Aalborg Universitet

Privacy Analysis in Mobile Social Networks

the Influential Factors for Disclosure of Personal Data

Sapuppo, Antonio

Privacy analysis in mobile social networks: the influential factors for disclosure of personal data

Antonio Sapuppo

Aalborg University,
Center for Communication, Media and Information Technologies,
Sydhavnsgade 17-19, Frederikskaj 12, Copenhagen 2450, Denmark
E-mail: antoniosapuppo@gmail.com

Abstract: Nowadays, mobile social networks are capable of promoting social networking benefits during physical meetings, in order to leverage interpersonal affinities not only among acquaintances, but also between strangers. Due to their foundation on automated sharing of personal data in the physical surroundings of the user, these networks are subject to crucial privacy threats. Privacy management systems must be capable of accurate selection of data disclosure according to human data sensitivity evaluation. Therefore, it is crucial to research and comprehend individual's personal information disclosure decisions happening in ordinary human communication. Consequently, in this paper we provide insight into influential factors of human data disclosure decisions, by presenting and analyzing results of an empirical investigation comprising of two online surveys. We focus on the following influential factors: inquirer, purpose of disclosure, access & control of the disclosed information, location familiarity and current activity of the user. This research can serve as relevant input for the design of privacy management models in mobile social networks.

Keywords: privacy; information disclosure; mobile computing; social networks; social and proximity interactions; ubiquitous computing.

Biographical notes: Antonio Sapuppo is a PhD from 2009 at the Aalborg University in Copenhagen (Denmark). He holds a master degree in software engineering from the Aalborg University (Denmark) and a bachelor degree in computer science at the University of Studies of Catania (Italy). During 2008 and 2009, he worked as research engineer at the Copenhagen University College of Engineering (Denmark) to design, implement and test a mobile social network, called Spiderweb. During fall 2011, he visited the Auckland University of Technology (New Zealand), focusing on ubiquitous social computing research area. His main interest areas are privacy, social networks and ubiquitous computing.

1 Introduction

The development of Internet reduced the distance between people living in different parts of the world by providing an innovative communication infrastructure. Soon on the basis of this technology new services have been developed, which improved the communication between people. Online social networks (in the following referred to as OSNs), such as Orkut, MySpace and Facebook, share a common characteristic: they enable people to create a virtual social network. By using OSNs services, users can stay in touch with friends from the whole world, share pictures, talk, chat, send messages and look for new acquaintances. The success of OSNs, the wide spread of mobile phones and the current development of numerous information and communication technologies allowed to create similar services also for mobile terminals (Counts and Fisher, 2008; Ziv and Mulloth, 2007).

Notably, mobile devices are not just entry points to existing online social networks, but they also offer new networking services due to their advanced technological capabilities. In fact, thanks to the wireless technologies of mobile devices, they enable Opportunistic Networks (in the following referred to as ONs). In ONs, nodes are wirelessly interconnected and have the possibility to identify each other as well as share data in peer-to-peer networks with communication links created in ad hoc manner (Lilien et al., 2006).

The integration of ONs with OSNs enables mobile social networks users to exploit social networking benefits in the physical world, rather than just

in the virtual world. This integration has been previously introduced as Local Social Networks (in the following referred to as LSNs) (Sapuppo and Sørensen, 2011; Sapuppo, 2010). LSN is a distributed network architecture in which nodes are linked to online social networks profiles and wirelessly interconnected to exchange personalized contents. LSNs target at developing possible advantageous relationships (e.g. friendships, partnerships, business relations) during physical meetings between people who do not know each other, but probably they should (Eagle and Pentland, 2005; Sapuppo and Sørensen, 2011).

When transferring OSNs benefits to the physical world, the privacy threats are indisputably increased due to support of face-to-face interactions between strangers during physical meetings. While the risk of unintentional information sharing is similar in virtual and physical worlds, the consequences of such disclosure are more crucial in the physical world. For example, when LSN users disclose their personal information, the shared data is tied to a physical person and immediately available for the recipient (Sapuppo and Sørensen, 2011). Thus, the information disclosure can be directly translated into physical contact and potentially undesired or unpleasant face-to-face interactions. In order to address these privacy concerns, privacy management systems should protect users' personal privacy as individuals do in ordinary human interactions and ensure accuracy of selective disclosure of personal information (Bunnig, 2009a,b; Hong et al., 2004). In fact, during face-to-face communication, people intuitively evaluate various determinants and unconsciously choose what personal information to share. In order to help privacy management systems to attempt to act as the real user would, it is necessary to gain an extensive comprehension of variation of human data sensitivity that affects information disclosure under different circumstances. The factors that might influence users personal data disclosure decisions must be depicted and evaluated for enabling privacy management systems to take automated data disclosure decisions.

In the past work, the identity of the inquirer was identified as the primary index for selection of data disclosure decisions (Lederer et al., 2003a; Davis and Gutwin, 2005; Olson et al., 2005; Jones et al., 2004; Consolvo et al., 2005). However, mobile social networks, such as LSNs, advance the attention to other factors as crucial determinants for data disclosure, due to their primary focus on relationship initiation between strangers (Sapuppo and Sørensen, 2011). Consequently, in this paper we firstly identify the relevant influential factors that might impact users' personal data disclosure decisions in LSNs. Afterwards, we present results of an empirical investigation comprising 2 online surveys to evaluate the identified influential factors. We collected more than 100 responses in each of the surveys and we applied the Wilcoxon Signed Rank statistical test to examine whether the identified influential factors impact on users' personal data disclosure decisions. The results

of our analysis can provide significant input for the design and development of privacy management systems for mobile social networks.

