A Privacy Protection Model for Online Social Networks

Javed Ahmed

University of Luxembourg, Luxembourg

Abstract. Online Social Networks (OSNs) have become an important part of daily digital interactions for more than half a billion users around the world. Unconstrained by physical spaces, OSNs offer to web users new interesting means to communicate, interact, and socialize. While these networks make frequent data sharing and inter-user communications instantly possible, privacy-related issues are their obvious much discussed immediate consequences. Recent research identifies a growing privacy problem that exists within OSNs. Several studies have shown how easily strangers can extract personal data about users from the OSNs. There is need for automatic and easy to use privacy protection mechanism. We propose social interaction based audience segregation model for online social networks. Our model uses type, frequency, and initiation factor of social interactions to calculate relationship strength. This model mimics real life interaction patterns and makes online social networks more privacy friendly.

1 Introduction

The Internet has become an inevitable part of lives of people today. Online social networks (Facebook, LinkedIn, Twitter etc) are top most visited sites on internet.[1] These sites are an easy and cost effective way for people to reach out to their classmates, friends and family from across the globe. A large percentage of success of these social networking sites can be attributed to a fact that they give users the opportunity to create their own space and a great way to connect with likeminded people, learn and share knowledge. Online social networks are one of the most popular fora for self representation and user interactions. Individuals join social networks to present themselves. In OSNs user can present themselves by constructing a profile. A profile is a digital representation of an OSN user. A Profile contains huge amount of personal information about the user. According to Grimmelmann [1] Facebook knows an immense amount about its users. A fully filled-out Facebook profile contains about 40 pieces of recognizably personal information, by the time you are done, Facebook has a reasonably comprehensive snapshot both of who you are and of who you know. Additionally, these users are engaged in various social interactions with other users. All these activities are recorded on these platforms which can be easily

[1] Alexa http:/www.alexa.com/topsites

analyzed, manipulated, systematized, formalized, classified, and aggregated [2]. This poses a serious privacy threat to OSN users, and that is the main reason privacy is hotly debated topic in research literature [3] [4] [5] [6] [7]. After an extensive analysis of the articles on privacy issues in online social networks, we conclude that OSNs users are unable to control privacy vulnerabilities due to following reasons:

Inflexible Privacy Tools: Privacy tools in online social networks are not flexible enough to protect user data. Most online social networks only allow users to make their data either public or private. Facebook is one of the few online social networks that provide detailed privacy settings. However, privacy interface is too complicated to most of the normal users. The current interface has limited visual feedback, confusing language, and promotes a poor mental model of how the settings affect the profile. Even after modifying settings, users can experience difficulty in ensuring that their settings match the actual desired outcome [8] [9].

Risky Friends: Although friends can enrich the social graph of users, they can also be a source of privacy risk, because a new relationship always implies the release of some personal information to the new friend as well as to friends of the new friend, which are strangers for the user. Online social network users cannot control what others reveal about them. It is possible for information to be passed on without one's consent [10] [11]. For example, Javed and Serena are friends in online social network. Serena is very careful about the privacy. She adopts a policy that conceals all her friends from public. On the other hand, Javed uses a weaker policy that allows any users to view his friends. In this case, Serena's relationship with Javed can still be learned through Javed. We say that privacy conflict occurs as Serena's restrictive policy is violated by Javed's weaker privacy policy. This shows that the user can only control one direction of an inherently bidirectional relationship.

Third Party Applications: Online social networks offer open platforms to enable third party developers to build applications which provide seamless integration of profile data to third party applications. These applications pose serious privacy risk for online social network users because installed applications receive the privileges equal to owner of the profile and can access user's profile data. Third party application developers have access to user's data regardless of the actual application needs. Facebook additionally gives social applications second degree access which means if Javed installs a social application then the application can also request information about Javed's friends and fellow network members. Moreover, users have no control over how third party companies use their personal information [12]

OSN Service Providers: OSN service providers have too much control over user information. The individuals accept terms of privacy policy before creating account on such services. By accepting the terms of the policy, OSN user volunteer to relinquish some known right or privileges they may have. OSN users are unaware of how their personal information is being used, and it is unclear to the users if the OSN is respecting its privacy policy. More-

over, personal information of the user is retained even after the user decide to delete his account. Some of the privacy threat related to OSN service providers are data retention, targeted marketing, selling of data etc [13].

