Association for Information Systems
AIS Electronic Library (AISeL)

WHICEB 2020 Proceedings | Wuhan International Conference on e-Business

Summer 7-5-2020

Exploring Privacy-traces of Users from Online Community: A Case Study of Diabetes Topic Discussions

Wenxiu Chen
School of Information, Central University of Finance and Economics, China

Xi Wang
School of Information, Central University of Finance and Economics, China, xiwang@cufe.edu.cn

Xinyi Li
School of Information, Central University of Finance and Economics, China

Yahan Wang
School of Information, Central University of Finance and Economics, China

Follow this and additional works at: https://aisel.aisnet.org/whiceb2020

Recommended Citation
Chen, Wenxiu; Wang, Xi; Li, Xinyi; and Wang, Yahan, "Exploring Privacy-traces of Users from Online Community: A Case Study of Diabetes Topic Discussions" (2020). *WHICEB 2020 Proceedings*. 65.
https://aisel.aisnet.org/whiceb2020/65

This material is brought to you by the Wuhan International Conference on e-Business at AIS Electronic Library (AISeL). It has been accepted for inclusion in WHICEB 2020 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Exploring Privacy-traces of Users from Online Community: A Case Study of Diabetes Topic Discussions

Wenxiu Chen, Xi Wang, Xinyi Li and Yahan Wang*
School of Information, Central University of Finance and Economics, China

Abstract: Online health communities (OHCs) have already become essential medium for people to obtain medical knowledge, share experiences and emotions. OHC users are able to post user-generated content (UGC) to interact with each other. However, the large amount of UGC may lead to personal information even privacy disclosed online. Although such disclosure may help users to trade some social support, which is the basis of sustaining a successful OHC, the users should be aware of the risks of leaving such traces online. This study selects a popular online Q & A community "Zhihu" in China as the research target. By collecting all questions and corresponding answers from 4 diabetes sub-communities, we would like to identify online privacy-traces of users from UGC. According to the theory of Communication Privacy Management, we build an explanatory model to understand user behaviors of concealing or revealing private information from the aspects of user characteristics, peer attention, and social support effects.

Keywords: online health community, User-generated Content (UGC), Communication Privacy Management theory, privacy leaking behavior

1. INTRODUCTION

By the end of year 2018, the quantity of Chinese internet users has reached 829 millions, in which Health and Medical ranked highest (72.89%) among all mobile searching topics [1]. With the rapid development of information technologies and increasing attention of individuals towards health issues, online health communities (OHCs) have become valuable venues for people to communicate and receive social support [2]. OHCs allow users to disclose health related information and share life stories with online strangers, which makes it hard to avoid leaving privacy traces [3]. Nowadays more and more people started to pay attention to security of privacy, but majority of people did not take steps away from revealing personal information for trading relatively rewards, which is also known as the "privacy paradox" [4]. In terms of the OHCs, this problem seems more serious since the healthcare related personal information is more sensitive and valuable.

Prior studies about OHCs have worked on user motivations towards knowledge sharing or social support. For example, the health professionals and regular users have different intentions in sharing knowledge in OHCs—the regular users are more influenced by the altruism, reciprocity, and empathy [5]. Researchers also focused on the participants' activities from the social support perspective and discussed the mechanism of social support on user churn [6]. The investigations of self-disclosure in OHCs are emerging as well, such as analyzing the privacy control mechanisms [7]. However, few studies have conduced work on exploring privacy-traces of users from the online community and further identify users' privacy-leaking trajectories, especially with the consideration of users' aggregated behaviors over many different sub-communities.

In this study, we developed a conceptual model about identifying effect of influential factors on users' privacy disclosure behaviors. By analyzing data scraped from Zhihu, we would like to identify users' privacy-traces left online. The managerial implication of the study is to improve the design of online community information system and to lower the risks of users at the meantime. For example, the community may adjust the privacy policy and operating strategy while participants need to improve consciousness of privacy protection.

