AGENT-BASED SAFETY PROTECTION IN ONLINE SOCIAL NETWORKS

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF COMPUTER AND INFORMATION SCIENCES

Supervisors

Dr. Quan Bai

Dr. Jiamou Liu

July 2017

By

Yingying Tao

School of Engineering, Computer and Mathematical Sciences

Copyright

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

Signature of candidate

Acknowledgements

Research is life. Life is research. Masters study is a unique experience in my life which represents transformation from learning courses to doing research.

I would like to express my sincere appreciation to my supervisors Dr. Quan Bai and Dr. Jiamou Liu. They always give me very important suggestions, not only tell me how to do that, but also guide me to think about why to do that. They push me to become creative during the process of research. Besides, they require me to have a critical thinking method whether in research or during meetings. Creativity and critical thinking are the enormous wealth for me. My supervisors have helped me a lot since the first time they taught me papers in my undergraduate studies. They continually support and encourage me whatever in research or life. Under their supervision, I have tried my best in the Master studies. I am grateful to my supervisors.

I would like to thanks my parents. They always support me and provide the strong backing for me. I can focus on my research. Without their unconditional encouragement and support, I could not have finished my studies.

I would also like to give my great appreciation to all members of Department of Computer and Mathematical Sciences for their help and assistance. Especially lab members of Artificial Intelligence. My thanks also go to Auckland University of Technology and Collaborative Program between Auckland university of Technology and China Jiliang University. You have given me the chance to be a better person.

Finally, I give my gratitude to all the people who have helped me since the first time

I came into the area of Computer and Information Sciences.

Abstract

With the development of the Internet, more and more people actively interact with others via online social networks. Potentially, people can hide themselves in the dark and continually gather information from other users from the Internet. To assist individual users to protect their privacy and security, in this study a computational approach for abnormal attention detection will be presented. The proposed approach can detect abnormal attention from the local view of a user, without invading other people's privacy. We then move on to focus on the online interpersonal surveillance which is an excessive, unreciprocated and persistent attention. We address the issue of interpersonal surveillance by asking the question, "who is surveiling you through social networking?". This is a challenging question, as interpersonal surveillance is a victim-defined behaviour and often occurs without a visible trace. Viewing a network as interconnected agents who interact through posting and reading information, we provide a measure to quantify the level of attention a person pays towards another from a global view. This measure allows us to capture online interpersonal surveillance. Through theoretical and experimental analyses, we show that our method distinguishes and detects online interpersonal surveillance.

Publications

Jiamou Liu, Yingying Tao and Quan Bai. Towards Exposing Cyberstalkers in Online Social Networks. The 14th Pacific Rim International Conference on Artificial Intelligence (PRICAI), 2016

Yingying Tao, Quan Bai, Jiamou Liu and Hangxia Zhou. Detecting Abnormal Attention in Online Social Networks from Local Views. The 2nd IEEE International Conference on Agents (ICA), 2017

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Online social networks (OSNs) have become extremely popular in our daily lives. Through OSNs, users can easily establish new friendships, share their stories and opinions. With the convenience and fast development of online communication, there is also danger lurking in the dark. People post a vast amount of personal information in OSNs, which if abused, could lead to tragic outcomes.

Nowadays, many OSNs have tens of millions of registered users. With the increasing usage of OSNs, many users have unknowingly exposed their privacy to threats. Phishing attacks and spammer threats not only attack themselves but also their friends(Fire, Goldschmidt & Elovici, 2014). Many people use OSNs for uploading photos of themselves and their friends(Acquisti, Gross & Stutzman, 2011). Such information can be used to create a biometric database, which can be used to identify OSNs users without their consent. People stand naked in front of the OSNs, their personal private information exposed to the public. Moreover, people have to face the fake profiles threats, identity clone attacks, inference attacks and information or location leakage attacks. More alarmingly, some users may gain control over their targets exploiting the information gathered from social networks, all without the targets' awareness (McFarlane & Bocij, 2003).

Online safety and privacy protection is an important research field in computer science (Fire et al., 2014). When a person posts a message on an OSN, the readers of the message stay anonymous. Moreover, the person may even not be aware of who has read the message.

In 1999, a report by Attorney General J. Reno asserted that cyberstalking had been an increasingly prevalent practice (Reno, 1999). Since then, cyberstalking has attracted a growing interest (Al−Khateeb et al., 2015; Parsons-Pollard & Moriarty, 2009; Spitzberg & Hoobler, 2002). Existing research mainly focused on understanding the drive behind stalkers and their impact on the victims (Dreßing, Bailer, Anders, Wagner & Gallas, 2014; McFarlane & Bocij, 2003), as well as deriving behavioral taxonomies on how stalkers use online means to gather information and harass victims (Hitchcock, 2003; Goodno, 2007). Interpersonal surveillance also exists in contexts outside of cyberstalking. People commonly monitor activities of others through OSNs (Joinson, 2008). In a 2011 study, around 67% of college students responded using Facebook to monitor their ex-romantic partners (Lyndon, Bonds-Raacket & Cratty, 2011). Numerous studies also discussed the role of online surveillance in romantic relationships (Tokunaga, 2011, 2016; Fox, Warber & Makstaller, 2013; LeFebvre, Blackburn & Brody, 2015).

People may easily hide themselves in OSNs. How can we find the person who pays excessive attention to you in OSNs? This is a challenging question, as attention often occurs without a visible trace. Besides, it is more complicated to detect the online interpersonal surveillance which is an excessive, unreciprocated and persistent attention.

1.1 Research Motivations and Objectives

The issue of cyberspace security becomes more and more important in OSNs. In terms of the special situation in OSNs that the information may relate to users' personal

privacy which increases the difficulties to find the person who pays excessive attention to others. Moreover, it is hard to detect users who surveil others. So in this thesis, we focus on two aspects. One is how to detect abnormal attention which is regarded as excessive attention one user pays to another. Another one is how to detect interpersonal surveillance among users which represents excessive, unreciprocated and persistent attention.

In order to detect abnormal attention, we propose a fuzzy logic-based approach from agents' local views which can protect personal privacy. We aim to detect a user who pays excessive attention to another, which is far beyond the level of their real relationships. Based on the motivations above, we define three factors Individual Physical Distance, Social Interaction Distance and Message Reading ratio to establish a social attention model from local views.

In terms of detecting interpersonal surveillance, we find only rely on agents' local views, it is very hard to detect interpersonal surveillance. Then we propose a Markov chain attention model base on the global view to measure online interpersonal surveillance which is excessive, unreciprocated and persistent attention. We define message-based attention index and network-based attention index as two factors in the Markov chain model.

1.2 Research Methodology

In this section we introduce the research methodology we used in our research. Fig. 1.1 shows the detailed methodology of each step in our research. At first, we review existing researches and methods about problems in the online social networks such as cybercrime, cyberbullying, cyberstalking and so on. Then we define the core concept of our research which is attention. We want to quantify how much attention one user pays to another. We want to know who pays excessive attention to others and who not only

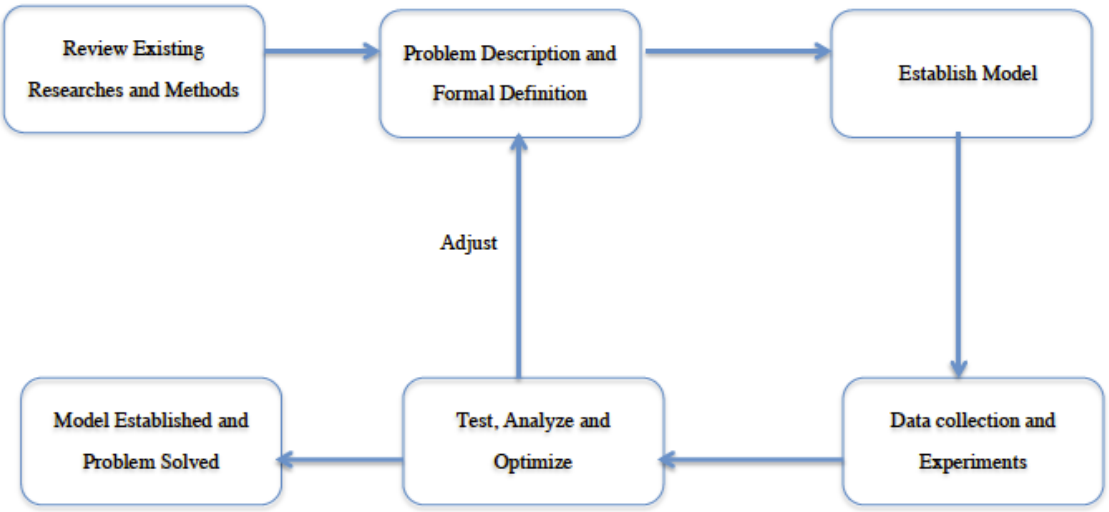

Figure 1.1: Research Methodology

pays excessive attention but also pays unreciprocated and persistent attention to others. So the main research problems are detecting abnormal attention and online interpersonal surveillance among users. After that, we establish a computational model for detecting abnormal attention and online interpersonal surveillance. We collect network data and generated information data, then we test our model in the generated and real-world graph. According to the result of our experiments, we adjust problem descriptions and formal definitions to optimize our model and let them easily and clearly distinguish abnormal attention and interpersonal surveillance. Finally, the model is established and the problem is solved.

1.3 Problem Description and Formal Definitions

1.3.1 Modelling of User Interactions in OSNs

Publishing profiles, blogging, commenting, posting texts, photos, and videos, which are only some of the forms of interactions among users of an OSN. We need an abstraction

that encapsulates these forms of interactions in a general yet conceptually simple framework.

An OSN consists of ties between its members; these ties denote friendship or a "following" relationship from one individual to another. Such ties represent visible connections and form the basis of information dissipation. However, such ties are not the only form of interactions. People communicate and access others' messages on the Web, regardless of whether the two parties have a visible tie.

Hence online interaction fundamentally departs from physical interaction and any model should take into account not only the visible connections between users but also interactions among unconnected users.

With this view in mind, we introduce the following formal model to describe user interactions in an OSN. The user interactions model consists of a directed graph where nodes represent agents (i.e., users). Directed links represent established connections between two users; a mutual relationship can be represented by a pair of directed links. Each agent in this network has the ability to carry out three actions: *posting* messages, *interaction* with other users like clicking *like* buttons or leaving comments and *reading* messages which friends post. Once a message is posted, it can be read or interacted by others in cyberspace. We assume that there is a universal set of messages that can be posted, read and interacted with users over the network.

Definition 1 *A* social network *is defined as a directed graph* $G = (V, E)$*, where* V *is a set of* nodes *(i.e.,* agents*) and* $E \subseteq V^2$ *is a set of* directed edges *denoting relationships among agents.* $E = E^P \cup E^I \cup E^R$*.* E^P *is a set of Physical Links.* E^I *is a set of Interactive Links.* E^R *is a set of Reading Links.*

In Definition 1, a Physical Link e^P_{ij} ($e^P_{ij} \in E^P$) denotes a physical connection (e.g., Twitter "follow") between Agents v_i and v_j in G. An Interactive Link e^I_{ij} ($e^I_{ij} \in E^I$) denotes an interactive behaviour between Agents v_i and v_j. A Reading Link e^R_{ij} ($e^R_{ij} \in$

E^R) denotes that v_i once read an article or post published by v_j.

Agent behaviours: Let M be a set of *messages* propagated in the network G. In the proposed model, an Agent v_i has three types of social behaviours associated with a message m ($m \in M$). $m \in \mathsf{post}(v_i)$ denotes that m is posted by v_i. $m \in \mathsf{read}(v_i)$ denotes that m is read by v_i. $m \in \mathsf{interact}(v_i)$ indicates that v_i has interactions with other agents with m, e.g., commenting or liking m.

1.3.2 Social Attention

Attention is a core concept in this research. It refers to an invested interest from one person to another and measures the intensity of interactions. An acceptable amount of attention has generally a positive effect to the target person; as it could be viewed as the result of increased personal influence which leads to new personal ties or opportunities. However, excessive attention may lead to negative effects and potential risks. An extremely high level of attention often means one user pays an abnormal attention to another. Furthermore, if a user pays excessive attention to another in an unreciprocated level and persistence, it can be defined as interpersonal surveillance.

1.4 Major Contributions of Thesis

We argue that social network knowledge could contribute to the modelling, simulation, and detection of abnormal attention and interpersonal surveillance. In particular, we view an OSN as a crowd of autonomous agents whose behaviors are affected by actions of others through interpersonal ties. In order to develop an effective means to mitigate the risks posed by abnormal attention and interpersonal surveillance, one needs to characterize its behavioral traits. We approach this attention problem from a network-centric perspective which is based on the local view and global view.

The main contribution is five-fold: (1) We present abnormal attention and interpersonal surveillance as a computational problem in the framework of social network analysis from a local view and a global view. This is a novel initiative, and could lead to a wealth of new research problems towards online safety. (2) Towards modelling abnormal attention, we use fuzzy logic to detect potential abnormal attention from individual users' local views. (3) Towards modelling interpersonal surveillance, we give a quantitative definition of the attention a person pays to others. This definition consists of two parts, the first is based on messages retrieved by the user, and the second is a link-based centrality that takes into account interpersonal ties. Our study relies on a novel measure of social attention which takes into account not only individuals' reading of others' messages, but also the network topology. We contrast these two notions using theoretical analysis on several special network structures. (4) We present surveillance index, which measures the excessive, unreciprocated and persistent information gathering behaviours that are typical in interpersonal surveillance. (5) We present a simulation framework capturing a range of agent behaviors. Through experimental analysis, our simulation clearly demonstrates the validity of our surveillance definition and differentiates non-surveillance behaviors from surveillance behaviors.

1.5 Thesis Organisation

The remainder of this thesis is organised as follows.

In Chapter 2, we present an overview of related works in Online Social Networks (OSNs). We focus on the privacy and security problems in the OSNs. Then we review other researches on how to define attention and allocate agents' attention. We face the big and serious challenges such as cyberbullying and cyberstalking with the development of OSNs.

In Chapter 3, we use fuzzy logic to establish a model to detect abnormal attention

from agents' local views. We define Individual Physical Distance, Social Interaction Distance and Message Reading Ratio in this model. Then we collect network topology data from a website and evaluate our model in the experiments. We set two experiments, one proofing our model fits the different social status users while the other detecting abnormal attention. Then we analyze the experiments results and successfully detect abnormal attention.

In Chapter 4, we calculate Messsage-based attention and Network-based attention index. Then we define a Markov-chain model to capture the network topology effect on social attention. We study Markov-chain model in special network typologies. Then we present the measure of interpersonal surveillance from global view and it is reasonable to view as excessive, unreciprocated and persistent attention. After that, we perform simulation and experimental results on both synthesized and real-world networks. We evaluate our model from four experiments: average attention index, in-degree and attention index, network density and surveillance index and number of messages and surveillance index. Moreover, we use precision and recall to measure the performance of our proposed method for surveillance detection.

In Chapter 5, we conclude our current findings, results and limitations. We want to carry on this research in the future.

Chapter 2

Literature Review

2.1 Privacy and Security

A large amount of personal information are disclosed by OSNs users themselves. Users have to face to the threats of leaking personal information. OSNs can be the target of different types of attacks such as identity theft, phishing, spamming and clickjacking, to name a few.