The rest of the paper is structured as follows: firstly, we introduce the potential influential factors for the disclosure of personal information in mobile social networks. In Section 3, we present the design and methodology of the two surveys, which investigate the relevant influential factors in LSNs. Further, the information about the participants is provided in Section 4. In Section 5, we present and discuss the results of the empirical investigation. Final conclusions and recommendations for future work are drawn in Section 6.

2 Human data disclosure

The core foundation of mobile social networks, such as LSNs, is based on automated sharing of users' personal data. Surely, the amount of disclosed information is directly proportional to networking benefits. The optimal outcome would be achieved by sharing as much as possible personal information (e.g. the full user profile). However, this would result in jeopardy of users' privacy and a compromise is necessary. It can be achieved by following the assumption that the sensitivity of the users' personal information is not stable; it may vary depending on different circumstances in which the user is involved (Lederer et al., 2003a; Wright et al., 2009; Sapuppo and Sørensen, 2011). Consequently, only information that is relevant, but not sensitive in specific circumstances should be disclosed at a time (Bunnig, 2009a,b; Kapadia et al., 2007; Langheinrich, 2001; Yee, 2010). Therefore, no standard rules can be applied for all the cases of disclosure of users' personal data (Altman, 1975, 1977; Palen and Dourish, 2003).

In previous studies (Bunnig, 2009a; Jendricke et al., 2002; Lederer et al., 2003a; Sapuppo and Sørensen, 2011), the sensitivity of personal information was assumed to vary depending on the inquirer and the situation determinants. The inquirer is considered to be the individual that the user is interacting with and the situation is defined according to the circumstances at that time.

Lederer et al determined the identity of the inquirer to be the most important factor, influencing the users' data disclosure decisions, followed by the situation as parameter of secondary significance (Lederer et al., 2003a). Based on these findings several privacy management models have been designed for disclosure of personal information: Faces (Lederer et al., 2003b), Precision Dial (Lederer et al., 2004), Diverged Personalities (Sapuppo and Sørensen, 2011) and Disclosure Decision Model (Bunnig, 2009a,b; Bunnig and Cap, 2009).

In (Davis and Gutwin, 2005), the authors provided further insight into the inquirer influential factor by carrying out a survey to investigate the nature of

relationships between the users as a crucial determinant. Their results showed that users differentiate choices of disclosure of personal information upon relationships with the inquirer. Additionally to Davis and Gutwin, other studies (Olson et al., 2005; Jones et al., 2004) highlighted the relevance of users' clustering into a manageable categories of inquirers (e.g. friends, families, co-workers, etc) in social location disclosure applications.

Even if defining the inquirer as a crucial parameter, Consolvo et al emphasized that knowing the particular reason of data disclosure would significantly motivate users to share their personal information (Consolvo et al., 2005). Other studies as well researched (Byun et al., 2005; Byun and Li, 2008) and applied (Tian et al., 2009; Agrawal et al., 2002; Bresciani et al., 2004; Massacci et al., 2006) the purpose of disclosure as a crucial determinant.

Additionally to the purpose of disclosure, Consolvo et al investigated the granularity of the disclosed information, which refers to the extent of details of shared data. The results showed that users tend not to differentiate granularity of disclosed information in order to protect their data privacy. In the majority of the cases, users either choose to disclose detailed information or they do not disclose anything at all. However, when they decide to disclose not detailed set of information, they do so because they assume that it is more useful for the inquirer, rather than for preserving their privacy (Consolvo et al., 2005).

Finally, anonymity can also be considered to be a relevant influential factor for sharing of personal information in mobile social networks. Being anonymous is defined as the state of not being identifiable within a set of subjects, due to removal of connections between the data owner and information. In (Langheinrich, 2001), the author discussed that having the possibility to remain anonymous would significantly increase users' data privacy protection. Consequently, applications of anonymity might allow users to feel safer and thus influence users' personal information decisions.

Additionally to the previously introduced influential factors, in this paper we draw the attention to other potential determinants that might impact human data disclosure decisions in mobile social networks, which are following defined:

- Location familiarity: it is considered to be the users' familiarity with his current location (e.g. home, parents' place, work environments, social environments, holiday environments, etc.);

- Current activity: it refers to the current action of the user (e.g. working, relaxing, shopping, etc.);

- Access & Control: it regards empowering users to add, remove or modify any information disclosed at any time, i.e. enabling to control other people's access to one's personal data even after the actual disclosure.

Importantly, access & control should be considered as an essential privacy protection principle for personal data disclosure (Directive, 1995; Gregg, 1975; Langheinrich, 2001). This principle is of crucial importance for avoiding potential future privacy threats, because a set of data, given up freely today, might create major user's privacy concerns in the future. Moreover, mobile social networks are becoming increasingly complex, thus users might feel that they are losing control over their personal data after the actual disclosure.

To the best of our knowledge, access & control of the disclosed information, user's current activity and location familiarity influential factors were not previously empirically investigated in regard to the disclosure of personal information in mobile social networks. Moreover, we did not observe other research considering additional influential factors for personal data disclosure in mobile social networks apart from the ones discussed in this section.

3 Design of the surveys

In order to gain insight into human data sensitivity, we asked surveys' participants to indicate personal information that they would like to share in different circumstances of their lives. The participants were informed that sharing of personal data is motivated by potential networking benefits, provided in return to disclosed information. Naturally, the benefits would be directly proportional to the amount of shared information, thus respondents were asked to compromise between privacy risks and potential benefits.

