AUTOMATED FRAMEWORK TO IMPROVE USERS' AWARENESS AND CATEGORIZE FRIENDS ON ONLINE SOCIAL NETWORKS

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Rahaf Barakat

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Program:
Software Engineering

July 2015

Fargo, North Dakota

North Dakota State University
Graduate School

Title

AUTOMATED FRAMEWORK TO IMPROVE USERS' AWARENESS AND CATEGORIZE FRIENDS ON ONLINE SOCIAL NETWORKS

By

Rahaf Barakat

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Kenneth Magel
Chair

Dr. Jun Kong

Dr. Sameer Abufardeh

Dr. Rhonda Magel

Approved:

7/ 30 / 2015
Date

Dr. Brian Slator
Department Chair

ABSTRACT

The popularity of online social networks has brought up new privacy threats. These threats often arise after users willingly, but unwittingly reveal their information to a wider group of people than they actually intended. Moreover, the well adapted “friends-based” privacy control has proven to be ill-equipped to prevent dynamic information disclosure, such as in user text posts. Ironically, it fails to capture the dynamic nature of this data by reducing the problem to manual privacy management which is time-consuming, tiresome and error-prone task.

This dissertation identifies an important problem with posting on social networks and proposes a unique two phase approach to the problem. First, we suggest an additional layer of security be added to social networking sites. This layer includes a framework for natural language to automatically check texts to be posted by the user and detect dangerous information disclosure so it warns the user. A set of detection rules have been developed for this purpose and tested with over 16,000 Facebook posts to confirm the detection quality. The results showed that our approach has an 85% detection rate which outperforms other existing approaches. Second, we propose utilizing trust between friends as currency to access dangerous posts. The unique feature of our approach is that the trust value is related to the absence of interaction on the given topic. To approach our goal, we defined trust metrics that can be used to determine trustworthy friends in terms of the given topic. In addition, we built a tool which calculates the metrics automatically, and then generates a list of trusted friends. Our experiments show that our approach has reasonably acceptable performance in terms of predicting friends’ interactions for the given posts. Finally, we performed some data analysis on a small set of user interaction records on Facebook to show that friends’ interaction could be triggered by certain topics.

ACKNOWLEDGMENTS

Without the support, encouragement, and friendly care of a number of people, this dissertation would not have been possible. It is my pleasure to have the opportunity to express my gratitude to many of them here. First of all, it behooves me to thank my advisor, Dr. Kenneth Magel. I am substantially indebted to him for the constant support and supervision he provided me through out by means of rejuvenating my thoughts with bright insights and new ideas and thus guided me through to the successful completion of my dissertation. His patience, flexibility, genuine caring and concern, and faith in me during the dissertation process enabled me to attend to life while also earning my Ph.D., for this I cannot thank him enough. I am forever grateful. Thank you, Dr. Magel!

Second, I would like to express my gratitude to my co-advisor Dr. Samer Abufardeh for his enthusiasm and advice on my research. I have benefited greatly from his discussion and support with sorting out technical problems. Thirdly, I am very grateful to the remaining members of my Ph.D. committee, Dr. Jun Kong and Dr. Rhonda Magel, for their interest in my research and for their valuable suggestions and feedback. My gratitude also extends to Dr. Kendall Nygard for his continued support and friendly care. He was one of the first friendly faces to greet me when I began this doctoral program, and he always has been a tremendous help. A very special thanks to Dr. Oksana Myronovych, I cannot begin to express my gratitude for the encouragement and motivation, I received from her, and for her help with professional and personal matters at each and every step of my research, which in many ways helped me achieve my cherished goal.

The most special thanks go to my beloved husband, Amro. What can I say? You have been central to my completion of this dissertation as you have given me confidence and motivation in so many ways. You took care of everything else without complaining when I needed to single-mindedly focus on completing my dissertation. You went through every excruciating step and mood change with me. You were the voice of reason when I became irrational and the practical one when I wanted to conquer the world in a day. Through your love, patience, support and unwavering belief in me, I have been able to complete this long dissertation journey. I would not have completed this journey without you by my side. I love you and am forever indebted to you for giving me life, your love, and your heart.

To my beloved children, Maisa and Sami. When I look into your eyes, I am driven to work for a bright future for you. You bring much joy and happiness into my life. I will love you forever. I will always be by your side loving and supporting you through your life journeys.

An honorable mention goes to my family and friends for their unwavering support and love especially my Mom, Mai. She instilled many admirable qualities in me and has given me a good foundation with which to meet life. She taught me about hard work and self-respect, about persistence and about how be independent. She sacrificed too much of her life so I could have a preferential education.

Last but not least, I would like to express my gratitude and obligation to Allah (God), for answering my prayers and for giving me the strength and patience to accomplish this research. I could have never completed this work without my faith in you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

LIST OF EQUATIONS

CHAPTER 1. INTRODUCTION

The power of broadcasting information to everyone, anywhere, at no cost is the basis of online social networks (OSNs). Users of OSNs start out by creating a profile; a virtual representation of each user, and then using the services provided by the OSNs to broadcast their information. For example, on Facebook, users upload photos, videos, share links, or write short responses the question “What’s on your mind?” These pieces of information are called status updates and are referred to in the literature as dynamic data. While status updates are considered a very effective mechanism to reach out to all people simultaneously, the process of writing the status update and broadcasting it has negative aspects.

First, the lack of physical contact on OSNs lowers people’s natural defense, leading them into disclosing personal information that they would never think of publicly revealing. Second, it is an overwhelming experience when people are trying to decide on whom to broadcast their status updates. For instance, some people may not feel comfortable and safe sharing the details of their life with everyone on their social network. Unfortunately, the large number of social links that users establish on these sites, along with the manual privacy controls, make managing the privacy of dynamic data a very challenging task to users. For instance, users need to configure their privacy decision with every post they make. As a result, OSNs users have become victims of unexpected consequences.

In real life, sharing information with people is a selective process. For example, some messages are dangerous and should be restricted to only trusted people, while others are not dangerous and therefore can be shared with everyone. Therefore, we believe that if a privacy control for OSNs was built to simulate this selective process, it could insure the safety of the

poster and that the messages are delivered to the right audience. To approach this goal, we present in this dissertation a way in which a privacy framework could be used to detect and warn users of possible dangerous information disclosure in their text posts and automatically assist them with deciding to whom to reveal this potentially dangerous information. The main contributions of this dissertation toward the future of OSNs are to (1) improve users' awareness about dangerous information disclosure in their text posts; (2) reduce the burden of configuring privacy settings for dynamic data on OSNs.

In this dissertation, we propose a framework to automatically check and identify messages that a OSN user has prepared for posting that may contain potentially revealing information. For example, the message might indicate that the poster is presently not at home, and will not be at home for several more hours. Such a message could alert a possible thief to a vacant home as an easy opportunity for a break in. After identifying and warning the poster and before the message is actually posted, we would recommend restricting the post to only trusted recipients and provide the poster with the opportunity to modify the post or select a subset of recipients. While dangerous information people broadcast on OSNs falls under multiple categories e.g., identity, location, and work, the focus of this dissertation is on revealing location information. These little tidbits of information have been reportedly used easily by criminals and stalkers to learn more about people's patterns so they know where to find them. Our approach has two major components, namely *Awareness Systems* and *Circles of Trust*.

The *Awareness System* addresses the gap between the users' mental model and countermeasures against revealing dangerous information. Our approach to improve users' awareness on OSNs involves a detection system that uses a combination of natural language analysis and tag-based detection rules. These rules were developed based on experimenting with

16,000 real Facebook accounts. Unlike other existing approaches, our approach can cope with a wide variety of revealing expression and detect the maximum number of dangerous pieces of information that may reveal the time and/or location of user activities and social plans. To evaluate the success of our approach, we developed a tool the employs the detection rules, and we collected a 16,000 real Facebook posts. The detection rate based on the collected posts is 85% in term of true/false, which outperforms other existing approaches.

The second component involves placement of people (friends) into sets herein called *Circles of Trust* whereby potentially unwise messages can be restricted to the appropriate Circle of Trust. Finding the right audience is a central aspect of OSN privacy control. It has been reported in the literature that the number of active links (friends who interact with the user) is significantly lower that the number of social links (people on the friends list). In response, our approach sets out to automatically infer trusted friends based on their interaction with the user. Moreover, our approach for suggesting trusted friends depends on the content of the post as the content has the potential to be either, dangerous, so that it should be seen by only trusted friends, or not dangerous, so that it can be seen by all friends.

To approach our goal, we suggest a list of trust metrics that can be used to identify trusted friends. These metrics employ explicit interaction methods that OSN users normally use to interact with each other, namely: likes, comments, replies, and tags. Then we developed a tool that can extract these metrics automaticity from interactions files to suggest trusted friends. To evaluate the success of our approach, we first performed data analysis on small interaction data sets that consisted of friends, posts and interaction between them. Our data analysis provides high level characteristics of the interaction data sets, which confirms: (1) the number of active links is significantly lower than social links; and (2) interaction methods do not have equal effect

when determining trusted friends. For example, likes and comments are more popular among social network users than tags, and tags are more popular than replies.

Using the developed tool and the collected data set we show that location topics evoke interaction from certain audiences. These findings could indicate (1) interaction on social network could be triggered by the content of the posts; and (2) limiting information disclosure to only a subset of active friends does not compromise the enjoyment of social media. Finally, evaluating our approach on the comment metric shows that our approach has a reasonably acceptable error rate in terms of predicting future friends' interactions for given posts.

Problem Background

Online social networks (OSNs) are popular Internet-based and mobile-based applications for communication, interaction and information sharing of user-generated content (Pontes et al., 2012). While, OSNs can be broken up into many categories, where many networks fall under more than one category, they could be divided based on their purposes; business-oriented, entertainment, dating, professional, friends etc. Friends OSNs e.g., Facebook, Myspace and Twitter, allow people to stay in touch with their real-life friends, make new online friends, or look for people with similar interests and ideas to connect with. The popularity of these sites has skyrocketed in recent years. As of winter 2014, Facebook, still the most popular social network in the world, had over a one billion monthly active users worldwide (Facebook, 2015). Moreover, it is the second most visited website on the Internet after Google, the popular search engine (Popular websites, 2015). According to IACP (2010), the average amount of time per day that Americans spend on Facebook is 40 minutes.

While OSNs are not free of the common threats that any distributed application on the Internet may face, they have unique characteristics that brought up new types of privacy threats.

These sites not only allow people to interact with strangers, but they also enable users to mass broadcast their personal information easily and more explicitly with a wider group of people than they would do in real life. This has resulted in connections between individuals and information revelation that would not have been made otherwise. Thus, protecting the privacy of users' information on OSNs is a very important issue.

OSNs share several core features; they ask the user to create a profile; an online representation of the user. Profiles are used for multiple purposes including forming social links by "friending" other users. Required profile information includes, but are not limited to, a user's full name, full date of birth, gender, e-mail address, phone number, etc. Each profile includes a "wall" which serves as the primary asynchronous content-sharing mechanism between friends. The information that people share about themselves on their wall e.g., videos, photos, interests, location, status update, are called dynamic data. Sharing dynamic data allows the user to stay connected with other users of the same OSN and remain socially active. However, many other parties besides faithful friends are interested in the information people share about themselves e.g., data mining companies, marketing companies, insurance companies, identity thieves, stalkers, robbers, sexual predators, etc.

Information revelation can take multiple forms on OSNs. One, is through user profile information e.g., user name, home address, relationship status, age, gender etc. Most people realize the importance of protecting the privacy of their profile information. For example, a recent study done by Pew Research Center on 802 teens, showed that 60% of teen Facebook users take an array of steps to keep their profile data private (Lenhart & Madden, 2007). In addition, this type of information is limited and relatively static and can be managed to some extent with the well-adapted friends-based privacy control. For example, when a user creates a

profile on Facebook the visibly of all profile items is set to “public” by default. However, the user can change the privacy settings for any profile item to: only me, friends, friends of friends, or custom privacy settings.

The second type of information revelation is through status updates e.g., text posts, photos, videos etc. People tend to lose sight of any potential danger they might face by revealing a great deal of their personal information through status updates. People share this type of information under the assumption that it is useless and no one is trying to deceive them. This goes hand in hand with the poor functionality of the existing friends-based privacy control in managing the privacy of dynamic data. The well-adopted friends-based privacy controls have several flaws that make it time-consuming, tiresome and error-prone for the users when they try to configure their dynamic data privacy settings.

First, it fails to capture the characteristics of the relationship among people by enforcing binary relationship among users e.g., friend/unfriend. The relationships on social network sites are more complex than the traditional social network model studied in social science (Golbeck & Hendler, 2006). Unlike the relationship on social network sites, in real life people can make a variety of assertions about their relationships with others. For example, people can explicitly state how much they trust the person to whom they are connected. Recently some social network sites such as Facebook started to allow users to state the type of the relationship they have with another user e.g., close friends, parents of, brothers, which can be used to some extend to determine the level of trust in the relationship.

Second, it requires the users to manually manage their friends lists. For example, users need to create different lists and then manually sort their friends into these lists. This could become a tiresome and error-prone task where users establish hundreds to thousands of social

links with other users. For example, 94% of teenagers have over 400 friends on Facebook (Sterling, 2013).

Finally, it does not support the frequent content updates by reducing the problem to manual privacy settings management. For instance, users should manually consider their privacy settings with every post they make. The poor functionality of the existing privacy control is well emphasized in the literature and the consequences are well-reported in the media. As a result, users of OSN are being exposed to different types of threats ranging from identity theft to sexual attack (Poovy, 2010; Kreiser, 2006; Police, 2010; CBS News, 2010; Lehrman, 2010).

How the Privacy of Status Updates is being Handled

The following section sheds light on the importance of protecting the privacy of status updates and then explains how some of the major OSNs handle the privacy of their users' status updates:

Status Updates

Status updates are the means of communication on OSNs. Generally, a status update could be defined as a short, text-based message telling noteworthy stories from users' life. For example, "just got a speeding ticket," "going on a vacation to LA," or "I got engaged." Some might use status updates to express their opinion on a recent event, to open a discussion, or to ask questions. While, others might use it as a sharing medium to post photos, links, quotes etc., information that people share through status updates are usually brief and informative and intended for specific audiences.

The number of status updates have increased dramatically in recent years, with more people adopting OSNs as their main method of communication. For example in 2012, approximately 175 million tweets were sent each day (IACP, 2010). This large amount of

information could be used to form a rich information base for strangers and criminals about the user. Moreover, OSNs as a medium for communication are not only embraced by teenagers and young adults, but also by elderly people. Therefore, protecting the privacy of status updates and limiting this type of information disclosure is very important.

How Privacy of Dynamic Data is being Handled

In this section we provide a quick overview of major OSNs and their privacy settings for dynamic data. A list of major OSNs are provided in Figure 1 along with the percentage of online adults who use these sites (Duggan, Ellison, Lampe, Lenhart, & Madden, 2015).

Facebook

Facebook users can share their status updates with five different sets of their friends.

- *Public:* everyone on the Internet
- *Friends:* only people with direct connection with the user (friendship connection)
- *Friends except Acquaintances:* where acquaintances are manually selected
- *Only Me:* only the user can see his/her status update
- *Custom:* manually created lists of friends

The default privacy setting for dynamic data on Facebook is configured to Friends. In order to protect their privacy while posting status updates, users have to manually create custom lists of friends, such as family, close friends, classmates, high school friends, acquaintances, etc. However, Facebook users tend to have an average of over 400 friends, so assigning each friend to one or multiple lists can become burdensome very quickly. Thus, the feature is not effectively used by Facebook users.

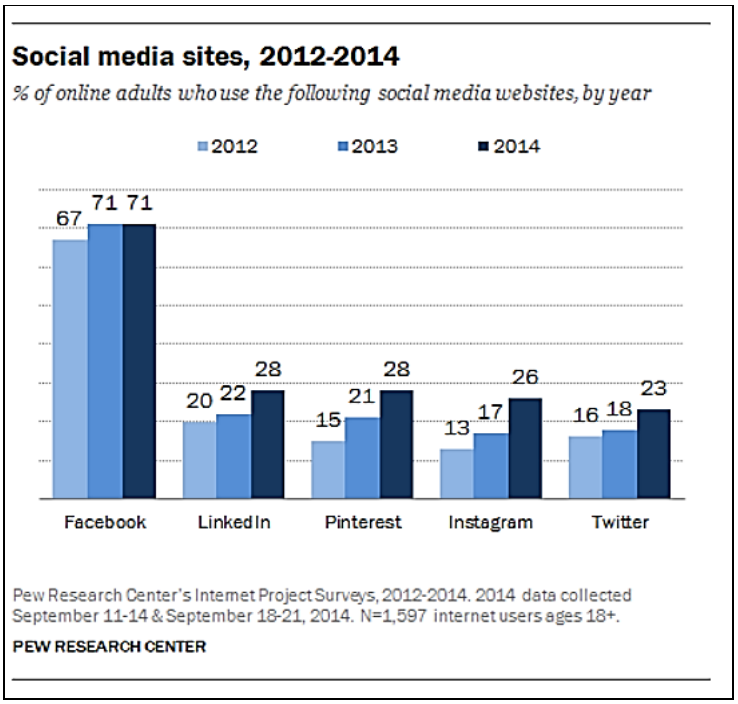

Figure 1. Major Social Sites and Their Users' Percentage. Adapted from Social media update 2014, by Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Pew Research Center, Retrieved from http://www.pewinternet.org/2015/01/09/ social-media-update-2014/, p. 2.

Google+

Started in 2011, Google+ is considered one of the most recent social networks. Although Google+ is still not considered to be one of the major online social networks, we chose to discuss its privacy control because of its similarity to Facebook. Google+ main design principle is based on friend circles. In Google+ everyone is assigned to one or multiple friend circles. While, status updates can be shared with particular circles, public, or extended circles (circles of circles or friends of friends), the main problem of this privacy setting control is that it is almost as demanding as using the Facebook friends' list feature. Despite its user-friendly graphical user interface (GUI), users on Google+ have to create circles and manually assign users to circles, which is also a tiresome and error-prone task.

Twitter

Twitter, a broadcasting or microblogging platform, was officially launched in July 2006. Tweeting on Twitter is a form of instant messaging where users daily broadcast brief messages about themselves or the world of up to 140 characters. Twitter is simple to use as a broadcaster and a receiver. People join Twitter to communicate with users who have similar interests, regardless of whether they know each other or not. This fact makes Twitter an even more dangerous environment to share users' personal information such as location of activities and social plans.

In contrast to Facebook and Google+, Twitter, uses a binary approach to manage their users' status updates. In other words, after writing a tweet the user can make this tweet, public or protected. Public tweets will be broadcasted to everyone, whether or not they have a Twitter account. Protected tweets are broadcasted to only approved followers. The default privacy setting is set to public.

LinkedIn

LinkedIn, is an online social network for work professionals, officially launched in May 2003. The LinkedIn updates are equivalents to Facebook's status updates or Twitter's tweet. For a long time, LinkedIn was considered a safe social network. For example, there are no games to jeopardize users' privacy settings. However, during the partnership between Twitter and LinkedIn, users were allowed to tweet their LinkedIn updates or steam their tweets to their LinkedIn profile. Many privacy experts expressed their concern about the users' privacy on LinkedIn because many of the users' tweets were not business appropriate.

LinkedIn has a very unique characteristic that does not exist in other social networks. It tells its users how many people have viewed their profile including their names (if their privacy settings allow). Moreover, on LinkedIn, a user's social network is made up of:

- 1st degree: this group includes the people that the user is directly connected with on LinkedIn because they accepted the user's invitation or vice versa.
- 2nd degree: this group includes people who are connected with the user's 1st degree connections. The user chose to connect with users in this group by clicking the "connect" option.
- 3rd degree: this group includes people who are connected to the user's 2nd degree connections.

Similar to Facebook privacy control, LinkedIn users can stop their updates from being broadcasted to others on LinkedIn by manually choosing one of the following privacy settings, namely: your network, your connections, or everyone; where the "connections" option includes only 1st degree members.

Pinterest

Pinterest is a popular social network that was officially launched in 2010. Pinterest has attracted people who love to share their stories in a visual way. Users of Pinterest can upload, save, sort, and manage images, known as pins. On Pinterest, the possibilities are unlimited where users can create boards for any imaginable topic. For example, creating a board to collaborate with co-workers where everyone's idea is in one place. When users see a good pin, they can like it, leave a comment, or re-pin it to their boards. While most pins are photos, users can also pin videos. To help expand their network, Pinterest users can tweet their pins or share them on Facebook.

Like any other growing social platform, Pinterest has its share of threads and security holes. e.g., scam pins, collaborator hijacking, and fake accounts. Maybe one of the very interesting privacy threads comes from Pinterest's earliest terms of services, where Pinterest ambiguously self-imposed ownership of users' content; however, these terms of service were later ended by Pinterest. Moreover, Pinterest was the only social network with no privacy options when it was first launched. Pinterest is still young and immature, so the level of its privacy control is much less than the ones users have on Facebook, Google+, LinkedIn and Twitter.

Recently, Pinterest allowed its users to create a secret board. These boards are only visible to the pins' owner and to followers the user adds manually. Otherwise, users' boards are visible to all followers. In 2014, Pinterest launched a new feature that allows its users to follow a topic feed. For example, before this feature, the users' home feeds were populated with content pinned by people they followed. After this change, users can receive updates from categories they chose to follow as well.

Instagram

Instagram is also a visual social network that allows its users to transfer mobile photos to professional snapshots by applying digital filters and a caption, and then instantly share these photos and/or videos on multiple social networks, such as Facebook and Twitter. Moreover, Instagram makes the uploading and sharing process easy, fast and efficient. While Instagram and Pinterest both fall under the category of visual social network, they are used differently by their users. For example, Instagram users look for a more personal experience than Pinterest by transferring everyday photos into professional shots and share them with friends and family. However, unlike photos on Pinterest that are linked to external websites to refer to the source of the images, content on Instagram is often authentic.

Similar to Twitter and Pinterest, Instagram adopts a binary approach to manage their users' updates. For example, the default privacy settings upon joining Instagram are set to public, which means photos shared by the user are available to everyone on Instagram to see. Users have the option to change their default to private in which case, only approved followers will be able to see the users' photos. Moreover, if the location service is enabled on Instagram and the user profile is public, then everyone can see where the photo was taken.

Motivations

This section explains the motivating factors behind the work in this dissertation:

Popularity and Privacy Threads of OSNs

Online social networks such as Facebook, MySpace, Google+ and Twitter, have grown in popularity since they were first introduced on a large scale in late 1990. These sites have become one of many assets of life, and they continue to revolutionize the way people communicate, share information, and do business. While they are great places to interact with people, stay connect, and be socially active, they brought up new privacy threats. These threats often arise after users willingly, but unwittingly share their information with a wider group of people than they actually intended.

Safety

On sites that promote a very open community, users become unaware of the potential danger they face by sharing a great deal of information through text posts that reach untrusted friends. A recent study published in 2013 by Pew Research Center about teens on social media and their privacy showed that 25% to 36% of friends that teens had created relationships with people they had never met, not including celebrities (Madden et al. 2013). Without effectively addressing these safety problems, social network users could easily become exposed to various

kinds of threats ranging from identify theft to sexual assault. Studies from the *Journal of Adolescent Health* have shown that 82% of sex predators have found minor's likes and dislikes on OSNs (Mitchell, Finkelhor, Jones, & Wolak, 2010). While, 65% were able to locate their victims using OSNs where they found home and school information, and 26% have even been able to locate victims at their exact time and location through OSN sites (Mitchell, Finkelhor, Jones, & Wolak, 2010).

> Kids are going to post photos and personal information. At least they should be aware of the risks. At least they should use the privacy tools built into the sites to keep people they don't know from accessing their information and their data. And at least moms and dads should learn about what they are doing.
>
> — Ernie Allen, President & CEO, National Center for Missing & Exploited Children (Enough Is Enough, 2014, p. 1)

Others were victims of robbery after Facebook posts in which they mentioned when they would be away from home (Police, 2010; CBS News, 2010). In 2013, Keri McMullen posted on her Facebook that she was going to watch a band with her fiancé. It took the burglars a phone call to learn about the time of the show to plan the robbery (CBS News, 2010). While in 2010, a group of thieves were able to rob 50 homes successfully based on Facebook status that people posted. These posts indicated when users are going to be away from home which gave the robbers an easy access to the vacant homes (Police, 2010).

Some were victims of identity thefts. In 2008, an attacker successfully hacked into Sarah Palin's e-mail to reset her e-mail password after using pieces of personal information that she revealed in different posts on her social network (Poovy, 2010). Others lost their jobs and employment opportunities. This type of information leak has become a concern not only to

individuals, but also to government agencies (DeCicco, 2008; Maranmins, 2009), universities (Barnes, 2006) and more (Lehrman, 2010). Thus, finding a technique to alert users about dangerous information they share and to assist them with the decision with whom they share this information is very important.

News Feed Filtering

On Facebook, dynamic data is generated under the name of the newsfeed. Facebook users have an average of 229 friends, while teens have an average of over 400 friends. Users' newsfeeds could easily get over populated with irrelevant contents with each friend generating only one status update a day. Therefore, if we look at the newsfeed section from the receiver's point of view, it becomes an overwhelming mission to go through all status updates with the possibility of overlooking relevant and interesting posts. Therefore, our approach could be also used as a filtering mechanism from the receiver's point of view to show only relevant and interesting updates.

Existing Approaches to Protecting User Privacy in OSNs

The relationship between the privacy of information and the use of social network is subtle. Research in this area has considered different aspects. However, for many years, researchers have been studying and suggesting techniques to protect user privacy by focusing on static disclosure control, yet little has been done in the research community on detecting information disclosure in user text posts and countermeasures against this type of information leak.

Problem Statement and Objectives

The motivation behind this dissertation is concentrated on two components. The first component is to improve users' awareness on online social network about dangerous information

disclosure in their text posts such that users' privacy and safety are protected. The second component is to change how text posts are broadcasted on OSNs, so that the burden of configuring privacy settings for dynamic data on OSNs is reduced.

This dissertation looks particularly at how to prevent information revelation in text posts, and how to conveniently limit the broadcasting of this information to only trusted friends. While dangerous information can take multiple forms on social networks e.g., identity, work, family, social plans and activities, we are particularly interested in information that might reveal the time and location of users' activities and social plans because this information can be easily found and traced by criminals. When detecting dangerous information, unlike existing approaches that construct costly large knowledge-based or user profile information, in our work, we define a set of tag-based detection rules that can be used along with natural language parsers to automatically detect location revealing information from text posts.

In addition, this dissertation aims to provide privacy control for text posts that is flexible and configurable depending on the captured content. This is because some content has the potential to be either dangerous, so that it should be shared with only trusted friends; or not dangerous, so that it can be shared with everyone. In this dissertation, we particularity look at the concept of trust on social networks and whether or not interaction data in social networks can be used to recommend to one user how much to trust another. We want to test our hypothesis that interaction between pairs on OSNs is triggered by the content of the posts.

While we can distinguish between two types of trust, namely, *topical trust* (how much a user trusts another user with respect to the topic of the post) and *absolute trust* (how much a user trust another user independently from the topic of the post), our notion of trust is closer to the first, which has not been presented in previous work. In our work, we define interaction-based

trust metrics to measure the subjective value of trust. We are interested in explicit interaction methods on OSNs such as comments, replies, likes and tags.

The objectives of this work are two-fold: (1) protect the privacy of the users of online social networks. For that we developed a new approach for detecting dangerous information from text posts without using a large database nor collecting users' profile information while still detecting at least 85% of text posts that reveal the time and location of users' activities and social plans; and (2) remaining flexible enough to allow meaningful interactions between the users where users have the ability to control the privacy of their information in a much greater degree of granularity with content being visible to only trusted friends on a particular topic. Therefore, our approach to configure the privacy of dynamic data depends on the content of the messages.

This dissertation concentrates on developing a privacy management model that can be used to (1) protect the safety and privacy of social network users; and (2) provide privacy management module for dynamic data that is flexible and easy to use.

Contributions

This dissertation makes the following contributions:

1. Propose a new approach to automatically check text messages that the poster has prepared for posting and detect a small set of potential dangerous information from the post.
2. Design a useful set of detection rules that can be used with a natural language parser to detect the time and location of users' activities and social plans. These rules were developed based on an exploratory experimental study on 16,000 real Facebook text posts collected from participants. These experiments involved analyzing the

combination of special keywords in status updates that could expose location revealing information.

3. Evaluate the proposed approach by building a tool that uses the detection rules. The tool was tested with over 16000 real Facebook posts yielding approximately an 85% detection rate in terms of true/false.
4. Presents software architecture for developing a privacy management system for dynamic data.
5. Develop trust metrics that can be used to quantify the subjective value of trust between a user and friends on social networking sites. Our trust metrics considered the interaction intensity between users with respect to a given topic.
6. Conduct an experimental study on a set of interaction records derived from real Facebook users that shows users' posts, friends and interaction between them. The experimental results showed that our approach has reasonably acceptable performance in term of predicting future friends' interactions.

Dissertation Outline

The rest of this dissertation is organized as follows: Chapter 2 presents related work in the areas of privacy and trust on online social networks. Chapter 3 describes the two-phase proposed approach taken in this research. The chapter starts by an overview of the proposed privacy control, then, it explains the steps of the two-phase approach followed by a section that presents the tools we developed to support this research. The chapter ends with a small example to explain the proposed approach. Chapter 4 explains the experiments done in this research to evaluate our two-phase approach and presents the results with some analysis of those experiments. Chapter 5 concludes the dissertation and proposes follow up research on this work.

CHAPTER 2. LITERATURE REVIEW

Background

This research is related to online social networks (OSNs), dynamic data on OSNs, privacy and its role on OSNs, and trust in distributed environments. This section introduces background materials of these areas. This section is organized as follows. In the next section, we begin by framing the concept of OSNs and provide a brief summary on the rise of OSNs. We then navigate the benefits and drawbacks of OSNs. Next we introduce data that OSNs operate on. In this section we distinguish between two types of data, namely, static and dynamic. We also highlight existing research on OSN analysis where we provide a brief introduction to the graph notation used to model social networks. Benefits are followed by a discussion of risks and challenges as we frame the concept of privacy and its role on OSNs, and discuss potential threats to users' privacy on OSNs. In the data on OSNs section, we frame the concept of trust, its properties, and its role on social networking sites.

Online Social Networks (OSNs)

An online social network is a way to transmit or share information with people using the Internet were everyone has the opportunity to create and share. Sharing means that the information is easily accessible, easy to comment on, always available and, last but not least, there are no fees associated with viewing the media contents. Kaplan and Haenlein (2010) defined OSNs as a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and allow the creation and exchange of user-generated content. While Collin, Richardson and Third (2011) defined social networking services as web-based services that allow individuals to: (1) create an online public or semi-public profile upon

joining the social network; (2) articulate a list of other users with whom they share a connection; and (3) view and browse their list of connections and those made by others within the same network.

Kietzmann, Hermkens, McCarthy, & Silvester (2011) created a framework that defines social media by using seven functional building blocks: identity, conversations, sharing, presence, relationships, reputation and groups as illustrated in Figure 2. The identity block, represents the extent to which users reveal their identity on OSNs. Identifying information include, name, date of birth, gender, location, profession, education or any other information that can be used to identify the user. The conversation block, represents the way that users communicate with each other on OSNs. For example, having many users facilitates conversation among users and/or groups. These conversations happen for many reasons such as, meeting like-minded people, finding partner, discussing political events etc. While the presence block, refers to the accessibility of information from which you can learn where someone is virtually and/or physically. The relationship block is pretty straightforward in that it refers to connections between users of OSNs. The study stated that the reputation block is a matter of trust and can have different meanings on different social media platforms. The sharing block implies exchanging data between two individuals. Finally, the groups block refers to what communities and sub communities exist on OSNs. As the user's social network become bigger and bigger it becomes important to sort people into different groups. This exemplifies Facebook's public, private, and custom groups.

