An approach for Threat detection in Temporal Social Networks

[1]Swetha G,
Asst. Professor, CSE Department, CMRIT, swetha9g@gmail.com .
[2]*Dr.S.Arvind,*
Professor, CSE Department, CMRIT-Hyd, scarvi@rediffmail.com

***Abstract*— Temporal social network exhibits evolution of a network with different scales of dynamics over a period of time. Threat detection along with its time of occurrence is a challenging task in these time varying networks. The impact of the incremental changes in the network over time on community strength must be analyzed to capture its behavior. In this paper, a methodology is proposed to study the effect of the incremental links on strength of the detected communities within the network. It also includes a method to identify a set of instances during which an effective information flow would lead to a threat. It also demonstrates the utilization of the precise temporal information associated with each interaction. Proposed approach is applied on an email network and identified the instances over which a threat is possible.**

I. INTRODUCTION

In general, a static network encompasses unified information of its complete growth which is not adequate to find out dynamic nature of its activities. A temporal network would be thought of a set of static networks that represent activities along with their time of occurrences. Activities in temporal social networks are not likely to be stable during each period of their evolution. An incremental link at an instant may in turn makes the network further complex to interpret.

Time is the foremost parameter that plays central role in temporal social networks analysis. It is important to examine sequence of activities and their occurrences if a network is capable to store activities over time. Recent technological developments produce data in volumes at an ever increasing rate, social media is one such that demands new methods that can take account of the precise temporal information. An email network or instant message communication can be represented as a sequence of timestamped directed edges, one for each interaction. However, methods are required to perform further analysis over its evolution process in order to determine time-bounded patterns and to capture its dynamics.

In this paper, we proposed a methodology to detect threats in temporal social networks along with their occurrences. Number of time-windows considered within the temporal length of observation period determines number of iterations. In each iteration, the network with the accumulated incremental changes occurred within each time-window is considered for community detection [1, 2]. Modularity, which is the most popular parameter to measure the community strength, is computed in each iteration. We also proposed a method that determines the extent to which new links merely pertain to a time-window contributes to strength of the community detected over the network. Proposed method also identifies a set of instances over which an effective information flow is observed that lead to threat.

This paper is organized as follows. Section II provides an overview of community structure in temporal networks and related work. Proposed method is presented in Section III. Section IV shows experimental results. Section V concludes the paper.

II. RELATED WORK

Palla et al. proposed a methodology [3] to detect the communities based on Clique Percolation Method and this is applied over two pairs of graphs at consecutive time steps to identify community-centric events. This approach was focused on collaboration and mobile phone-call networks. Tantipathananandh et al. also proposed a similar model [4] using graph coloring method which is NP-hard in finding communities in dynamic graphs. So, a heuristic technique is employed to match the set of nodes of two consecutive time steps. However, this technique performs well on small network datasets.

Asur et al. described a strategy [5] that identifies community events by considering non-overlapping snapshots of interaction graphs. This was applied to the pharmaceuticals datasets to characterize behavioral patterns of individuals and communities over time.

Rosvall et al. proposed an approach [6] for identifying changes occurred in dynamic networks. It involves statistical methods to create significant clusters for each time step graph based on the parametric bootstrap. The significant structural changes over time are highlighted and summarized using alluvial diagrams. This was applied on citation network covering various fields with interesting patterns.

Based on the above study we understood that:

- The analysis over the aggregated networks is not capable to reveal pattern of the activities, as it does not consider the temporal ordering of links.
- Strength pattern of the detected communities are better analyzed over varying timelines.
- Number of links established at an instance is merely not sufficient to determine variations in strength of the communities detected in the network.

III. PROPOSED METHOD

In this paper, we have extended the most popular community detection algorithm proposed by Newman-Girvan to temporal social network using time-dependent adjacency matrix. Strength of the detected communities is measured using modularity parameter.

