ISSN No:-2456-2165

A Survey on Online Social Network Anomaly Detection

Naznin Sultana
Department of Computer Science and Engineering
Daffodil International University, Bangladesh.

Prof. Sellappan Palaniappan
Department of Information Technology,
Malaysia University of Science & Technology, Malaysia

Abstract:- The increasing popularity of online social networks in different domains, have made convenient platforms for people to share, communicate, and collaborate with each other, which at the same time poses significant challenges and threats as many malicious behaviors, such as bullying, planning terror attacks, stealing personal information, profile cloning, social phishing, and neighborhood attacks, and physical threats. These abnormal activities in social networks are called anomalies. As social networks become a convenient platform for such malicious activities, it is extremely important to detect these activities as accurately and early as possible to avert potential threats and ensure the safety of social networking. Many researchers are now studying the usage of these networks to detect anomalous activities. There are many anomaly detection techniques available. Some of these are developed for generic applications while others are developed for specific applications. This survey provides a comprehensive overview and discussion of recent researches on different types of anomalies and their novel categorizations based on several characteristics. It also presents a structured overview of the various state-of-the-art methods for anomaly detection. Some challenging issues of existing state-of-the-art-methods for anomaly detection are also addressed. Finally, the paper concludes with future directions and research areas.

Keywords:- Online Social Networks (OSN), Anomaly Detection, Data Mining, Network Outlier Detection

I. INTRODUCTION

Today's Information Age is characterized by an explosion of huge volumes of raw data, particularly data that involves social interactions in large groups. As online social networks provide a hangout space for many people, using this technology enables them to communicate with their friends and share their personal information such as photos and videos. With this influx of data, we can discover many useful patterns such as information on individual or group behavior and their interconnections. This also opens the door for harmful and/or illicit activities. These activities are also referred to as anomalies [1].

Anomalies are defined as deviations from the normal or expected behavior. Different authors have defined anomalies in different ways but in general, anomalies can be defined as "patterns in data that do not conform to a well-defined notion of normal behavior". Anomalies emerge in online social networks as a result of specific people or gatherings of people rolling out sudden surge in their interactions or communicating in a way that extraordinarily contrasts them from their peers. The effects of this strange phenomenon can be seen in the subsequent network structure.

Today, there are many social networks such as Facebook, Twitter, LinkedIn, Viber, WhatsApp, YouTube, and We Chat that aim to form a cyber-network amongst users. As individuals spenda lot of time on digital devices, it is easy to study their behaviors in a variety of contexts and extract a range of features. However, we can only get a few relevant social groups that can be analyzed from these networks, whereas interesting social patterns are likely to emerge in the everyday context, such as workplace politics and other substantial situation.

Data mining and knowledge discovery are popular research topics. The branch of data mining that deals with discovering abnormal occurrences in databases is called anomaly detection. This has many high-impact applications in domains such as security, finance, healthcare, and law enforcement. There are many techniques available for detecting anomalous behavior in online social networks. Every online social network has its own structure and nature. This survey discusses both traditional and modern ways to detect anomalous users from online social networks.

There are many survey articles based on anomaly and outlier detection, but they generally focus on point anomaly of multi-dimensional data instances. For example, Chandola et al. [2]discussesoutlier detection techniques; Schubert et al. [3], local outlier detection techniques; Zimek et al. [4], outlier detection in high dimensions. Others discuss detecting network intrusion and network failure [5], credit card fraud[6], auto insurance fraud [7], email and Web spam [8], opinion deception and reviews spam [9], health insurance claim errors[10], accounting inefficiencies[11], auction fraud [11], tax evasion [12][13], customer activity monitoring and user profiling [14][15], click fraud [16][17], securities fraud [18], malicious cargo shipments [5][19] malware/spyware detection [20], false advertising [20], insider threat [21], image/video surveillance[22][23], health insurance claim errors[10], and accounting inefficiencies [11]. As anomalous events occur in wide-range of application domains, fraud detection has led to

several studies in these areas[24][25].

The rest of this paper is organized as follows: Sections 2 and 3 contain novel categorization of anomalies on the basis of a number of parameters and a brief overview of existing techniques for social network anomaly detection respectively. The output of anomaly detection and different challenges of anomaly detection are discussed in Sections 4 and 5 respectively. Finally, Section 6 presents future directions.

II. TYPES OF ANOMALIES

Anomalies can be classified into various categories. This section describes the various types of anomalies with examples that can be found in social networks.

➢ *Based on nature of anomalies*

Though anomalies can be classified into three main categories based upon the nature and scope of anomalies[2], there are actually four types of anomalies in social networks:

A. Point anomalies

A point anomaly refers to detecting an anomalous data instance in the data. Point anomaly detection aims to detect suspicious individuals, whose behavioral patterns deviate significantly from the general public. It is also referred to as global anomaly if a data object (i.e. a point) shows a different behavior than that of the rest of the data. For example, we assume that for a normal network every node must have at least two neighbors linked to it. Fig. 1 shows two groups V1 and V2 where nodes in GroupV2 form such type of network and thus represent a normal behavior whereas group V1 contains isolated points because of their dissimilar behavior to other nodes, so they are seen to represent an anomalous behavior.

Fig. 1:- Point anomalies.

Similarly, we may also have local anomalies which are studied relative to their local neighborhood only. For example, if we group a set of individuals based on their links in the network as friends and check their income (some parameter), a particular individual, let's say A, might be having a fairly low income compared to his friends suspecting a local anomaly while overall in the global context his income might be insignificant as many people may have similar income representing a normal behavior. This behavior is depicted in Fig. 2(a) and 2(b).

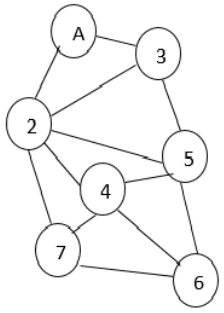

Fig. 2a:- Groups on the basis of friendship links.

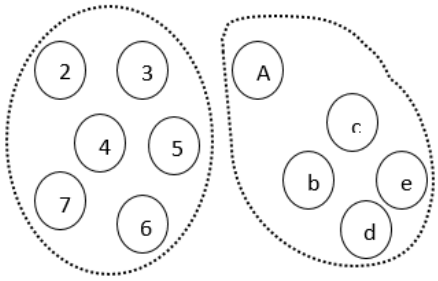

Fig. 2b:- Groups according to income.

Although being the simplest kind of anomaly to be detected yet, a major problem associated with detecting point anomaly is finding a suitable measurement in deviation of the object from other objects.

B. Contextual anomalies

A contextual anomaly *is* also referred to as conditional anomaly. It refers to a data instance which is considered anomalous in a specific context, but not in others. For instance, Fig. 3 shows temperature time series which shows the monthly temperature of an area over last few years. A temperature of 35F might be normal during the winter (at time t1) at that place, but the same value during summer (at time t2) would be an anomaly. Though the temperature at timet1 is same as that at time t2 but occurs in a different context and hence is considered as an anomaly [2].

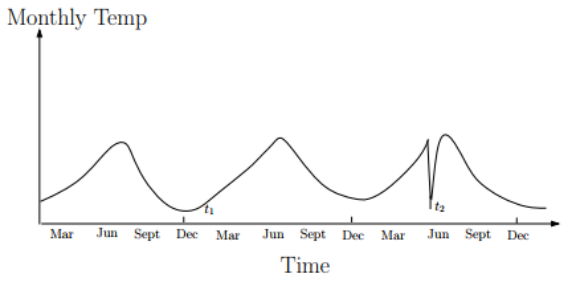

Fig. 3:- Contextual anomaly t2 in a temperature time series

In a contextual anomaly data instance is defined by using contextual attributes and behavioral attributes[26][27][28][29][30], where contextual attributes

ISSN No:-2456-2165

only define the context of the object and behavioral attributes imply the characteristics of an object to identify an anomalous behavior of an object with respect to its context. A contextual anomaly occurs at a certain time or certain region, e.g. large spike at the middle of night.

