A Comprehensive Survey of Time Series Anomaly Detection in Online Social Network Data

Md Rafiqul Islam
Lecturer, Department of
Computer Science & Engineering,
City University, Bangladesh

Naznin Sultana
Assistant prof. Department of
Computer Science & Engineering
Daffodil Int. University,
Bangladesh

Mohammad Ali Moni
School of Biomedical Science
University of Sydney, Australia

Prohollad Chandra Sarkar
Department of Computer Science & Engineering
City University, Bangladesh

Bushra Rahman
Assistant Prof. Department of Computer Science &
Engineering,
Bangladesh Naval Academy

ABSTRACT

In the field of data mining, the social network is one of the complex systems that poses significant challenges in this area. Time series anomaly detection is one of the critical applications. Recent developments in the quantitative analysis of social networks, based largely on graph theory, have been successfully used in various types of time series data. In this paper, we review the studies on graph theory to investigate and analyze time series social networks data including different efficient and scalable experimental modalities. We provide some applications, challenging issues and existing methods for time series anomaly detection.

Keywords

Social networks, Time Series Analysis, Anomaly Detection

1. INTRODUCTION

Social networks have been an upcoming research field over the last few years, and they have gained an established position globally. Social networks provide online hang out space for everybody and using this technology anybody can communicate with their interested friend and share their information, photos, and videos. But this prominent technology also opens the door for unlawful activities. These illicit activities are also alluded to as anomalies. Anomalies emerge in online social networks as a result of specific people, or gatherings of people, rolling out sudden improvements in their examples of connection or communicating in a way that extraordinarily contrasts from their companions. The effects of this strange conduct can be seen in the subsequent network structure. In the field of mathematics, graph theory is a major area to model relations between objects and to represent a connected network structure. From the past decade, researchers are using graph theory to quantify aspects such as similarity, hierarchy and network efficiency of complex network structure in many other fields.

Anomalies are typically defined in terms of deviation from some expected behavior. In general, the definitions of anomalies as “patterns in data that do not conform to a well-defined notion of normal behavior”. Anomalies can be categorized into three classes: point anomalies, contextual anomalies, and collective anomalies [2]. This review focuses on time series anomalies because time series anomalies are also referred to as point anomalies.

A record of phenomenon irregularly varying with time is referred to as time series. The analysis of time series is critical for many different scientific areas as often it contains some anomalies. However, it is necessary to carefully examine graphs of the data as a first step of the time series analysis. For example, to experiments the time series analysis of social networks, S. Asur and B. A. Huberman [1] collected the number of tweets about a specific 'point from Twitter. Also in their trials, each question is characterized by the Twitter handle of a prevalent artist recorded on the site reverbnation.com, and they recover all tweets that straightforwardly address this craftsman. The Twitter time series at that point comprises of the number of such tweets. A case is appeared in Figure 1.

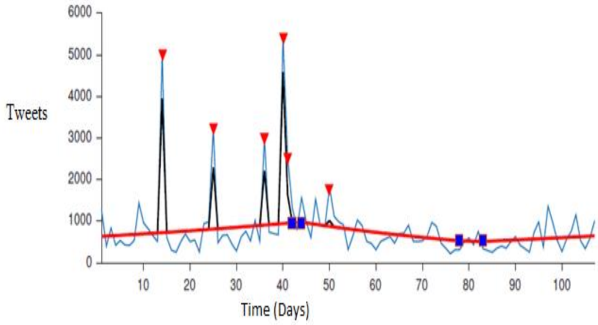

Fig 1: Time series analysis of twitter data. Image courtesy of [1]

Over the past decade, many of the researchers are working to improve their understanding of how the social network work and how they can detect the anomalies. A large number of prominent techniques have been applied to detect time series anomalies in the social network. Time series anomaly detection contains many challenging issues. A predetermined number of tasks have been done on applying time arrangement oddity recognition systems to online social networks because of the issues, for example, precision, computational multifaceted nature, and protection, absence of marked datasets and absence of adequate data. Different methods (distance-based, distribution-based, and clustering based.) have been applied to find anomalies in time series data. But these methods do not work well.

The major contributions of this survey paper are as follows:

(1) We aim to familiarize time series anomaly detection in the social network. We describe how the anomalies can be identified and how time series anomalies can be detected from the social network.

(2) We discuss many of the relevant works that have shown how various anomaly detection techniques of the social network have already been used to detect anomalies.

(3) Finally, with these ideas established we then explain the contributions of our proposed method which has been created in understanding social network. Besides, we discuss how this technology can help ameliorate the future directions of the social network research.

The remainder of the paper is organized as follows: Section II presents the background study of time series anomaly detection in the social network whereas deeper insight into current state of the art is in Section III. The proposed method is presented in Section IV, and its discussion in Section V. Finally, the conclusion is provided in Section VI.

2. BACKGROUND STUDY

2.1 Social Network

A Social network is an online platform which provides an accessible space for everyone, especially young adults. They use this technology to socialize with absorbed accompany and acquaintances, and to allotment photos, information, and videos. This controlling phenomenon, which captures the structure and dynamics of person-to-technology and person-to-person interaction, is being used for numerous purposes such as education, business, medical, telemarketing, ball and adulterous activities. Social networks have been upcoming research areas over the last few years and they have increased a recognized position globally. In general, social networks are self-organizing, emergent, and complex, such that a worldwide articular pattern appears from the local alteration of the elements that make up the system. These patterns become added credible as network size increases. Some of the most popular social network websites are Facebook, Google+, LinkedIn, Twitter, Messenger, Viber, WhatsApp, YouTube, WeChat etc. Sometimes these popular social networks also open the door for anomalies. Anomalies in online social networks can announce irregular, and generally illegal behavior. Detection of such anomalies has been acclimated to analyze awful individuals, including spammers, animal predators, and online fraudsters [3, 4]. It is truly difficult to take care of an anomaly detection problem in a general frame. In this manner, an anomaly detection method should be developed and modified for a particular application by embracing thoughts from various teaches, for example, insights, machine learning, and data mining.