Identity theft is a type of attack on OSNs in which the adversary attempts to collect personal information of OSN users so that he can impersonate the victim of the attack in order to gain some benefits or harm the victim. Different methods can be used to launch identity theft attacks, including accepting friend requests from unknown people, sharing account details with others, clicking links that lead the user to other websites, downloading free applications, low privacy settings and so on. Phishing is the most common method in identity theft attack. The fraudulent user attempts to steal the victim's personal information like his credentials or credit card information.

As for the spamming attacks, spammers send out a huge number of emails to advertise, sell their products and gain sensitive information about the users (e.g., username and password) by pretending as a trusted party (e.g., banks or online payment processors).

With this approach, the attacker sends random friend requests to the members of the OSNs target community waiting for them to accept their requests.

Clickjacking is an example of confused deputy problem. A malicious user deceives OSN users into clicking on a link, which is different from what the users expect it to be. In other word, the webpage that is showed to the user is different from the page where the user's action is taking place.

Online safety and privacy has also been an active field in computer science. There are major efforts on the detection of phishing, spamming, cloning, and bots on OSNs (Fire et al., 2014).

2.2 Attention Analysis

With the emergence of online social networks and increased enthusiastic users, information dissemination has been transformed. Users rely on their relationships for gaining information. Relationships represents how two users are close, which decides the speed of propagation in the OSNs (Jiang, Hegde, Masoulié & Towsley, 2013). Information propagation in the OSNs can be regarded as the attention propagation.

Backstrom et al. propose a measure that an individual divides his or her attention across contacts for analysing personal networks that addresses a dimension distinct from network size and composition (Backstrom, Bakshy, Kleinberg, Lento & Rosenn, 2011). Some people focus most of their attention to a small circle of close friends, while others disperse their attention more broadly over a large set. They capture users' behaviours as the different modalities of attention. They define Messages, Comments, Wall Posts, Profile Views and Photo Views as a way to measure how users allocate attention across friends.

Jiang et al. propose the "plus-one" mechanism to identify optimal allocations of limited frequency among neighbours for each user in the network. This mechanism

is based on using incentives as a form of feedback for reallocating attention. They consider that there are different types of information from different sources. At first, they consider an asynchronous pull model, where each user contacts a neighbor after a random delay and pulls any content available from that neighbor. Further, they assume that each user has a limited budget of attention (Jiang et al., 2013). They take the approach of conceiving a general model for studying the balance of attention and an analysis for several network topologies.

Bernstein et al. quantify the attention from an invisible audience who is listening to the content which users shared in OSNs. They use survey and analyze the large-scale log data to compare users' perceptions and real attention gained from their actual audience on Facebook. Researchers analyze audience logs for 222,000 Facebook users' from friend count, likes and comments (Bernstein, Bakshy, Burke & Karrer, 2013).

2.3 Cyberbullying Detection

Cyberbullying is a type of attack that takes place by sending out harmful or offensive material, including text and images, to targeted users using the internet, cell phones, video game systems, or other technology, according to the National Crime Prevention Council (*http://www.ncpc.org/topics/by-audience/teens/protect-yourself/cyberbullying*, n.d.). Once such material spreads among a large group of OSN users, it is very hard to remove them from the network.

Nahar et al. find predators and victims by determining the most active users in the form of the OSN (Nahar, Li & Pang, 2013). They propose a novel statistical detection approach, which is based on the weighted Term Frequency/Inverse Document Frequency (TFIDF) scheme on bullying-like features. Research on cyberbullying detection associates the theory of communication and text mining methods to differentiate between predator and victim conversations, as applied to one-to-one communications

like in a chat-log dataset (Kontostathis, 2009).

Yin et al. use a supervised learning approach for detecting cyberbullying. They define the detection of cyberbullying as a classification problem with two classes: positive class for documents which contain harassment and negative class for documents which do not contain harassment. The researchers use local features, sentiment features, and context features as attributes to classify each document belongs to one of the two classes. As for the local features, it can be extracted from users' post itself. They use each distinct term as one feature and calculate a TFIDF value for each feature. Then they capture the sentiment features like second person pronouns, all other pronouns, foul language and so on. After that, they identify other features as contextual features which can distinguish harassment-like posts from real harassment posts. They collect experiments datasets from CAW 2.0 workshop and perform them, which shows the significantly improved performance than the basic TFIDF model (Yin et al., 2009).

Cyberbullying can be defined as 'sending or posting harmful or cruel text or images using the Internet or other digital communication devices' (Willard, 2005). It can happen in various formats including flaming, harassment, cyberstalking and so on. Cyberbullying can lead to stalking and death threats. Unlike face to face bullying, in the cyberspace, people always think they can do and say anything they want. In addition, they can easily cover themselves.

Combating cyberbullying is an extremely difficult task. Many bullies are anonymous. Further, they have free-speech rights, so it is very difficult to take down a website (Willard, 2005).

2.4 Cyberstalking Description and Detection

The Internet grants people access to a vast amount of personal information, which if abused for diabolical reasons, could lead to tragic outcomes. On October 2, 2012,

15 year old Canadian girl Amanda Todd killed herself at her home, after a three-year tormenting experience of being stalked, harassed and bullied on OSNs by a stranger (Todd, 2012). Since the age of 12, Amanda had been stalked online by a stranger she met at an Internet chat room, who collected details of Amanda's life and convinced her to bare her breasts on camera. The stranger later harassed Amanda with this photo, following Amanda as she changed her online accounts and switched schools, while befriending Amanda's schoolmates, which caused Amanda's mental breakdown and eventual suicide.

Amanda is only one of countless victims of *cyberstalking* on OSNs in the last 15 years. The term "cyberstalking" refers to persistent monitoring, information gathering, identity theft, threatening, vandalism, which happens through online social medias, which may be used to threaten or harass a network user (Fire et al., 2014). Along with cyberbullying and online predators, cyberstalking amounts to a new form of threats that exploits cyberspace. Similar to physical stalking, cyberstalking targets especially vulnerable groups such as children and women (McQuade, Rogers, Gentry & Fisk, 2012).

The conventional, physical *stalking* refers to the stealthy and persistent pursuit of a specific person with unwanted and obsessive attention which results in harassment. The person who carries out the act of stalking is referred to as a *stalker* or the *perpetrator*, and the person being stalked is the *victim*. *Cyberstalking* is the form of stalking that occurs in cyberspace. It is believed that cyberstalking may result in a similar, if not higher, level of threatening as conventional stalking; indeed, these two types of behaviors share a number of common traits (Pittaro, 2007):

- Both behaviors are characterized by obsessive attention from the perpetrator to the victim, which includes monitoring and information gathering.
- Both behaviors are mainly driven by the perpetrator's desire and need to gain

power, control and influence over the victim.

- Both behaviors are *victim-defined*, that is, the level of seriousness of a particular stalking incident is determined by how much intimidation the victim perceives (Reno, 1999).

However, cyberstalking should not be regarded simply as an extension of conventional stalking, but rather a different deviant behavior in its own form (Bocij & McFarlane, 2002):

- Firstly, conventional stalking includes some clearly defined, detectable actions such as physically following the victim home, vandalizing the victims properties and sending gifts, leaving a physical trail of evidences. Cyberstalking, on the other hand, is much harder to define: Here, a victim may purposely expose private information on OSNs, making monitoring extremely easy; any online user may derive information regarding another's occupation, age, and address with little effort. Furthermore a stalker may also use identity masks, making information gathering unnoticeable.

- Secondly, in most conventional stalking incidents, the stalker and the victim are within each other's social or physical periphery: Either they have a prior relationship (whether real or perceived), or they live or work within relatively close proximity of each other. For cyberstalking, the victim is more often chosen at random and may occur between two people with arbitrary physical distance.

- Thirdly, cyberstalkers often employ tools and techniques that are unique to the use of the Internet. For example, a number of incidents involve stalkers carrying their deeds using Trojan software, email spamming or phishing techniques (Al−Khateeb et al., 2015), all of which are challenging technical online threats themselves.

Due to the reasons above, the tasks of detection, prevention, and forensics of cyberstalking becomes considerably more challenging than for conventional stalking. Despite serious efforts from academics, there have not been major technical advancements that effectively prevent people from cyberstalking. The most widely used methodology is *profiling*: by gathering statistical information of cyberstalkers and victims (such as age, gender, occupation, drug abuse history, etc.), the method aims to capture the likelihood of an individual of being a stalker. While profiling provides certain indication of the general phenomenon of cyberstalking, it is far from an effective method for prediction and detection. Based solely on statistical information, however, this method is far from an effective method for detection.

Towards a Formal Definition. One of the most comprehensive definitions of cyberstalking was offered by Bocij and McFarlane in (Bocij & McFarlane, 2002):

> *"A group of behaviours in which an individual, group of individuals or organisation uses information technology to harass one or more individuals. Such behaviour may include, but are not limited to, the transmission of threats and false accusations, identity theft, data theft, damage to data or equipment, computer monitoring and the solicitation of minors for sexual purposes. "*

The definition entails a large sum of deviant behaviors, which vary by nature and require different countermeasures. Therefore it is difficult to use one single mathematical framework to capture the entire plethora of behaviors. As a pilot study, we can only focus on a particular aspect of cyberstalking. In an earlier study (Meloy, 1999), Maloy provides a more "strip-down" definition, which asserts cyberstalking as consisting of two major functions:

1. The stalker gathers private information on the target to further a pursuit; and

2. The stalker communicate (in real time or not) with the target to implicitly or explicitly threaten or to induce fear.

2.5 Agent-Based Modelling

Agent-based modelling (ABM) is a widely used simulation modelling technique, used to solve real-world problems such as stock markets, consumer markets, the threats of bio-warfare and so on. The agent-based modelling has been defined by Bonabeau (Bonabeau, 2002) as "a system is modeled as a collection of autonomous decision-making entities called agents. Each agent individually assesses its situation and makes a decision on the basis of a set of rules." An agent is identifiable, autonomous, situated and flexible (Macal & North, 2009). In the simplest level, an agent-based model consists of a system of agents and the relationships between them. Though the agent-based model is established in the simply way, it still can indicate the complex behaviours patterns and gain the valuable information about the dynamics of the real-world system (Bonabeau, 2002).

In terms of benefits of ABM, it can capture both individual and organizational behaviours of the system entities. Besides, ABM provides a natural description of a system, which is flexible to be observed along multiple dimensions (Bonabeau, 2002).

The ABM can be used for modelling social processes, where agents represents the people or groups of people, and agent relationships represent the processes of social interactions (Gilbert & Troitzsch, 2005).

Macal et al. consider ABM as a powerful tool for modelling complex systems. Especially as the systems are too complex, only by using ABM, we can get some assumptions and take a more realistic view of systems. In addition, with the development of computational power, we can compute large-scale micro-simulation models (Macal & North, 2009).

It is hard for traditional modelling approaches to capture the complexities of OSNs. Hence, in this thesis, we adopt ABM model users' behaviours.

2.6 Summary of Literature Review

In this chapter, we review the existing problems in OSNs. Every user wants their personal privacy to be protected and at the same time the ability to easily share their feelings and stories in OSNs. However, some specific users hide themselves in the dark while continually gathering information from their interested users which causes the security problems.

We review researches which are related to attention analysis. They focus on how to allocate attention and the budget of attention. Then we review the cyberbullying problems. Cyberbullying consists of text-based harassments, visible traces that are amenable to text mining and machine learning. After that, we review the problem of cyberstalking, which is a form of stalking that occurs in the cyberspace.

All of these researchers falls into the broader area of cyber-safety and privacy in computer science. However we have found no technical breakthrough specifically focusing on abnormal attention detection and the problem of online interpersonal surveillance. Online interpersonal surveillance is different from cyberstalking as the former is much more hidden. This may be due to the complex nature of the type of online user interactions and the actions generally associated with relationships among users. As pointed out in (Pittaro, 2007), despite decades of criminological research, there has not been a generally agreeable definition of abnormal attention and online interpersonal surveillance. The goal of our result is to bridge the gap by investigating abnormal attention and interpersonal surveillance in a formal, computational perspective and hopefully develop useful technologies that help with the prediction and detection of them. Moreover, we aim to use agent-based approach to detect abnormal attention and

online interpersonal surveillance from agent local views and a global view respectively. Due to the complex and heterogeneous of agents' behaviours in OSNs, agent-based approach is preferred to be used in our research.

Chapter 3

Detecting Abnormal Attention in Online Social Networks from Local Views

3.1 Introduction

Abnormal attention is regarded as a user paying excessive attention to another, which is far beyond the level of their real relationship. It is hard for a user to detect potential risks without the knowledge of the entire OSN, e.g., the network topology, other users' activities, etc. In addition, messages posted by users are related to their personal privacy and cyber security. It is almost impossible for a normal user to gain the information about other users' behaviours from a global view without the invasion of other users' privacy. Hence, a major challenging issue is how a user can know who has paid how much attention to him/her from a local view. In this chapter, we propose a computational model to facilitate a user to analyse received attention and detect potential abnormal attention in OSNs from his/her local view only. We only take users' local views into consideration to protect other users' privacy in the OSNs. Meanwhile, in most OSNs

applications, it is very hard to have a global view for an individual user.

3.1.1 Agent-based Modelling for Abnormal Attention Detection

In this thesis, an online social network is considered as a Multi-agent System (MAS), which consists of a number of agents. Human users are represented as agents, which can take different types of actions, including posting articles, leaving comments to other people's posts, or reading posts. These actions result in different types of mutual relationships among users (agents). Some of these relationships are not formed through interactions, or with the awareness of the users. For example, a post on Weibo[1] can be viewed by any registered users regardless of whether the viewer has followed the owner of the post or not. Similarly, cyberspace supports communications between two users regardless whether they know each other or not. With this view in mind, in the proposed approach, various interactions and relationships are considered.

Agent local view: As mentioned earlier, it is hard for an individual user to possess a global view in an OSN. Hence, in the proposed approach, each agent analyses received attention based on its local view. The local view of Agent v_i contains the information about the behaviours taken by other agents, which are associated with itself. Namely, v_i is aware of who once interacted with it, and who once read its posts. However, v_i is not aware of other agents' actions which are not associated with itself or its posts. As shown in Fig. 3.1 (refer to Section 1.3), there are three types of links among agents, i.e., Physical Links, Interactive Links and Reading Links. The three types of links are corresponding to the three actions taken by the users, i.e., post messages, read messages and interact with others. Moreover, each agent can only make decisions based on its local view. For example, in Fig. 3.1, the area of red oval is the local view of Agent V_2. Namely, Agent V_2 can only collect information from its "neighbours", i.e., agents who

[1] `www.weibo.com`

interacted with it and/or read its posts.

