Distinguishing Malicious Community behaviour in Social Networks for Cyber Security

Sashaank Pejathaya Murali
Department of Computer Engineering
SSN College of Engineering Chennai, India

Abstract— this paper is an attempt to distinguish users' emotions or views for the fabrication of communities over social networks for eff icient circumstantial consciousness. I trust that the mined patterns in user sentiments can behave as measures of likely threats in cybernetics. This paper proposes a novel data mining outlook to identify groups and trace these communities to efficiently determine users that impact the kinetics of a community over time.

Key words: Community Detection, Social Networks, Mining, Tags, Clustering

I. INTRODUCTION

The current growth of both social networking and social media as a means of contact cannot be ignored for its outreach. This constantly developing and dynamic cyber system provides a surge of information that can be subjugated to augment the situational consciousness. Accessible methods to achieve the awareness work by modeling the construction of user groups. These practices utilize visualization tools to compute associations between user profiles within a group and across groups. In a social network, users are associated with each other if they share similarities in interests referred to as topics of interest that are dynamic and subject to change due to the user's activities. These changes have to be made track of, in order to create a circumstantial awareness. There is a call for data mining and machine learning for a more spontaneous web that can acclimatize to an individual's wants. One can anticipate the categorization of a personage over cyberspace by the data he has and shares with others. Handling these challenges is indispensable to instituting a well-associated semantic web.

This paper centers around the withdrawal, deduction, and testing of spaces for the detection and categorization of groups in a social network. It emphasizes on techniques that utilize the conduct of individuals over a social network. It exhibits a novel graph technique that uses/traces user behavior patterns for eff icient community detection. Finally, it focuses on future and potential applications of the proposed system.

II. STIMULUS

The main reason for clustering tags is to extract semantics from the web. Grouping tags by identifying associated tags has been done using association rules (depends on user defined parameters) and clustering techniques (require number of clusters as input). The demerit of conventional clustering is that, the overlapping of tags isn't considered, but just the displaced set of the tags. User profiling is determining user interests by analyzing the user activity on the net or by taking into consideration the different views/opinions of the users and forming a cluster of similar views.

Thus, when there are multiple users with similar interests, then there is overlapping of tags. But, I propose a methodology that identifies distinct set of users, even when they are overlapping.

III. ANALYSIS OF PATTERNS

To relate or filter users, we must first find out and extract the best features that bring out the similarity among them. Thus we must find different users that share similar topics of interest. To identify these similarities, tags are used. These tags are nothing but keywords. For the semantic information that is needed for these tags, the profiles of the users are taken into account. Thus, similar communities are formed into a cluster that consists of nodes. So each cluster is a network of similar nodes. This is essentially, community detection.

I propound a graph-based data withdrawal method to extort relevant features that identify the tendency of users in a social network to relate to other users with similar interests. Clusters of tags that overlap are typically the groups of similar interests.

IV. PROPOSED AGENDA

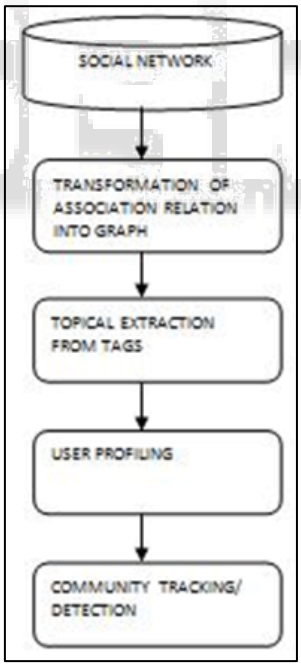

Fig. 1: Mining Step

- In the propounded methodology, first we take into account three entities from the social network database, namely,
  1) Users/Clients
  2) Tags(Keywords)
  3) Association between the users and tags
- Then we take the association relation and transform it in the form of a graph.
- Further, we perform topic extraction from the relevant features from tags.

- Then, we perform user profiling, based on tag clustering.
- Finally, we track communities using mining and association rules.

In the mining step, we perform following:

- User profiling.
- Link settlement between user profiles based on similar interests.
- Community detection.

A. Transformation

The association relation is modeled into a graph.

B. Topical Extraction from Tags

Topic extraction is a process in which we search for topic features (i.e. a set of tags), which can be used to form user profiles. For this, we follow a graph-based modeling approach to extract topics from the tag graph in the form of cliques (subset of the vertices of graph such that every pair of vertices are adjacent). Each tag in the cluster is connected to every other tag of the community, forming a group of tags which is typically a clique. All these tags in a clique refer to similar topics of interest.

C. User Profiling

We haul out graphs of user interests by building user profiles using a method that includes plotting users to interests/topics in the form of cliques. In a user profile/interest graph, similarity between two users who are recognized by an interest topic that is a vector is computed by the dot product between these vectors. The interest graph, which is represented by an incidence matrix, is a compilation of all user vectors. This graph is subjected to the Louvain community detection algorithm to distinguish communities from the social network.

D. Community Detection

Thus, all such users who belong to a cluster of tags, they have coinciding topics of interest. The reason for choosing the Louvain algorithm, among other available algorithms, is that this algorithm is a maximization-based community detection algorithm and it does not need parameters. On usage of this algorithm on the graph, we get groups of similar interest. The similarity in the tag patterns is the reason for the cluster alignment. Conventional mining algorithms, using association rules centers on discovering relationships among entities. Recent research has seen algorithms performing mining based on spatial and temporal parameters. The application of such algorithms is beneficial, and can be used to analyze malicious communities over time.

V. Concluding Remarks

The proposed agenda takes into concern that topics of interest that are shared are identified as overlying clusters of tags from a hierarchy modeled that can be used to mold links between users. From these, we unearth the social community graphs. In cyber security, this can be used to detect who have malevolent targets on social media.

Acknowledgment

I am greatly indebted to my college and my department professors for inculcating and instilling ideals into me that drove me to writing this paper.

References

[1] Wang, X., Tang, L., Gao, H. and Liu, H. (2010). Discovering overlping groups in social media, in Proceedings of the 2010 IEEE International Conference on Data Mining, ICDM '10 (IEEE Computer Society, Washington, DC, USA), pp. 569–578, doi:10.1109/ ICDM.2010.48.

[2] Rees, B. S. and Gallagher, K. B. (2012). Overlapping community detection using a community optimized graph swarm, SNAM 2, 4, pp. 405–417.

[3] Hua, G. and Haughton, D. (2012). A network analysis of an online expertise sharing community, Social Netw. Analys. Mining 2, 4, pp. 291–303.

[4] Kashoob, S. and Caverlee, J. (2012). Temporal dynamics of communities in social bookmarking systems, Social Netw. Analys. Mining 2, 4, pp. 387–404.

[5] Kas, M., Carley, K. M. and Carley, L. R. (2012). Trends in science networks: Understanding structures and statistics of scientific networks, Social Netw. Analys. Mining 2, 2, pp. 169–187.

[6] Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks, Phys. Rev. E 69, p. 066133, doi:10.1103/PhysRevE.69.066133.