Enos, James R., and Roshanak Nilchiani et al. [1] has proposed social network analysis to comprehend the interoperability related to the DoD arrangement of frameworks. It applied a few centrality measurements to a system of Major Defence Acquisition Programs (MDAPs) to evaluate the interoperability of individual frameworks inside the arrangement of frameworks. In particular, it inspects the contrasts between the degree, closeness, and eigenvector centrality measurements to distinguish which metric best speaks to the interoperability of individual frameworks. Gardounis, Fotios, Heap-Yih Chong, and Xiangyu Wang et al. [2] has purposed an SNA applied a system to dissect venture forms and their changeability as an outcome of BIM implementation. The proposed point of view features applied endeavors for upgrading and connecting holes apparent in the current methodologies. Such holes prompt a failure to oblige the mind-boggling frameworks surfacing among partaking groups and their refined needs, likewise influencing the incorporation of various mechanical interfaces and the coordination of the viable trade of huge scale data.

Paul, Padma Polash, Marina L. Gavrilova, and Reda Alhajj et al. [3] has presented the idea of the social network classifier that can autonomously order an on-screen character from the relationship among performers. They have tried the likelihood of utilizing the social network classifier as a supporting classifier. It can be utilized to enhance the certainty level of different classifiers paying little heed to their temperament. Himaja, N., and G. Murali et al. [4] has suggested a general straggler-aware execution technique, SAE, to help the assessment transporter inside the cloud. It introduces a novel computational disintegration strategy that causes straggling capacity extraction strategies into more magnificent grained sub systems that are then distributed over groups of PCs for parallel execution. The test comes about demonstrate that SAE can pace up the assessment by method for as much as 1.77 occasions in correlation with best in class arrangements.

2.1.1 Graph theory and Social Network Analysis

Graphs are simple models of complex structures, authentic as a set of nodes or vertices V affiliated by a subset of edges E. Mathematically, this can be represented as a graph, G = (V, E), area E is a subset of non-zero elements in the V × V adjacency matrix. When a complex arrangement declared mathematically by graph approach can be called a circuitous or complex network.

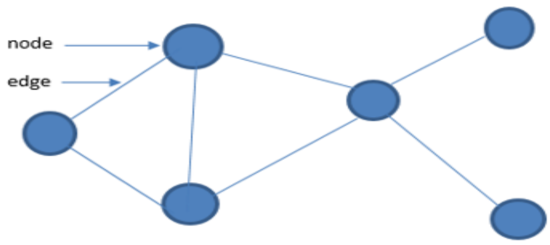

Fig 2: Graphical representation of graph

At present, complex networks have become of above absorption in the area of biological, technological, and amusing sciences such as the science of ecological networks, amusing networks, accord networks, the World Wide Web, and neuroscience. And graph approach examines of complex networks has also accustomed cogent advances in the compassionate of this field.

Awasthi, Abhishek [1] has designed a Genetic Clustering Algorithm which can be utilized for a) recognizing every one of the brunches in a chart and b) the group of only one given hub. The Genetic algorithm utilized for the grouping issue additionally consolidates the Island demonstrates with movement making the calculation much proficient than a solitary popular Genetic algorithm. The execution of the island show with movement can be upgraded substantially more by parallel programming. Doostmohammadian, Mohammadreza, and Usman A. Khan [2] has built up the vital conditions for circulated discernibleness of social networks modeled as LSI frameworks. They have a portray essential perceptions, special order, and system network that empower every operator to deduce any social marvels advancing over a given social digraph. Specifically, they demonstrated that the circulated recognisability requires no a bigger number of perceptions than the brought together case; nonetheless, it requires certain characterization and availability necessities on the specialists watching those states. Cutillo, Leucio Antonio, Refik Molva, and Melek Onen [3] has brook down the connection between the social network graph topology and the achievable security. They have watched three measurements, in particular degree dispersion, bunching coefficient and blending time, and demonstrate that they give key bits of knowledge on the protection level of the OSN.

They proposed how to misuse these experiences for the outline of future protection agreeable OSN. Patil, Neha An., and Amitkumar S. Manekar [4] has actualized and enhanced runtime post or message channel for a social network application. Because of enhanced manage set, it is demonstrated that different sorts of assaults as for private information of the client and malignant social exercises are avoided. They have made a constant application to break down framework execution and it is discovered its execution is expanded essentially when contrasted with other online networking like Facebook, Twitter, LinkedIn and so forth.; Mathematicians Andrew Beveridge and Jie Shan published Network of Thrones in Math Horizon Magazine where they analyzed graph theory and then applied social network analysis algorithms to the network to find the most important characters in the network and a community detection algorithm to find clusters of characters [5].

