Journal of Network and Computer Applications 68 (2016) 213–229

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Review

Mining social networks for anomalies: Methods and challenges

P.V. Bindu *, P. Santhi Thilagam

Department of Computer Science and Engineering, National Institute of Technology Karnataka, Surathkal, India

ARTICLE INFO

Article history:
Received 20 October 2015
Received in revised form
5 February 2016
Accepted 25 February 2016
Available online 16 April 2016

Keywords:
Anomaly detection
Online social networks
Graph mining
Graph anomaly detection
Outlier detection
Social Network Analysis

ABSTRACT

Online social networks have received a dramatic increase of interest in the last decade due to the growth of Internet and Web 2.0. They are among the most popular sites on the Internet that are being used in almost all areas of life including education, medical, entertainment, business, and telemarketing. Unfortunately, they have become primary targets for malicious users who attempt to perform illegal activities and cause harm to other users. The unusual behavior of such users can be identified by using anomaly detection techniques. Anomaly detection in social networks refers to the problem of identifying the strange and unexpected behavior of users by exploring the patterns hidden in the networks, as the patterns of interaction of such users deviate significantly from the normal users of the networks. Even though a multitude of anomaly detection methods have been developed for different problem settings, this field is still relatively young and rapidly growing. Hence, there is a growing need for an organized study of the work done in the area of anomaly detection in social networks. In this paper, we provide a comprehensive review of a large set of methods for mining social networks for anomalies by providing a multi-level taxonomy to categorize the existing techniques based on the nature of input network, the type of anomalies they detect, and the underlying anomaly detection approach. In addition, this paper highlights the various application scenarios where these methods have been used, and explores the research challenges and open issues in this field.

Contents

* Corresponding author.
E-mail addresses: bindupv007@gmail.com (P.V. Bindu), santhi@nitk.ac.in (P.S. Thilagam).

http://dx.doi.org/10.1016/j.jnca.2016.02.021
1084-8045/

1. Introduction

The past decade has witnessed a dramatic increase of interest in online social networks due to the growth of Internet and Web 2.0. The social networking websites enable users to interact with each other easily regardless of the geographical locations and hence, they have received tremendous attention by both academic and industries. The openness of online social networks allows the users to gather a large amount of information about other users. Regrettably, this huge amount of information as well as the ease with which one can navigate through the network, attract the interest of malicious users. As a result, the misuse of the social networks and services has also increased and has opened up the door for many illegal activities and security issues such as identity theft, cyber attacks, organized crimes, bullying, spamming, fraudulent information dissemination, and even terrorist attack planning (Keyvanpour et al., 2014; Liu and Chawla, 2015; Yu et al., 2015).

Anomalies are patterns whose behavior varies significantly from majority of the data (Chandola et al., 2009). In social networks arena, an anomaly is an unexpected behavior of a user or group of users whose behavior is unusual compared to the normal behavior of the users in the network. More specifically, anomalies in social networks occur when the behavior of users forms unusual network patterns (Savage et al., 2014). For example, in an online social network, majority of the users follow the pattern of "friends of friends are often friends", and the minority follow either the "cliques or near-cliques" or the "stars or near-stars" pattern (Akoglu et al., 2010, Hassanzadeh and Nayak, 2013a,b; Hassanzadeh et al., 2012). These two minority patterns in which all neighbors are mostly interconnected or mostly disconnected can be possible candidates for anomalies, as only a modest percentage of users follow this behavior. Even though not all anomalies are malicious, they can signify strange behavior such as credit card fraud, campaign donation irregularities, electronic auction fraud, e-mail spam and phishing (Akoglu et al., 2010), and many others. Therefore, it is extremely important to detect these irregular behaviors.

Anomaly detection in social networks refers to the problem of spotting the strange and unexpected behavior of users by exploring the patterns hidden in the networks. It has become a significant task in Social Network Analysis (SNA). In order to detect anomalies in a social network, we need to analyze the interactions among different users in the network. This makes the anomaly detection problem on social networks different from other forms of traditional anomaly detection.

A social network is normally modeled as a graph, with nodes denoting the users and the edges denoting the relationships. As a result of the robustness of the graph representation, it is hard for an adversary to fake or alter it. Therefore, the anomalous parts of the social network can be identified by using graph mining techniques. Graph mining-based anomaly detection in social networks is an emerging area of research that has got a strong foundation in classical graph theory as well as sociological aspects such as how users interact with each other and group together. Even though a plethora of work has been done on anomaly detection on datasets containing multi-dimensional data instances, anomaly detection using graph mining techniques has received attention only in the recent years.

In this paper, we provide a comprehensive and systematic review of the research works done in the area of mining social networks for anomalies. Even though the emphasis of this paper is to review the anomaly detection techniques proposed in the last ten years, we have also described few of the earlier works that are formative to this area. The main contributions of this paper can be summarized as follows:

- Identifying the key aspects associated with the problem of anomaly detection in social networks.
- Providing a multi-level taxonomy to categorize the existing anomaly detection techniques based on (i) nature of input network, (ii) types of anomalies, and (iii) anomaly detection approach.
- Providing a comprehensive review of the state of the art in anomaly detection using the above taxonomy.
- Highlighting various real-world application scenarios where the social network anomaly detection methods have been used.
- Exploring the research challenges and open problems in the area of anomaly detection in social networks.

The rest of the paper is organized as follows. Section 2 presents the background topics related to mining social networks for anomalies. Section 3 describes some of the specific application domains of anomaly detection in social networks and Section 4 discusses the existing related survey articles. Section 5 describes the different aspects of anomaly detection in social networks. Sections 6 and 7 review the various anomaly detection methods on static and dynamic social networks respectively. After presenting a