The different circumstances, presented to the respondents, were defined according to the influential factors, outlined in Section 2. However, anonymity and granularity of disclosed information influential factors were not included in this analysis, as we focus on investigating information disclosure in LSNs. The granularity of the disclosed information is often applied in mobile social networks in relation to disclosure of social locations among acquaintances, e.g. extent of details of current geo location: country, city, neighborhood, exact address where I am now (Iachello et al., 2005; Smith et al., 2005; Barkhuus et al., 2008). However, the main target of LSNs is to promote potential networking benefits between strangers by exploiting ONs. In ONs the disclosed information is restricted to the range of the wireless technology adopted. Particularly, users are notified about the presence of other LSN users only when they are in the proximity. When they move away, their location information is not available anymore, unless they re-enter into each other's wireless range. Therefore, the granularity of the disclosed information was not further investigated in this paper. Moreover, anonymity influential factor was not included in this research because LSNs users must be identifiable, i.e. they must allow other users to link their profiles to

real people. If anonymity would be applied in LSNs, it would result in significant losses of potential networking possibilities.

In the following we provide a detailed description of the design of the two surveys, which researched on the remaining influential factors, introduced in Section 2.

3.1 Survey I

In the first survey we investigated whether the location familiarity and current activity of the user can be considered as relevant determinants for data disclosure decisions in mobile social networks. First, we researched whether the time that the user had previously spent in his current location could influence the amount of disclosed information. For example, we examined if the user would differentiate his data disclosure choices between places where has spent a lot of time (e.g. a bar of his home town) and unfamiliar locations (e.g. a bar during a holiday). Secondly, we analyzed whether the user's current activity might influence users data disclosure decisions. For example, while working the user might be more motivated to share data related to working activities (e.g. professional abilities) in comparison to data related to social interactions (e.g. music taste).

In order to study those influential factors, we grouped the most common life situations into five categories, and asked the participants to indicate, which information they would like to disclose when they are facing those situations:

- Family places: these environments can be considered to be places where the user or her family members live (e.g. parents' apartment, uncles' apartment, etc). Thus, it was assumed that users would encounter their family members as well as family members' acquaintances, who could also be strangers for them;

- Social environments: these environments refer to the places where the users spend their leisure time, e.g. restaurants, bars, theaters in his home city. Thus, it was assumed that they would encounter friends and strangers;

- Holiday: similarly to the social environments, holiday environments are considered to be social leisure places, however the users' encounters and activities are occurring outside their home city;

- Work environments: these environments can be considered to be the ordinary employment places of the users, such as university, office, etc. Thus, users would mainly encounter co-workers and strangers, associated to their employment activities;

- Work trip: similarly to work environments, during work trips the users were assumed to encounter colleagues and strangers, associated to their employment activities, however these encounters

and activities were occurring outside their regular work place.

3.2 Survey II

In the second survey, we investigated whether inquirer, access & control and purpose of disclosure can be considered as relevant determinants for data sharing in mobile social networks. In mobile social networks inquirers can be generally categorized into friends and strangers segments. In this investigation we target at the latter segment of inquirers, namely strangers, due to focus of LSNs. Consequently, we chose to investigate the following two concepts in the analysis of inquirer as influential factor. Firstly, we analyzed whether knowing the number of mutual friends between the inquirer and the user a priori any data sharing, might impact his data disclosure decisions. Further, we also researched whether being familiar strangers with the inquirer could be considered as a relevant determinant. Two people are identified as familiar strangers if they encounter each other regularly without interacting or forming an explicit relationship of social nature (Milgram, 1977). Moreover, we researched access & control as determinant factor by investigating whether clearly emphasizing access & control rights might influence the users' data disclosure decisions. Finally, we also analyzed whether users' disclosure decisions might be affected by knowing beforehand what potential benefits they could get for disclosing their personal information to strangers.

In order to research these influential factors, we asked respondents to select their personal information that they would like to disclose in different scenarios. It was emphasized that the exchange of personal data would be automated, thus it would not interfere with the user's current activity. The relevant information that could be applied for networking with other users could be retrieved and used even at a later time. All the scenarios, presented in this survey, were indicated to be occurring in a social environment. Particularly, respondents were asked to imagine to be in a bar of their home city, drinking a coffee with friends. The respondents decided what to disclose to different inquirers, who were strangers for them. A priori any data disclosure decision, some information about the inquirer was known. Particularly, at least the basic information set about the inquirer, consisting of name, surname and portrait, was available in all the scenarios, which are following presented:

- Basic scenario: the respondents did not know so much about the inquirer. Particularly, only the basic information set was available a priori any data disclosure;

- Familiar strangers scenario: the respondents knew the basic information set and the number of previous encounters with the inquirer a priori any data disclosure. Notably, encountering does not necessarily imply interaction - they may have just

passed by each other without noticing. Specifically, in this scenario the respondents had already encountered the inquirer 280 times;

- Mutual friends scenario: additionally to the basic information set, the respondents knew the number of mutual friends with the inquirer a priori any data disclosure. Specifically, in this scenario respondents had 15 mutual friends with the inquirer;

- Access & Control scenario: the respondents only knew the basic information set about the inquirer a priori any data disclosure. Moreover, it was explicitly emphasized that they can always edit/delete their disclosed personal information. Thus, respondents were empowered to control the inquirer's access to their disclosed personal data at any time in the future;

- Purpose scenario: additionally to the basic set of information, the respondents also knew other personal data regarding the inquirer a priori any data disclosure. This data indicated that the inquirer was a project manager in a major company within the respondent's professional area. Moreover, the inquirer's professional targets were also available beforehand and particularly matching with the ones of the respondents.