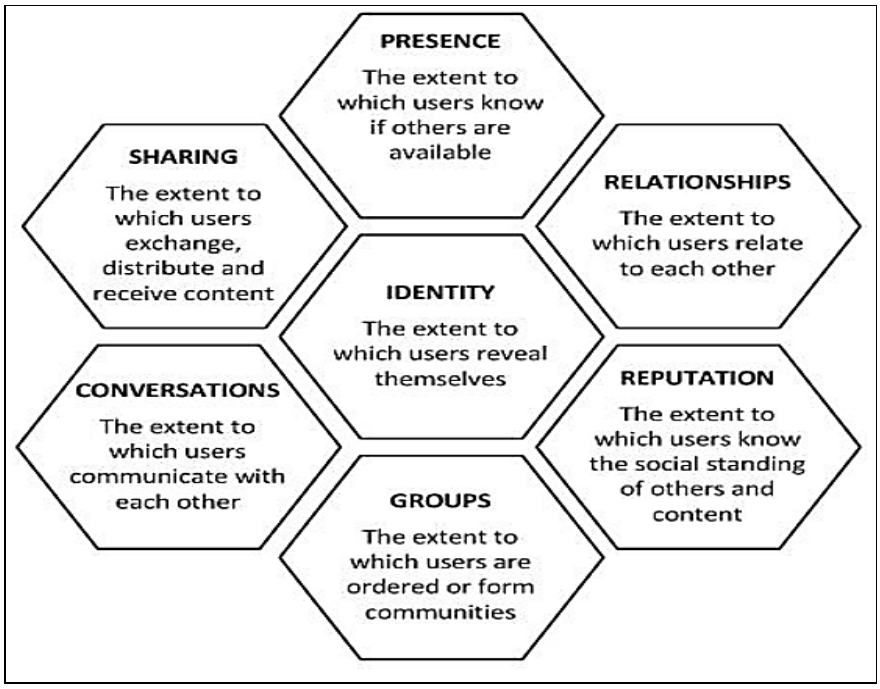

Figure 2. The Seven Building Blocks of OSNs. Adapted from Social media? Get serious! Understanding the functional building blocks of social media, by Kietzmann, J. H., Hermkens, K., McCarthy, I., & Silvester, B. (2011). *Business Horizons 54*(3): p. 243.

Schrape (2011) examines the relationship between OSNs and mass media from systems-theoretical perspective. The study concluded that OSNs are significantly different from mass media in many aspects. The user-generated content on OSNs distinguishes it from the content created by professionals such as journalists or paid content providers. These sites normally contain text, images, audio, video, and the like known as status updates. In addition, the users' relationships and activities are also different. The study done by Kaplan and Haenlein (2010) relies on the two key elements of social media, namely, *media research* (social presence, media richness) and *social process* (self-disclosure, self-presentation) to divide social media into six types. These six types are: collaborative projects, blogs, content communities, social networking sites, virtual game worlds and virtual social worlds.

Although people have been using the Internet to connect with each other since the early 1980s, it is only in the last decade that OSNs have emerged into proliferation. In 1994, Geocities, the first social networking site, was lunched to allow its users to connect with people who have similar interests and hobbies. AOL, which was probably the portent to today OSNs sites, was launched in 1997. AOL allowed its users to create a profile; a personal representation on the social network site, and share details about themselves. SixDegree was launched shortly after AOL; however, its philosophy of encouraging the forming of social relationships with random members led to its failure. It was completely shut down in 2002. Other sites followed suit a few years later, sites such as: Classmates, BlackPlanet, and Freindzy. Classmates was a hit immediately as it allowed its users to connect with old classmates, crushes, pals and even bullies. Friendster, was the first modern social networking site and it was basically a dating website. It was created in 2002 and attracted three million users in the first three months. MySpace was lunched around the same time after 10 days of coding and was soon more popular than Friendster. MySpace gave its users more freedom in customizing their online environment e.g., music and videos. Even though MySpace is no longer the top OSN in the world, it still has 90 million registered users.

While there is a great amount of literature on OSN classification, in this dissertation we adopted classifying social networking sites based on their purposes e.g., dating, religion, friends, business etc. In this research, we are mainly interested in the top friends-based social networks such as, Facebook, Google+, Twitter, etc. These sites allow their users to connect with old friends, make new friends, or connect with people who have similar interests and hobbies. Social networking sites arguably have become an extension of an individual's real life as people spend more time on these sites than any other activities on the web such as e-mail, games, and music.

However, online social networks have benefits and drawbacks that we will discuss in the following sections.

Benefits of OSNs

Participating in OSNs has many benefits. Without a doubt, keeping in touch and developing new relationships are the major benefits of OSNs. However, these sites have lots of other benefits encompassing a range of disciplines including education (Munoz & Towner, 2009), sociology, political science, culture study and health (Collin, Richardson, & Third, 2011). For example, these sites provide tremendous advantages for creating social capital (e.g., cause, beliefs, advocacy) enabling that particular society to function effectively (Valkenburg, Peter, & Schouten, 2006). However, the real benefit of OSNs comes from the value of the social relation between individuals or groups. Valkenburg, Peter, and Schouten's (2006) study found that the more frequent users participate in the social site, the more relationships they establish and the more positive feedback they receive from these relationships. Participation and positive feedback that users receive on OSNs help them in many ways to improve their well-being. In fact, social sites seem to be very helpful for people who have low life satisfaction and low self-esteem when they receive positive feedback on social sites (Ellison, Steinfield, & Lampe, 2007).

Moreover, these sites bring the benefit of staying informed about the world. Users of these sites share what interests them such as social events, news, and much more. These stories eventually make their way to people who have never read them before. Overall, this could help people to connect to a larger pool of new information and opinion. In fact, it was reported that after analyzing the Asian market, that Asians trust people's opinion on social networks more than traditional media (e.g., TV ads, magazines ads, and newspapers ads) (Asian media, 2012). Several social sites act as applications platforms. For example, users can find dozens of

applications ranging from games, quizzes, surveys, and restaurant reviews etc. The benefits of these sites go beyond individuals. For instance, Collin et al.'s (2011) study shows that using social networking sites can help organizations increase their transparency.

Risks and Challenges of OSNs

On the other hand, when participating and engaging in social networking sites, users also navigate a range of risks and challenges. Issues related to young people's well-being and their safety are one of the major concerns of parents (Rashid et al., 2009; Penna, Clark, & Mohay, 2010, Nadali, Murad, Sharef, Mustapha, & Shojaee, 2013; Cyberbullying statistics, 2014; Poeter, 2011). OSNs have created a very rich environment for bullying, in fact, 81% of youth believe cyberbullying to be much easier than bullying in person (Cyberbullying statistics, 2014). A Consumer Reposts Statistic from Internet Safety 101 shows that at least one million children have been victims of cyberbullying on Facebook alone (Cyberbullying statistics, 2014). Other concerns are often based on the fact that children and young people lack awareness of the public nature of the Internet (Acquisti, & Gross, 2006; Stutzman, 2006; Utz, & Krämer, 2009). For example, the study done by Acquisti and Gross (2006) analyzed the Facebook profiles of more than 400 students and reported that only a small percentage of young people considered changing their default privacy settings which gives access to their personal information to all users of the social network. For instance, on Facebook, the default privacy settings are set for public/everyone. While others fear that young people's use of social networking sites may affect their ability to construct supportive real-life friendships with others as in involvement in traditional institutes such as schools, sports clubs, and families (Flad, 2010).

While these concerns have dominated both the public debate and policy-making in recent years, they are not the only risks that people face on social networking sites. Other risks may

include, spams, scams, phishing, clickjacking, malicious applications, identity thefts, kidnapping, sexual abuse, robberies, suicide, stalking, losing jobs etc. Although the risks are real and the consequences could be extremely serious, we need to accept the reality that social networking sites are part of many people's lives as a part of their daily routine. Therefore, we need to seek ways to promote the positive impact of these sites and find a convenient privacy mechanism to improve users' awareness on social sites and help them behave in a more secure manner, which is the ultimate goal of this dissertation.

Data on OSNs

In essence, OSNs operate on two fundamental levels: information and the social graph as illustrated in Figure 3. Information is user-generated data, namely, profile data (static data), and status updates (dynamic data). While social graph is a reflection of the social links between people on the social networking sites e.g., friendship. In this section, we provide a brief introduction to each information type and shed light on some basic concepts of graph theory that are normally used to analyze this type of data.

Static or profile data

This is an online representation of the user. Normally, this is a self-description data provided by the user upon joining the social network. Some data may include, first and last name, date of birth, phone number, relationship status, home address, work information, etc. Static data does not change very often as it contains core aspects of a person. The privacy of this type of data could be managed to some extend with existing privacy controls the OSNs provide to their users.

Dynamic data

This data type represents the person's behavior on OSNs over time. Depending on the type of the social network, the forms of the user-generated communicated data often involve: text messages, photos, videos, tags, interests and hobbies, groups, tags, and preferences. This type of information has special characteristics that distinguish it from static data: (1) it changes rapidly, e.g., few content updates every day; (2) it is often personal; (3) it is relevant to a small subset of people; (4) it is short lived; and finally (5) managing the privacy of this type of data is hard with existing privacy controls.

Social graph

This is a network of individuals and their connections to other people in their life. These connections could be of multiple types; for example, friend, college, fan, professional, strangers, acquaintance, and many other types. These relationships are maintained by the individuals themselves. The dynamic of the social graph often changes over time, and the size of the social graph tends to grow to a point that it becomes very complex to maintain.

A collection of profiles and connections can be modeled by a graph. Generally, a network can be defined as a set of items (nodes) interconnected via edges (links). Online social networks are special implementations of the general network concept and can be defined as a group of two or more nodes interconnected via relationships. Nodes on online social networks could represent people, organizations, brands, etc. Relationships with other users are various e.g., friendship, family, professional, or acquaintances. Graph is the common way to represent a network.

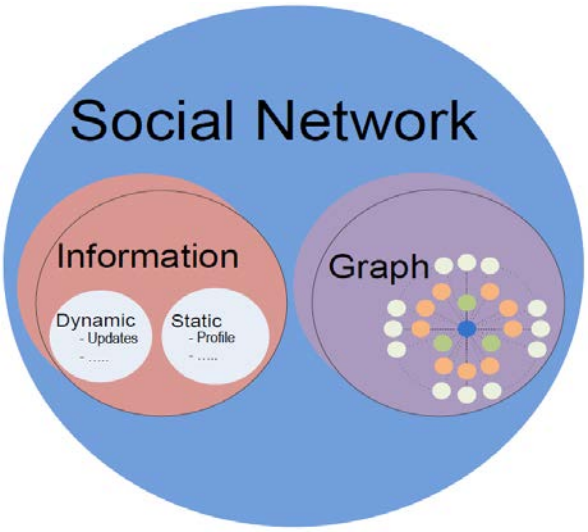

Figure 3. Data on Online Social Networks. Adapted from Privacy management for online social networks. Ph.D. Thesis, by Baatarjav, E.-A. (2013). University of North Texas, Denton, TX., p.18.

Therefore, OSNs can be modeled as a graph *G (N, E)*, that consists of set of nodes *N = {1,…, n}*, where *N* are the users interacting through the social networks, and edges *E* is the set of relationships linking those users. Figure 4 shows a simple graph showing the relationship between different nodes on a social network.

The adjacency matrix (Hsu, Lancaster, Paradesi, & Weninger, 2007) is $n \times n\ E = [E_{ij}]_{ij \in N}$ where $E_{ij \in} \{0, 1\}$ indicates the relationship status for node *i* and *j*. *1* indicates a relationship and 0 indicates no relationship. Figure 3 presents a simple directed graph showing the interaction between four users. where each node represents a person, and each link represents interaction between two nodes and therefore the communication flow from one node to another. Also, the OSN could be represented as undirected graph which captures the connections between the individuals (nodes) but does not show the information flow between individuals.

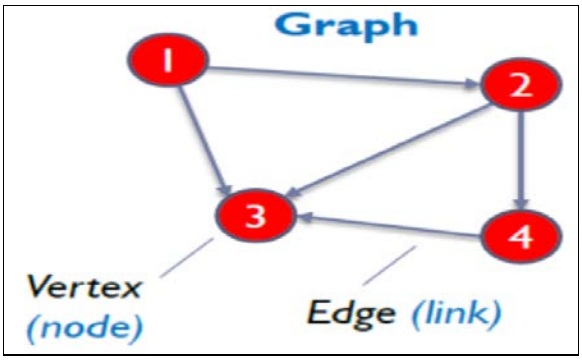

Figure 4. A Simple Graph Showing Relationship between Different Nodes on Social Network. Adapted from Trust in online social networks (Master thesis), by Volakis, N. (2011), School of Informatics, University of Edinburgh, Edinburgh, Scotland, p. 17.

Some basic concepts of graph theory (Ruohonen, 2013) that are needed to perform social network analysis and identity "key players" within a given network include:

- *Size*: number of nodes in the network.
- *Density*: number of connections that exist in the network.
- *Out-degree*: sum of connections from one node to other nodes in the network.
- *In-degree*: sum of connections to a node.
- *Degree centrality*: sum of connections from or to a node. For example, this measure is very helpful in identifying which nodes are important when spreading information. This could also influence the popularity of the node, which in turn is a strong indicator of trust and trustworthiness of the node.
- *Walk/Path*: is a finite sequence of nodes (vertex) and edges (link) that begins and ends with nodes. For example, a walk in a graph $G=(V, E)$ is v_{i0} , e_{j1} , v_{i1}, e_{j2}, . . . , e_{jk} , v_{ik}. A variation of the walk/path is the "geodesic distance" also known as shortest

distance, which is the number of edges in the shortest possible walk from one node to another. This measure is useful when determining the speed of communication.

- *Betweeness centrality*: is the sum of all of the shortest paths that access a specific node divided by the sum of all possible shortest paths that lie within the network. This number represents how frequently an actor is between other actors' geodesic paths. This measure is important in identifying which nodes are acting as brokers between sub-networks.
- *Closeness centrality*: is the distance of one node to all others in the network. This measure could determine how quickly a node is able to reach all other nodes in the network.
- *Adjacency matrix*: which is a means of representing which vertices (or nodes) of a graph are adjacent to which other vertices. It aims to show the link degree for each node and how they are connected. Edges are considered adjacent when they share a common end node (vertex). Table 1 shows the corresponding adjacent matrix for the given graph in Figure 4. In Table 1, rows and columns labeled by graph vertex, with a 1 or 0 in positon (v_i, v_j) according to whether v_i and v_j are adjacent or not.
- *Strong and Weak ties*: this is very important concept in network analysis to reason about highly connected world. Identifying a strong and weak tie could be achieved by assigning weight to each edge. This weight could be a function of many different things according to the situation; for example, it could represent the frequency of interaction between two nodes, or the frequency of interaction over a period of time, the nature of interaction, or any other attributes of the node or the tie.

Table 1.

Adjacency Matrix for the Graph in Figure 4.

Vertex	**1**	**2**	**3**	**4**
1	-	1	1	0
2	0	-	1	1
3	0	0	-	0
4	0	0	1	-

Note: Adapted from Trust in online social networks (Master thesis), by Volakis, N. 2011, School of Informatics, University of Edinburgh, Edinburgh, Scotland. Retrieved from, http://www.inf.ed.ac.uk/publications/thesis/online/IM110932.pdf, p. 18.

OSNs provide a rich source of behavioral data. Profile data, users' generated content and connections can be collected using automated collection techniques (e.g., social network crawlers) or through a database provided directly by the social network provider. Existing researches on OSNs data analysis could be categorized to make an analysis of users' characteristics such as in (Shin & Lee, 2012; Hughes, Rowe, Batey & Lee, 2012; Brake, 2012; Strater & Richter, 2007; Trammell & Keshelashvili, 2005; Vasalou, Joinson, Banziger, Goldie, & Pitt, 2008; Crawford, 2009), and content analysis to find useful information in huge amounts of data such as in (Clements, de Vries, & Reinders, 2010; Cliff, Lampe, Ellison, & Steinfield, 2007; Kim, Rawashdeh, Alghamdi, & El Saddik, 2012; Golder, Wilkinson, & Huberman, 2007; Qi, Aggarwal, Tian, Ji, & Huang, 2012; Lee, & Ma, 2012; Zhou, Lawless, & Wade, 2012; Kim, Roczniak, Levy, & Saddik, 2012; Xiong, Liu, Zhang, Zhu, & Zhang, 2012; Aiello et al., 2012; Liu, Zhu et al., 2012; Moturu & Liu, 2011; Grassi, Cambria, Hussain, &Piazza, 2011). Researchers analyze this data to explore different patterns such as friending, usage, sharing, and interaction between users; for example, (Golder et al., 2007) analyzed exchanged messages

between Facebook users during a 26 month interval to reveal the number of daily and weekly usage including seasonal variations. While the study found in (Cliff et al., 2007) used a database of Facebook users' profiles to explore the relationship between profile items and number of social links. The study reported that certain profile items which reduce transactional costs such as having an e-mail address, are associated with a larger number of social links. Other researchers have also examined the network structure of the friendship (Hsu et al., 2007; Backstrom, Huttenlocher, Kleinberg, & Lan, 2006; Herring et al., 2007; Kumar, Novak, & Tomkins, 2006; Liben-Nowell, Novak, Kumar, Raghavan, & Tomkins, 2005). Network visualization is also a very popular type of analysis that has been done on this type of large data (Heer & Boyd, 2005; Paolillo & Wright, 2005).

Privacy on OSNs

This section explains the difficulties involved in protecting users' privacy on online social networks. In general, privacy as a whole concept is defined as "the state or condition of being free from being observed or disturbed by other people" (Andress, 2014, p. 96). However, the word privacy has different meanings as it could range from physical privacy to information privacy, each with their own definition.

Privacy on OSNs falls under the information privacy category. Information privacy was defined by Kang (1998) as "an individual's claim to control the terms under which personal information–information identifiable to the individual–is acquired, disclosed or used" (p. 1205). On OSNs privacy means keeping the information in its intended scope, where the scope is defined by three factors (Beye et al., 2010):

- The number of people who can see the shared information.
- The extend of usage allowed

- And the lifetime of the information. In other words, the period of time of when this information can be available to view by others.

Definition. In this dissertation we chose to define *information privacy* on OSNs as: the right of an individual to (1) be aware of information disclosure in their status updates that might place them at risk; (2) be able to easily configure their privacy settings to determine the group of people who are allowed to know their personal information; and (3) decide what type of information each group have the right to see.

In our approach we try to address these three areas.

Research on Privacy on OSNs

The relationship between privacy of information and the use of OSNs is subtle. Researches on privacy on OSNs have considered the following different aspects:

Identify privacy breaches of OSNs

Weiss (2008) compared the traditional approach to privacy with the new privacy challenges that OSNs have brought. Unlike, the traditional Web, the data on OSNs is closely linked and available to wider groups of people. Therefore, traditional access controls that use predefined set of rules for predefined set of objects is not enough to protect users' privacy on OSNs. Gross and Acquisti's (2005) work aims to distinguish between three privacy boundaries that people normally struggle with, namely, disclosure boundary, identity boundary, and temporal boundary. The disclosure boundary expresses the control of tension between private and public, where identity boundary distinguishes between self-representation, personal life and public face. On the other hand, temporal boundary means managing past information disclosure with future expectation as people's behaviors may change over time

Online social networks raised specific types of privacy concerns due to their unique characteristics of handling personal data and allowing people to publicize their information easily and to more recipients. Lehrman (2010); Rosenblum (2007); Lipford, Besmer, and Watson (2008) reported that OSNs users do not change their privacy settings because it is a time-consuming process and they assume their profiles are private and that no one is trying to deceive them. Another OSN privacy breach could take place when shared information is moved beyond its intended scope. For example, the information is shared with a party for whom it was not intended. Other privacy breaches could take place when the shared information is abused by the party with whom it was shared, or the information was accessed after its intended lifetime (Maximilien et al., 2009).

Surveying information disclosure on OSN

Social network providers normally assist their users with privacy controls that are relatively simple to use, but too coarse. Existing privacy controls could be generally described as a list of items (e.g., profile items) that has check boxes to control their visibility levels to either private, friends only, or public. Some social network providers provide few more options (e.g., high school friends, best friends) which give the users options to include or exclude certain groups of these items.

A key study done by Gross and Acquisti (2005) analyzed 4,000 user profiles on Facebook, the online social network, and reported that 89% of profiles revealed real names, 87.8% date of birth, and 50.8% current residence. On a comparison between Facebook and Myspace, Dwyer, Hiltz & Passerini (2007) found that Facebook beats Myspace for its better capability to deal with privacy. In "*Involuntary Information Leakage in Social Network Services*," Lam, Chen and Chen (2008) analyzed 592,548 profiles on one of the largest

Taiwanese social networks. The study found that regardless of the users' effort in protecting the revelation of their personal information, first name of 72%, full name of 30%, and age of 15% of users could be easily inferred. Also, at least the name of one school attended could be inferred for 42% of users.

Devine (2008) reported that the profile information of people on OSNs, once it is gathered, can form a valuable source of information for data mining and commercial companies. Lewis, Kaufman & Christakis of Harvard University (2008) studied college students' privacy settings and found that 39% of user profiles included phone numbers and 28% included partners' name. In Lenhart and Madden's (2007) study they examined how teens understand their privacy settings and reported that teens are taking steps to protect their privacy by keeping their profile information visible only to their trusted friends. However, 32% of teens have been contacted by total strangers and 31% have friends they did meet before.

Although all social network sites publish their own privacy policies, these sites are largely devoid of security standards and practices. For example, privacy statements are often confusing, unclear, and inconsistent, as they appear to be more concerned with protecting the social network provider than protecting user privacy and security. To date, there are no specific regulations for social network sites, as they are treated as an online database of information. The technical report on the European Commission pertaining to the legal framework for the fundamental rights to protect personal data argues for change in the light of new technologies and globalization, so that the social network providers can be treated as data collectors under the data protection directive (Ngoc, Echizen, Kamiyama, & Yoshiura, 2010). Adapting those changes should make social network sites more privacy friendly without disturbing the service offered to the users.

Privacy threats based on attack model on OSNs

Beye et al. (2010) have distinguished between two types of threats. First, user-related privacy threats. These threats involve disclosing information to fellow users on OSNs e.g., other users posting information about you, not being able to hide specific information from specific people, or simply a stranger view your private information. Second, provider-related privacy threats. These threats originated from revealing information to the social network provider such as the selling of data, targeted marketing, and data retention. For example, when posting to social networking sites it is often hard to remove this information as those providers like to store this information forever.

Accorsi, Zimmermann and Muller (2012) discussed the threat of inference and the difficulties of controlling this type of threat on OSNs. The study distinguished between data the user provided explicitly to the social network provider e.g., profile data and data results from users' activity on OSNs such as incidental data; data the others disclose about the user. The Becker and Chen (2009) study emphasized the possibilities of inferring private undisclosed information from information disclosed by others on OSN.

Another threat deals with the family of attack methods for de-anonymizing social networks e.g., active and passive attack, phishing and spamming (Backstrom, Dwork, & Kleinberg, 2007; Nuha, Molok, Chang, & Ahmad, 2010; Korolova, Motwani, Nabar, & Xu, 2008). In the active attack, the attempt is to create a pattern of links among targeted accounts to make it visible in the de-anonymized graph. While in the targeted attack, the adversary can target specific users to gain access to the user's network. Table 2 aims to provide an overview of privacy concerns of user data on social sites, where many existing research areas are trying to mitigate some of the aforementioned threats (Beye et al., 2010).

Trust on Online Social Networks

The notion of trust is a multidisciplinary concept as it has been used in different disciplines such as, psychology, sociology, economy, organizational behavior, communication, and computer science to represent different concepts. For example, within studied subfields of computer science, trust means different things, such as a descriptor of security and encryption (Cheng, 1998), or a name for digital signatures in authentication methods (Ansper, Buldas, Roos, & Willemson, 2001), and game theory. Trust has different types or facets e.g., calculative, relational, emotional, cognitive, institutional, dispositional (Sherchan, Nepal, & Paris, 2013). For instance, a relational trust is built over time. Employee relationships are an example of relational trust. While cognitive trust is based on reason and relational behavior. For instance, positive referrals can increase cognitive trust.

Moreover, trust has different properties, e.g., context specific, dynamic, propagative, non-transitive, composable, subjective, asymmetric, self-reinforcing, event sensitive (Sherchan et al., 2013). For example, dynamic property of trust means that trust could decrease or increase with new experience. While propagative property of trust adapt the will know FOAF (Friend of a Friend) model (Kalemi & Martiri, 2011), where a user can derive some amount of trust in mutual friends. For example, propagated trust has been used to fight spams in e-mails and in web page ranking. The self-reinforcing property emphasizes the fact that members may not interact if the trust between them is low.

Table 2.

Privacy Concerns for Users' Data on OSNs.

Privacy Concerns	Data Types	Profiles	Connections	Messages	Multi-media	Tags	Preferences	Groups	Behavioral information	Login credentials
User related	Stranger view private info	●	●	●	●	●	●	●	●	•
	Unable to hide info from specific friend / group	●	●	●	●	●	•	•	•	•
	Other users posting information about you	•	•	●	●	●	•	•	•	•
Provider related	Data retention	●	●	●	●	●	●	●	●	•
	OSN employee browsing private info	●	●	●	●	●	●	●	●	●
	Selling of data	●	●	●	●	•	●	•	●	•
	Targeted marketing	●	•	•	•	•	●	•	●	•

Note: Concern in this table is high (●), medium (●), and low (•).Adapted from Literature overview - Privacy in online social networks, by Beye, M., Jeckmans, A. J. P., Erkin, Z., Hartel, P. H., Lagendijk, R. L., & Tang, Q. 2010, *Technical Report TR-CTIT-10-36*, Centre for Telematics and Information Technology University of Twente, Enschede, Netherlands, p. 11.

Existing trust models could be classified under three categories, namely, statistical and machine-based learning techniques (Zhao & Pan, 2014; Yu, & Singh, 2002; Josang, Hayward, & Pope, 2006), heuristics-based techniques (Huynh, Jennings, & Shadbolt, 2006; Malik, Akbar, & Bouguettaya, 2009), and finally behavior-based techniques (Adali et al., 2010). However, across all disciples, trust relationship has been characterized by two aspects, namely, risk and interdependence where the relationship is not considered a trust relationship, if the

aforementioned two conditions are not satisfied. The risk taking in the relationship involves the "trustor's belief about the likelihood of gains or losses outside of considerations that involve the relationship with the particular trustee" (Mayer, Davis, & Schoorman, 1995, p. 719). Where interdependency is the reliance on others to accomplish a certain task. For instance, in a workplace, two parties cannot achieve the desired result without relying on each other.

The following is a universally accepted concept of trust "a belief or expectation about the other (trusted) party, or as a willingness to rely on another party, coupled with a sense of vulnerability or risk if the trust is violated" (Grabner-Kräuter, 2009, p. 506). Figure 5 provides an overview of trust origin, leading to different types of trust with different properties and trust models (Sherchan et al., 2013).

In this dissertation we chose trust in social network sites as a very specific area to study the larger issues of trust, its properties, and its role in online social networks. The vision of social network sites when they were first lunched was initially limited to being used to connect friends and families; however, the public accessibility to these sites and the ability to easily share opinions, thoughts, experience and information, has led to a phenomenal growth of users in recent years.

While privacy is a very important consideration for users, trust among users on OSNs is a crucial factor when trying to balance between the open nature of these sites and privacy protection of users (Haythornthwaite, 2005; Sherchan et al., 2013). In fact, trust is one of the fundamental concepts in social activities. The ability to evaluate the strength of trust in the existing relationship helps create an environment where users can share their thoughts and ideas in an open honest way without concern about their safety and privacy. It is also important to note that although OSNs allow for connecting and communication with known friends, they also

provide anonymity. This anonymity could lead to misleading information or hindering verification mechanisms. In this case of uncertainty, trust can play an important role to reduce complexity on the decision of exchanging or sharing information on social sites.

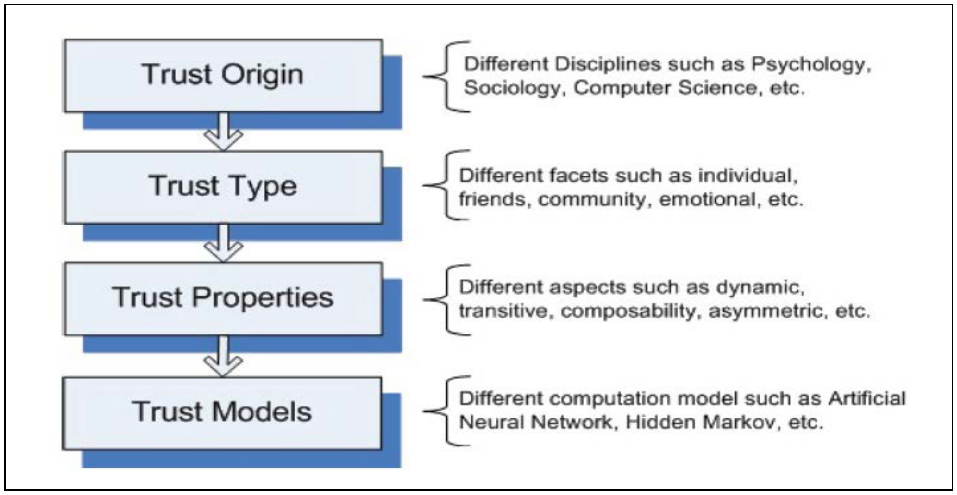

Figure 5. Trust Definitions and Measurements. Adapted from A survey of trust in social networks, by Sherchan, W., Nepal, S., & Paris, C. (2013, August). *ACM Computing Surveys (CSUR) 45*(4), p. 4.

Figure 6, provides an overview of types of trust in online social networks (Grabner-Kräuter & Bitter, 2015), where trust could be categorized to dispositional trust, macro-level trust, and micro level trust with interplay between them. The dispositional trust is a relatively stable construct as it is based on early trust-related experiences or generalized expectations about trustworthiness of other people. For example, trust in another person upon initially encountering them even if no interaction has yet taken place. While macro-level trust is where the network provider, the Internet, and the Web 2.0 technologies can all be considered as objects of trust. In macro-level trust the expectation that the device of the system will behave in a particular manner to faithfully fulfill its intended purpose. Finally, the micro-level trust is formed by participants of the social network. The micro level trust could be defined as the process in which trust

emergence depends on the beliefs of social network users about certain characteristics of other users such as integrity and honesty, which impact their trusting intention and behaviors. Micro-level trust on social sites is based on interaction between users of the sites and its strength can be determined based on the frequency of interaction. Our work in this dissertation falls under the micro-level trust on OSNs.

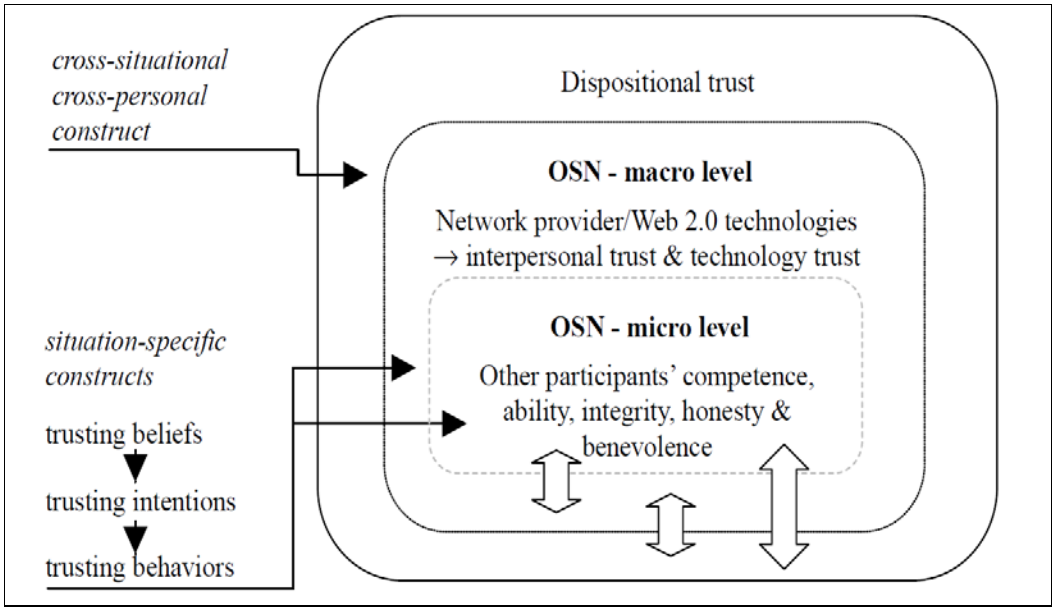

Figure 6. Types of Trust in Online Social Networks. Adapted from Trust in online social networks: A multifaceted perspective, by Grabner-Kräuter, S. & Bitter. S. (2015), *Forum for Social Economics, 44*(1), p. 52.

