



Don't Overshare, Don't Undershare:

a Goldilocks Solution for Inferring Sharing Preferences and Social Relationships from Phone Logs

Jason Wiese, Jun-Ki Min, Jason Hong, John Zimmerman



Problem Statement

Privacy controls are infrequently set or changed in online social sharing systems, resulting in static privacy settings that seldom match the user's real preferences [4]. Both over- and under-sharing in these systems can have negative social outcomes. For example, undersharing may limit friendship building and maintainance [2], but when information is made accessible to people that it was not intended for, negative consequences include damaging relationships and losing jobs [1]. Furthermore, as social sharing becomes more passive and continuous (such as Facebook's "frictionless sharing"), these negative outcomes are exacerbated.

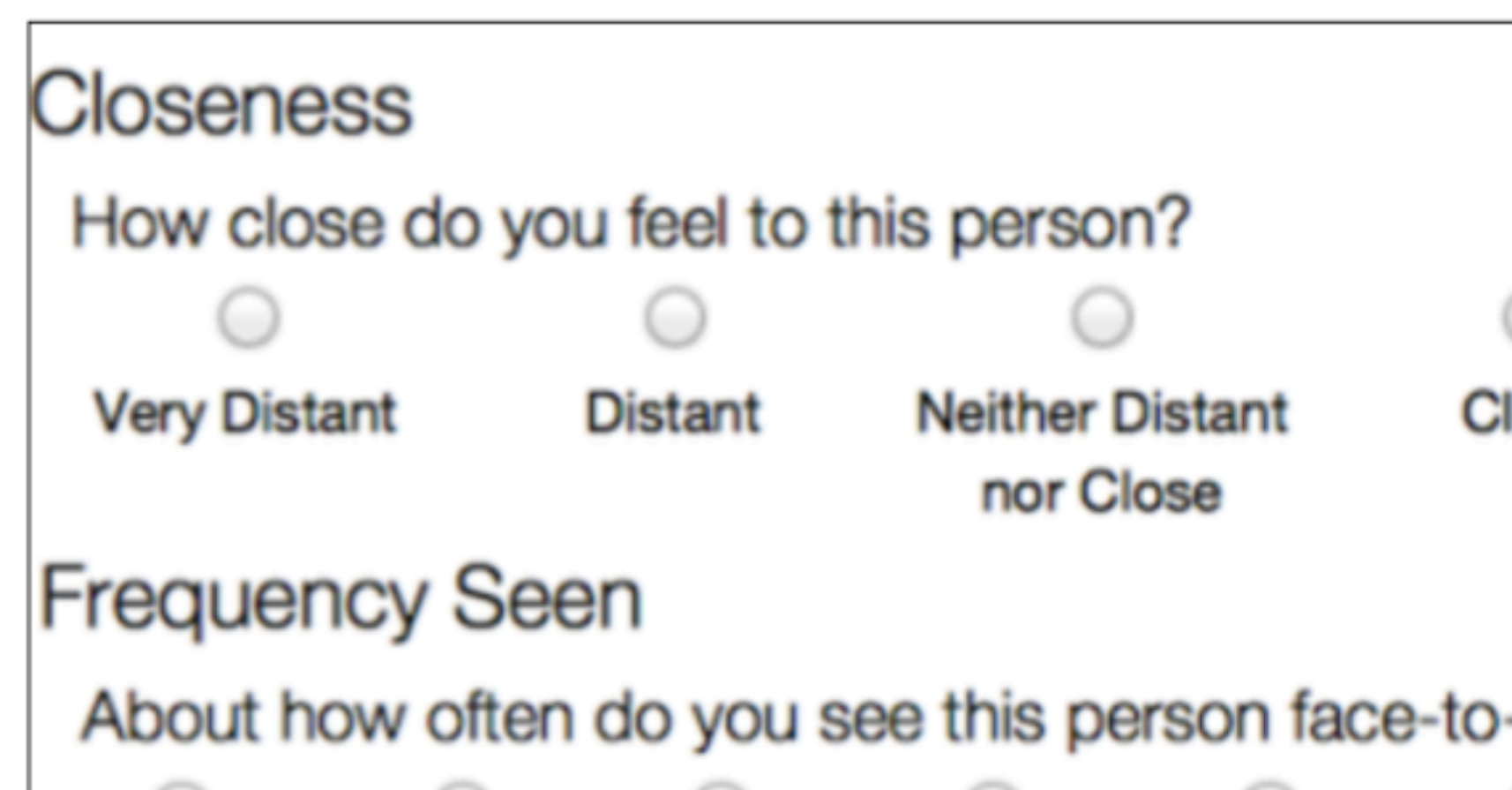
The goal of this work is to partially automate the process of specifying sharing preferences by inferring dimensions of the social relationships that the user has with each of her friends.