C. Collective/Group anomalies

A collective anomaly refers to detecting an anomaly that calls a collection of data instances as anomalous with regard to the whole data set. The data instances themselves may perhaps not be anomalous if they do not occur together as a group. Noble & Cook [31] investigated collective anomalies for graph data. They proposed a general framework for the algorithms categorized under various settings: unsupervised versus (semi-)supervised approaches, for static versus dynamic graphs, for attributed versus plain graphs and also highlighted the effectiveness, scalability, generality and robustness aspects of the methods.

Group/collective anomalies occur whenever a collection of data objects as a whole behaves differently than others, whereas the individual data objects within this group may not be anomalous whenever we treat them individually. For example, in Fig.4, we may assume a set of students who reserved as eat for a particular course and if one of them leaves a course, it may be considered as normal but if multiple students start leaving the course then the complete group is considered as anomalous which is represented by G in Figure 4. Collective anomalies are used only for related data instances. They have two variations: events in unexpected order (e.g. breaking rhythm in ECG) and unexpected value combinations (e.g. buying large number of expensive items).

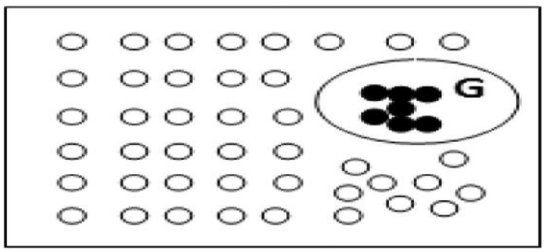

Fig. 4:- Collective anomalies

D. Horizontal anomalies

According to Gao J et al. [32], another type of anomaly has evolved in social networks, which is called horizontal anomaly which depicts the presence of anomalies based upon the different sources of data available. For example, the same user may present in different communities on different social networks. For instance, the horizontal anomaly detection is that for a better business marketing, one wants to find out the person who bought quite different items compared to his peers in the same social community by analyzing data from users' purchase history and friendship networks.

- *Based on static/dynamic nature of network/graph structure*

Considering the network structure being used, Savage D et al.[33]classified anomalies as being either static or dynamic.

A. Static anomalies

A static anomaly occurs with respect to the remainder of the network ignoring the time factor. As static networks allow the changes to happen slowly over time, so only the current behavior of anode is analyzed with respect to others in the network.

B. Dynamic anomalies

Dynamic networks such as mobile applications, allow faster communications and continuous changes in the networks. A dynamic anomaly exists with respect to previous network behavior in which changes occur in the network with the passage of time.

C. Based on information available in network/graph structure

Based on the type of information available at a node or an edge of a graph structure, Savage D et al. [33] categorized anomalies as labeled or unlabeled.

D. Labeled anomalies

Labeled anomalies are related to both the structure of the network and the information gathered from the vertex or edge attributes. There are some other classifications of labeled anomalies:

- Static labelled anomalies

When along with the network structure labels on the vertices and edges are also considered, then the anomalous substructures found are referred to as static labeled anomalies. For example, to detect opinion spam which involves the fake product reviews, static labeled anomalies are used.

- Dynamic labelled anomalies

Dynamic networks are worked upon by considering the structure of the network at fixed time intervals and treating them in the same way as for a static network. When anomalous behavior in a dynamic network is observed by considering labels of the vertices and edges also then the anomalies observed are classified as dynamic labeled anomalies.

E. Unlabeled anomalies

For unlabeled anomalies no attribute of a node or an edge is taken into consideration. It only considers the network structure. Like labeled anomalies unlabeled anomalies is also classified as static and dynamic unlabeled anomalies:

- Static unlabeled anomalies

This type of anomaly occurs when the behavior of an individual remains static and the attributes such as the age of individuals involved, type of interactions, and its duration are

ignored due to unlabeled nature of the network. Only the pattern of interaction took place is important.

- Dynamic unlabeled anomalies

Dynamic unlabeled anomalies can be found when the behavior of the data object is different with respect to the previous time period relative to the network structure. This type of anomaly arises when the dynamic networks change with time. For example, while considering only the pattern of interactions, there is maximum of six ways in which a maximal clique can evolve: shrinking, growing, splitting, merging, appearing or vanishing [34]. All of these involve studying the network structure with respect to the network structure prevalent at some previous time period.

F. Based on behavior

Based on anomalous behavior another two classes of anomalies namely, ‘‘white crow anomalies “and ‘‘in-disguise anomalies” was proposed by Chenet al. [34].

G. White crow anomaly

As the name implies, it is totally an unusual/uncommon thing. It arises when one data object deviates significantly from other observations. For example, while examining the graduating student record, if a record is found where the age of a student is entered as 80years, which is impossible, then it is taken as a white crow anomaly.

H. In-disguise anomaly

It is considered as a small deviation from the normal pattern, defined by Eberle W et al. [35]. For example, anyone attempting to peep into someone’s social network account would not want to get caught; therefore, he will try to behave in the same manner as a normal user. Such anomalies are recognized through strange patterns, which also include uncommon nodes or entity alterations.

I. Based on structural operations on network/graph structure

When dealing with the graphical structures in social networks, anomalies can be classified according to the graphical properties as well. Eberle and Holder[35] classified three possible graph anomalies: insertions, modifications and deletions.

- *Insertion*

Insertion deals with the existence of an unexpected vertex or an edge in the graph.

- *Modification*

Modification deals with the presence of an unexpected label on a vertex or an edge.

- *Deletion*

Deletion involves the absence of an expected vertex or an edge. Sometimes, it even incorporates the concept of dangling edges i.e. with the deletion of a particular vertex all the adjacent edges to it may also have been deleted.

J. Based on interaction pattern in network/graph structure

Based on the types of interaction pattern and links among nodes in graph anomalies can be categorized in three ways [36]:

- *Near Stars/Cliques*

As the presence of completely disconnected (Near Stars) or all connected neighbors (Near Clique) in a graph structure is very uncommon so it is considered as anomalies.

- *Heavy locality*

Sometimes heavyweight around a particular area or a group represents suspicious activity and hence it may suggest the presence of anomaly.

- *Particular dominant links*

An unexpected presence of heavy load at a particular node or link as compared to other node6s or links specifies an unusual activity.

III. OUTPUT OF ANOMALY DETECTION

The main objective of anomaly detection is to detect data points in data that does not fit well with the rest of the data. But it is often very hard to find training data, and even when they are found, most anomalies are in the order of 1:1000 to 1:10^6 events where classes are not balanced. Moreover, in some cases the most of the data are auto correlated. Hence, the precision of the output predicted from a given model (i.e. how likely it is to be true) and recall (how much anomalies the model can catch) trade-offs are different from normal classification use cases. In general, we consider scores and labels as the output of most of the anomaly detection techniques.

- Scores: Scoring technique assigns an anomaly score to each instance in the test data depending on the degree to which that instance is considered as anomaly. Thus the output from this technique is a list of anomalies from a given network which are ranked by their score. Here we can also use a cut-off threshold to select the anomalies.
- Labels: This technique assigns a label to each test instance either as normal or anomalous. Scoring based anomaly detection techniques allow the analyst to use a domain specific threshold to select the most relevant anomalies. Whereas this technique provides binary labels to the test instances which does not allowus to make such a choice for threshold detection.

