SocialViz: Understanding Privacy Through History and Context

Morris Hopkins, Mauricio Castaneda, Swapneel Sheth, Gail Kaiser
Department of Computer Science
Columbia University
New York, NY, USA
{mah2250, mc3683}@columbia.edu, {swapneel, kaiser}@cs.columbia.edu

ABSTRACT

Social network platforms provide users with a rich set of options for sharing information and communicating with platform connections. However, they provide users with few options to view and understand the overall privacy settings for content being shared. In particular, current privacy configuration systems do not allow users to view their settings in the context of a larger peer group, nor do they provide users with a historical view of their privacy settings. As a result there is often a disparity between what content users believe they are sharing, and what is actually being made visible to other users within the network. To address these issues, SocialViz, our end-user oriented tool, provides a simplified understanding of privacy settings through visualizations showing both a contextualized overview of a user's settings and a category specific view illustrating how the privacy of a single category has changed over time. This is done with data visualizations constructed from crowd sourced data collected from 512 users over an 17 month period. Our tool allows end-users to easily view and understand their privacy, to determine if, and how their settings should change. The preliminary results of interviews conducted to test the usefulness of our user oriented tool show that 80% of users find such a tool helpful and 70% reported that they would use such a tool. A video demonstration of our tool can be found at the following url: `http://youtu.be/kmNOnMtEuTM`.

Categories and Subject Descriptors

K.4.1 [**Public Policy Issues**]: Privacy; H.3.3 [**Information Search and Retrieval**]: Information filtering

General Terms

Human Factors

Keywords

privacy, empirical studies, crowdsourcing, data visualization, social networking

1. INTRODUCTION

The past decade has seen an increase in the number of social networks, from traditional ones like Facebook, to more professional ones like LinkedIn. These platforms contain vast amounts of personal data about their users, something that is being viewed as a major concern [13]. At the same time, privacy controls for these systems are complicated and undergo frequent changes, making it hard for users to navigate [5] and configure their settings to limit access to their private information. Platforms such as Facebook are known to have silently modified policies, defaulting users to opt-in to services, making previously private information public [4].

Current privacy configuration systems do not allow users to see their settings in the context of a larger peer group. Additionally, current systems do not provide users with a historical view of their privacy settings. Previous research studies in this field have tried to improve user understanding of privacy by providing users with an outsiders view of their data [8]. In our approach, we try to make it easier for users to understand their settings in context of their social circles - family, friends, colleagues, peer groups - rather than in isolation.

This project introduces a tool that uses crowd-sourced data-driven visualizations to improve users' understanding of social network privacy. These visualizations present users settings in the context of peer data. Additionally, they provide historical data that allows end users to view how their settings have evolved over time. This history helps users recognize when their settings have changed so that they can make better informed decisions when changing their privacy settings.

2. RELATED WORK

There has been other research focused on developing tools for simplifying user understanding and configuration of privacy. In this section we will present examples of such research. An automatic system developed by Fang and LeFevre [3] can configure a user's privacy settings in social networking sites, by using a machine learning model that does not require much user input. This system allows users to have automatically customized privacy settings. Another system, developed by Tootoonchian et al. [11], called Lockr, improves privacy of decentralized and centralized on-line content sharing systems by separating the content that is being stored with other functionality provided by the social networking site. An approach using game theory was developed

by Squicciarini et al. [10] where, by extending the notion of content ownership, they proposed a solution that offers an automated way to share images. This system models the problem of collaborative enforcement of privacy policies on shared data. Toubiana et al. [12] developed a geo-location aided tool that allows users to define their photo tagging preferences at the time a picture is taken. The system enforces tagging preferences without revealing user's identities. A tool called PrivAware, developed by Becker and Chen [1] can detect, and reports on, unintended information loss on on-line social networks. The tool also presents users with actions to mitigate privacy risks. Socially acceptable privacy management tools were investigated by Lipford et al. [8], by developing a prototype where profile information is presented in an audience-oriented view. StakeSource is a tool by Lim et al. [7] that helps to gather crowd sourced recommendations about stakeholders of a project using social network analysis and to ensure that all concerns are addressed by providing necessary information to software engineers. In this paper we propose a contextualized view of privacy to assist user configuration.