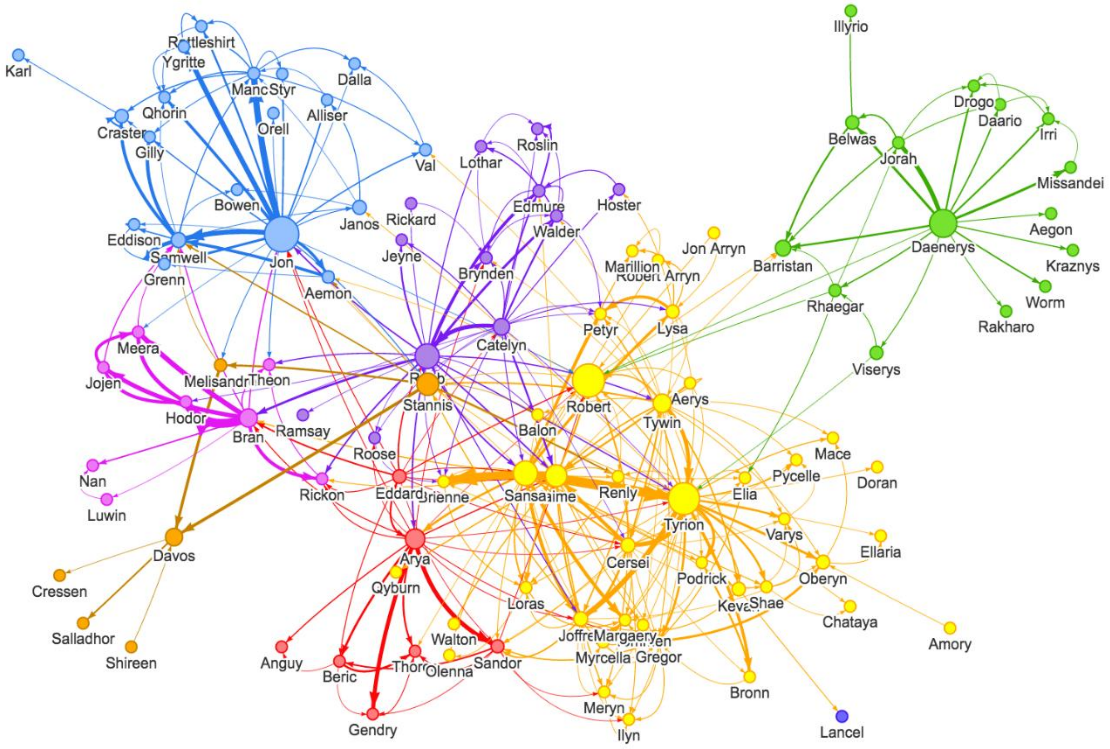

Fig 3: Graph theory analysis in social network. Image courtesy of [5]

2.1.2 Social Network Research with Facebook and Twitter

X. Zhang, W. Li, H. Huang, C.-T. Nguyen, X. Chen, X. Wang, et al., [6] illustrated that online social networks (OSNs) have become a major platform for people to obtain information and to interact with their friends. They have proposed a framework to study the influence of happiness in OSNs. Moreover, they developed a regression model and a greedy algorithm to detect the high influence users known as emotion representatives. By using a small number of detected emotion representatives as features to train prediction models, they showed that their proposed framework achieves good performance in predicting the happiness states of the whole online social network users. T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk [7] has planned to think about the attributes of online depression communities (CLINICAL) in correlation with those joining other online groups (CONTROL). They utilize machine learning and factual techniques to separate online messages between depression and control communities utilizing inclination, psycholinguistic procedures and substance themes removed from the posts produced by individuals from these groups. All viewpoints including the state of mood, the composed substance and composing style are observed to be altogether extraordinary between two sorts of communities. Clear segregation between composing styles and substance, with great prescient power is an imperative stride in understanding social media and its use in mental health.

M. De Choudhury, S. Checks, and E. Horvitz [8] has presented a work on utilizing a crowdsourcing strategy to fabricate an expansive corpus of postings on Twitter that have been shared by people determined to have clinical depression. Next, they build up a probabilistic model prepared on this corpus to decide whether posts could show despondency. The model use flag of the social movement, feeling, and dialect showed on Twitter. Utilizing the model, they present an online networking depression file that may serve to portray levels of sadness in populaces. Geographical, statistic and occasional examples of sorrow given by the measure affirm mental discoveries and connect very with misery insights announced by the Centers for Disease Control and Prevention (CDC). O. L. Haimson, N. Andalibi, M. De Choudhury, and G. R. Hayes [9] has shown how media philosophies around Facebook separation exposures change; yet individuals expect others hold comparable convictions about what is fitting. They add to self-divulgence and online personality writing by distinguishing two new ways individuals participate in revelation and self-introduction via web-based networking media: declarations, which feature how web-based social networking can fill in as productive one-tom any exposure sources, and private status change practices, a reflexive method for self-introduction. Understanding separation divulgences gives knowledge into planning web-based social networking to better empower clients to discover bolster amid troublesome life changes.

K. Saha, I. Weber, M. L. Birnbaum, and M. De Choudhury Saha et al.,[10] has pointed to this investigation were to (1) build a record that deliberate the consciousness of various statistic bunches around schizophrenia-related data on Facebook; (2) consider how this list contrasted crosswise over statistic gatherings and how it corresponded with correlative Web-based (Google Trends) and non– Web-based factors about populace prosperity (emotional well-being markers and framework), and (3) look at the relationship of Facebook determined schizophrenia list with different sorts of online movement and in addition disconnected health and mental health and indicators. K. Garimella, I. Weber, and M. De Choudhury [11] has exactly inspected the part of a recently presented Twitter highlight, 'cite retweets' (or 'quote RTs') in political talk, particularly whether it has prompted enhanced, common, and adjusted trade. Quote RTs enable clients to cite the tweet they retweet, while including a short remark. They investigation utilizing substance, system and group marked information demonstrates that the element has expanded political talk and its dispersion, contrasted with existing highlights. They talk about the ramifications of our discoveries in comprehension and lessening on the web polarization.

2.2 Time Series Analysis

Time-series data is a type of temporal data which is naturally high dimensional and large in data size. Time-series data are of interest due to their ubiquity in various areas ranging from science, engineering, business, finance, economics, healthcare, to government. While each time-series is consisting of a large number of data points it can also be seen as a single object. Genshiro Kitagawa, [12] described the definition of time series data in his book "Introduction to Time Series Modeling" as follows: A record of phenomenon irregularly varying with time is called time series. He also provided the examples of time series data. Typical time series examples are economic data like stock prices; meteorological data like temperature or rainfall and also medical data. Besides he described different classifications of time series. Following categories are Continuous and discrete time series, Univariate and multivariate time series, Stationary and nonstationary time series, Linear and nonlinear time series. Shumway and Stoffer [13] appealed that the examination of investigational information, which depends on perceptions at various time focuses, "prompts new and remarkable issues in measurable displaying and derivation". The investigation of such time series is beside huge for a wide range of scientific areas.