4 Participants of the surveys

The two questionnaires were distributed to 500 potential respondents. The distribution of the questionnaires was limited to online social networks users. We determined this category to be the most relevant because of their advanced experience with personal data disclosure in online social networks sites, even if the perceptions of data disclosure might vary between virtual and physical worlds. Due to anonymity of the responses and different timeframes of the surveys, it cannot be ensured that the respondents of both surveys completely match, however a significant overlap is expected.

Respondents were asked to provide information about their demographics characteristics. We focused on three demographic features, namely gender, age and occupation, which were further applied for clustering purpose. Moreover, respondents were asked to indicate their privacy settings in their main OSN site, such as visibility of their user profile, pictures, posts to the other users. Based on these answers, we were able to observe patterns among data disclosure attitudes. Consequently, we classified the participants into three privacy clusters, following the Westin/Harris privacy segmentation model (Westin, 1991):

- Fundamentalists: these respondents were extremely concerned about sharing their personal data with any other online social networks users (friends or strangers);

- Pragmatists: they also cared about the disclosure of their personal information. However, they often had specific concerns and particular strategies for addressing them. For example, this category of respondents generally preferred sharing personal information only among their friends;

- Unconcerned: these respondents were trusting online social networks sites and believing that the privacy of their data was not jeopardized. Thus, they were willing to share their personal data not only with people who were their friends, but as well with users who were complete strangers to them.

In the following we present an overview of the demographic information as well as privacy clusters of the respondents in both surveys.

4.1 Respondents of the first survey

In total we received 121 complete answers for the first survey, which composed the sample. In the following we present the demographic characteristics of the respondents:

- Gender: 54.5% of the respondents were males and 45.5% were females;

- Age: 64.5% of the respondents were between 26 and 35 years old, 28.1% were younger than 26 years and 7.4% were older than 35 years;

- Occupation: 75.2% of the respondents were working and the 24.8% were studying at the time of the survey.

In regard to privacy clusters, the sample of the first survey was composed as follows:

- Fundamentalists: 10.7% of the respondents;

- Pragmatists: 74.4% of the respondents;

- Unconcerned: 14.9% of the respondents.

4.2 Respondents of the second survey

The sample of the second survey was composed of 101 answers. The demographic characteristics of the respondents are following presented:

- Gender: 67.3% of the respondents were males and 32.7% were females;

- Age: 57.4% of the respondents were between 26 and 35 years old, 33.7% were younger than 26 years and 8.9% were older than 35 years;

- Occupation: 58.4% of the respondents were working and the 41.6% were studying at the time of the survey.

Moreover, following we present the privacy clusters of the sample of the second survey:

- Fundamentalists: 17.8% of the respondents;

- Pragmatists: 64.4% of the respondents;

- Unconcerned: 17.8% of the respondents.

5   Survey results and discussion

In order to investigate the influential factors, defined in Section 2, we relied on statistical characteristics and methods. First, we tested if the responses, grouped by different clusters, were normally distributed. We found out that many datasets of both surveys were not normally distributed. Consequently, we focused on analysis based on non-parametric statistical tests, due to expected higher precision of the results in comparison to the parametric tests (Moore and McCabe, 2005). Specifically, the Wilcoxon Signed Rank (in the following referred to as WSR) test was applied to examine the surveys' results by comparing two datasets and evaluate whether their population means differ (Wilcoxon, 1945). When statistically comparing two samples, they are considered to be statistically different if the *p-value* is observed to be less than the critical significance level, commonly set to *0.05*. However, when analyzing more than 2 datasets, to evaluate our results we used the Bonferroni correction in order to avoid potential type I errors (Weisstein, 2004). In these cases, the critical significance level is decreased to $0.05/n$, in which n is the total number of comparisons.

The following sections summarize the major results of our investigation. At the beginning we present and discuss results of the first survey followed be the ones of the second survey. Results are classified according to the privacy segmentation as well as the demographic characteristics, introduced in Section 4.

5.1   Results of the survey I

In this section we investigate the impact of user's location familiarity and current activity influential factors for the disclosure of personal information in mobile social networks.

5.1.1   Location familiarity

In order to evaluate the location familiarity influential factor, respondents were asked to choose which kind of personal information they would like to share in different locations, as described in Section 3.1. The selection of personal data to be disclosed was limited to a dataset composed of 28 different types of personal information.

Table 1 presents the standard deviations (σ), means (μ) and medians ($\tilde{x}$) of amounts of data shared in different user's locations. The mean and median results highlighted that respondents tend to share more personal information in familiar locations such as family places and work environments in comparison to less familiar places as work trip and holiday locations. This

inclination can be explained by the fact that the users spend the majority of their time in these places and thus they develop an unconscious trust in more familiar environments.

Indisputably, work environments and work trip comprise similar conditions because in both circumstances the user is still in his professional environment. However, the user's familiarity with these locations is notably different and it motivates significantly lower data sharing preferences in work trip in comparison to work environments. Similar results were also observed when analyzing social environments and holiday locations, however with lower overall impact of the location familiarity influential factor.

On the contrary, locations that comprise different conditions (i.e. working and leisure), such as holiday and work trip, presented relevantly low differences between the amounts of shared data in all the clusters. This inclination can be explained by the fact that both locations can be considered to be unfamiliar to the user as he/she is outside of his/her ordinary environment.

Table 2 shows results of statistical WSR test, comparing data disclosure in different locations. To account for multiple testing, we used the Bonferroni correction and considered significant only those p-values for which $P < 0.05/10 = 0.005$. As a result, we observed common statistically significant differences of data disclosure between all users' locations, except *Social Environments - Work Trip* and *Holiday - Work Trip*.