3. APPROACH

Simplifying user understanding of privacy on social networks must address multiple problems. First, current systems don't allow users to see their settings in the context of the settings of a larger peer group. Without this information, users lack the reference needed to make a fully-informed decision. Second, current systems don't allow users to view historical data about their privacy. Being able to see history allows users to verify their settings are persistent, and allows users to more quickly understand and address problems as they arise. In this section, we outline the motivating research questions and the approach we selected to address these questions.

3.1 Research Questions

Research has shown that end-users, our intended audience, are more interested in transparency than encryption or other software engineering techniques [9]. Taking this into consideration, our tool provides end users with additional information not currently available on social network platforms, with the aim of improving their understanding of privacy. Moreover, we wanted to determine if crowd sourced data could be used to simplify this process. Our tool provides end users with visualizations that can be tailored to their individual needs.

We focused on the following research questions:

- **RQ 1:** Can crowd sourced data visualizations improve transparency of end user privacy settings?
- **RQ 2:** Does providing context and history improve user understanding of privacy?

4. IMPLEMENTATION

This section outlines what dataset was used to build the tool, the visualizations that were selected to give the users a better interpretation of their privacy settings, and the technologies that were used to build this tool.

4.1 Dataset

The data used to feed information into this tool was collected since May 1, 2012. This information was obtained using an automated retrieval tool using the Facebook Graph API. There were an average of 154 scans made per user, on a total of 512 through November 2013. The scans were not constant due to incomplete or badly formated responses, or API outages during this time period. Each scan contains information about the magnitude of the information being shared per category (e.g., number of friends, number of check-ins, etc.), totaling 41 categories. Using this information, we can determine how much data a specific user is sharing, and by combining the data across multiple users, we can determine how much an individual category is being shared.

4.2 Individual Visualizations

The first visualization in this group consists of a line graph that shows an individual user's data from each category over time (Figure 1). Users can confirm their privacy settings, or notice any unexpected changes, by glancing at this visualization. The x-axis shows changes in time, while the y-axis represents how much information is being shared for the selected category. Categories are differentiated by using different colors and distinct anchor shapes. When a specific category returned an error using the Facebook API, the data was mapped to -10. This graph allows a user to have a snapshot in time of what information is being shared.

The second visualization consists of a word cloud. Users can quickly and easily understand what categories are being shared with this visualization. The bigger a word is (the larger its font), the more that category is being shared. If a users do not share a specific category, or if an API error was returned for all users, the title for that category is struck-out, as seen in the "inbox" category. Both visualizations in this category grant users a measure of how much information they are sharing for a specific category. This visualization was omitted due to space limitations. For more information please refer to the full report [9].

4.3 Contextualized Visualizations

The contextualized visualizations allows an individual user to view it's privacy configuration in the context of a larger group of users. The first visualization in this group (Figure 2) consists of a series of donut charts, where each donut represents the ratio of users from the selected group that share each category. The blue portion represents how many users are sharing a category, the orange represents how many users are not sharing data for a category, and the gray represents the percentage of users for which the Facebook Graph API returned an error. API errors usually indicate whether the category has been restricted, or if a user has restricted sharing. Each individual user may know if they are sharing an individual category by looking at the color-coded dot next to the donut, or by identifying the faded out donuts in the donut chart. This visualization was developed because it provides information about what privacy settings are trending for a specific group.

Users were also provided with the ability to view how their privacy settings for a particular category has changed over time. This visualization (Figure 3), includes information about the categories shared by many users over time. The y-axis represents the magnitude of the information being shared, for example, the number of friends a user has, and

