# **Technical Report**

# **SocialViz: Understanding Privacy Through History and Context**

## Priyank Singhal, Swapneel Sheth, Gail Kaiser

### Department of Computer Science, Columbia University, New York

### ps2721@columbia.edu, {swapneel, kaiser}@cs.columbia.edu

**1. Introduction**

Social networks contain vast amounts of personal data about their users, something that is being viewed as a major concern [1]. Privacy controls for these systems are complicated and undergo frequent changes, making it hard for users to navigate [2] and conﬁgure their privacy settings. Platforms such as Facebook are known to often silently modify policies, defaulting users to opt-in to services, making previously private information public [3]. 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.

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. Additionally, it provides historical data that allows end users to view how their settings have evolved over time, helping them recognize changes and make better informed decisions.

**2. Objectives**

The primary objective of my work was to assist the ongoing project in the following tasks:

2.1. *Online Survey*

The team created an online survey [5] on privacy concerns in social networks which was being used to gather information and data points about user habits, services they use, and their views on privacy. The goal was to collect more responses (at least 50). In the past semesters, the team had collected about 40 responses and the aim was to get more responses for better data analysis.

2.2. *Follow up interviews*

To get a more detailed sense of users’ understanding of privacy and to get feedback about the visual tool, we decided to conduct in-person follow up interviews. These would help verify the information we previously collected. The goal was to conduct 10-15 more interviews (about 10 had already been conducted previously).

2.3. *Submit paper to conference*

The team had been working on a paper submission for the International World Wide Web Conference, 2014. WWW is one of the premier global forum for discussion and debate in regards to the standardization of its associated technologies and their on society and culture. My goal was to assist the team with compiling the information, editing and formatting and ensuring timely submission of the paper.

2.4. *Data Analysis*

The most important objective of the project was actually trying to find interesting trends and verifying hypothesis from the collected information. The goal was also to run statistical data analysis tests to try and validate the significance of the responses.

**3. Methodology**

This section describes the methodology and steps taken to complete each of the objectives.

3.1. *Online Survey*

The first part of my project was to collect more responses. I did so by reaching out to various sections of people, including friends, family, professional connections and students at Columbia. I also tried to gather respondents from online communities dedicated to discussion about privacy and enthusiasts on Reddit, Facebook and Google+. Overall, we managed to collect 104 responses, out of which my contribution was 60 new ones.

3.2. *Follow up interviews*

The follow up interviews were a more detailed way of getting feedback on our study. The team had created a tool that analysis a user’s Facebook profile for visibility of various different categories such as posts, friends, location, photos, contact information, etc. It then creates some easy to understand infographics to give users social and historical context on their privacy settings. This helps to augment their decision in changing privacy settings as they feel are best suited. The follow up also asks more granular information about the type of privacy control mechanisms users prefer. Overall, we managed to collect 17 respondents, with my contribution being 7 users.

3.3. *Paper submission to ICSE*

I worked on formatting and editing the paper for the WWW ’14 conference, my major contribution was towards the submission for International Conference on Software Engineering (ICSE) ’14. The team decided to submit to formal demo workshop, which included a 4 page paper along with a video of the tool in action. I was responsible for drafting the introduction, related work, evaluation and conclusion sections.

3.4. *Data Analysis*

From the data collected in the online and the follow up surveys, I tried to come up with some hypothesis about certain trends. For these hypothesis, I then queried data using an in-built tool in our survey software. Then, exporting this data to Excel, I reformatted it to make it usable. Finally, after verifying the hypotheses, I tried to find their statistical significance. This was done using 2 types of statistical analysis:

3.4.1. *T-test*

A t-test is a statistical hypothesis test in which the test statistic follows a Student's t distribution [7] if the null hypothesis is supported. It can be used to determine if two sets of data are significantly different from each other, and is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. More specifically, the paired t-test was used since our data is a sample of matched pairs of similar units.

3.4.2. *Z-test*

A Z-test [8] is a statistical test in which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. Because of the central limit theorem, many test statistics are approximately normally distributed for large samples. Many statistical tests can be conveniently performed as approximate Z-tests if the sample size is large (generally >30) or the population variance known. Here again, the paired sample z-test is more applicable for our dataset.

**4. Data Analysis**

This section describes a few hypotheses that we considered, and running tests to prove if they are statistically significant.

Hypothesis 1:

**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 set- tings. 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 [9].

**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 con- textual and historical privacy data that end users can tailor to their individual needs. This makes user data more trans- parent and allows users to respond quickly to events that may aﬀect their privacy. Though our tool is still a proto- type, it does indicate that context and history improve end user understanding of privacy.

**References**

[1] 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, pages 265–274, New York, NY, USA, 2009. ACM.

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

[3] D. Fletcher. How Facebook is redeﬁning privacy, 2010.

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

[5] Social Networks and Online Privacy Concerns, Dept. of Computer Science, Columbia University, 2013. http://psl.cs.columbia.edu/wp-content/uploads/2013/10/Survey.pdf

[6] Follow-up Social Networks and Online Privacy Concerns Survey, Dept. of Computer Science, Columbia University, 2013. http://psl.cs.columbia.edu/wp-content/uploads/2013/10/Follow-up.pdf

[7] Student’s t-test, Wikipedia. http://en.wikipedia.org/wiki/Student's\_t-test

[8] Z-test, Wikipedia. http://en.wikipedia.org/wiki/Z-test

[9] 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.