Resources of Trust Information on OSNs

In Sherchan et al.'s (2013) trust information collection model shows where trust information on OSNs can be collected from three main resources, namely, attitudes, experiences, and behaviors:

- Attitudes: represent the members' degree of like or dislike for something such as, a person, place, thing, event, etc. Little research effort has focused on deriving trust

from attitudes information on OSNs. Attitudes' judgments (positive or negative) are developed based on:

  - Affect: emotional response that indicates personal preference
  - Behavior: tendency of the individual
  - Cognition: beliefs about an object.

- Experiences: describe the perception of members through interactions where experience must be measured through feedback.
- Behaviors: are the most important resource for trust information on OSNs. Behavior information is derived from patterns of interactions between users. Therefore, trust in a person or community is measured by the frequency of interactions.

Within the literature, efforts heretofore have been focused on using behavior information to derive trust aside from other aforementioned aspects (Caverlee, Liu, & Webb, 2008; Adali et al., 2010; Yan, Niemi, Dong, & Yu, 2008; Nepal, Sherchan, & Bouguettaya, 2010). Figure 7, illustrates the resources of trust information on OSNs as suggested in (Sherchan et al., 2013).

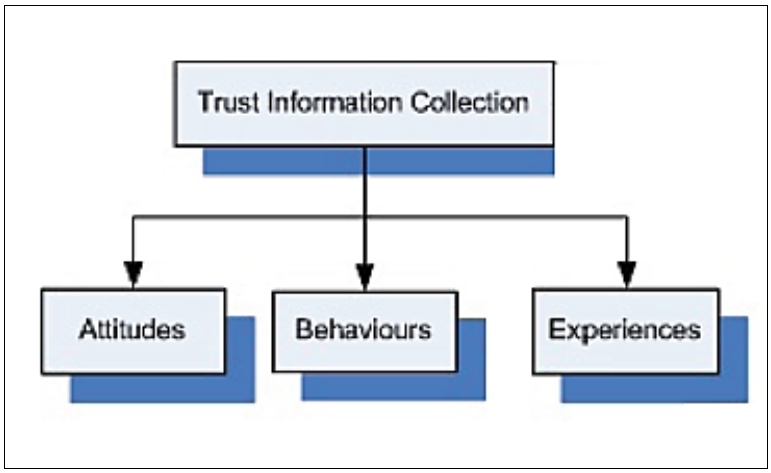

Figure 7. Resources for Trust Information on Online Social Networks. Adapted from A survey of trust in social networks, by Sherchan, W., Nepal, S., & Paris, C. (2013, August), *ACM Computing Surveys (CSUR) 45*(4), p. 2.

Related Work

After providing the necessary background information to this research, this section discusses closely related work. This dissertation concerns privacy and information disclosure on OSNs, detecting information disclosure from text posts, and trust on OSNs as a countermeasure against information disclosure. Associated literature on these areas is discussed in this section. This section is organized as follows: In the section on privacy management below, we present a review of existing techniques to mitigate privacy issues on OSNs including awareness and regulations, privacy settings and management, and countermeasures against information disclosure in text posts. Privacy techniques related to decentralization and encryptions are out of the scope of this dissertation. In the section on computational models, we present trust in the context of OSNs, and review existing approaches to trust computation and evaluation. These approaches are categorized as network structure-based trust, interaction-based trust, and hybrid-based trust.

Privacy Management on OSNs

Privacy as a concept is becoming increasingly more important in the era of OSNs where people share and communicate personal information. Related privacy techniques to protect users' safety and privacy on social networks can be roughly categorized to include awareness and regulations, privacy management for static data, and privacy management for dynamic data.

Awareness and regulations

Researches in these categories are mainly non-technical as they emphasize the importance of the users' awareness of privacy issues related to OSNs and the rules of establishing laws, policies, and regulations. Laws and regulations involve areas such as, market-regulations, self-regulations, and mandatory government rules (Chen & Shi, 2009). While

proposed techniques to raise user awareness focus on finding methods to enhance the user's awareness of privacy issues on online social networks.

Kang and Kagal (2010) proposed a mechanism to prevent the data on users' profiles from being abused. This mechanism focuses on using descriptive icons on user profiles to determine what is acceptable to do with this data and what is not. However, no further technical support was provided. Figure 8, shows a snapshot of five pictures that the user can use to restrict the abuse of their profile data. While the study in (Onwuasoanya, Skornyakov, & Post, 2008) recommends that users should group their friends on Facebook continuously to be able to manage their privacy setting effectivity for each group. For example, allow specific profile items to be seen by a specific group. The goal of this study is to improve users' awareness by providing an easy and intuitive method to manage their privacy settings; however, the proposed approach is manually configured.

In a different approach to improve users' awareness, the work in Goecks, Edwards, and Mynatt (2009) proposes the use of social collaboration to manage privacy settings e.g., when the users are not sure of how to manage their privacy settings, they can simply choose to follow the community's major privacy decisions. According to Lipford, Besmer, and Watson (2008), user awareness of his/her privacy can be enhanced by showing the user the consequences of his/her actions. The study showed that most users did not think of the consequences of their activities on OSNs until reminded of the public nature of these sites such as the public reading their profile information.

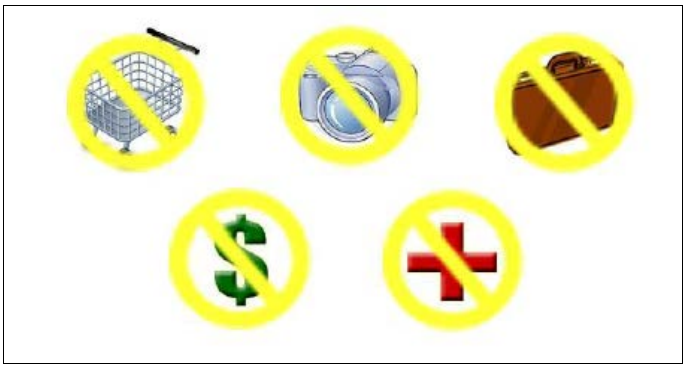

Figure 8. The Picture of Five Restrictions: No-commercial, No-deception, No-employment, No-financial, No-medical. Adapted from Understanding privacy settings in Facebook with an audience view, by Lipford, H. R., Besmer, A., & Watson, J. (2008), In *UPSEC'08: Proceedings of the 1st Conference on Usability, Psychology, and Security,* (pp. 1-8), Berkeley, CA, USA, p. 2.

Unlike our approach to improve users' awareness on OSNs, these approaches lack the power and ability to enforce the proposed changes. First of all, laws and regulations are used to solve problems after they go wrong, while in our approach, we try to provide a mechanism to prevent privacy violations by automatic detection of dangerous information. In addition, these techniques are only applicable to static profile data, e.g., username and birth data, while our approach is applicable to dynamic data. These approaches require users' manual configuration, which is a time-consuming process; however, our approach is applied automatically upon the user creating a text post.

Privacy managements for static data

Researchers of this category are mainly concerned with static disclosure control to protect the static data (profile data). Their focus is on two main things: (1) give the user more control over their privacy settings; and (2) make managing privacy settings easier for the users to configure. The central question in this category of privacy research is how to assist the user with

an appropriate tool for finer-tuning privacy control, without overburdening the user and without compromising the enjoyment of using social network sites.

Reclaim Privacy (Cloak, 2015) is a JavaScript privacy tool on Facebook that scans all users' privacy settings for profile items. The tool then suggests a level for each setting that would provide the most privacy. A rule-based tool proposed by Zakaria, Lau, Alias, and Husain (2011) helps children manage their Facebook privacy settings and serves as a monitoring mechanism for their parents. Bonneau, Anderson and Church (2009) proposed a privacy suite where a user can adapte the privacy settings of another user, e.g., privacy expert or trusted friend. Becker and Chen's (2009) PrivAware is a tool to measure privacy risks and suggests user actions to alleviate risks. The tool can query the profile information of the users' direct friends to measure the inference level for the user's profile items. The privacy risk is measured based on an algorithm that selects the most frequent attribute value among the inferred friends. To prevent profile information inference, PrivAware suggests either deleting risky friends, or moving them to a private group. Another approach involves assigning a privacy score for each profile item based on its sensitivity and then a profile item visibility score is calculated.

While these techniques are relatively cheap to implement their main focus is on the static disclosure control. In our case, our approach in this dissertation focuses on the privacy of dynamic data and on protecting users' privacy by preventing information revelation from the dynamic disclosure control, e.g., text on social network sites.

Countermeasures against information revelation in text posts on OSNs

Due to the difficulty of controlling information disclosure in personal posts on social networks, research has not emphasized countermeasures against revealing this type of information. Methods for detecting sensitive words in user's text post on OSNs were proposed in

(Hirose, Utsumi, Echizen, & Yoshiura, 2012; Watanabe & Yoshiura, 2010; Nguyen-Son et al. 2012; Kataoka, Utsumi, Hirose, & Yoshiura, 2010). In these closely related works, sensitive information was defined as any information about a user that could be used to identify the user. As a countermeasure against information discourse in text posts, automatic text anonymization was used to limit information disclosure.

A private information detector (PID), uses a combination of users' pre-collected profile data and data available on the Internet to retrieve sensitive information in users' text posts. The PID implementation is illustrated in Figure 9. Its algorithm checks the text that the user wants to post on OSNs and detects potential disclosure of sensitive information so it can warn the user or suggest modifying the text automatically. Detecting private information is done by comparing the text against pre-collected profile information. The algorithm works on generating all possible keyword combinations from the users' text posts. The algorithm then uses Google, the popular search engine, once to search for each keyword combination. The PID, then retrieves the first 24 links for each keyword combination and then applies co-occurrence analysis to determine the frequency of two words appearing in the same text. The algorithm then calculates the total risk score of the text post to determine whether the post contains sensitive information about the user. Technically speaking, the algorithm works as following:

- If T is the input post, $A = \{a_0, a_1, a_2...\}$ is set of attributes from user's profile, $P = \Theta(A,t) = \{P_i\}$ where P is a set of sensitive phrases. Θ is the function that detects sensitive phrases on the basis of the relationship between user's profile item, which is previously collected, and text in the user's post.
- $S = \sum_{i=1}^{m} \binom{n}{i}$ is all possible keywords combination in the sensitive phrases.

- To detected private information revelation, the algorithm will use the search engine S time, once for each combination.
- For each combination, the algorithm will check the first 24 retrieved links for the number of times the sensitive phrase appears; co-occurrence analysis.
- Finally, the algorithm calculates the total risk score (TRS), where $TRS = 1/S \sum_1^S (\text{vi})$ and v_i is number of appearance of each sensitive phrase on the search engine.

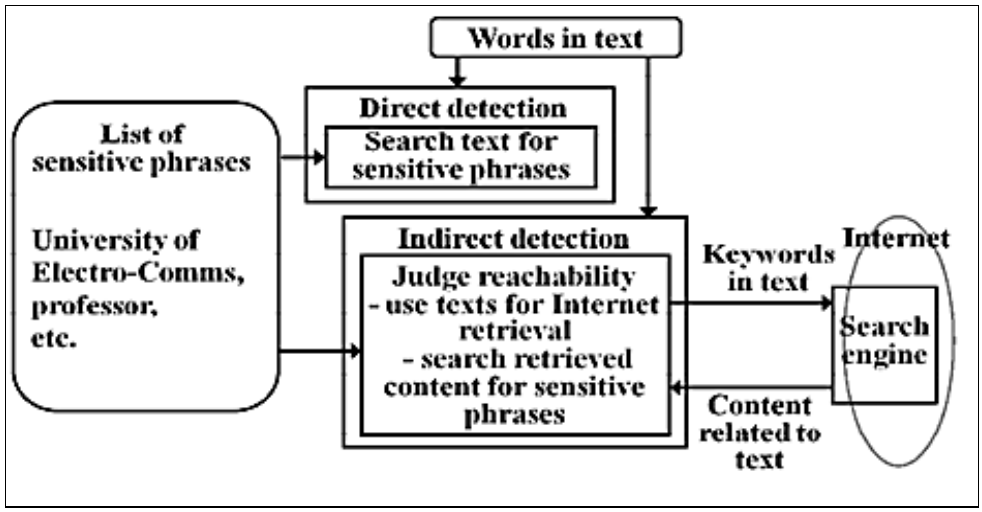

Figure 9. Private Information Detector (PID). Adapted from A private information detector for controlling circulation of private information through social networks, by Hirose, M., Utsumi, A., Echizen, I., & Yoshiura, H. (2012), Seventh International Conference on Availability, Reliability and Security. Tokyo, Japan, p. 475.

As a second step against sensitive text information revelation, transformation algorithms are applied to the text, namely, omitting words algorithms and amending sentences. While this approach is useful to some extent, it is not helpful in detecting dangerous posts as simple as "going out for run," or "heading to Hawaii for spring break!" These posts do not contain

sensitive information that could be used directly to identify the user. Therefore, according to the PID, the TRS would be 0 and the system would not detect these risky posts.

Furthermore, PID does not advise users on whom they may safely allow to see their private information. Instead, it advises the user on either omitting sensitive words or anonymizing the text. Anonymizing the text is approached by generalizing the sensitive information in the text. For example, if the input text is "I am studying at NDSU" and the user affiliation can be revealed from NDSU, then the algorithm will suggest anonymizing the text by generalizing the sensitive word to a more abstract word, e.g., use the word "University" instead of "NDSU." However, anonymizing the text could result in unnatural posts or grammatically incomplete sentences, which sacrifices the enjoyment of communication on OSNs. Moreover, the PID detection algorithm uses user profile information to judge sensitive information revelation in text. This approach could lead to inaccurate results in case the user chose to provide false information. Moreover, this could lead to security breaches if users' profile information database was hacked. Finally, PID has a high number of false alarms 56%, which could be annoying to the users.

As an extension to the work in (Nguyen-Son et al. 2012; Kataoka et al., 2010), an algorithm is proposed to identify the disclosure, the friend who disclosed information about the user on the OSNs. The algorithm has two steps; first, it applies text generalization technique to anonymize each combination of sensitive phrases detected in the user's post. Second, it applies synonym techniques on the generalized text to write different sentences for each friend so that the disclosure (the person who disclosed the information) could be detected. In a previous study (Kataoka et al., 2010) synonyms were used as fingerprints to identify the disclosure, e.g., the person who disclosed the user information. Figure 10 illustrates sensitive phrase detection in

input text on social network sites, where the user's profile information is used to judge sensitivity. Figure 11 illustrates the generalization schema for each phrase in the input text that can be used to identify the user. While Figure 12 illustrates creating synonym phrases for generalization so each friend can receive a different text.

The illustration below is based on the given post: "Many computer applications interest me very much. Therefore, I am studying computer science at Stanford. Because I am from Tokyo."

User's Profile		Input phrases
First name	a_0 ={Adam}	Many
Last name	a_1 ={Ebert}	computer
Favorite	a_2 ={Football}	...
University	a_3 ={**A university founded in 1891**}	**Stanford**
Nickname	...	...
Prefecture	a_5 ={**Shinjuku, Tokyo**}	**Tokyo**
...	...	...

Figure 10. Sensitive Phrase Detection for the Given Post. Adapted from Automatic anonymization of natural languages texts posted on social networking services and automatic detection of disclosure, Nguyen-Son, H.-Q., Nguyen, Q.-B., Tran, M.-T. Nguyen, D.-T., Yoshiura, H., & Echizen, I. (2012, August 20-24). *Seventh International Conference on Availability, Reliability and Security*. (pp. 358-364). Prague, Czeck, Republic, p. 734.

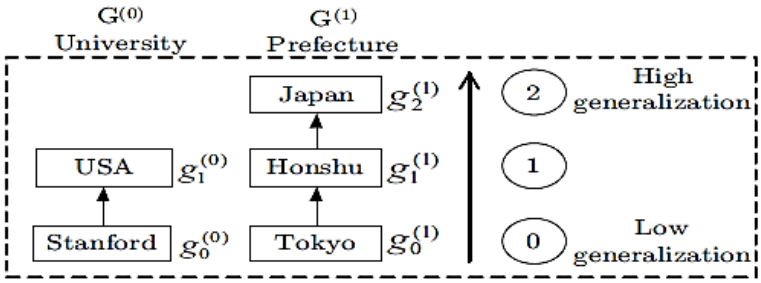

Figure 11. Generalization Schemas for the Two Identifiers in the Given Post. Adapted from Automatic anonymization of natural languages texts posted on social networking services and automatic detection of disclosure, Nguyen-Son, H.-Q., Nguyen, Q.-B., Tran, M.-T. Nguyen, D.-T., Yoshiura, H., & Echizen, I. (2012, August 20-24). *Seventh International Conference on Availability, Reliability and Security*. (pp. 358-364). Prague, Czeck, Republic, p. 736.

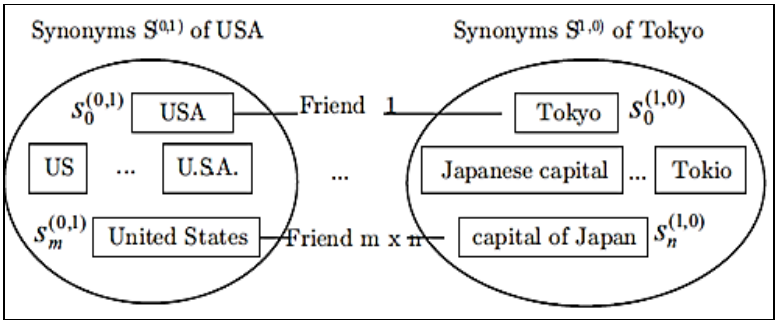

Figure 12. Synonyms for Generalization {USA, Tokyo}. Adapted from Automatic anonymization of natural languages texts posted on social networking services and automatic detection of disclosure, Nguyen-Son, H.-Q., Nguyen, Q.-B., Tran, M.-T. Nguyen, D.-T., Yoshiura, H., & Echizen, I. (2012, August 20-24). *Seventh International Conference on Availability, Reliability and Security*. (pp. 358-364). Prague, Czeck, Republic, p. 737.

Hart, Castille, Johnson, and Stent (2009) proposed a tag-based access control method as a plug-in to WordPress, the online blogging tool (Wordpress, 2015). The proposed approach allows bloggers to use tags to describe their privacy policies. The system applies machine-

learning techniques to apply policies to newly created blogs. While this is an evolutionary approach, it does not suit the dynamic nature of user generated content on OSNs such as, Facebook or Twitter. OSNs are open worlds where people post numerous amounts of information every day. Therefore, it is virtually impossible for users to know what topics they will talk about next and in which context, and thus anticipate all tags for all potentially revealing information. Further, in their proposed approach, the blogger has to pick an average of nine tags for each blog in order for the system to provide accurate results. People on social sites tend to post many posts a day on a variety of topics, so assigning tags for each post could very easily become a cumbersome task. Moreover, with an average number of over 400 friends on Facebook, deciding on how to assign friends to tags would be a time-consuming and error-prone task.

Trust Computation Models on OSNs

Trust on OSNs is relatively a new and evolving research topic. Existing trust models on OSNs can be categorized as graph-based, interaction-based, and hybrid. The main focus of this dissertation is on interaction-based models to evaluate trust. Therefore, this section will present a review of literature of major work only under the graph-based category and the interaction-based with special focus on related literature on trust models and metrics using interaction to compute trust.

Graph-based trust models

The main idea of the graph-based trust is to model the social network as a graph then apply data analysis measurements to infer the trust relationships from the graph. Basic concepts about modeling OSNs as graph and basic concepts of graph theory was introduced earlier in this chapter as a background information to this section.

In graph-based trust models, people are represented as nodes and the amount of trust they have for each other is represented as edges. Representing the social network as a graph allows analyzing the network using tools of mathematical graph theory to estimate trust in a path, trust in a graph, and trust in a target node. Usually, a trust network is created for each member. There are a number of ways to think about how a network structure (graph) matters when determining trust between two nodes. The literature contains examples of each.

First, the structure of the network can be an indicator for influence and popularity of a node which are two important factors when assessing the level of trust in a node. For instance, the node degree is an important factor that impacts the trust and trustworthiness of a node. When the node's degree increases (number of connections) the level of trust a member can have in another member on the social network also increases. The observation from the study in Busken's work (1998) concludes that a node's level of trust increases when (1) the node has a higher out-degree; (2) the node directs its tie toward nodes with higher out-degree; (3) the node's location is in the central location of the network, centralization; and finally, (4) the average level of trust of all nodes decreases in the central location of the network.

Major approaches for trust evaluation that leverage the structure of the network and requires feedback from users on interactions to estimate trust include (Golbeck, Parsia, & Hendler, 2003; Golbeck, 2005; Zhang, Chen, & Wu, 2006; Kim, 2008; Carminati, Ferrari, & Perego, 2009). In an attempt to create a trust network on the sematic web, Golbeck et al., (2003) proposed a method to create a trust network on semantic web. The method proposes extending FOAF ontology, which is a simple ontology for representing information about people and who they know. Their proposal suggests adding a model for trust rating on a scale of 1-9, which allows users to annotate their relationship with information about how much they trust their

friends. The trust rating value, can be used to infer the trust value between two nodes that are not directly connected. For example, by selecting two individuals the source (*Node i*) and the sink (*Node s*), we can recommend to the source how much to trust the sink as illustrated in Figure 13.

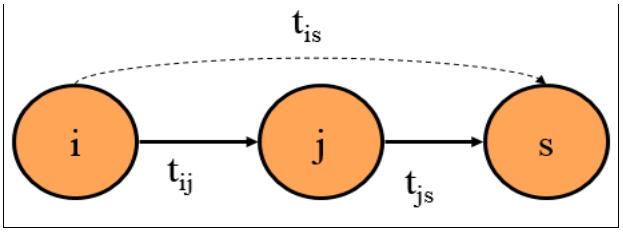

Figure 13. Inferring Trust between Two Not Directly Connected Nodes *I* and *S*.

The proposed algorithm works as follows:

- If the source node does not know the sink node, then the source node asks all its friends' nodes how much to trust the sink. Then the trust value is computed using the weighted edges of the graph according to Formula 1.

$$t_{is} = \frac{\sum_{j=0}^{n} \begin{cases} (t_{js} * t_{ij}) & \text{if } t_{ij} >= t_{js} \\ (t_{ij}^2) & \text{if } t_{ij} < t_{js} \end{cases}}{\sum_{j=0}^{n} t_{ij}} \quad (1)$$

- Neighbors' nodes repeat the same process if they do not have direct rating for the sink.

The experiment results show that trust is computed accurately within about 10% using the proposed algorithm. As an extension to their work, an algorithm was proposed by Golbeck (2005). TidalTrust is an algorithm for inferring the trust relationship between two nodes on the social network using FOAF vocabulary. Their approach suggests that individuals with a higher

trust rating are more likely to agree on the trustworthiness of a third party. The Film Trust project used the proposed algorithm as an application of trust.

For example,

- Suppose Bob (B) has three trusted friends (A1, A2, A3) who rated the movie (M).
- If Bob does not know the movie (M), the TidalTrust algorithm is used to calculate the rating based on the trust information in the network as illustrated in Figure 14.

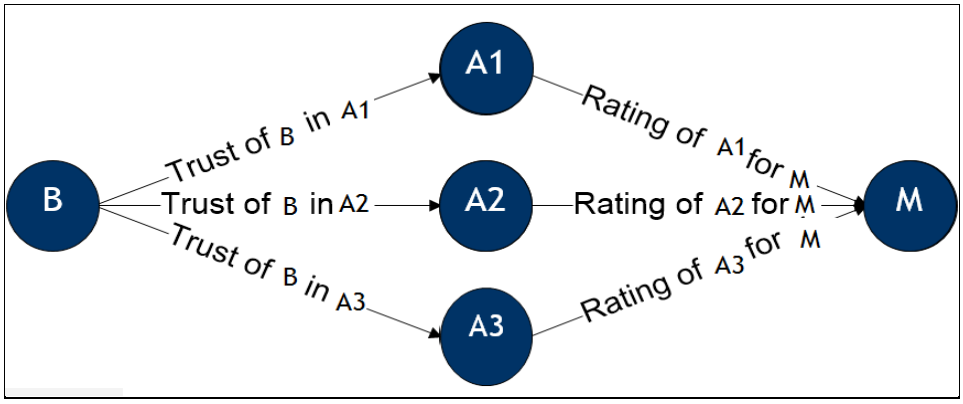

Figure 14. Trust Network Visualization.

- The Algorithm for calculating the recommended rating is recursive and it uses the following formula, Formula 2.

$$r_{sm} = \frac{\sum_{i \in S} t_{si} \cdot r_{im}}{\sum_{i \in S} t_{si}} \quad (2)$$

Where:

- s: is a node in a set of nodes S
- r_{sm}: is a rating inferred by s for the movie m
- i: describes intermediate nodes

- t_{si}: describes the trust of s in i
- r_{im}: is the rating of the movie m assigned by i

Results show that most accurate results come from the highest trusted individuals. Moreover, the accuracy of the algorithm decreases with the path length. Some drawbacks of the proposed algorithms are: (1) it does not deal with uncertainty; (2) the calculated rating can be heavily influenced when the number of poorly trusted friends increases. This could result in misleading results because some recommenders might have malicious intentions; and finally (3) It does not update trust values in recommenders.

Existing graph-based approaches and techniques that leverage network structures to calculate trust have a few limitations. First, they capture only a few aspects of trust notion, namely, how individuals are related to each other, and how trust flows through the network. Therefore, they fail to capture other important computational properties of trust, e.g., temporal and context dependency. In other words, these approaches are static, whereas trust changes over time. For instance, some friends might not be trusted at the time when the friendship was established, but that trust could change over time. Second, actual interactions between individuals are a very important indicator of trust on OSNs. For example, the volume, the frequency of interaction, and even the nature of interaction negative/positive have impact. Third, the effectiveness of these approaches depends on the connectivity of the trust network and can perform poorly when the connectivity is sparse, which is the case of online social networks.

From another point of view, the volume of interactions between two members and the frequency of interaction convey valuable information in terms of trust on social networks. Researchers have repeatedly cast doubt on the practices of inferring meaningful relationships from analyzing social network connections alone and. in addition, they suggested that in order

for social application to reflect real users' activities rather than social linkage alone, social application should be designed with interactions graphs in mind.

Interaction-based trust models

The main idea of interaction-based trust models is to compute trust value using the interactions of a user with other users within the social network. In an attempt to build a web of trust based on interactions between two users in the context of online product review, Liu, Lim et al. (2008) proposed a semi-supervised approach that automatically predicts trust from interactions between a pair of users. Their approach is based on the observation that a member trusts another member because of the latter's good reputation, or because of good interactions between the pair. For that purpose, they developed two taxonomies of trust factors, namely, taxonomy of user factors and taxonomy of interaction factor, to predict trust among users in online communities. The user factors taxonomy refers to features associated with the shared data of a given user, e.g., reviews, posted comments, and rating, with metrics such as the number of first reviews, the number of reviews posts, and review frequency. The pair factors taxonomy includes different types of interactions that could take place between users, e.g., between writers and raters, writers and writers, and raters and raters.

Their empirical studies showed that predicting trust among users is effective using the trained-classifier. While this work is evolutionary, there is no evidence or supported study that their approach and their taxonomies are applicable to OSNs such as Facebook. Their approach was evaluated using Epinions; a large product review and rating community that support different types of interactions.

In order to build trust communities in OSNs, a number of trust models has been proposed in the literature (Hamdi, GancÌ§arski, Bouzeghoub, & BenYahia, 2012; Caverlee et al., 2008;

Caverlee, Liu, & Webb, 2010; Nepal & Sherchan, 2011; Ali, Villegas, & Maheswaran, 2007; Moalla, Hamdi, & Defude, 2010). In closely related work, STrust, a reputation-based model, was proposed in (Nepal & Sherchan, 2011). STrust is an interaction-based trust model for OSNs where trust level consists of two types, namely, the popularity trust and the engagement trust. These two types of trust distinguish between two types of community members: trusted and trusting members. The combination of the popularity and the engagement trust is used to determine the social trust of a community.

The popularity trust is based on the trustworthiness of a member in a community; for instance, members who continually receive positive comments on their posts. Whereas engagement trust is based on how much a member trusts other members in the community; for example, members who frequently give others positive comments on their posts. Finally, the social trust is the sum of popularity trust and engagement trust. Figure 18, presents a STrust interaction model where individuals are represented as nodes and interaction methods are represented as arrows.

All interactions come from the user performing different types of activities, e.g., reading, commenting, rating, viewing etc. In addition, STrust distinguishes between active interactions and passive interactions. Active interactions are visible to the community, e.g., commenting and rating, while passive interactions are not visible to the community, e.g., reading and logging. In Figure 15, each arrow provides information about popularity trust and engagement trust on the other side.

Let’s consider user F as an example. The user has two outgoing arrows that represent F’s engagement trust, and two incoming ones that represent F’s popularity trust. In another example, the arrows between users C and E provide information about the engagement trust between the

two users. Solid lines represent active interactions, and dotted lines represent passive interactions. Whereas, the interactions between two nodes can be either positive (+) or negative (-) where passive interactions (e.g., reading) are always considered positive.

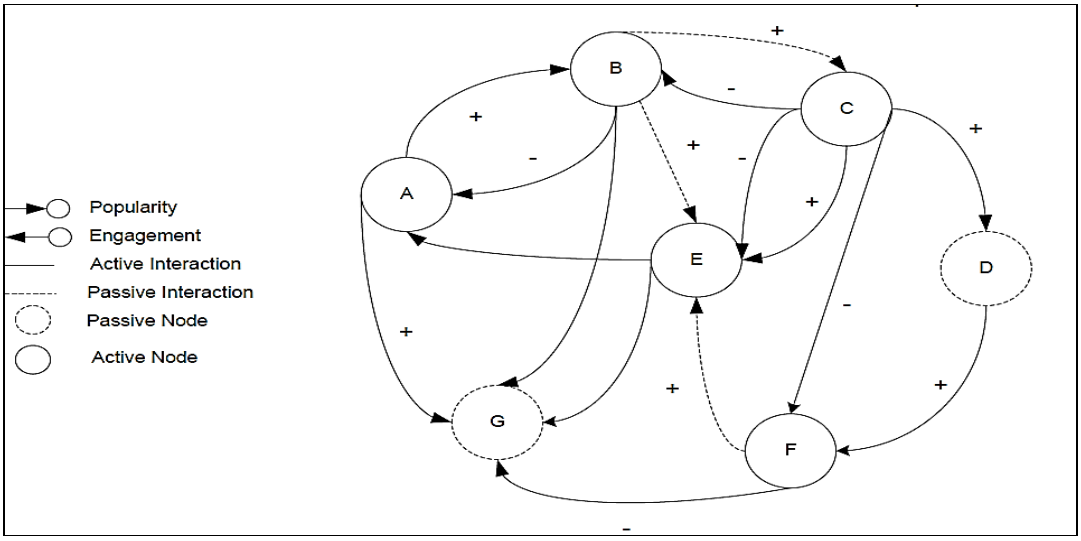

Figure 15. STrust Interaction Model. Adapted from STrust: A trust model for social networks, by Nepal, S. & Sherchan, W. (2011). In *Proceedings of the 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications* (pp. 841-846). Changsha, China, p. 844.

The benefits of separating these two types of trust is to recommend two different types of people, namely, the leaders and the mentors. The popularity trust can be used to recommend the leader, while the engagement trust can be used to recommend the mentor. The leader is a trustworthy member recognizable from metrics such as the number of positive feedback/opinions of their posts, and/or how many members follow them. The mentor is a member who would like to actively engage in the community. The mentor is recognized from metrics such as, how many members they follow, and/or how many posts they comment on.

STrust aims to encourage positive interactions to increase the social capital and, as a result, to increase the social trust. Technically speaking, the popularity trust (PopTrust) for a member $u_i \in U$ in a particular context (x) is calculated using Formula 3:

$$PopTrust(u_i^x) = \frac{\sum_{j=1}^{|M|-1} \frac{|PT_{ij}^{x+}|+1}{|PT_{ij}^{x+}|+|PT_{ij}^{x-}|+2}}{|M|-1} \quad (3)$$

Where:

- $|PT_{ij}^{+}|$ is the total number of positive interaction for a member with other members
- $|PT_{ij}^{-}|$ is the total number of negative interaction for a member with other members
- $|\mathrm{X}|$ is the context of trust e.g., live chat, comment etc.
- $|\mathrm{A}|$ represents the number of activities in each context
- $|M|$ is the total number of members in the community

Similarly, the engagement trust is calculated using the following Formula 4:

$$EngTrust(u_i^x) = \frac{\sum_{j=1}^{|M|-1} \frac{|ET_{ij}^{x+}|+1}{|ET_{ij}^{x+}|+|ET_{ij}^{x-}|+2}}{|M|-1} \quad (4)$$

Finally, the social trust of an individual in the community is given by:

$$STrust(u_i) = \alpha.PopTrust(u_i) + (1-\alpha).EngTrust(u_i) \quad (5)$$

Where α represents a value of a weight in the range [0…1] if:

- α = 1, the social trust of an individual determines how much other members trust that individual.
- α = 0, the social trust of an individual determines how much this individual trusts other members in the community.