Similarly to results presented in Table 1, all the clusters presented statistical differences between *Work Environments - Work Trip*, except of the respondents older than 35 years ($> 35.p = .007$). As well, many clusters also presented statistical differences between *Social Environments - Holiday* locations. Moreover, evidence towards equal amount of data sharing were observed in *Holiday - Work Trip*, as tests of statistical differences between those locations presented considerably high *p-values*.

Comparing the responses of different clusters, no relevant differences were observed between males and females, except in *Social environments - Holiday* ($Fema.p = .003$, $Male.p = .006$). Moreover, it can be noticed that pragmatists, employed as well as respondents between 26 and 35 years old were more affected by the familiarity of user's location factor in comparison to the other relevant clusters. For example, data sharing results in *Work Environments - Social Environments* ($Pra.p = 26\text{-}35.p = Empl.p = .000$) and *Social environments - Holiday* ($Pra.p = 0.001$; $26\text{-}35.p = Empl.p = .000$) presented statistically significant differences in contrast to the other relevant clusters. Finally, no statistical differences were observed in any location comparison among the respondents older than 35 years.

Table 1 Descriptive statistics of information disclosure in different locations

			Family places	Work Env.	Social Env.	Holiday	Work trip
Privacy	Fund.	σ	5.64	3.26	6.47	5.43	5.66
		μ	17.54	12.46	6.85	6.23	5.92
		$\tilde{x}$	19	12	4	5	6
	Prag.	σ	7.42	5.88	6.50	6.34	6.20
		μ	18.97	15.82	12.02	10.69	10.68
		$\tilde{x}$	18.50	16	12	10.50	11
	Unco.	σ	4.19	4.16	6.07	5.34	5.58
		μ	23.89	19.61	15.83	14.67	13.94
		$\tilde{x}$	24.50	19	17.50	14.50	14.50
Gender	Male	σ	7.10	5.56	6.57	6.34	6.32
		μ	19.83	16.48	13.23	11.89	11.38
		$\tilde{x}$	20.50	16	14	12	11.50
	Fema.	σ	7.07	5.87	6.78	6.34	6.31
		μ	19.20	15.47	10.60	9.49	9.78
		$\tilde{x}$	21	15	10	8	10
Age	< 26	σ	8.56	5.78	6.19	5.64	5.53
		μ	18.24	14.88	12.12	11.67	10.85
		$\tilde{x}$	19	14	12	11	10
	26-35	σ	6.41	5.55	6.67	6.35	6.11
		μ	19.73	16.15	11.90	10.22	10.30
		$\tilde{x}$	20	16	12	10	11
	> 35	σ	6.10	6.07	9.99	9.39	10.40
		μ	22.67	19.11	12.89	12.78	13
		$\tilde{x}$	25	21	17	16	14
Occupation	Stud.	σ	7.51	5.91	6.31	5.80	6.50
		μ	19.67	15.17	12.37	11	10.40
		$\tilde{x}$	20	15	13	11.50	10
	Empl.	σ	6.96	5.64	6.94	6.65	6.32
		μ	19.51	16.31	11.92	10.74	10.74
		$\tilde{x}$	21	16	12	11	11

Table 2 Results of Wilcoxon Signed Rank test for information disclosure in different locations

		FP-WE	FP-SE	FP-H	FP-WT	WE-SE	WE-H	WE-WT	SE-H	SE-WT	H-WT
Privacy	Fund.	.004	.001	.001	.001	.009	.003	.002	.324	.301	.611
	Prag.	.000	.000	.000	.000	.000	.000	.000	.001	.050	.603
	Unco.	.002	.000	.001	.000	.032	.003	.000	.082	.138	.507
Gender	Male	.000	.000	.000	.000	.001	.000	.000	.006	.009	.486
	Fema.	.000	.000	.000	.000	.000	.000	.000	.003	.364	.463
Age	< 26	.005	.000	.000	.000	.022	.010	.000	.421	.273	.493
	26-35	.000	.000	.000	.000	.000	.000	.000	.000	.012	.554
	> 35	.051	.008	.008	.008	.011	.008	.007	.483	.999	.892
Occup.	Stud.	.000	.000	.000	.000	.016	.000	.000	.108	.119	.699
	Empl.	.000	.000	.000	.000	.000	.000	.000	.000	.036	.753

FP: Family Places; **WE**: Work Environments; **SE**: Social Environments; **H**: Holiday; **WT**: Work Trip.

5.1.2 Current activity

In order to evaluate the user's current activity influential factor, we focused on two different datasets: data related to work activities (DWA) and data related to social interactions (DSI). Both datasets were composed of nine different types of personal information, which were subsets of the full dataset of the first survey. For example, data related to working activities is employer, work phone number, career skills and abilities, while examples of data related to social interactions are relation status, food taste, interests, etc. In this analysis we compared these two datasets in their associated environments, i.e. social and work environments.

In Table 3 we present the standard deviations (σ), means (μ) and medians ($\tilde{x}$) of the amounts of DWA and DSI, disclosed in both work and social environments. The results showed significantly different

sharing preferences between the two analyzed datasets. In work environments, the mean and median values of DWA were considerably higher than DSI and notably close to the possible maximum amount of shared data, i.e. 9. Similar patterns were also observed in regard to social environments, in which DSI achieved higher sharing rate in comparison to DWA. However, the current activity influential factor presented a lower impact in social environments as the difference between sharing of DSI and DWA was considerably lower than the one in work environments.