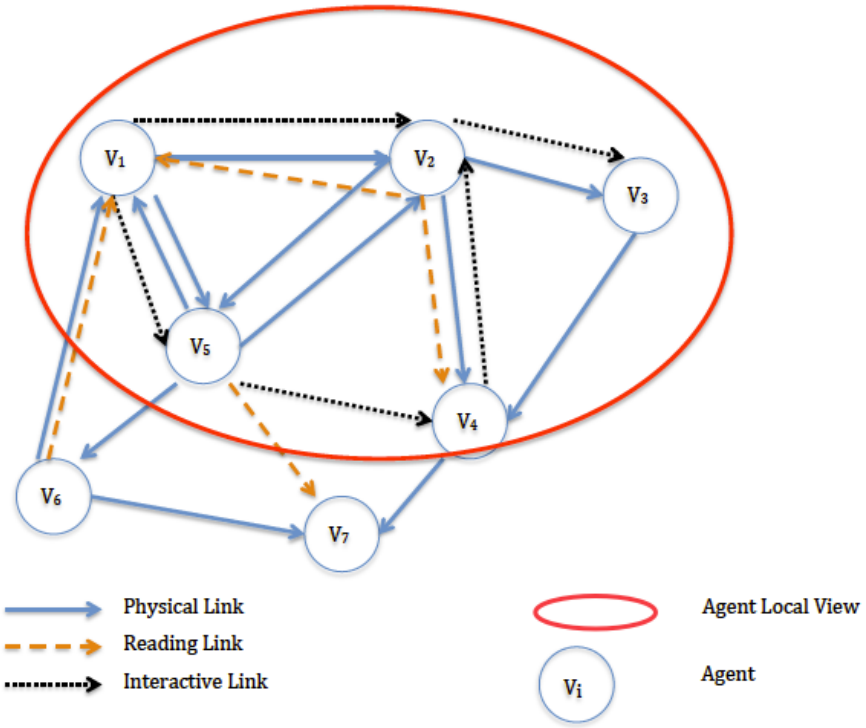

Figure 3.1: Relationships among Agents and Agent Local View

The rest of this chapter is organised as follows. We formally describe the attention model from agent local views in Section 3.2. In Section 3.3, we proposed a fuzzy-logic based approach for detecting abnormal attention. Then, we discuss simulation and experiments conducted in this research in Section 3.4. Finally, we conclude and mention the future directions in Section 3.5.

3.2 Social Attention Model from Local Views

Definition 2 *Individual Physical Distance (PD_{ij}) between Agents v_i and v_j is defined as the number of agents (nodes) that the shortest path between v_i and v_j travels through.*

We suppose that the weight of each Physical Link equals to 1, and use the Floyd-Warshall's shortest path algorithm (Floyd, 1962) to calculate the Physical Distance

(PD_{ij}) between two agents. If PD_{ij} equals to 1, it shows Agents v_i and v_j are the closest friends. When the value of the distance becomes larger, it means that the relationship between v_i and v_j is further. For example, in Fig. 3.2, PD_{12} = 1, PD_{13} = 1, PD_{23} = 1, PD_{34} = 1 and PD_{14} = 2.

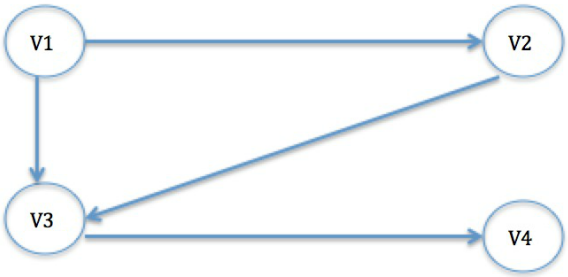

Figure 3.2: The Example of PD_{ij} in 4 nodes Graph

Definition 3 *Social Interaction Distance* (ID_{ij}) *is the measure of the interaction frequency between Agents* v_i *and* v_j. ID_{ij} *can be calculated by using Equation 3.1.*

$$ID_{ij} = \frac{I_{ij}}{I_j} \tag{3.1}$$

In Equation 3.1, I_{ij} is the amount of interactions from v_i to v_j. I_j is the total amount of interactions from all agents to v_j. The Social Interaction Distance between v_i and v_j (ID_{ij}) is the ratio of I_{ij} against I_j.

Definition 4 *The Message Reading Ratio* RM_{ij} *is the ratio of messages that Agent* v_i *read out of the total posted messages of* v_j. RM_{ij} *can be calculated by using Equation 3.2.*

$$RM_{ij} = \frac{|R_{ij}|}{|P_j|} \tag{3.2}$$

In Equation 3.2, P_j represents the set of messages which Agent v_j posted. R_{ij} represents the set of messages posted by Agent v_j which were read by v_i. RM_{ij} is between 0 and 1. When RM_{ij} equals to 0, it means v_i has not read any messages that v_j posted.

When RM_{ij} equals to 1, it means v_i reads all messages that v_j posted. If v_i reads most messages v_j posted (i.e., $RM_{ij} \approx 1$), it indicates that v_i pays much attention to v_j. However, in such situation, we still cannot make the conclusion that v_i pays abnormal attention to v_j. We have to consider their Physical Distance and Interaction Distance, as v_i and v_j can be close friends or always frequently interact with each other. Therefore, PD_{ij}, ID_{ij} and RM_{ij} are all affecting the detection result of abnormal attention.

3.3 Fuzzy-based Abnormal Attention Detection

As discussed in the previous section, PD_{ij}, ID_{ij} and RM_{ij} all need to be considered for abnormal attention detection. In this research, we adopt fuzzy logic (L. A. Zadeh, 1965) (L. A. Zadeh, 1978), and aim to establish a computational model to capture the linguistic states of these three factors.

In the proposed fuzzy-based approach, PD_{ij}, ID_{ij} and RM_{ij} are input parameters. The output from the fuzzy approach is Excessive Attention Index (EA_{ij}), which indicates the excessive attention v_i pays to v_j at time stamp t.

3.3.1 Fuzzy Membership Functions

Input Parameters

We regard Physical Distance (PD_{ij}), Interaction Distance (ID_{ij}) and Message Reading Ratio (RM_{ij}) as three input parameters in the fuzzy approach.

In terms of PD_{ij}, we define three fuzzy sets, i.e., “Close”, “Medium” and “Far”, to capture the linguistic meanings of the Physical Distance. The membership functions are shown from Equations 3.3 to 3.5.

$$F_{PDClose}(x) = \begin{cases} 1, & x \in [0,1] \\ -x+2, & x \in [1,2] \end{cases} \tag{3.3}$$

$$F_{PDMedium}(x) = \begin{cases} x-1, & x \in [1,2] \\ 1, & x \in [2,4] \\ -x+5, & x \in [4,5] \end{cases} \tag{3.4}$$

$$F_{PDFar}(x) = \begin{cases} x-4, & x \in [4,5] \\ 1, & x \in [5,\infty) \end{cases} \tag{3.5}$$

ID_{ij} is fuzzified based on the following three fuzzy sets: “Frequent”, “Moderate” and “Seldom”. The membership functions for the fuzzy sets are shown from Equations 3.6 to 3.8, respectively.

$$F_{IDSeldom}(x) = \begin{cases} 1, & x \in [0,0.1] \\ -10x+2, & x \in [0.1,0.2] \end{cases} \tag{3.6}$$

$$F_{IDModerate}(x) = \begin{cases} 10x-\frac{3}{2}, & x \in [0.15,0.25] \\ 1, & x \in [0.25,0.4] \\ -5x+3, & x \in [0.4,0.6] \end{cases} \tag{3.7}$$

$$F_{IDFrequent}(x) = \begin{cases} 10x-5, & x \in [0.5,0.6] \\ 1, & x \in [0.6,1.0) \end{cases} \tag{3.8}$$

Apart from that, for RM_{ij}, linguistic meanings are defined as “High", “Medium" and “Low". The membership functions for these three fuzzy sets are defined from

Equations 3.9 to 3.11.

$$F_{RMLow}(x) = \begin{cases} 1, & x \in [0,0.2] \\ -10x + 3, & x \in [0.2,0.3] \end{cases} \tag{3.9}$$

$$F_{RMMiddle}(x) = \begin{cases} 10x - \frac{5}{2}, & x \in [0.25,0.35] \\ 1, & x \in [0.35,0.55] \\ -10x - \frac{13}{2}, & x \in [0.55,0.65] \end{cases} \tag{3.10}$$

$$F_{RMHigh}(x) = \begin{cases} 10x - 6, & x \in [0.6,0.7] \\ 1, & x \in [0.7,1.0] \end{cases} \tag{3.11}$$

Output Parameter

Excessive Attention Index (EA_{ij}) is the output parameter in the fuzzy approach. We fuzzify it based on the five fuzzy sets, i.e., “low”, “more or less low”, “medium”, “high”, “very high”, and their degrees of memberships are calculated from Equations 3.12 to 3.16.

$$F_{AILow}(x) = \begin{cases} 1, & x \in [0,0.08] \\ -\frac{100}{7}x + \frac{15}{7}, & x \in [0.08,0.15] \end{cases} \tag{3.12}$$

$$F_{AILessLow}(x) = \begin{cases} \frac{25}{3}x - \frac{2}{3}, & x \in [0.08,0.2] \\ 1, & x \in [0.2,0.3] \\ -20x + 7, & x \in [0.3,0.35] \end{cases} \tag{3.13}$$

$$F_{AIMedium}(x) = \begin{cases} 10x - 3, & x \in [0.3,0.4] \\ 1, & x \in [0.4,0.55] \\ -10x + \frac{13}{2}, & x \in [0.55,0.65] \end{cases} \tag{3.14}$$

$$F_{AIHigh}(x) = \begin{cases} 10x - 6, & x \in [0.6,0.7] \\ 1, & x \in [0.7,0.8] \\ -20x + 17, & x \in [0.8,0.85] \end{cases} \tag{3.15}$$

$$F_{AIVeryHigh}(x) = \begin{cases} \frac{20}{3}x - 5, & x \in [0.75,0.9] \\ 1, & x \in [0.9,1.0] \end{cases} \tag{3.16}$$

3.3.2 Fuzzy Inference

We perform the fuzzy reasoning to evaluate the Excessive Attention Index between two agents based on the three fuzzy inputs. The fuzzy rules are represented by a three dimensional matrix, and shown from Tables 3.1 to 3.3 (each table is for PD_{ij} with a particular value).

Table 3.1: Fuzzy Rule Base Matrix when PD_{ij} is "Close"

ID_{ij} \ RM_{ij}	High	Middle	Low
Frequent	Low	Low	Low
Moderate	More or less low	Low	Low
Seldom	More or less low	Low	Low

Based on Tables 3.1 to 3.3, we can find the fuzzy rules to infer the attention index as "low", "more or less low", "medium", "high" and "very high".

We adopt one of the most commonly used compositional operation Max-min operation (L. Zadeh, 1973), to calculate EA_{ij}^{t} values. The output membership degree

Table 3.2: Fuzzy Rule Base Matrix when PD_{ij} is "Medium"

ID_{ij} \ RM_{ij}	High	Middle	Low
Frequent	Medium	More or less low	More or less low
Moderate	High	Medium	More or less low
Seldom	High	Medium	Medium

Table 3.3: Fuzzy Rule Base Matrix when PD_{ij} is "Far"

ID_{ij} \ RM_{ij}	High	Middle	Low
Frequent	High	Medium	More or less low
Moderate	Very High	High	Medium
Seldom	Very High	Very High	High

$\mu(EA_{ij}^t)$ can be calculated from Equations 3.17 to 3.18.

$$\begin{aligned} Min_\delta = &(\mu_\alpha(PD_{ij}), \mu_\beta(ID_{ij}), \\ &\mu_\gamma(RM_{ij})) \end{aligned} \tag{3.17}$$

$$\mu(EA_{ij}^t) = Max(Min_\delta) \tag{3.18}$$

3.3.3 Defuzzification

In the previous subsection, we define the linguistic states mapped to the fuzzy sets and fuzzy rules. We also need to obtain a real value for each time stamp. Here, we adopt Center of Area (COA) defuzzification method (Wang, 1992) to calculate the value. The defuzzification equation is shown in Equation3.19.

$$EA_{ij}^t = \frac{\sum_{i=1}^{N} y_i \times \mu_{EA_{ij}^t}(y_i)}{\sum_{i=1}^{N} \mu_{EA_{ij}^t}(y_i)} \tag{3.19}$$

, where y_i is the output from the output membership functions (refer to Equations 3.12 to 3.16); $\mu_{EA^t_{ij}}$ is the membership degree (refer to Equation 3.18). The real value of EA^t_{ij} can be calculated. N is the total number of fuzzy rules y_i satisfies.

Furthermore, the Average Excessive Attention Index v_j received from all other agents in the t^{th} time stamp $\overline{EA^t_j}$ can be calculated by using Equation 3.20.

$$\overline{EA^t_j} = \frac{\sum_{i=i}^{N} EA^t_{ij} - Max_{EA^t_{ij}} - Min_{EA^t_{ij}}}{N-2} \tag{3.20}$$

, where $Max_{EA^t_{ij}}$ is the maximum value of EA^t_{ij} from all agents to v_j in the t^{th} time stamp. $Min_{EA^t_{ij}}$ is the minimum value of EA^t_{ij} from all agents to v_j.

Definition 5 *The Difference Excessive Attention Index DEA^t_{ij} is the difference of EA^t_{ij} and $\overline{EA^t_j}$ at the t^{th} time stamp. DEA^t_{ij} can be calculated by using Equation 3.21.*

$$DEA^t_{ij} = \begin{cases} \dfrac{EA^t_{ij} - \overline{EA^t_j}}{\overline{EA^t_j}} & EA^t_{ij} > \overline{EA^t_j} \\ 0 & EA^t_{ij} \leq \overline{EA^t_j} \end{cases} \tag{3.21}$$

We consider that DEA^t_{ij} indicates the degree of abnormal attention v_i pays to v_j at time stamp t. From Equation 3.21, it can be seen that EA^t_{ij} contributes to DEA^t_{ij} when it is greater than $\overline{EA^t_j}$. Namely, DEA^t_{ij} is 0 when the attention paid by v_i to v_j is less than average.

Definition 6 *The Accumulated Difference Excessive Attention Index ($ADEA^t_{ij}$) from v_i to v_j at time stamp $t = n$ ($n \geq 0$) is the accumulated value of DEA^t_{ij}. It indicates the accumulated abnormal attention v_i pays to v_j in the period from $t = 0$ to $t = n$. It can be calculated by using Equation 3.22.*

$$ADEA^t_{ij} = \sum_{t=0}^{n} e^{-\frac{\Delta t}{\lambda}} DEA^t_{ij} \tag{3.22}$$

, where Δt is the time difference between the t^{th} time stamp and the beginning. We also introduce a diminishing factor, i.e., $e^{-\frac{\Delta t}{\lambda}}$, to gradually decrease the impact of abnormal attention over time. We adopt weighted moving average method (Holt, 2004), in order to strengthen the influence of data which is close to the current time stamp and eliminate the effect of out of date data.

$ADEA_{ij}^{t}$ indicates the accumulated abnormal attention from v_i to v_j, when $ADEA_{ij}^{t}$ becomes larger and quickly increases, it means v_i pays abnormal attention to v_j continually in the period of Δt.

3.4 Experiments and Analysis

We performed experiments on synthetic data generated based on a real-world network topology. We collected network topology data from website and generated data as users' behaviours in OSNs. In the experiments, we considered two types of users, i.e., ordinary users and star users. Ordinary users follow others and at the same time are followed by other people. Ordinary users' indegrees are no larger than their outdegrees. Star users want their fans to pay high attention to them, and their indegrees are much larger than their outdegrees. Ordinary users and star users represent two typical types of users in OSNs according to the number of users they followed or followed by others. We aim to indicate our model can fit different types of users in different situations.

In the experiments, the network topology is extracted from the Blogs network graph from the Konect dataset (*Konect Network Dataset*, n.d.). This directed network contains front-page hyperlinks between blogs in the 2004 US election. It contains 1224 vertices and 19025 edges.

