Anomaly Detection in Online Social Network: A Survey

Ketan Anand,[1] Jay Kumar[2], Kunal Anand[1]
[1]K L University, Vaddeswaram, Guntur, Andhra Pradesh
[2]North Bihar Power Distribution Corporation Limited, Bihar
ketananand13@gmail.com, jay.fst.iud@gmail.com, kunaljitm@yahoo.com

***Abstract*—Anomaly in Online Social Network can be referred as abnormal or unexpected behavior which deviates from majority of users. Due to popularity of social networking sites such as Facebook, Twitter etc., malicious activities have increased in recent past. Anomaly detection has become an important area for researchers to be looked upon. This survey gives an overview of existing techniques, which is further kept under two different types, structural based and behavioral based techniques for anomaly detection in social network. It also discusses major problem/s relating anomaly detection.**

***Index Terms*—OSN, Anomaly Detection, Structural based anomaly detection, Behavioral based anomaly detection**.

I. INTRODUCTION

Online Social Network (OSN) can be defined as the web services, which provide sharing of files and interactive re-lationship between people in a virtual society. For example, Facebook, Google+, Twitter, etc. It enables people to stay in touch with their contacts, reconnect with old acquaintances and establish new relationships with another people based on shared features such as communities, hobbies, interests and overlap in friendship circles.

Online Social Network has gone through a tremendous growth due to its core features such as: Personal Space Management, Social Connections, Means of Communication, and Exploring Digital Social Space.

Due to huge growth of social media and online social system, many social networks have become a key target for malicious individuals. Lot of malicious activities have been reported in the recent past. Some of which are as follows:

- Sybil Attacks
- Cloning (creating fake profiles)
- Spamming
- Private Information Inference

According to Chandola et al. [9], anomalies can be referred as outliers, novelties, noise, deviations, and exception. Anomaly Detection is the identification of items, events or observations which do not conform to an expected pattern. Anomaly detection is applicable in a variety of domains, such as intrusion detection, fraud detection, fault detection, system health monitoring, event detection in sensor networks, etc.

According to Doostari et al. [5], anomalies in an Online Social Network can be defined as an observation that deviates from majority of observation. In other words, an unexpected or irregular behavior that deviate from majority of users in social network. Therefore, we can say that" Anomaly detection in OSN can signify various malicious activities".

Malicious activities in online social networks are not been limited to simple spamming, but have been transformed to relatively more intelligent attacks, which tend to break the privacy of users. The detection of malicious activities is very important to prevent privacy breach of users in OSN.

As per Savage et al. [9], anomalies that occur in social networks can be characterized as being dynamic or static, and as labelled or unlabeled. Anomalies can be further classified as occurring with respect to a global or local context across a network unit which is known as minimal anomalous unit. Global anomalies occur with respect to entire network while local anomalies occur with respect to only close neighbors. Lot of anomaly detection methods has been proposed by using data mining techniques. For example, parametric (Barnet et al. [11], Hawkins et al. [12]) and non-parametric method (Manson et al. [13], Papadimitriou et al. [14]).

Several techniques have been proposed in literature for the detection of the anomalies in OSN. Two important approaches among them are behavioral based and structural based approach. Behavioral based approach is based on the characterization of user behavior in Online Social Network while Structural based approach is basically based on Social Network structure which is created due to some malicious activities in OSNs.

Further, this survey paper is divided into four sections. Section I explains introduction about the survey, Section II depicts about the behavioral based approach. Section III discusses about structural properties approach and finally Section IV provides the final inference.

II. BEHAVIORAL BASED APPROACH

This approach focusses on mining the behavior of the user i.e. "usage pattern" of users. It also focuses the way in which interactions occurs between pairs of individuals and other individuals in the system.

Doostari et al. [5], has proposed a new hybrid approach to detect anomaly in OSN which uses both parametric and non-parametric approaches based on fuzzy logic. In OSN, cliques are suitable place for malicious user to fulfil their sabotage intent. It is very difficult to evaluate whole social network. Hence Clique is used as a base structure for detection of anomalies. OSN contents change rapidly and members of different network or sub-network have different behaviors based on their culture and thoughts. Therefore, anomalies for every network is different from others. This uncertainty and dynamic situation of social network environment is in adaptation with fuzzy logic principles. Proposed approach was examined in two datasets first was Iranian student social network and second was a simulated dataset which was examined in other researches. In first implementation, it detects 47 percent abnormal users, in second implementation this approach is compared with concluded results from similar researches thereby showing that the approach is better for this dataset.

