Graduate School ETD Form 9
(Revised 01/14)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Yue Zhang

Entitled
A CROSS-SITE STUDY OF USER BEHAVIOR AND PRIVACY PERCEPTION IN SOCIAL NETWORKS

For the degree of Master of Science

Is approved by the final examining committee:

Melissa Dark

Brandeis H. Marshall

Baijian Yang

To the best of my knowledge and as understood by the student in the *Thesis/Dissertation Agreement. Publication Delay, and Certification/Disclaimer (Graduate School Form 32)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Melissa Dark

Approved by Major Professor(s): _____

Approved by: Gene Spafford 07/08/2014

Head of the Department Graduate Program Date

A CROSS-SITE STUDY OF USER BEHAVIOR AND PRIVACY PERCEPTION IN SOCIAL NETWORKS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Yue Zhang

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

August 2014

Purdue University

West Lafayette, Indiana

To my Mom, Dad and Qi, for all the support and love they give me.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my committee. I am so fortunate to have the best committee to help me going through the process of inquiry. My Chair, Prof. Dark has been patiently giving me advice and leading me to the right direction. I would never forget her support, advice and encouragement. I want to thank Prof. Yang and Prof. Marshall for their unwavering help and invaluable input. I also want to thank Prof. Whitten for helping me going through the formatting process.

There are many who have supported the creation of this work, making it difficult to choose the words that truly express the heartfelt gratitude and appreciation. I feel I am fortunate to be surrounded by an amazing group of graduate students and faculties. Thank every one of them and thank Purdue for giving me such a wonderful experience.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

Table Page

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Appendix Figure

LIST OF ABBREVIATIONS

OSN - Online Social Networks

SSN - Social Security Number

ANOVA - Analysis of variance

URL - Uniform Resource Locator

IRB - Institutional Review Board

GLOSSARY

Privacy attitude	User's general privacy concern toward information privacy. To be specific, it means what information a user considers private and what information she considers public.
Privacy perception	Measures how a user perceives privacy risks while using social networks. In this thesis, such perception refers to how much a user trusts social networking sites.
Privacy behavior	A user's actual behavior that relates to privacy protection or indicates privacy awareness while using social networks. Such behavior includes changing privacy setting, using private profile, etc.
One-way ANOVA	A common technique used to compare means of two or more samples. It tests the null hypothesis that samples in two or more groups are drawn from populations with the same mean values.
Kruskal-Wallis test	Non-parametric equivalent of one-way ANOVA. It's used for testing differences of ordinal variables in this study.
Cohen's guideline	A guideline for interpreting correlation results. According to this guideline, r= 0.5, 0.3 and 0.1 represents large/medium/small correlation respectively (Cohen 1988).

ABSTRACT

Zhang, Yue. M.S., Purdue University, August 2014. A Cross-site Study of User Behavior and Privacy Perception in Social Networks. Major Professor: Melissa Dark.

While online social networking sites have brought convenience and diversity in people's social lives, they have also been the source for information leakage. Researchers have been looking for ways to balance user privacy protection and information disclosure. However, literature suggested that many users either failed to perceive privacy risks correctly or they failed to behave in accordance with privacy awareness even they have already perceived potential risks.

This thesis conducted a survey to measure social network users' privacy attitude, privacy perception and their actual behavior when using social networking sites. The survey targeted at three populations of different cultural contexts: U.S. college students, Chinese students in the U.S. and Chinese students in China. It also targeted at 6 populate sites – Facebook, Twitter. WhatsApp, RenRen, Weibo and WeChat.

Based on the survey results, this thesis conducted a cross-cultural and cross-site study to explore the relationships of social network users' privacy attitudes, privacy perceptions and various user behaviors. It also studied whether cultural contexts and the differences of sites had an impact on privacy attitude, perception and behavior.

CHAPTER 1. INTRODUCTION

The Internet's wide adoption has contributed to online social networking sites' thriving popularity. Facebook, for example had 1.3 billion monthly active users worldwide in 2014 (StatisticBrain, 2014) compared with a total of 835 million in 2012 (Internet World Status, 2012). The nature of social networks is to imitate real world social relationships by providing mechanisms for sharing information, creating personal profiles, establishing relations and communicating with each other. Not surprisingly, at the same time of using such services, people are giving out massive amount of information which may pose real threat to privacy. Documented threats include identity theft, digital stalking, and personalized spam. The problem becomes worse when most people are completely unaware of short-term and long term risks of sharing personal information without restricted access (Schrammel et al., 2009; Krishnamurthy et al., 2008; Acquisti et al.,2006) .