* Corresponding author. Email: xiwang@cufe.edu.cn

2. RELATED WORK

2.1 Privacy disclosure in online community

In OHCs, people tend to reveal their private information for different purposes. Inspired by previous studies, we divided the online private information into two types, personal identifiable information (PII) and personal health information (PHI). PII refers to the information which contains individual characteristics and could be used to distinguish, identify, contact, and locate an individual. PHI means the health-related information, including personal treatment information, personal health condition, personal living habits, personal emotional information, and other personal experience.

In online communities, people usually do not show the same extent of privacy behaviors as they do in offline contexts [8]. When people set their own profile in SNS, lots of PII would be displayed, such as gender, birthday, educational information and occupation. People who have a greater willingness to share personal information online would be more likely to meet the cyber-victimization and engage in other potentially risky behaviors [9]. Meanwhile, OHC users disclose their PHI in exchange for trading support or advices [10]. They share their medical histories, feelings and experiences with people who have suffered similar health problems to receive their support [11]. Also, health information includes both the cognitive aspect, including information for disease prevention and treatment, and the affective aspect, including information for coping with illness emotionally [12]. According to the prior studies, we listed categories of private information disclosed in OHCs in Table 1:

Table 1. Privacy Information in OHCs

Category	Explanation	Type
Personal Identifiable Information (PII)	It includes both users' demographic features, which includes name, gender, residence, education background, career experience, user's interests and social relationship.	PII
Personal Treatment Information (PTI)	It refers to the medical information, like prevention and treatment knowledge, medical instruments, hospital and doctor's recommendation.	PHI
Personal Health Condition (PHC)	It refers to objective statement of user's health condition, such as physical indicators level, complications, physical reaction after medication.	PHI
Personal Living Habits (PLH)	It refers to the strategy of fighting disease in daily life, such as users' eating habit, exercise plan.	PHI
Personal Emotional Information (PEI)	It refers to the mood and mental statement of participants.	PHI
Personal Other Experience (POE)	It refers to stories related to illness but not specifically on medical aspects.	PHI

In conclusion, there are various privacy information leakage in OHCs, closely related to people's online behavior. It's necessary to figure out the extent of privacy disclosure while participants are enjoying the convenience of OHCs. Hence, the first research question we want to address in this paper is:

RQ1 :What kind of private information would be exposed most in OHCs, and how likely would such information be exposed?

The Pew Research Center reported that roughly 60% Americans believe it is not possible to go through daily life without having their data being collected [13]. And we cannot ignore that online privacy literacy is an important mediator to a safe and privacy-enhancing online behavior [14].When people want to improve their privacy online, they need to pay more attention on their own behavior and privacy literacy as well. Accordingly, we propose the next research question:

RQ2: How does privacy disclosure related to the users' own online activities?

In this research, users' own online activities do not only include users' Q&A from 4 diabetes sub-communities, but also contain users' behaviors in other communities in "Zhihu", such as raising questions, providing answers, writing articles, and collecting columns. The success of OHCs in promoting health, however, depends not just on posting activity by participants, but, crucially, on whether or not responses are subsequently received [15]. Prior study has found that the informational support and emotional support are closely relevant to users' intention of PHI disclosure [16]. But from a community perspective, there are still some other factors that are not taken into account, such as other's attention, community policy. Therefore, we raise the last research question:

RQ3: What are the influential factors of users' privacy disclosure behavior from the perspective of the community, *and how do they impact?*

2.2 Communication privacy management theory

Communication Privacy Management Theory (CPM) points out that individuals developing rules could maximize the benefits while minimizing the risks of disclosure [17]. The author also explains how and why people decide to reveal or conceal private information across various contexts [17]. Although CPM was proposed to explain the people's privacy strategy in interpersonal relationships, it also could be applied to other expanding areas [18]. In online community, information revealing will bring both good and harm as in face-to-face contexts, including more social support, visibility, potential possibility of network tracking, advertisement recommendation, and more permeable privacy boundaries.

