Investigating Factors Affecting Personal Data Disclosure

Christos Perentis
Telecom Italia – SKIL & FBK
Via Sommarive, 18 – Povo
Trento, I-38123, Italy
perentis@fbk.eu

Michele Vescovi
Telecom Italia – SKIL
Via Sommarive, 18 – Povo
Trento, I-38123, Italy
michele.vescovi@telecomitalia.it

Bruno Lepri
FBK
Via Sommarive, 18 – Povo
Trento, I-38123, Italy
lepri@fbk.eu

ABSTRACT

Mobile devices, sensors and social networks have dramatically increased the collection and sharing of personal and contextual information of individuals. Hence, users constantly make disclosure decisions on the basis of a difficult trade-off between using services and data protection. Understanding the factors linked to the disclosure behavior of personal information is a step forward to assist users in their decisions. In this paper, we model the disclosure of personal information and investigate their relationships not only with demographic and self-reported individual characteristics, but also with real behavior inferred from mobile phone usage. Preliminary results show that real behavior captured from mobile data relates with actual sharing behavior, providing the basis for future predictive models.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: [human factors]; H.5.2 [**User Interfaces**]: [user-centered design]; K.4.1 [**Public Policy Issues**]: [privacy]

Keywords

Personal Data, Living Lab, Human Factors

1. INTRODUCTION

The wide-spread use of mobile devices and their capability of collecting information about human behavior have boosted personal data production. Moreover, services like online social networks and other popular mobile applications allow people to increasingly share their personal information (e.g. current location, activities, etc.) [6]. Subsequently, unprecedented privacy concerns are raised as users continuously deal with personal information disclosure decisions. Researchers have focused on the role of various factors linked to the attitudes towards data disclosure: e.g. interpersonal relationships [2]; user characteristics such as age [1], gender or personality traits [8]; the typology of the data to be disclosed [5]. Besides considering only demographic information, self-reported individual traits and privacy dispositions, our study considers real behavioral features about social interactions captured automatically by mobile phones like communication (calls and SMSs), locations and self-reported expenses captured by a mobile phone app. In order to investigate all the aforementioned factors, we undertook a field-study with a community of 63 subjects that were provided with (i) a smart phone incorporating a sensing software explicitly designed for collecting mobile phone data; and (ii) a Personal Data Store (PDS): a system meant to both enable individuals to control the disclosure of their data with the other members of the community and to keep track of their actual sharing behavior. A distinguishing feature of our approach is that we observe actual sharing behavior rather than attitudes expressed through questionnaires. Preliminary results indicate that sharing behavior is related with the individuals' traits as well as with communication and PDS usage features.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742757.

2. FIELD STUDY

In this work, we report a study conducted in 20 male and 43 female members of the MTL (`www.mobileterritoriallab.eu`) community. Participants' age ranged from 28 to 46 years old (mean = 38.67 and standard deviation = 3.34). They held a variety of occupations and education levels. All were savvy Android users. All participants lived in Italy and the vast majority were Italians.

Sensed Data. All users where provided with a smart phone equipped with a sensing software that runs in a passive manner and does not interfere with the normal usage of the phone. The collected data types were: i) Location (GPS), ii) Call & SMS logs, and (iii) daily expenses collected via mobile app. From the collected behavioral data we computed 3 features per data type to use in the analysis like: i) total number of: visited locations, calls (outgoing/incoming), SMS (sent/received), and expenses transactions, ii) unique number of: locations visited, Call contacts, SMS contacts, type & category of expenses, and iii) diversity of: locations, calls, SMS and expenses [3].

Personal Data Store. The system stores the participant's information and permits user to exercise full control on own data (e.g. collect, share, delete) [10]. By using the PDS *Sharing Area*, subjects can decide about whether and how to disclose their different data types (i.e. locations, calls & SMS, expenses) to the other participants choosing the desired disclosure level, distinguished into: (i) *Do Not Share*; (ii) *Share Anonymously*; (iii) *Share Not Anonymously*. Depending on these preferences the PDS makes user data visible and provides a relative feedback to enable comparisons between community members. A user not sharing data can access only the *Individual Views* (e.g. charts, pies) that aggregate user behavior. If one chooses *Share Anonymously*, the *Social Views* release user data to the community and user receives community feedback; both aggregated and anonymously. Finally, if the selection is *Share Not Anonymously*, detailed data with demographics are released. We also considered the role played by PDS usage by computing the total number of subjects accessed the: (i) PDS system, (ii) Individual Views and (iii) Social Views per data type.