IV. ANOMALY DETECTION TECHNIQUES IN SOCIAL NETWORKS

The increasing trend of social networks attracted their misuse by a number of malicious individuals as well as in groups. Hence, the detection of anomalous activities becomes

ISSN No:-2456-2165

the need ofhour. Sometimes, it becomes difficult to analyze the social networks because of their large size and complex nature. According to Singh L et al.[37], it becomes necessary to prune the networks to include only the most relevant and significant relationships. Usually, the presence of an anomaly is considered as a binary property in which anomaly is either present or not, and in some applications, the presence of anomaly is considered by giving the degree of being an outlier to each object in the dataset. As an example, Breunig et al. [38] referred this degree as Local Outlier Factor (LOF).

Every online social network has its own structure and nature and there are a variety of techniques to detect anomalous accounts from online social networks. These techniques are evaluated to find the best technique to detect anomalous users from all online social networks like Facebook, Twitter etc. The rest of this section will describe the various techniques associated with anomaly detection in OSN.

Based on the data mining approaches there are three techniques that can be used to detect anomalous users from online social networks [39].

- Supervised Learning Techniques.
- Unsupervised Learning Techniques.
- Semi-supervised Learning Techniques.

A. Supervised anomaly detection techniques

Supervised learning techniques are used to model both normal and abnormal behaviors. These techniques require pre-labeled data for anomaly detection classified as normal or abnormal. Different training models are used to identify the normal or abnormal data from the dataset. Supervised techniques work on two approaches:

- Training model is compared with the dataset to find analogues data from the dataset that is classified as normal data.
- Opposite to above method some anomalous data is compared against training model to find abnormal data from dataset.

B. Unsupervised anomaly detection techniques

Unsupervised anomaly detection techniques are used when labeled data objects are not available i.e. no predefined labels as ‘‘anomalous” or ‘‘normal” are attached to the data objects.

Unsupervised methods are usually studied as a clustering problem. These methods implicitly assume that the normal objects are a bit clustered forming one or more groups with distinct features whereas anomalies do not seem to behave in this manner. However, sometimes this assumption becomes wrong as many anomalies also make clusters with the similar pattern such as collective anomalies as shown in Fig. 4. So, in that case unsupervised methods work inefficiently by issuing a large number of false alarms especially when the normal objects are variedly scattered.

C. Semi-Supervised anomaly detection techniques

In semi-supervised techniques dataset is only labeled with one label as normal. Training model detects abnormal class by itself from dataset [40]. Since they do not require labels for the anomaly class, they are more widely applicable than supervised techniques. So, these methods are used when out of the complete data set only few instances of data labeled as normal are available and from this small amount of labeled data classifier can be constructed which then tries to label the rest of the unlabeled data.

There are so many approaches based on the three data mining approaches discussed above which are highly helpful in determining anomalies in social networks. In this section some of the prominent approaches with their advantages and disadvantages are discussed:

- Proximity based techniques
- Clustering based techniques
- Classification based techniques
- Behaviour-based techniques
- Structure based techniques
- Spectral based techniques
- Graph based techniques

A. Proximity based (or nearest neighbour based) anomaly detection techniques

The key idea of proximity based approach is, an object is anomalous if it is distant from most point. One of the simplest ways to measure whether an object is distant from most point is to use the distance to the k-nearest neighbor and the outlier score of an object is given by the distance to its k-nearest neighbor. However, a number of k-nearest neighbor methods can be used which make use of various measures such as distance, density and other similarity measures to determine the proximity between the nodes. Proximity-based methods can be mainly classified into the following two categories:

- *Distance based*

This method computes the anomaly score by using the distance of a data object to its k neighbors. Distance-based anomalies are considered as ‘‘global anomalies”. Generally, Euclidean or Mahalanobis distance is taken as the distance metrics.

- *Density-based*

Density-based approaches work by comparing the density of an object with density around its neighbors and

ISSN No:-2456-2165

computes anomaly score by using the relative density of each data object. For a normal object, both densities are assumed to be same whereas for anomalous objects they are different. The concept of relative density is often used to measure the degree of anomalous behavior of an object. Density-based methods overcome the difficulties associated with distance-based methods in detecting local anomalies.

B. Cluster based anomaly detection techniques

As stated by Berkhin[41], clustering is considered as an unsupervised learning of a hidden data concept. Cluster-based methods follow a simple assumption that usually anomalies either belong to a small sparse cluster or do not belong to any cluster whereas the normal objects are part of large and dense clusters. These approaches consider the presence of anomaly in the following three cases:

- If the object does not belong to any cluster.
- If the distance between object and cluster to which it is closest is large.
- If the object is a part of a small or sparse cluster, then not only the object but all the objects belonging to that cluster are considered as anomalous.

Clusters of the data objects can be constructed using numerous methods such as K-Means, K-Medoids[42] for small data sets; CLARA [43], CLARANS [44] for large data sets and BIRCH [45], Chameleon [46] for performing macro clustering on micro clusters.

*C. Classification based anomaly detection techniqu*es

Classification is defined by J. Han[47] as a supervised method with two steps: a learning step and a classification step. In the learning stage a trained set of labeled data instances are used to construct a classification model and in the classification step, the constructed model is used to predict the class labels for the data. Both the steps are respectively stated in the training and the testing stages. Classification based approaches can use either a one class model or a multiclass model[42].

One-class model help to identify new anomalous objects that are far from the other anomalous objects present in the given training dataset and the multiclass model is used when the available data objects not only belong to a single class but to multiple classes. Some of the examples of one class models used for anomaly detection are one-class SVM [48], Gaussian model description (GAUSSD) [42], Principal component analysis description (PCAD) [42], Parzen window classifier (PWC) [42] etc. In each of them, a decision boundary is set up. The data objects falling outside the decision boundary are treated as anomalous[49].

D. Behaviour-based anomaly detection techniques

Behavior-based techniques handle the behavioral properties of the users such as number and content of messages, number of likes or comments on a post, duration of a conversation, the content of the items shared, the status of the users etc. Some of the popular behavior-based techniques are discussed below:

- *Content-based filtering*

Content-based filtering is one of the prominent and well-known behavior-based approach in which anomalous behavioris detected by looking at the internal content of the sent and received messages. Vanetti et al. [50]proposed a Filtered Wall system in which certain set of filtering rules were used by the users to avoid unwanted and irrelevant posts from their walls. A Blacklist (BL) may be created using these filtering rules in which a number of constraints are imposed like ‘constraint on message creators’, ‘constraints on message contents’, and ‘action taken in the form of blocking, publishing or notification’ to handle attack. However, some smart malicious users are intelligent enough to bea fool and deceive others by behaving similarly to the legitimate users.

In social network scenario, two famous attacks, called Sybil attacks and cloning attacks are quite popular nowadays [51]. Though a number of techniques have been projected to handle such type of attacks yet most of them seem to fail because of one or the other reason. Some of the simple techniques such as clustering coefficient (CC) and voting scheme are botched by the spammers by behaving or creating a similar network structure to that of a normal user. In Clustering coefficient method, for normal users the value of clustering coefficient is high whereas that for spammers is close to 0. But in order to present themselves as legitimate, the spammers increase their CC value by making the neighborhood structure similar to that of the genuine users. Similarly, in voting schemes the illegitimate users make a number of fake profiles to increase the voting of a target post in the form of likes, views etc. or to avoid being classified as spam during voting. Even the advanced techniques such as honeypots proposed by Dagon D et al. [52]to detect the spammers fail to attract anomalous users in most of the situations.