These results were confirmed to be statistically significant by the WSR test. As shown in Table 4-A, we firstly compared amounts of DWA and DSI shared in work environments and afterwards in social environments. In both circumstances, all the clusters presented statistically significant differences between DWA and DSI, except respondents older than 35 years ($> 35.p = .748$) and the unconcerned privacy

Table 4 Results of Wilcoxon Signed Rank test for information disclosure during different activities

		(A) DWA-DSI		(B) WE-SE	
		WE	SE	DWA	DSI
Privacy	Fund.	.001	.024	.001	.559
	Prag.	.000	.000	.000	.101
	Unco.	.001	.050	.000	.647
Gender	Male	.000	.000	.000	.180
	Fema.	.000	.000	.000	.371
Age	< 26	.000	.000	.000	.138
	26-35	.000	.000	.000	.200
	> 35	.007	.748	.011	.104
Occupation	Stud.	.000	.000	.000	.013
	Empl.	.000	.000	.000	.536

WE: Work Environments; **SE**: Social Environments; **DWA**: Data related to Work Activities; **DSI**: Data related to Social Interactions.

cluster ($Unco.p = .050$) in social environments. However, it must be noted that respondents older than 35 years presented a strong evidence of similarity, while unconcerned privacy clusters showed relevant differences, even if not statistically significant.

In order to complement the investigation of the current user's activity as crucial determinant, we also compared data disclosure between work and social environments by testing separately DWA and DSI, as shown in Table 4-B. The results indisputably proved that the user's current activity factor had different impact on different data types, i.e. DWA and DSI. Notably, all the clusters differentiated DWA between work and social environments. However, the same tendency was not observed in regard to DSI. In fact, the only significant statistical difference was presented among the students ($Stud.p = .013$). No other significant differences among the clusters were observed, despite an overall increasing influence of user's current activity among the pragmatists privacy cluster and respondents older than 35 years ($Prag.p = .101; > 35.p = .104$). Consequently, results of Table 4-B proved that differences between DWA and DSI in Table 4-A were mainly caused by significant differentiation of DWA sharing preferences upon different activities.

Table 3 Descriptive statistics of information disclosure during different users' activities

			WE		SE	
			DWA	DSI	DWA	DSI
Privacy	Fund.	σ	1.33	1.30	1.94	2.73
		μ	7.46	3.23	1.62	3.15
		$\tilde{x}$	7	3	1	3
	Prag.	σ	2.00	2.33	2.18	2.48
		μ	7.48	4.66	3.12	5.17
		$\tilde{x}$	8	5	3	5.50
	Unco.	σ	0.75	1.98	2.34	2.24
		μ	8.72	5.94	4.94	6.22
		$\tilde{x}$	9	6	5.50	6.50
Gender	Male	σ	1.73	2.20	2.38	2.44
		μ	7.61	5.03	3.67	5.45
		$\tilde{x}$	8	5	3	6
	Fema.	σ	1.99	2.34	2.17	2.69
		μ	7.73	4.29	2.71	4.69
		$\tilde{x}$	9	5	3	4
Age	< 26	σ	1.73	2.30	2.02	2.48
		μ	7.48	4.36	3.30	5.09
		$\tilde{x}$	8	4	3	5
	26-35	σ	1.96	2.25	2.20	2.55
		μ	7.62	4.72	3.06	5.16
		$\tilde{x}$	8	5	3	5
	> 35	σ	0.71	2.45	3.91	3.32
		μ	8.67	5.67	4.44	4.67
		$\tilde{x}$	9	7	5	5
Occup.	Stud.	σ	1.78	2.29	2.10	2.27
		μ	7.53	4.47	3.07	5.57
		$\tilde{x}$	8	4	3	6
	Empl.	σ	1.88	2.29	2.40	2.66
		μ	7.70	4.77	3.29	4.96
		$\tilde{x}$	9	5	3	5

WE: Work Environments; **SE**: Social Environments; **DWA**: Data related to Work Activities; **DSI**: Data related to Social Interactions.

5.2 Results of the survey II

In this section we investigate the impact of the following influential factors for the disclosure of personal information in mobile social networks: inquirer, access & control and purpose of disclosure.

5.2.1 Inquirer and Access & Control

In order to research on the inquirer influential factor we investigated the impact of being familiar strangers as well as having mutual friends with the inquirer. Afterwards, we analyzed the influence of explicit

emphasis of access & control rights to the users. To evaluate the impact of these influential factors,

Table 5 Descriptive statistics of information disclosure in basic, familiar strangers mutual friends and access & control scenarios

			Basic	FS	MF	AC
Privacy	Fund.	σ	3.11	3.47	4.10	3.63
		μ	2.17	3.17	4	3.50
		$\tilde{x}$	0	3	3	4
	Prag.	σ	3.05	4	4.12	4.06
		μ	4.72	6.69	7.08	7.22
		$\tilde{x}$	5	7	7	7
	Unco.	σ	3.50	3.51	3.63	3.17
		μ	7.17	10.11	9.89	9.56
		$\tilde{x}$	6	10	10	9.50
Gender	Male	σ	3.52	4.26	4.53	4.45
		μ	4.96	6.91	7.32	7
		$\tilde{x}$	5	7	7	7
	Fema.	σ	3.32	4.48	4.03	3.82
		μ	4.18	6.18	6.42	6.91
		$\tilde{x}$	4	7	6	7
Age	< 26	σ	3.43	4.66	4.26	4.48
		μ	5.21	8.21	7.44	8.35
		$\tilde{x}$	5	7.5	7	9
	26-35	σ	3.41	3.98	4.46	3.88
		μ	4.17	5.67	6.67	6
		$\tilde{x}$	3.50	6	7	6
	> 35	σ	3.49	3.67	4.47	4.27
		μ	6.22	7.33	7.78	8
		$\tilde{x}$	6	8	9	7
Occup.	Stud.	σ	3.75	4.49	4.48	4.79
		μ	5.05	7.43	7.29	7.74
		$\tilde{x}$	5	7.50	7	8
	Empl.	σ	3.24	4.15	4.32	3.73
		μ	4.46	6.14	6.85	6.42
		$\tilde{x}$	5	6	7	7

FS: Familiar Strangers; **MF**: Mutual Friends; **AC**: Access & Control

we compared the responses in regard to the relevant scenarios, described in Section 3.2. We focused on additional personal information, which was not shared in the basic scenario, however it was preferred to be disclosed in the familiar strangers, mutual friends or access & control scenarios.