Procedure. The study took place for 15 weeks. Before the beginning of the study participants were requested to fill a survey about demographic information and other individual traits like: Big-5 personality traits [7], Locus of Control [4] and Privacy Concerns [9]. At the first week of the study, participants were asked to set their initial disclosure preference per data type using the Sharing Area of the PDS. From that time on, subjects were free to change their setting at will and at any time. At the end of the study individuals were asked to set their final sharing preferences on the PDS.

3. DATA ANALYSIS AND RESULTS

Our main goal was to model the disclosure of personal information. Hence, we constructed dependent variables taking into account the final disclosing choices that subjects set in the PDS; one for each data type: Location, Interactions (calls & SMS) and Expenses. We observe from the frequency tables that the *Do Not Share* choice has few occurrences (it was selected only by a couple of subjects), therefore we merged it with the closest option *Share Anonymously*.

Next, we focused only on share *Anonymously* and *Not Anonymously* choices considering three dichotomous variables, one per data type, depicting sharing preference: *Final Location*, *Final Interactions* and *Final Expenses*. Here, we report observed relationships between the dependent variables and i) the individual traits as well as ii) the real behavior (mobile phone logs and PDS logs). All the individual traits captured from surveys, the mobile phone data and the PDS logs are normalized scalar scores. We apply the Point-biserial r_{bp} coefficient to identify linear correlations between dichotomous categorical variables with scale variables.

Firstly, call diversity, which reflects how the communication time of an individual is distributed among its contacts, has a significant positive linear relationship with *Final Location* (r^2=0.335 with p=0.007), *Final Interactions* (r^2=0.333 with p=0.008) and *Final Expenses* (r^2=0.309 with p=0.014). This means that the higher the call diversity score is, the more tends someone to disclose (by choosing *Not Anonymously*) Location, Interactions and Expenses data types.

The *Final Location* and *Final Interactions* present significant positive linear relationships with how many times they accessed the Location (r^2=0.374 with p=0.002) and the Interaction Social Views (r^2=0.310 with p=0.013), respectively. A way to interpret this finding is that the subjects that are more interested in *Social Views* (i.e. they access them more often to compare their behavior with the rest members of the community) evaluate this feedback positively, so they tend to be more open in disclosing.

Regarding the individual characteristics and traits, we capture a significant negative linear relationship of Locus of control with the *Final Expenses* disclosure choice (r^2= −0.266 with p=0.035). High score of Locus of Control indicates that individuals do not feel in control of the events happening in their life (*externals*), while a low score shows the opposite (*internals*). Therefore, this negative relationship reveals that the more external the individual is, the less open shares (*Anonymously*) the expenses data. As multicollinearity will be a concern when building a model, the relations between all the features correlated with the dependent variables were examined and no significant associations were discovered. The preliminary findings discribed here indicate that real behavior features are useful in understanding disclosing preferences. Building on these results, we aim to formulate predictive models considering more individual traits and factors characterizing the underlying social relationships.

4. REFERENCES

[1] E. Christofides, A. Muise, and S. Desmarais. Hey mom, what's on your facebook? comparing facebook disclosure and privacy in adolescents and adults. *Social Psychology and Personality Science*, pages 48–55, 2011.

[2] S. Consolvo, I. E. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge. Location disclosure to social relations: why, when, & what people want to share. In *In SIGCHI on Human factors in computing systems*, pages 81–90. ACM, 2005.

[3] N. Eagle, M. Macy, and R. Claxton. Network diversity and economic development. *Science*, 328(5981):1029–1031, 2010.

[4] T. Farma and I. Cortinovis. Un questionario sul "locus control": A questionnaire on the locus of control: its use in the italian context). *Ricerca in Psicoterapia*, 3(2/3):147–155, 2000.

[5] B. P. Knijnenburg, A. Kobsa, and H. Jin. Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies*, 71(12):1144–1162, 2013.

[6] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, 2010.

[7] M. Perugini and L. Di Blas. *The Big Five Marker Scales (BFMS) and the italian AB5C taxonomy: Analyses from an emic-etic perspective*. Hogrefe & Huber Publishers, 2002.

[8] D. Quercia, D. B. Las Casas, J. P. Pesce, D. Stillwell, M. Kosinski, V. Almeida, and J. Crowcroft. Facebook and privacy: The balancing act of personality, gender, and relationship currency. In *ICWSM*, 2012.

[9] H. J. Smith, S. J. Milberg, and S. J. Burke. Information privacy: measuring individuals' concerns about organizational practices. *MIS quarterly*, pages 167–196, 1996.

[10] M. Vescovi, C. Perentis, C. Leonardi, B. Lepri, and C. Moiso. My data store: toward user awareness and control on personal data. In *UbiComp: Adjunct Publication*, pages 179–182. ACM, 2014.