Usable Privacy and Security, the central theme of our IGERT project, focuses on understanding and improving the mechanisms that humans use to manage privacy and security. This work aims to improve the usability of privacy settings in social networks by lowering the burden for users to configure and maintain privacy settings that match the user's preferences

Data Collection

We conducted two studies to explore the relationships between sharing preferences, tie strength and life facet (work/family/social), and calling/sms patterns. In both studies, For each each of 70 of their friends, participants provided:

- Sex of the Friend
- Age of the Friend
- Frequency Seen
- Frequency of Comm
- Sharing preferences
- Closeness
- Life facet
- Years known



Example questions from our custom data collection website

For Study 1, we asked the 39 participants to indicate their willingness to share information with each of their 70 friends in the context of 21 different information-sharing scenarios. This resulted in a total of 57330 sharing judgements captured by this study.

For Study 2, 40 participants provided us with their mobile call and SMS logs using an application that we developed for the Android operating system. Participants then used a custom web interface to ensure that there were no duplicate representations of their friends.

Study 1 Results: Relationships and Sharing Preferences

To understand the relationship between social relationships and sharing preferences, we conducted a mixed-model analysis of variance predicting sharing as the outcome variable. This analysis allowed us to explain and compare the variation in sharing using different combinations of independent variables. All regressions were done on a per-friend level of analysis; for these models, we took the mean sharing value across all scenarios for each friend (n=2730) and used features that described each relationship as effects in the models.

Model Description	R ² (variance explained)
Closeness Only	0.63
Life Facet (Fam/Work/Social) Only	0.48
Closeness & Life Facet	0.65

These results indicate that, while social closeness and life facet do not completely explain sharing preferences, there is a solid connection between these dimensions of relationship and the desire to share information with these friends.

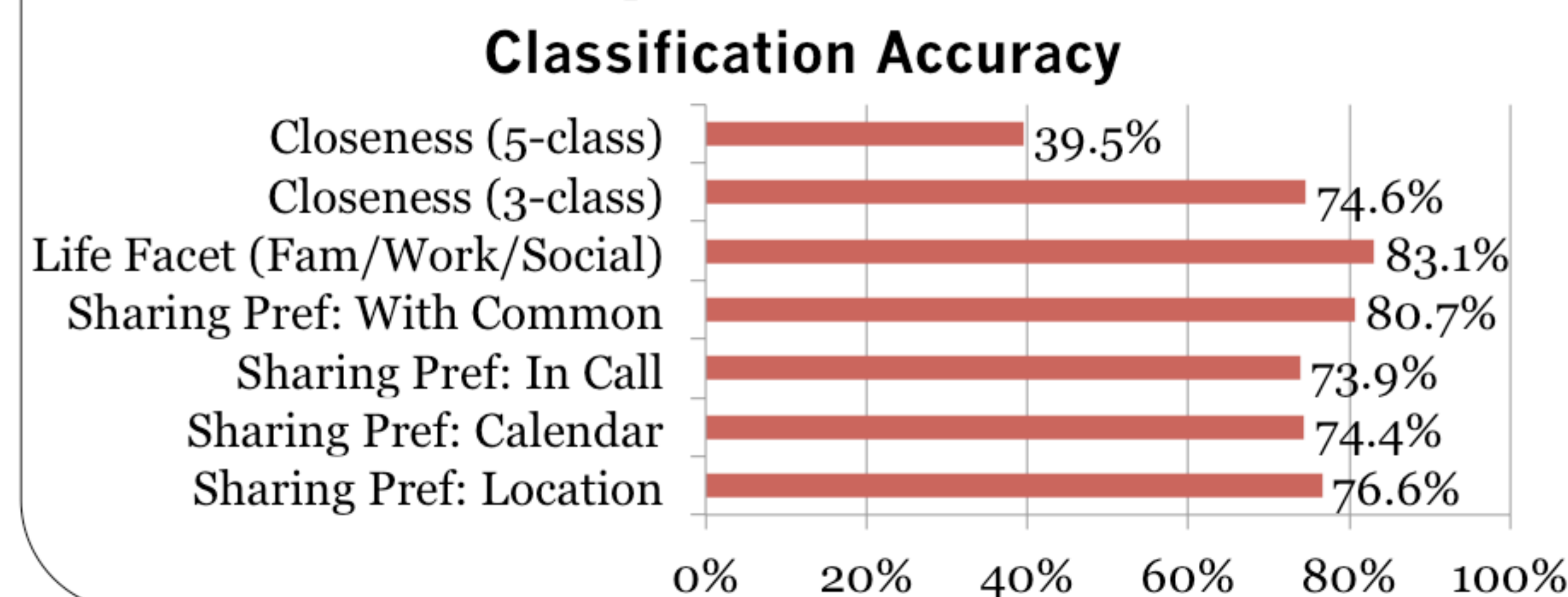
Study 2 Results: Inferring Relationships from Communication