To show the utility of Nepal and Sherchan's (2011) model and to analyze the sustainability of social networks using their model, they carried out a few experiments. In their experiments, they used a data set that represents interactions in an online community between students at the University of California. Their analysis concludes that the number of unique members (leaders and mentors) is less in highly interactive communities than in a community with lower interaction numbers.

However, their analysis has numerous of limitations. First, there is no reported method on how to distinguish a positive interaction from a negative one. Therefore, all interactions in their database are treated as positive interactions which violate the dynamic nature of interactions on social networks. Second, the effect of the context of interaction still needs to be studied as their data had no context information. Finally, the temporal effect of interactions still need to be studied. Therefore, we can conclude that their work does not capture all of the communication behavior by focusing only on a static snapshot of the social network.

Zhan and Fang's (2011) work calculated trust scores between two directly connected members from three different aspects, namely, profile similarity, information reliability, and social opinions. The system then returns a trust score representing the actual trust from one member to another. Their research is based on two hypothesis: (1) trust is transitive, e.g., if Alice trusts Bob and Bob trusts Dan, then Alice trusts Dan; and (2) the trust between two members are normally known, e.g., mutual rating between members. However, this is not realistic as it is hard

to get users to rate their degree of trust in their friends in public. Moreover, most OSNs providers do not provide explicit trust rating service for their users.

Adali et al.'s (2010) work gives measures for trust that can be applied to very rapid social networks, such as Twitter. In their work, the trust is evaluated based on two types of interactions between individuals, namely, conversation and propagation information from one member to another. Conversation trust specifies the frequency and the length of the conversation between two members. On the other hand, propagation information in OSNs indicates a high level of trust placed on the propagated information and on its source. Their analysis reveals that their two types of measures are tightly correlated in terms of members involved, communities formed, and actual forwarding behavior as an indicator of trust. However, unlike our approach that takes the context of the communication between two members to evaluate the trust relationship, they developed their statistical measures based on timing and sequence of communication, not textual content.

Conclusion

Dynamic data, such as text posts, are one type of information available on OSNs. Users of OSNs generate a large volume of text every day, and the question "To whom should I broadcast this information?" is being posed everyday by users. Unfortunately, existing privacy models on major OSNs fail to address how and to whom users' dynamic information should be broadcasted to. As a result, a variety of personal and sensitive information is being exposed to a wide range of people on OSNs. Therefore, in an effort to answer this question, we suggest a unique two phase privacy framework that can be used; first, to detect dangerous information disclosure from users' text post; and then, suggest a list of trusted friends to whom the user may safely broadcast his/her deemed dangerous posts.

Detecting information disclosure in dynamic data on OSNs is a very challenging and complicated task, thus, a limited number of researchers addressed detecting information revelation in users' text posts. Existing literature focused fundamentally on detecting identifiers from status updates. Identifiers are information that can be used to directly identify the person. Technically, these approaches use a pre-collected set of users' profile information and the search engine to determine whether the text post could be used to identify the person. Other approaches employ a large knowledge base of preexisting categorized words that can be used to determine the content category.

Our approach in detecting information disclosure in dynamic data differs fundamentally from the aforementioned approaches from two aspects, namely, its scope and its technical approach. First, the scope of our approach is to detect dangerous information, which could place a person or property at risk from a users' text post. This is a very important problem that has not been addressed before in the literature. Second, from a technical point of view, in our approach we do not collect users' profile information, neither do we use a large database to detect dangerous information revelation. Collecting users' profile information has multiple implications. First, it can be used primarily to reason out revealing user identify. Second, users' may express some discomfort toward collecting and storing their private data. Finally, users normally provide false profile information which affect the accuracy of the results.

While limited literature have suggested models to evaluate trust based on interactions between users on the web. Existing approaches are either not applicable to OSNs, or they have different objectives than our approach, which suggests leaders and mentors in communities. Therefore, our approach differs fundamentally from aforementioned references. First, our approach evaluates trust on more finely tuned interaction methods that are well known and used

by users on OSNs. Second, our approach in suggesting trusted friends is dynamic as it depends on the content. Some content are dangerous and should be restricted to only trusted friends, while other content are not dangerous and can be broadcasted to everyone on the social network. Content driven interaction-based trust as a mechanism to manage the privacy of dynamic information on OSNs has not been studied before.

CHAPTER 3. THE PROPOSED APPROACH

To overcome the previous limitations and in an effort to protect users' safety and privacy on online social networks, we propose a framework of natural language to automatically identify messages that a user of a social network such as Facebook or Twitter has prepared for posting that contains potentially dangerous information. For example, the message might indicate that the poster presently is not at home and will not be home for several more hours. Such a message could alert a possible thief to a vacant home as an easy opportunity for a break in. After identifying and warning the poster and before the message is actually posted, our work will partition the user's potential recipients into groups and suggest which group is safe for each post. Our approach is dynamic as the suggested list of trusted friends will change when the content of the status update changes.

Our proposal envisions a small group of types of dangerous information that should be restricted.

Definition. A dangerous post is a post that reveals information about someone, either the poster or another person, which could lead to that person being harassed or placing property or person at risk. Major categories of dangerous posts on social networks are: identity, work, location, day-to-day activities and social plans. Criminals often easily find the time and location of user's activities and social plans: therefore, this work focuses on detecting location revealing information and constructing the "Location circle of trust" with an intention to expand our approach to detecting other categories of dangerous information.

In this section, we discuss the proposed approach. Figure 16 provides an overview of the two major components, namely, the *Awareness System*, and the *Circles of Trust System*. The

Awareness System will address the gap between the user's mental model and the countermeasures against revealing dangerous information. The Awareness System detects dangerous information from the user's text post using a set of given detection rules. The system then sends a message to the user about the information disclosure. To mitigate the risk associated with revealing such information to all friends on the social network, the Circles of Trust System serves to slightly categorize content and to restrict the flow of information to only relevant trusted parties. The approach to do the initial assignment of people (friends on Facebook) to the "Location trust circle" is done automatically.

The first component is the Awareness System and it has three steps. First, the natural language parser grammatically parses the plain text input (user's text post). The result of this step is a part-of-speech tagged text. Second, the dangerous information extractor extracts dangerous information from the part-of-speech tagged text, if there is any, according to a provided set of detection rules. Third, the categorizer assigns the appropriate tag/topic category to the dangerous post. Tags allow the system to categorize the topic of dangerous post and are used in our approach as a basic means for restricting dangerous information to only relevant trusted parties.

The second component is the Circles of Trust which has four steps. First, the interaction records file is partitioned into topic groups. The output of this step is topically categorized interaction files. Second, the trust metrics extractor selects trust metrics from individual interaction files using the collected interaction records files. The output of this step is the trust metrics at different hierarchal levels including individuals' comments rate, appreciation rate, tagging rate, and threaded comments rate. Third, the trust rating calculator suggests weights for each trust metric and calculates individuals' trust scores. Fourth, the Circles of Trust generator suggests a threshold to be used in grouping trusted friends into the (Location circle of trust). The

last step is to automatically generate a list of trusted friends who can safely see the detected dangerous post. The following sections explain these steps in greater detail.

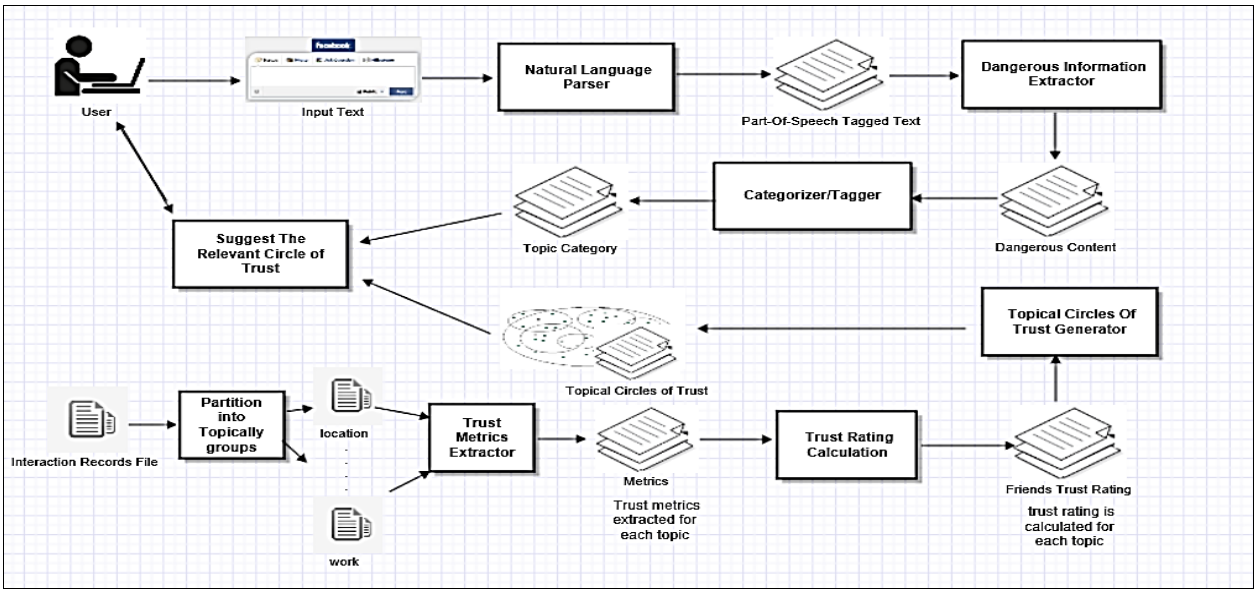

Figure 16. An Overview of the Proposed Approach.

The Awareness System

The first phase of our approach involves recognition and detection of key phrases in a post that might reveal the time and/or the location of the user's activities and social plans. Revealing this information on the social network might compromise the safety of the user or someone else.

Text Representation

We first define a representation for a text post. We define a "text post" (TP) as a sequence of words with an arbitrary length *N*. A text post (*tp*) can represent a single word as well as any combination of words $W = \{w_1, w_2, \ldots, w_n\}$. The text post can include symbols and emotional icons as well. For example, Input: tp_n = "going out for a run!" where tp_n denotes an input text post *tp* with *n* words.

Natural Language Parser

When the user uploads a text post to the status update section in the social network, the main process in the Awareness System reads and sends it to the natural language parsers, the Stanford Parser (Chen & Manning, 2014), which works out the grammatical structure of the sentence, e.g., identify which words are the subject or object of a verb. In our work, we use the Stanford Part-of-Speech Tagger (POS Tagger), which is software that reads plain text input in the English language and assigns a part of speech tag to each word or token. Figure 20 presents an example of the plain text input "with Carla and Linda at the mall" and output shows POS tagged text. Table 3, presents the list of tags and a description of the corresponding part of speech.

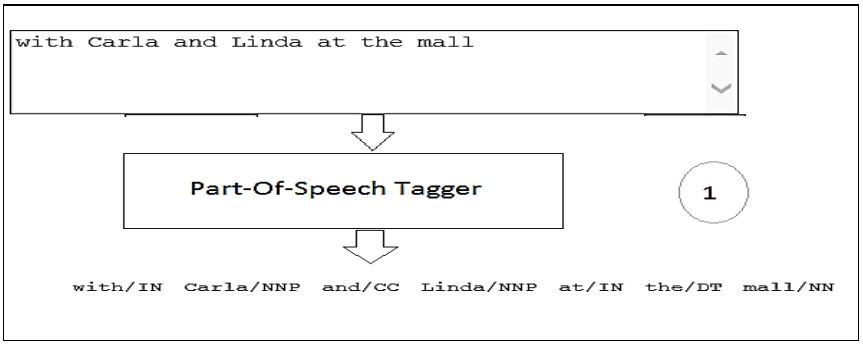

Figure 17. Example of Plain Text Input and Output Shows POS Tagged Text.

Table 3.

Alphabetical List of POS Tags.

No.	Tag	Part of Speech
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NP	Proper noun, singular
15.	NPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PP	Personal pronoun
19.	PP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection

Table 3. *Alphabetical List of POS Tags.(continued).*

No.	Tag	Part of Speech
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determinater
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Note: Adapted from A fast and accurate dependency parser using neural networks, by Chen, D. & Manning, C. D. (2014), *Proceedings of EMNLP*, 2014, p. 1.

Dangerous Information Extractor

The tagged text is then sent to the dangerous information extractor, which extracts temporal phrase patterns and evaluates whether the text is dangerous or not. The dangerous information extractor uses a list of extracted patterns stored in a knowledge base to determine whether a post is dangerous or not. These extracted patterns are selected based on the observation and experimental study we have performed on more than 16,000 real Facebook text posts collected from actual Facebook users who were willing to participate in our study after reading about our research.

The algorithm for extracting dangerous information runs as the user uploads their text post. The algorithm first uses the Stanford POS tagger library to pair every word in the post with its POS tagging. Second, the algorithm searches the database for any matching pattern. If the

detected keyword and its tags match with any detection pattern in our database, then the post is claimed to have dangerous information. Figure 18 illustrates the steps of the dangerous information extractor. First, it accepts the tagged text as an input, then it extracts any temporal data, and finally it matches the extracted temporal data with existing detection patterns in our database to determine whether the text post has dangerous information or not.

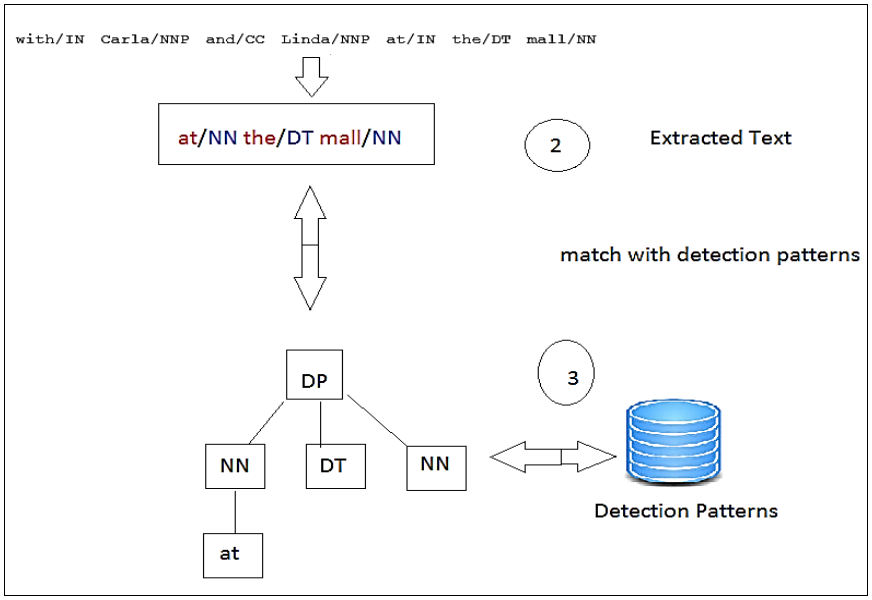

Figure 18. An Example of Dangerous Information Extracting.

Our detection patterns can be categorized as follows:

- Subset of time and location prepositions followed one specific tag, e.g., “to” preposition followed by a proper noun.
- Subset of time and location prepositions followed by combinations of two specific tags, e.g., “at” preposition followed by determiner followed by proper noun.
- Keyword, e.g., the symbol “@” was used in multiple posts instead of the “at” preposition to reveal the location of the user.
- Specific verbs followed by one specific tags, e.g., “heading” followed by noun.

- Specific verbs followed by two specific tags e.g., “going” followed by “to” followed by determinate.
- Other phrases which are commonly used on social networks, e.g., “on my way home.”

Table 4 presents a snapshot of the detection rules that we developed to be used by our detection tool. Figure 19 provides a snapshot of the detection tool we have developed. While Figure 20 provides a snapshot of the class diagram of our the proposed tool.

Table 4.

List of Detection Rules.

Keywords	Keyword Suffixes Tags
Prepositions	
at	
1	at/IN + CD
2	at/IN + NN
3	at/IN + NNP
4	at/IN + NNPS
5	at/IN + DT + NN
6	at/IN + DT + NNP
7	at/IN + DT + CD
8	at/IN + DT + NNS
9	at/IN + DT + NNPS
in	
10	in/IN + NNP
11	in/IN + CD
12	in/IN + few/JJ + days/NNS
13	in/IN + downtown/JJ + Chicago/NNP
14	in/IN + JJ + NN
15	in/IN + the/DT + theater/NNP
16	NN,NN
17	NN,NNP
on	
18	NNP
19	CD

Table 4. *List of Detection Rules (continued).*

Keywords	Keyword Suffixes Tags
Prepositions	
to	
20	on/IN + NNP
21	on/IN + CD
for	
22	for/IN + three/CD + weeks/NNS
23	for/IN + few/JJ + days/NNS
24	for/IN + Sunday/NNP
Verbs	
going	
25	TO,DT
26	TO,NNP
heading	
27	NN
28	RP
29	TO,DT
30	TO,NNP
leaving	
31	TO,DT
32	TO,NNP
33	NN
moving	
34	TO,DT
35	TO,NNP
36	TO,NN
hitting	
37	DT
head	
38	RB
39	NN
40	TO,NNP
41	TO,NN
Phrases	
1	**on my way**
2	**out for**
3	**made it into**

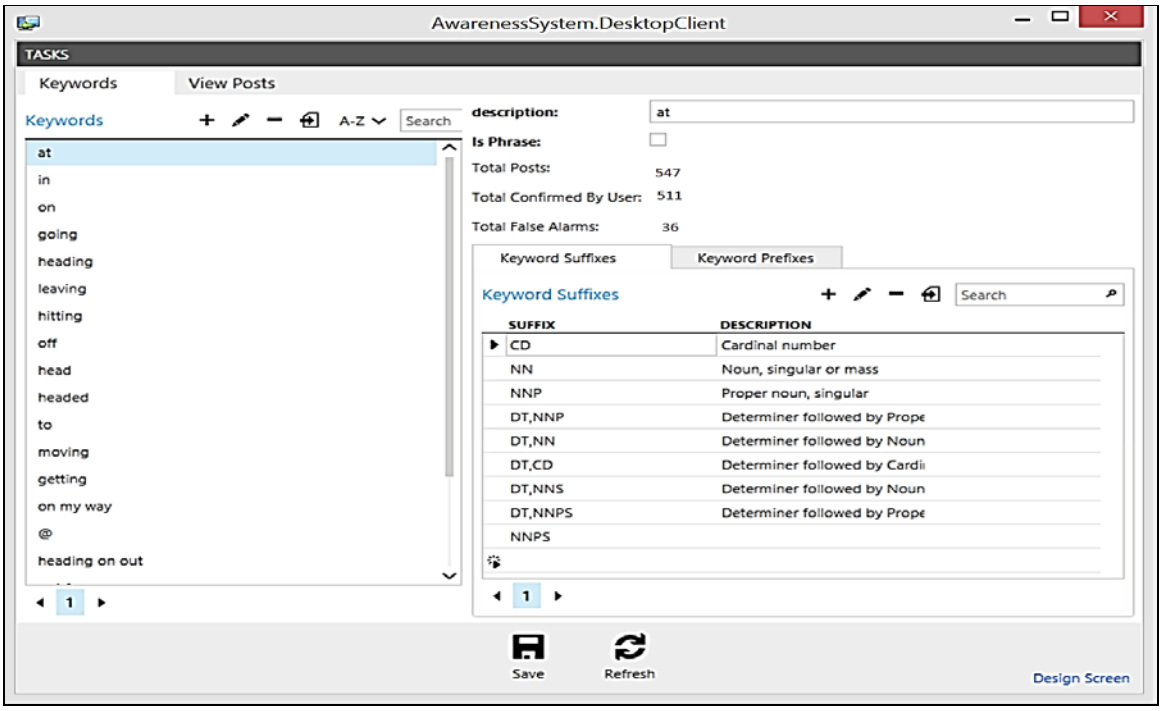

Figure 19. A Snapshot of the Detection Tool.

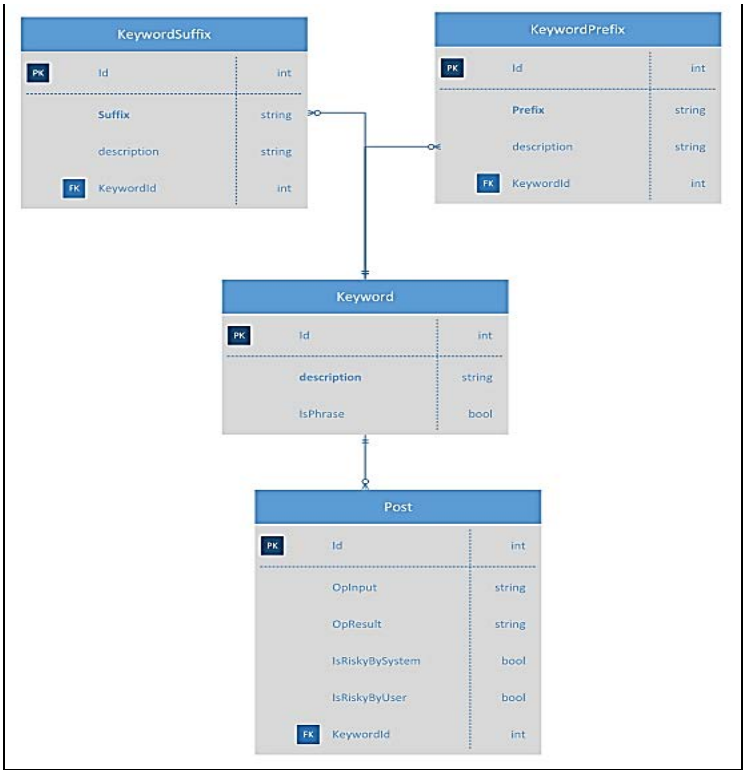

Figure 20. Awareness System ERD Diagram.

Information Categorizer/Tagger

The primary purpose of the categorizer/tagger is to assign the appropriate topic category to the detected dangerous post. Categories/tags also serve as the skeleton of our privacy control, clueing in both the detected dangerous post and the relevant circle of trust. Tags or categories are used in our approach as a basic means of restricting information to only relevant parties. Figure 21 provides an illustration of this concept.

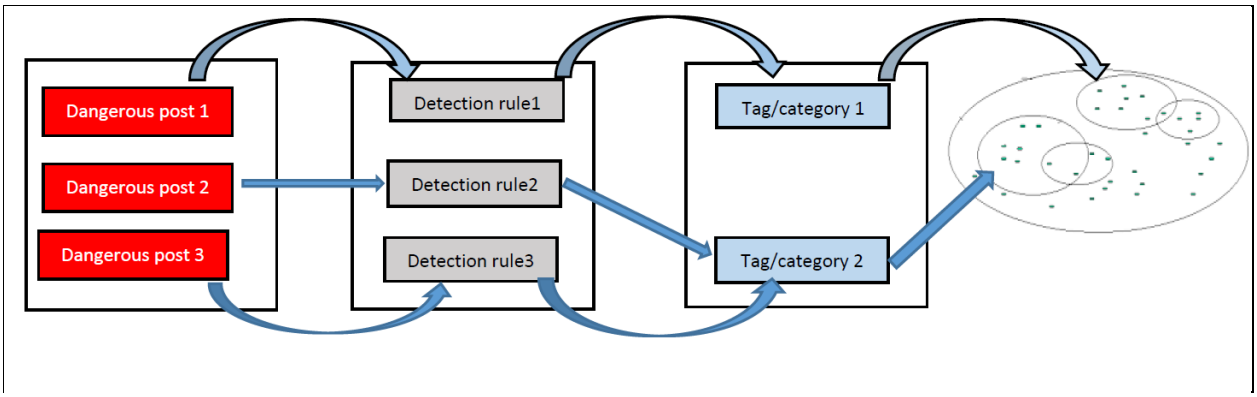

Figure 21. Tagging/Categorizing in Our Proposed Approach.

Figure 22, provides a snapshot of the output of the Awareness System tool that we developed to detect location revealing information from a users' text post.

- *OP input*: is the text post that the user wants to post on the social network.
- *OP Result*: is the parsed and tagged text by the part-of-speech tagger.
- *Is Risky By System*: represents that the detection system detected that this post may contain revealing information about the user's location.
- *Is Risky By the User*: represents the user's confirmation to classify this post as a dangerous post as claimed by the system. This feature is used to evaluate the detection accuracy of the tool as described in Chapter 4.

- *Keyword*: is the keyword that was detected by the system to be used in identifying dangerous information.
- *Detection Rule*: is the detection rule used to identify the dangerous information in the text.
- *Assigned Tag*: is the category of the detection rule.

It is important to note that the scope of our approach is to detect revealing information related to the location and time of users' activities and social plan from the text post. Therefore, all detection rules in our system are assigned the "Location" tag. However, the way our system is developed will allow for easy extensibility in the future as we have a plan to develop more detection rules to detect other types of dangerous information from text posts and therefore assign different tags/categories to newly added rules.

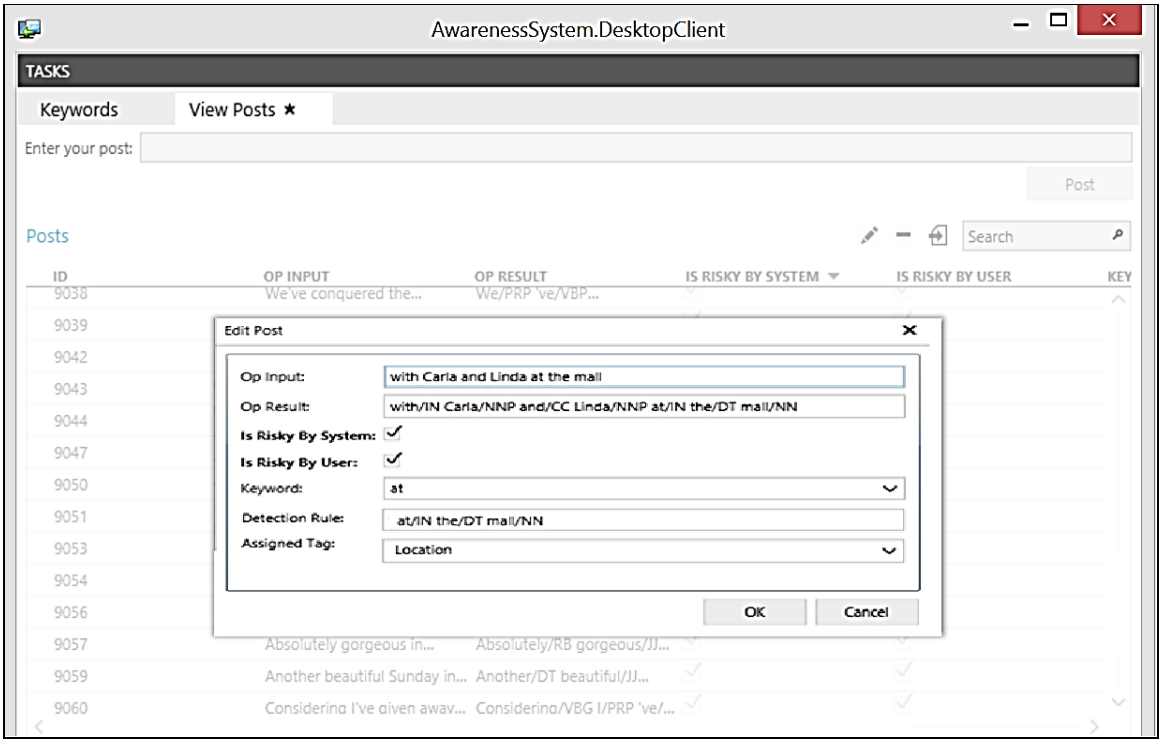

Figure 22. A Snapshot of the Output of Our Detection Tool.

The Circles of Trust

The objective of this component is to improve the existing trend of sharing dynamic information on social networking sites toward a trusted-friends paradigm. Figure 23 illustrates the difference between the two methods. In Figure 23 (a) the user broadcasts the status updates to all friends (social links) on the social networks. In our proposed approach in Figure 23 (b), the user only broadcasts the status update that may reveal the time and/or location of his activities and social plans only to trusted friends. The approach to do the initial assignment of people to the Location circle is done automatically.

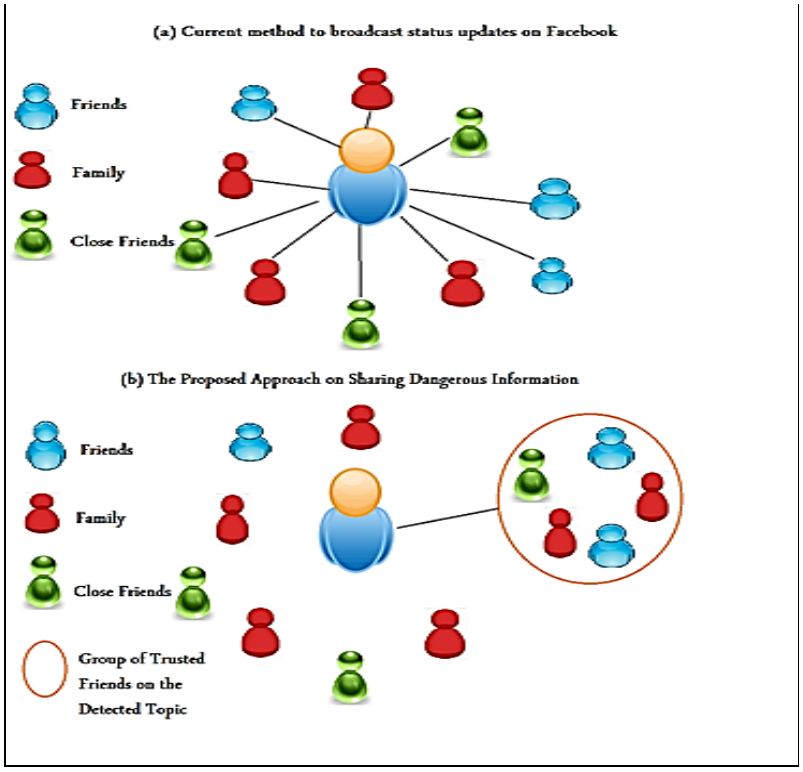

Figure 23. The Difference between Current and Proposed Methods for Sharing Status Updates.

Trust Sources on Online Social Networks

Members of online social networks, such as Facebook, could have a variety of information sources that they can use to identify friends' trustworthiness. Each source represents a factor for trust establishment between a user and a friend. For example, in Figure 24, Alice can use different methods to evaluate Bob's trustworthiness within the social network as follows:

First, Alice can ask Bob for his credentials to confirm his identity, e.g., skills or other properties. Then Alice can check the validity of Bob's credentials from the authority that issued Bobs' credentials (e.g., employer). However, this technique is time-consuming due to the large number of social links that users establish on these sites. In addition, this technique does not fit the dynamic nature of the data on social networks.

Second, Alice can use recommendations of other members she trusts. However, trust is a personal preference and is not transitive in nature. For example, if Alice trusts Sarah and Sarah trusts Bob that does not mean that Alice can trust Bob. Besides, this method lacks the gradient we have in our proposed approach which is the context of trust.

Third, Alice can confirm from members she trusts whether Bob is skillful in certain disciplines and the level of his expertise. However, this approach might be a better fit for professional social networks such as, LinkedIn where the relationship is based on the member's professional information. This approach is useless in a friends-based social network such as Facebook, Google+ or Twitter.