Many studies have analyzed online users' behavior based on CPM. There are findings indicate that online consumers do set boundaries around their privacy information and develop regulations to choose when to disclose information in agreement with CPM theory [19]. Besides, an empirical investigation on Twitter derived from CPM theory shows that "Control and Boundary Rules" of private information on Twitter predict user daily online time significantly [20]. Users' privacy concern and behavior researches based on CPM theory are arousing as well. Some researchers analyze the narratives and posts of a pro-anorexia website, "prettythin.com", using CPM theory as a framework to explain how people balance their competing needs for revealing and concealing [21].

In terms of this study, we use the categorization criteria for different influence factors provided by CPM and propose hypotheses based on CPM. In the privacy information control part, two types of criteria, core and catalyst play a decisive role in privacy rules. In other words, the criteria influence people on whether to share or conceal their privacy information [18]. Combining the factors mentioned before in CPM with some new elements relevant to nowadays users' online privacy disclosure behavior, we design the new core and catalyst criteria.

Table 2. Criteria

Category	Explanation	Type
User Characteristics	It covers the gender and culture factors, like education background, business, residence.	Core
Peer Attention	This means the degree of attention of other users to someone. Usually the more follower count people get, the more possible to be visited. And the voted-up count, thanked count, collected count could also measure the other users' acknowledgement.	Catalyst
Information Support	It refers to the help in aspect of medical information and other objective health-related info, and the answer count and comment count beneath user questions could measure this.	Catalyst
Emotional Support	It refers to the help from emotional level, such as encouragement, empathy. The answer count and comment count beneath user questions could measure this.	Catalyst

As CPM has verified that gender has a great impact on privacy management in off-line communication [17], we assume that it also works in online environment, so here is the first hypothesis:

HP1: Females are more likely to reveal their privacy information than males in OHC.

In core criteria, education level cannot be ignored too. People with more knowledge will have a better grasp of the privacy policy and will also be more aware of the dangers of privacy leak. We assume that users with a higher education level would show more privacy literacy and we propose the second hypothesis:

HP2: Education level is negatively correlated to privacy disclosure in OHC.

In terms of catalyst criteria, when people gain more peer attention or support, the ratio of risk-benefit in privacy revealing would fluctuate. Previous study has noted that through answering, commenting and liking, people are connected and forming a social network, in which the number of friends is related to his or her online activities. In this study, we intend to use the count of followers and accumulative likes received as indicators of peer attention and try to identify their impact on privacy disclosure. Meanwhile, some researchers have analyzed users' behavior from a social support perspective and indicated that what users get from the community would influence their behavior as well [6]. As it has been mentioned in CPM that people will change or create privacy rules in order to response to the new situation [17], we take peer attention and support into account and propose the following hypotheses:

HP3: Peer attention is positively related to privacy disclosure in OHC.

HP4: Information support is positively related to privacy disclosure in OHC.

HP5: Emotional support is positively related to privacy disclosure in OHC.

According to statements above, we build a conceptual model on identifying influential factors on users' privacy disclosure behaviors (shown as Figure 1).