- *Principal Component Analysis*

Recently, an unsupervised statistical anomaly detection technique known as Principle component analysis (PCA) was used by Viswanath et al. [26] to detect the anomalous behavior in individuals. Unlabeled Facebook dataset was used and a number of fake and compromised users were identified. The criteria for normal and anomalous distributions were judged by observing the ‘like’ activities of the users, such as by studying the pages ‘liked’ by a user or the number of posts/pages liked by the user at a particular time period. Besides, a significant contribution made by them was the detection of click spams which is highly prevalent nowadays in ads where the users are unintentionally made to click on the spam links which seem to be genuine or some sort

of malware hacks a person's account and clicks 'likes', posts comments or reviews without the knowledge of the user. By experiments, it was inferred that most of the clicks on such sites were done by anomalous users. Xiao et al.[53] used the profile information of users to detect fake accounts in online social networks using certain supervised machine learning techniques for feature extraction and cluster building. It is one of the efficient and faster methods so far to identify fake accounts.

E. Structure-based anomaly detection techniques

Structure-based methods work on the basic principle of using structural properties to find out the normal and anomalous users by checking their characteristics. A particular graph metric is figured out for different nodes or structures and the nodes showing different values than other users are considered as anomalous. The properties or metrics used may range from the simple properties such as the number of nodes, edges to highly complex centrality measures. Just like supervised learning, here also a predefined normal pattern is already known and any deviation from that known pattern depicts the anomalous behavior.

The structural properties have been used by most of the researchers working in social network domain to define a number of new approaches for identifying anomalies in online social networks. As an example, Link mining, used by Getoor and Diehl [54] studies the structural properties of the networks to predict different behaviors of individuals in social networks. They covered eight link mining tasks with their respective algorithms and grouped the defined tasks into three categories, namely object-related, link-related and graph-related. By analyzing the association between different nodes it is usually found that the linked objects often have a set of correlated attributes. In other words, connectivity of two users can be checked by examining the common properties and what is usually observed is that the objects sharing some sort of common features are often found to be linked with each other. However, most of the structure based link prediction methods show poor performance because of the assumption of prediction of future relationships likely to occur [55]. Earlier Rattigan et al. [56] proposed another advanced task such as anomalous link discovery (ALD), which involve only the prediction of anomalous relationships rather than all the involved relationships. It was seen that almost every prediction model performed quite well for ALD.

In social networks, link prediction is highly useful for detecting friendship links between different users as such techniques are a good way to examine connected, missing and corrupted links [57]. These techniques help to identify dynamic unlabeled anomalies by predicting future events and analyzing previous network behavior which is a prerequisite for dynamic anomalies. Shrivastava et al. [58]proposed a generic approach for detection of attacks, named as Random Link Attacks (RLAs). The basic motive behind such an attack resembles that of the Sybil attacks. These attacks are quite prominent in email spams, virtual marketing etc. with a fact that the victims are chosen randomly with each one having the equal probability to be a victim. This helps to analyze and detect the attacks efficiently as for an attacker by assuming that the structure of a set of random nodes in its neighborhood will be quite different from that of a good node. A set of two properties namely, a clustering test and a neighborhood independent test are conducted on the suspicious nodes which after creating groups mark them as anomalous. Two heuristic algorithms GREEDY and TRWALK algorithm were proposed to detect the attackers.

Many already existing node-based and egonet-based features were studied recursively by Henderson et al.[59]. Some aggregate values were calculated on the already existing characteristics. Neighborhood information was retrieved using both node and egonet-based features and behavioral information was extracted using recursive features. Akoglu et al.[36]utilized another structure-based approach in which a number of pattern and law discoveries were used to detect different types of anomalies in social network graph. To spot some abnormal nodes especially in weighted graphs an Oddball algorithm was proposed by them. A set of features were grouped into certain set of carefully chosen pairs and anomalous behavior was analyzed by examining the group structure. Groups were formed where the patterns of normal behavior (power laws) were observed and the points deviated from discovered patterns were considered as anomalous. A number of anomalous relationships were also observed namely Near Stars or Near Cliques, Heavy Vicinities and Dominant Edges. Hassanzadeh et al. [60]used the power laws defined in Oddball algorithm to analyze the relationship among various social network metrics, thereby detecting the anomalous relationships between different users. Among the different metrics used it was seen that the relationship between number of edges and average betweenness centrality of a user's direct neighborhood helped to better predict the anomalous nodes. Similarly, Rezaei et al. [61] used the same approach for some Twitter dataset and predict Near Star/Clique behavior by analyzing the number of nodes and edges behavior.

F. Spectral based anomaly detection techniques

Spectral anomaly detection techniques help in detecting anomalies using some spectral characteristics in the spectral space of a graph. Different complex measures such as eigenvalues or eigenvectors applicable to the adjacency matrix[13] or different hyper graph algorithms used for Laplacian graphs [62] are focused on these methods. In most of the techniques, a social network graph is partitioned into different groups or communities and this partitioning is done either by eliminating the links between different nodes or by using certain clustering/classification algorithms and measures. Even some of the advanced techniques use the structural concept of centrality. For example, community structures were worked upon by Girvan and Newman [63]. As

shown in Fig. 5, communities in the form of different friendship groups were created in which the strength of links between the nodes within a community or friendship group is dense whereas among different groups is sparse.

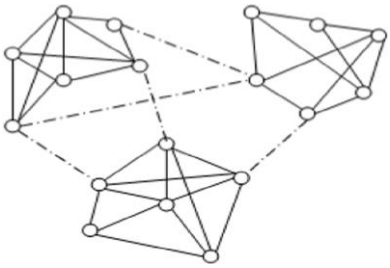

Fig. 5:- Friendship links depicting centrality also

The concept of betweenness centrality formulated by Freeman [64] is modified to work for edges instead of vertices to find the number of shortest paths between a set of vertices that pass through the edge under consideration. The edges with the high value of betweenness centrality state the points where a network is expected to break and hence are separated. Generally, in online social networks, high betweenness centrality is found to be at the intersection of densely connected network groups. As a result, a number of significant groups could be determined by removing the set of links from a graph, a concept also used by Newman [65]. Ying et al. [13] identified the malicious nodes by computing the spectral coordinates or the spectra i.e. the eigenvalues or eigenvectors for both the normal and anomalous user with a special reference to RLA's. The use of RLA's was stressed upon because of the absence of prior knowledge regarding which node is the attacker and which one the victim node. The presence of fake links or nodes affects the value of the graph spectra. Spectral coordinates of a victim node are used to analyze the interdependency between the victim and the attacker nodes, thereby calculating the spectral coordinates for attacking nodes. It was observed that malicious users govern the attack set and each attacking node is linked to a number of victim nodes as shown in Fig. 6.