A. Time-dependent adjacency matrix

In general, each interaction in social network contains its time of occurrence along with the involved participants. Social interactions can also be represented as an adjacency matrix that shows time of an interaction occurred between any two persons. This is called time-dependent adjacency matrix (TAM) which is used in finding a community structure that takes care of temporal ordering of interactions.

B. Community detection in temporal networks

We represent a temporal network, denoted by G, as an ordered collection of M time step networks $\{G_1, G_2,.., G_M\}$ defined over N nodes, where each network G_m has the links occurred at time $t_m = t_{m-1} + k$ where $m=1, \ldots, M$. and k is the time-window. Here, temporal length of the observation period, denoted by p, is $(t_M - t_1)$ for G. In other words, the value of k is the ratio of p to the number of networks (M) to be considered for analysis. Eventually, the total number of networks over which analysis has to be performed is M which is less than the total instances occurred during the evolution.

There exist various approaches [7, 8, 9, 10] that are suitable for some problems particularly graph bisection or vertex similarity measures, but these are not applicable for general cases [11]. Newman-Girvan community detection method [12, 13] is extended for temporal networks as shown in Algorithm 1. Community structure is detected in each network that has accumulated directed links based on the defined interval using Algorithm 1.

Algorithm 1 Community detection in temporal networks

Input: Time-dependent adjacency matrix (TAM), number of communities.

Output: Nodes belong to each community.

1) Betweenness metric is computed for each directed link in the network.
2) Identify a link with the highest betweenness value and remove it from the network.
3) Recalculate betweenness for all remaining links.
4) Repeat from step 2 until no link has dominant betweenness value.

Shortest-path betweenness is adopted in this algorithm.

C. Strength of the detected community

Newman-Girvan community detection algorithm formally segregates the nodes of any given network into communities without further categorization of the detected communities. Properties of the detected communities must be analyzed during its evolution to capture the behavior of the network. Modularity is the most popular parameter used to determine strength of the detected community in a network which is represented by the following formula

$$Q = \Sigma_i(e_{ii} - a_i^2)$$

Where e_{ij} is the fraction of edgesin the network that connect vertices in community i to those in community j, and let $a_i = \Sigma_j e_{ij}$. Q is the fraction of edges that fall within communities, minus the expected value of the same quantity if edges fall at random without regard for the community structure. In practice, modularity values fall in the range from about *0.3* to *0.7* are considered strong. However, higher values are rare.

IV. EVALUTION

Here, we evaluated the strength of the detected communities over the email network data sets.

A. Analysis

It is very important to observe the variation in strength of the detected community in a temporal network during its evolution. It is not guaranteed that strength of the detected community will be improved proportionately with an increase in the number of links. The proposed methodology enforces to analyze the behavior of a temporal network by varying duration of time-windows over the complete observation period. Identifying the point of occurrences wherein a sudden change in strength of the community is observed would be vital information in capturing its dynamics. Therefore, analysis of a sequence of networks defined over different intervals is essential in capturing the dynamics of a temporal network.

B. Experiments and results

An email database of four departments in a large European research institution [14] over 803 days is considered. These e-mails only represent interaction between institution core members along with time of exchange, but not with the rest of the world. Each file consists of a set of entries such as (i, j, t) that represents a person i sent an e-mail to person j at time t.

The members of the department 1 (named *D1)* are 309 that represent as nodes participated in email communication. Time-dependent adjacency matrices (TAM) that represent each department email communication are shown in the figure 1. The value in each entry in the TAM represents the day in which an interaction occurred which is normalized and shown in the range from 1 to 64 days.

Communities are detected along with their strengths by choosing three time-windows with duration such as 7, 15 and 30 days over the fixed temporal length of the observation period (803 days). Trend in strength of the detected communities in terms of modularity over different time windows with respect to the incremental links of *D1* are shown in figure 2, 3and 4. It is also evident from the figure 2, 3 and 4 that there is no email communication occurred from 560th day onwards in *D1*.

