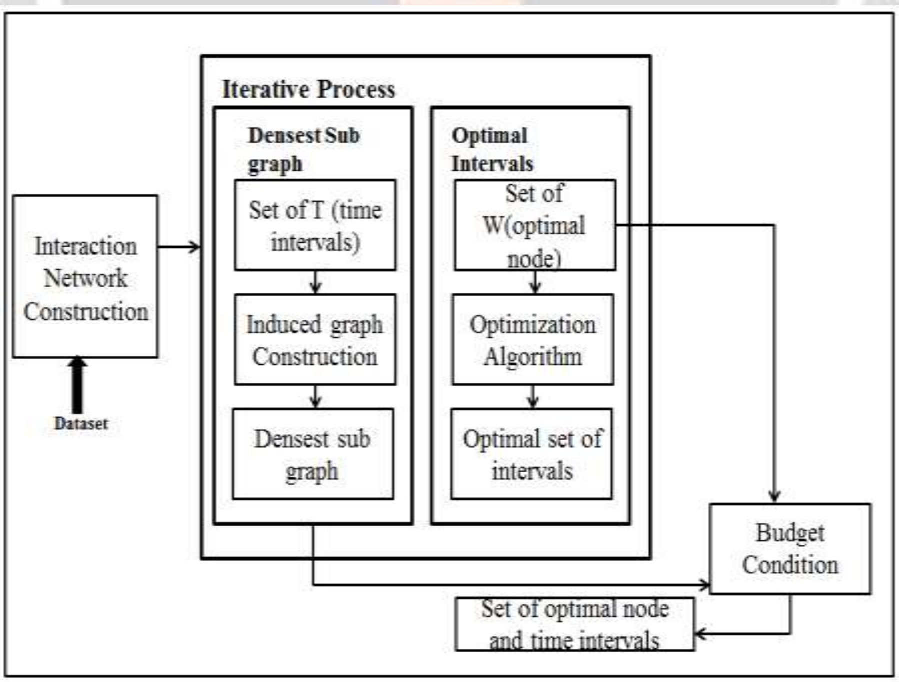

Fig -1: System architecture

3.1 Prerequisites:

For a given interaction network which have subset of nodes, an induced interaction network G(W) = (W, E(W)) is defined, such that E(W) consists of intercommunication whose both endpoints are incorporated in nodes W,

$$E(W) = \{(u, v, t) \epsilon E | u, v \in W\}$$

.

Time intervals [s, f] are defined, where s ∈ R is the beginning point, and f ∈ R is the finish point of the interval. And thus the span can be defined as, span (τ) =f-s. Span (τ) is defined as,

$$\text{span}(\tau) = \sum_{i=1}^{k} \text{span}(T_i)$$

Spliced interaction network can be defined as G (T) = (V, E (T)), where E (T) is communication that occur at time interval T,

$$E(T) = \{(u, v, t) \epsilon\ E \mid s \le t \le f\}$$

Interaction network consists of spliced and induced network which provides two different aspects foe selecting subsets of interactions. One aspect is based on subsets of nodes and the other aspect is based on time intervals. Thus from above definition we can define dynamic community as G = (V, E), where it has subset of nodes 'W' and 'T' as time intervals. Density is defined as,

$$d(H) = 2|F| / |V|$$

Processing of proposed work takes places in following way:

A. **Interaction Network:** Interaction network consists of spliced and induced network which provides two different aspects for selecting subsets of interactions. One aspect is based on subsets of nodes and the other aspect is based on time intervals.

B. **Iterative process:** Algorithm starts by accepting time Interval set τ_0, and obtains results (W_i, T_i) by repeatedly solving the two problems i.e. Charikar and binary until convergence. This process has the feature that both the algorithms outputs accurate solution if it gets output obtained by other algorithm as input to other. Input is time intervals τ_0, and output is (W_i, T_i).

C. **Densest sub-graph:** This algorithm is given by Charikar [11] which deletes the vertex with smallest degree in an iterative process which in turn returns sequence of subgraphs. Among the obtained results it selects one which has maximum density. Input is fixed set of time intervals and output is densest sub graph.

D. **Optimal intervals:** This algorithm has 2 conditions based on budgets; first condition imposes τ to stay below K, whereas second budget imposes time span condition.

E. **Budget Condition:** This algorithm has 2 conditions based on budgets; first condition imposes τ to stay below whereas second budget imposes time span condition. Assuming W is given as input, we find dense dynamic community for K and B budgets and quality score q.

4. CONCLUSION

In this review paper, several existing techniques have studied and analyzed in section II. The study focuses on the importance of detecting communities in a dynamically changing social network. Various methods for detecting the communities in dynamic social network are discussed and a comparative study is done which shows the central idea and weaknesses of various methods. Discovering condensed communities in the interaction networks that consists of time stamped information regarding all the interaction among nodes is a base problem in interaction networks. The method is designed for the problem of discovering communities in condensed subgraph whose edges arise in less time intervals. Issue defined was demonstrated to be NP-hard and efficient algorithms were developed for discovering condensed sub graphs. Charikar's algorithm helped us to obtain the set of nodes and optimization algorithm i.e. Binary algorithm gave us output as an optimum set of time intervals. These two algorithms worked in an iterative process and gave an optimal set of nodes and intervals as an output.

5. REFERENCES

[1] L. Backstrom, D. P. Huttenlocher, J. M. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In KDD, 2006.

[2] Asur, S. Parthasarathy, and D. Ucar. An event-based framework for characterizing the evolutionary behavior of interaction graphs. TKDD, 3(4), 2009.

[3] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In KDD, 2006.

[4] P. Bogdanov, M. Mongiov, and A. K. Singh. Mining heavy subgraphs in timeevolving networks. In ICDM, 2011.

[5] P. Winter,"Steiner problem in networks: a survey J", Networks, vol. 17, pp. 129-167, 1987.

[6] Khawla Asmi,A novel approach based on the minimum spanning tree to discover communities in social networks.

[7] Abdelghani, B. and Anasse, B. SFLOSCAN: A Biologically-Inspired Data Mining Framework for Community Identification in Dynamic Social Networks, IEEE Symposium Series on Computational Intelligence (SSCI 2011).

[8] Jingyong Li, Lan Huang, Tian Bai, Zhe Wang, CDBIA: A dynamic community detection method based on incremental analysis, International Conference on Systems and Informatics - 2012).

[9] Chang-Dong, W., Jian-Huang, L., Philip S. Y. NEIWalk: Community Discovery in Dynamic Content-Based Networks, IEEE transactions on knowledge and data engineering, Vol. 26, No. 7, July 2014.

[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] M. Charikar. Greedy approximation algorithms for finding dense components in a graph. In APPROX, 2000.