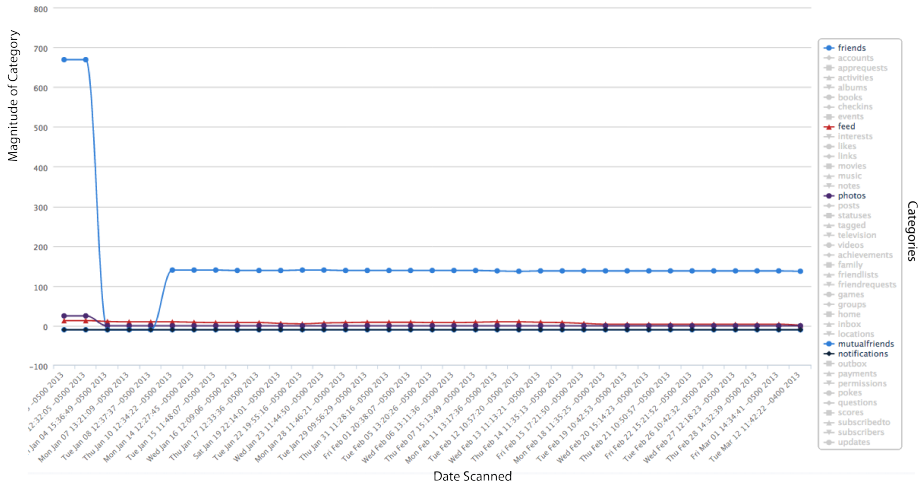

Figure 1: Individual visualization: Visualization of a single user's data from all categories presented over time.

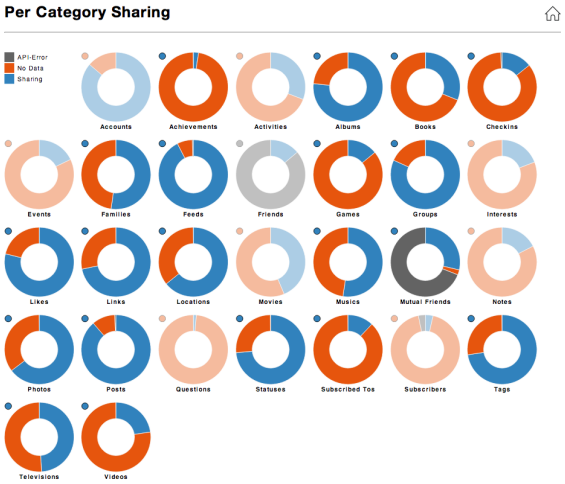

Figure 2: Contextualized visualization: A single user's data from as a snapshot in time (small dot top left corner) presented in the context of the same data collected from 512 users over 17 months (the larger donut)

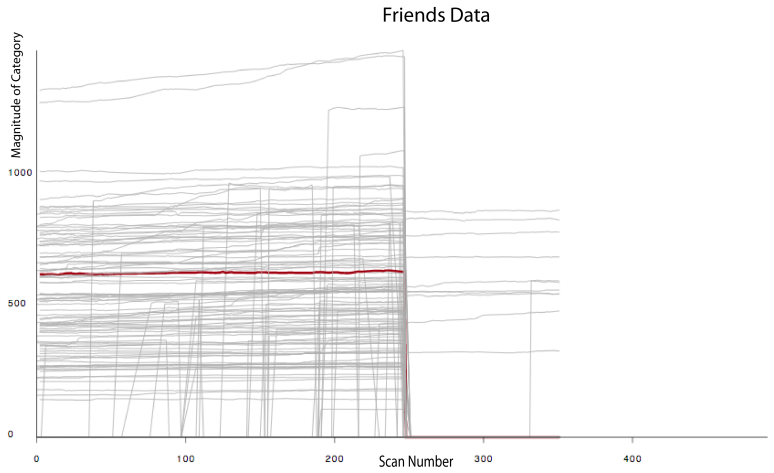

Figure 3: Contextualized visualization: a single user's data, for a single category presented over time, in the context of the same data collected from 512 users over a 17 month period

the date corresponding for that measurement, represented along the x-axis. A highlighted line can be used to further narrow a specific user's data within the larger group. As before, where there was no data available (Facebook API error), we designated a special value, mapping this case to -10. Using this visualization a user can have a good measure of how much information they are sharing compared to a larger group. and it is useful for detecting policy changes, or abrupt changes in a specific users privacy settings.