We are addressing risky friends threat to privacy of OSNs users, and propose interaction based audience segregation model for online social networks. The main motivation for this research is providing users with audience segregation model which mimic real life interaction patterns. In everyday life individuals have fine grain control over what kind of information is presented to different audiences. Mirroring similar strategy for online social networks can enhance privacy and give user more control on his personal information. The binary nature of a relationship in OSNs make privacy uncontrollable. The relationship strength is crucial factor to decide what to reveal and whom to reveal. This research is step towards providing OSN users with tools to manage their relationship in similar ways as they do in real life. This doctoral synopsis is organized as follows. Section 2 discusses research problem and identify research questions to address this problem. Section 3 presents preliminary interactions based audience segregation model, and Section 4 covers state of art related to the privacy problem of OSNs. Finally, Section 5 concludes the doctoral synopsis providing directions for future work.

2 Research Problem

Exponential growth of online social networks resulted in fundamental shift in status of end users. Individual end users become content managers instead of just being content consumers. Today, for every single piece of data shared on OSNs, the uploader must decide which of his friends should be able to access the data. In OSNs, term "friend" has become all-encompassing, it has become increasingly difficult for users to control which friends get to see what personal information. Several studies on Facebook usage have shown that the average number of friends per user is approximately 150. Anyone can make a request to join a user's friend circle–family members, colleagues, classmates, acquaintances, strangers etc. Current literature support the claim that users are willing to add strangers to their friend circle [14]. However, allowing strangers to join user's friend circle can lead to a number of privacy risks [10]. Most of the OSNs provide users with binary relational ties (e.g., friends or stranger) [15]. This binary indicator provides only a coarse indication of the nature of the relationship. In reality human relationships are much more complicated than a single binary relational tie. There is need for segregation of friends according to the strength of relational ties. Some of the social networking sites have begun providing friend-lists feature, in order to help users in organizing a large friend network into groups. Grouping several hundred friends into different lists, however, can be a laborious process; on what basis should users construct the friend-lists? And even if the user were to group friends into lists, are these lists meaningful for setting privacy policies? To alleviate the burden of constructing meaningful lists manually, we propose interaction based audience segregation model for online

social networks. The estimation of friendship interaction intensity among OSN users and its classification based on different level of intensity can be quite useful for identifying privacy threat from individuals added as friends. The social web is kind of virtual society that exhibits many of the characteristics of real societies in term of forming relationships and how those relationships are utilized. In real societies, the relationship strength is a crucial factor for individuals while deciding the boundaries of their privacy. Moreover, this subjective feeling is quite efficiently utilized by humans to decide various other privacy related aspects such as what to reveal and whom to reveal. The main question for this research is how interactions of users determine tie strength and implement privacy in online social networks. More specifically, we want to explore whether a user's interaction with his friends can be used as a basis for making data access decision for that user. To answer this question, we need to understand nature of privacy in online social networks and dynamics of interactions intensity for OSN users. We break main research question into three sub questions:

- How to measure privacy risk associated with social graph of OSN users?
- How to construct interaction graph by quantifying users interactions in OSN?
- How to segregate audience on the basis of interaction graph in OSN?

From our first research question, we quantify the privacy risk attributed to friend relationship in online social networks. We show that risky friends can reveal user personal information unintentionally in online social networks. Second research question deals with user's interaction patterns in online social networks. We show that users tend to interact mostly with small subset of friends, often having no interactions with majority of their friends in online social networks. This cast doubts on the practice of extracting meaningful relationships from social graphs. We suggest interaction based model for validating user relationships in online social networks. Third research question deals with audience segregation. We consider social interactions as currency to estimate friendship strength and perform audience segregation. Providing users with audience segregation mechanism would improve the quality of interactions and self presentations.

3 Our Approach

We propose interaction based audience segregation model for online social networks. We consider interaction intensity as a proxy for relationship quality. It is used as currency for making data access decisions in online social networks. Current online social networks assume binary, symmetric relationship of equal value between all directly connected OSN users. In real world an individual has relationships of vary quality with his friends. Providing OSN users a mechanism which mimic real life interaction patterns to larger extent would improve self presentation, and reduce privacy risks. It will also enable users to avoid social convergence, and provide users opportunity to present different sides of themselves to different audiences. Our model considers several factors to identify relationship quality such as type, frequency and interaction initiation. We

describe in detail all these aspects of interactions to understand the usefulness of our approach.