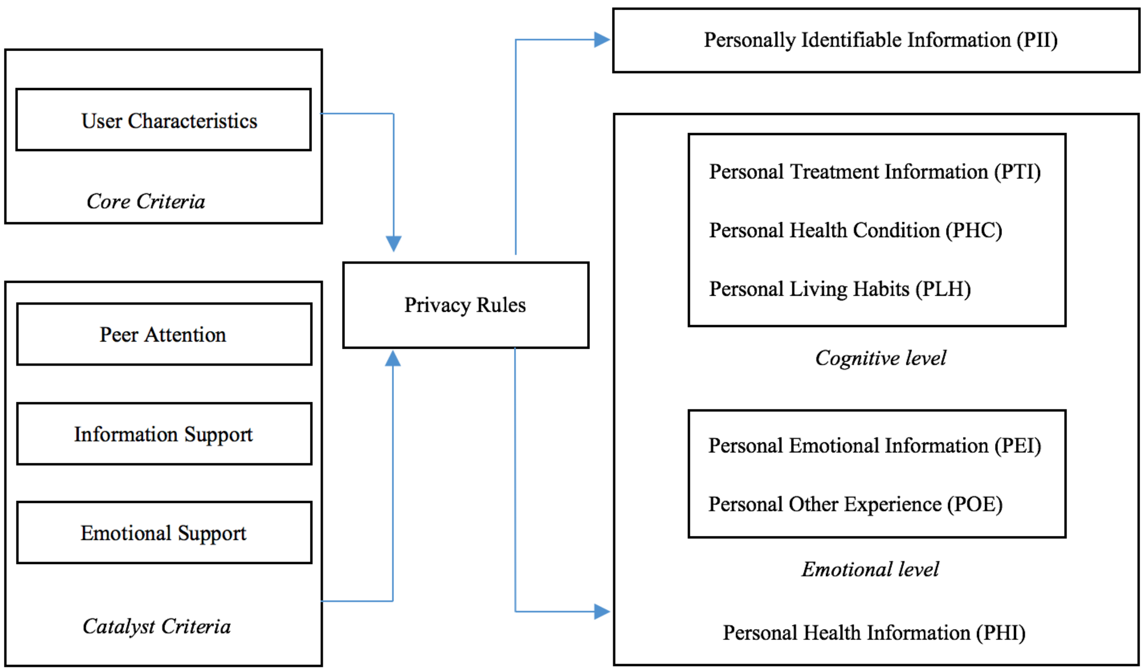

Figure 1. The mechanism of OHCs user's privacy disclosure

3. DATA COLLECTION

"Zhihu" is a well-known online Q & A community, whose diabetes sub-communities are also very popular, the following figures show us what these communities look like and the main activities of people in them.

Figure 2. The questions and answers in diabetes sub-community of "Zhihu"

In order to get an overall comprehension of users' privacy disclosure in their own generated contents, we collected all questions and corresponding answers from 4 sub communities of "Zhihu", the subject of which is diabetes, and including diabetes subtopics, insulin, diabetic diet, diabetic complications as well.

Table 3. Question Number of Different Sub-communities

Theme	Question Number	Theme	Question Number
Diabetes	5,096	Diabetic Diet	326
Insulin	543	Diabetic Complications	878
Grand total	6,843		

Furthermore, we have gathered all the relevant 6,782 non-anonymous respondents' public information from "Zhihu", such as questions, answers, articles, interests and basic profiles.

Table 4. User Generated Contents

Theme	Number	Content
Basic Profiles	6,471	1. Gender, business, education background, employments info, self-description, location. 2. Collected count, thanked count, voted-up count, follower count. 3. Following column count, following question count, following topic count, following count. 4. Questions count, answers count, columns count, lives count, independent articles count, pins count.
Interest	6,412	1. Following topics name and introduction. 2. Following columns title and description. 3. Following questions information. 4. Public favorite pages lists name and introduction. 5. Sponsoring lives subject and description.
Questions	6,469	User question id, title, answer count, comment count, follower count, belonging topics, time.
Answers	6,460	Question title, answer contents, comment count, voted-up count, collected count, thanked count, time.
Articles	6,433	Article title, article contents, comment count, voted-up count, belonging column, time.

Figure 3. The main activities of people in diabetes sub-community of “Zhihu

4. PRELIMINARY RESULTS AND FOLLOWING RESEARCH AGENDA

This study conducted a rough summary on prior works of privacy disclosure in online community and the CPM theory. We identified a research gap of OHCs users’ privacy management and further built a conceptual framework based on CPM theory. We assumed that user characteristics, peer attention, information support and emotional support effect people’s decision on whether to conceal or reveal their privacy information.