Table 5 presents the standard deviations (σ), means (μ) and medians ($\tilde{x}$) of the amounts of data shared in the analyzed scenarios. The median and mean results in familiar strangers, mutual friends and access & control scenarios presented higher data sharing preferences in comparison to the basic scenario, where no connections with the inquirer were highlighted and no access & control rights were emphasized. Moreover, it was observed that unconcerned privacy cluster and respondents younger than 26 years were the most impacted by the influential factors. In fact, they presented higher mean and median differences between the basic and other scenarios in comparison to the other clusters.

Table 6 presents WSR test results obtained by comparing the responses associated to the four scenarios. To account for multiple testing between different scenarios, we used the Bonferroni correction and considered significant only those p-values for which $P < 0.05/6 = 0.008$. In regard to the inquirer influential factor, all the clusters generally presented significant differences between sharing of personal information in the basic and mutual friends/familiar strangers scenarios. Particularly, statistically significant differences were not observed only among fundamentalists privacy cluster in *Basic - Familiar Strangers* ($Fund.p = .011$) as well as respondents older than 35 years in *Basic - Familiar Strangers* ($> 35.p = .039$) and *Basic - Mutual Friends* ($> 35.p = .026$). Moreover, when comparing mutual friends and familiar strangers scenarios, significant differences were observed only among working respondents ($Empl.p = .003$) and respondents between 26 and 35 years old ($26\text{-}35.p = .000$). No other significant differences were presented, despite an overall

Table 6 Results of Wilcoxon Signed Rank test for information disclosure in basic, familiar strangers, mutual friends and access & control scenarios

		Basic-FS	Basic-MF	Basic-AC	FS-MF	FS-AC	MF-AC
Privacy	Fund.	.011	.003	.026	.107	.905	.402
	Prag.	.000	.000	.000	.064	.092	.655
	Unco.	.001	.001	.001	.908	.302	.430
Gender	Male	.000	.000	.000	.104	.935	.234
	Fema.	.000	.000	.000	138	.121	.343
Age	< 26	.000	.000	.000	.292	.913	.057
	26-35	.000	.000	.000	.000	.322	.031
	> 35	.039	.026	.026	.279	.245	.786
Occupation	Stud.	.000	.000	.000	.903	.680	.228
	Empl.	.000	.000	.000	.003	.301	.188

FS: Familiar Strangers; **MF**: Mutual Friends; **AC**: Access & Control

increase of the difference between *Familiar Strangers - Mutual Friends* among the pragmatists privacy cluster ($Prag.p = .064$).

Similarly to the previous results, *Basic - Access & Control* also presented statistically significant evidence of differences among the clusters, except of respondents older than 35 years and fundamentalists ($> 35.p = Fund.p = .026$). Moreover, no statistically significant differences were observed when comparing access & control with familiar strangers/mutual friends scenarios, despite an overall increase of differences in *Familiar Strangers - Access & Control* among pragmatists and females ($Prag.p = .092; Fema.p = .121$) and in *Mutual Friends -Access & Control* among respondents younger than 35 years ($< 26.p = .057; 26-35 = .031$).

Comparing the privacy and demographic clusters, the fundamentalists were the only privacy cluster not affected by familiar strangers and access & control factors. We did not observe any differences between clusters within gender and occupation segments. Moreover, respondents older than 35 years were not influenced by any factor presented in this section in contrast to the other age groups.

Finally, these results confirmed that when access & control rights were clearly emphasized, the respondents were motivated to disclose more personal information as they might feel not to lose control over their personal data, even after actual disclosure.

5.2.2 Purpose of disclosure

In order to evaluate the impact of the purpose of disclosure influential factor, we compared data sharing preferences in basic and purpose scenarios. As described in Section 3.2, the purpose scenario focused on potential professional networking benefits, even if it occurred in a social environment. This test was limited to personal information related to work activities. Specifically, the chosen dataset was composed of 7 different types of user's personal information, which was a subset of the full dataset of the second survey. Examples of data related to work activities are employer, career skills and abilities, education details, etc.

Table 7 presents the standard deviations (σ), means (μ) and medians ($\tilde{x}$) of amounts of information shared in basic and purpose scenarios. In the purpose scenario, the values of mean and median were considerably higher than in basic scenario and very close to the maximum possible amount of shared data, i.e. 7. Comparing the responses of different demographic and privacy clusters, we did not observe important mean and median differences within the age and gender segments, however fundamentalists and employed respondents were the most impacted by the purpose of disclosure in comparison to the other relevant clusters.