The type of interaction is quite important in order to calculate friendship strength because an individual choose an interaction type according to the nature of relationship with its target audience. Hence, the interaction type defines the intimacy, openness, sensitivity as well as strength of relationship between communicating parties. Some of the interaction types are preferred to communicate with close friends, whereas the others to interact with ordinary friends. Hence, all interaction types cannot be given similar weight in estimation of relationship strength. Each interaction type is given a numerical weight in order to increase or decrease its contribution in development relationship strength. Our computation model take into considerations latent as well as active interaction types. The latent interactions are non-reciprocal in nature such as profile visits, whereas active interactions are visible actions such as wall posts and comments. The active interactions can be further classified into real time as well as non-real time interactions. The real time interaction requires the presence of interacting parties and examples of such interaction is chatting. Private messaging and status updates can be classified as non-real time interactions. Apart from active interactions based measures, we can use latent interactions to calculate friendship strength. Latent interactions are more prevalent and frequent in online social networks. Profile visits is a latent interaction and it is very frequent in nature in online social networks. It can be a measure for friendship strength estimation. Mutual friends can be another important measure for friendship strength estimation. Many common friends lead to the fact that individuals are strongly connected with each other, or they share same context such colleagues, family etc.

The interaction count refer to the total number of interactions between an individual and his friends within certain period of time. The frequency of interaction demonstrates the willingness of the user to communicate with his friends. The interaction initiation aspect is very important to understand relationship strength. Some times an individual user is spammed with a lot of interactions initiated by his friends, but his response to that communication determines his willingness to interact. So, we categorize interactions initiation factor in following two ways.

Initiated Interactions These interactions are initiated by the user with his friends. These interactions have more weight in developing relationship strength because the user is willing to communicate and collaborate with his friends.

Received Interactions These interactions are received by the user from his social circle. These interactions have less weight in developing relationship strength because willingness of communication and collaboration is coming from friends. We chose to focus on interactions initiated by the user to limit the inflationary effect of message senders. Some users can artificially boost their status with a particular friend by frequently interact with him.

We consider interactions as a very strong indicator for audience segregation. Our model calculates interaction intensity that can be useful in audience segregation.

4 Related Work

The development of usable, fine grained tools for protecting personal data is serious emerging problem in online social networks. Kelley et al. [16] have done preliminary work towards investigating how users create friend groups in Facebook. They have examined four different methods of friend grouping and their results show that the type of mechanism used, affects the groups created. Their findings lead to a number of recommendations for designing group-based privacy controls for online social networks. Adu-Oppong et al. [17] have proposed partitioning a user's friends into lists based on communities extracted automatically from the network, as a way to simplify the specification of privacy policies. Mazzia et al. [18] built a policy visualization tool that extracts and presents the user's communities to help him in managing his group based privacy policies. Squicciarini et al. [19] [20] propose an approach to facilitate online social network users to group their contacts into social circles with common interests. The authors design a multi-criteria model that takes into account multiple aspects of user's profiles, and automatically groups each user's contacts into social circles with common characteristics. Users in the same social circle (group) have similar behavior, such as similar education background, hobbies, and similar privacy preferences. The authors further propose an approach to recommend privacy policies for newly uploaded data items or newly added contacts. Fang et al. [21] [22] propose the privacy wizard for social networking sites. The goal of the wizard is to automatically configure a user's privacy settings with minimal effort from the user. The wizard is based on the underlying observation that real users conceive their privacy preferences based on an implicit structure. Thus, after asking the user a limited number of carefully chosen questions, it is usually possible to build a machine learning model that accurately predicts the user's privacy preferences. Cetto et al.[23] introduce a serious game that allows its users to playfully increase their privacy awareness on Facebook. The conceptual design of the game is based on two foundations: firstly, an in-depth understanding of privacy awareness as the match or mismatch between perceived and actual visibility of shared items. Secondly, an inductive learning approach that allows its users to experiment and play with their own Facebook data in order to actively learn about the visibility of their personal items.

One of the research studies closely related to our work is done by Lerone et al. [24]. The authors have introduced interaction count based approach to determine relationship strength. In this approach, the authors simply take into consideration three types of interactions and count them in order to calculate relationship strength. The interaction intensity model by Lerone et al. [24] do not differentiate interactions on the basis of initiative, so it is possible that a malicious user intentionally perform larger number of interactions to get access to

user's sensitive profile information. Our model takes into consideration this issue and resolve it by assign more weight to interactions initiated by user himself. Our interaction intensity model has another advantage over Lerone's model that we consider all possible type of interactions.

