Computers and Electrical Engineering 57 (2017) 294–310

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

A comparative analysis of structural graph metrics to identify anomalies in online social networks☆

Ravneet Kaur*, Sarbjeet Singh

Department of Computer Science and Engineering, UIET, Panjab University, Chandigarh, India

ARTICLE INFO

Article history:
Received 21 January 2016
Revised 12 November 2016
Accepted 14 November 2016
Available online 23 November 2016

Keywords:
Anomaly
Betweenness centrality
Brokerage
Clique
Online social networks
Star networks

ABSTRACT

Social networks are becoming vulnerable to a number of fraudulent attacks and mischievous activities due to their widespread use and increasing popularity. So, detection of anomalous activities, especially in social networks, is essentially required as it helps to identify important and significant information regarding the behavior of anomalous users. In order to detect anomalies in social networks, researchers have mainly relied on the use of behavior and structure based approaches. Working in the similar direction, we extend the graph structure based approach by introducing and analyzing important graph metrics to detect anomalous activities. The comparison and effectiveness of measures have been presented on the basis of statistical measures like precision, recall and F-score, as well as on the basis of calculated anomalous scores. Theoretical and empirical evaluation reveals that the relationship between brokerage and number of edges helps to detect and correctly rank maximum number of anomalies.

1. Introduction

Online social networks (OSN) have emerged as an interesting area of research seeking immense attention from researchers. At present, almost every domain is linked in one or the other way with social networks, be it entertainment, education, trading, business, communication and many others. OSNs have made an influence on each of them. But undoubtedly, this rising popularity has also made these networks prone to a number of social crimes. Many terrorist organizations, unethical corporations and fraudulent users are shifting onto social networks to commit unlawful activities. Therefore, analysis of social networks for the detection of some unusual and mischievous activities, often referred to as anomalies, is the need of the hour. Also, simply learning about the structure of data without spotting what stands out in the data, does not add much benefit to the analysis. This “stand out” behavior is what is referred to as an anomaly, and is defined as an unusual activity of a user with respect to others or one’s own past behavior. In the context of social networks, Savage et al. [1] formally defined the term anomaly as “regions of the network whose structure differs from that expected under the normal model”. Therefore, in simple terms, some kind of quantitative or qualitative features of a user’s behavior that are inconsistent with the rest of the users/itself are considered anomalous in online social networks.

The “unusual and inconsistent” behavior that this paper addresses is the presence of near star and clique pattern. In general, a star topology represents the network in which any two nodes in the network are connected with each other through one common node only. In star networks, nodes in the network do not have any connected edges between them,

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. M. S. Kumar.

* Corresponding author.
E-mail addresses: ravneets48@gmail.com (R. Kaur), sarbjeet@pu.ac.in (S. Singh).

http://dx.doi.org/10.1016/j.compeleceng.2016.11.018
0045-7906/

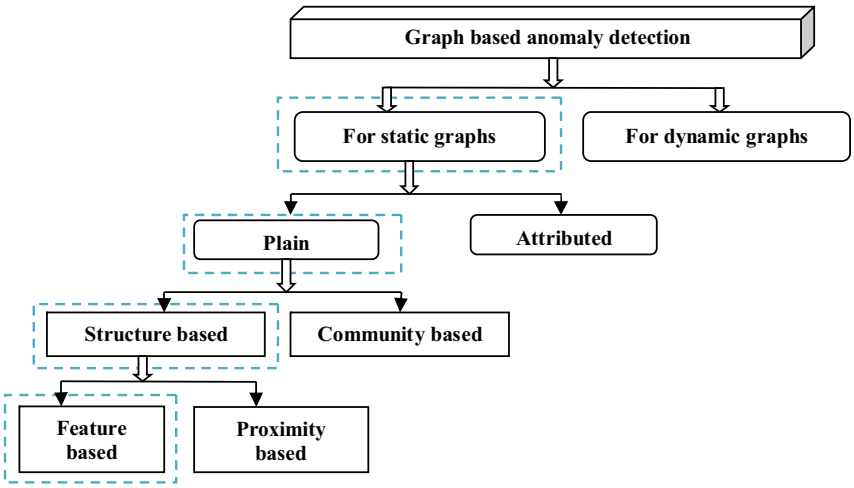

Fig. 1. Graph based anomaly detection domains [2].

Table 1
Terminologies used in graph based anomaly detection systems.

Factors	Type	Description
Type of graph	Static	Static graphs represent the data at a particular snapshot of time.
	Dynamic	Dynamic graphs deal with the time varying data.
Data in graph (both static and dynamic)	Plain	Plain graphs handle the unlabeled/unweighted graphs without considering any weights/features on the nodes or edges.
	Attributed	Attributed graphs represent the labeled and weighted graphs in which nodes/edges have weights assigned to them.
Patterns observed (in plain graphs)	Structure based	Presence of structure based patterns in plain graph involves the analysis of the patterns pertaining to the overall structure of a graph (e.g. density of the complete graph).
	Community based	Detection and analysis of community based patterns involve the patterns followed by a closed community or cluster of nodes and edges with similar behavior (e.g. density of a particular community or cluster of users).
Attributes used (for structure based patterns)	Feature based	Feature based attributes include the study of graph-centric features (e.g. node degree, centrality)
	Proximity based	Proximity based features include the detection of closeness of nodes to detect associations in the graph

except the common node. On the other hand, clique represents the presence of a dense structure in which most of the nodes in the network have connections among each other. A common trend usually followed in all the social networks is that “friends of friends are often friends”, but there are some nodes which follow the near star or clique pattern, depicting complete disconnections or extreme connections in their friend circle, which are considered as anomalous. For example, a star topology in social networks states an uncommon friendship pattern that could be related to a celebrity or an influential person. A financial company can easily use such type of information for advertising their products in the influential persons' network. Nowadays, review websites like Epinions, Amazon, Qype are grabbing huge attention from fraudulent users to commit opinion spams. Opinion spam involves writing up of fake reviews for a product so as to enhance or mangle the reputation of a vendor. In order to present themselves as legitimate, such users usually befool the public by making a lot of connections among themselves, thereby forming a clique structure.

In order to detect anomalous activities in social networks, it is a common practice to observe the interaction patterns among different users. As social networks are well represented as graphs, such interactive patterns can be effectively studied by analyzing the structural properties of the graphs. Also, graphical analysis portrays a number of advantages that make them a well sought out option to be considered. Easy and friendly representation of graph data makes the implementation of various techniques quite easy [2]. Moreover, the presence of autocorrelation and interdependence among the nodes along with the robustness and ubiquitous trend of graph based methods makes them more effective and difficult for an adversary to fake or alter them.

However, graph based anomaly detection in itself is a very vast domain. As stated by Akoglu et al. [2], depending upon the type and structure of a graph, graph based anomaly detection techniques can be further studied and applied at different granularity levels. According to the problem being addressed in this paper, only those sub modules, that are applicable to our work, have been drilled down in Fig. 1 and a brief description of these modules have been highlighted in Table 1. For a detailed study of various graph structure methods, it is suggested to refer to the survey paper by Akoglu et al. [2].