Users can have normal behaviours and abnormal behaviours in the network. For normal behaviours, users randomly post messages, read messages and interact with others. For abnormal behaviours, users still randomly post messages, but purposely

read messages and interact with some particular users. For example, we assume v_i pays abnormal attention to v_j. v_i reads most messages posted by v_j and his closest friends. At the same time, v_i seldom has interactions with v_j. Namely, v_i hides himself/herself in the social network but continually and incredibly reads the messages from v_j.

3.4.1 Experiment 1: Ordinary and Star Users under the Normal Attention

Experiment 1 aims to indicate our model fits users in different social statuses. We generate normal behaviours for all agents. They all randomly post messages, read messages and interact with each other. Then we selected a star user v_s and a ordinary user v_o, and analyse their received attention from one particular user v_i.

In Fig. 3.3, we plot the Excessive Attention Index (EA^t_{ij}) from the local view of a star user and an ordinary user. The horizontal axis represents the time stamps while the vertical axis indicates the ranges of Excessive Attention Index EA^t_{ij} which users received from other users. The plot clearly demonstrates that the star user received much higher attention than the ordinary user. The value of EA^t_{is} is from 35 to 40, while the value of EA^tio is from 0 to 5. The value of EA^t_{is} is nearly 9 times higher than the value of EA^t_{io} in each time stamp. This is because that, when a user is a star user in the network, it means that the user has high in-degree and more users will pay higher attention to him/her. In Fig. 3.4, we plot the Difference Excessive Attention Index DEA^t_{ij}, which calculates the differential ratio of EA^t_{ij} and $\overline{EA^t_j}$. Though the curve of ordinary user has sharply increased and decreased, the value of DEA^t_{ij} always stays in at low level which is less than 0.8. For the star user, the DEA^t_{ij} values are also at a low level, even when the EA^t_{ij} values are high (refer to Fig. 3.3).

Fig. 3.5 plots the Accumulated Difference Excessive Attention Index ($ADEA^t_{ij}$) that the star user and ordinary user received from one agent. The curves are corresponding

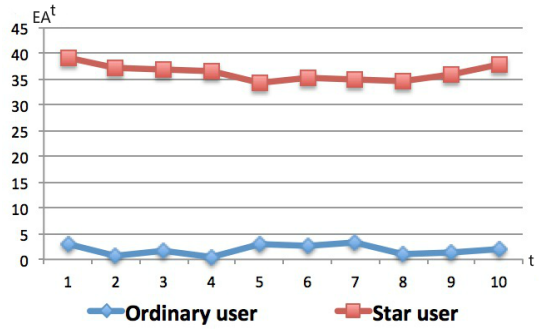

Figure 3.3: The Excessive Attention Index star user and ordinary user received from normal behaviours

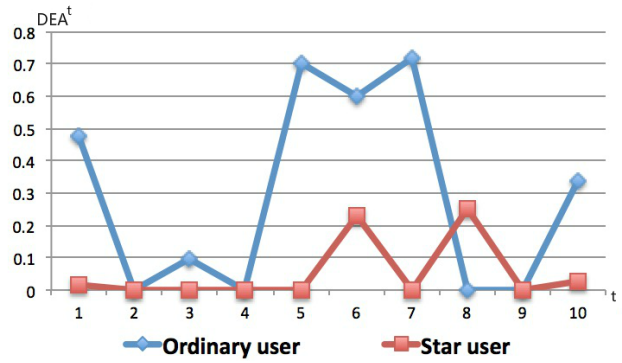

Figure 3.4: The Difference Excessive Attention star user and ordinary user received from normal behaviours

to the tendency of the star user and ordinary user respectively. For the star user, the values are between 0 to 0.5. For the ordinary user, though the $ADEA_{ij}^{t}$ value is higher than the star user, it is below 2 and fluctuates which is in the normal range.

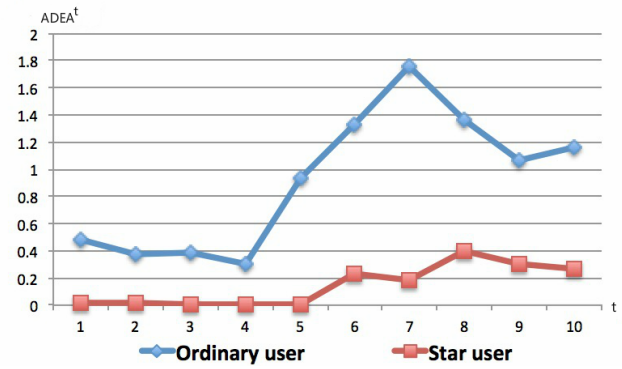

Figure 3.5: The Accumulated Difference Excessive Attention star user and ordinary user received from normal behaviours

3.4.2 Experiment 2: Detect Abnormal Attention

In Experiment 2, we focus on how to detect abnormal attention. We set one user v_a pays abnormal attention to another user v_b while all other agents (e.g., v_c) have normal behaviours in the network.

Fig. 3.6 shows the values of Excessive Attention Index of users under normal attention (EA^t_{ac}) and abnormal attention (EA^t_{ab}). We compare the value of the top line which indicates the agent received attention from abnormal behaviours with the value of the bottom line which is from normal behaviours. The line corresponding to the agent who has abnormal behaviours on another agent always has higher EA^t_{ab} value than the one who has normal behaviours in each time stamp.

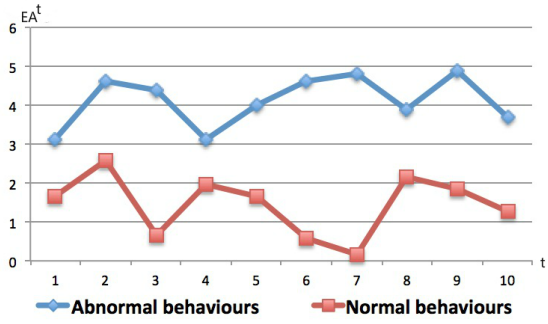

Figure 3.6: The Excessive Attention Index from abnormal behaviours and normal behaviours

Fig. 3.7 shows the Difference Excessive Attention Index. If a user has abnormal behaviours, he might pay higher attention to his target. The value of DEA^t_{ab} of abnormal behaviours are all positive, it means the user always pays higher attention to the target and higher than the target's social status. The value of DEA^t_{ac} of normal behaviours are located around 0 in the 6 time stamps during the 10 time stamps, which indicates the user pays normal attention to the target.

In Fig. 3.8, we plot the Accumulated Difference Excessive Attention Index versus time. With the time goes by, both two curves show the different tendency. Since the beginning, the two curves start from a low value which represents that the agent received nearly similar Excessive Attention Index from abnormal behaviours and normal

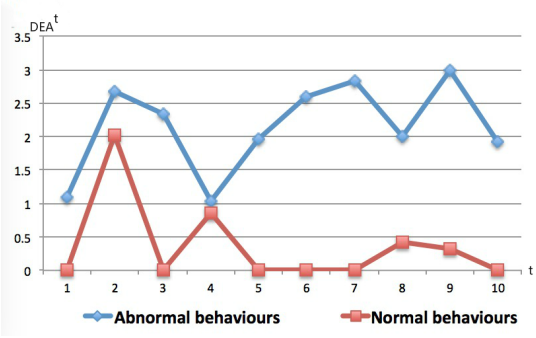

Figure 3.7: The Difference Excessive Attention Index from abnormal behaviours and normal behaviours

behaviours. The value of the top line which represents abnormal behaviours dramatically increases in a continuous style, which represents v_i continuously pays higher attention to the target. The value of the bottom line which represents normal behaviours has fluctuated which has slowly increased and slightly decreased, also it stays between 0 and 1.5. Hence our model clearly detects the abnormal attention in the network.

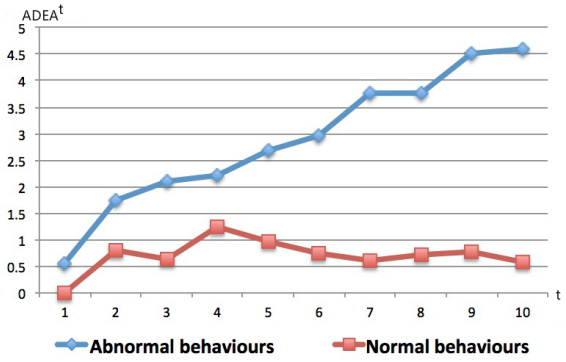

Figure 3.8: The Accumulated Difference Excessive Attention Index from abnormal behaviours and normal behaviours

3.5 Summary

In this chapter, we proposed a fuzzy logic-based approach for detecting abnormal attention in OSNs. Experimental results demonstrated that the proposed approach can effectively detect abnormal attention in different situations with local information of

individual users. Namely, it can achieve abnormal attention detection without a global view or the invasion of other users' privacy. We claim that the proposed approach is more suitable for real-world OSN applications.

Chapter 4

A Measure of Online Interpersonal Surveillance from Global Views

4.1 Introduction

In Chapter 3, we focused on detecting the abnormal attention from agents' local views. Though we protect personal privacy, it is extremely hard to detect online interpersonal surveillance behaviours from local views. Online interpersonal surveillance refers to the excessive, unreciprocated and persistent pursuit of a person with abnormal attention through cyberspace. It does not necessarily lead to harassment and harm (Spitzberg & Hoobler, 2002). Moreover, the surveillance may not all be negative as in the case of e.g. government surveilling potential criminal activities. For convenience, we call the person who carries out the monitoring a watcher, and the person being watched the target. We aim to detect interpersonal surveillance from global views and suppose we can gain the whole information from OSNs.

Investigators of online interpersonal surveillance face several severe challenges: (1) Online users may purposely expose private information online, making surveillance extremely easy; any online user may derive information regarding another's occupation,

habits, and location with little effort. (2) The watcher may be at an arbitrary physical and social distance from the target, making detection extremely difficult. (3) Interpersonal surveillance is largely *victim-defined*, that is, the level of seriousness of a particular incident is determined by how much intimidation the target perceives upon its exposure (Reno, 1999); hence the boundary between surveillance and benevolent information seeking is hard to define. Despite serious research efforts, there have not been major technical advancements in the detection and prevention of interpersonal surveillance (Pittaro, 2007). In view that anyone has the potential to be a watcher, one needs a *behavioral* approach rather than summarizing personal attributes.

Hence, we need to collect the detailed propagated information and users' behaviours in the OSNs from a global view. If we only focus on a user's local view which is based on the user himself/herself, the user only knows what happened to his/her neighbours. We want to detect the agent who is surveilling others. In this situation, we have to let users know not only their neighbours' behaviours but also behaviours of their neighbours' neighbours. Namely, only from a global view, we can well define interpersonal surveillance.

In this Chapter, we focus on the problem of Online Interpersonal Surveillance which is a more serious problem than a user paying abnormal attention to others. Towards modelling Interpersonal Surveillance, we give a quantitative definition of the attention from a global view that a user pays to others from message-based and Network-based aspects. We suppose we can collect the whole information that posted and read by the whole users in OSNs.

The rest of this chapter is organised as follows: Section 4.2 focuses on social attention from a global view, which forms the core of our model. We support our model with several case studies on special network topologies. Section 4.3 presents the measure of interpersonal surveillance. Section 4.4 discusses simulation and experimental results on both synthesized and real-world networks. Finally, Section 4.5 closes this chapter.

4.2 Social Attention Model from Global Views

We consider two perspectives of social attention from agent a to b. The first is a micro perspective: the attention a pays to b depends on the posts and reads of a and b. The second is a macro perspective: the link structure of the network affects attention from a to b. We propose a definition that takes into account both factors.

As we have defined the model of social network and user interactions in previous Chapter, we further require that for all $v \in V$, $\mathsf{post}(v) \cap \mathsf{read}(v) = \varnothing$. It is important to note that while $post(v)$ are usually maintained and accessible by OSNs, extracting $read(v)$ is much harder. As studied in (Tagarelli & Interdonato, 2015), a large amount of Internet users are apparently inactive and restricts their online activities to reading others' posts. Nevertheless, it is not unimaginable to build functionalities into an OSN which records when a user clicks on another's profile or posts (Chaabane, Kaafar & Boreli, 2012). Therefore, we abstractly characterize the states of an OSN using the *post* and *read* functions and assume them accessible through the network logs.

Message-based attention In the micro level; we measure attention by comparing the set $\mathsf{post}(b)$ against the set $\mathsf{read}(a)$. If $\mathsf{post}(b)$ and $\mathsf{read}(a)$ are similar, then the agent a reads mostly the posts of b, which naturally implies that a pays attention to the messages of b. To measure similarity between two sets, we adopt the well-known *Jaccard distance*.

Definition 7 *Given an instance* $s = (\mathsf{post}, \mathsf{read})$ *of a network* G*, the* $\mathsf{M}\text{-index}$ *from* a *to* b *is defined as*

$$m_s(a,b) = \begin{cases} \frac{|\mathsf{read}(a) \cap \mathsf{post}(b)|}{|\mathsf{read}(a) \cup \mathsf{post}(b)|} & \text{if } a \neq b, \\ 1 & \text{otherwise} \end{cases} \tag{4.1}$$

Network-based attention The $\mathsf{M}\text{-index}$ only measures an agent's interests in the messages posted by another agent, which is not necessarily identical to the attention on

the target agent. For example, a may read messages posted by b in the hope to monitor a friend c of b; while no message posted by c is read by a, a's primary intention is the surveillance of c. This method of "indirect" information gathering has been regarded as a great privacy risk (Mislove, Viswanath, Gummadi & Druschel, 2010). We next define a Markov chain model to capture the network effect on social attention.

4.2.1 Network topology effect of social attention

A *stochastic (row) vector* is a vector whose entries are non-negative real numbers and add up to 1. A *stochastic matrix* is a square matrix where each row is a stochastic vector.

Definition 8 *A (finite state, time-homogeneous)* Markov chain *is a tuple* $M = (Q, P)$ *where* $Q = \{q_1, q_2, \ldots, q_n\}$ *is a finite set of states and* $P = [P_{ij}]$ *is a* $n \times n$ *stochastic matrix, called the* transition matrix *of* M.

The Markov chain M represents discrete-time stochastic process $X_0, X_1, \ldots$ where $P_{ij} \in [0, 1]$ represents the *transition probability* of the process from state q_i to state q_j. Let $\vec{v}_t = (v_{t,1}, \ldots, v_{t,n})$ denote the probability distribution of the stochastic process at time t, i.e., $v_{t,i}$ represents the probability that at time t the process is in state q_i. Then $\vec{v}_t = \vec{v}_{t-1} \cdot P$. A state $q \in Q$ is said to be *accessible* from a state p, written $p \to q$, if there is $t > 0$ with $\mathsf{Pr}(X_t = q \mid X_0 = p) > 0$. Two states p, q are said to be in the same *communication class* if both $p \to q$ and $q \to p$. A communication class is *closed* if the probability of leaving the class is zero.

The Markov chain (Q, P) is *primitive* if there is some $t > 0$ such that every entry in the tth-power P^t is positive. When the Markov chain consists of more than one communication class, we say that a communication class C is *primitive* if the sub-matrix of P restricted to elements in C is primitive.

Definition 9 *The Markov chain* $M = (Q, P)$ *is* centered *if only one communication class of* M *is closed, and the only closed communication class is primitive.*

In particular, any primitive Markov chain is centered. The following is a special case of the Perron-Frobenius theorem, a classical result on the limiting behavior of linear systems.