4.4 Technologies

The tool described in this article was built using a number of available technologies in order to assure a fast development time, and a responsive interface. The dataset mentioned earlier is stored as a group of text documents, each named with a number that identifies each user. This data was imported into a PostgreSQL database, so we could have a logical grouping of the data that was easily and quickly accessible. In the database, the information is stored using a star schema that allows simpler queries and query performance gains. The information is then loaded into a Ruby on Rails application that allow quick development cycles, as well as a web based environment. In order to provide users with a clean, user friendly, and intuitive interface, we used D3.js [2], a data visualization library.

5. EVALUATION

In order to validate the usefulness of our tool and to gain additional insight into how an end user would interact with the application, we conducted interviews that included questions and think aloud responses. From these interviews 80% of the respondents found the visualizations useful and most users agreed that the results shown in the donut visualization did correspond to what they believed their privacy settings to be. Additionally, 70% of respondents said they would use this application at least once, and 90% of the respondents said they would like to get notifications every time the application identifies a change in their privacy settings. Most respondents agreed that the tools successfully improved their understanding of privacy. Additional information and a summary of our research data can be found on our website [6].

6. CONCLUSION

Current social network platforms do not provide end users with the information needed to make informed decisions regarding their privacy. In particular, they do not have a contextualized view of user settings, without out which users lack important reference. Additionally, platforms do not provide end-users with a historical view which would allow them to confirm that their settings remain accurate and persistent. Our tool addresses these issues by presenting contextual and historical privacy data that end users can tailor to their individual needs. This makes user data more transparent and allows users to respond quickly to events that may affect their privacy. Though our tool is still a prototype, it does indicate that context and history improve end user understanding of privacy.

7. ACKNOWLEDGMENTS

We would like to thank all the participants that made this research possible by agreeing to participate our interviews. We would also like to thank Priyank Singhal for his contributions to the project. The authors are members of the Programming Systems Laboratory funded in part by NSF CCF-1302269, CCF-1161079, NSF CNS-0905246, and NIH U54 CA12185.

8. REFERENCES

[1] J. L. Becker and H. Chen. *Measuring privacy risk in online social networks*. PhD thesis, University of California, Davis, 2009.

[2] D3.js. http://d3js.org/.

[3] L. Fang and K. LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 351–360, New York, NY, USA, 2010. ACM.

[4] D. Fletcher. How facebook is redefining privacy, 2010.

[5] V. Goel. Facebook to update privacy policy, but adjusting settings is no easier. *The New York Times*, August 2013. Accessed October 01, 2013.

[6] M. Hopkins, M. Castaneda, S. Sheth, and G. Kaiser. N Heads are Better than One. Technical Report cucs-025-13, Dept. of Computer Science, Columbia University, 2013. `http://mice.cs.columbia.edu/getTechreport.php?techreportID=1552`.

[7] Q. Lim and Finkelstein. Stakesource: Harnessing the power of crowdsourcing and social networks in stakeholder analysis. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering*, ICSE '10, pages 239–242, New York, NY, USA, 2010. ACM.

[8] H. R. Lipford, A. Besmer, and J. Watson. Understanding privacy settings in facebook with an audience view. *UPSEC*, 8:1–8, 2008.

[9] S. Sheth, G. Kaiser, and W. Maalej. Us and Them — A Study of Privacy Requirements Across North America, Asia, and Europe. Technical Report cucs-024-13, Dept. of Computer Science, Columbia University, 2013. `https://mice.cs.columbia.edu/getTechreport.php?techreportID=1551`.

[10] A. C. Squicciarini, M. Shehab, and F. Paci. Collective privacy management in social networks. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 521–530, New York, NY, USA, 2009. ACM.

[11] A. Tootoonchian, S. Saroiu, Y. Ganjali, and A. Wolman. Lockr: better privacy for social networks. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, CoNEXT '09, pages 169–180, New York, NY, USA, 2009. ACM.

[12] V. Toubiana, V. Verdot, B. Christophe, and M. Boussard. Photo-tape: user privacy preferences in photo tagging. In *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion, pages 617–618, New York, NY, USA, 2012. ACM.

[13] A. L. Young and A. Quan-Haase. Information revelation and internet privacy concerns on social network sites: a case study of facebook. In *Proceedings of the fourth international conference on Communities and technologies*, C&T '09, pages 265–274, New York, NY, USA, 2009. ACM.