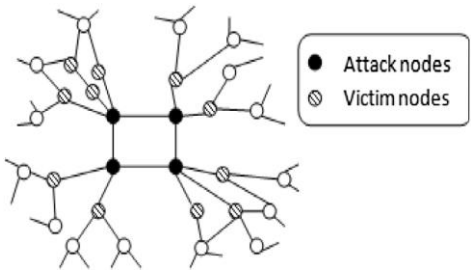

Fig. 6:- Describing relationship between attacking and victim nodes

G. Graph-based anomaly detection techniques

Based on the nature and type of anomaly being studied a variety of graph-based techniques have been proposed and implemented in the social network domain. Savage et al. [33] surveyed on different techniques for each of the static/dynamic unlabeled/labeled anomalies. For detecting dynamic unlabeled anomalies some new techniques such as Bayesian analysis and scan statistical approaches (mainly applicable to hyper-graphs) are used. In case of labeled anomalies, a number of techniques have been proposed for static and dynamic networks. As an example, for the detection of opinion spam a belief propagation method has been applied which deals with a set of hidden labels. Another approach called Trust Rank were proposed by Z. Gyöngyi et al. [66]. The fundamental principle behind this method is that trustworthy pages are unlikely to be linked immediately or within a predefined range to spam pages. One of the prominent methods employed for such static labeled anomalies is the use of information theory, a quantitative measure such as entropy to detect the anomalies. In most of the approaches, the network structure is considered as static for a fixed time period and in order to add the dynamic concept the behavior of different nodes/modules is compared at different time intervals. Akoglu et al. [5]also surveyed on different graph-based anomaly detection methods for static/dynamic and labeled/unlabeled constraints. In each network structure, different quantitative and qualitative techniques have been categorized into different sub-modules such as structure-based, window based, community-based and feature based. Moreover, researchers have described number of real-world applications where graph-based anomaly detection methods could be fit, for example, opinion spams, auction networks, social networks, telecommunication networks, cybercrimes, security networks, trading networksetc. Recently, there has been an inclination toward detecting anomalies in dynamic networks. So, a number of researchers are adding dynamic concept into their research work. For example, a number of anomaly detection techniques especially related to dynamic networks are recently surveyed by Ranshous et al. [67]. For instance, a scoring function is used to identify various types of anomalies. Categorization of anomalous behavior is based upon the scoring function being used along with the application area under consideration. Vigliotti and Hankin[68] used the most significant and pertinent subset of nodes to detect anomalous patterns in dynamic networks. In their work, the experiments were performed on the temporal networks.

Lately, community outliers have gained much attention and a number of approaches have been proposed for them. Harenberg et al. [69] studied various disjoint and overlapping community detection techniques used in large-scale networks. Disjoint communities involve the participation of an individual node in at most one community whereas in overlapping communities a node can participate in multiple communities. For the detection of disjoint communities

ISSN No:-2456-2165

different clustering or graph partitioning algorithms are frequently used. Similarly, the detection of overlapping communities makes use of various block modeling, clustering, or clique extraction methods. Gao et al. [70]also worked a lot in the field of detecting community anomalies differentiating them from local and global anomalies. A simple approach to detect community anomalies is to make use of the approaches used for both the local and global anomalies i.e. DNODA (for local anomalies) and GLODA (for global anomalies).

Gao et al.[70] proposed an advancement in the above approach by integrating both the network and data object information to detect the community anomalies. The proposed approach is called Community Outlier Detection algorithm(CODA) which makes use of a probabilistic mixture model designed for multivariate data objects (objects with multiple attributes). Statistical anomaly detection approaches were used to detect the community anomalies where different distributions were analyzed on the associated data and it is assumed that normal data objects will follow the defined distribution whereas anomalous objects will deviate from it or follow some other distributions. A set of hidden variables for data objects, and HiddenMarkov Random Field (HMRF) for the network links, are worked upon by the defined ICM and EM-based algorithms. In order to make it more effective a set of hyper graphparameters like, threshold (indicating few anomalies for its high value and more anomalies for the low value), link importance(for the prediction of confidence level), number of components(small determining global anomalies and large the local ones) were also defined and used.

TABLE 1 summarizes the advantages and disadvantages of different anomaly detection techniques that have been discussed in this section

Anomaly Detection Techniques	Advantages	Disadvantages
PROXIMITY (NEAREST NEIGHBOR) BASED • Using Distance to *kth* Nearest Neighbor • Using Relative Density	1. Proximity-based techniques are unsupervised in nature and are purely data-driven. 2. In terms of missed anomalies semi-supervised techniques perform better than unsupervised techniques. 3. Adapting NN based techniques to a different data type is straightforward, and primarily requires defining an appropriate distance measure for the given data.	1. Handling and detection of anomalies become difficult when we have several regions with widely differing densities. 2. It becomes difficult to detect the group anomalies if they are present close to each other. 3. Proximity-based methods are highly dependent on the proximity measures used for their efficient working which might not be available in certain situations.
CLUSTERING BASED	1. Cluster based techniques are basically an unsupervised technique. 2. This kind of techniques can sometimes be adapted to other complex data types by simply plugging in a clustering algorithm that can handle the particular data type. 3. The testing phase for clustering based techniques is fast.	1Incur high computation cost when the clusters are to be found before detecting anomalies. 2. Computational complexity for such methods is highest of all the data mining methods applied. 3. Clustering approaches are a costly procedure for large datasets. 4. Sometimes clustering process involves anomalous objects depicting similar behavior and hence forming the clusters. 5. As anomalies follow a presumption to be belonging to either no cluster or a small cluster, so, objects in the above encountered clusters might be considered as normal.
CLASSIFICATION BASED • Neural Networks Based • Bayesian Networks Based • Support Vector Machines Based • Rule-Based	1. Use of powerful algorithms can distinguish between instances belonging to different classes. 2. The testing phase of classification based techniques is fast.	1. Multi-class classification techniques depend on the availability of accurate labels for various normal classes, which is sometimes difficult to determine. 2. Classification based techniques assign a label to each test instance, which can also become a disadvantage when a meaningful anomaly score is desired for the test instances. 3. One of the concerns for the classification method is the heavy dependency and reliability on training data which if not properly available may lead to the degradation of performance. 4. Many of the times we may encounter a class

		imbalance problem in which only a few objects represent the main class.
BEHAVIOR BASED	1. This technique is a faster and efficient to identify fake accounts as it only uses the attributes entered by a user during registration i.e. profile creation. 2. The employed technique is a first in its form to detect the clusters of fake accounts usually created by a single user on a particular social network thereby superseding the existing techniques which only work and make a deduction for a single account.	1. Though this technique can detect the clusters of fake accounts usually created by a single user on a particular social network, this system was found to restrict around 2,50,000 fake accounts. 2. Some smart malicious users are intelligent enough to befool and deceive others by behaving similarly to the legitimate users. For example, in social network scenario, two of the famous attacks called Sybil attacks and cloning attacks are quite popular nowadays [42]. Though a number of techniques have been projected to handle such type of attacks yet most of them seem to fail because of one or the other reason.
STRUCTURE BASED	1. These techniques help to identify dynamic unlabeled anomalies by predicting future events and analyzing previous network behavior which is a prerequisite for dynamic anomalies.	1. Most of the structure based link prediction methods show poor performance because of the involvement of prediction of future relationships likely to occur. Earlier also a number of advanced tasks such as anomalous link discovery (ALD) were proposed which involved only the prediction of anomalous relationships rather than all the involved relationships [56].
SPECTRAL BASED	1. Spectral based techniques primarily focused on dimensionality reduction features and so these techniques are suitable for handling high dimensional datasets. Moreover, they can also be used as a pre-processing step followed by application of any existing anomaly detection technique in the transformed space. 2. Spectral techniques can be used in an unsupervised setting.	1. Spectral techniques are useful only if the normal and anomalous instances are separable in the lower dimensional embedding of the data. 2. Spectral techniques typically have high computational complexity.
GRAPH-BASED	1. Majority of the relational data can be thought of as inter-dependent, which may help to find anomalies in large interconnected networks. 2. Problems in anomaly detection domain are mostly relational in nature. The nature of anomalies could exhibit themselves as relational. These phenomena can be easily and efficiently modeled as a graph. 3. Graph has a powerful representation which is used to represent inter-dependencies by the introduction of links (or edges) between the related objects very efficiently. 4. Representation of rich datasets is permitted by the graphical representation which enables the incorporation of the node and edge attributes. 5. Adversarial robustness, which is very important, is provided by graph.	1. Though these techniques require the use of time-window for feature extraction, discovering of different types of outliers in the graph sequence or computation of the normal graph node activity, but it is very difficult to choose the window size. 2. Most methods ignore the cost aspects of information. These costs, on the other hand, may exhibit themselves in various forms with varying levels.