2.2.1 Types of Time Series Data

Periodic and Synchronous: This is the easiest setting where each it has a steady time period (p) and each of the time series is transiently adjusted (begin from a similar time example).

Aperiodic and Synchronous: The time arrangement doesn't have any periodicity, yet they are transiently adjusted.

Occasional and Asynchronous: Each time arrangement has a particular time period, yet they are not transiently adjusted.

Aperiodic and Asynchronous: The time arrangement neither has periodicity, nor are they transiently adjusted.

2.3 Anomaly Detection

The Anomalies are about real as far as a deviation from some acknowledged conduct and the discovery of anomalies tries to discover anomalous subsequences in an arrangement [7]. In our regular day to day existence, abnormality identification systems are utilized expressly or verifiably to identify divergences from what is ordinary or anticipated. Chandola, V., A. Banerjee, and V. Kumar [2] described the definition of non-network based anomaly detection defined anomalies as "patterns in data that do not conform to a well-defined notion of normal behavior". Another recent analysis, Hodge, V. and J. Austin, [14] defines anomalies as "an observation which appears to be inconsistent with the remainder of that set of data". Keogh, E., et al [15] illustrated another definition of anomalies, the time-series discords, are authentic as subsequences that are maximally different from all the actual subsequences. This definition is able to abduction the concept of a lot of unusual subsequences within a time series and its different constant is the appropriate length of the subsequences.

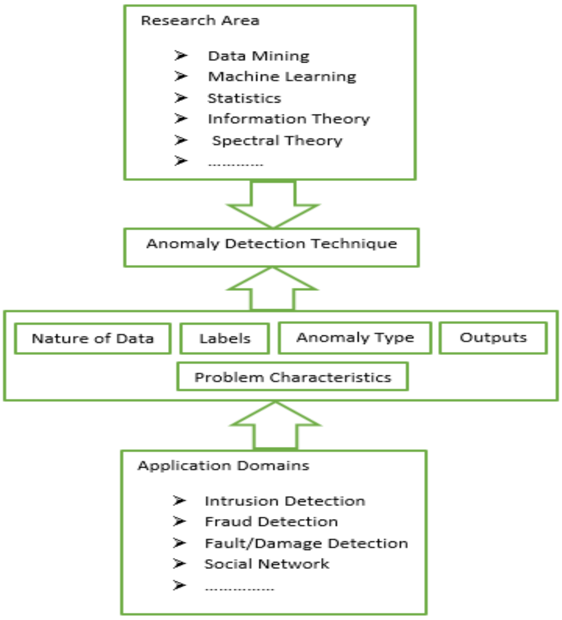

Fig 4: Main Elements Associated with an Anomaly Detection Technique. Image courtesy of [2]

Anomaly detection is an important problem that has been studied within various research areas as shown in Figure 1. Chandola, et al. [2]. They include online social networks, fraud, industrial damage, critical systems, image processing insurance, healthcare, military and bank. Anomaly detection, also named as outlier detection, refers to detecting patterns which do not accede to accepted behaviors. Besides, especially anomaly detection is coming to play an increasingly great role in the social network. Detecting time series anomalies is an important issue in an area as it could be either a sign of a cogent botheration or carrying advantageous advice for the analyzer.

2.4 Anomaly Detection Techniques

2.4.1 The Supervised Anomaly Detection

Supervised methods are also known as classification methods. In the supervised mode, anomaly detection techniques expect that there exists a training data set in which cases have been marked into normal and abnormality classes [16]. Theoretically, supervised methods provide better detection rate than semi-supervised and unsupervised methods, since they have access to more information. However, there exist some technical issues, which make these methods seem not accurate as they are supposed to be. The first issue is the shortage of a training data set that covers all areas. Moreover, obtaining accurate labels is a challenge and the training sets usually contain some noises that result in higher false alarm rates. The most common supervised algorithms are, Supervised Neural Networks, Support Vector Machines (SVM), k-Nearest Neighbors, Bayesian Networks and Decision Tree.

2.4.2 Semi-Supervised Anomaly Detection

The semi-supervised anomaly detection techniques that operate in a semi-supervised mode, expect that there is a training subset in which cases have been marked for only the normal class [17]. Since they do not require labels for the anomaly class, they are more widely applicable than supervised techniques. For example, in spacecraft fault detection, an anomaly scenario would signify an accident, which is not easy to model. The typical approach used in such techniques is to build a model for the class corresponding to normal behavior, and use the model to identify anomalies in the test data. A limited set of anomaly detection techniques exist that assumes the availability of only the anomaly instances for training. Such techniques are not commonly used, primarily because it is difficult to obtain a training data set which covers every possible anomalous behavior that can occur in the data.

2.4.3 Un-Supervised Anomaly Detection

In the unsupervised mode, anomaly detection techniques needn't bother with training data. This sort of techniques, which is broadly utilized, certainly expects that typical cases are more common than anomalies in the data [18]. If this assumption is not true, then such techniques suffer from high false alarm rate. Many semi-supervised techniques can be adapted to operate in an unsupervised mode by using a sample of the unlabeled data set as training data. Such adaptation assumes that the test data contains very few anomalies and the model learnt during training is robust to these few anomalies.