Theorem 1 (Perron-Frobenius (Meyer, 2000)) *Let $M = (Q, P)$ be a centered Markov chain and the initial state of the chain $\overrightarrow{v_0}$. There is a unique row vector $\overrightarrow{\pi_M}$, called* stationary distribution *of M, such that $\lim_{t\to\infty} P^t = \overrightarrow{1} \cdot \overrightarrow{\pi_M}$ where $\overrightarrow{1}$ is the column vector with all entries 1 and $\overrightarrow{\pi_M}$ denotes a stochastic vector where $\overrightarrow{\pi_M} \cdot P = \overrightarrow{\pi_M}$.*

Now we introduce a Markov chain model of social attention. Let $G = (V, E)$ be a network with instance s. In Table 4.1, we summarize the variables used in this section.

Table 4.1: Variables and their definitions

a, b, c	Agents in V
$m_s(a, b)$	the M-index of a, b
E_a	the set of edges $E \setminus \{ba \mid b \in V\}$
r	correlation ratio in the range $(0, 1]$
$\deg_a(b)$	Relative degree of b=$\lvert\{c \mid bc \in E_a\}\rvert$+$r\cdot\lvert\{c \mid bc \notin E_a\}\rvert$
$T_{s,a}$	Attention transfer matrix for a in instance s
$\delta_a(b)$	$\lvert\{c \mid bc \in E_a\}\rvert / \deg_a(b)$
$W_{s,a}(b)$	$\sum_{bc\in E_a}(m_s(a, c) + 1)$
$W'_{s,a}(b)$	$\sum_{bc\notin E_a} m_s(a, c)$

Fix an agent a, we aim to measure how much attention a pays to other agents. We conceptualize the following *attention allocation stochastic process*: For every $b \in V$, define a random variable $X_{a,b,t}$ for the amount of attention a pays to b at step t. At step $t = 0$, we assign $X_{a,a,0} = 1$ and $X_{a,b,0} = 0$ for $b \neq a$. Suppose the attention at step $t \geq 0$ is determined. We define the allocation at step $t + 1$ based on two intuitions:

1. The first intuition is about M-index. Let c be an agent such that $m_s(a, c) > 0$. At step $t + 1$, a certain amount of attention of a will be moved from any agent $b \in V$ to c; this amount depends on the value of $m_s(a, c)$: a higher value of $m_s(a, c)$ means c receives more attention.

2. The second intuition is about link structure of G. If bc is an edge in G. At step $t+1$, a certain amount of attention of a will be moved from b to c. The amount depends on the out-degree of b as well as the M-index.

The stochastic process does not involve incoming transitions to a itself and thus the edge set becomes E_a. If $bc \notin E_a$, we think of b, c as linked by an *absent tie*: certain attention of b from a may still be transferred to c (Granovetter, 1973). However, the effect of such absent ties is naturally weaker than the actual edges. Hence we use the correlation ratio r to define the *relative degree* $\deg(b)$ of b as in Table 4.1. We define the matrix *attention transfer matrix* $T_{s,a}$ as follows: For any $b, c \in V$,

(1) if $bc \in E_a$, the percentage of attention transferred from b to c is $T_s(a)[b,c] = \delta_a(b) \cdot \frac{m_s(a,c)+1}{W_{s,a}(b)}$ Note that if $m_s(a,c) = 0$, $\delta_a(b)/W_{s,a}(b)$ of the attention on b is still transferred to c due to the edge bc;

(2) if $bc \notin E_a$, the percentage of attention transferred from b to c is $T_s(a)[b,c] = (1-\delta_a(b)) \cdot \frac{m_s(a,c)}{W'_a(b)}$. Contrary to (1), if $m_s(a,c) = 0$, then no attention is transferred to c.

Lemma 1 *For any network G, any instance s of G and $a \in V$, the attention transfer matrix $T_{s,a}$ is a stochastic matrix.*

Proof 1 *Take an agent $b \in V$. By definition, $W_{s,a}(b) > 0$ if and only if b has an outgoing edge. If b has no outgoing edge, then $\sum_{bc \in E_a} T_{s,a}[b,c] = 0 = \delta_a(b)$. If b has some outgoing edges,*

$$\sum_{bc \in E_a} T_{s,a}[b,c] = \frac{\delta_a(b)}{W_{s,a}(b)} \cdot \sum_{bc \in E_a} (m_s(a,c)+1) = \delta_a(b)$$
$$\sum_{bc \notin E_a} T_{s,a}[b,c] = \frac{1-\delta_a(b)}{W'_{s,a}(b)} \cdot \sum_{bc \notin E_a} m_s(a,c) = 1-\delta_a(b)$$

Thus $\sum_{c \in V} T_{s,a}[b,c] = 1$.

Lemma 1 implies that $(V, T_{s,a})$ is a Markov chain.

Lemma 2 *The Markov chain $(V, T_{s,a})$ is centered.*

Proof 2 *For any $b \in V$, since $T_{s,a}[b, a] = 1 - \delta_a(b) > 0$, $b \to a$. Let $C_a \subseteq V$ be the set of all agents accessible from a. Then C_a forms a closed communication class. Also, as the diagonal entry in $T_{s,a}$ corresponding to a is positive, C_a is primitive. Furthermore, any other communication class in the Markov chain is not closed as they can all access a. Hence the Markov chain is centered.*

Combining Lemma 2 and Theorem 1, we obtain:

Theorem 2 *For any agent a, the Markov chain $(V, T_{s,a})$ has a stationary distribution $\overrightarrow{\pi_a}$.*

Definition 10 *The* attention $A_s(a, b)$ *that agent a pays to b in instance s is defined as the entry $\overrightarrow{\pi_a}(b)$ in the stationary distribution $\overrightarrow{\pi_a}$.*

4.2.2 Case Studies: Attention in Special Network Topologies

We study social attention in special network topologies. Through the case studies, we show that our model stays consistent with intuition. Since the total attention of an agent a sums to 1, we are only interested in the ratio $A_s(a, b) : A_s(a, c)$ between the attention that a pays to b and to c.

Isolated agent networks

We first look at a network that does not contain any edges, i.e., $E = \varnothing$. In this case, M-index naturally becomes the sole indicator of attention, as shown in the next theorem.

Theorem 3 *Let G be $(V, \varnothing)$. For any $a \in V$, the attention that a pays to others satisfies:* $\forall b, c \in V,\ A_s(a, b) : A_s(a, c) = m_s(a, b) : m_s(a, c)$

Proof 3 *Suppose* $V = \{1, \ldots, n\}$. *Fix* $a \in V$. *Let* $\sigma_a = \sum_{c \in V} m_s(a, c)$. *Then* $W_{s,a}(b) = 0$ *and* $W'_{s,a}(b) = \sigma_a$ *for any* $b \in V$. *The* b*th row in the matrix* $T_{s,a}$ *is* $\left(\frac{m_s(a,1)}{\sigma_a}, \frac{m_s(a,2)}{\sigma_a}, \ldots, \frac{m_s(a,n)}{\sigma_a}\right)$. *Let* $\overrightarrow{\pi_a}$ *be the stationary vector of* $T_{s,a}$. *We have* $\overrightarrow{\pi_a} \cdot T_{s,a} = \overrightarrow{\pi_a}$. *Hence for any* $b \in V$,

$$\begin{aligned}\pi_{a,b} &= \pi_{a,1} \cdot \frac{m_s(a,b)}{\sigma_a} + \cdots + \pi_{a,n} \cdot \frac{m_s(a,b)}{\sigma_a} \\ &= \frac{(\pi_{a,1} + \cdots + \pi_{a,n})}{\sigma_a} \cdot m_s(a,b)\end{aligned}$$

Therefore $\forall b, c \in V$, $\pi_{a,b} : \pi_{b,c} = m_s(a,b) : m_s(a,c)$.

Loner with a clique

We consider network G that contains an isolated agent 0 and n others $1, 2, \ldots, n$ who form a complete graph of order n (i.e. an n-*clique*). Detached from all others, 0 is a *loner* in the network. Suppose 0 pays equal amount of attention to messages posted by any other agent, i.e., for some fixed value k, $\forall 1 \leq b \leq n, m_s(0,b) = k$.

Theorem 4 *Suppose* G *is a loner with clique as above. Then* $A_s(0,0) : A_s(0,b) = 2r(kn+1) : k(1+k)(2r+n-1)$ *for any agent* $b \neq 0$.

Proof 4 *Since 0 is an isolated node,* $W_{s,0}(0) = 0$ *and* $W'_{s,0}(0) = kn + 1$. *Hence the first row of the matrix* $T_{s,0}$ *is* $\left(\frac{1}{kn+1}, \frac{k}{kn+1}, \ldots, \frac{k}{kn+1}\right)$.

Take any $b \in \{1, \ldots, n\}$. *Since* b *has an edge to any other nodes in* $\{1, \ldots, n\}$, $W_0(b) = (n-1)(k+1)$ *and* $W'_{s,0}(b) = 1 + k$. *Furthermore,* $\delta_0(b) = \frac{n-1}{2r+n-1}$. *Hence* $T_{s,0}[b,0] = \frac{2r}{(1+k)(2r+n-1)}$, $T_{s,0}[b,b] = \frac{2rk}{(1+k)(2r+n-1)}$ *and for all* $1 \leq c \neq b \leq n$, *we have*

$$T_{s,0}[b,c] = \frac{(n-1)(k+1)}{(n-1)(k+1)(2r+n-1)} = \frac{1}{2r+n-1}.$$

Let x *be the amount of attention* 0 *pays to herself, and* y *be the amount of attention* 0

pays to others. Then we have

$$(x, y, \ldots, y) \cdot T_{s,0} = (x, y, \ldots, y),$$

corresponding to the following system of linear equations:

$$\begin{cases} \frac{1}{kn+1} \cdot x + \frac{2rn}{(1+k)(2r+n-1)} \cdot y = x \\ \frac{k}{kn+1} \cdot x + \frac{2rk}{(1+k)(2r+n-1)} \cdot y + \frac{n-1}{2r+n-1} \cdot y = y \end{cases}$$

Solving this system of linear equations, we get the desired result $x : y = 2r(kn+1) : k(1+k)(2r+n-1)$

To interpret Theorem 4, we consider two special cases:

(a) Suppose $k = 0$. Then 0 pays no attention to the messages posted by any agent in the n-clique. In this case, 0 only pays attention to herself as $\alpha_s(0,0) = 1$ and $\forall 1 \leq b \leq n, A_s(0,a) = 0$.

(b) Suppose $k = 1$. Then 0 pays full attention to others' posts. Now, $2r(kn+1) : (1+k)(2r+n-1) = rn+r : 2r+n-1 \leq 1$, which means the attention $A_s(0,0) \leq A_s(0,b)$ for any $1 \leq b \leq n$. In particular, as r gets closer to 0, the attention $A_s(0,0)$ 0 pays to herself decreases towards 0.

Loner with a star

An (in-coming) *star network* contains a node v with no outgoing edge, and all other nodes link to v via edges, i.e., $E = \{uv \mid u \neq v, u \in V\}$. Thus v is the *center*. Suppose G contains an isolated agent 0 and n other agents $1, 2, \ldots, n$ which form a star with center 1. We study attention from 0 and assume that $m_s(0,1) = 0$ and $m_s(0,a) = k$ for all $2 \leq a \leq n$. Despite not reading any message posted by 1, 0 may still pay attention to 1 as he reads messages from people linking to 1. This is verified by the next theorem.

Theorem 5 *Suppose G consists of a loner and a star as described above. Then for all agents $2 \le b \le n$, $A_s(0,0) : A_s(0,1) : A_s(0,b) = 1 + rn : k(n-1) : k(1+rn)$.*

Proof 5 *Since 0 and 1 do not have any outgoing edge, $W_{s,0}(0) = W_{s,0}(1) = 0$ and $W'_{s,0}(0) = W'_{s,0}(1) = k(n-1)+1$. Then the first and the second row of the matrix $T_{s,0}$ are both*

$$\left(\frac{1}{kn-k+1}, 0, \frac{k}{kn-k+1}, \ldots, \frac{k}{kn-k+1}\right)$$

For any $2 \le b \le n$, $W_{s,0}(b) = 1$, $W'_{s,0}(b) = k(n-1)+1$ and $\delta_0(b) = \frac{1}{rn+1}$. Thus the row corresponding to b in $T_{s,0}$ is

$$\left(\frac{rn}{(rn+1)(kn-k+1)}, \frac{1}{rn+1}, \frac{rnk}{(rn+1)(kn-k+1)}, \ldots \frac{rnk}{(rn+1)(kn-k+1)}\right)$$

Let $x = A_s(0,0)$, $y = A_s(0,1)$ and $z = A_s(0,2)$. Then

$$(x, y, z, \ldots, z) \cdot T_{s,0} = (x, y, z, \ldots, z)$$

Hence we arrive at the linear system

$$\begin{cases} \frac{x}{kn-k+1} + \frac{y}{kn-k+1} + \frac{zrn(n-1)}{(rn+1)(kn-k+1)} = x \\ \frac{z(n-1)}{rn+1} = y \\ \frac{xk}{kn-k+1} + \frac{yk}{kn-k+1} + \frac{zrnk(n-1)}{(rn+1)(kn-k+1)} = z \end{cases}$$

whose solution gives us $x : y : z = rn+1 : k(n-1) : k(rn+1)$, as required.

When $k = 1$, the ratio in Theorem 5 becomes $rn+1 : n-1 : rn+1$. In particular, when $r < 1$ and $n \ge \frac{2}{1-r}$, $k(n-1) \ge 1 + rn$. This shows that when the star is large enough, 0 puts more attention to the center 1 than to other agents.

Chain network

Lastly we study *chain networks*, which consist of nodes $0, 1, 2, \ldots, n$ and directed edges $(i-1)i$ for all $1 \le i \le n$. We assume that $m_s(0, a) = 1$ for all $a \in V$.

Theorem 6 (Chain network) *Suppose G is a chain network as above. Then $A_s(0,0) : A_s(0,1) : \cdots : A_s(0,n)$ is $1 : (1+\lambda) : (1+\lambda+\lambda^2) : \cdots : 1+\lambda+\cdots+\lambda^n$ where $\lambda = (1-r)/(rn+1)$.*

Proof 6 *For any $0 \le i < n$, $W_{s,0}(i) = 1$ and $W'_{s,0}(i) = n$ and $\delta_0(0) = \frac{1}{rn+1}$; therefore the ith row of the matrix $T_{s,0}$ is*

$$\left(\underbrace{\frac{r}{rn+1}, \ldots, \frac{r}{rn+1}}_{i+1}, \frac{1}{rn+1}, \underbrace{\frac{r}{rn+1}, \ldots, \frac{r}{rn+1}}_{n-i-1} \right)$$

We also have $W_{s,0}(0) = 0$ and $W'_{s,0}(n) = n+1$. Thus the last row of $T_{s,0}$ is

$$\left(\frac{1}{n+1}, \frac{1}{n+1}, \ldots, \frac{1}{n+1} \right)$$

Let $c_0, c_1, \ldots, c_n$ be the attention that 0 pays to $0, 1, 2, \ldots, n$, respectively. Then $(c_0, c_1, \ldots, c_n) \cdot T_{s,0} = (c_0, c_1, \ldots, c_n)$. Therefore

$$\begin{aligned} c_0 &= \frac{r}{rn+1}(c_0 + c_1 + \cdots + c_{n-1}) + \frac{c_n}{n+1} \\ c_1 &= \frac{r}{rn+1}(c_0 + c_1 + \cdots + c_{n-1}) + \frac{c_n}{n+1} + \frac{1-r}{rn+1}c_0 \end{aligned}$$

Hence $\frac{1-r}{rn+1}c_0 = c_1 - c_0$ and thus $c_1 = (\lambda+1)c_0$. Similarly, for all $1 \le i \le n$, we get $c_i = \frac{1-\lambda^{i+1}}{1-\lambda}c_0 = (1+\lambda+\lambda^2+\cdots+\lambda^i)c_0$, as required.

