
Understanding ethical concerns in social media privacy studies

Sam McNeilly

School of Computer Science
University of St Andrews
St Andrews, Fife, Scotland
sm2269@st-andrews.ac.uk

Luke Hutton

School of Computer Science
University of St Andrews
St Andrews, Fife, Scotland
lh49@st-andrews.ac.uk

Tristan Henderson

School of Computer Science
University of St Andrews
St Andrews, Fife, Scotland
tnhh@st-andrews.ac.uk

Abstract

There are myriad ethical considerations with conducting social media studies, in particular those investigating privacy concerns in such sites. We are interested in understanding how to address these concerns, and in particular wish to discuss our empirical work at this workshop and how to progress further in this space.

Author Keywords

ethics, informed consent, Facebook

ACM Classification Keywords

K.4.1 [Public Policy Issues]: Privacy

General Terms

Social media, privacy, research

Introduction

Understanding privacy in social network sites (SNSs) is a large and important area of research in computer science and many other fields. The ethical considerations of such research are numerous and complicated. Our position is that understanding how to address such considerations will improve measurement, and therefore our understanding, of networked social privacy. In this paper we briefly discuss some empirical work that we have conducted to replicate two existing studies in an attempt

to understand SNS users' privacy concerns about sharing data with researchers, rather than with other SNS users.

Ethical considerations

Ethical considerations with SNS research in part result from the large number of key actors in SNS research: the participants, their friends, other SNS users, researchers, and the SNS providers. Our particular interest is the relationship between SNS users and researchers, which is an active area of discussion. For instance, Neuhaus and Webmoor propose "agile ethics" for academic researchers, as such researchers need to take more care with SNS data than the SNS providers themselves [4]. Zimmer notes that just because users share data on an SNS, this does not mean that they are fair game for researchers [8].

A study of consent

One particular ethical consideration in SNS research is the issue of informed consent. It has been argued that obtaining informed consent is impractical, or perhaps inappropriate [6]. But the ease at which one can collect data from an SNS means that data might be collected, without consent, which SNS users might be unwilling to share with researchers. To examine this, we conducted a study which aimed to replicate two previous Facebook studies:

- The Harvard *T3* study [3]. This study used student research assistants' Facebook accounts to retrieve information from an entire Harvard cohort's Facebook accounts without their knowledge.
- Our own *LocShare* study [1]. This study used a small number of participants, all of whom gave informed consent, and explored location-sharing preferences and concerns on Facebook.

These two studies were chosen due to their different ethical approaches, in terms of how informed consent was obtained, and how the use of personal data was communicated to participants.

81 participants were recruited from the US and UK using a combination of Facebook advertisements, mailing lists and snowball recruitment. Each participant was preselected into one of two groups, representing the two emulated studies. They were presented with some information which deceived them into believing that they were participating in a study researching health (*T3*) or mobile information dissemination (*LocShare*). The deception was explained at the end of the study. Participants were then presented with 100 pieces of information from their Facebook accounts, and asked whether they were willing to share this with the researchers (e.g., Fig. 1).

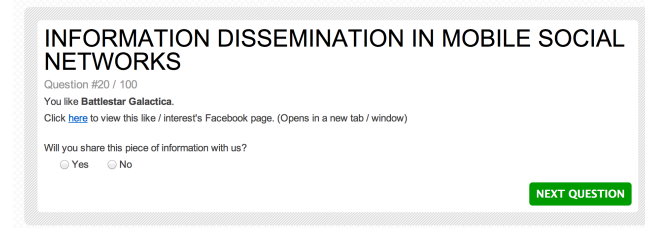


Figure 1: An example of the Facebook information that participants were asked to share in our study.

Participants were willing to share different types of information in different ways (Fig. 2), confirmed by a chi-square test of independence ($\chi^2 = 20.8$, $N = 5379$, $p < 0.05$).

Likes and interests, as used in the *T3* study, were indeed shared, and in both studies. Photos were less likely to be shared, and location information varied between the two

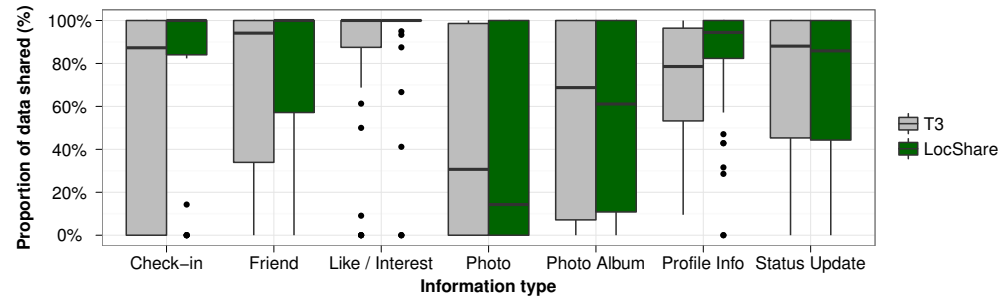


Figure 2: Different types of information were shared with researchers in different ways. Photos were shared less often, and locations were shared more often in the LocShare study.

studies, being more likely to be shared in the *LocShare* study, where location might be deemed more appropriate (as per the norms of Nissenbaum’s contextual integrity [5]). For all information types, we see significant internal variation, with some consistent themes emerging. Many participants either shared all or nothing, which underscores the difficulty of objectively assessing the sensitivity of such information. For example, while privacy concerns with location data are well understood, the high sharing rate of check-ins during this study suggests its sensitivity may decay with time. Conversely, the relatively low willingness to share photos could indicate people are embarrassed by the content of the images, particularly younger people who may regret some of the photos they have been previously captured in [7]. This poses a challenge for research of this nature, where the sensitivity of historic social network data is a moving target.