In terms of data analysis, we are going to take the following approaches. First, we aim to apply the text mining on all posts and tag them on the basis of the health privacy information definition in OHCs in table 1. Based on all the questions and answers in 4 communities we’ve mentioned above, we want to analyze the disclosure of private information and make response to our first research question, finding out the extent of privacy information disclosure. Second, we will analyze 6,782 non-anonymous respondents based on all of their interests, questions, answers, articles, rather than behaviors in the diabetes communities only, exploring the possibility of privacy disclosure and correlation between their own activities and privacy disclosure degree through a logistic regression. Third, according to the conceptual model we have built, we want to focus on the members who have both raised questions and provided answers in the 4 research communities and explore the impact of feedback from other users in the community in different time periods, separately calculating the index value of user characteristics, peer attention, information support, emotional support. Meanwhile, we want to set up the revealing degree of PII and PHI as the dependent variables, with the criteria as independent variables, identifying the influential mechanism of impact factors and calculating the coefficient of influence between different factors by conducting regression analysis. Furthermore, we will test some moderating effects cause the catalyst criteria's performance may differ due to the different core criteria.

ACKNOWLEDGEMENT

The study is supported by Program for Innovation Research in Central University of Finance and Economics.

REFERENCES

[1] China Association for Science and Technology. (2020). Chinese Internet Users' Popular Science Demand Search Report(Q1 2019). http://www.cast.org.cn/art/2019/4/26/art_1281_94546.html.

[2] Yan, Z., Wang, T., Chen, Y., & Zhang, H. (2016). Knowledge sharing in online health communities: A social exchange theory perspective. Information & Management, 53(5), 643-653.

[3] Acquisti, A. (2004, May). Privacy in electronic commerce and the economics of immediate gratification. In Proceedings of the 5th ACM conference on Electronic commerce (pp. 21-29). ACM.

[4] Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Computers & security, 64, 122-134.

[5] Zhang, X., Liu, S., Deng, Z., & Chen, X. (2017). Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. Computers in Human Behavior, 75, 797-810.

[6] Wang, X., Zhao, K., & Street, N. (2017). Analyzing and predicting user participations in online health communities: a social support perspective. Journal of medical Internet research, 19(4), e130.

[7] Lopez, O., & Hess, T. J. (2019). Spurring Participation: The Effect of Privacy Control Mechanisms on Trust and Self-Disclosure in Online Health Communities.

[8] Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. First Monday, 11(9).

[9] Liau, A. K., Khoo, A., & Hwaang, P. (2005). Factors influencing adolescents engagement in risky internet behavior. CyberPsychology & Behavior, 8(6), 513-520.

[10] Xiao, N., Sharman, R., Rao, H. R., & Upadhyaya, S. (2014). Factors influencing online health information search: An empirical analysis of a national cancer-related survey. Decision Support Systems, 57, 417-427.

[11] Oh, S. (2012). The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. Journal of the American Society for Information Science and Technology, 63(3), 543-557.

[12] Wilson, T. D. (1999). Models in information behaviour research. Journal of documentation, 55(3), 249-270.

[13] Pew Research Center.(2019).Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information.

[14] Bartsch, M., & Dienlin, T. (2016). Control your Facebook: An analysis of online privacy literacy. Computers in Human Behavior, 56, 147-154.

[15] Chen, L., Baird, A., & Straub, D. (2019). A linguistic signaling model of social support exchange in online health communities. Decision Support Systems, 113233.

[16] Zhang, X., Liu, S., Chen, X., Wang, L., Gao, B., & Zhu, Q. (2018). Health information privacy concerns, antecedents, and information disclosure intention in online health communities. Information & Management, 55(4), 482-493.

[17] Petronio, S. (2002). Boundaries of privacy: Dialectics of disclosure. Suny Press.

[18] Petronio, S. (2015). Communication privacy management theory. The International Encyclopedia of Interpersonal Communication, 1-9.

[19] Metzger, M. J. (2007). Communication privacy management in electronic commerce. Journal of Computer-Mediated Communication, 12(2), 335-361.

[20] Yang, K. C., Pulido, A., & Yowei, K. A. N. G. (2016). Exploring the Relationship between Privacy Concerns and Social Media Use among College Students: A Communication Privacy Management Perspective. Intercultural Communication Studies, 25(2).

[21] Herrman, A. R., & Tenzek, K. E. (2017). Communication privacy management: A thematic analysis of revealing and concealing eating disorders in an online community. Qualitative Research Reports in Communication, 18(1), 54-63.