COMPUTERS & SECURITY 72 (2018) 60–73

Available online at www.sciencedirect.com

ScienceDirect

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

Computers & Security

Combating the evolving spammers in online social networks

Qiang Fu [a], Bo Feng [a], Dong Guo [a,b], Qiang Li [a,b,*]

[a] College of Computer Science and Technology, Jilin University, Changchun, China

[b] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, China

ARTICLE INFO

Article history:
Received 17 February 2017
Received in revised form 22 August 2017
Accepted 22 August 2017
Available online 7 September 2017

Keywords:
Online social networks
Spammer detection
Temporal evolution
Machine learning
Classification

ABSTRACT

Online social networks, such as Facebook and Sina Weibo, have become the most popular platforms for information sharing and social activities. Spammers have utilized social networks as a new way to spread spam information using fake accounts. Many detection methods have been proposed to solve this problem, and have been proved to be successful to some extent. However, as the spammers' strategies for evading detection evolve, many existing methods lose their efficacy. A major limitation of previous approaches is that they are using the features from a static time point to detect spammers, without considering temporal factors. In this study, we approach the challenge of spammer detection by leveraging the temporal evolution patterns of users. We propose a dynamic metric to measure the change in users' activities and design new features to quantify users' evolution patterns. Then we develop a framework by combining unsupervised and supervised learning to distinguish between spammers and legitimate users. We test our method on a real world dataset with a large number of users. The evaluation results show that our approach can efficiently distinguish the difference between spammers and legitimate users regarding temporal evolution patterns. It also demonstrates the high level of similarity in the spammers' temporal evolution patterns. Compared with other detection methods, our method can achieve better performance. To the best of our knowledge, our study is the first to provide a generic and efficient framework to depict the evolutional pattern of users. It can handle the problem of spammers updating their strategies to evade detection and is a valuable reference for this research field.

1. Introduction

Online Social Networks (OSNs), such as Twitter, Facebook, and Sina Weibo, have become an essential part of people's daily life. Facebook, the most popular social network worldwide, has a monthly average of 1.51 billion active users (Statista, 2016). With their increasing influence among users, OSNs have become an ideal platform for spammers to spread spams. The term "spammers" mainly refers to users that initialize unsolicited social relationships or send unsolicited messages through fake accounts, social bots or spam applications (Yang et al., 2014). The types of attacks that are launched by spam include, but are not limited to, product advertisements, phishing attacks, and drive-by-download attacks. Spam can reduce users' social media experiences by annoying them with content that they

* *Corresponding author.*
E-mail addresses: fuqiang15@mails.jlu.edu.cn (Q. Fu), fengbo16@mails.jlu.edu.cn (B. Feng), guodong@jlu.edu.cn (D. Guo), li_qiang@jlu.edu.cn (Q. Li).

http://dx.doi.org/10.1016/j.cose.2017.08.014
0167-4048/

are not interested in. Furthermore, it can lead to privacy leaks or economic losses if users are tricked to the phishing websites. Hence, accurately detecting spammers to make online social networks more user-friendly and secure is one of the most serious issues in existing OSNs.

The primary challenge of detecting spammers is that they are upgrading their spam strategies rapidly to race with the development of detection systems (Yang et al., 2013). For methods that use common features based on user profiles and message content, such as Chu et al. (2012); Egele et al. (2013); Gao et al. (2012), spammers can evade being detected by purchasing followers or using tools to post messages with the same meaning but different words automatically (Yang et al., 2013). Yang et al. (2012) found that spammers tend to be interconnected, forming account communities, thus rendering certain advanced features for detecting spammers such as Clustering Coefficient, ineffective. The primary assumption of the PageRank-based method is that there are a limited number of the edges maintaining reciprocal social relationships between spammers and legitimate users, yet the evidence that legitimate users follow spammers more than expected has been found. Ghosh et al. (2012) found that a small fraction of users, known as social capitalists, follow back anyone who follows them to increase their reputation. Yang et al. (2012) also discovered supporter accounts that help spammers avoid detection by increasing their followers, allowing them to prey on more victims.