Fire et al. [3], has proposed a supervised learning and graph theory method to detect fake profiles in OSN. A classifier is used in this method which requires a set of positive and negative examples for its training by supervised learning. For each social network decision tree(J48) and Naïve Bayes classifiers is constructed for detecting fake profiles in the social network. These classifiers are applied on three different databases of social network (namely academia, google+ and Anybeat), for each classifier false positive rate, f-measure, and AUC (area under ROC curve) is measured. This algorithm finds number of fake profiles in these databases that shows good performance of this method.

Bhat et al. [8], has proposed a community shield approach to detect the Sybil and cloning attacks in OSN. It splits the interaction network into communities. It uses OCTracker algorithm to detect node level communities as cores, non-cores and outliers. a classifier is trained using a set of extracted features (topological and community based features) that classify unlabeled nodes of the interaction graph. To detect cloning Community-tracking algorithm can be used because it can detect merged region which is created due to greedy behavior of attackers that create more dense regions in the network by adding more nodes and cross-links that cause adjacent communities to merge.

Viswanath et al. [10], has proposed Principle Component Analysis (PCA) approach that differentiates bad behavior of user from normal behavior. PCA is the technique used, that accurately identifies significant deviation from normal users. PCA uses three input features to capture user behaviors in OSN such as: temporal, spatio-temporal and multiple features. Two years (2011-13) Facebook like activity is analyzed using PCA. it shows that PCA approach is more effective than naive approaches capturing complex normal user behavior patterns to correctly flag misbehaving users. PCA also shows robustness as detection accuracy does not change with the different choices of k (principal component) value. It is also scalable as space requirement for training is O(n+m) where, n is no. of inputs and m is no. of users in training set. Sahlabadi et al. [7], has proposed a process which uses mining approach, that defines the behavioral pattern in order to detect anomalous behaviors. This method develops a control system for the monitoring of users' activities to detect threats. By logging activities and comparing the actions of the users against predefined model. The system can detect suspicious activities. This method uses three algorithms used for anomaly detection in process mining. First algorithm is based on sampling. It makes a sample population from log file based on sampling factor. Second algorithm is based on threshold division of the log files into frequent and infrequent sections. The last algorithm is based on iteration it works with threshold which is used to categorize the traces into normal and anomalous. Genetic mining is the key feature of this proposed approach because it can mine all major types of pattern structures having low performances. The idea of using genetic process mining for finding suitable model in social network websites was introduced first time in this study.

III. Structure Based Approach

In this approach, focus is put to find some special type of network structure in social network graph, which is constructed due to abnormal/malicious activities in online social network. Structural properties of graph play an important role in this approach.

Social networks can be represented using a graph with vertices (Node) and edges (links) between the vertices rep-resenting interaction. Akoglu et al. [1] represented, anomalies in graph termed as anomalous nodes, follow a specific pattern which is different from normal network structures like clique, stars, heavy vicinity and dominated links. In their proposed method, four features are extracted. These features are combined which is able to detect new patterns in the graph. Finally, score of individual node is calculated based on the observation from power law to detect most anomalous node in the graph.

Dynamic Networks such as social network that is evolving day-by-day can be highly clustered. Chen et al. [4], has proposed Detection of communities in this type of networks would be a better approach for anomaly detection. Communities in dynamic network is a very important property of evolutionary networks. Community detection based approach detect communities in the dynamic network. This approach is based on the graph of dynamic networks. In graph, there are two major anomalies namely White Crow and In-Disguise. White Crow anomaly is an observation that substantially deviates from other observations, while in-disguise anomaly the deviations from the normal pattern is minor. Communities are the maximal clique in graph. This approach can detect six types of possible community-based anomalies in evolutionary networks. Altshuler et al. [3], has proposed a Structural based approach which is used in the scenario of Mobile networks. It examines the detection of anomalous events in such networks using a comprehensive dataset containing entire internal calls as well as many of the incoming and outgoing calls with in major mobile carrier in a west European country, for a period