Table 1:- Advantages & Disadvantages of different Anomaly Detection Techniques

Table 2- summarizes the different existing algorithms for anomaly detection techniques for three primary anomaly types:

Point Anomaly Detection
• Activity-based Point Anomaly -Bayes one-step Markov[Schonlau et al. (2001)] [71], multi-step Markov [Ju and Vardi (2001)] [72] -Compression [Schonlau et al. (2001)] [71] -Poisson process [Ihler et al. (2006)] [73] -Probabilistic suffix tree (PST) [Sun et al. (2006)] [74] -Temporal dependence [Qiu et al. (2012)] [75] • Graph-based Point Anomaly (Static graph) -Power law [Akoglu and McGlohon (2010)] [36]; [Akoglu et al. (2015)] [5] -Random walk [Moonesinghe and Tan (2008)] [76]; [Sun et al. (2008)] [77] -Hyper-graph [Silva and Willett (2008b,a)][78][79] -Spatial auto-correlation [Sun and Chawla (2004)] [80] -Time series data analysis of graph -ARIMA process [Pincombe (2005)] [81] -Graph eigenvectors [Ide and Kashima, 2004)] [82] -Graph Scope: Minimum description length (MDL) [Sun et al. (2007)] [83] -Window based approach: Scan statistics [Park et al., (2008)] [84]
Collective/Group Anomaly Detection
• Activity-based Group Anomaly Detection -Scan statistics [Das et al. (2009)] [48] -Density estimation -Multinomial genre model (MGM) [Xiong et al. (2011a)] [85] -Flexible genre model (FGM) [Xiong et al. (2011b)] [86] -Group Latent Anomaly Detection model(GLAD) [Rose et al. (2015)] [87] -One class support measure machine (OCSMM) [Muandet and Scholkopf (2013)] [88] • Static Graph-based Group Anomaly Detection -Minimum description length (MDL) [Chakrabarti (2004)] [89]; [Lin and Chalupsky (2003)] [90]; [Rattigan and Jensen (2005)] [56] -Anomalous substructure [Noble and Cook (2003)] [31]; [Eber le and Holder (2007)] [35] -Tensor decomposition [Maruhashi et al. (2011)] [91] • Dynamic Graph-based Group Anomaly Detection -Bipartite graph [Friedland and Jensen (2007)] [92]; [Liu et al. (2008)] [93] -T-partite graph [Xu et al. (2007)] [94]; [Kim and Han 2009)] [95] -Counting process [Heard et al. (2010)] [96]
Contextual Anomaly Detection
Contextual anomalies are calculated by focusing on segments of data (e.g. spatial area, graphs, sequences, customer segment) and applying collective anomaly techniques within each segment independently.

Table 2:- List of Algorithms for Anomaly Detection

V. CHALLENGES IN SOCIAL MEDIA ANLOMALY DETECTION

Most existing approaches to anomaly detection suffer from a series of following shortcomings:

- Sensitiveness: high false alarm rate.
- Interpretation: statistical test results with very limited insights about the detected anomaly.
- Scalability: challenging for high-dimensional streaming data.
- Heterogeneous data with rich and complex information.
- Beyond the typical iid assumptions.
- Very limited labelled examples or benchmark datasets.
- Varieties and dynamics in anomalies.

Besides, there are some other challenges especially for group anomaly detection:

- Two forms of data co-exist in social media: one is the point-wise data, which characterize the features of an individual person. The other is pair-wise relational data, which describe the properties of social ties. So, it is important to take into account both point-wise and pair-wise data during anomaly detection. For example, teams with the same composition of member skills can perform very differently depending on the pattern of relationships among the members [97].
- Group anomaly is usually more precise than an individual anomaly. At the individual level, the activities might appear to be normal [98], but when we consider a group the same activities may appear as abnormal. Therefore, existing anomaly detection algorithms that deal with individual or point anomaly usually fails when it is related to a group.
- Empirical studies in social media analysis suggest the dynamic nature of individual network positions [99]. People's activities and communications changes constantly over time and we can hardly know the groups beforehand. Thus developing method that can be easily generalized to the dynamic setting is critical to anomaly detection that evolves social media data.

VI. CONCLUSION

This paper has reviewed the basic concepts of anomalies, different types of anomalies in social networks, and various possible anomaly detection techniques with their advantages and disadvantages. In this survey we have tried to provide a detailed discussion on different ways in which the problem of anomaly detection has been formulated in literature, and also have attempted to give a literature review of the various techniques. As it is very difficult to cover each and every technique in this review paper, best efforts have been made to cover the most important ones. Ideally, the main purpose of a comprehensive survey on anomaly detection is not only to understand particular or different existing anomaly detection techniques, but also to provide a comparative study on various techniques which we have tried to present in this paper. In spite of enormous work that has been done so far in the anomaly detection domain there remains a number of shortcomings that could be addressed and worked upon in future. There are also few works that have been done so far on contextual and collective anomaly detection techniques in several domains, so there are huge scopes for development of new techniques/insights in this area, especially using graph theory. Another upcoming area where anomaly detection is finding more and more applicability is in complex systems. An example of such system would be an aircraft system with multiple components. Anomaly detection in such systems involves modeling the interaction between various components [2]. There alsoremains a scope for the exploration of graph metrics in behavior based, structure-based or spectral based anomaly detection techniques that could be used to detect some new kinds of anomalies present in online social networks which are still undiscovered.

REFERENCES

[1] B. D. Patil and P. R. K. Bedi, "Survey on Anomaly Detection Techniques in Social Networking," vol. 3, no. 11, pp. 1573–1576, 2014.
[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Comput. Surv., vol. 41, no. September, pp. 1–58, 2009.
[3] H. Kriegel, E. Schubert, and A. Zimek, "LoOP: Local Outlier Probabilities," dl.acm.org, pp. 1649–1652.
[4] A. Zimek, E. Schubert, and H. P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," Statistical Analysis and Data Mining, vol. 5, no. 5. pp. 363–387, 2012.
[5] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," Data Min. Knowl. Discov., vol. 29, no. 3, pp. 626–688, May 2015.
[6] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decis. Support Syst., vol. 50, no. 3, pp. 602–613, 2011.
[7] C. Phua, A. Damminda, and L. Vincent, "Minority Report in Fraud Detection: Classification of Skewed Data," Acm sigkdd Explor. Newsl., vol. 6, no. January, pp. 50–59, 2004.
[8] C. Castillo, D. Donato, A. Gionis, V. Murdock, and F. Silvestri, "Know your neighbors: web spam detection using the web topology," in Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, 2007, pp. 423–430.
[9] M. Ott, C. Cardie, and J. Hancock, "Estimating the prevalence of deception in online review communities," in Proceedings of the 21st international conference on World Wide Web - WWW '12, 2012, p. 201.
[10] M. Kumar, R. Ghani, and Z.-S. Mei, "Data mining to predict and prevent errors in health insurance claims processing," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10, 2010, p. 65.
[11] M. McGlohon, S. Bay, M. G. M. Anderle, D. M. Steier, and C. Faloutsos, "SNARE: A Link Analytic System for Graph Labeling and Risk Detection," Kdd-09 15Th Acm Sigkdd Conf. Knowl. Discov. Data Min., pp. 1265–1273, 2009.
[12] N. Abe et al., "Optimizing debt collections using constrained reinforcement learning," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10, 2010, p. 75.