Waqar et al. [25] extend work of Lerone et al. by applying data mining model to calculate relationship strength for online social networks, Whereas, this data mining model is not validated on real OSNs data. The authors also conduct online survey to analyze Facebook user's interaction behavior with their friends. Xiang et al. [15] propose a model to infer relationship strength based on profile similarity and interaction activity. The authors compute three features to determine profile similarity. These features are: common group, common friends, and logarithms of the normalized counts of common networks. In addition to profile similarity features, the authors consider wall posting, and photo tagging for interaction activity. Our approach is different from their approach because of two reasons: 1. We take into consideration broader set of interactions types. 2. We develop intensity scale for all interaction types. This intensity scale has vital role in computation of relationship strength.

Lizi et al. [26] propose interaction ranking based trustworthy friend recommendation model. This model is able to effectively recommend trustworthy friends to community members by taking into consideration four interaction attributes: reply frequency, comment length, time difference, and domain similarity. Another interesting work by the authors [27] propose trust ranking based recommendation model for suggesting the most trustworthy community members. The authors investigate four new interaction attributes that influence trust in virtual communities. These interaction attributes are interaction quality, seriousness in interactions, consistency over a long period, and common interest. The author's hypothesis is validated by processing real data collected from Slashdot. A recent work related to friend recommendation is done by Zhao et al. [28]. The authors propose scalable and explainable friend recommendation model for social network systems. This model takes multiple relationship factors into account such as common friends, common followed users, common followers, and common joined groups of the target user and the candidate for friend recommendation. Our research work is not focused on recommending new friends, but identifying the strength of relationship among existing friends.

The majority of online social networks offer second degree access which means a friend of a friend is able to access the user's personal information. According to Cuneyt et al. [10] friends can be source of privacy risk because this relationship always implies the release of some personal information not only to friends, but also to friends of a friend, which are strangers for the users. Akcora et al.[11] propose a risk measure for OSNs. The aim is to associate a risk level with social network users in order to provide other users with a measure of how much it might be risky, in terms of disclosure of private information, to have interactions with them. The authors compute risk levels based on similarity and benefit measures, by also taking into account the user risk attitudes. In particular, The authors adopt an active learning approach for risk estimation, where user risk

attitude is learned from few required user interactions. Another interesting fact demonstrated by Frank et al. [29] that more users are willing to divulge personal details to an adversary if there is a mutual friend connected to the adversary and the user. Christo et al. [30] show that users tend to interact mostly with small subset of friends, often having no interactions with up to 50 percent of their friends. The authors suggest a model for representing user relationships based on user interactions. Existing research literature supports our idea that all friends should not be give equal access to user personal information, but access to personal information should be administrated based on relationship strength among online social network users.

5 Conclusion and Future Work

We propose social interaction based audience segregation model which mimic real life interaction patterns to larger extent. We also identify the impact of various social interactions available to users in online social networks. There are three main innovative aspects of our model. First of all, it consider all possible of set interactions among friends. Secondly, the model considers the direction of interaction either from user to friend or vice versa. Finally, all interaction types are assigned a numerical weight in order to increase or decrease its contribution in interaction intensity calculation based on its importance in the development of relationship ties.

In future, we plan to conduct formal study of user interaction behavior and sharing patterns. This study will provide us basis for assigning different weight to social interactions and ranking profile items on the basis of their sensitivity. In the next phase, we will develop formal model and proof of concept prototype to validate of our hypothesis.

References

1. James Grimmelmann. Facebook and the social dynamics of privacy. *Iowa Law Review*, 95(4):1–52, 2009.
2. Bibi van den Berg, Stefanie Pötzsch, Ronald Leenes, Katrin Borcea-Pfitzmann, and Filipe Beato. Privacy in social software. In *Privacy and Identity Management for Life*, pages 33–60. Springer, 2011.
3. Justin Lee Becker and Hao Chen. *Measuring privacy risk in online social networks*. PhD thesis, University of California, Davis, 2009.
4. Ai Ho, Abdou Maiga, and Esma Aïmeur. Privacy protection issues in social networking sites. In *Computer Systems and Applications, 2009. AICCSA 2009. IEEE/ACS International Conference on*, pages 271–278. IEEE, 2009.
5. Giles Hogben. Security issues and recommendations for online social networks. *ENISA position paper*, 2007.
6. Balachander Krishnamurthy and Craig E Wills. Characterizing privacy in online social networks. In *Proceedings of the first workshop on Online social networks*, pages 37–42. ACM, 2008.