TABLE I
COMPARISON OF ANOMALY DETECTION ALGORITHMS

Sr No.	Method's Name	Description
1	Anomaly Detection in Cliques of OSNs using Fuzzy Node-Fuzzy Graph	It uses anomaly detection algorithm that is hybrid approach from parametric and non-parametric methods based on fuzzy logic. Clique is taken as a base structure. This approach integrates different methods for anomaly detection which proves quite advantageous upon previous ones. Also social network graph is sparse and the computational overhead is not a limitation in this approach. The dynamicity of this approach makes it practical for real time anomaly detections.
2	Stranger Intrusion Detection	This algorithm which is based solely on the topology of the social network, detects users who randomly connect to others by detecting the anomalies in that networks topology.
3	Communities Against Deception in OSNs	Fake profile and sybil node detection based on community structure and merging pattern.
4	Detecting anomalous behavior in OSNs	It shows that using social features of the network (namely, focusing on network hubs) prediction accuracy of anomalous events can be significantly increased. In order to detect anomalies in the networks hubs, it uses the Local-Outlier-Factor (LOF) anomaly detection algorithm. Specifically, focusing on the neighborhood around a hub (the connections among the alters) detects events external to the network that provoke spreading communication within the network.
5	PCA	It achieves a detection rate of over ***66%*** (covering more than ***94%*** of misbehavior) with less than ***0:3%*** false positives. PCA provides a more systematic and general frame-work for modeling user behavior in social networks, and in fact, PCA-based approach could leverage the user behavior features.
6	Process Mining Technique	This technique comprises of Process mining algorithm and Metric quantifying the manipulation of model structure in order to measure the extent to which the trace fits the model. Most of previous approaches suffer the lack of model mutations to achieve norms but this method has advantage over this.
7	Anomaly Detection in large graph	A novel, hybrid method for outlier node detection. The major types of anomalous nodes spotted are as:Near-cliques and stars, heavy-vicinities and dominant heavy links. A fast, hybrid and unsupervised method to detect abnormal nodes in weighted graphs. Possible approximations in feature extraction that provides speed-up, keeping accuracy at as high as 9

of roughly three years. This method applied for emergency detection using real world data. Anomalies categorize into three groups: Concerts and festivals, Small exposure events and Large exposure events. Each day is rated between 0-1 as per its anomalousness. 1 for anomalous day and 0 otherwise. Approx. 300-800 edges have been monitored for the three types of anomalous events. Using social structure properties of social network, accuracy of anomalous events has significantly increased.

Hassanzadeh et al. [2], has proposed a Local graph properties are used that refer single node (ego), its neighborhood (an egonet) and 2-neighborhood (super egonet). Betweeness centrality and average betweenness centrality of user's egonet and community cohesiveness of the user's super egonet is used as measures for detecting anomalies based on the structure of user links. Method evaluation is performed on existing data collected from three online social networks (Facebook, Orkut and Flickr).

IV. CONCLUSION

In this paper, we have reviewed a small but growing number of solutions to detecting anomalies in online social networks. Existing methods can be characterized as structure and behavior based techniques. In structure based technique special type of network structure which is an indication of abnormal behavior is discussed. For example, Srivastava et

al., has taken star like structure as an indication of abnormal behavior while Akoglu has taken near-star and near-cliques as an indication of malicious behavior in online social network. Behavioral based technique is based on the user behavior in social network such as anomaly detection in clique based on user behavior (Dosstari et al. [5]) and PCA based method which uses activity log of users to detect anomalies in online social network (Viswanath et al. [10]).

A little work has been done in this area so, lack of papers clearly describes the difficulty arises in this field. In near future, requirements for anomaly detection in social networks will be advance due to larger volumes of data and increasingly complex behaviors will be taken into consideration. We can apply multiple detection techniques across numerous graph properties, so, that it may be possible to identify core links between the actual behaviors of interest and resulting changes in network properties, and to determine features that clearly differentiate malicious and normal behaviors.

Anomaly detection in online social network is a challenging task. Some major challenges we have discussed are the selection of suitable feature that define the normal behavior of a user in social network for anomaly detection. It is very important to choose best feature which describes the normal user behavior in online social network environment. Another major challenge for detecting anomalies in online social network is the evaluation and comparison of different methods. Many existing approaches have been developed with specific problem domains and data formats in mind. Moreover, there is a lack of publicly available data sets with known ground truths.