[13] X. Ying, X. Wu, and D. Barbará, "Spectrum based fraud detection in social networks," in Proceedings - International Conference on Data Engineering, 2011, pp. 912–923.
[14] T. Fawcett and F. Provost, "Combining Data Mining and Machine Learning for Effective User Profiling," in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), 1996, pp. 8–13.
[15] T. Fawcett and F. Provost, "Activity monitoring: Noticing interesting changes in behavior," Proc. fifth ACM SIGKDD Int. Conf. Knowl. Discov. data Min., vol. 1, no. 212, pp. 53–62, 1999.
[16] B. J. Jansen and U. 2007, "Click fraud," ieeexplore.ieee.org.
[17] N. Kshetri, "The economics of click fraud," IEEE Secur. Priv., vol. 8, no. 3, pp. 45–53, 2010.
[18] J. Neville, Ö. Şimşek, D. Jensen, J. Komoroske, K. Palmer, and H. Goldberg, "Using relational knowledge discovery to prevent securities fraud," in Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining - KDD '05, 2005, p. 449.
[19] R. Hassanzadeh, "a Nomaly D Etection I N O Nline S Ocial N Etworks : U Sing Data - Mining T Echniques and F Uzzy," no. November, 2014.
[20] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: A survey," Data Min. Knowl. Discov., vol. 29, no. 3, pp. 626–688, 2015.
[21] W. Eberle and L. Holder, "Graph-Based Approaches to Insider Threat Detection," dl.acm.org.
[22] U. Damnjanovic, F. Virginia, E. Izquierdo, and J. Martínez, "Event detection and clustering for surveillance video summarization," in 9th international workshop on image analysis for multimedia interactive services, 2008, pp. 63–66.
[23] B. Krausz and R. Herpers, "MetroSurv: detecting events in subway stations," Multimed. Tools Appl., vol. 50, no. 1, pp. 123–147, Oct. 2010.
[24] U. Flegel, J. Vayssiere, and G. Bitz, "A State of the Art Survey of Fraud Detection Technology," Insid. Threat. Cyber Secur., vol. 49, pp. 73–84, 2010.
[25] S. Wang, "A comprehensive survey of data mining-based accounting-fraud detection research," in 2010 International Conference on Intelligent Computation Technology and Automation, ICICTA 2010, 2010, vol. 1, pp. 50–53.
[26] B. Viswanath et al., "Towards Detecting Anomalous User Behavior in Online Social Networks," 23rd USENIX Secur. Symp., pp. 223–238, 2014.
[27] P. Bogdanov, M. Busch, J. Moehlis, A. K. Singh, and B. K. Szymanski, "The Social Media Genome-Modeling Individual Topic-Specific Behavior in Social Media," 2013 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min., pp. 236–242, 2013.
[28] M. A. Ahmad, B. Keegan, A. Roy, D. Williams, J. Srivastava, and N. Contractor, "Guilt by Association? Network Based Propagation Approaches for Gold Farmer Detection," ieeexplore.ieee.org, pp. 121–126, 2013.
[29] S. O'Banion and L. Birnbaum, "Using explicit linguistic expressions of preference in social media to predict voting behavior," 2013 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min., pp. 207–214, 2013.
[30] B. Pan, L. Zhang, and K. Smith, "A Mixed-Method Study of User Behavior and Usability ON an Online Travel Agency," Inf. Technol. Tour., vol. 13, no. April, pp. 353–364, 2012.
[31] C. Noble, … D. C. conference on K. discovery, and undefined 2003, "Graph-based anomaly detection," dl.acm.org.
[32] J. Gao, N. Du, W. Fan, D. Turaga, S. Parthasarathy, and J. Han, "A multi-graph spectral framework for mining multi-source anomalies," in Graph Embedding for Pattern Analysis, New York, NY: Springer New York, 2013, pp. 205–227.
[33] D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Anomaly detection in online social networks," Soc. Networks, vol. 39, no. 1, pp. 62–70, 2014.
[34] Z. Chen, W. Hendrix, and N. F. Samatova, "Community-based anomaly detection in evolutionary networks," J. Intell. Inf. Syst., vol. 39, no. 1, pp. 59–85, 2012.
[35] W. Eberle and L. Holder, "Anomaly detection in data represented as graphs," Intell. Data Anal., vol. 11, pp. 663–689, 2007.
[36] L. Akoglu, M. McGlohon, and C. Faloutsos, "OddBall: Spotting anomalies in weighted graphs," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2010, vol. 6119 LNAI, no. PART 2, pp. 410–421.
[37] L. Singh, L. Getoor, and L. Licamele, "Pruning social networks using structural properties and descriptive attributes," Data Mining, Fifth IEEE Int. Conf., 2005.
[38] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying Density-Based Local Outliers," Proc. 2000 Acm Sigmod Int. Conf. Manag. Data, pp. 1–12, 2000.
[39] V. J. Hodge and J. Austin, "A Survey of Outlier Detection Methodoligies," Artif. Intell. Rev., vol. 22, no. 1969, pp. 85–126, 2004.
[40] C. Aggarwal and P. Yu, "Outlier detection for high dimensional data," Icdm, 2012.
[41] P. Berkhin, "Survey of Clustering Data Mining Techniques," Group. Multidimens. Data Recent Adv. Clust., pp. 25–71, 2006.
[42] R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," Egypt. Informatics J., vol. 17, no. 2, pp. 199–216, 2016.
[43] L. Kaufman, … P. R. in data: an introduction to, and U. 2008, "Clustering Large Applications (Program