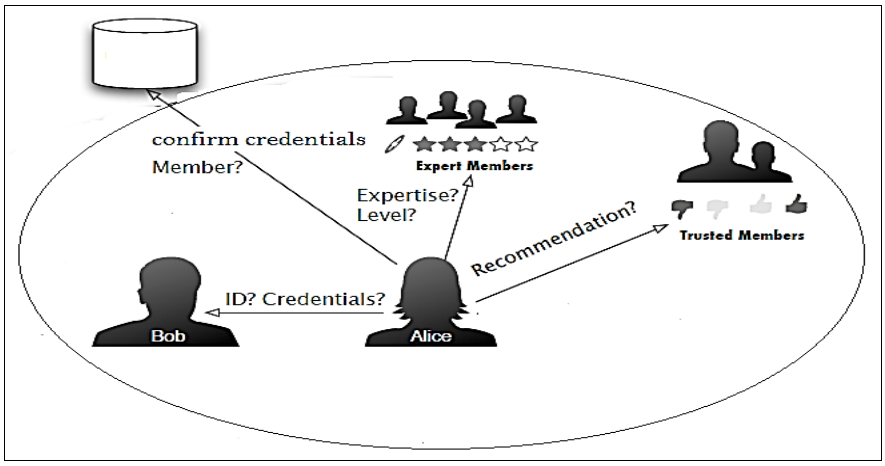

Figure 24. Trust Sources Diversity within Online Social Networks. Adapted and modified from Social-compliance in trust management within virtual communities, by Yaich, R., Boissier, O., Jaillon, P., & Picard, G. (2011, August 22-27). In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (pp.322-325). Lyon, France, p. 3.

Trust-based privacy control for protecting dynamic data on social networks requires much more maneuvering as the content of this data can contain a wide range of topics. Some topics may not expose the user to risks and therefore can be seen by everyone on the social network, while some topics might put the user at risk such as being attacked or robbed. Trust-based privacy control for managing access and denial can be a very complicated task due to a lot of factors that can be used to define personal trust preferences. However, we believe that configuring an interaction-based trust model for managing the privacy of dynamic data has to depend on the context of the interaction between a user and their friends.

Methodology

To reiterate the objective of the problem, when a user posts status updates that potentially have dangerous information, the proposed system would suggest sharing these posts with only a set of trusted friends and/or to whom the status updates are relevant. This is a better approach than broadcasting the updates to a user's entire list of friends on the social network. This way the

system protects the user's safety and privacy. Besides, our approach could also work as a filter to reduce the amount of irrelevant status updates that clutter friends' news feeds from the receiver's point of view. Thus, the problem remains that when a user posts a status update, how would he or she select a set of trusted friends who do not have malicious intentions and are interested in reading that status update?

The selected approach to this problem was to collect a set of files that contain users' posts, friends and interaction between them to identify list of friends with whom the user is comfortable to share and discuss a specific topic. Figure 25 shows the overall structure of our proposed framework. In our approach, trust is related to the absence of interaction as there would be no need to trust anyone whose activities were not continually visible and whose thought processes were not transparent. It is natural to trust people we communicate with more often than with people who we do not.

The privacy control for suggesting the set of friends is dynamic. This means that when a user decides to post a text message (status update), the system had to: (1) check the content to be posted to gauge whether it is dangerous or not; and then (2) suggest a different set of trusted friends depending on the content of the status update. The content of the status updates can be about a general topic and therefore could be broadcasted to active links. On the other hand, the status updates content could be about any defined dangerous topic on the social network, e.g., location, work, family, or identity and, therefore, could be broadcasted to only active links on this particular topic. The assumption here was that friends who show continuous signs of interaction with the user can be considered trusted friends, because friends who do not show any signs of interaction are: (1) either not interested in the user's status updates; or (2) not active on

the social network in general; or (3) they have malicious intentions and do not want to be noticed.

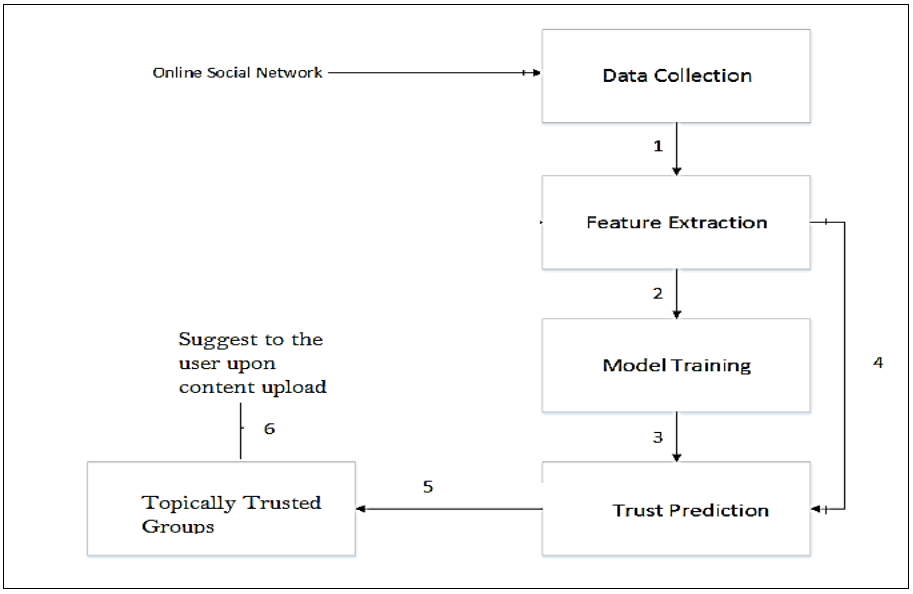

Figure 25. An Overall Structure for the Proposed Approach.

First, we start by defining a representation for a circle of trust.

Definition. A circle of trust (*COT*) is a collection of friends or users $U = \{u_1, u_2, \ldots u_n\}$ that are trusted with specific dangerous topic t_n € $T = \{t_1, t_2, \ldots, t_n\}$ by showing continuous signs of interaction with the poster about that particular topic t_n and, therefore, are permitted to view that poster's content, which was deemed dangerous by the awareness system and any other general posts. The circle of trust is denoted by $COT_{dtn}(U)$ where a user u_n can be in more than one circle of trust. The user may not be assigned to any circle as well, and in this case the user will not be permitted to see any post that falls under the dangerous post categories.

These circles serve as a mechanism to automatically categorize users' contacts on the social network into topically trusted groups and to automatically control who sees relevant dangerous posts. A friend who is trusted in work revealing information may not necessarily be

trusted in location revealing posts. In addition, our approach can be used as a mechanism to filter the news feed from the viewer's point of view. For example, some friends may get annoyed by the amount of the day-to-day activities that a particular user loves to talk about. Therefore, using the Circles of Trust will allow the user to target both trusted and interested audiences and make sure that the poster stays relevant to friends who are seeing his/her posts. Figure 26 provides an illustration of this concept.

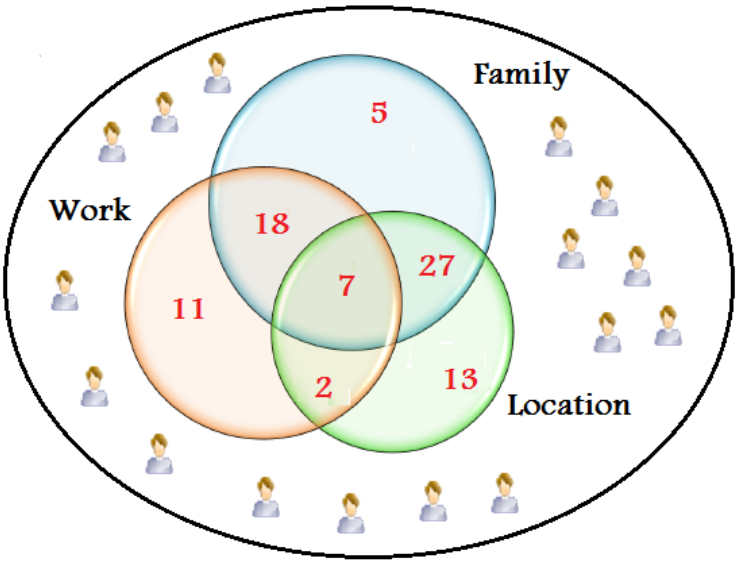

Figure 26. Illustration of Circles of Trust in Our Proposed Approach.

Major categories of dangerous information that people share on social networks are: identity, family, work, and location revealing information. Therefore, each of the inner circles, as illustrated in Figure 26, will represent one of these categories, e.g., Location circle, Family circle, and Work circle etc. Each of these circles will contain people whose interactions with the poster are topically motivated. In other words, their interaction is centered on a specific topic such as the time and location of the user's activities and social plans.

These inner circles may intersect with each other as some users might belong to more than one circle if they have shown signs of interaction with the poster about more than one topic. People in a particular circle will be allowed to see text posts related to that particular topic as well as any other not dangerous social posts on a variety of other topics such as sports, politics etc. When a user belongs to more than one circle, then he/she will be permitted to see dangerous posts under the topical category of these circles as well as any other non-dangerous posts.

Based on their interaction behavior, some people may not be qualified to be grouped in any circle. Therefore, privacy setting for friends in this category will be set to automatically prevent any post deemed dangerous from being visible to them. On the other hand, they will still be able to see other non-dangerous posts.

Topical Groups Extractor

To be able to construct the Circles of Trust, we need to consider a set of users' interactions records on different dangerous topics, e.g., work, family, and location. However, due to limited resources and the size of the interaction data set, this dissertation focuses mainly on location reveling information. Therefore, the aim of this step is to extract a set of interaction records on the "location" topic.

Since the messages in our data set were not topically categorized and since the scope of our approach was to detect and restrict dangerous text posts that could possibly reveal the time and location of a user's activities and social plans, we used our Awareness System detection tool. The Awareness System was explained earlier in this section. Its detection system was used to extract text messages about the topic "location." We also collected these posts' related interaction records from the data set. In total, we used a set of 46 detection rules and phrases that social network users use repeatedly to indicate their current location or future destination. Text

messages were considered to be about the location topic when they satisfy any of the listed detection rules or phrases in Table 4. Since the Awareness System has a detection rate of 85% and 11% false alarms, we also needed to manually adjust the extracted data set to include the 5% undetected location-revealing posts and remove the 11% mistakenly deemed dangerous posts.

Trust Metrics Extractor

The aim of this step is to identify a set of trust metrics and to extract those metrics automatically from users interaction records file. We made the assumption that the existing interactions between two users on the social network on a particular topic can influence the trust between them on that particular topic. Thus, we first needed to identify metrics that can be used to model this interaction-based trust. These metrics needed to be applied directly to the OSN environment.

Interaction methods on social networks could be categorized to implicit interaction methods such as, chatting and messages, and explicit interaction such as likes, comments, sharing, tagging etc. However, giving the restriction on collecting data related to implicit interaction methods, we limited the focus of our research to explicit interaction methods, such as comments, likes, tags, and replies. Moreover, our interaction-based trust metrics included the number of positive and negative feedbacks given from friend *y* to user *x*. We also provided different metrics for each comment structure, namely, the flat comments discussion structure and the threaded comments discussion structure.

Each node (user) can observe and record the interaction behavior of its neighbor nodes (friends) with whom it maintains a direct relationship (friendship) towards its content. Therefore, interaction-based trust metrics were defined explicitly as comments metric, hieratical threaded discussion metric, appreciation or like metric, and tagging metric.

First, we provide a definition for the trust-based social network then we proceed to explain and discuss each of the four defined metrics in detail in the following sections:

Trust-based social network

Definition. The trust-based social network is a subset of the original social network, where members can trust each other and safely share their personal or sensitive information without privacy concerns.

The trust-based social network is denoted as $OSN = (V, E, TL, T)$, is a trust-based network model consisting of social graph, trust between users, and the topic of trust,

Where:

- V is the set of vertex, where a vertex in the graph stands for a user of the social network.
- E: is the set of edges, where a direct edge from a user x to a user y indicates the existence of social link between a user x and a user y. For example, x is a friend with y.
- TL: the trust level set which contains the trust level that one user has on another user is defined as:

$$TL = \{ (x, y, TS(x, y, t)) \mid x \in V,\ y \in V,\ x \neq y \}$$
$$\text{And } t \in \{'Location', 'identity', 'work', 'family'\} \quad (6)$$

A trust feature vector from user x to user y is defined as:

$$v(x, y) = (W_{fdr}\, CR_t(x, y),\ W_{hdr}\, HDR_t(x, y),\ W_{ar}\, AR_t(x, y),\ W_{tr}\, TR_t(x, y)) \quad (7)$$

Where:

- $0 \leq W_{fdr}, W_{hdr}, W_{ar}, W_{tr} \leq 1$ and
- $0 \leq W_{fdr} + W_{hdr} + W_{ar} + W_{tr} \leq 1$ and

- $CR_t(x, y)$ denotes the Comments Trust Metric
- $HDR_t(x, y)$ denotes the Hierarchical Threaded Discussion Trust Metric
- $AR_t(x, y)$ denotes the Appreciation or like Trust Metric
- $TR_t(x, y)$ denoted the Tagging Trust Metric

The elements of the trust vector are listed in Table 5, and then each element is discussed in detail in the following sections:

Table 5.

Elements of Trust Vector.

Metric	**Description**
$CR_t(x, y)$	The ratio of $noc_t(x, y)$ and $tnotp_t(x)$
$HDR_t(x, y)$	The ratio of $notc_t(x, y)$ and $noc_t(x, y)$
$AR_t(x, y)$	The ratio of $nol_t(x, y)$ and $tnotp_t(x)$
$TR_t(x, y)$	The ratio of $not_t(x, y)$ and $tnotp_t(x)$

Comments metric

The number of comments from a friend *y* on the text post of *x* on topic *t* implies that friend *y* is interested in forming a bond with user *x* by taking the time to write a comment on the content of *x*. As a result, this bond may lead to a trust relationship between the pair on that particular topic. Figure 27 illustrates this concept.

Therefore, the rate of comments on topic *t* of user *x* made by *y* is given by:

$$CR_t(x, y) = \frac{noc_t(x, y)}{tnotp_t(x)} \tag{8}$$

Where:

- $noc_t(x, y)$: is the number of comments made by user *y* on the text post of *x* about the topic *t*.
- $tnotp_t(x)$: is the number of total text posts posted by user *x* on the social network about the topic *t* e.g., the topic location.

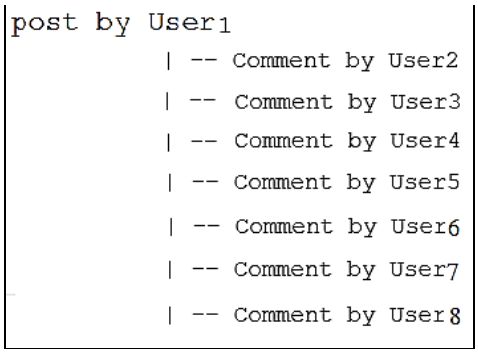

Figure 27. An Example of Commenting System on Facebook.

Hierarchical threaded discussion metric

We first presented the threaded comments concept on Facebook, also known as replies, then we explained why it is an important trust metric. Each post or comment on an original post will contain a "reply" button underneath it, next to the "like" button. If a user replies to a comment, their reply will appear beneath the comment and indented slightly. Or, if someone has already replied to the comment, beneath the previous reply. Figure 30 illustrates the concept of hierarchal threaded comments or replies on Facebook.

Under the assumption that trust was a key factor to trigger the discussion process between the pair on the social network, we believed that hierarchal threaded discussion was an important factor in determining trusted friend status. Moreover, longer discussion may intrinsically incorporate more ideas, and hence, might be another interesting reason to collect metrics of a

post. Therefore, in our work we also consider the hierarchal threaded comments (replies) between the pair on a particular post to calculate the level of trust. The individual with whom you discuss things with more, you trust more.

Figure 28 illustrates this concept of a comment thread, such as the comments thread between User1 and User2. This threaded comment connects multiple comments with links such that each comment can be linked to the comment it is in reply to, ensuring future readers can more easily follow the conversation and add their constructive thoughts to the discussion.

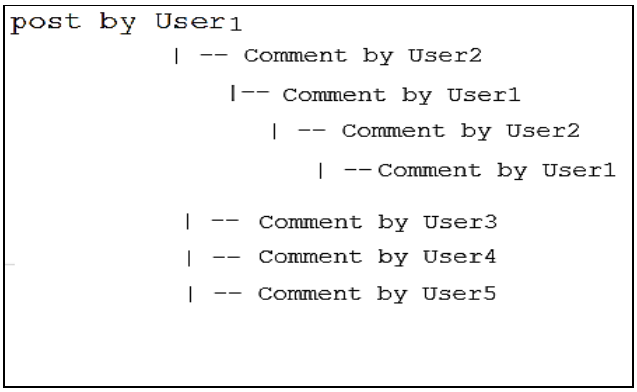

Figure 28. An Example of Hierarchal Threaded Comments.

The rate of the hierarchical threaded discussion on topic *t* between user *x* and a friend *y* is defined as:

$$HDR_{t}(x,y) = \frac{notc_{t}(x,y)}{tnoc_{t}(x,y)} \quad (9)$$

Where:

- $notc_{t}(x,y)$: is the number of threaded comments that user *y* made on the text post *p* of *x*.

- $tnoc_t(x, y)$: is the number of comments that user *y* made on the text content of user x.

Appreciation metric

The number of "likes" received from a friend y to the text posts written by the user x is also an indicator that the y is interested in x's contents. So, the rate of appreciation or like of the text posts *p* about the topic *t* of the content of *x* by *y* is:

$$AR_t(x, y) = \frac{nol_t(x, y)}{tnotp_t(x)} \quad (10)$$

Where:

- $nol_t(x, y)$: is the number of likes made by user *y* on the text post of *x* about the topic *t*.
- $tnotp_t(x)$: is the number of total text posts posted by user *x* on the social network about the topic *t*.

Tagging metric

We first presented the concept of tagging on OSNs, then we explained why it is an important trust metric. Users of social networks such as Facebook can use the symbol "@" followed by the person's name of whom they wish to identify in their post, photo or status update that they share. A tag may also notify that tagged person that he/she has been mentioned or referred to in a post or photo. Tagging also provides a link back to their profile. Moreover, the post a person is tagged in will also be added to that person's profile or timeline on Facebook. Therefore, it will be displayed on the user's wall also and be available for the user's friends to see. There are two types of tagging on Facebook, namely, photo tag and status update tags. Photo

tag is tagging someone in a photo to identify them in the photo. While status update tag is tagging someone in a status update to make sure that they see the post. Based on the scope of our research, we were primarily interested in status update tags.

Tagging is an important indicator of an existing trust relationship between the pair. Tagging might be an indicator of many location and activity related exchanges that the tagger and the tagged person are sharing:

- It might indicate the tagger (the owner of the post) and the tagged person might be going to the same location sometime in the future.
- Or they might be planning to attend the same social event.
- Or they might be currently both at the same location
- Tagging could also indicate a link to people in the real world.
- It could also mean that the two users share some interest or hobbies which could also influence the trust relationship.

The rate of tagging of a friend *y* by the user *x* about the topic *t* is defined as:

$$TR_t(x,y) = \frac{not_t(x,y)}{tnotp_t(x)} \tag{11}$$

Where:

- $not_t(x,y)$: is the number of tags made by user *x* to a friend *y* about the topic *t*.
- $tnotp_t(x)$: is the number of total text posts written by user *x* on the social network about the topic *t*.

Weighted metrics

The next part in our approach was to weight each trust metric as not all metrics are evenly influential when determining trusted friends. For example, some interaction methods are

popular among social network users such as, likes and comments. Some other interaction features such as reply (threaded comments) have been recently introduced to the users of Facebook. Reply features were added in May 2013. On the other hand, people use tags rarely and for socially close friends. Therefore, this section provides insights into each metric weight calculations.

With the above information being calculated for each friend the comments metric weight was defined as:

$$W_{fdr} = \sum_{k=1}^{n} \frac{CR_t(x,k)}{n} \tag{12}$$

Where:

- n: the number of friends (neighbor nodes)
- $CR_t(x,k)$: is the metric rate of a friend k on the content of user x

While, the threaded comments metric weight is defined as:

$$W_{hdr} = \sum_{k=1}^{n} \frac{HDR_t(x,k)}{n} \tag{13}$$

And, the appreciation metric weight is defined as:

$$W_{ar} = \sum_{k=1}^{n} \frac{AR_t(x,k)}{n} \tag{14}$$

Finally, the tagged metric weight is defined as:

$$W_{tr} = \sum_{k=1}^{n} \frac{TR(x,k)}{n} \tag{15}$$

This approach in assigning weights gives higher weight to the metric that is used more frequently by friends in a set of status updates. On the other hand, it gives lower weight to a metric that is not commonly used by friends in a set of status updates. In addition, this approach is adaptable to each user interaction behavior with friends. Trust is a personal preference; for example, users may consider some interaction method to be more important than the others in term of determining trustworthy friends. To allow users to reflect their view of trust on social networks, we suggest that users be given the flexibility to change the metrics' weight to reflect their view of trust on OSNs.

Topical Trust Scores Calculator

In this step, we use the combination of metrics defined in the previous steps to calculate individual trust scores. The combination of metrics can be used to examine the correlation between these metrics and the trusted users' prediction rate. The trust score of a user on a particular topic specifies the circle of trust that the user belongs to and, therefore, the eligibility of a user to see future posts under this category.

Each node (e.g., user) in the social network, can compute the level of trust with each of its neighbors, (e.g., friends) with respect to a particular topic, which binds the pair as follows:

$$TS\ (x, y, t) = T_{max}\ (W_{fdr}\ *\ CR_t(x, y) + W_{hdr}\ *\ HDR_t(x, y) + W_{ar}\ *AR_t(x, y) + W_{tr} * TR_t(x, y)) \quad (16)$$

Where:

- W_n : is the weight of each interaction metric, which comprises a percentage of the maximum possible rating. Users might be also given a choice to define their desired trust level by giving weights for each interaction-based trust metric.
- $T_{max} > 0$: is a predefined constant representing maximum trust rating.

The computed direct topical trust value belongs to the continuous 0-100 range. As the number of interactions (that is the number of comments, hierarchal threaded comments, likes and tags) increases by friend *y* to the text contents put by user *x* on the topic *t* as *y* became more trustworthy for *x*. Each node can compute the level of trust with its neighbors about the topic *t* and this trust level is not symmetric. Therefore, each node follows the same steps as previously explained to compute the level of trust with its neighbors.

This trust computation allows: (1) to consequently isolate malicious users, with no signs of interactions, from seeing posts that may reveal dangerous information; and (2) know benevolent friends who consistently show signs of interaction with the user; and finally (3) allow the user to isolate trusted audiences in a particular topic to ensure the privacy and safety of users and the relevancy of the content to the viewer.

Circles of Trust Generator

With the above information having been calculated by each of the four metric modules, the aim of this step is to determine a friend's eligibility to view another user's location revealing content based on their interaction behavior with the user on the topic. We use a single threshold across all users to cut off the list of suggested trusted friends. The threshold is the average trust scores calculated in the previous step. With the above trust scores being calculated, the threshold used to separate trusted friends from non-trusted is defined as:

$$threshold\ (x) = \sum_{k=1}^{n} \frac{T(x,k,t)}{n} \tag{17}$$

Where:

- *n:* the number of friends (neighbor nodes)
- $T(x,k,t)$: individual trust scores

A friend's *y* eligibility to view future location revealing posts written by the user *x* will be determined as follows:

$$T\ (x, y, t) \geq threshold(x) \tag{18}$$

Where:

- $threshold(x)$: is owner's threshold or the trust rating requirement for location posts of user *x* out of a maximum trust rating of 100.

Circles of trust are dynamic in nature as they must be calculated upon the user logging into the social network. This module introduces a new way to characterize social network dynamics as the placement of people into the Circles of Trust will change as their topical interaction behavior changes over time.

Display Example

The goal of this section is to illustrate the proposed approach using a simple example in which, we, as the user, made a total of 57 location revealing posts. Friend 1 through 7 represents the list of friends who showed signs of interaction on location revealing posts. This list is different from the friends list in that it contains a subset of the user's friends or social links. For example, the user may have a 100 social links (friends), however, only a subset of these social links showed signs of interaction on a specific topic. The following steps show the process of how trusted friends are identified on the "location" topic:

Step1

We are assuming that the user's friends' interaction information is calculated and aggregated as given in Table 6. Where:

- Comments: is the total number of comments that each friend has made on all location revealing posts.

- Likes: is the total number of likes that each friend has made on all location revealing posts.
- Tags: is the total number of location revealing posts that each friend was tagged in.
- Threaded comments: is the total number of conversations that each friend was involved in with the poster (the owner of the post).

Table 6.

An Example of Interaction Information of User's Friends.

Friends	Comments	Likes	Tags	Threaded Comments
Friend 1	30	50	0	0
Friend 2	15	18	0	0
Friend 3	48	15	12	0
Friend 4	33	48	0	0
Friend 5	18	30	1	0
Friend 6	32	57	23	0
Friend 7	15	18	12	0
Friend 8	18	33	0	0
Friend 9	1	4	0	0
Friend 10	8	0	0	0
Friend 11	19	19	0	2
Friend 12	1	22	0	1
Friend 13	32	14	0	0
Friend 14	9	8	0	0
Friend 15	17	25	0	0
Friend 16	30	51	0	0
Friend 17	12	30	0	0

Step 2

Individual trust metrics are then calculated according to the earlier formulas presented in Table 7.

Table 7.

An Example of Computed Trust Metrics.

Friends	Comments Rate	Likes Rate	Tags Rate	Threaded Comments Rate
Friend 1	0.53	0.88	0	0
Friend 2	0.27	0.32	0	0
Friend 3	0.85	0.27	0.22	0
Friend 4	0.58	0.85	0	0
Friend 5	0.32	0.53	0.02	0
Friend 6	0.57	1.00	0.41	0
Friend 7	0.27	0.32	0.22	0
Friend 8	0.32	0.58	0	0
Friend 9	0.02	0.08	0	0
Friend 10	0.15	0	0	0
Friend 11	0.34	0.34	0	0.11
Friend 12	0.02	0.39	0	1.00
Friend 13	0.57	0.25	0	0
Friend 14	0.16	0.15	0	0
Friend 15	0.30	0.44	0	0
Friend 16	0.53	0.90	0	0
Friend 17	0.22	0.53	0	0

Step 3

Weights for each trust metric are calculated according to the weight formulas earlier presented in this chapter and are given in Table 8. Our weights calculation approach insures that the most influential trust metric is given the highest weight. As we can see in Table 4, most friends have used the commonly known interaction method "like." Therefore, this metric was assigned the highest weight when calculating trust scores. Comment interaction is the second most frequently used interaction method. Therefore, it will be assigned less weight than the like method, but more weight than the tags and threaded comments interaction methods. Replies or threaded comments are the least used interaction method by Facebook users. We believe that this is due to the fact that this feature was added shortly before we collected our data set.

Table 8.

An Example of Trust Metrics Weights.

Comments Metric Weight	**Likes Metric Weight**	**Tags Metric Weight**	**Threaded Comments Metric Weight**
0.35	0.46	0.065	0.115

Step 4

With the above weights being assigned to each interaction method, trust score, or the threshold, and list of trusted friends are presented in Table 9. TRUE, indicates that the friend is trusted, therefore is eligible to see future dangerous posts made by the user. While FALSE indicates that the friend is not trusted on this particular topic (location), therefore, he/she is not eligible to see posts under this category.

Table 9.

An Example of Trust Scores and Suggested List of Trusted Friends.

Friends	Trust Scores	List of Trusted Friends
Friend 1	59.03	TRUE
Friend 2	24.17	**FALSE**
Friend 3	43.60	TRUE
Friend 4	59.40	TRUE
Friend 5	35.71	TRUE
Friend 6	68.62	TRUE
Friend 7	25.60	**FALSE**
Friend 8	37.88	TRUE
Friend 9	4.38	**FALSE**
Friend 10	5.25	**FALSE**
Friend 11	28.81	**FALSE**
Friend 12	30.14	**FALSE**
Friend 13	31.45	**FALSE**
Friend 14	12.50	**FALSE**
Friend 15	30.74	**FALSE**
Friend 16	59.95	TRUE
Friend 17	32.08	**FALSE**
Threshold:	**34.67**	

Tool Implementation

To support our research, we built a Topical Circles of Trust (TCT) tool. As a first step, the tool starts by reading the collected interaction data sets and importing this information into the database. Second, the tool extracts the trust metrics automatically, calculates initial metrics weights, and initial trust scores (threshold) for each user. Third, the tool then compares

individual trust scores against threshold to suggest the group of trusted and interested friends with respect to the location topic. In this section, we present the tool architecture. Then, we provide the database design and the class diagram of the developed tool. Next, we proceed to provide an explanation of the developed methods. Finally, we provide a description of the user interface.

Tool architecture

Figure 29 shows a high level description of the **TCT** tool. The tool works as follows:

- First, the **TCT** tool reads the Excel data set and imports it into the database.
- Second, the **TCT** directly invokes the detection tool, developed in phase one, to detect location revealing posts and their interaction records.
- At detection, location revealing posts are assigned the "Location" tag to help identify these posts from other general posts.
- The tool uses the imported training interaction data set to extract information about posts, commenters, likes, tagged friends and threaded comments between the user and friends.
- The **TCT** directly invokes the trust metrics formulas in the source code to compute an individual comments rate, likes rate, tagged rate, and threaded comments rate between the user and each friend.
- After extracting trust metrics information, the tool then computes and assigns weights to the trust metrics.
- When the previous information has been calculated, the tool computes the initial trust score (threshold). The output of this step is the suggested weights for each trust

metric and the corresponding trust score (threshold) that should be used to partition friends into trusted/not trusted groups.

- Finally, TCT suggests a list of trusted friends that are eligible to read any future location revealing posts posted by the user.

As illustrated in Figure 29, the TCT invokes two different tools:

- TransferPostData: we developed this tool to read the collected interaction data set from the Excel file and import the data into the TCT's database. Figure 30 provides a snapshot of tool execution on one of the participant's data.
- Detection Tool: the TranfearPostdata invokes directly the detection tool to determine whether the post is location revealing or not. Figure 31 provides a snapshot of the invocation of the detection tool on the collected posts.

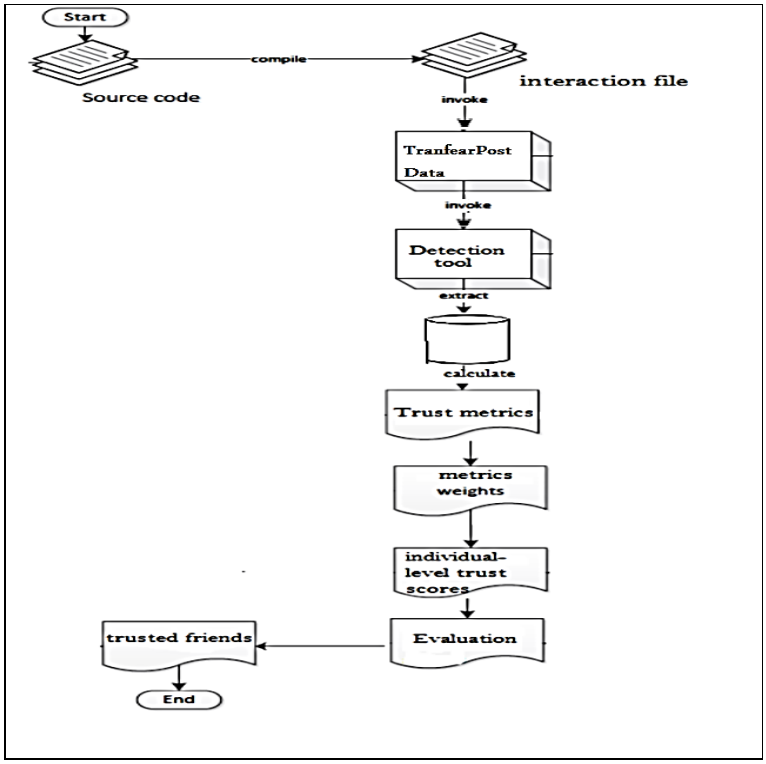

Figure 29. High Level Description of TCT Tool.

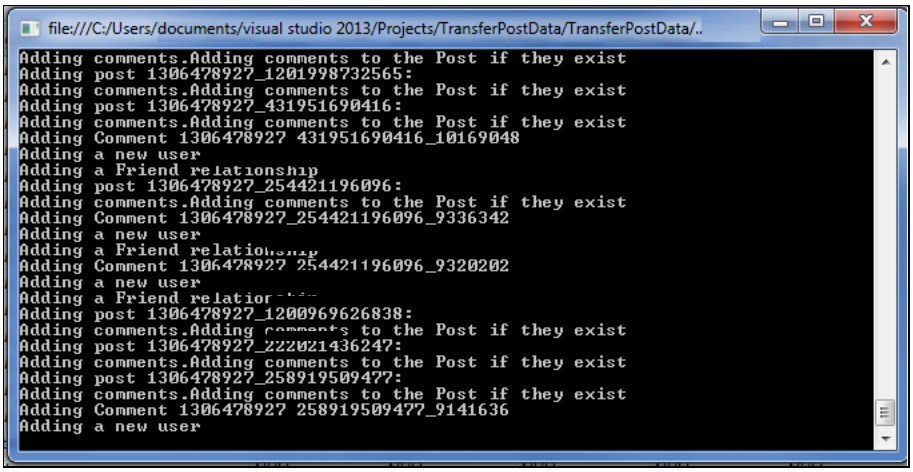

Figure 30. A Snapshot of the TransferPostData Tool Running.