2.5 Anomaly Detection and Online Social Network

2.5.1 Online Behavior's

Online behaviors refer to constant activities shaped in the cyberspace which can be an impression of offline exercises [19]. Computer interceded communication by means of online social networks assumes a key part of this sort of conduct. Therefore, it is critical first to comprehend the idea of the medium and its suggestions. A social network can be viewed as a Computer intervened communications benefit that enables individuals to make a profile (public or semi-public), share and build up association among each other, and explore through their rundown of associations [20]. This facility in addition to absence of physical association with obscure people requesting friendship enables online predators and criminals to ply their commerce. These sorts of associations can be fabricated naturally by computer software or different means, for example, enlisting in a foundation email framework or tolerating fellowship from somebody just met. These associations can't be denied or created by clients regardless of the possibility that they need to do as such. In this way, they are the most dependable information which can be utilized for examining client behaviors in online social networks.

Sparks users are the most dynamic clients of social media. Inviters are somewhat dynamic users who are occupied with articulating their disconnected groups into on the web and efficiently associated with welcoming and urging their companions to join the online services. They are most persuasive users in the way of life of the system [21]. These users with abnormal state of trust to OSNs effortlessly share their own data, cross through the system, and openly disseminate their own points of interest. An inviter sends requests to non-members to join online social networks. Linkers are users who interface themselves to other users [22] play a great part in the development of the online social networks by effectively adding to make associations with different users. They are the most dynamic and powerful users of their gatherings and groups.

Social networks as one of the effective and immediate self-expression method. Newcomer' users are kinds of passive users of a social network who joined recently to an online social network. A linker sends requests to existing users including inviters. Stars are users who discriminately associate with other users with no associations between them. As indicated by a current report, the average users, for example, in Facebook can be classified as respectably dynamic user regarding sending requests, posting substance, and enjoying the substance of their companions [23]. It implies the normal user gets a larger number of requests and substance than they send. Female, old and unmarried users are more vulnerable to be affected contrasted with male, youthful, and married users individually [24].

Creating trust is a standout amongst the essential issues in online activities [25]. The obscurity of recognizable proof is utilized by predators to create trust and closeness considerably speedier online than in a face-to-face relationship. Online predators are expert in finding helpless focuses by gathering individual data, seeking inside profiles, playing as a confided in gathering, and utilizing counterfeit characters. The online predators can be arranged into three gatherings. To start with, security predators [25, 26], these gatherings utilize OSNs to break the protection of alternate users. For example, they may utilize open Twitter channel to find where somebody eats with another person. Second, sexual stalkers [27, 28], they utilize OSNs to acquire sexual contact with someone else in a savage way. Third, financial predators [29], they utilize OSNs to get the financial information, for example, offer fake cash making plans.

2.5.2 Types of Anomalies

Akoglu, et al.[30] and Faloutsos [31] displayed online social networks with graph theory. Faloutsos [31] recommends to utilizing Eigen Spokes, the greatest magnitude projection along the singular vector, to spot anomalies. These are the individuals who utilize similitude between prompted sub-graphs and close inner circles topology as a technique for distinguishing abnormality. Faloutsos demonstrates that factions and star topology can potentially identify with suspicious exercises in financial predators, Facebook, and Twitter-like systems. Akoglu, et al. [30] likewise portrays outliers as star or close star, faction or close inner circle, overwhelming region, and predominant edge. The star or close star topology is a sort of graph network whose nodes are associated with a central node like a hub. The hub gives a mutual point to alternate nodes, which have no or least associations with each other. The star idea can be identified with online social networks: a client interfaces with alternate users unpredictably as there are no associations between the associated users and the central one. The faction or close inner circle topology (or a total subgraph) is a subset of a graph network in which each at least two nodes are associated with an edge. This idea in online social networks refers to groups of individuals, every one of whom knows each other.

2.5.3 OSN Anomaly Detection Challenges

In spite of the fact that outlier detection idea appears to be extremely straightforward, it is a critical issue to comprehend particularly in online social networks when we are managing human behaviors. The difficulties include low accuracy, time and computational challenges, unlabeled datasets, protection, temporal speed, lack of accessibility of adequate data, and no steady definition for an anomaly across over various online behaviors [22, 32, 33]. Finding a labelled dataset in any social network is difficult because of the security of users. None of the online social network suppliers is quick to make their users' data accessible to the general population because of the legitimate issues. Lack of adequate data, for example, exchanged messages between users in existing datasets makes it considerably harder for anomaly detection techniques. In addition, OSNs providers, after some time, continue adding new features and abilities to their system to react to new requests. This brings new difficulties for outlier detection techniques keeping in mind the end goal to accomplish their objective.

Moreover, various structures and motivations behind existing on the social networks bring irregularity between meanings of abnormalities that is difficult to overcome by a generalized method. Discussion of an online social network's graph can be constituted in two classifications: local and global [34]. This basic division is vital as it can be utilized as a systematic use to get inside user behaviors. The nearby perspective of user network focuses on removing rules about users conduct; the worldwide view concentrates on generalizing the extracted manages as examples for separating user's online behavior. For example, the local and global perspectives of users' friendship systems can be used to produce utilization designs. As indicated by the anomaly definition [2], if the use example of a user takes after the regular use design characterized by the global view, the user behavior can be classified "normal", else it can be called "anomalous".