Privacy preserving methods which aim at anonymizing the social graph (Machanavajjhala et al., 2007; Li et al., 2007; Sweeney, 2002), privacy setting management (Squicciarini et al, 2012) or raising privacy awareness by evaluating user's privacy score (Liu et al, 2010), etc. have not been proven successful in protecting user's sensitive information or changing user's information disclosure behavior.

This thesis argued that in order to preserve privacy in social networks, the differences and interactions of multiple social networks should be considered as well as the differences of user's privacy attitudes to better define and mitigate privacy risks. This study took the first step towards evaluating and preserving privacy by studying the differences and interactions of user's privacy, attitude, perception and behavior variables in different social networking sites. It also studied whether people of different cultural contexts would perceive or behave differently when using social networks.

1.1 Motivation and Objectives

Recent literature has identified that using multiple social networks have become an emerging threat to user privacy. The study by Irani et al. (2011) has shown that the more social network a user uses, the more information can be potentially leaked. They argued that, because different social networks have different privacy protections, the risk of information leakage may be dependent on the "weakest point" in the social network ecosystem. Malhora et al. (2012) successfully linked the different profiles in different social networks that belonged to the same user which demonstrated the threats for those users who used multiple social network services.

Therefore, it would no longer be valid that privacy protection can be contained within the boundary of each social network. The information flow among social networks enables profit-seeking individuals or organizations to collect as much "digital footprints" (Irani et al. 2011) as possible by integrating a user's information that she has disclosed from all the sites that she uses.

Though such an emerging risk has been raised for several years, few literatures have focused on such topic and no solutions that attempt to preserve privacy across multiple social networks that have been proven effective.

The above literature assumed that the privacy risks came from the fact that users disclosed different information in different social networks. However, arguments such as "users do not behave inconsistently nor they have inconsistent profiles in different social networks" or "users may not care about the information they have provided at all" may easily debunk the above assumption. To find out a solution that adapts to real life scenario, the fundamental understanding of why and how the usage of different social networks poses threat to privacy is necessary.

Therefore, the main objective of this study is to understand how and why user's privacy perceptions and behaviors differ in different social networks and among different cultural contexts and how the privacy perception influences the information disclosure behavior.

1.2 Significance

As discussed above, previous studies have assumed that users behave differently in social networking sites, however, none of them sufficiently justified their argument either because of lacking empirical data to support their argument or they fail to explore the reasons behind the differences of information disclosure.

Schrammel et al. (2012) took a site-centered approach to explore the difference of information disclosure on different types of social networks. This aggregated approach failed to distinguish the difference within the same type of social networks. Wang's work

on the other hand, did discover the difference of user's tagging behavior between two popular bookmarking websites. However, this research only focuses on only two tagging sites and such an ad-hoc result can't be justifiably generalized to other sites.

To the best of our knowledge, a cross-site study of privacy attitude, perception and behavior on different social networks has not yet been conducted studied the impact of cultural contexts on social network users which have been rarely documented.

Another significance of the study was that it provided an up-to-date survey that investigates diverse aspects of social network privacy. Hopefully it could help researchers in this field better understand the usages, perceptions, attitudes and behaviors of social network users.

1.3 Research Questions

The questions central to this research are as follows:

1. What are users' privacy attitudes when they use the social networks? (e.g. what information do they consider private and what is not?) Are they different?
2. Does culture background have a significant impact on social network users' privacy perceptions and behaviors?
3. What's the relationship among a user's privacy attitudes, perceptions, and behavior in a specific site?

1.4 Assumption

The following assumptions are inherent to this study:

1. The participants are assumed to be honest and to have a basic understanding of the definition of privacy and social networks without major confusion.
2. The participants are assumed to be able to use the Qualtrics online survey system and to navigate and answer the questions correctly.
3. The participants will not retake the survey as not to disproportionally affect the outcomes.
4. The participants are representative of the study population.
5. The survey provides adequate information for the research questions.

1.5 Limitations

The study has the following limitations:

1. The self-reported survey may be biased as the actual behavior of the respondents on social networks may be different from what they reported. The self-reported survey may be biased also because the non-respondents may be more concerned about privacy.
2. Social networking is fast-evolving. The popularity of each site rises and falls. The results of this study including the survey itself are prone to be dated.
3. The design of the survey questions tried to capture the general characteristics of each sites. In many cases, they were of coarse-grained. The profile preference question for example, only had two options – public or private