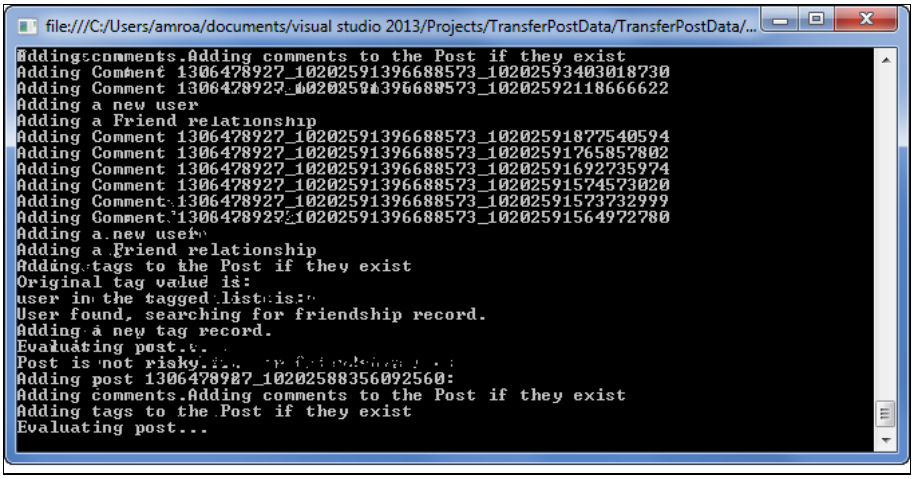

Figure 31. A Snapshot of the Detection Tool Running to Evaluated if the Post is Dangerous or Not.

Database design

The **TCT** tool extracts all friends interaction records on the user's posts and saves the result in the database. Figure 32 shows the database design of the **TCT** tool. When extracting and importing post information into our database, we only consider posts that are made by the profile owner. In other words, we do not consider posts made by friends on the user's wall.

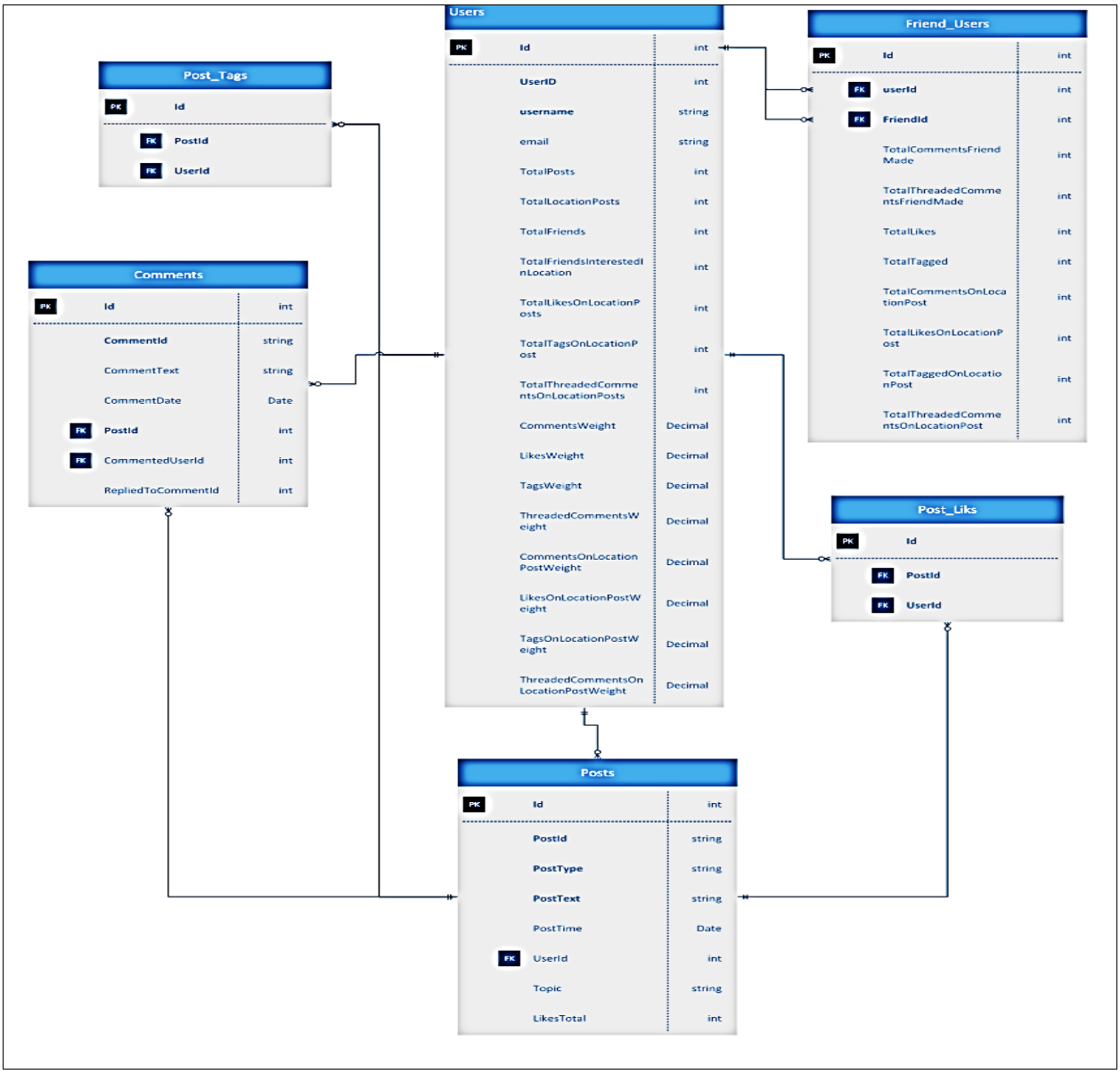

Figure 32. The TCT Database Design.

Class diagram

After extracting all of the interaction records of friends, TCT computes the trust metrics, namely, comments, likes, tags, and threaded comments. Then the TCT tool computes a weight for each metric. Finally, TCT suggests a threshold (trust score) to be used to separate trusted friends from non-trusted friends. The class diagram of the TCT is displayed in Figure 33.

The TCT implementation program consists of four classes and the following major functions:

- TotalPosts(): this method calculates the total number of text posts made by the user.
- TotalFriends(): this method calculates the total number of friends who interacted with the user.
- AverageCommentsRate(): this method calculated the average rate of the comment metric.
- AverageLikesRate(): This method calculates the average rate of the like metric.
- AverageTagsRate(): This method calculates the average rate of the tag metric.
- AverageThreadedCommentsRate(): this method calculates the rate of the threaded comments metric.
- The same aforementioned methods are repeated for the location revealing posts.
- TotalCommentsFriendMade(): This method calculates the total number of comments per friend.
- TotalThreadedCommentsFriendMade(): This method calculates the total number of threaded comments per friend.
- TotalLikes(): This method calculate the total number of likes per friend
- TotalTags(): This method calculate total number of tags per friend.
- CommentsRate(): This method calculates the average number of comments made by user's friends.
- LikesRate(): this method calculates the average number of likes made by user's friends.
- TagsRate(): this method calculates the average number of tags.

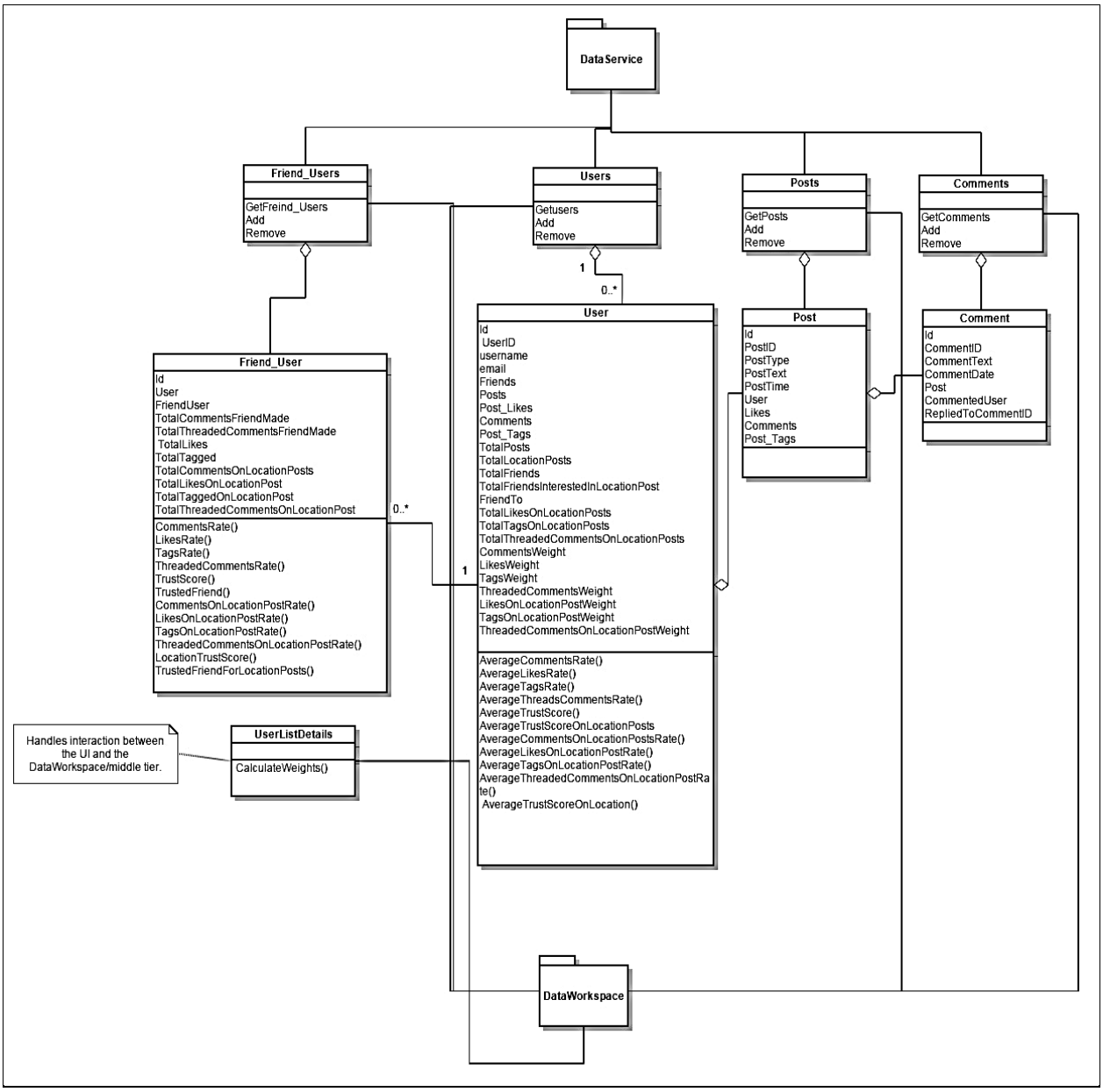

Figure 33. TCT Class Diagram.

Graphical User Interface (GUI)

This section provides a description for the TCT Graphical User Interface (GUI). The GUI of the TCT output interaction information is in a two section user interface, namely, users section, and friends section. An example of the TCT tool output is provided in Figure 34.

*Users section***.** This section provides information about the users and their interaction information such as:

- User ID: a unique identifier for each user.
- Total number of text posts: this column displays the total number of text posts made by the user.
- Total Location posts: this column displays the total number of location revealing posts that were made by the user. These posts are identified by using the detection tool from phase one, and these location posts are a subset of the total text posts made by the user.
- Total Friends: this column displays the total number of friends who appeared in our collected interaction data set. This number reveals the total number of friends who showed signs of interaction with the user using any of the listed interaction methods, namely, comments, replies and tags.
- Total Friends Interested in Location posts: this column displays the total number of friends who interacted with the user on the detected location revealing posts.

Friends section. This section provides information about the user's friends and their interaction information as follows:

- Friend ID: is a unique identifier for each friend in our database
- Total Comments Friend Made: total number of comments made by a friend on the user's posts
- Total Threaded Comments Friend Made: total number of threaded comments made by a friend on the user's posts.
- Total tagged: total number of tags that the user made for a friend.

- Total Comments on Location posts: total number of comments a friend made on the location revealing posts.
- Total tagged on Location: total number of tags made for a friend on the location revealing posts.
- Total Threaded Comments on Location: total number of threaded comments a friend made on the location revealing posts.

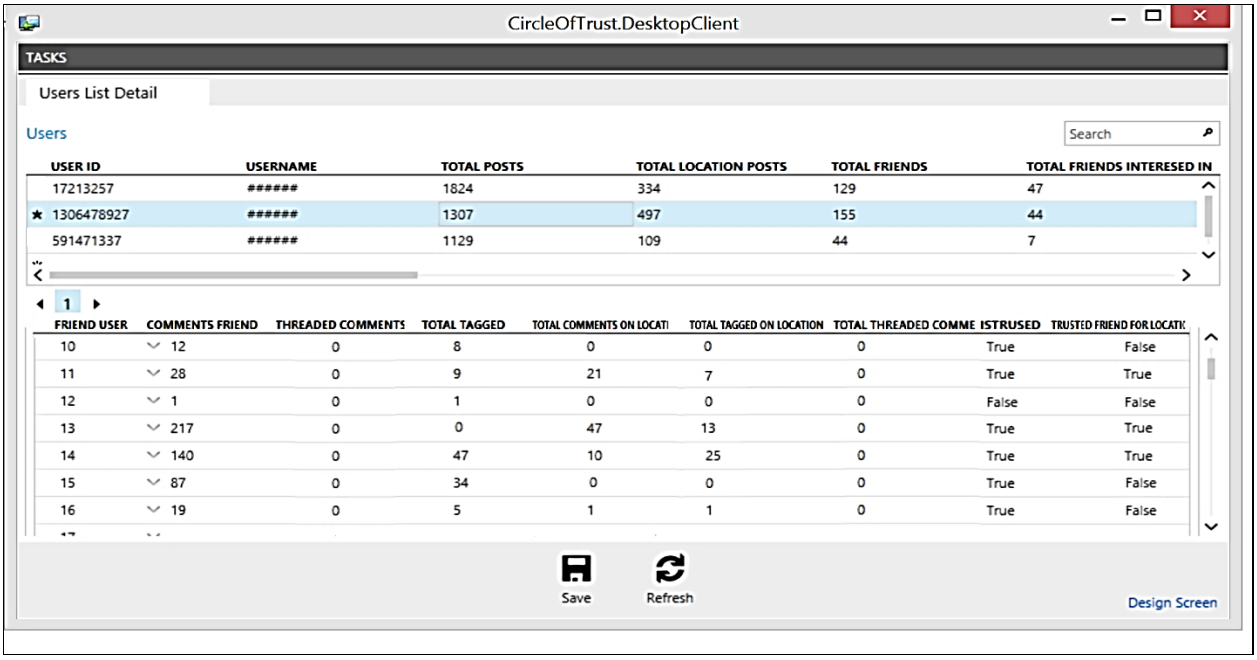

Figure 34. A Snapshot of the TCT Output Results.

Demonstration

The goal of this section was to illustrate the proposed tool using a simple example. Our objectives were to (1) warn the user about dangerous information disclosure in the text post; and (2) reduce the burden of configuring privacy settings of dynamic data on OSNs. Therefore, our goal was to make the process as efficient and easy as possible. This section walks us through the privacy steps that a social network user will go through when trying to broadcast a status update.

An assumption was make here that the user interaction information with friends was pre-collected and used to identify trust friends.

Suppose the user has successfully logged into the social network and wants to share a status update:

Step 1. Suppose that user wrote the following message: “Going to the game tonight!” in the status update section.

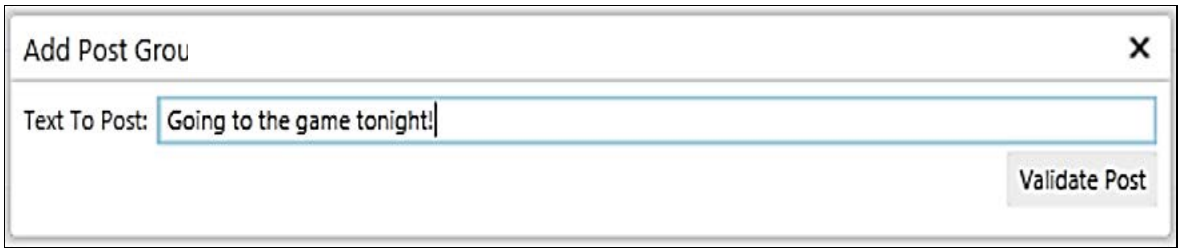

Figure 35. A Snapshot of Text Post Input on the Social Network.

After the user finishes writing the text and before the text is actually posted on the social network, the proposed tool will give the user the opportunity to validate the text against any dangerous information disclosure in the text post. Therefore, the user has two options:

- Post: the post option is not showing in the dialog box above. However, the post option will allow the user to post the text according to his/her predefined privacy setting on the social network. In this case, the text will not run through our detection and validation tool.
- Validate post: the validate post will allow the user to check the text to be posted on the social network against dangerous information disclosure. When the user clicks this option, the detection tool will be invoked to evaluate whether the text contains any dangerous information that might compromise the safety of the poster.

Step 2. Under the assumption that the user chose to validate the text post, the detection tool will run. Notice the text indicates that the user is planning to attend the game tonight. Therefore, according to our tool, this text should be detected as a dangerous (risky) message to be posted in an open environment such as a social network. This post might place the poster at risk of being attacked, robbed etc.

After running the text through the detection tool, the following message will appear to the user to warn him/her about dangerous information being detected in the text post:

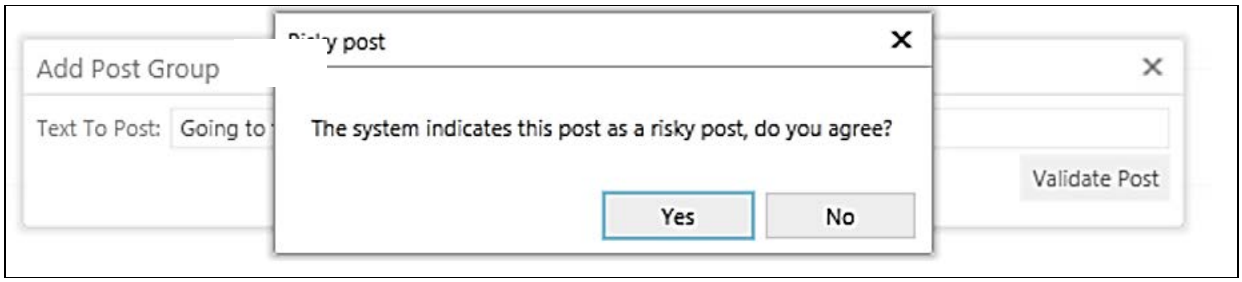

Figure 36. A Snapshot of the Warning Message.

The user has two options:

- Yes: this option indicates that the user agrees with the system that the post contains dangerous information. If the user clicks “yes,” the system will direct the user to Step 3.
- No: this option means that the post was mistakenly identified as a dangerous post while it actually is not. Therefore, if the user chooses this option the system will automatically direct the user to the first dialog box where the user can proceed to post the text message on the social network according to his/her predefined privacy settings.

Step 3. As a countermeasure against revealing such dangerous information to all friends on the social network, our tool suggests that the user limit the visibility of this information to a subset of predefined trusted friends. We believe that this solution is a good trade-off between users' engagement and the privacy wish of users to have more control over their data on online social networks. Therefore, the message in Figure 37 will be displayed to the user.

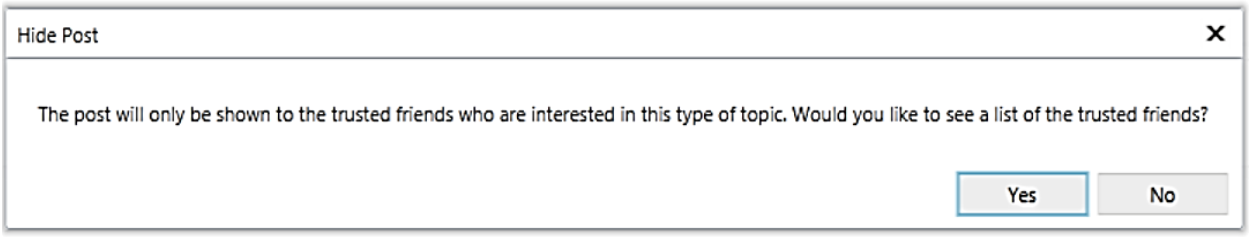

Figure 37. A Snapshot of the Suggestion Message to Limit Information Disclosure to Trusted Friends.

The user has two options:

- Yes: this option indicates that the user wants to see the list of trusted friends that was automatically identified by the system. Therefore clicking this option will direct the user to Step 4.
- No: this option indicates that the user does not want to see the automatically identified list of trusted friends. Clicking this option will automatically broadcast the dangerous post to the group of trusted friends.

Step 4. The goal of this step is to allow the user to see the automatically defined group of trusted friends. The user might want to adjust this list such as adding or removing a specific friend from the trusted group. Therefore, under the assumption that the user chose yes from Step 3, the group of trusted friends will be displayed as in Figure 38.

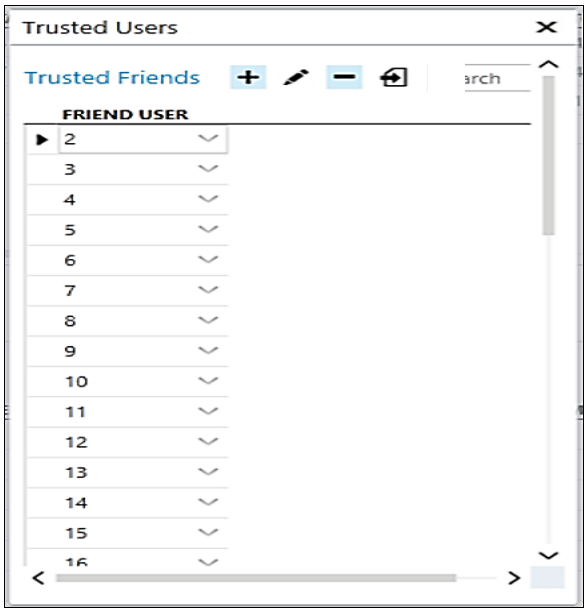

Figure 38. A Snapshot of Trusted Friends Group.

The user has the following options:

- Add a friend to the trusted group
- Remove a friend from the trusted group
- Quick search for a specific friend in the list

After making the desired changes the user can save the changes and then click OK to broadcast the post to the trusted group.

CHAPTER 4. EXPERIMENTAL EVALUATION

In order to evaluate and asses our approach, we carried out a sequence of experiments using a real Facebook data set gathered from participants in our study. Table 10 shows a summary of the data set used to evaluate our two phase approach, namely, the Awareness System and the Circles of Trust. Our purpose here was threefold: (1) show the accuracy of our Awareness System in detecting the time and/or location of users' activities and social plans from text posts; (2) show that friends' interaction could be triggered by post content; and (3) examine the accuracy of our Circles of Trust is predicting the group of user's trusted friends.

In this chapter, we:

- Describe the data we used in our evaluation
- Present the results of the proposed approach
- Discuss the results

Data Set under Test

Much of the challenge in undertaking this research was during the data collection process. After obtaining approval from the **I**nstitutional **R**eview **B**oard (**IRB)** for the protection of human participants in research, we recruited participants from all over the United States who were at least 18 years old. Upon signing up for the study and giving informed consent, participants authorized us to collect their Facebook activity data. To protect the participants' privacy, any information that could be used to identify the participants were deleted from the data set. Moreover, all demographic information was deleted permanently from the collected data set before storing the data set in the database. Unlike other standard Facebook applications, the application that we used to collect the data required neither passwords nor user names from

participants. For privacy protection, participants and their Facebook friends were assigned unique IDs, and only these IDs are kept as an identifier in the extracted data set.

Table 10.

Quantities of Facebook Text Status Updates Collected from Participants.

Item	Count
Text status updates	16,071

Members (participants) were selected randomly. We believed that this had the advantage of observing our models in different settings. For example, teenagers post more frequently on Facebook compared to old adults. Moreover, teenagers use brief and revealing text messages while, older adults tend to write longer messages. In addition, teenagers and young adults have more social connections than older adults; therefore, more interactions with their friends. Our data set did not contain information on whether an interaction was positive or negative. Therefore, we assumed all interactions were positive. For example, a comment might indicate a positive interaction between the user and a friend and might indicate a negative interaction. However, classifying the type of interaction was out of the scope of this study. We assumed that negative interactions will result in reduction in the number of interactions from that friend on the content of the user over time, and therefore will result in excluding the friend from the list of trusted people who can read future posts on this content.

Limitations

Our measurement methodology had the following limitations:

- In Facebook, users can update their status in many ways (e.g., text messages, link, photo, and video). While text messages are a popular method for updating status on

social networking sites in general, such as on Twitter, we do not know whether it is representative of other forms of location revealing information. For example, a user may share a video of himself on a Disney cruise, which would be considered revealing information of the user's current location. However, our approach cannot detect this type of location revealing information.

- In Facebook, users can interact in many ways (e.g., private messages, photo tagging, applications, closed groups, and chatting). While wall posting is one of the most popular methods of user interactions with friends, we did not know if it was representative of other forms of interaction which may affect the trust value. For example, a friend may continuously interact with the user through the chat or the messaging services and still show no explicit signs of interaction on the user's status updates. However, our approach focused mainly on explicit interaction methods on OSNs.
- Our data set was limited to the subset of people who were willing to participate in our study. Our data set was based on a single social network; Facebook; however, a recent study has demonstrated similarity across multiple social networks. That gave us confidence that our findings to some extent might be generalized to other similar networks such as Google+ or Twitter.
- The collected interaction data sets do not include a "Date Joined" field, nor a "Friendship Date" for when the friendship was established. We did not have access to such data, and we could not make an estimation of these dates from the collected data. Therefore, the fairness of compression between newly added friends and previously added friends was not discussed in this dissertation. For example,

previously added friends had more opportunities to interact with the user due to the fact that they have established the friendship on Facebook for a longer period of time.

- We collected "tags" and "replies" data associated with each post. These are another form of friends' explicit interactions on Facebook and may give insights into friends' trustworthiness. However, "replies" were recent additions to Facebook. For example, the "replies" feature was added to Facebook in March 2013 and the data was collected in 2014. Therefore, "replies" data was not sufficient and therefore was excluded from our experiments.
- The data collected through NCapture (QSR, 2015) does not have information on the likers (individuals who liked specific posts). It has only aggregated data on how much likes each post has gained. Due to limited resources and manpower constraints, performing manual collection to identify likers of each post was not feasible; therefore, we focused solely on the number of likes per post, with a vision to expand our analysis in the near future to include likers.
- Finally, the number of posts that we can capture was limited and was determined by Facebook. The exact number of posts collected from each participant was controlled by Facebook through mechanism unknown to us and it may vary depending on:
    - Number of available posts.
    - The privacy setting of the posting user.

Experiments and Results

In this section, we report on the experiments conducted on the data retrieved from Facebook social network.

Experiments on the Awareness System

In the first part of our experiments, we aimed to show the accuracy of our detection rules in term of a true/false detection rate. As mentioned earlier in Chapter 3, the Awareness System tool employs these rules to detect any time and/or location that would reveal information about a user's activities and/or social plans from a text message.

The system accepts a text message as an input, the text message could be a combination of one or more words and it could also include symbols and emotional icons. If the tool detects any revealing information, then it asks the user to confirm that the post is indeed dangerous. If the user confirms that the post is dangerous, then the system will increment the counter of the detected dangerous post. Otherwise, the counter of "false alarm" will be incremented. If the system fails to detect a post that contains dangerous information, then the undetected "dangerous posts" counter will be incremented. By following this supervised method, we were able to know the number of revealing posts, number of detected posts, number of undetected posts, and the number of false alarms in order to evaluate the accuracy of our detection approach. Table 11 summarizes our findings on both text status updates and check-in messages.

As we can see, 19% of the collected posts had information about the user's current and/or future location, activities or social plans. For example: travel plans, summer vacation destination, the time of visiting family or friends, and some other social plans like attending a soccer game, visiting the movie theater, hiking trips, shopping plans etc. Moreover, 3% of the collected information were also under the check-in category, which was explained earlier in Chapter 3. Our system was able to detect 85% of these dangerous posts which outperforms another existing approach that has a 44% detection rate.

Table 11.

Awareness System True/False Detection Rate.

Dangerous Messages Percentage	**Detection Percentage**	**False Alarms Percentage**	**Not Detected Percentage**
19%	~85%	~11%	5%

Check-in messages percentage	**Detection Percentage**	**False Alarms Percentage**	**Not Detected Percentage**
3%	100%	0	0

Table 12 shows samples of anonymized text messages taken from our data set that were detected by our system; the applied detection rule is bold-italic print:

The system makes occasional mistakes by identifying some messages as dangerous while they are actually not. For example, “in Finland we say: happy friends’ day.” The system detected this post as dangerous according to the rule number 10, which is the proposition ‘in’ followed by ‘proper noun.’ Although, this post fulfilled one of the detection rules employed by the system, the post was actually not dangerous as it did not reveal any location information about the user or anyone else. Table 13 shows few anonymized text messages taken from our data set that were mistakenly detected by our system. The rule used in detecting dangerous information is bold-italic print. Text posts that generated false alarms were mostly characterized as long conversational style sentences.

Table 12.

Samples of Dangerous Posts Detected by Our System.

Samples of Detected Status Updates
On my way to***/TO*** Twin***/NNP*** Falls for a lovely morning hike! :-)
Having lots of great times with family and friends at***/IN*** the***/DT*** lake***/NN***
Is **@** home with cold
Enjoying some tasty Ramen with my awesome sister before we ***head*** downtown!***/NN*** Yummy!
Time to ***head*** home***/NN*** after an awesome weekend
is on his way back to the Airport to ***head*** home***/NN***
I will ***head*** out***/RP*** and get me some Mega Millions tickets
In***/IN*** Lincoln***/NNP***, Nebraska
Sarah and I are in***/IN*** Portland***/NNP***, scoping out the books at Powell's!
is in***/IN*** Vermont***/NNP*** visiting family
Tonight I will be at***/IN*** Fargo***/NNP*** Civic Center
Eating all natural dinner at***/IN*** Canteen***/NNP*** with friends
With Carla & Linda at***/IN*** the***/DT*** mall***/NN***
On a road trip to***/TO*** Indiana***/NNP*** … life is beautiful
With Daniel at***/IN*** the***/DT*** Matrix***/NNP*** class :-RRB- in***/IN*** Omaha***/NNP*** ,/,
On my way home. It was great seeing family and friends after not seeing many of them for almost a year. Hope everyone had a good Christmas and hope you all have a Happy New Year!
On my way to sunny LA…

Table 13.

Samples of Mistakenly Detected Posts by Our System.

Samples of Mistakenly Detected Non Dangerous Status Updates
In***/IN*** Finland***/NNP*** we say: 'Happy Friend's Day'. This is to all of my wonderful friends and friendships all around the world. You all mean a lot to me. Thank you for your friendship. Miss you, friends.
Out. Of. Control. US All must speak out. Moms Demand Action for Gun Sense in***/IN*** America***/NNP*** needs dads, grand moms and granddads saying loud and strong: this is madness.
another reason why I do not shop at***/IN*** Hobby***/NNP*** Lobby***/NNP***

On the other hand, the system failed to detect 5% of the dangerous posts. These posts actually included dangerous information were the user activities and social plans were revealed. These posts where claimed as dangerous by the user at the time of posting. For example, in this post: "five more days and Vermont is calling," we can infer that the poster (user) may go to Vermont anytime during the coming five days. However, with the current set of detection rules the system was not able to identify this as a dangerous post. Table 14 shows samples of anonymized text messages that were not detected by our system. An inference of the user's current location, future plans, vacation destination, trip planning, and more can be made from reading each of these messages. However, our system currently does not have the capability to detect these dangerous posts.