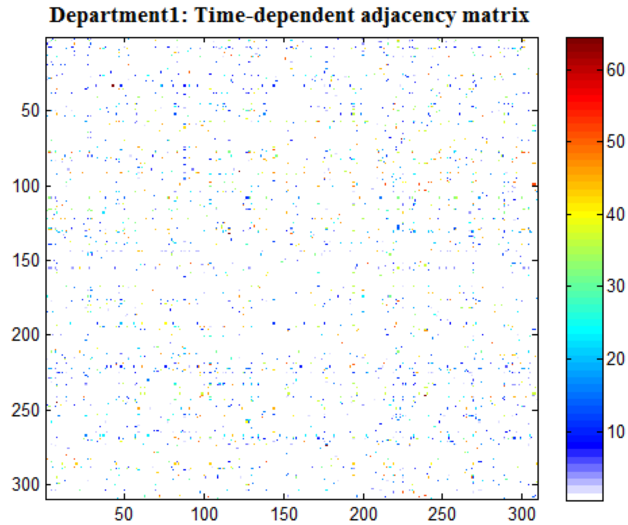

Fig 1. Time-dependent adjacency matrix (TAM) that represent Email communication

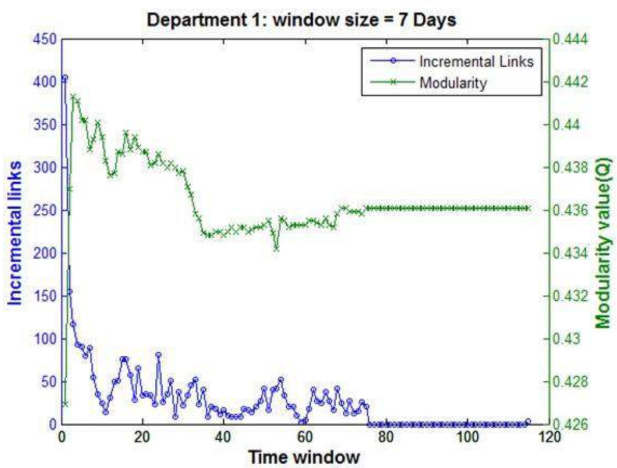

Fig 2. Modularity over incremental links accumulated over 7 days

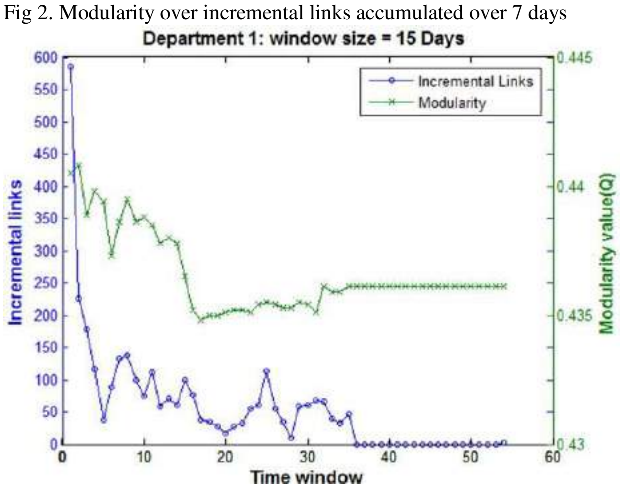

Fig 3. Modularity over incremental links accumulated over 15 days

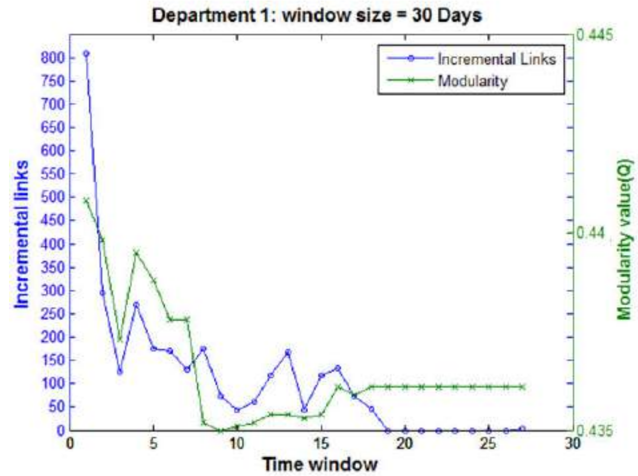

Fig 4. Modularity over incremental links accumulated over 30 days

It is observed from the figure 2, 3 and 4 that the lowest strength of the detected community in *D1* is observed during 50th week. There is a gradual decrease observed in strength of the detected communities within *D1* for a period of 5 weeks i.e., from 30th week to 35th week. These periods are not vulnerable for threat. In other words, rest all periods wherein the modularity values show high are categorized as threat zones during the evolution.

However, the detected communities within *D1* are strong throughout the considered observation period as the modularity values are above *0.4*.

V. Conclusion

The Newman-Girvan community detection finding method applicable for static networks is successfully extended to temporal social networks using Time-dependency adjacency matrix. We have studied the variations in strength of the detected communities along with its incremental links to capture dynamics of the network. More importantly, it also discovers a set of nodes that had an effective communication along with its time of occurrence.

The proposed methodology shown the utilization of the precise temporal information associated with each interaction along with its strength pattern. It is also able to identify a set of instances during which information flow is effective by observing the pattern of strength of the detected communities.

This methodology is easily adoptable to analyze any kind of time-varying networks. A suitable clustering method can also be developed based on the objective of the analysis.

Acknowledgment