The privacy settings attached to information seems to have had an effect (Fig. 3), with a chi-square test of independence showing significant differences between sharing rates for different privacy settings ($\chi^2 = 48.27$, N

= 5379, $p < 0.05$). Information that participants had already selected to be shared with everyone was more likely to be shared with researchers, and information which was originally published with a ‘custom’ privacy setting, indicating it was only shared with a subset of friends, was shared less frequently with researchers. This also suggests that privacy settings are fairly robust to capturing people’s contextual norms over time.

Note that this is preliminary analysis and our work is ongoing.

Contribution to workshop

We would like to discuss the following issues in the workshop:

- What are the risks associated with empirical evaluation of ethical considerations?
- Is it meaningful to replicate studies, and to compare results across studies, in the way that we have done? What are meaningful cross-study metrics?

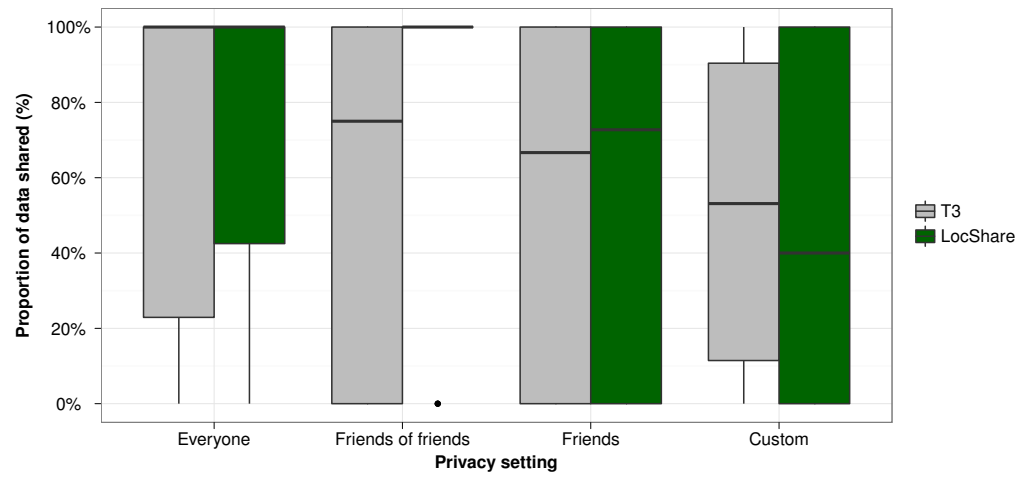


Figure 3: Information that was public was more likely to be shared with researchers.

- Can we use contextual integrity to minimise ethical considerations in social media privacy studies?
- What do our results mean for data sharing? Is it possible to share data from one study for use in other studies? How should we explain this to participants?
- We have developed an architecture for privacy-aware data collection from SNSs for social media studies [2]. This was used to collect data in the study described previously. We do not know if our architecture is sufficiently general enough for other researchers, and would welcome feedback.

Acknowledgements

This work was sponsored by the EPSRC Framework for Responsible Research & Innovation in IT.

References

- [1] Ben Abdesslem, F., Parris, I., and Henderson, T. Mobile experience sampling: Reaching the parts of Facebook other methods cannot reach. In *Proceedings of the Privacy and Usability Methods Pow-Wow (PUMP)* (Sept. 2010). Online at <http://scone.cs.st-andrews.ac.uk/pump2010/papers/benabdesslem.pdf>.
- [2] Hutton, L., and Henderson, T. An architecture for ethical and privacy-sensitive social network experiments. In *Proceedings of the ACM SIGMETRICS Workshop on Privacy and Anonymity*

- for the Digital Economy* (June 2012). Online at <http://pade12.mytestbed.net/pade12-final4.pdf>.
- [3] Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., and Christakis, N. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks* 30, 4 (Oct. 2008), 330–342. doi:[10.1016/j.socnet.2008.07.002](https://doi.org/10.1016/j.socnet.2008.07.002).
- [4] Neuhaus, F., and Webmoor, T. Agile ethics for massified research and visualization. *Information, Communication & Society* 15, 1 (Feb. 2012), 43–65. doi:[10.1080/1369118X.2011.616519](https://doi.org/10.1080/1369118X.2011.616519).
- [5] Nissenbaum, H. F. Privacy as contextual integrity. *Washington Law Review* 79, 1 (Feb. 2004), 119–157. Online at <http://ssrn.com/abstract=534622>.
- [6] Solberg, L. Data mining on Facebook: A free space for researchers or an IRB nightmare? *University of Illinois Journal of Law, Technology & Policy* 2010, 2 (2010). Online at <http://www.jltp.uiuc.edu/works/Solberg.htm>.
- [7] Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., and Cranor, L. F. “I regretted the minute I pressed share”: a qualitative study of regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security, SOUPS '11* (July 2011). doi:[10.1145/2078827.2078841](https://doi.org/10.1145/2078827.2078841).
- [8] Zimmer, M. “But the data is already public”: on the ethics of research in Facebook. *Ethics and Information Technology* 12, 4 (Dec. 2010), 313–325. doi:[10.1007/s10676-010-9227-5](https://doi.org/10.1007/s10676-010-9227-5).