By Theorem 6, when $r < 1$, 0 pays more attention to nodes further down the chain, i.e., $A_{s,0}(i) < A_{s,0}(i+1)$ for $i \ge 1$. This is consistent with intuition, as the agents form a

chain of "following" relation. Furthermore, as the index i gets larger, the attention that 0 pays to i converges to $\lim_{n\to\infty} \sum_{i=0}^{n} \lambda^i A_s(0,0) = \frac{rn+1}{rn+r} A_s(0,0)$.

4.3 Measure Online Interpersonal Surveillance

Past research identified a common trait of watchers as emotionally distant loners who want to seek the attention and companionship of others (Spitzberg & Hoobler, 2002; Pittaro, 2007; McFarlane & Bocij, 2003). The problem lies in the fact that these individuals often become infatuated or obsessed with the target, but such a feeling is not reciprocated. Hence, it is reasonable to view interpersonal surveillance as *excessive* , *persistent* and *unreciprocated* attention. We next address each of these three attributes.

4.3.1 Excessive attention

As mentioned in Section 4.1, the judgement of excessive attention is victim-defined: For those who enjoy being in the spotlight (such as an entertainer), a great amount of attention is not problematic; for those who usually get little attention (such as a loner), even a small amount of exposure may seem to be too much. Hence, when judging whether an agent a pays excessive attention to another agent b, we should consider how much attention b generally receives.

Definition 11 *The* average attention *to agent b equals to*

$$A_s(b) = \frac{\sum_{c \in V} A_s(c,b)}{|V|} \tag{4.2}$$

The relativized attention *that a pays to b equals to*

$$A'_s(a,b) = \frac{A_s(a,b)}{A_s(b)} \tag{4.3}$$

4.3.2 Unreciprocated attention

A large amount of attention is not necessarily problematic. For example, the relation between intimate partners, closed friends, or family relatives, when the attention is mutual. It is only when the attention is unreciprocated that the agent feels unfulfilled. We can find that for all people in the world that it is not.

The *reciprocity* from agent a to b is Unreciprocated attention ($R_s(a,b)$). It can be calculated by using Equation 4.4

$$R_s(a,b) = A'_s(a,b) - A'_s(b,a) \tag{4.4}$$

Mutual attention between a, b happens in an instance s when $R_s(a,b) = 0$. A positive reciprocity of attention means a pays more attention to b than the attention that b pays to a.

4.3.3 Persistent attention

Surveillance consists of persistent attention that lasts a prolonged period of time. To capture the temporal effect of surveillance, we accumulate the attention from one agent to another over all instances of the network. For simplicity, we assume that the network G stays unchanged. Given a sequence of instances $s_0, s_1, \ldots$ of G, we define the *level of surveillance* from agent a to agent b as follows.

Definition 12 *At time stamp $t \geq 0$, the (t-step)* surveillance index $S_t(a,b)$ *from a to b is defined as*

$$R_{s_t}(a,b) + \frac{1}{2}R_{s_{t-1}}(a,b) + \cdots + \frac{1}{t+1}R_{s_0}(a,b). \tag{4.5}$$

Note that we impose a diminishing effect: the effect of attention in an earlier time will gradually decrease over time. The surveillance index is a numerical measure of the extend to which a is *watching* b. When a is not watching b, then we expect that the

reciprocity $R_s(a, b)$ to fluctuate around 0. When a is watching b, however, a pays a persistent, unreciprocated and excessive amount of attention to j, which would result in $R_s(a, b)$ to be consistently positive. Thus the surveillance index $S_t(a, b)$ will rise with t.

Definition 13 *Suppose G is a network and $s_0, s_1, s_2, s_3, \ldots$ is an infinite sequence of instances of G. For any agents a, b in G, we say that a is a* watcher *of b if*

$$\limsup_{t\to\infty} S_t(a, b) = \infty. \tag{4.6}$$

4.4 Simulation and Experiments

We performed simulation and experiments on both generated random networks and real world networks.

4.4.1 Simulating OSN

We used three random graph models to generate networks:

1) The first type (ER) is the Erdös-Rényi random graph model (Erdös & Rényi, 1959). The model takes as parameters a number n of agents and $p \in [0, 1]$. Edges are set up between nodes with probability p.

2) The second type (SF) is a scale-free graph model. Here, the fraction of nodes having k edges is proportional to $k^{-\gamma}$ for some $\gamma > 1$ (hence the degree distribution follows a power law) (Hein, Schwind & König, 2006). Our model is similar in spirit to Barabási-Albert preferential attachment model (Barabási & Albert, 1999) but builds a directed graph. The model takes as parameters a number n of agents and a smaller number $0 < m < n$. Initially, the procedure creates n agents $\{1, \ldots, n\}$ and links

consecutive agents with edges so that they form a cycle of length n. The procedure then links each node a to m randomly selected nodes, each time the probability of adding an edge ab is proportional to the number of incoming edges to b. This creates a *cumulated advantage* effect where the agent with a lot of incoming edges is more likely to get new edges.

3) The third type (SW) is the Watts-Strogatz model that builds a graph with small-world property (Watts & Strogatz, 1998). Small-world property is common in social networks and implies that the graph has small average path length and high clustering coefficient. The model takes as parameters a number n and a number $k < n$. The procedure first creates a ring lattice of size n where each node has degree k and then rewires each edge ab to another edge ac with probability p.

We defined instances of networks by generating $\mathsf{post}(a)$ and $\mathsf{read}(a)$ for each agent a. The set $\mathsf{post}(a)$ is randomly selected from a large pool of messages. For any pair (a, b) of distinct agents, we specify two types of reading behaviors:

1) The first type is *benevolent reading*. Here a does not purposely collect information from b and therefore the set $\mathsf{read}(a)$ is randomly picked.

2) The second type is *surveillance*. Here a purposefully collects information regarding b. In this case, a will tend to read more posts of b, as well as posts of those that are close b. Hence we compute the *distance* $\mathsf{dist}(c, b)$ from any node c to b. The likelihood of a reading a post by c decreases as $\mathsf{dist}(c, b)$ increases. To simulate a's read, we apply a *stochastic urn process* (Johnson & Kotz, 1977): We create a stochastic urn which initially contains all messages. In each iteration, we do the following:

1. randomly pick a message ω from the urn,
2. find the agent c with the smallest $\mathsf{dist}(c, b)$ where $\omega \in \mathsf{post}(c)$,
3. put a copy of each message $\omega' \in \mathsf{post}(d)$ such that $\mathsf{dist}(d, b) \leq \mathsf{dist}(c, b)$.

Experiment 1: Network typologies effect on average attention

We generate all three types of networks that contain 500 agents and instances where every agent practices benevolent reading (over 1000 messages). For the ER and SW graphs, the probability p is 50% and for the SF graph, each node makes 200 edges to others. This experiment looks at the average attention $A_s(a)$ received by each agent a in the network; See from Fig. 4.1 to Fig. 4.3. The horizontal axis represents the ranges of $A_s(a)$ while the vertical axis indicates the number of nodes in each range. The plot for $A_s(a)$ clearly indicates a power law distribution for the SF graph, but not for the ER and SW graphs. This shows that the scale-free property of the SF graph clearly affects the allocation of attention.

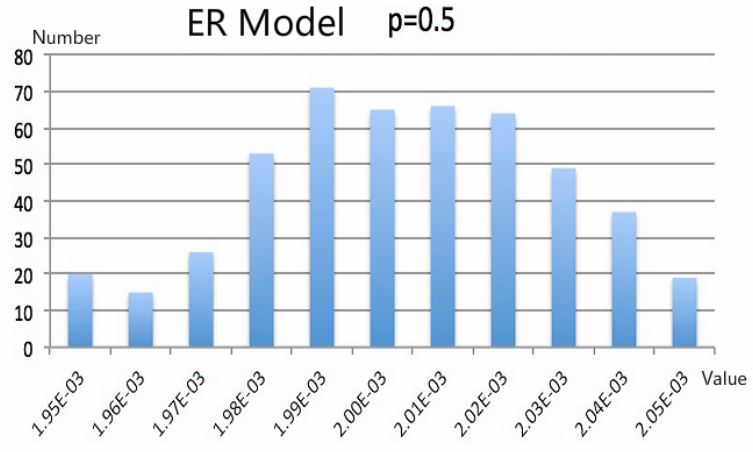

Figure 4.1: Distribution of average attention $A_s(a)$ in ER graph.

Experiment 2: In-degree effect on attention index

We then compare the amount of attention to agents with high and low in-degrees. When an agent has a high in-degree, it means that the agent is at the center of a large social circle and hence should also receive high attention. From Fig. 4.4 to Fig. 4.5, we plot the attention (in an SF graph) to two agents, one with in-degree 33 while the other has in-degree 398. The plot clearly demonstrates that a node's in-degree significantly

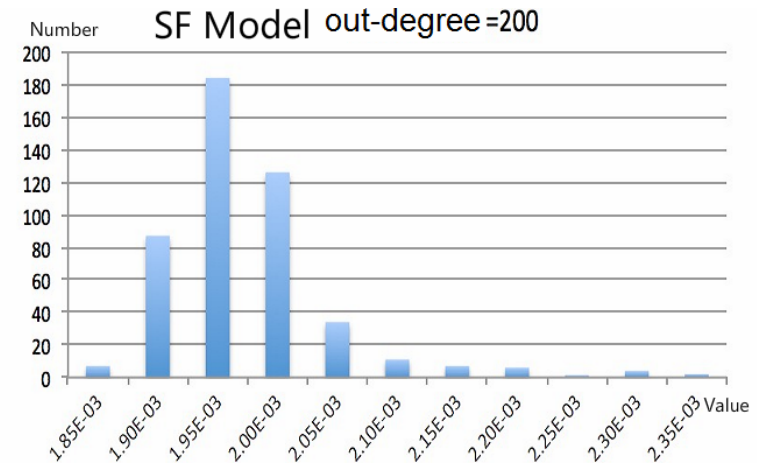

Figure 4.2: Distribution of average attention $A_s(a)$ in SF graph.

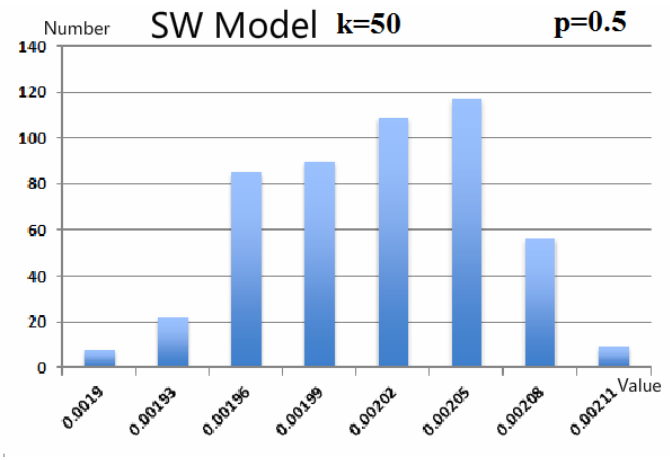

Figure 4.3: Distribution of average attention $A_s(a)$ in SW graph.

affects the attention to the agent: while nodes with higher in-degree gets consistently high attention from those around them, the attention from others to a node with lower in-degree varies considerably. The node with a higher degree gets consistently high degree from all others, while the node with a lower degree gets uneven attention from others.

Experiment 3: Network density effect on surveillance index

This experiment tests how network density impacts the surveillance index. Density refers to the number of edges in the graph. A high density means that in general users

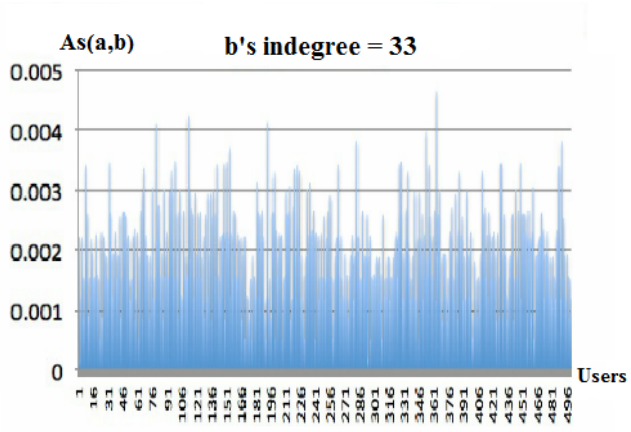

Figure 4.4: Attention to an agent with in-degree 33.

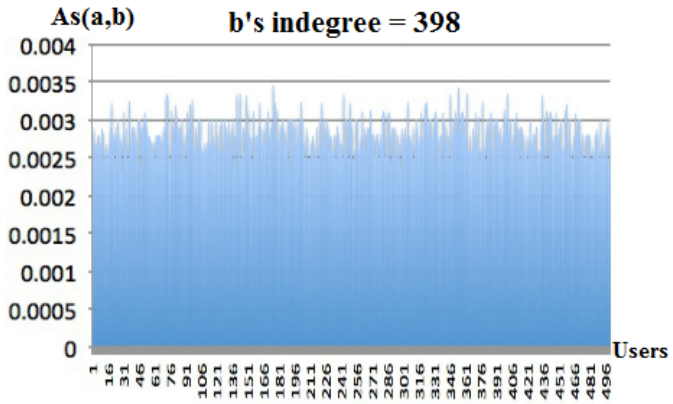

Figure 4.5: Attention to an agent with in-degree 398.

have more interpersonal ties and thus density indicates the level of user interactions in an OSN. We generate all types of random graphs with different parameters. In each of the graphs, we set one agent watching another agent, while all other agents are acting benevolently. See Fig. 4.6 to Fig. 4.8for plots of surveillance index against time. The line corresponding to the watcher dramatically increases in a continuous fashion, while all other line fluctuates around 0. Hence our model clearly separates the simulated cases of benevolent and surveillance behaviors.

For the ER model, we create graphs with edge probability $p \in \{0.15, 0.5, 0.85\}$. The result shows that as the graph gets denser, the surveillance index in both the benevolent

and surveillance cases gets smaller. This may be due to the increases in the average attention to all agents. For the SF model, we create graphs with different out-degree of nodes $m \in \{100, 200, 300\}$. For the SW model, we create graphs with different average degree of nodes $k \in \{20, 100, 180\}$. There is no significant change in the surveillance index in both the benevolent and surveillance cases.