7. Chi Zhang, Jinyuan Sun, Xiaoyan Zhu, and Yuguang Fang. Privacy and security for online social networks: challenges and opportunities. *Network, IEEE*, 24(4):13–18, 2010.
8. Heather Richter Lipford, Andrew Besmer, and Jason Watson. Understanding privacy settings in facebook with an audience view. *UPSEC*, 8:1–8, 2008.
9. Cuneyt Gurcan Akcora and Elena Ferrari. Graphical user interfaces for privacy settings.
10. Cuneyt Gurcan Akcora, Barbara Carminati, and Elena Ferrari. Risks of friendships on social networks. *arXiv preprint arXiv:1210.3234*, 2012.
11. Cuneyt Gurcan Akcora, Barbara Carminati, and Elena Ferrari. Privacy in social networks: How risky is your social graph? In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, pages 9–19. IEEE, 2012.
12. Javed Ahmed and Zubair Ahmed Shaikh. Privacy issues in social networking platforms: comparative study of facebook developers platform and opensocial. In *Computer Networks and Information Technology (ICCNIT), 2011 International Conference on*, pages 179–183. IEEE, 2011.
13. Michael Beye, Arjan JP Jeckmans, Zekeriya Erkin, Pieter Hartel, Reginald L Lagendijk, and Qiang Tang. Privacy in online social networks. In *Computational Social Networks*, pages 87–113. Springer, 2012.
14. Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80. ACM, 2005.
15. Rongjing Xiang, Jennifer Neville, and Monica Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 981–990. ACM, 2010.
16. Patrick Gage Kelley, Robin Brewer, Yael Mayer, Lorrie Faith Cranor, and Norman Sadeh. An investigation into facebook friend grouping. In *Human-Computer Interaction–INTERACT 2011*, pages 216–233. Springer, 2011.
17. Fabeah Adu-Oppong, Casey K Gardiner, Apu Kapadia, and Patrick P Tsang. Social circles: Tackling privacy in social networks. In *Symposium on Usable Privacy and Security (SOUPS)*, 2008.
18. Alessandra Mazzia, Kristen LeFevre, and Eytan Adar. The pviz comprehension tool for social network privacy settings. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 13. ACM, 2012.
19. Anna Cinzia Squicciarini, Dan Lin, Sushama Karumanchi, and Nicole DeSisto. Automatic social group organization and privacy management. In *CollaborateCom*, pages 89–96, 2012.
20. Anna Squicciarini, Sushama Karumanchi, Dan Lin, and Nicole DeSisto. Identifying hidden social circles for advanced privacy configuration. *Computers & Security*, 41:40–51, 2014.
21. Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web*, pages 351–360. ACM, 2010.
22. Lujun Fang, Heedo Kim, Kristen LeFevre, and Aaron Tami. A privacy recommendation wizard for users of social networking sites. In *Proceedings of the 17th ACM conference on Computer and communications security*, pages 630–632. ACM, 2010.
23. Alexandra Cetto, Michael Netter, Günther Pernul, Christian Richthammer, Moritz Riesner, Christian Roth, and Johannes Sänger. Friend inspector: A serious game to enhance privacy awareness in social networks. *arXiv preprint arXiv:1402.5878*, 2014.

24. Lerone Banks and Shyhtsun Felix Wu. All friends are not created equal: An interaction intensity based approach to privacy in online social networks. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 970–974. IEEE, 2009.
25. Waqar Ahmad, Asim Riaz, Henric Johnson, and Niklas Lavesson. Predicting friendship intensity in online social networks. In *21st International Tyrrhenian Workshop on Digital Communications*, 2010.
26. Lizi Zhang, Hui Fang, Wee Keong Ng, and Jie Zhang. Intrank: Interaction ranking-based trustworthy friend recommendation. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, pages 266–273. IEEE, 2011.
27. Lizi Zhang, Cheun Pin Tan, Siyi Li, Hui Fang, Pramodh Rai, Yao Chen, Rohit Luthra, Wee Keong Ng, and Jie Zhang. The influence of interaction attributes on trust in virtual communities. In *Advances in User Modeling*, pages 268–279. Springer, 2012.
28. Zhao Du, Lantao Hu, Xiaolong Fu, and Yongqi Liu. Scalable and explainable friend recommendation in campus social network system. In *Frontier and Future Development of Information Technology in Medicine and Education*, pages 457–466. Springer, 2014.
29. Frank Nagle and Lisa Singh. Can friends be trusted? exploring privacy in online social networks. In *Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in*, pages 312–315. IEEE, 2009.
30. Christo Wilson, Bryce Boe, Alessandra Sala, Krishna PN Puttaswamy, and Ben Y Zhao. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, pages 205–218. Acm, 2009.