3. ANALYSIS OF THE TIME SERIES ANOMALY DETECTION IN SOCIAL NETWORKS

To detect anomalies, recently researchers have proposed the subject of using graphs to identify properties of the social network. Wasserman, S. and K. Faust, 1994. [35] Illustrated in their famous book that the graph could be characterized a social network basic of Twitter, where each node denotes to a user, and the edges denote the status if a user is a friend with another. Diestel, R., Graph theory,2005. [36] discussed graph that can be represented as a pair G = (V, E). The elements of V are the so-called nodes or points, denoted as v of the graph G and the elements of E are the edges between the nodes. The edges between two nodes can be directed or undirected. Directed express that the edges are only pointing in one direction, whereas in an undirected graph, the edges are pointing in both directions (Newman, M.E., 2003.) [37]. Kwak, H., et al. 2010. [38] used twitter as a social networks and provided a deeper insight into the dynamics of social networks. They told that presently Twitter is one of the biggest and most influenced social networks. It is a microblogging service, which allows users to post status messages with a maximum of 140 characters. Kitagawa, G., 2010. On research group [12] defined time series as a record of phenomenon irregularly varying with time is called time series and provided examples of time series data like temperature or rainfall; economic data like stock prices and also medical data. To detect time series anomalies from the social network, (Rawlings, J.O., S.G. Pantula, and D.A. Dickey, 2001) [39] used regression model which describe the behavior of a random variable of interest. This variable can be the number of stock price, an average number of the temperature or rainfall etc. (Rawlings, J.O., S.G. Pantula, and D.A. Dickey, 2001) [39] extended their regression model equation to identify each independent variable and its regression coefficient. They concluded that it is also possible to state the multiple linear models. Gottman, J.M.J.M., 2009[4] stated an Autoregressive Regression model where the model goes back p time units in the regression to have the ability to predict. To detect time series anomalies, they added a new value in their model as dependent variables. Hayes, A.F., 2006. [40]; Pan, Z. and J.M. McLeod, 1991. [41]; Ritchie, L.D. and V. Price, 1991[42]; Wang, S. and P. Groth, 2010. [43] used Multilevel Regression model for time series anomaly detection in their own works which examine independent and interactive effects of variables. According to authors the crossing of levels of analysis was possible by using this model.

3.1 Applications of time series anomaly detections

The analysis of time series anomaly detection on scientific applications can be examined of the diverse fields. Kitagawa [12] points out that it is also essential to carefully examine graphs of the data as a first step of the time series analysis. So it is easier to identify the next step of the analysis and find appropriate strategies for statistical modeling. Some of the important applications of time series anomaly detections are as follows.

Table 1: Application of time series anomaly detections

	Applications	References
Time series anomaly detection	Heart beat pulses	[44, 45]
	Social network	[46, 47]
	Flight sequences data from aircrafts	[48, 49]
	Shape of medical data	[50-52]
	Complex brain network	[53, 54]
	Outlier light curves of periodic stars	[55, 56]
	Eco-system	[57]
	Attack detection in recommender systems	[58]

Table 1 summarizes the different time series anomaly detection applications obtained from various references. Each of the references has described briefly how to detect anomalies from these applications.

3.2 Challenges of time series anomaly detection

There are some major challenges related with anomaly detection for time series are:

a. There are many different means in which an anomaly occurring in a time series may be identified. An event within a time series may be anomalous; a subsequence within a time series may be anomalous; or an entire time series may be anomalous with respect to a set of normal time series.

b. The training and test time series can be of different lengths.

c. For detecting anomalous subsequence, the real length of the subsequence is generally unknown.

d. Best similarity/distance measures which can be used for different types of time series is not easy to determine. Simple measures like Euclidean distance do not consistently perform well as they are awful acute to outliers and they also cannot be used if the time alternation are of different lengths.

e. Analyzing performances of many anomaly detection algorithms are awful affected to noise in the time series data, as detect appropriate anomalies from noise is an arduous task.

f. Time series in absolute applications are usually continued and as the length increases the computational complication also increases.

3.3 Methods for time series anomaly detection

A large number of approaches have been developed to study time series anomaly detection of social networks and each of these technologies aims to aid in examine and assessing the extent of the social network. Although these techniques are able to detect anomalies from the social network, sometimes they are not able to provide the clear results. They have not the capacity to cover the entire social network rather they provide only information. Moreover, many of the anomaly detection algorithms apprehend multiple time series to be at a comparable scale in consequence while for a lot of the data it is not true.

Table 2: List of time series anomaly detection methods

Method	Description	Refs
Principal Component Analysis (PCA)	Used for identifing unusual time series in a large collections of time series.	[59]
RLPSVDD (Relaxed of linear programming Support Vector Data Description)	RLPSVDD solves a linear programming problem to provide a flexible data description for time series anomaly detection.	[60]
Decision tree classifier	Automatically extracting and summarizing reports	[61]
Latent Dirichlet Allocation (LDA)	Uses an abnormality estimation scheme based on probabilistic topic modeling and seasonal-trend decomposition to find and examine relevant message subsets.	[62]
Two-stage approach	Used for anomaly detection in large dynamic networks, in a context where in principle any type of anomaly should be detected.	[46]
EFS (Evolving fuzzy ststem)	Automatically analyze the user profiles in real times of a specific community of users. This analysis includes the detection of outliers, the clustering of profiles and their classification.	[63]
Statistical modeling	Used for the early and accurate anomaly detection from the time series of the negative tweets.	[64]
Regression Model	Describe the behavior	[6]
Multiple Linear Regression model	Describe the behavior of random variables	[6]
Autoregressive model	Ability to predict	[7]
Multilevel Regression Model	Examine independent and interactive effects of variables	[8-11]
Gaussian Mixed Model	Support categorical and continues values	[65]
ARIMA	Able to cluster seasonality patterns	[66]
Markov Chain	Multiple variable support	[67]
Hidden Markov Chain	Able to capture the dependencies between variables	[68]

Table 2 summarizes the different time series anomaly detection approaches references. Each of the references has applied their approaches and discussed briefly how to detect time series anomalies. Moreover, they have compared their method with other approaches.

4. DISCUSSION AND LIMITAIONS

Nowadays, Time series anomaly detection in the social network is a novel research field globally. Although many researchers already engaged with this research field still there are some challenging issues need to be identified. In this paper, we have surveyed a number of methods for detecting time series anomalies in social networks. We found that these different methods can be usefully categorized based on characterization of anomalies as being static or dynamic and labeled or unlabeled. Depending on this characterization, different features of the network may be examined, and for this we have suggested Model based clustering algorithm. A future aspect could be to apply this unique algorithm or framework to further and may be larger datasets to get better insights. Based on these new datasets and algorithm it would also be a good idea to extend the social network properties. As a result of this it would also be possible to compare the determined results better and have more possibilities to find anomalies.