As conventional detecting methods cannot cope with the new strategies adopted by spammers, researchers have proposed some new approaches to meet these challenges. For example, Yang et al. (2013) used some features that are more sophisticated than the previous ones to improve the efficiency of machine learning classifiers. Boshmaf et al. (2016), based on the information of victims who are benign social network users and have mutual connections with spammers, made the PageRank-based method more robust. These studies provide a deeper insight into the difference between spammers and legitimate users and improve detection accuracy. However, whether an individual user is a spammer is inferred by these methods based on user characteristics at a single instant of time. Their real-world data about users are collected at a single point of time, and the experiments are conducted and evaluated from the perspective of a static social network. In fact, social networks are constantly changing, and spammers may be able to improve the effectiveness of an attack through persistent efforts (Liu et al., 2015). Therefore, as spammers evolve their strategies to evade detection, the capacity of these approaches to efficiently detect them becomes dubious.

In this study, we introduce temporal factors into the detection of spammers by inspecting the activities of users over an extended period of time and offer a detecting framework to identify the spammers that evade detection by changing their strategies. Intuitionally, even if many spammers can make their accounts appear like legitimate user accounts at some static time points to avoid being detected, it is impossible for them to manipulate the dynamic changing process of features over an extended period of time due to the high cost (Yang et al., 2013).

To achieve our research goals, we collect the profiles of a vast number of social network users and track their activities over a series of points of time. A window-based dynamic metric is used to assess the temporal evolution patterns of users and uncover a clear distinction between legitimate users and spammers concerning different aspects of the temporal patterns. Based on the dynamic metric, new temporal user features are designed to detect spammers. Instead of using these features to identify spammers directly, we investigate the similarity in the temporal patterns of different spammers, and conduct a clustering algorithm (Maulik and Bandyopadhyay, 2000) on users by abstracting their dynamic metrics into feature vectors. The results indicate that it is relatively easier to group spammers into the same cluster. We combine the new features with the clustering results to build a machine learning classifier for accurate detection of spammers. Finally, we evaluate our method using the real-world dataset and demonstrate the effectiveness of our approach by comparing it with two conventional spammer detection methods.

The contributions of this paper are three-folds:

- We propose a dynamic metric model based on sliding window to measure the dynamic changes of social networks users' activities.
- On the basis of the dynamic metric, we design new features to describe the temporal evolution patterns and come up with a novel framework combining unsupervised clustering and supervised classification to detect spammers in OSNs.
- We implement our approach and evaluate it using a real-world dataset. Compared to conventional methods, we are able to achieve better performance.

The remainder of this paper is organized as follows: Section 2 covers related works on spammer detection in online social networks. Section 3 provides the motivation and some assumptions used in this study. Section 4 provides further details about the proposed dynamic metric. Section 5 describes our approach for detecting spammers in online social networks. Section 6 explains the experiments conducted and results of the evaluation of our approach. Section 7 concludes the paper.

2. Related work

Due to the inundation of spams in online social networks, many studies have been conducted aiming to investigate spammers in different kinds of social networks. Some previous works focused on characterizing spammers (Almaatouq et al., 2014; Gao et al., 2010; Grier et al., 2010; Stringhini et al., 2010; Thomas et al., 2011). After collecting a certain amount of data of spammer accounts, the characteristics of spammers were analyzed in these studies from different aspects, such as individual profiles, propagative behavior, message contents and social relationships. Stringhini et al. (2010) collected the data related to spammers by using Honeypot and then divided them into different categories according to their behavior patterns. Yang et al. (2014) deploy social honeypots on Twitter with diverse strategies to trap spammers, consequently revealing the preferences of spammers when they are finding their targets. These works showed the fundamental characteristics of spammers from a variety of aspects, laid the foundation for the follow-up