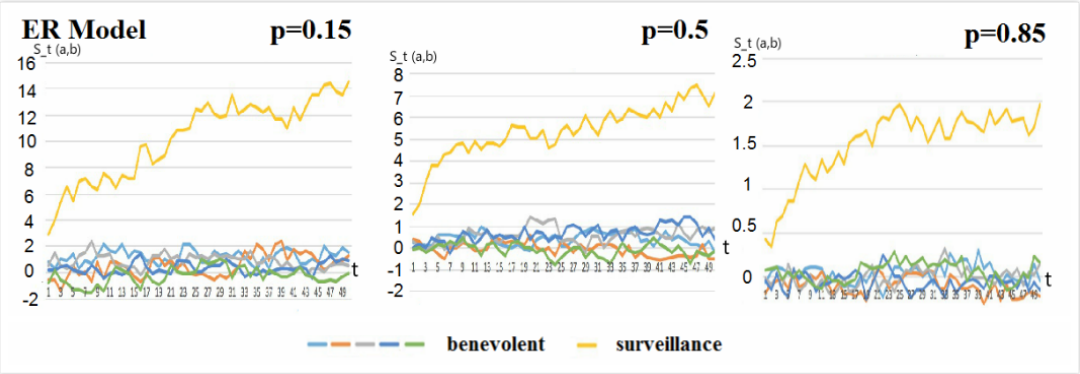

Figure 4.6: Surveillance index versus time. While all benevolent users' surveillance index fluctuates around 0, the watcher's surveillance index significantly and continuously increases and clearly shows divergence. The is one agent watching another while all others are benevolent in ER model.

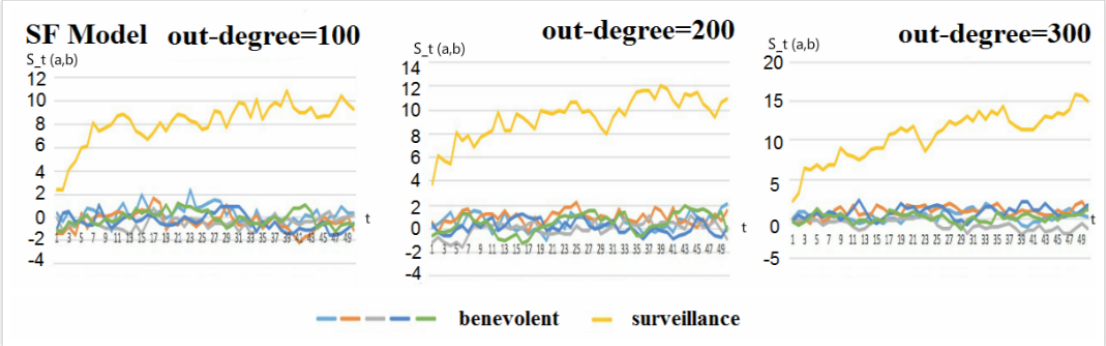

Figure 4.7: Surveillance index versus time. While all benevolent users' surveillance index fluctuates around 0, the watcher's surveillance index significantly and continuously increases and clearly shows divergence. One agent is watching another while all others are benevolent in SF model.

Experiment 4: Number of messages effect on surveillance index

The number of messages also measures the level of activities of an OSN. This experiment aims to find out the effect of the amount of messages on the surveillance index. We use

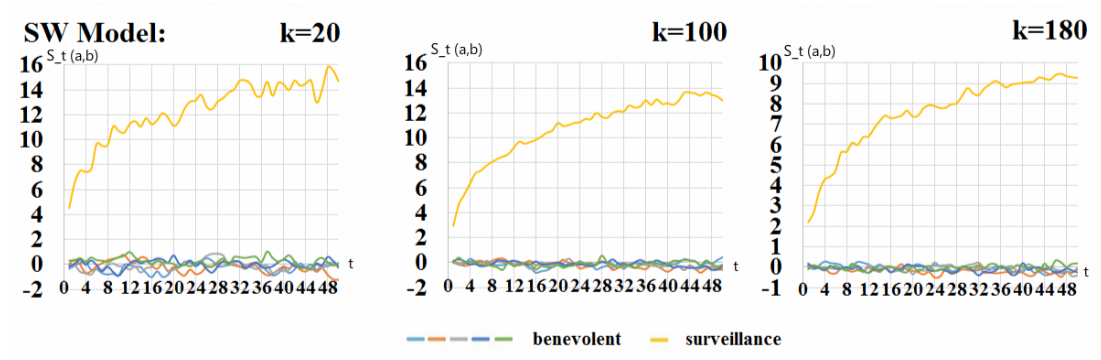

Figure 4.8: Surveillance index versus time. While all benevolent users' surveillance index fluctuates around 0, the watcher's surveillance index significantly and continuously increases and clearly shows divergence. One agent is watching another while all others are benevolent in SW model.

the SF model with an out-degree of 100 for each node. The number of total messages ranges from 100, 500 to 10000. The results in Fig. 4.9 shows that as the total number of message increases, the range of surveillance index also increases. However, the general shape of the curve does not differ much.

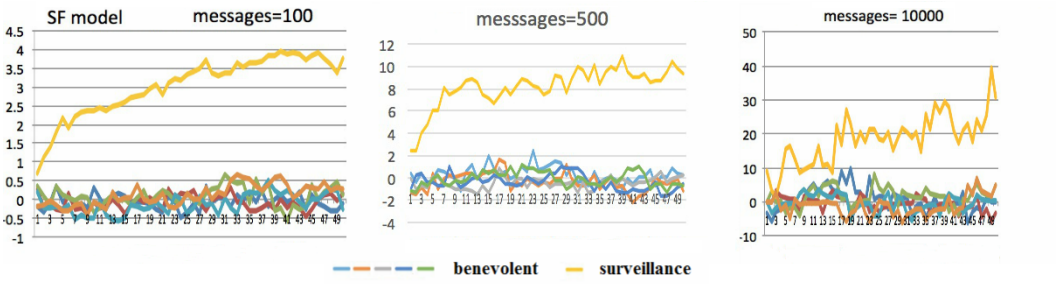

Figure 4.9: SF model with $m = 100$ and total message =100, 500, 10000. Again the watcher is clearly identifiable from the group as it rising continuously with time.

4.4.2 Surveillance Detection

When an agent a watches another agent b, the surveillance index $S_t(a, b)$ will appear to be "unusually high" as t gets large, i.e., a watcher will be the *outliers* in the distribution of surveillance indices. We, therefore, ask if it is possible to automatically detect surveillance behaviors based on this intuition.

In our experiments, we randomly select a number of agents and let them watch

randomly selected victims. Fix an agent b and consider the distribution of surveillance indices from all agents to b (given a specific time t), i.e., $\{S_t(a,b) \mid a \in V\}$. Suppose this distribution is normal, the probability density of surveillance index to b is given by the following function.

Definition 14 *The probability density of surveillance index can be defined as follows: the probability that $S_t(a,b) = x$ is*

$$f_a(x \mid \mu_b, \sigma_b^2) = \frac{1}{\sqrt{2\sigma_b^2\pi}} e^{-\frac{(x-\mu_b)^2}{2\sigma_b^2}} \tag{4.7}$$

where a, b are agents, μ_b, σ_b are the mean and standard deviation of all agents' surveillance indices to b, respectively.

Equation 4.7 shows the probability that an agent b gets certain surveillance index from another agent a. For each b, we set a threshold to represent a boundary β_b of b's tolerance to surveillance behaviors. The detection procedure for each agent b is therefore the following:

1. Compute $S_t(a,b)$ for each $a \in V$
2. If the probability that $S_t(a,b)$ has the current value (computed by (4.7)) is below β_b, declare that a is watching b

We test the performance of the detection algorithm on both random networks and real-world social networks (*Konect Network Dataset*, n.d.)[1] (with simulated instances): a Haggle network (HA), Jazz musician collaboration network (JZ) and a physician social network (MI) and a manufacturing emails network (RA). Details of the networks are shown in Table 4.2.

We use *precision* and *recall* to measure the performance of our proposed method for surveillance detection. Let D be the set of all detected watchers and S be the watchers

[1] All real-world datasets are retrieved from: http://konect.uni-koblenz.de/

Table 4.2: Details of real world networks

Networks	HA	JZ	MI	RA
Number of nodes	274	198	286	167
Number of links	28244	2742	1098	82927
Average degree	206	27	9	993

identified by the algorithm. The precision and recall of surveillance detection are, respectively,

$$\mathsf{precision} = \frac{|D \cap S|}{|D|} \text{ and } \mathsf{recall}(a) = \frac{|D \cap S|}{|S|}$$

In the following experiments, we use a fixed setup to generate the OSN unless otherwise stated. The fixed setup is: 1000 messages in total; $|\mathsf{post}(a)| = |\mathsf{read}(a)| = 1000$ for any agent a; the correlation ratio $r = 0.65$.

Experiment 5: Random networks

We first generate 10 instances of SF networks with the number of agents $n = 200$ and $m = 100$. The number of watchers is chosen from the range $\{10, 20, \cdots, 100\}$ and the number of time period is taken from the range $\{10, 20, \cdots, 50\}$. We then generate 10 instances of each type of ER, SF and SW networks with varying parameters: edge probability, agent number and rewiring probability. The boundary of any agent is set to 1.0×10^{-4}. As shown in Fig. 4.10 and Fig. 4.11, the recall in all cases are close to 1 which means that all watchers are correctly detected. Precision slightly increases as more watchers appear in the network. In most cases, precision is higher than 70%. In Fig. 4.10 (b), we see that the detection method has good performance already when $t = 10$.

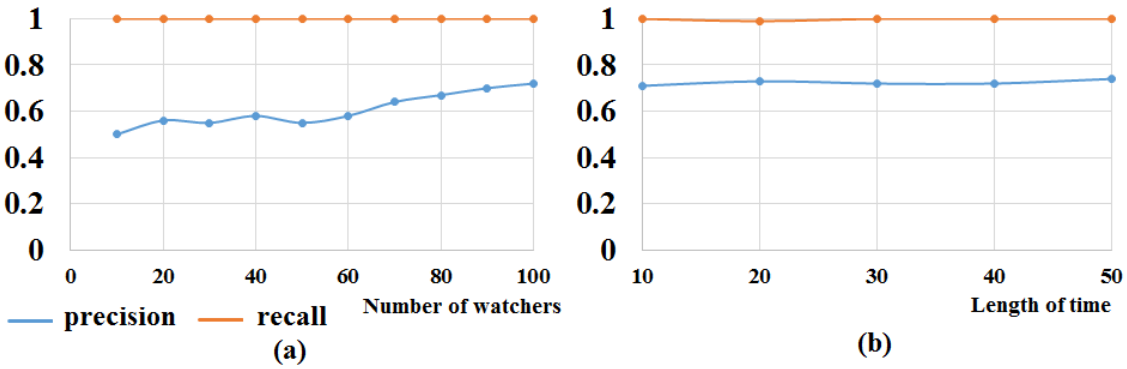

Figure 4.10: (a) shows performance affected by number of watchers in SF networks; (b) shows performance affected by the length of a time period in SF networks

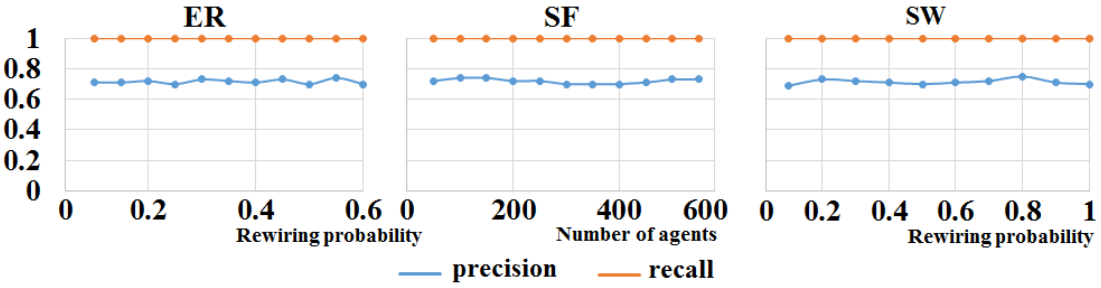

Figure 4.11: Performance of random networks

Experiment 6: Real-world networks – Surveillance index and detection

We use the fixed setup to experiment on the mentioned real networks and set the boundaries of agents 4.0×10^{-6}. They all successfully detect watchers in networks with high precision and recall as shown in Table 4.3. We further generate 10 instances for each of the real-world networks with changing number of watchers. In Fig. 4.12, we see improvements on performance as the network contains more watchers.

Table 4.3: The performance of surveillance detection on real-world networks

	MI	RA	HA	JZ
Precision	0.73	0.79	0.74	0.84
Recall	0.92	0.96	1.0	0.99

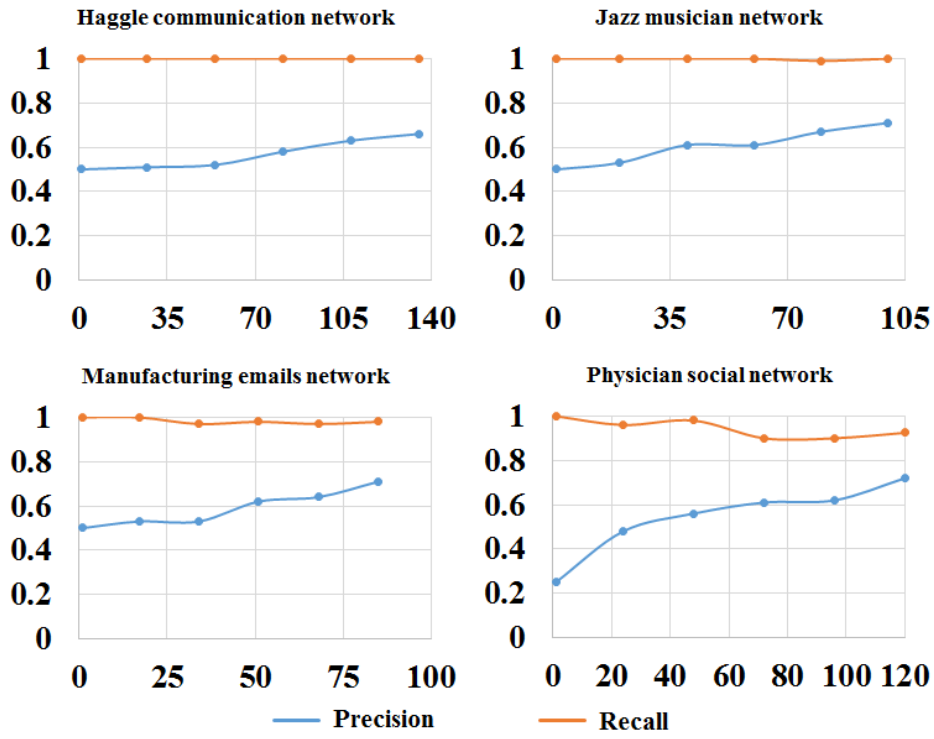

Figure 4.12: Performance of real networks

Experiment 7: Real-world networks – Simulation with surveillance breaks

In real life, watchers do not necessarily intensively gather victims' information across all time instances. Rather, a watcher a may perform surveillance towards b in certain instances while acting benevolently in other instances. Therefore, we suppose that watchers have a interval of benevolent readings, called *surveillance breaks*, between two consecutive instances of surveillance. For each mentioned real-world network, for any length of surveillance breaks in $\{1, 5, 10\}$, we compute the watcher's (and others') surveillance indices to the victim. As shown in Fig. 4.13 to Fig. 4.16, watchers exhibit much higher surveillance index in comparison to non-watchers' when the surveillance break has length 1. When the surveillance breaks have length 5, the surveillance indices display some clear "spikes", which indicate periods of intensive surveillance. When the surveillance breaks have length 10, the watchers' surveillance indices become more indistinguishable from other agents' surveillance indices.