CLARA),” in Wiley Online Library, pp. 126–163.
[44] R. T. Ng and J. Han, “CLARANS: A method for clustering objects for spatial data mining,” IEEE Trans. Knowl. Data Eng., vol. 14, no. 5, pp. 1003–1016, 2002.
[45] D. HaiZhou and L. YongBin, “An Improved BIRCH Clustering Algorithm and Application in Thermal Power,” Web Inf. Syst. Min. (WISM), 2010 Int. Conf., vol. 1, pp. 53–56, 2010.
[46] G. Karypis, E. Han, and V. Kumar, “Chameleon: Hierarchical clustering using dynamic modeling,” Computer (Long. Beach. Calif)., vol. 32, no. 8, pp. 68–75, 2002.
[47] J. Han, “Data mining : concepts and techniques,” J. Chem. Inf. Model., pp. 101–103, 2012.
[48] K. Das and J. Schneider, “Detecting anomalous records in categorical datasets,” in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’07, 2007, p. 220.
[49] R. Kaur and S. Singh, “A survey of data mining and social network analysis based anomaly detection techniques,” Egypt. Informatics J., vol. 17, no. 2, pp. 199–216, Jul. 2016.
[50] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, “Content-based filtering in on-line social networks,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2011, vol. 6549 LNAI, pp. 127–140.
[51] S. Y. Bhat and M. Abulaish, “Communities A gainst Deception in Online Social Networks 1 The Platform 2 The Mischef,” abulaish.com, vol. 2014, no. 2, pp. 8–16, 2014.
[52] D. Dagon et al., “HoneyStat: Local Worm Detection Using Honeypots,” Recent Adv. Intrusion Detect., pp. 39–58, 2004.
[53] C. Xiao, D. M. Freeman, and T. Hwa, “Detecting Clusters of Fake Accounts in Online Social Networks,” in Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security - AISec ’15, 2015, pp. 91–101.
[54] L. Getoor and C. P. Diehl, “Link Mining: A Survey,” dl.acm.org.
[55] D. Liben-Nowell, J. K.-S. and Technology, and undefined 2007, “The link-prediction problem for social networks,” Wiley Online Libr.
[56] M. Rattigan, D. J.-A. S. E. Newsletter, and U. 2005, “The case for anomalous link discovery,” dl.acm.org.
[57] E. Zheleva, L. Getoor, J. Golbeck, and U. Kuter, “Using friendship ties and family circles for link prediction,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2009, vol. 5498 LNAI, pp. 97–113.
[58] N. Shrivastava, A. Majumder, and R. Rastogi, “Mining (social) network graphs to detect random link attacks,” in Proceedings - International Conference on Data Engineering, 2008, pp. 486–495.
[59] K. Henderson, B. Gallagher, T. Eliassi-rad, and C. Faloutsos, “It ’s Who You Know: Graph Mining Using Recursive Structural Features,” Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. data Min., pp. 663–671, 2011.
[60] R. Hassanzadeh, R. Nayak, and D. Stebila, “Analyzing the effectiveness of graph metrics for anomaly detection in online social networks,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, vol. 7651 LNCS, pp. 624–630.
[61] A. Rezaei, Z. M. Kasirun, V. A. Rohani, and T. Khodadadi, “Anomaly detection in Online Social Networks using structure-based technique,” 2013 IEEE Third Int. Conf. Inf. Sci. Technol., no. October, pp. 619–622, 2013.
[62] S. Agarwal, K. Branson, and S. Belongie, “Higher order learning with graphs,” in Proceedings of the 23rd international conference on Machine learning - ICML ’06, 2006, pp. 17–24.
[63] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” Proc. Natl. Acad. Sci., vol. 99, no. 12, pp. 7821–7826, Jun. 2002.
[64] L. C. Freeman, “A Set of Measures of Centrality Based on Betweenness,” Sociometry, vol. 40, no. 1, p. 35, Mar. 1977.
[65] M. E. J. Newman, “Detecting community structure in networks,” Eur. Phys. J. B - Condens. Matter, vol. 38, no. 2, pp. 321–330, Mar. 2004.
[66] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen, “Combating web spam with TrustRank,” Proc. Thirtieth Int. Conf. Very large data bases, vol. 30, pp. 576–587, 2004.
[67] S. Ranshous, S. Shen, D. Koutra, S. Harenberg, C. Faloutsos, and N. F. Samatova, “Anomaly detection in dynamic networks: a survey,” Wiley Interdiscip. Rev. Comput. Stat., vol. 7, no. 3, pp. 223–247, May 2015.
[68] M. G. Vigliotti and C. Hankin, “Discovery of anomalous behaviour in temporal networks,” Soc. Networks, vol. 41, pp. 18–25, 2015.
[69] S. Harenberg et al., “Community detection in large-scale networks: a survey and empirical evaluation,” Wiley Interdiscip. Rev. Comput. Stat., vol. 6, no. 6, pp. 426–439, Nov. 2014.
[70] J. Gao, F. Liang, W. Fan, C. Wang, Y. Sun, and J. Han, “On community outliers and their efficient detection in information networks,” in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’10, 2010, p. 813.
[71] M. Schonlau, W. DuMouchel, W. Ju, A. Karr, M. T.-S. science, and undefined 2001, “Computer intrusion: Detecting masquerades,” JSTOR.
[72] W.-H. Ju and Y. Vardi, “A Hybrid High-Order

Markov Chain Model for Computer Intrusion Detection," J. Comput. Graph. Stat., vol. 10, no. 2, pp. 277–295, Jun. 2001.
[73] A. Ihler, J. Hutchins, and P. Smyth, "Adaptive event detection with time - Varying poisson processes," dl.acm.org, vol. 2006, pp. 207–216, 2006.
[74] P. Sun, S. Chawla, and B. Arunasalam, "Mining for Outliers in Sequential Databases," Sort, pp. 94–105, 2006.
[75] H. Qiu, Y. Liu, N. A. Subrahmanya, and W. Li, "Granger causality for time-series anomaly detection," in Proceedings - IEEE International Conference on Data Mining, ICDM, 2012, pp. 1074–1079.
[76] H. D. K. Moonesinghe and P.-N. Tan, "OutRank: A GRAPH-BASED OUTLIER DETECTION FRAMEWORK USING RANDOM WALK," Int. J. Artif. Intell. Tools, vol. XX, no. X, pp. 1–18, 2007.
[77] J. Sun et al., "Neighorhood Formation and Anomaly Detection in Bipartite Graph," Proc. 2008 SIAM Intl. Conf. Data Min., pp. 0–7, 2008.
[78] J. Silva and R. Willett, "Hypergraph-based anomaly detection in very large networks," pp. 1–21, 2008.
[79] J. Silva and R. Willett, "Detection of anomalous meetings in a social network," in CISS 2008, The 42nd Annual Conference on Information Sciences and Systems, 2008, pp. 636–641.
[80] P. Sun, … S. C.-'04. F. I. I. C., and undefined 2004, "On local spatial outliers," ieeexplore.ieee.org.
[81] B. Pincombe, "Anomaly detection in time series of graphs using ARMA processes," ASOR Bull., pp. 1–10, 2005.
[82] T. IDÉ and H. KASHIMA, "Eigenspace-based anomaly detection in computer systems," in Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04, 2004, p. 440.
[83] J. Sun, C. Faloutsos, S. Papadimitriou, and P. S. Yu, "GraphScope: parameter-free mining of large time-evolving graphs," in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 2007, pp. 687–696.
[84] N. Heard, D. Weston, … K. P.-T. A. of A., and undefined 2010, "Bayesian anomaly detection methods for social networks," projecteuclid.org.
[85] L. Xiong, B. Poczos, J. Schneider, A. Connolly, and J. Vanderplas, "Hierarchical probabilistic models for group anomaly detection," JMLR WCP Proc. Int. Conf. Artif. Intell. Stat. AISTATS, vol. 15, pp. 789–797, 2011.
[86] L. Xiong, B. Poczos, and J. Schneider, "Group Anomaly Detection using Flexible Genre Models," in Advances in Neural Information Processing Systems, 2011, pp. 1071–1079.
[87] R. Yu, X. He, and Y. Liu, "GLAD: Group anomaly detection in social media analysis," Trans. Knowl. Discov. from Data, vol. 10, no. 2, p. 18:1-18:21, 2015.
[88] K. Muandet and B. Schölkopf, "One-Class Support Measure Machines for Group Anomaly Detection," Mar. 2013.
[89] D. Chakrabarti, "AutoPart: Parameter-Free Graph Partitioning and Outlier Detection," 2004, pp. 112–124.
[90] S. Lin and H. Chalupsky, "Unsupervised link discovery in multi-relational data via rarity analysis," in ieeexplore.ieee.org, 2003, pp. 171–178.
[91] K. Maruhashi, F. Guo, and C. Faloutsos, "MultiAspectForensics: Pattern Mining on Large-scale Heterogeneous Networks with Tensor Analysis," ieeexplore.ieee.org.
[92] L. Friedland and D. Jensen, "Finding tribes: identifying close-knit individuals from employment patterns," in KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 2007, pp. 290–299.
[93] Z. Liu, J. X. Yu, Y. Ke, X. Lin, and L. Chen, "Spotting significant changing subgraphs in evolving graphs," in Proceedings - IEEE International Conference on Data Mining, ICDM, 2008, pp. 917–922.
[94] X. Xu, N. Yuruk, Z. Feng, and T. A. J. Schweiger, "SCAN," in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '07, 2007, p. 824.
[95] M. Kim, J. H.-D. Science, and undefined 2009, "CHRONICLE: A Two-Stage Density-Based Clustering Algorithm for Dynamic Networks.," Springer.
[96] N. a. Heard, D. J. Weston, K. Platanioti, and D. J. Hand, "Bayesian anomaly detection methods for social networks," Ann. Appl. Stat., vol. 4, no. 2, pp. 645–662, 2010.
[97] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network Analysis in the Social Sciences," vol. 892, 2009.
[98] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection," ACM Comput. Surv., vol. 41, no. 3, pp. 1–58, 2009.
[99] G. Kossinets and D. J. Watts, "Empirical Analysis of an Evolving," Science (80-.)., vol. 311, no. January, pp. 88–91, 2006.