How do we decide how much to reveal?

(Hint: Our privacy behavior might be socially constructed.)

Aylin Caliskan-Islam
Drexel University
ac993@drexel.edu

How do we decide how much to share online given that information can spread to millions in large social networks? Is it always our own decision or are we influenced by our friends? Let's isolate this problem to one variable, private information. How much private information are we sharing in our posts and are we the only authority controlling how much private information to divulge in our text messages? Understanding how privacy behavior is formed could give us key insights for choosing our privacy settings, friends circles, and how much privacy to sacrifice in social networks. Before analyzing end users' privacy behavior, we had the intuition that privacy behavior might be under the effect of network phenomena. Christakis and Fowler's network analytics studies [2] showing that obesity spreads through social ties and smoking cessation is a collective behavior [3], influenced us to further investigate network properties of privacy behavior.

In a recent paper that appeared at the 2014 Workshop on Privacy in the Electronic Society [1], we present a novel method for quantifying privacy behavior of users by using machine learning classifiers and natural-language processing techniques including topic categorization, named entity recognition, and semantic classification. Following the intuition that some textual data is more private than others, we had Amazon Mechanical Turk workers label tweets of hundreds of users as private or not based on nine privacy categories that were influenced by Wang et al.'s Facebook regrets categories [7] and Sleeper et al.'s Twitter regrets categories [5]. These labels were used to associate a privacy score with each user to reflect the amount of private information they reveal. We trained a machine learning classifier based on the calculated privacy scores to predict the privacy scores of 2,000 Twitter users whose data were collected through the Twitter API.

The supervised machine learning classifier detects the privacy scores of hundreds of labeled users with 95% accuracy in a few minutes, whereas manual analysis does not scale. After predicting the privacy scores of thousands of Twitter users, we found that there is a correlation between the privacy score of a user and those of her friends. There is even a higher correlation of privacy score between a user and the other users mentioned in her tweets. People with similar privacy scores appear in groups. The possible causal relationships in this phenomenon need further exploration.

The ability to automatically quantify private information disclosure and compute privacy scores provides a potentially useful method for users, researchers, and companies. A user can make sharing decisions in a more informed manner if the privacy risk associated with each friend is known. For example, she can take privacy scores into account when constructing friend lists. Researchers who study people's use of social media can also use the privacy score calculation method for a fine grained analysis of individual privacy behavior. Which type of textual data, namely messages, status updates, mentions, or comments have more private information?

Social media companies could tailor "nudges" based on users' (and their friends') privacy scores. For example, a social network could alert the user when she is about to share content that appears to be highly private with a group of friends that includes users with low privacy scores. A recent study by Wang et al. [6] on privacy nudges show promising results on preventing unintended disclosure and associated regret. As Garg et al. [4] demonstrate, outcomes of privacy management can be improved at a lower overall cost if peers, as a community, are empowered by appropriate technical and policy mechanisms. Social media companies are also in a position to run controlled experiments to determine if privacy behaviors are indeed contagious.

We are planning to do another privacy analytics study after obtaining IRB approval to learn more about how people are influenced to reveal private information and the effects of Facebook's default newsfeed algorithm. The correlation between the privacy score of a user and her friends gives a starting point for investigating the causal factors behind self-disclosure. Better understanding these factors can help effectively design privacy enhancing technologies and target educational interventions.

This material is based on work supported by the National Science Foundation under grant 1253418. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation.

1. REFERENCES

[1] A. Caliskan Islam, J. Walsh, and R. Greenstadt. Privacy detective: Detecting private information and collective privacy behavior in a large social network. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 35–46. ACM, 2014.

[2] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

[3] N. A. Christakis and J. H. Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258, 2008.

[4] V. Garg, S. Patil, A. Kapadia, and L. J. Camp. Peer-produced privacy protection. In *International Symposium on Technology and Society (ISTAS)*, pages 147–154. IEEE, 2013.

[5] M. Sleeper, J. Cranshaw, P. G. Kelley, B. Ur, A. Acquisti, L. F. Cranor, and N. Sadeh. i read my twitter the next morning and was astonished: a conversational perspective on twitter regrets. In *Proceedings of the 2013 ACM annual conference on Human factors in computing systems*, pages 3277–3286. ACM, 2013.

[6] Y. Wang, P. G. Leon, K. Scott, X. Chen, A. Acquisti, and L. F. Cranor. Privacy nudges for social media: An exploratory facebook study. In *Proceedings of the 22Nd International Conference on World Wide Web Companion*, WWW '13 Companion, pages 763–770, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[7] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor. "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, pages 10:1–10:16, New York, NY, USA, 2011. ACM.