In this paper, we have surveyed several methods and approaches for anomaly detection in online social networks. We found that useful characterization of anomaly, based on network structure and user behavior. Savage et al. [9], have discussed a novel anomaly known as horizontal anomaly, it occurs when the characteristics or behaviors of an entity varies depending on the source of the data. For example, a user of social media may have a similar set of friends or followers across a few platforms (e.g. Twitter, Facebook, etc.), but for one platform (Google+ for example) has a markedly different set of acquaintances. Methods for detecting this type of anomaly are beyond the scope of this paper. Finally, if the requirements of social network analysis continue to grow, more challenges can occur in big data analysis. For this, we need to develop a novel algorithm to analyze complex behavior in social network.

In future, anomaly detection may become an important research area and will play a role of immense importance in securing the privacy of users and detecting malicious activities such as spamming, cloning, sybils etc. Sophisticated heuristic algorithms that analyze the complex behavior of user based on new feature space or based on characterization of user activities can be future work in this area.

REFERENCES

[1] Leman Akoglu, Mary Mcglohon, and Christos Faloutsos, "Anomaly detection in large graphs," In CMU-CS-09-173 Technical Report, 2009.

[2] Reza Hassanzadeh, Richi Nayak, and Douglas Stebila, "Analyzing the Effectiveness of Graph Metrics for Anomaly Detection in Online Social Networks," Web Information Systems Engineering - WISE 2012, vol. 7651, pp. 624-630, 2012.

[3] Michael Fire, Gilad Katz, and Yuval Elovici , "Strangers Intrusion Detection - Detecting Spammers and Fake Profiles in Social Networks Based on Topology Anomalies," ASE Human Journal, 2012.

[4] Zhengzhang Chen, William Hendrix and Nagiza F. Samatova , "Community-Based Anomaly Detection in Evolutionary Networks," Jour-nal of Intelligent Information Systems , vol. 39, iss. 1, pp. 59-85 , 2012.

[5] M.A. Doostari, Ramin Zeinali, Hamed Lashkari, and Mehrana Ajamza-mani, "Anomaly Detection in Cliques of Online Social Networks Using Fuzzy Node-Fuzzy Graph," Journal of Basic and Applied Scientific Research, vol 3 (8), pp. 614-626, 2013.

[6] Yuval Elovici, Alfred Bruckstein, Alex (Sandy) Pentland, and David Lazer, "Detecting Anomalous Behaviors Using Structural Properties of Social Networks," Social Computing, Behavioral-Cultural Modeling and Prediction, vol. 7812, pp. 433-440, 2013.

[7] Mahdi Sahlabadi, Ravie Chandren Muniyandi and Zarina Shukur, "De-tecting Abnormal Behavior in Social Network Websites by Using A Process Mining Technique," Journal of Computer Science, vol 10(3), pp. 393-402, 2014.

[8] S.Y. Bhat and M. Abulaish, "Communities Against Deception in Online Social Networks," Computer fraud Security, vol. 2014, Issue 2, pp. 8-16, 2014.

[9] David Savage, Xiuzhen Zhang, Xinghuo Yu, Pauline Chou, and Qingmai Wang, "Anomaly Detection in Online Social Network," Social Network, vol. 39, pp. 62-70, 2014.

[10] Bimal Viswanath, M. Ahmad Bashir, Mark Crovella, Saikat Guha, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove, "Towards Detecting Anomalous User Behavior in Online Social Net-works," 23rd USENIX Security Symposium (USENIX Security 14), pp. 223-238, 2014.

[11] V. Barnett and Lewis T, "Outliers in statistical data",j. wiley & sons 1994, xvii. 582 pp., 49.95. Pincus, R, 1995.

[12] Simon Hawkins, Hongxing He, Graham Williams, and Rohan Baxter, "Outlier detection using replicator neural networks," In Data warehousing and knowledge discovery, pages 170-180. Springer, 2002.

[13] Graeme Manson, Gareth Pierce, and Keith Worden, "On the long-term stability of normal condition for damage detection in a composite panel," Key Engineering Materials, 204:359370, 2001.

[14] Panagiotis Papadimitriou, Ali Dasdan, and Hector Garcia-Molina, "Web graph similarity for anomaly detection," Journal of Internet Services and Applications,1(1):1930, 2010.