A Statistical Infinite Feature Cascade-Based Approach to Anomaly Detection for Dynamic Social Networks

Yasser Yasami , Farshad Safaei

PII: S0140-3664(16)30621-1
DOI: 10.1016/j.comcom.2016.11.010
Reference: COMCOM 5413

To appear in: *Computer Communications*

Received date: 21 September 2015
Revised date: 7 November 2016
Accepted date: 24 November 2016

Please cite this article as: Yasser Yasami , Farshad Safaei , A Statistical Infinite Feature Cascade-Based Approach to Anomaly Detection for Dynamic Social Networks, *Computer Communications* (2016), doi: 10.1016/j.comcom.2016.11.010

A Statistical Infinite Feature Cascade-Based Approach to Anomaly Detection for Dynamic Social Networks

Yasser Yasami and Farshad Safaei

Faculty of Computer Science and Engineering, Shahid Beheshti University G.C., Evin 1983963113, Tehran, IRAN

{y_yasami, f_safaei}@sbu.ac.ir

Abstract

The development of methods for anomaly detection in dynamic ubiquitous online social networks is critical to coincide with the growth in social network usage. This paper presents a novel statistical approach to anomaly detection in dynamic social networks. The approach relies upon the fact that the network dynamics can be driven by microscopic features of each node that dynamically cascade to neighboring nodes over time. The proposed approach consists of two main components: (1) normal modeling component and (2) anomaly detection component. The former component is involved in three main processes, governing the network dynamics. The first process is the features' birth, death, and lifetime, which is assumed to follow a realistic statistical distribution in this paper for the very first time. The second process is the evolution of nodes' features that is modeled by an Infinite Factorial Hidden Markov Model (IFHMM), considering feature cascade. The feature cascade is a phenomenon that explicitly describes how the past features of each node affect the features of its neighboring nodes in future. The third process modeled in this paper is the relationship between nodes' features and link generation in dynamic social networks. The latter component of the proposed approach provides a new method to quantize deviation of network dynamics from the normal behavior. Some Markov Chain Monte Carlo (MCMC) sampling strategies have been used to simulate parameters of the proposed model, given social network data. The proposed anomaly detection approach is validated by experiments on synthetic and real social network datasets. Experimental results show that this approach outperforms other related approaches in terms of some statistical performance measures, especially applied to binary normal-abnormal classification test.

Keywords: Dynamic Social Networks, Anomaly Detection, Feature Cascade, Statistical Modeling,

1. Introduction

Computer and communication networks are the main noteworthy infrastructures of the information society since they have contributed to the proper functioning of many serious services. Internet as the heart of modern computer communication infrastructures should be secure, resilient, and as close to human social communication paradigms as possible. Accordingly, providing computer networking services through the Internet to facilitate the emerging forms of computer-mediated social communications seems to be inevitable. Online Social Networks as new phenomena, applying the computer and communication networks infrastructures, have affected our lives in various ways. However, the growth of social networks applications comes with an increase in the prevalence of malicious activities using social network services [1]. Consequently, the development of methods for anomaly detection is critical to coincide with the growth in social network usage.

Anomaly detection refers to monitoring the behavior of a system and flagging significant deviations from the normal behavior as anomalies. Recently, anomaly detection has been used for identifying attacks in computer and communications networks [2-4], malicious activities in computer systems [5, 6], fatigue-cracks in electromechanical systems [7], misuses in Web [8], outlier identification in relational data [9, 10] and malicious activities in online social networks [11-13]. Anomalies in social networks can denote to irregular and mostly illegitimate behavior, and so it could be beneficial to be detected.

There are some similarities between anomaly detection and two related well-known problems in social network research: missing link prediction [14-16] and future link recommendation [17, 18]. The former focuses on predicting which links are likely to be present, even if not observed. The latter is focused on predicting new relationships, *e.g.* new collaborations given a history of past projects, while the main purpose of anomaly detection task is to identify anomalous links, given a network with objects and their relationships. Anomalous links denote to missing links with high likelihood as well as the existing links being statistically unlikely. We contend that, often, the most interesting links are those that are