Based on a broad variety of references (see our paper for more details [3]), we defined five factors that characterize communication patterns in the context of social relationships:

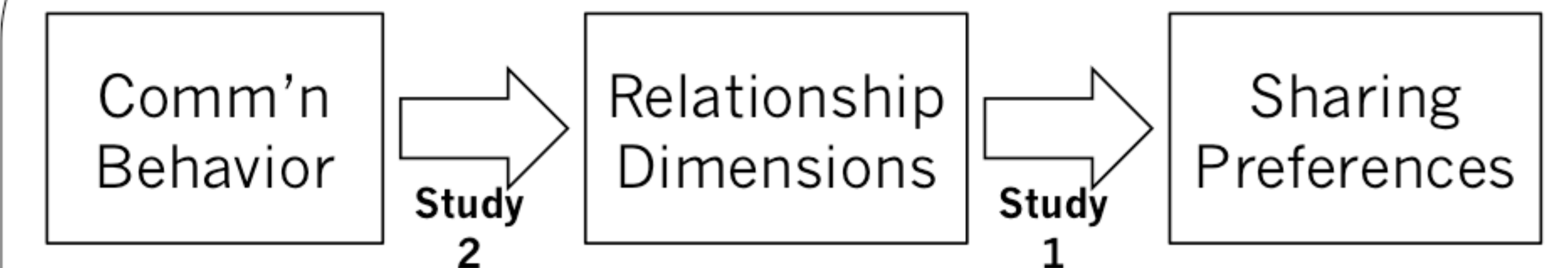
- **Intensity and regularity** – eg: the number and duration of calls
- **Temporal tendency** – eg: specific time of day, day of week
- **Channel selection/avoidance** – eg: incoming vs outgoing calls
- **Maintenance cost** –eg: # of recent calls vs # of long-term calls

With these factors in mind, we extracted 153 features based on call logs, SMS logs, and contact list. We used these features as an input to the Weka Toolkit's SVM SMO algorithm to infer both closeness and life facet. In addition, we asked 4 sharing questions of these participants, and constructed models to predict those as well.

Classification results (below) indicate that closeness (74.6%) and life facet (83.1%) can be inferred with moderate accuracy, but that these models are still far from perfect.



Discussion



Study 1 demonstrated that existing dimensions for describing social relationships are useful for helping to characterize social sharing preferences. Study 2 demonstrated a proof of concept that observable factors of communication behavior can be used to infer these dimensions of social relationships. Together, these results lead to several important insights and implications

- 1. Models do not need to be perfect to be useful** – While the ideal situation is for sharing preferences to be automatically configured and always correct, this approach represents a marked improvement over the current, fully manual approach to privacy settings. Even with inaccuracies, if these models were deployed today they would demonstrate an improvement over the common current approaches of *share with everybody* or *share with nobody*.
- 2. More complete communication behavior will likely improve these models** – Both a dataset that extends further back in time and data from more communication sources (e.g. Skype, IM, etc) are likely to improve the results of these models
- 3. Identifying additional relationship dimensions to model** – In this work we have examined two important dimensions of personal relationships (closeness and life facet), but it is likely that communication behavior also is indicative of other dimensions of these relationships.
- 4. As online sharing becomes more automated and less based on explicit action, inferring sharing preferences will become more important** – Data is increasingly shared on our behalf (as in Facebook's "frictionless sharing" or Google Latitude), privacy decisions cannot be expected to remain static, and instead must adjust to the situation at hand

References

- [1] J. Binder, A. Howes, and A. Sutcliffe. The problem of conflicting social spheres: Effects of network structure on experienced tension in social network sites. In CHI '09, pages 965–974, New York, NY, USA, 2009. ACM.
- [2] N. Ellison, C. Steinfield, & C. Lampe. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), 1143-1168. 2007.
- [3] J. Min, J. Wiese, J. Hong, and J. Zimmerman. Mining smartphone data to classify life-facets of social relationships. In CSCW '13, pages 285-294. New York, NY, USA. 2013. ACM.
- [4] K. Strater and H. R. Lipford. Strategies and struggles with privacy in an online social networking community. In BCS-HCI '08, pages 111–119, Swinton, UK, UK, 2008. British Computer Society.