5. CONCLUSION AND FUTURE WORK

Time series anomaly detection in social networks has been played an important role globally. In this article we described some review of studies and identify some challenging issues of time series analysis. Besides, detecting these challenging problems we also show some novel algorithm. We hope these techniques will be played an increasingly important rule in the evolvement of the social network in near future.

6. REFERENCES

[1] S. Asur and B. A. Huberman, "Predicting the future with social media," in Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on, 2010, pp. 492-499.

[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM computing surveys (CSUR), vol. 41, p. 15, 2009.

[3] B. A. Huberman, D. M. Romero, and F. Wu, "Social networks that matter: Twitter under the microscope," 2008.

[4] J. M. J. M. Gottman, Time-series analysisa comprehensive introduction for social scientists, 1981.

[5] A. Beveridge and J. Shan, "Network of thrones," Math Horizons, vol. 23, pp. 18-22, 2016.

[6] X. Zhang, W. Li, H. Huang, C.-T. Nguyen, X. Chen, X. Wang, et al., "Predicting Happiness State Based on Emotion Representative Mining in Online Social Networks," in Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2017, pp. 381-394.

[7] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, "Affective and content analysis of online depression communities," IEEE Transactions on Affective Computing, vol. 5, pp. 217-226, 2014.

[8] M. De Choudhury, S. Counts, and E. Horvitz, "Social media as a measurement tool of depression in populations," in Proceedings of the 5th Annual ACM Web Science Conference, 2013, pp. 47-56.

[9] O. L. Haimson, N. Andalibi, M. De Choudhury, and G. R. Hayes, "Relationship breakup disclosures and media ideologies on Facebook," New Media & Society, p. 1461444817711402, 2017.

[10] K. Saha, I. Weber, M. L. Birnbaum, and M. De Choudhury, "Characterizing Awareness of Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates," Journal of medical Internet research, vol. 19, 2017.

[11] K. Garimella, I. Weber, and M. De Choudhury, "Quote RTs on Twitter: usage of the new feature for political discourse," in Proceedings of the 8th ACM Conference on Web Science, 2016, pp. 200-204.

[12] G. Kitagawa, "Introducing to Time Series Modeling, Chapman & Hall," ed: USA, CRC Press, 2010.

[13] R. H. Shumway and D. S. Stoffer, Time series analysis and its applications: with R examples: Springer Science & Business Media, 2010.

[14] V. Hodge and J. Austin, "A survey of outlier detection methodologies," Artificial intelligence review, vol. 22, pp. 85-126, 2004.

[15] E. Keogh, J. Lin, S.-H. Lee, and H. Van Herle, "Finding the most unusual time series subsequence: algorithms and applications," Knowledge and Information Systems, vol. 11, pp. 1-27, 2007.

[16] R. Hassanzadeh, "Anomaly detection in online social networks: using data-mining techniques and fuzzy logic," Queensland University of Technology, 2014.

[17] G. Blanchard, G. Lee, and C. Scott, "Semi-supervised novelty detection," Journal of Machine Learning Research, vol. 11, pp. 2973-3009, 2010.

[18] W. Chimphlee, A. H. Abdullah, M. N. M. Sap, S. Chimphlee, and S. Srinoy, "Unsupervised clustering methods for identifying rare events in anomaly detection," a a, vol. 2, p. 1, 2005.

[19] L. Tang, "Online Friendship," Encyclopedia of Cyber Behavior, pp. 412-421, 2012.

[20] D. Centola, "The spread of behavior in an online social network experiment," science, vol. 329, pp. 1194-1197, 2010.

[21] A. Gupta, K. P. Sycara, G. J. Gordon, and A. Hefny, "Exploring friend's influence in cultures in Twitter," in Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013, pp. 584-591.

[22] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in Link mining: models, algorithms, and applications, ed: Springer, 2010, pp. 337-357.

[23] K. N. Hampton, L. S. Goulet, C. Marlow, and L. Rainie, "Why most Facebook users get more than they give," Pew Internet & American Life Project, vol. 3, pp. 1-40, 2012.

[24] S. Aral and D. Walker, "Identifying influential and susceptible members of social networks," Science, vol. 337, pp. 337-341, 2012.

[25] K.-K. R. Choo and A. I. o. Criminology, Online child grooming: a literature review on the misuse of social networking sites for grooming children for sexual offences vol. 103: Australian Institute of Criminology Canberra, 2009.

[26] J. Hitchcock, "Cyberbullies, online predators, and what to do about them," Multimedia and Internet@ Schools, vol. 14, p. 13, 2007.

[27] J. Wolak, D. Finkelhor, K. J. Mitchell, and M. L. Ybarra, "Online “predators” and their victims: Myths, realities, and implications for prevention and treatment," 2010.

[28] I. R. Berson, "Grooming cybervictims: The psychosocial effects of online exploitation for youth," Journal of School Violence, vol. 2, pp. 5-18, 2003.

[29] R. Bapna, "When snipers become predators: can mechanism design save online auctions?," Communications of the ACM, vol. 46, pp. 152-158, 2003.

[30] L. Akoglu, M. McGlohon, and C. Faloutsos, "Oddball: Spotting anomalies in weighted graphs," Advances in Knowledge Discovery and Data Mining, pp. 410-421, 2010.

[31] C. Faloutsos, "Large graph mining: patterns, cascades, fraud detection, and algorithms," in Proceedings of the 23rd international conference on World wide web, 2014, pp. 1-2.

[32] M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou, "Walking in facebook: A case study of unbiased sampling of osns," in Infocom, 2010 Proceedings IEEE, 2010, pp. 1-9.

[33] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in Proceedings of the 2005 ACM workshop on Privacy in the electronic society, 2005, pp. 71-80.

[34] M. E. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks," Proceedings of the National Academy of Sciences, vol. 99, pp. 2566-2572, 2002.