The author is grateful to D Rajesh Reddy, Scientist, Advanced Data processing Research Institute (ADRIN), Department of Space, Secunderabad for his valuable guidance and suggestions to accomplish this work.

REFERENCES

[1] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Defining and identifying communities in networks, Proc. Natl. Acad. Sci. USA 101, 2658-2663, 2004.

[2] Vincenzo Nicosia, John Tang , Mirco Musolesi , Giovanni Russo , Cecilia Mascolo , Vito Latora Components in time-varying graphs, American Institute of Physics, 2012.

[3] Palla, G., Barab´asi, A, Vicsek, T., Quantifying social group evolution. Nature, 446, 664–667, 2007.

[4] Tantipathananandh C., Berger-Wolf T., Kempe D, A framework for community identification in dynamic social networks, In Proc. 13th ACM SIGKDD International conference on Knowledge Discovery and Data mining, 717–726, ACM, 2007.

[5] Asur S, Parthasarathy S, Ucar D, An event-based framework for characterizing the evolutionary behavior of interaction graphs, In Proc. 13th ACM SIGKDD international conference on Knowledge Discovery and Data mining, 921, ACM, 2007

[6] Rosvall, M. & Bergstrom, C.T. (2010). Mapping change in large networks. PLoS ONE, 5, e8694.

[7] B. W. Kernighan and S. Lin, An efficient heuristic procedurefor partitioning graphs. Bell System Technical Journal 49, 291-307, 1970.

[8] M. Fiedler, Algebraic connectivity of graphs. Czech.Math. J. 23, 298-305, 1973.

[9] A. Pothen, H. Simon, and K.-P. Liou, Partitioning sparsematrices with eigenvectors of graphs, SIAM J. Matrix Anal. Appl. 11, 430-452, 1990.

[10] J. Scott, Social Network Analysis: A Handbook. Sage, London, 2nd edition, 2000.

[11] M. E. J. Newman, Detecting community structure in networks, Eur. Phys. J. B 38, 321-330, 2004.

[12] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," Phys. Rev. E, vol. 69, p. 026113, Feb 2004.

[13] M. E. J. Newman, "Modularity and community structure in networks," Proceedings of the National Academy of Sciences, vol. 103, no. 23, pp. 8577–8582, 2006.

[14] Email-data source: https://snap.stanford.edu/data/email-Eu-core.html