We strongly believe that detecting similar text messages will be a challenge. In our approach, we counted on the time and location prepositions to detect temporal data. The general characteristics of undetected dangerous posts are (1) the absence of the time and location propositions and (2) short, brief and informative.

Table 14.

Samples of Dangerous Posts that Our System Failed to Detect.

Samples of Undetected Dangerous Status Updates
Group is ready to race the kid's marathon tomorrow.
Five more days and Vermont calling
Walking 'round the bend north of Lindenwold with the spaniels
And we're off! Sam and Sandy's...and a cast of dozens...Excellent Adventure. Brace yourself, Italy, here we come Italy!
Packing and getting ready for a long day traveling home!!!
Today I will go have a berry chill
France here we come!

Experiments on the Circles of Trust

Although users on OSNs generate a massive amount of text status updates every day, the relationship between topic "Location" and the audiences' responsiveness have not been studied before. Therefore, our work on analyzing the collected interaction data set began with the goal to understand this relationship. As of 2014 (the time of the data collection), Facebook was the largest most studied social network. Therefore, we acquired interaction data from this platform, and later we proceeded to do our analysis. The data set has a total of 16,000 posts which include 4,859 location revealing status updates and their corresponding likes, comments, tags, and replies.

In this section, we start by providing insights into the data used. Then we explain the experiments we ran and discuss the results.

The analysis consists of:

- High-level characteristics of collected data:
    - Comparison between social links and active links
    - Distribution of interaction methods per user
- Impact of the location topic on friends interaction

Our findings have the following implication: the Location topic evokes interaction from certain audiences, while the interaction from other friends on the user's friends list may be poorer. Therefore, we believe that privacy control for dynamic data on OSNs that limits information disclosure to this subset of friends will be an effective approach to safe communication on social networks without compromising users' safety and privacy.

Data used

This section explains the collection process and the structure of the collected interactive data set. We collected a set of interaction records between users and friends. To be able to collect the data set, we used NCapture, which is a free web browser extension, developed by QSR (2015). It can be used to capture web-page content. For example, posts, comments etc. Each interaction data set contains Facebook wall posts, associated with the profile owner, and their related information such as commenters, and tagged friends etc. Currently and due to the recent changes that Facebook has made on its API, NCapture lost the ability to collect Facebook posts as a data set.

Table 15 provides a simplified example of an anonymized data set containing Facebook data. The first row contains a post made by the profile owner "Mike Jones" and the next three rows are comments made by the owner's friends on that post, while the last row contains a post that has no comments.

Table 15.

Simplified Data Set of Facebook Posts and Their Related Comments.

Posted by Username	Post	Commenter Username	Comment Text
Mike Jones	I'm heading to a workshop on rainwater tank installation.		
		Mary Smith	You'll have to tell me about it afterwards.
		Carlos Garcia	I've been thinking of installing one too.
		Mike Jones	Great workshop. That's my next project.
Mike Jones	I've ordered a tank. It's getting delivered next week.		

To be able to understand the thread of the conversation and the context of the reply, we include the "replies to comments" column. Figure 39 illustrates a Facebook one post (1) and four comments (2, 3, 4, and 5). There is also a threaded comment between (1) and (2), as (1) replied to comment (2) and then (2) replied back to comment (1).

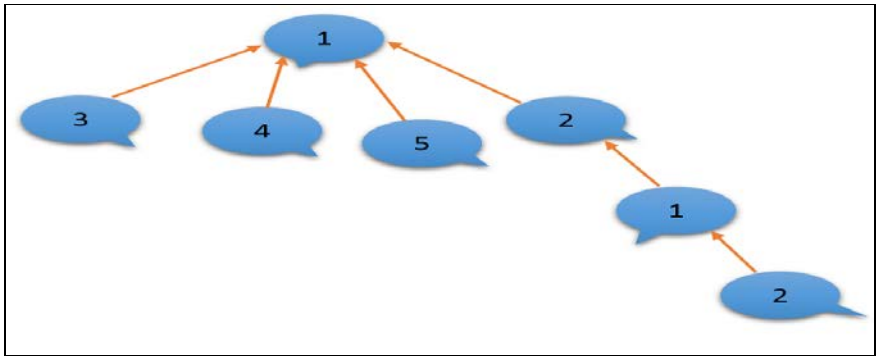

Figure 39. An Illustration of Threaded Comments on Facebook.

In our data set, we have an **In Reply To ID** column that will link back to the reply to the comment that it was replying to. Table 16 is a simplified data set that corresponds to the example in Figure 34.

Table 16.

A Simplified Data Set of Hieratical Threaded Facebook Comments.

Post ID	Posted by username	Comment ID	Commenter Username	In Reply ID
130-76931	1			
		130-76931-052	2	
		130-76931-054	3	
		130-76931-242	4	
		130-76931-312	5	
		130-76931-448	1	130-76931-052
		130-76931-003	2	130-76931-448

Figure 40 provides a snapshot of an anonymized input interaction data set for one of the participants in our research.

In our data set, the collected data are of four types:

- Video
- Link
- Status update (text)
- Check-in

Row ID	Post ID	Posted By	Tagged	Type	Likes	Comment ID	Commenter Username	In Reply To ID	Comment Likes
254	1306478927_10202492308011418	#####	#####,#####,#####,#####,#####	status	60				
255	1306478927_10202492308011418					1306478927_10202492308011418_10202294635663984	#####		0
256	1306478927_10202492308011418					1306478927_10202492308011418_10202294689023818	#####		1
257	1306478927_10202492308011418					1306478927_10202492308011418_10202294675903490	#####		2
258	1306478927_10202492308011418					1306478927_10202492308011418_10202291820232100	#####		2
259	1306478927_10202492308011418					1306478927_10202492308011418_10202287641647638	#####		1
260	1306478927_10202492308011418					1306478927_10202492308011418_10202287052112900	#####		1
261	1306478927_10202492308011418					1306478927_10202492308011418_10202286734544961	#####		1
262	1306478927_10202492308011418					1306478927_10202492308011418_10202286664543211	#####		1
263	1306478927_10202492308011418					1306478927_10202492308011418_10202286438219053	#####		1
264	1306478927_10202492308011418					1306478927_10202492308011418_10202286209891845	#####		1
265	1306478927_10202492308011418					1306478927_10202492308011418_10202285939405083	#####		1
266	1306478927_10202492308011418					1306478927_10202492308011418_10202285908244304	######		1
267	1306478927_10202492308011418					1306478927_10202492308011418_10202285882323656	######		1
268	1306478927_10202492308011418					1306478927_10202492308011418_10202285855482985	######		1
269	1306478927_10202492308011418					1306478927_10202492308011418_10202285835322481	######		4
270	1306478927_10202492308011418					1306478927_10202492308011418_10202285809001823	######		0
271	1306478927_10202492308011418					1306478927_10202492308011418_10202285794601463	######		1
272	1306478927_10202492308011418					1306478927_10202492308011418_10202285733759942	######		2
273	1306478927_10202492308011418					1306478927_10202492308011418_10202285730799868	######		2
274	1306478927_10202492308011418					1306478927_10202492308011418_10202285711599388	######		1
275	1306478927_10202492308011418					1306478927_10202492308011418_10202285697319031	######		3
276	1306478927_10202492308011418					1306478927_10202492308011418_10202285689878845	######		2
277	1306478927_10202489561142748	#####		status	10				
278	1306478927_10202489517981669	#####	#####, #####	status	8				
279	1306478927_10202489051810015	#####		status	14				
280	1306478927_10202459646474900	#####		status	2				
281	1306478927_10202459646474900					1306478927_10202459646474900_10202459744397348	#####		0
282	1306478927_10202459646474900					1306478927_10202459646474900_10202459717076665	#####		0

Figure 40. A Simplified and Anonymized Snapshot of an Input Interaction Data Set.

While the video and links types were out of the scope of this dissertation, the Check In datatype was considered in our approach. We first explained the check in feature on Facebook, and then explained how this type of data was considered in our approach.

In 2010, Facebook announced the launch of Facebook places services, well-known as Check-in. This service allows smart phone users to use GPS on their phone and let their friends know exactly where they are. Users who want to announce their location to their friends can click the “check in” button to see a list of nearby places, then the user can choose the place that matches where they are. Besides, this service will allow the user to see if any other friends are currently at the same place. This service also allows users to “tag” any Facebook friends who are with them. After the user “Checks in,” Facebook will announce the user’s location in the

friends' News Feed on Facebook. Using this service will allow all the users' social links to know the current location and place of the user.

Based on the scope of our work this data is also considered dangerous as it reveals the time and location of a user's activity. Therefore, we also considered collecting this data in our work. Table 17 shows the difference between a text status update and a check-in status update. Where the username is the actual username of the profile owner.

Table 17.

Difference between Text Statuses Update and Check-in Data in Our Data Set.

Status update	**Check-in**
At the mall shopping	***Username Check in at*** West Acers Mall
Doing some shopping at the bike store	***Username Check in at*** Velco Cult Bike Shop
Going with my kids on a hiking trip	***Username Check in at*** Multnomah Falls
Watching Iron Man at the theater	***Username Check in at*** Marcus Theater
I'm having a nice meal with my friends at the restaurant	***Username Check in at*** Elliott's Oyster House.

Table 18 provides a detailed description of the data set fields, where each file in our data set represents a user's posts, friends, and the interaction between them.

Table 18.

Description of Data Fields in the Data Set.

Name	Description	Type
RowID	Newly generated ID for each row in the data set	Int32
UserID	ID of the user who made the post	String
PostID	The post ID. It is a dash separated string that has the following style, *UserID-ID where:* *UserID is a the user of the post owner* *ID: is a newly created ID for each post*	String
Post Text	The main body of the message in the post	String
CommentID	The comment ID. It is a dash separated string that has the following style *PostID-ID where:* *PostID: corresponds to the PostID for which the comment is made on* *ID: is a newly generated ID for each post*	String
Commenter ID	ID of the user who made the comment	String
Tagged	Comma-separated list of users who were tagged in a specific post	cvs [String]
In Reply ID	This filed is used to identify threaded comments and it has the following style, *CommentID-ID where:* CommentID: *corresponds to the CommentID for which the comment is made on* *ID: newly generated ID for each reply*	
Likes	Total number of likes that a post has received	Int
Comments	Total number of comments that a post has received	Int

High-level characteristics of data

We first present two high-level characteristics of the collected data set. The analysis presented in this section consists of:

- Comparison between the size of social links and active links
- Distribution of interaction methods

Social links vs. active links. First, we examine the difference in size between the friends list and list of active friends who showed signs of interaction with the user in our data set. Using the collected data sets, we constructed the list of active friends for each user. A friend is considered active if he/she interacted with the user at least once using comments, tags, or replies. We performed this analysis to prove our hypothesis that users on Facebook only interact with a small subset of their social links (friends), opposed to the null hypothesis that assumes users interact with all their social links. Therefore, a privacy management that limits information disclosure to a subset of active friends will not compromise the enjoyment of social network.

In Figure 41, we show for each user:

- Total number of friends: this is the number of all friends on each user's friends list. This number is obtained from the users' profile information on Facebook.
- Total number of active links: friends who interacted with the user on text posts at least once.

This analysis shows that users interact only with a small percentage of their friends when they post text posts; 33% based on the data in our data set. Therefore, we can conclude that even though users in our data set have established an average of 309 social links (friend relationship), they only interacted with a subset of these friends. Our findings confirm the findings from

previous research (Sterling, 2013) that have shown that the size of the activity network is significantly smaller than the size of the social network.

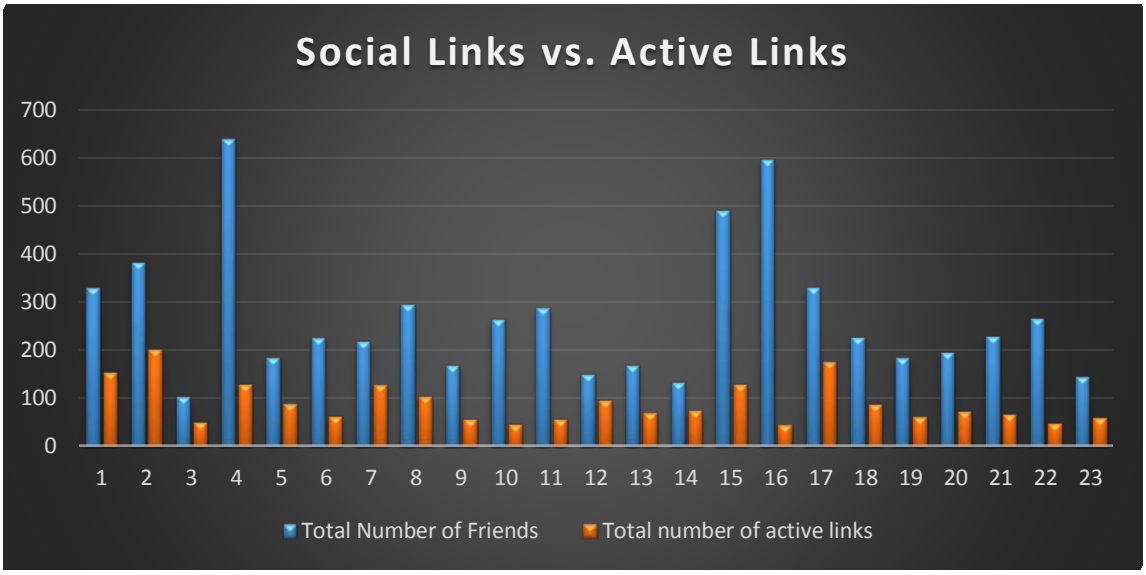

Figure 41. A Comparison between Social Links and Active Links.

Distribution of interaction methods. Second, we have also examined the difference between friends' engagement in each interaction method. On Facebook users' wall posts, friends are able to interact through different types of interaction methods such as like, comments, or replies. We analyzed the data sets to understand friends' engagement in the aforementioned interaction methods. The null hypothesis assumes that all interaction methods effects are equal when determining the trustworthiness of friends, against the alternative hypothesis that the interaction methods effects are not all equal when determining trusted friends. Figure 42 shows friends' interaction rate on text post.

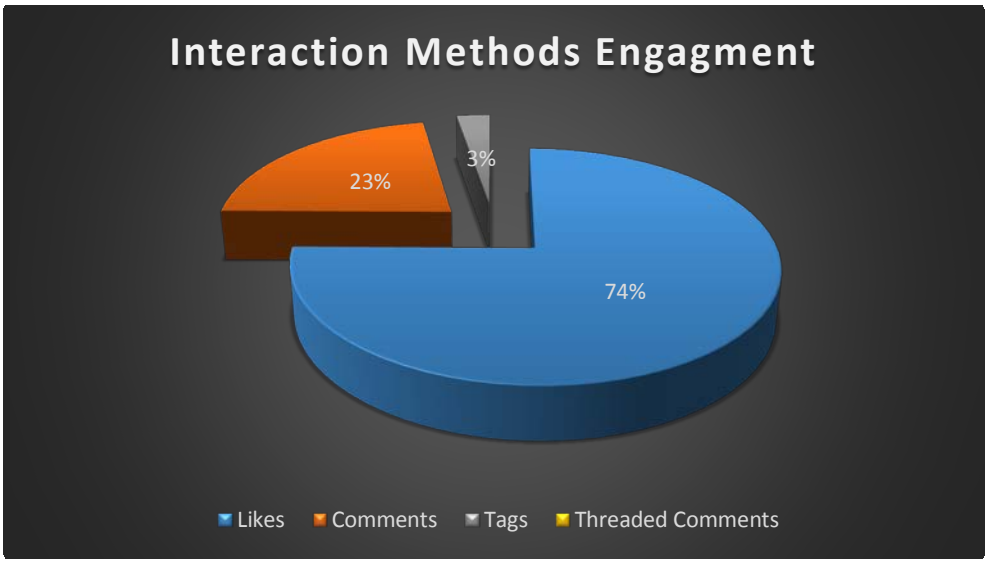

Figure 42. Interaction Methods Engagement Percentage.

We observed a clear common pattern among users' friends:

- Friends are more engaged in liking user's text posts than commenting on posts.
- Friends are more engaged in commenting on posts than getting involved in a threaded comments conversation.
- Friends are more engaged in tagging than in getting involved in a threaded comments conversation.

Therefore, we concluded that like was the most adapted explicit interaction method between Facebook users in our data set. Like on OSNs becomes a form of confirmation from a friend that he/she has seen the status updates of the user. On the other hand, comment is still the most adapted form of explicit interaction between users on Facebook right after like. While the percentage of users' engagement in comments are less than likes, we believe that comments are more valuable than like when determining trusted friends. The interesting fact about comment is that it reflects the opinion of the commenter. Moreover, comments could be studied and analyzed to suggest positive or negative interaction between users.

The percentage of tags compared to likes and comments is significantly low. This is due to the fact that the use of tagging in Facebook is more often associated with photo status updates. In our study, we only focused on text status updates. However, tagging a friend in a location revealing post, or check-in status update could be interpreted in a different way to reflect the existence of the tagged friend and the user at same event or location or their plans to be in the same location in the future. As we mentioned earlier in this section, reply/threaded comment is a new feature that was added to Facebook shortly before the collection of our data set; therefore, replies data in our data set were not conclusive.

Location topic and friends' interaction

First, in this analysis we examined the difference in size between the list of active friends and the list of friends who interacted with the user on the location topic. Using the collected data sets we constructed both lists. A friend is considered active if he/she interacted with the user at least once using comments, tags, or replies. While a friend is consider an active user on the location topic if he/she interacted with the user on the location topic at least once using the aforementioned interaction methods.

We performed this analysis to prove our hypothesis that the location topic triggers the response from certain active links, opposed to the null hypothesis that all active links interact with the user equally on all topics. Figure 43 shows the total number of active links against the total number of friends who interacted with the user on the location revealing posts.

The analysis shows that the location revealing posts evoke a response from a small percentage of active links on the user's social network. Based on our collected data set, only 19% of active links interacted with the user on location revealing posts. On the other hand, 81% of active links did not show signs of interactions with the user on location posts. This could be an

indicator that not all active friends are interested in day-to-day activities and social plans that the user likes to talk about in his/her status update. Moreover, this finding could mean that restricting location information disclosure to the subset of friends who interact with the user on this topic might not compromise the enjoyment of social network.

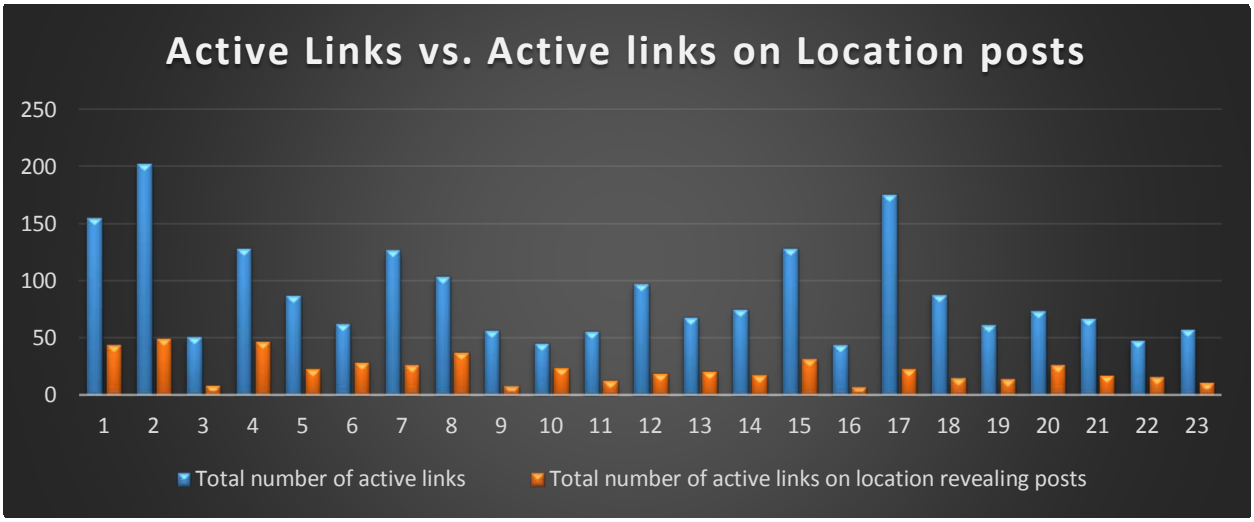

Figure 43. A Comparison between Active Links and Active Links on Location Revealing Posts.

Evaluating the Circles of Trust Prediction Rate

The goal of our experiment was to test the ability of our proposed approach to use a user's existing interaction data set to predict his/her future friends' interaction. We only choose to perform this evaluation using the comment metric due to insufficient data on other interaction methods in our data set. We evaluated our approach in terms of its error rate (the percentage of friends classified incorrectly).

We used a training data set for each participant to build the initial model. The system accepted the user's interaction data set and output a list of friends who we predicted would interact with the user. Thus, the problem remains is how accurate our model is in predicting friends who will explicitly interact with the user on location revealing content.

Methodology

The first option was to evaluate our approach on the training data set. However, this was generally not a good idea since we built our model based on this data set. The problem with this approach was, as is often the case, the approach will perform very well on this data set, but may not give an accurate account of how well the model would perform on unseen data.

To be able to evaluate how well the model would perform on new data, we needed to apply the model to such data. In doing so, we can then obtain the model's error rate. This proportion of our observations is where the model and the actual pre-known outcome differ. Therefore, our approach to this problem was to use a Validation data set to test the accuracy of our model. The training and the evaluation data sets have pre-known information about posts and commenters on the posts.

The justification that is often given in the literature is to divide the data to 80% training data and 20% test data to better model any underlying distribution. Therefore, we divided each data set into training and validation as follows:

- First, we ordered the data chronically from older to most recent posts.
- Second, we used the last 20% of each participant's posts as a validation data set while the other 80% were used to train the model.

We chose to order the data chronically and then use older interaction records as our training data set, under two assumptions: (1) most users' friendships have been established closer to the date of the first post appearance in the data set, and (2) friends who did not interact with the user in the past would not interact in the future.

For each participant, we compared the suggested list of trusted friends using the training data set against the pre-known list of friends who actually commented on the user's location

revealing posts in the evaluation data set. To evaluate the overall performance of our system on the evaluation data set, we used the overall error rate. The overall error rate and the approach accuracy were defined as follows:

- Overall error rate = (sum of misclassified records) / (total records).
- Accuracy = 1- error rate

Using the confusion matrix, the average system accuracy was 61%. While the average system error rate was 39%. We also had 3% of friends who did not appear in our training data set and made their first appearance in the evaluation data set.

We believe that this 3% newly appeared friends are a result of our approach in dividing the data set into training and evaluations. For example, this 3% could include friends who were recently added by the user on the friends list. And therefore, this could be their first attempt to interact with the user in the evaluation data set. Unfortunately, in our data set we did not have information about the date of friendship to be used in making such a conclusion. This number could also indicate the dynamics of the change of the interaction behavior of some friends. For example, some friends might have recently started being active on Facebook, or might have recently started interacting with the user. As these friends continue interacting with the user on the same topic, their interaction rate will get higher, and they will be included in the circle of trust. The opposite is also true, the absence of friends' interaction on the given topic would cause their exclusion from the circle of trust over time.

Threats to Validity

Although the results of our experiments were encouraging, there were multiple threats to validity that exist which could limit the generality of the results. Limitations of our study were listed earlier in this chapter. One very important threat, is the small size of our data set. Our data

set was limited to a small number of Facebook users who were willing to participate. Therefore, large scale assessment is needed in order to get conclusive results to generalize our findings. Second, in our study we only focused on a one type of status update that could be used to reveal users' location, which is text post. Therefore, we do not know if this type of status update is representative of other forms of location revealing information, such as, photos and links. Third, our evaluation methodology was limited to the comment interaction method only, which represent only a 23% of interactions distribution among friends. Unfortunately, our data set did not contain individual records for likers, neither was there enough data on tags and threaded comments. Forth, our evaluation methodology did not address the fairness of comparisons between the interaction behaviors of brand-new friends and previously added friends. For example, one could expect that previously added friends most likely interacted with the user more than newly added friends. That is due to the fact that they have had the opportunity to interact on more posts. Finally, the results of our Circles of Trust model were inconclusive when the data set did not contain active users and active friends.

CHAPTER 5. CONCLUSION AND FUTURE WORK

In this chapter we provide a conclusion of our findings and contributions, we then discuss how our proposed approach can be adapted to the major online social networks (OSNs) we have discussed in Chapter 1 of this dissertation. Finally, we discuss our future work.

Conclusion

OSNs are being used for broadcasting information among all generations. While these sites are a great place for people to stay connected and socially active, there are still privacy issues involving the broadcasting of dynamic data. Existing OSNs lack the privacy control needed to manage the broadcasting of this type of data.

In this work, we proposed an approach to manage the privacy of dynamic data on online social networks, such as Facebook. Our goals are to: (1) detect dangerous information from a user's text post that might reveal the location and/or time of the user's activities and social plans; and then (2) provide a list of trusted friends who might be interested in these posts and with whom the user can safely share these dangerous posts. Dangerous information that people post on OSNs can take multiple forms. However, in our work we focused mainly on location revealing information as these little tidbits of information can be used easily by criminals and stalkers to learn more about a person's patterns so they know where to find them. Therefore, it is best to limit the visibility of this information to only trusted relevant parties. Our approach is divided into two phases, namely, the Awareness System, and the Circles of Trust.

The first phase of our approach involves recognition and detection of key phrases in a post that might reveal the time and/or the location of the user's activities and social plans. Revealing this information to all friends on the social network might compromise the safety of

the user or someone else. To reiterate the objective of this problem, we used a combination of natural language parser and tagged-based detection rules. This is a better approach than employing a large knowledge base of pre-categorized words. This phase is divided into three steps. First, the natural language parser grammatically parses the plain text input (user text post). The result of this step is a part-of-speech tagged text. Second, the dangerous information extractor extracts dangerous information from the part-of-speech tagged text, if there is any, according to a provided set of detection rules. Third, the categorizer assigns the appropriate tag/topic category to the dangerous post. Tags allow the system to categorize the topic of the dangerous post and are used in our approach as a basic means for restricting dangerous information to only relevant trusted parties.

Based on experimenting with 16,000 Facebook text posts, we were able to develop a set of detection rules that could be used to detect the revelation of time and/or of users' activities and social plans in their text posts. To evaluate the effectiveness of our approach, we built the Awareness System to implement the proposed approach. The tool uses a combination of Stanford natural language parser and a developed detection rule. Our approach was tested on 16,000 Facebook text posts and has an 85% detection rate in term of true/false, 5% undetected dangerous posts. Therefore, our goal of detecting dangerous information that might reveal the time and/or location of users' activities and social plans has been researched. Our approach has a room for improvements which we plan on doing in future work.

The second phase of our approach aims to improve an existing trend of sharing dynamic information on social networks toward a trusted-friends paradigm. To reiterate the objective of the problem, when a user posts status updates that potentially has dangerous information, the proposed approach would suggest the user share these posts with only a set of trusted friends and

to whom the status updates are relevant. This is a better approach than broadcasting the updates to a user's entire friends list. Our approach in defining trusted friends is unique as it depends on (1) the content of dynamic information; and (2) the context of interaction between user and friends. This phase is divided into four steps. First, friends' interaction records are partitioned into topical groups. Second, a trust metric extractor pulls trust metrics from the categorized interaction records. Third, the trust rating calculator suggests weights for each trust metric and calculates individuals' trust scores. Fourth, the Circles of Trust generator suggest a threshold to be used in grouping trusted friends into the (Location circle of trust). The last step is to automatically generate a list of trusted friends who can safely see the detected dangerous post. The interaction-based trust metrics used in our work are defined and selected to cover different types of explicit interaction methods that people use to interact with each other on Facebook. We defined four interaction-based trust metrics, namely, likes metric, comments metric, tagging metric, and threaded comments/reply metric.

To test our hypothesis, we collected a set of interaction records that show users' friends, posts and interaction between them. We carried out some data analysis on the collected data, which revealed that: (1) the size of the active links is significantly lower than the size of social links; (2) interaction methods are not evenly weighted when determining trustworthiness; and (3) the location topic evokes interaction from certain active links. The results of the study do support the proposed hypothesis that friends' interactions on online social networks could be triggered by the content of the status updates. Therefore, when a user posts a status update, only a subset of the user's active friends can be selected to receive relevant updates.

Moreover, to evaluate the accuracy of our approach in suggesting trusted friends, we built the topic related Circles of Trust tool to calculate the interaction-based trust metrics

automatically and to implement the proposed approach. The output of the tool is a suggested list of trusted friends on the Location topic that we predict will interact with the user on future posts. Due to insufficient data in our collected data set on all suggested interaction methods and to accurately measure the performance of our system, we evaluated our approach on the comments metric. The sample data was further processed into two sets; training and evaluation. The training data set was used to do the initial assignment of friends into the Location Circle of Trust. The evaluation data set was used to evaluate the accuracy of our model in terms of error rate. Our approach has shown to have a reasonably acceptable error rate of 39% and accuracy of 61%. The results are encouraging. However, future studies need to be done on a larger data set to generalize our findings. Moreover, future studies are needed on the remaining suggested interaction methods.

How Our Approach Can Be Adapted to Other OSNs

While we used Facebook, the most popular online social network as our case study, our research efforts were abstract and generalized as much as possible so the solution approach could be applicable to other similar networks. In this section, we briefly discuss how our approach can be adapted to each of the major OSNs we discussed earlier in Chapter 1.

Detection Approach

First, the power of our Natural Language Processing (NLP) detection approach comes from the fact that it can be applied to any text message on any of the social networks. For example, on Facebook, Google+, Twitter, and LinkedIn, users can write short brief text messages under different names such as, status updates, posts, tweets etc. Therefore, our approach can be used to detect any location revealing information from users' text messages on any of the aforementioned social networks and warn the user about such disclosure. However, our approach

might be less useful on social networks that are different in nature than Facebook, Google+ and Twitter.

For example, our approach might be useless when applied to LinkedIn, a professional online social network. People use LinkedIn and similar networks for professional networking and business situations such as, finding jobs, hiring people, keeping tabs on their businesses. Therefore, information related to users' activities and social plans are rarely shared on these sites. On the other hand, Pinterest and Instagram have a unique characteristic that makes our approach not applicable. Visual in nature, Pinterest and Instagram emphasize the sharing of users' thoughts, ideas and even social events by using photos. For example, with the Place Pin feature on Pinterest, where users can collaborate on a group board and plan ideas for weekend activities around the town, users' locations could be identifies. When a user pins a photo, a story or an article from a website, they can add descriptive text to their pins. Therefore, our detection approach might be useful, to some extent, if it is applied carefully to some to these short, descriptive text messages that are attached to the photos.

Limiting Information Disclosure to Trusted Parties

After warning the user about dangerous information disclosure in their text post, and before the text is actually posted, we also suggest that the user limit this type of information disclosure to only a subset of trusted friends. Our approach to limit information disclosure is based on identifying trusted friends from their interaction behavior with the user. In this section, we explain how this approach could be applied to other OSNs:

Google+

Although Google+ has some features that Facebook does not have such as Hangouts on Air and Party Mode, Google+ by far is one of the most similar social networks to Facebook.

Similar to Facebook, after writing a post, Google+ users are faced by the challenge of manually choosing their audiences, called circles. Similar to Facebook, friends interact with the poster thought likes, comments, shares and tags. Therefore, our approach in identifying trusted friends on Facebook is directly applicable to Google+.

Twitter

Followers on Twitter express their approval of a tweet's content by retweeting or marking a tweet as "favorite." However, "retweet" and "favorite" are different terminologies to the same interaction methods offered on Facebook, namely, "share" and "like." Unlike Facebook where friends can express their opinion about the user's content by leaving comments and/or replies, followers on Twitter can express their opinion about a tweet's content by using "reply" only. On the other hand, "tags" on Facebook are called "mention" on Twitter.

Therefore, to protect users' safety and privacy on Twitter, our dangerous information detection approach could be applied directly to users' text-based tweets. On the other hand, our approach to suggest trusted followers need to be alternated so that it fits the way people interact on Twitter. For example, retweet or (share) is a new metric that we did not address in our approach that needed to be considered on Twitter when suggesting the group of trusted followers. Moreover, the commenting system/comments in Twitter are used in a different way than Facebook; therefore, our commenting metric needs to be modified.