[35] S. Wasserman and K. Faust, Social network analysis: Methods and applications vol. 8: Cambridge university press, 1994.

[36] R. Diestel, "Graph theory, ser," Graduate Texts in Mathematics. Springer-Verlag, Heidelberg, vol. 173, 2005.

[37] M. E. Newman, "The structure and function of complex networks," SIAM review, vol. 45, pp. 167-256, 2003.

[38] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," in Proceedings of the 19th international conference on World wide web, 2010, pp. 591-600.

[39] J. O. Rawlings, S. G. Pantula, and D. A. Dickey, Applied regression analysis: a research tool: Springer Science & Business Media, 2001.

[40] A. F. Hayes, "A primer on multilevel modeling," Human communication research, vol. 32, pp. 385-410, 2006.

[41] Z. Pan and J. M. McLeod, "Multilevel analysis in mass communication research," Communication research, vol. 18, pp. 140-173, 1991.

[42] L. D. Ritchie and V. Price, "Of matters micro and macro: Special issues for communication research," Communication Research, vol. 18, pp. 133-139, 1991.

[43] S. Wang and P. Groth, "Measuring the dynamic bi-directional influence between content and social networks," The Semantic Web–ISWC 2010, pp. 814-829, 2010.

[44] M. C. Chuah and F. Fu, "ECG anomaly detection via time series analysis," in International Symposium on Parallel and Distributed Processing and Applications, 2007, pp. 123-135.

[45] S. Akhavan and G. Calva, "‘Automatic Anomaly Detection in ECG Signal by Fuzzy Decision Making," in Proceedings of 6th International Conference on Fuzzy Theory and Technology, 1998, pp. 96-98.

[46] N. A. Heard, D. J. Weston, K. Platanioti, and D. J. Hand, "Bayesian anomaly detection methods for social networks," The Annals of Applied Statistics, vol. 4, pp. 645-662, 2010.

[47] D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Anomaly detection in online social networks," Social Networks, vol. 39, pp. 62-70, 2014.

[48] V. Rajagopalan and A. Ray, "Symbolic time series analysis via wavelet-based partitioning," Signal Processing, vol. 86, pp. 3309-3320, 2006.

[49] D. K. Tolani, M. Yasar, A. Ray, and V. Yang, "Anomaly Detection in Aircraft Gas Turbine Engines," JACIC, vol. 3, pp. 44-51, 2006.

[50] M. Bicego and V. Murino, "Investigating hidden Markov models' capabilities in 2D shape classification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, pp. 281-286, 2004.

[51] Z. Liu, J. X. Yu, L. Chen, and D. Wu, "Detection of shape anomalies: A probabilistic approach using hidden markov models," in Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on, 2008, pp. 1325-1327.

[52] L. Wei, E. Keogh, and X. Xi, "Saxually explicit images: Finding unusual shapes," in Data Mining, 2006. ICDM'06. Sixth International Conference on, 2006, pp. 711-720.

[53] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in Proceedings, 2015, p. 89.

[54] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," Nature reviews. Neuroscience, vol. 10, p. 186, 2009.

[55] P. Protopapas, J. Giammarco, L. Faccioli, M. Struble, R. Dave, and C. Alcock, "Finding outlier light curves in catalogues of periodic variable stars," Monthly Notices of the Royal Astronomical Society, vol. 369, pp. 677-696, 2006.

[56] U. Rebbapragada, P. Protopapas, C. E. Brodley, and C. Alcock, "Finding anomalous periodic time series," Machine learning, vol. 74, pp. 281-313, 2009.

[57] H. Cheng, P.-N. Tan, C. Potter, and S. Klooster, "Detection and characterization of anomalies in multivariate time series," in Proceedings of the 2009 SIAM International Conference on Data Mining, 2009, pp. 413-424.

[58] S. Zhang, A. Chakrabarti, J. Ford, and F. Makedon, "Attack detection in time series for recommender systems," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 809-814.

[59] R. J. Hyndman, E. Wang, and N. Laptev, "Large-scale unusual time series detection," in Data Mining Workshop (ICDMW), 2015 IEEE International Conference on, 2015, pp. 1616-1619.

[60] C. Huang, G. Min, Y. Wu, Y. Ying, K. Pei, and Z. Xiang, "Time Series Anomaly Detection for Trustworthy Services in Cloud Computing Systems," IEEE Transactions on Big Data, 2017.

[61] J. Krumm and E. Horvitz, "Eyewitness: Identifying local events via space-time signals in twitter feeds," in Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2015, p. 20.

[62] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, et al., "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," in Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on, 2012, pp. 143-152.

[63] J. A. Iglesias, A. García-Cuerva, A. Ledezma, and A. Sanchis, "Social network analysis: Evolving Twitter mining," in Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on, 2016, pp. 001809-001814.

[64] K. Ikeda, G. Hattori, C. Ono, H. Asoh, and T. Higashino, "Early detection method of service quality reduction based on linguistic and time series analysis of twitter," in Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on, 2013, pp. 825-830.

[65] C. Biernacki, G. Celeux, and G. Govaert, "Assessing a mixture model for clustering with the integrated completed likelihood," IEEE transactions on pattern analysis and machine intelligence, vol. 22, pp. 719-725, 2000.

[66] M. Corduas and D. Piccolo, "Time series clustering and classification by the autoregressive metric," Computational Statistics & Data Analysis, vol. 52, pp. 1860-1872, 2008.

[67] M. Ramoni, P. Sebastiani, and P. Cohen, "Multivariate clustering by dynamics," in AAAI/IAAI, 2000, pp. 633-638.

[68] M. Bicego, V. Murino, and M. A. Figueiredo, "Similarity-based clustering of sequences using hidden Markov models," in International Workshop on Machine Learning and Data Mining in Pattern Recognition, 2003, pp. 86-95.