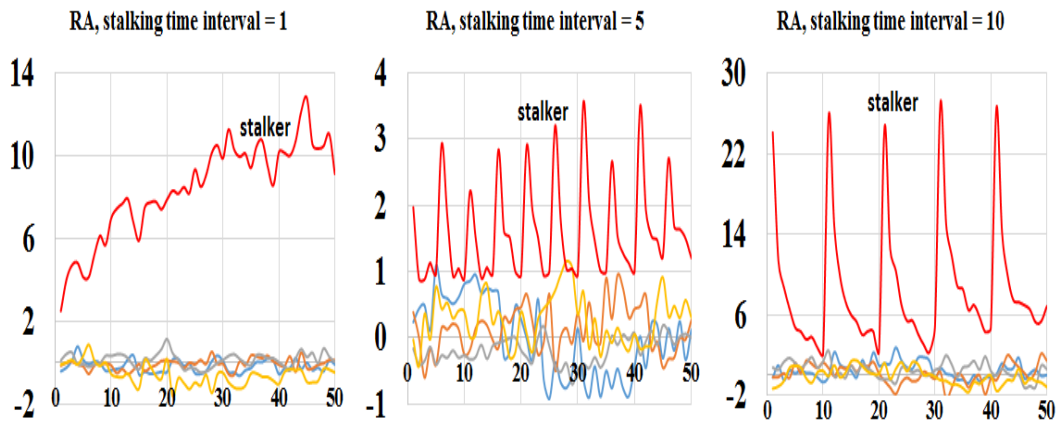

Figure 4.13: Experiment 7.1-RA network.

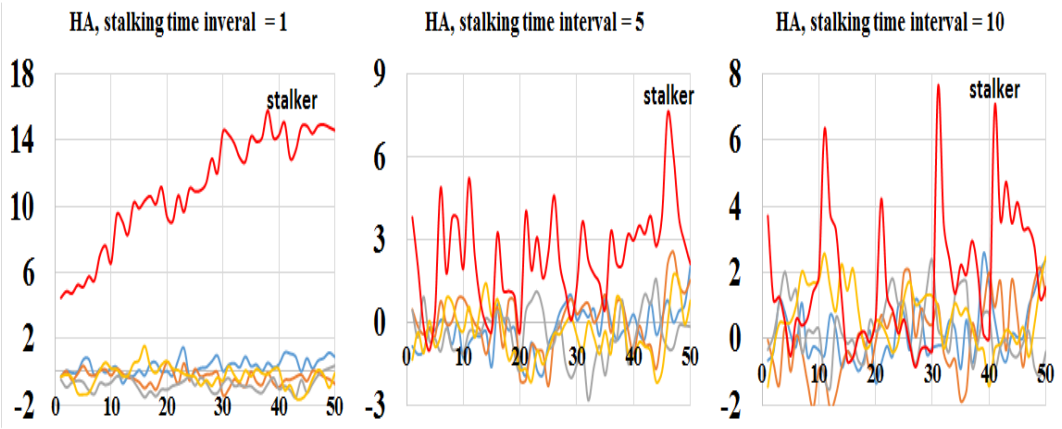

Figure 4.14: Experiment 7.2-HA network.

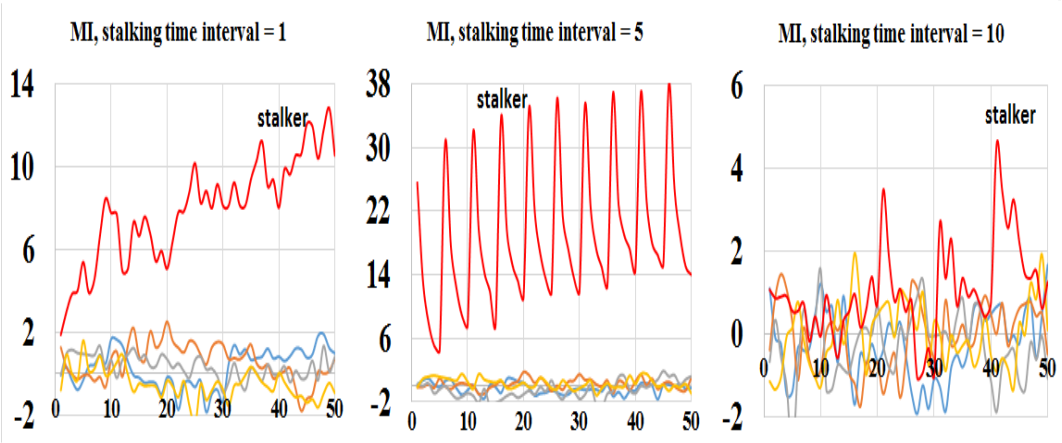

Figure 4.15: Experiment 7.3-MI network.

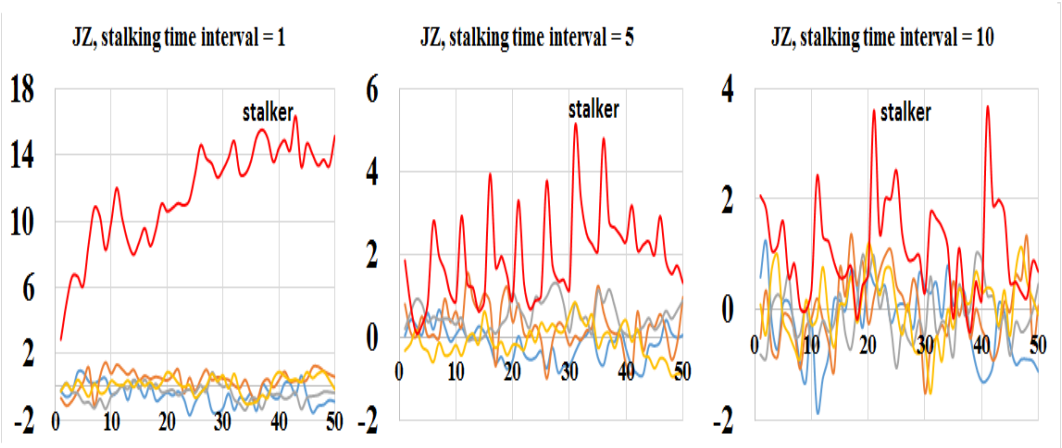

Figure 4.16: Experiment 7.4-JZ network.

4.5 Summary

In this chapter, we propose a new initiative towards a behavioral model of interpersonal surveillance from a global view. By presenting interpersonal surveillance as a computational problem, our research aims to contribute to the analysis, measurement and detection of this important online user behavior. Using a novel, network-based model of attention, we are able to define surveillance index, which is shown to differentiate benevolent behavior from surveillance behavior.

Chapter 5

Conclusion and Future Work

An Online Social Network is regarded as a Multi-agent System, which consists of a number of agents. Human users are considered as agents, which can take different types of actions. The social network is defined as a directed graph in our research. Then we modelled user interactions in OSNs like posting messages, interaction with other users and reading messages.

Attention is an important and core concept in Online Social Networks. We defined the attention as an invested interest from one person to another. In order to detect who pays abnormal attention to others, we established fuzzy-logic based computational model. At the same time, we only took users' local views into consideration to protect other users' privacy. Then we considered more serious online interpersonal surveillance situations, users may surveil others in OSNs. Online interpersonal surveillance is regarded as a user paying excessive, unreciprocated and persistent attention to others. We supposed we can collect the whole information from OSNs and each user knows each others' behaviours. We proposed a measure of interpersonal surveillance, giving a numerical scale of the interpersonal surveillance, hence defining precisely what it means for an individual to watch another.

There are also remaining a large number of interesting future works: 1) In real life

any message carries a meaning, i.e., a piece of information that relates to certain topics. By incorporating text mining techniques one may abstract and classify messages into topics and improve the current model of social attention. 2) One important property of real life social networks is that a user can hide behind multiple anonymous identities. When surveil a target, each online identity of the watcher may engage in a collective effort in collecting information, hence making the problem much harder. An effective detection tool should take into account the identity masks of the users. 3) When surveil a target, a watcher not only collects current post of the target but also retrieves historical posts. More work is needed to extend the current attention model to historical data.

References

Acquisti, A., Gross, R. & Stutzman, F. (2011). Faces of facebook: Privacy in the age of augmented reality.

Al−Khateeb, H., Alhaboby, Z. A., Barnes, J., Brown, A., Brown, R., Cobley, P., ... Shukla, M. (2015). *A practical guide to coping with cyberstalking*. Andrews UK Limited, 2015.

Backstrom, L., Bakshy, E., Kleinberg, J. M., Lento, T. M. & Rosenn, I. (2011). Center of attention: How facebook users allocate attention across friends. *ICWSM*, *11*, 23.

Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *science*, *286*(5439), 509–512.

Bernstein, M. S., Bakshy, E., Burke, M. & Karrer, B. (2013). Quantifying the invisible audience in social networks. In *Proceedings of acm sigchi conference on human factors in computing systems (chi 2013)* (pp. 21–30). ACM.

Bocij, P. & McFarlane, L. (2002). Online harassment: Towards a definition of cyberstalking. *Prison Service Journal*, *139*, 31–38.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*(suppl 3), 7280–7287.

Chaabane, A., Kaafar, M. A. & Boreli, R. (2012). Big friend is watching you: Analyzing online social networks tracking capabilities. In *Proceedings of the 2012 acm workshop on workshop on online social networks* (pp. 7–12).

Dreßing, H., Bailer, J., Anders, A., Wagner, H. & Gallas, C. (2014). Cyberstalking in a large sample of social network users: Prevalence, characteristics, and impact upon victims. *Cyberpsychology, Behavior, and Social Networking*, *17*(2), 61–67.

Erdös, P. & Rényi, A. (1959). On random graphs i. *Publ. Math. Debrecen*, *6*, 290–297.

Fire, M., Goldschmidt, R. & Elovici, Y. (2014). Online social networks: threats and solutions. *IEEE Communications Surveys & Tutorials*, *16*(4), 2019–2036.

Floyd, R. W. (1962). Algorithm 97: shortest path. *Communications of the ACM*, *5*(6), 345.

Fox, J., Warber, K. M. & Makstaller, D. C. (2013). The role of facebook in romantic relationship development: An exploration of knapp′s relational stage model. *Journal of Social and Personal Relationships*, *30*, 771–794.

Gilbert, N. & Troitzsch, K. (2005). *Simulation for the social scientist*. McGraw-Hill Education (UK).

Goodno, N. (2007). Cyberstalking, a new crime: Evaluating the effectiveness of current state and federal laws. *Missouri Law Review*, *72*.

Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, *78*(6), 1360–1380.

Hein, O., Schwind, M. & König, W. (2006). Scale-free networks. *Wirtschaftsinformatik*, *48*(4), 267–275.

Hitchcock, J. A. (2003). *Cyberstalking and law enforcement*. URL: http://www. theiacp. org.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, *20*(1), 5–10.

(n.d.). Retrieved from `http://www.ncpc.org/topics/by-audience/teens/protect-yourself/cyberbullying`

Jiang, B., Hegde, N., Masoulié, L. & Towsley, D. (2013). How to optimally allocate your budget of attention in social networks. In *Proceedings of the ieee infocom 2013* (pp. 2373–2381). IEEE.

Johnson, N. L. & Kotz, S. (1977). *Urn models and their application; an approach to modern discrete probability theory*. New York, NY (USA) Wiley.

Joinson, A. (2008). 'looking at', 'looking up' or 'keeping up with' people? motives and uses of facebook. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1027–1036).

Konect network dataset. (n.d.). Retrieved from `http://konect.uni-koblenz.de/networks/`

Kontostathis, A. (2009). Chatcoder: Toward the tracking and categorization of internet predators. In *Proc. text mining workshop 2009 held in conjunction with the ninth siam international conference on data mining (sdm 2009). sparks, nv. may 2009.*

LeFebvre, L., Blackburn, K. & Brody, N. (2015). Navigating romantic relationships on facebook: Extending the relationship dissolution model to social networking environments. *Journal of Social and Personal Relationships*, *32*, 78–98.

Lyndon, A., Bonds-Raacket, J. & Cratty, A. D. (2011). College students' facebook stalking of ex-partners. *Cyberpsychology, Behavior, and Social Networking*, *14*, 711–716.

Macal, C. M. & North, M. J. (2009). Agent-based modeling and simulation. In *Winter simulation conference* (pp. 86–98).

McFarlane, L. & Bocij, P. (2003). An exploration of predatory behaviour in cyberspace: Towards a typology of cyberstalkers. *First Monday*, *8*(9).

McQuade, S., Rogers, M. K., Gentry, S. E. & Fisk, N. W. (2012). *Cyberstalking and cyberbullying*. Chelsea House Pub (Library).

Meloy, J. R. (1999). Stalking: An old behavior, a new crime. *Psychiatric Clinics of North America*, *22*(1), 85–99.

Meyer, C. (2000). *Matrix analysis and applied linear algebra*. SIAM.

Mislove, A., Viswanath, B., Gummadi, K. P. & Druschel, P. (2010). You are who you know: inferring user profiles in online social networks. In *Proceedings of the third acm international conference on web search and data mining (wsdm'10)* (pp. 251–260).

Nahar, V., Li, X. & Pang, C. (2013). An effective approach for cyberbullying detection. *Communications in Information Science and Management Engineering*, *3*(5), 238.

Parsons-Pollard, N. & Moriarty, L. J. (2009). Cyberstalking: Utilizing what we do know. *Victims and Offenders*, *4*(4), 435–441.

Pittaro, M. L. (2007). Cyber stalking: An analysis of online harassment and intimidation. *International Journal of Cyber Criminology*, *1*(2), 180–197.

Reno, J. (1999). Cyberstalking: A new challenge for law enforcement and industry. *US Department of Justice. Retrieved June*, *25*, 2009.

Spitzberg, B. H. & Hoobler, G. (2002). Cyberstalking and the technologies of interpersonal terrorism. *New media & society*, *4*(1), 71–92.

Tagarelli, A. & Interdonato, R. (2015). Time-aware analysis and ranking of lurkers in social networks. *Social Network Analysis and Mining*, *5*(46), 1–23.

Todd, A. (2012). My story: Struggling, bullying, suicide, self harm. *YouTube. h ttps://www. youtube. com/watch*.

Tokunaga, R. (2011). Social networking site or social surveillance site? understanding the use of interpersonal electronic surveillance in romantic relationships. *Computers in Human Behavior*, *27*(2), 705-713.

Tokunaga, R. (2016). Interpersonal surveillance over social network sites: Applying a theory of negative relational maintenance and the investment model. *Journal of Social and Personal Relationships*, *33*(2), 171–190.

Wang, L. X. (1992). Fuzzy systems are universal approximators. In *Fuzzy systems, 1992., ieee international conference on* (pp. 1163–1170).

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440-442.

Willard, N. (2005). An educator's guide to cyberbullying and cyberthreats. Retrieved from `http://cyberbully.org/docs/cbcteducator.pdf`

Yin, D., Xue, Z., Hong, L., Davison, B. D., Kontostathis, A. & Edwards, L. (2009). Detection of harassment on web 2.0. *Proceedings of the Content Analysis in the WEB*, *2*, 1–7.

Zadeh, L. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetic*, *3*, 28–44.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338–353.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, *1*(1), 3–28.