LinkedIn

While Facebook and LinkedIn are both social networks, they play different roles in people's lives. For example, Facebook is a personal network, while LinkedIn is a professional network. Therefore, the type of data that people share on LinkedIn is different from that on Facebook. For example, if a person is on a vacation in Europe and sharing his/her family photos,

that person is more likely to share this information on Facebook. On the other hand, revealing location information on LinkedIn takes another form such as revealing information about attending a conference, a recruiting event, or a sales event etc. Therefore our dangerous information detection approach is directly applicable to LinkedIn text-based updates.

On the hand, social links on LinkedIn are often with a complete stranger such as with recruiters and sales people etc. Moreover, it is used by people seeking jobs or new career opportunities which do not happen often or on a daily basis. Therefore, even though a LinkedIn user can interact through like, share, and comments, we believe that these interaction information are not sufficient to suggest to a user how much to trust another LinkedIn user. For example, due to the absence of interaction between a recruiter and his/her connections, a recruiter announcing the time and location of a recruiting event on LinkedIn may lose all his potential recipients according to our approach. Deeper study needs to be done on LinkedIn to understand what information could be used to derive the value of trust.

Pinterest

It has been said, "A picture is worth a thousand words." Pinterest is a very unique and interesting social network with its visually focused networking approach. People join Pinterest to get inspiring ideas, promote their brands, and of course, to connect with others. While people do not share text-based information, photos that people share on Pinterest are open for interpretation. For example, users may not directly reveal their location through text-based messages, but photos they pin might be used in a way to make an inference about their location and activities patterns. Our approach in detecting information disclosure on Pinterest and similar social networks is limited to description text that users attached to photos.

The type of social connections that users establish on Pinterest is mostly with complete strangers. For example, upon signing up, Pinterest automatically add people who match users' interests to the list of people they follow. Unlike Facebook and Google+, Pinterest;s mission is focused on creating and sharing pins rather than getting to know follow pinners. Users interact on Pinterest through likes, comments, shares, tags and repins. However, due to the visual nature of the OSN, users' engagement in term of comments is very small. For example, only 0.6% of Pinterest users commented on pins. On the other hand, 83% of users' engagement on Pinterest is through the re-pinning method, while only 15% interacted using like. While our approach can be used to some extent in identifying trusted friends based on their interaction patterns, deeper study needs to be done to understand the type of social links that users establish on Pinterest, and the type of data that could be used to derive the value of trust and therefore suggest to a user who to trust. For example, the empirical study of users' behaviors on Pinterest (Feng; Cong, Chen, & Yu, 2013) reported that the most distinguishing characteristic of Pinterest users is their focus on everyday lives and their willingness to collect photos about decorations and designs, food and fashion.

Future Work

The new type of information that OSNs are increasingly encouraging people to share, will cause implications in a variety of arenas, e.g., privacy, security, safety, social relationships, marketing, economy. Therefore, we look for tremendous research opportunities there. In the long term, we plan to continue to study how the visibility of OSNs affects the safety and privacy of people and the role of information technology in addressing those challenges. Moreover, we plan on conducting a study on how to build a universal privacy approach for OSNs. In the short term, there are several directions to extend our current work.

First, in order to broadcast the right information to the right audiences, we need to expand our detection and categorization approach to cover other types of dangerous information people share on OSNs such as work and family related posts. First, we plan on exploring possible methods to expand our current approach to detect work and family revealing information.

In our current approach, we use Natural Language Processing (NLP) to learn and observe the semantics of the language on online social networks. Our approach in combining the use of NLP and tagged-based detection rules in detecting location revealing information benefited greatly from the location and time prepositions in users' text posts. However, this task will be very challenging to extend to other types of dangerous information, such as work and family. The data on online social networks has very unique characteristics such as abbreviations and the absence of consistently in following grammatical rules. In other words, this means that the informality of user-generated content on OSNs could create a completely new subfield of existing NLP studies.

Another alternative approach to NLP is to expand our detection approach to parse the text posts and then match its content against a set of predefined and categorized words. This approach would rely on an internal default dictionary. This dictionary would define a group of words that could be used to indicate a particular type of revealing dangerous information.

Second, when it comes to partitioning users' friends, we plan on testing our approach on a larger data set. Moreover, we also plan to expand the testing of our approach to cover the other interaction methods, namely, like, tags, and replies. In addition, we expect to enhance the accuracy of our current method of suggesting trusted friends. We plan on doing this by grouping friends' interaction records into groups based on the date of the friendship establishment date. In this case our algorithm can compare friends' interaction behavior to the subset of friends who

have all established the friendship with the user in the same month, or year. This will address the unfairness of comparison that we currently have in our approach. For example, it is unfair to compare the interaction behavior of a brand new friend with a friend who has been added a long time ago.

Third, while our approach suggests grouping friends based on a static snapshot of the friends' interaction behaviors with the user, we still need to capture the dynamic of the interaction behavior over time. For example, some friends might gradually get interested in a user's posts about his daily activities. Therefore, we need to suggest a mechanism to allow untrusted friends to express their change of interaction behavior with the user. One way of doing this is by studying the interaction behavior from both sides, namely, user-friends and friends-user. For example, if the user continually interacts with untrusted friends, this might be an indication that the absence of interaction is only from one side for some reason. In this case, we can suggest to the user to include this friend in his/her circle of trust.

Finally, our privacy model should be designed to learn from trial and error, since there is a chance of broadcasting the information to the wrong audiences. One possible approach to solve this issue is to incorporate users' feedbacks in the beginning, so our system can learn the patterns over time and the users do not have to correct the results.

REFERENCES

Accorsi, R., Zimmermann, C., & Muller, G. (2012). On taming the inference threat in social networks. In *1st International Workshop on Privacy and Data Protection Technology* (PDPT), Amsterdam. Netherlands.

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In P.Golle & G.Danezis (Eds.). *Proceedings of 6th Workshop on Privacy Enhancing Technologies* (pp. 36-58). Cambridge, UK: Robinson College.

Adali, S., Escriva, R., Goldberg, M. K., Hayvanovych, M. Magdon-Ismail, M., Szymanski, B. K., Wallace, W. A., & Williams, G. (2010). Measuring behavioral trust in social networks. In *Proceedings of the IEEE international conference on intelligence and security informatics*. (pp. 150-152). Vancouver, BC, Canada.

Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., & Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web 6*(2): 1-33.

Ali, B., Villegas, W., & Maheswaran, M. (2007, October 22-25). A trust based approach for protecting user data in social networks. In K. A. Lyons and C. Couturier, (Eds.). *Proceedings of the 2007 conference of the center for advanced studies on collaborative research*, CASCON' 07, (pp. 288-293). Richmond Hill, Ontario, Canada: IBM.

Andress, J. (2014). The basics of information security: Understanding the fundamentals of InfoSec in theory and practice. Rockland, MA: Syngress Publishing.

Ansper, A., Buldas, A., Roos, M., & Willemson, J. (2001, February 13-15). Efficient long-term validation of digital signatures, In *Proceedings of the 4th International Workshop on Practice and Theory in Public Key Cryptography: Public Key Cryptography*, (pp. 402-415). Cheju Island, Korea.

Asian media landscape is turning digital. How can marketers maximize their opportunities? (2012). Retrieved from http://www.nielsen.com/content/dam/corporate/au/en/reports/2012/changing-asian-media-landscape-feb2012.pdf

Baatarjav, E.-A. (2013). Privacy management for online social networks. Ph.D. Thesis. University of North Texas, Denton, TX.

Backstrom, L., Dwork, C., & Kleinberg, J. (2007, May 8-12). Wherefore art thou: Anonymized social networks, hidden patterns, and structural steganography, *Proceedings of the 16th international conference on World Wide Web*. Banff, Albert, Canada.

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: Membership, growth, and evolution. *Proceedings of 12th international conference on knowledge discovery in data mining* (pp. 44-54). New York: ACM Press.

Barnes. S. B., (2006, September). Privacy paradox: Social networking in the United States. *First Monday*, 11(9), Retrieved from http://firstmonday.org/article/view/1394/1312

Becker, J., & Chen, H. (2009, May). Measuring privacy risk in online social networks. In *Proceedings of W2SP 2009: Web 2.0 Security and Privacy*. Oakland, CA.

Beye, M., Jeckmans, A. J. P., Erkin, Z., Hartel, P. H., Lagendijk, R. L., & Tang, Q. (2010). Literature overview - Privacy in online social networks. *Technical Report TR-CTIT-10-36*, Centre for Telematics and Information Technology University of Twente, Enschede, Netherlands

Bonneau, J., Anderson, J., & Church, L. (2009). Privacy suites: Shared privacy for social networks, In *ACM International Conf. Proc. of the 5th Symposium on Usable Privacy and Security* (pp. 1-2). Mountain View, CA.

Boyd, D. M. & Ellison, N. B. (2007). Social network sites: Definitions, history, and scholarships. *Journal of Computer-Mediated Communication 13*(1), 210-230.

Brake, D. R. (2012). Who do they think they're talking to? Framings of the audience by social media users. *International Journal of Communication. V6*: 1056-1076.

Buskens, V. (1998). The social structure of trust. *Social Network. 20*(3), 265-289.

Butler, E., McCann, E., & Thomas, J. (2011). Privacy setting awareness on Facebook and its effect on user-posted content. *Human Communication 14*(1): 39-55.

Carminati, B., Ferrari, E., & Perego, A. (2009). Enforcing access control in web-based social networks, *ACM Transactions on Information & System Security, 13*(1): 1-38.

Caverlee, J., Liu, L., & Webb, S. (2008). Social trust: Tamper-resilient trust establishment in online communities. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'08)*. (pp. 104-114). New York: ACM Press.

Caverlee, J., Liu, L., & Webb, S. (2010). The social trust framework for trusted social information management: Architecture and algorithms. *Information Sciences 180*(1): 95-112.

CBS News (2010, March 25). Facebook "Friend" Suspected in Burglary. Retrieved from http://www.cbsnews.com/news/facebook-friend-suspected-in-burglary/

Chen, D. & Manning, C. D. (2014). A fast and accurate dependency parser using neural networks. *Proceedings of EMNLP, 2014* Retrieved from http://nlp.stanford.edu/software/lex-parser.shtml

Chen, X. & Shi, S. (2009). A literature review of privacy research on social network sites, In *MINES '09 1*, (pp. 93-97).

Cheng, P. (1998). A security architecture for the internet protocol, *IBM Systems Journal 37*(1), 42-60.

Clements, M., de Vries, A. P., & Reinders, M. J. T. (2010, July). The influence of personalization on tag query length in social media search. *Information Processing and Management: an International Journal, 46*(4), 403-412.

Cliff A., Lampe, C., Ellison, N., & Steinfield, C. (2007, April 28-May 3). A familiar face(book): Profile elements as signals in an online social network. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* San Jose, CA.

Cloak. (2015). Reclaim your privacy. Retrieved from http://www.reclaimprivacy.org

Collin, P., Richardson, I., & Third, A. (2011). The benefits of social networking services. *Cooperative Research Centre for Young People, Technology and Wellbeing*. Retrieved from http://www.fya.org.au/wp-content/uploads/2010/07/The-Benefits-of-Social-Networking-Services.pdf

Crawford, K. (2009). Following you: Disciplines of listening in social media. *Continuum-Journal of Media & Cultural Studies 23*(4): 525-535.

DeCicco, N. (2008, June 9). MySpace, Facebook, more pose challenge. Travis Air Force Base News. Retrieved from http://www.travis.af.mil/news/story.asp?id=123101954

Devine, S. M. (2008). Anti-social networking: Exploiting the trusting environment of Web 2.0. *Network Security 11*: 4-7.

Dhami, A., N. Agarwal, T. K. Chakraborty, B. Singh, P., & Minj, J. (2013). Impact of trust, security and privacy concerns in social networking: An exploratory study to understand the pattern of information revelation in Facebook. *3rd IEEE International Advance Computing Conference (IACC)* (pp. 465-469). Ghaziabad, Uttar Pradesh, India.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Social media update 2014, Pew Research Center, Retrieved from http://www.pewinternet.org/2015/01/09/social-media-update-2014/

Dwyer, C., Hiltz, S. R., & Passerini, K. (2007). Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. AMCIS 2007 Proceedings Paper 339. Keystone, CO.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4), (pp. 1143- 1168). Retrieved from http://blackwell-synergy.com/doi/abs/ 10.1111/j.1083-6101.2007.00367.x doi: 10.1111/j.1083-6101.2007.00367

Enough Is Enough. (2014). Cyberbullying Statistics. Retrieved from http://www.internetsafety101.org/cyberbullyingstatistics.htm

Facebook, (2015). News room. http://newsroom.fb.com/

Feng, Z., Cong, F., Chen, K., & Yu, Y. (2013, November 17-20). An empirical study of user behaviors on Pinterest social network. *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE/WIC/ACM International Joint Conferences* (pp. 402-409). Atlanta, GA.

Flad, K., (2010). *The influence of social networking participation on student academic performance across gender lines*. (Counselor Education Master's Theses). College at Brockport: State University of New York. Retrieved from http://digitalcommons.brockport.edu/cgi/viewcontent.cgi?article=1030&context=edc_theses

Goecks, J., Edwards, W. K., & Mynatt, E. D. (2009, July 15-17). Challenges in supporting end-user privacy and security management with social navigation. In *Proceedings of the 5th Symposium on Usable Privacy and Security*. Article No. 5. Retrieved from http://dl.acm.org/citation.cfm?id=1572539

Golbeck, J. (2005). Personalizing applications through integration of inferred trust values in semantic web-based social networks. In Processing of the ISWC Semantic Network Analysis Workshop. Galway, Ireland.

Golbeck, J. & Hendler, J. (2006). Inferring binary trust relationships in web-based social networks. *ACM Transactions on Internet Technology*, volume pp. 497-529.

Golbeck, J., Parsia, B., & Hendler, J. (2003, August 27-29). Trust networks on the semantic web. In *Proceedings of the 7th International Workshop on Cooperative Intelligent Agents* (pp. 238-249). Helsinki, Finland.

Golder, S. A., Wilkinson, D., & Huberman, B. A. (2007). Rhythms of social interaction: Messaging within a massive online network. *Third International Conference on Communities and Technologies, 2007*. Retrieved from http://www.hpl.hp.com/research/idl/papers/facebook/facebook.pdf

Grabner-Kräuter, S. (2009). Web 2.0 social networks: The role of trust. *Journal of Business Ethics 90*(4): 505-522.

Grabner-Kräuter, S. & Bitter. S. (2015). Trust in online social networks: A multifaceted perspective, *Forum for Social Economics, 44*(1), 48-68.

Grassi, M., Cambria, E. Hussain, A., & Piazza, F. (2011). Sentic web: A new paradigm for managing social media affective information. *Cognitive Computation 3*(3): 480-489.

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks: The 2005 ACM workshop on privacy. In *The Electronic Society*, 71-80.

Hamdi, S., Gancarski, A. L., Bouzeghoub, A., & BenYahia, S. (2012, June 25-27). Iris: A novel method of direct trust computation for generating trusted social networks. In *Proceedings of the 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications* (pp. 616-623). Liverpool, UK: TrustCom 2012.

Hart, M., Castille, C., Johnson, R., & Stent, A. (2009, August 29). Usable privacy controls for blogs, In *Proceedings of the International Conference on Computational Science and Engineering* (pp. 401-408). Vancouver, Canada.

Haythornthwaite, C. (2005). Social networks and internet connectivity effects. *Inf. Comm. Soc. 8*(2), 125-147.

Heer, J., & Boyd, D. (2005). Vizster: Visualizing online social networks. In *Proceedings of Symposium on Information Visualization* (pp.33-40). Minneapolis, MN: IEEE Press.

Herring, S. C., Paolillo, J. C., Vielba, I. R., Kouper, I., Wright, E., Stoerger, S., Scheidt, L. A., & Clark, B. (2007). Language networks on LiveJournal. In *Proceedings of the Fortieth Hawai'i International Conference on System Sciences*. Los Alamitos, CA: IEEE Press.

Hirose, M., Utsumi, A., Echizen, I., & Yoshiura, H. (2012). A private information detector for controlling circulation of private information through social networks, (pp. 473-478) Seventh International Conference on Availability, Reliability and Security. Tokyo, Japan.

Hsu, W. H., Lancaster, J., Paradesi, M. S. R., & Weninger, T. (2007). Structural link analysis from user profiles and friends networks: A feature construction approach. Manhattan, KS: Kansas State University, Department of Computing and Information Sciences.

Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior 28*(2): 561-569.

Huynh, T. D., Jennings, N. R., & Shadbolt, N. R. (2006). Certified reputation: How an agent can trust a stranger. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)* (pp. 1217-1224). New York, NY,

IACP social media center. (2010). Retrieved from http://www.iacpsocialmedia.org/

Immediate ban of Internet social networking sites (SNS) on Marine Crops enterprise network (MCNE) NIPRINT. (2009, August 3). Retrieved from http://www.marines.mil/News/Messages/MessagesDisplay/tabid/13286/Article/112458/immediate-ban-of-internet-social-networking-sites-sns-on-marine-corps-enterpris.aspx

Josang, A., Hayward, R., & Pope, S. (2006). Trust network analysis with subjective logic. In *Proceedings of the 29th Australasian Computer Science Conference (ACSC'06)* (pp. 85-94). Hobart, Australia: Australian Computer Society.

Kalemi, E. & Martiri, E. (2011). FOAF-academic ontology: A vocabulary for the academic community. In *Proceedings of the 2011 Third Int. Conf. on Intelligent Networking and Collaborative Systems (INCOS '11)* (pp. 440-445). Washington, DC: IEEE Computer Society.

Kang, J. (1998). Information privacy in cyberspace transactions. *Stanford Law Review, 50* (p. 1193). Retrieved from SSRN, http://ssrn.com/abstract=631723

Kang, T. & Kagal, L. (2010, March 22-24). Enabling privacy-awareness in social networks. In *Intelligent Information Privacy Management Symposium at the AAAI Spring Symposium 2010*. Stanford, CA: Technical Report SS-10-05.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons 53*(1): 59-68. Retrieved from http://openmediart.com/log/pics/ sdarticle.pdf

Kataoka, H., Utsumi, A., Hirose, Y., & Yoshiura, H. (2010, March). Disclosure control of natural language information to enable secure and enjoyable communication over the internet. *International Journal of u- and e- Service, Science and Technology 3*(1), 178-188.

Kietzmann, J. H., Hermkens, K., McCarthy, I., & Silvester, B. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons 54*(3): 241-251.

Kim, H.-N., Rawashdeh, M., Alghamdi, A., & El Saddik, A. (2012, March). Folksonomy-based personalized search and ranking in social media services. *Information Systems* 37(1): 61-76.

Kim, H.-N., Roczniak, A., Levy, P., & Saddik A. (2012). Social media filtering based on collaborative tagging in semantic space. *Multimedia Tools and Applications 56*(1): 63-89.

Kim, Y. A. (2008). Building a web of trust without explicit trust ratings. In *Proceedings of the 24th IEEE International Conference on Data engineering Workshop* (pp. 531-536). Cancun, Mexico.

Korolova, A., Motwani, R., Nabar, S. U., & Xu, Y. (2008, October 26-30). Link privacy in social networks. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, (n.p.). Napa Valley, CA.

Kreiser, J. (2006). MySpace: Your kids' danger? Retrieved from http://www.cbsnews.com/news/myspace-your-kids-danger/

Kumar, R., Novak, J., & Tomkins, A. (2006). Structure and evolution of online social networks. In *Proceedings of 12th International Conference on Knowledge Discovery in Data Mining* (pp. 611-617). New York: ACM Press.

Lam, T.-F., Chen, K.-T., & Chen, L.-J. (2008, November 25-27). Involuntary information leakage in social network services. In *Proceedings of the 3rd International Workshop on Security: Advances in Information and Computer Security*, Kagawa, Japan.

Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior 28*(2): 331-339.

Lehrman, Y. (2010). The weakest link: The risks associated with social networking websites, New York Police Department, *Journal of Strategic Security JSS 3*(2): 63-72.

Lenhart, A., & Madden, M. (2007). Teens, privacy and online social networks. *Pew Internet and American Life Project Report*. Retrieved from http://www.pewinternet.org/2007/04/18/teens-privacy-and-online-social-networks

Lewis, K., Kaufman, J., & Christakis, N. (2008). Harvard University. The taste for privacy: An analysis of college student privacy settings in an online social journal of computer-mediated communication. *Journal of Computer-Mediated Communication 14*(1): 79 - 100.

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., & Tomkins, A. (2005). Geographic routing in social networks. In *Proceedings of National Academy of Sciences, 102*(33): 11,623-11,628.

Lipford, H. R., Besmer, A., & Watson, J. (2008). Understanding privacy settings in Facebook with an audience view. In *UPSEC'08: Proceedings of the 1st Conference on Usability, Psychology, and Security*, (pp. 1-8), Berkeley, CA, USA.

Liu, H., Lim, E., Lauw, H., Le, M., Sun, A., Srivastava, J., & Kim, Y. (2008, July 8). Predicting trusts among users of online communities: An epinions case study. In *Proceedings of the 9th ACM Conference on Electronic Commerce* (pp. 310-319). Chicago, IL: ACM.

Liu, L., Zhu, F., Jiang, M., Han, J., Sun, L., & Yang, S. (2012). Mining diversity on social media networks. *Multimedia Tools and Applications 56*(1): 179-182.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2013). Teens, social media, and privacy. Pew Research center. Retrieved from http://www.pewinternet.org/ 2013/05/21/teens-social-media-and-privacy/

Mahmood, S. (2012, November). New privacy threats for Facebook and Twitter users. In *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, Seventh International Conference, (pp. 164-169). Victoria, BC, Canada.

Malik, Z., Akbar, I., & Bouguettaya, A. (2009, November 24-27). Web services reputation assessment using a hidden Markov model. In *Proceedings of the 7th International Joint Conference on Service-Oriented Computing* (pp. 576-591). Berlin, Germany: Verlag.

Maximilien, E. M., Grandison, T., Sun, T., Richardson, D., Guo, S., & Liu, K. (2009). Privacy-as-a-service: Models, algorithms, and results on the Facebook platform. In processing of *Workshop Program - W2SP 2009 2*. Oakland, CA.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integration model of organizational trust, *Academy of Management Review 20*(3): 709-734.

Mitchell, K. J., Finkelhor,D., Jones, L.M., & Wolak, J. (2010). Use of social networking sites in online sex crimes against minors: An examination of national incidence and means of utilization. *Journal of Adolescent Health* 47(2): 183-190.

Moalla, S., Hamdi, S., & Defude, B. (2010, December 15-17). A new trust management model in p2p systems. In *Proceedings of the 6th IEEE International Conference on Signal-Image Technologies and Internet-Based System, SITIS'*, (pp. 241-246). Kuala Lumpur, Malaysia: IEEE Computer Society.

Moturu, S. T., & Liu, H. A. (2011). Quantifying the trustworthiness of social media content. *Distributed and Parallel Databases 29*(3): 239-260.

Munoz, C., & Towner, T. (2009). Opening Facebook: How to use Facebook in the college classroom. In Gibson, I. Weber, R., McFerrin, K., Carlsen, R., & Willis, D. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 2623-2627). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Nadali, S., Murad, M. A. A., Sharef, N. M., Mustapha, A., & Shojaee, S. (2013). A review of cyberbullying detection: An overview. *Intelligent Systems Design and Applications (ISDA), 13th International Conference* (pp. 325-330). Selangor, Malaysia: Universiti Putra Malaysia.

Nepal, S. & Sherchan, W. (2011). STrust: A trust model for social networks. In *Proceedings of the 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications* (pp. 841-846). Changsha, China.

Nepal, S., Sherchan, W., & Bouguettaya, A. (2010). A behavior based trust model for service web. In *Proceedings of the IEEE International Conference on Service-Oriented Computing and Applications (SOCA'10)* (pp. 1-4). Perth, WA: IEEE Computer Society.

Ngoc, T., Echizen, I., Kamiyama, K., & Yoshiura, H. (2010). New approach to quantification of privacy on social network sites. In *Proceedings of 24th IEEE International Conference on Advanced Information Networking and Applications* (pp. 556-564). Perth, WA.

Nguyen-Son, H.-Q., Nguyen, Q.-B., Tran, M.-T. Nguyen, D.-T., Yoshiura, H., & Echizen, I. (2012, August 20-24). Automatic anonymization of natural languages texts posted on social networking services and automatic detection of disclosure. *Seventh International Conference on Availability, Reliability and Security*. (pp. 358-364). Prague, Czeck, Republic.

Nuha, N., Molok, A., Chang, S., & Ahmad, A. (2010, November 30). Information leakage through online social networking: Opening the doorway for advanced persistence threats. Originally published in the *Proceedings of the 8th Australian Information Security Management Conference*, Edith Cowan University, Perth Western Australia.

Onwuasoanya, A., Skornyakov, M., & Post, J. (2008). Enhancing privacy on social networks by segregating different social spheres. *Rutgers Gov Sch Eng Technol Res J 3*: 1-10.

Paolillo, J. C., & Wright, E. (2005). Social network analysis on the semantic web: Techniques and challenges for visualizing FOAF. In V.Geroimenko & C.Chen (Eds.), *Visualizing the Semantic Web* (pp. 229-242). Berlin: Springer.

Penna, L., Clark, A., & Mohay, G. (2010). A framework for improved adolescent and child safety in MMOs. In: Memon, N., Alhajj, R. (Eds.) *2010 International Conference on Advances in Social Network Analysis and Mining (ASONAM 2010)* (pp. 33-40). IEEE Computer Society. Odense, Denmark: University of Southern Denmark.

Poeter, D., (2011). Study: A quarter of parents say their child involved in cyberbullying. Retrieved from http://www.pcmag.com/article2/0,2817,2388540,00

Police: Thieves robbed homes based on Facebook, social media sites. (2010). Retrieved from http://www.wmur.com/Police-Thieves-Robbed-Homes-Based-On-Facebook-Social-Media-Sites/11861116

Pontes, T., Magno, G., Vasconcelos, M., Gupta, A., Almeida, J., Kumaraguru, P., & Almeida, V. (2012, December 10). Beware of what you share: Inferring home location in social networks. In *Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops* (pp. 571-578). Brussels, Belgium.

Poovy, B. (2010). Man convicted of Palin email hacking. *The Sydney Morning Herald*. Associated Press. Retrieved from http://www.smh.com.au//breaking-news-world/man-convicted-of-palin-email-hacking-20100501-tzoi.html

Popular websites, list of most. (2015, August 14). Retrieved from https://en.wikipedia.org/wiki/List_of_most_popular_websites

Qi, G. J., Aggarwal, C., Tian, Q., Ji, H., & Huang, T. (2012). Exploring context and content links in social media: A latent space method. *IEEE Transactions on Pattern Analysis and Machine Intelligence 34*(5): 850-862.

QSR International. (2015). NCapture. Retrieved from http://www.qsrinternational.com/products_nvivo_add-ons.aspx

Rashid, A., Rayson, P., Greenwood, P., Walkerdine, J., Duquenoy, P., Watson, P., Brennan, M., & Jones, M. (2009, June). Isis: Protecting children in online social networks. In *At the International Conference on Advances in the Analysis of Online Paedophile Activity*, Paris, France.

Rosenblum, D. (2007). What anyone can know: The privacy risks of social networking sites. *IEEE Security and Privacy 5*(3): 40 -49.

Ruohonen, K. (2013). *Graph Theory*. Retrieved from http://math.tut.fi/~ruohonen/GT_English.pdf

Schrape, J. F. (2011). Social media, mass media and social reality construction. *Berliner Journal Fur Soziologie 21*(3): 407-429.

Sherchan, W., Nepal, S., & Paris, C. (2013, August). A survey of trust in social networks. *ACM Computing Surveys (CSUR) 45*(4):1-33.

Shin, H., & Lee, J. (2012, August). Impact and degree of user sociability in social media. *Information Sciences: an International Journal, 196:* 28-46.

Srivastava, A., & Geethakumari, G. (2013). A framework to customize privacy settings of online social network users. *IEEE Recent Advances in Intelligent Computational Systems (RAICS)* (pp. 187-192). Trivandrum, Kerala, India.

Sterling, G. (2013). Pew: 94% of teenagers use Facebook, have 425 Facebook friends, but Twitter and Instagram adoption way up, Pew Research Center. Retrieved from http://marketingland.com/pew-the-average-teenager-has-425-4-facebook-friends-44847

Strater, K., & Richter, H. (2007, July 18-20). Examining privacy and disclosure in a social networking community, In *Proceedings of the 3rd Symposium on Usable Privacy and Security*. (pp. 157-158). Pittsburgh, PA.

Stutzman, F. (2006). *Student life on the Facebook*. Retrieved from http://ibiblio.org/fred/facebook/stutzman_fbook.pdf

Trammell, K., & Keshelashvili, A. (2005). Examination of the new influences: A self-presentation study of A-list blogs. *Journalism and Mass Communication Quarterly 82*: 968-982.

Utz, S., & Krämer, N. (2009). The privacy paradox on social network sites revisited: The role of individual characteristics and group norms. *Cyberpsychology: Journal of Psychosocial Research in Cyberspace 3*(2): Article 1. Retrieved from http://www.cyberpsychology.eu/view.php?cisloclanku=2009111001&article=1

Valkenburg, P. M., Peter, J., & Schouten, A. P. (2006). Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *CyberPsychology Behavior, 9*(5), 584-590. Retrieved from http://www.liebertonline.com/doi/abs/10.1089/cpb.2006.9.584 doi:101089/cpb.2006.9.584

Vasalou, A., Joinson, A., Bänziger, T., Goldie, P., & Pitt, J. (2008). Avatars in social media: Balancing accuracy, playfulness and embodied messages. *International Journal of Human-Computer Studies 66*(11): 801-811.

Volakis, N. (2011). Trust in online social networks (Master thesis). School of Informatics, University of Edinburgh, Edinburgh, Scotland. Retrieved from, http://www.inf.ed.ac.uk/publications/thesis/online/IM110932.pdf

Wang, Y.-C., Burke, M., & Kraut, R. E. (2013, April 27-May 2). Gender, topic, and audience response: an analysis of user-generated content on Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paris, France.

Watanabe. N, & Yoshiura, H. (2010). Detecting revelation of private information on online social networks. *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*. 2010 Sixth International Conference (pp. 502-505). Darmstadt, Germany.

Weiss. S. (2008). The need for a paradigm shift in addressing privacy risks in social networking applications. In *The Future of Identity in the Information Society 262* (pp. 161-171). IFIP International Federation for Information Processing. Karlstad, Sweden.

Wordpress. (2015). Retrieved from https://wordpress.org/

Xiong, F., Liu, Y., Zhang, Z., Zhu, J., & Zhang, Y. (2012). An information diffusion model based on retweeting mechanism for online social media. *Physics Letters A 376*(30-31): 2103-2108.

Yaich, R., Boissier, O., Jaillon, P., & Picard, G. (2011, August 22-27). Social-compliance in trust management within virtual communities. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (pp.322-325). Lyon, France.

Yan, Z., Niemi, V., Dong, Y., & Yu, G. (2008). A user behavior based trust model for mobile applications. In *Proceedings of the 5th International Conference on Autonomic and Trusted Computing*. (pp. 455-469**).** Heidelberg, Germany**.**

Yu, B., & Singh, M. P. (2002). An evidential model of distributed reputation management. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'02)* (pp. 294-301). New York: ACM Press.

Zakaria, N., Lau, K. Y., Alias, N. M. A., & Husain, W. (2011). Protecting privacy of children in social networking sites with rule-based privacy tool. *High Capacity optimal Networks and Enabling Technologies (HONET)* (pp. 253-257). Riyadh, Saudi Arabia.

Zhan, J. & Fang,X. (2011). A novel trust computing system for social networks. In *Proceedings of the IEEE International Conference on Privacy, Security, Risk and Trust* (pp. 1284-1289). Boston, MA.

Zhang, Y., Chen, H., & Wu, Z. (2006). A social network-based trust model for the semantic web. In *Proceedings of the 6th International Conference on Autonomic and Trusted Computing* (pp. 183-192). Chicago, IL.

Zhao, K., & Pan, L. (2014). A machine learning based trust evaluation framework for online social networks. *Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE 13th International Conference (pp. 69-74). Beijing, China.

Zhou, D., Lawless, S., & Wade, V., (2012). Improving search via personalized query expansion using social media. *Information Retrieval for Social Media 15*(3-4): 218-242.