Table 8 illustrates the WSR test results, obtained by comparing the responses of the basic and purpose scenarios. It can be noticed that all the clusters presented

Table 7 Descriptive statistics of information disclosure in basic and purpose scenarios

			Basic	Purpose
Privacy	Fund.	σ	1.98	1.94
		μ	1.56	5.33
		$\tilde{x}$	0.50	6
	Prag.	σ	1.78	1.81
		μ	2.69	4.98
		$\tilde{x}$	3	5
	Unco.	σ	2.03	2.40
		μ	3.67	5
		$\tilde{x}$	3	5.50
Gender	Male	σ	1.96	2.03
		μ	2.71	5.01
		$\tilde{x}$	3	5.50
	Fema.	σ	1.99	1.73
		μ	2.48	5.12
		$\tilde{x}$	3	5
Age	< 26	σ	1.90	2.30
		μ	2.71	4.85
		$\tilde{x}$	3	6
	26-35	σ	1.95	1.71
		μ	2.43	5.14
		$\tilde{x}$	2	5
	> 35	σ	2.18	1.92
		μ	3.67	5.22
		$\tilde{x}$	4	5
Occupation	Stud.	σ	2.07	2.14
		μ	2.95	4.55
		$\tilde{x}$	3	5
	Empl.	σ	1.87	1.69
		μ	2.41	5.41
		$\tilde{x}$	2	6

statistically significant differences between the amounts of data, shared in basic and purpose scenarios.

Finally, results of Table 7 and Table 8 proved that all the users were willing to share more personal information, when they had reasons to disclose their data, e.g. they had the possibility to predict potential professional networking benefits.

Table 8 Results of Wilcoxon Signed Rank test for information disclosure in basic and purpose scenarios

		Basic- Purpose
Privacy	Fund.	.001
	Prag.	.000
	Unco.	.037
Gender	Male	.000
	Fema.	.000
Age	< 26	.000
	26-35	.000
	> 35	.014
Occupation	Stud.	.000
	Empl.	.000

6 Conclusions

In this paper we provided insight into personal data sensitivity in order to contribute to design of privacy management systems for mobile social networks. Firstly, we outlined the relevant influential factors that might impact users' personal information disclosure decisions. Afterwards, we empirically investigated the most relevant determinants for data disclosure in mobile social networks, which promote networking not only among acquaintances, but also between strangers with interpersonal affinities in the physical world.

According to our analysis, the purpose of data disclosure was found to be the most determinant factor for privacy preferences among the ones tested, as it is statistically proven to be affecting all the respondent clusters. Moreover, emphasizing access & control rights was proven to help users to feel more secure to share their personal information. Thus, we strongly encourage privacy designers to take into account purpose of data disclosure factor as primary index into users' privacy preferences as well as apply and clearly emphasize access & control rights.

Further, following the results of our research we also suggest designers of privacy systems to consider the other influential factors, however as indexes of secondary importance. Particularly, the location familiarity factor was commonly approved by all the respondents who presented tendency to be more open to share their personal information in more familiar locations. Moreover, our analysis proved that knowing beforehand information about the inquirer, such as number of mutual friends or previous encounters, relevantly impacted the information disclosure decisions. Finally, the investigation of the activity factor presented different impact in relation to different data types. This factor was observed to be significantly influential only on disclosure of data related to work activities. In fact, respondents did not significantly differentiate sharing of data, related to social interactions, between work and social environments.

In regard to the demographic and privacy clusters, our analysis did not show relevant differences between data sharing among male and female clusters. The pragmatists privacy cluster and respondents between 26 and 35 years old were overall the most affected by the influential factors, in comparison to the other relevant clusters. Furthermore, the fundamentalist privacy cluster was the most influenced by the purpose of disclosure determinant, even if they were generally only slightly impacted by the other factors. Finally, respondents older than 35 years did not present impact of any influential factors, except of the purpose of disclosure.

The results of our research strongly encourage further research on the influential factors, discussed in Section 2, for the disclosure of personal information in mobile social networks. Particularly, a qualitative investigation would be a relevant supplement to the results of the quantitative research, presented in this paper. Moreover,

the current mood of the user might also be considered as a determinant for data disclosure. This potential influential factor is suggested to be taken into account during qualitative investigations in the future work.

References

Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. (2002). Hippocratic databases. In *Proceedings of the 28th international conference on Very Large Data Bases*, pages 143–154. VLDB Endowment.

Altman, I. (1975). The environment and social behavior: Privacy, personal space, territory, and crowding.

Altman, I. (1977). Privacy regulation: culturally universal or culturally specific? *Journal of Social Issues*, 33(3):66–84.

Barkhuus, L., Brown, B., Bell, M., Sherwood, S., Hall, M., and Chalmers, M. (2008). From awareness to repartee: sharing location within social groups. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 497–506. ACM.

Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., and Mylopoulos, J. (2004). Tropos: An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems*, 8(3):203–236.

Bunnig, C. (2009a). Simulation and analysis of ad hoc privacy control in smart environments. *Intelligent Interactive Assistance and Mobile Multimedia Computing*, pages 307–318.

Bunnig, C. (2009b). Smart privacy management in ubiquitous computing environments. *Human Interface and the Management of Information.Information and Interaction*, pages 131–139.

Bunnig, C. and Cap, C. H. (2009). Ad hoc privacy management in ubiquitous computing environments. In *2009 Second International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services*, pages 85–90. IEEE.

Byun, J. W., Bertino, E., and Li, N. (2005). Purpose based access control of complex data for privacy protection. In *Proceedings of the tenth ACM symposium on Access control models and technologies*, pages 102–110. ACM.

Byun, J. W. and Li, N. (2008). Purpose based access control for privacy protection in relational database systems. *The VLDB Journal - The International Journal on Very Large Data Bases*, 17(4):603–619.

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

Counts, S. and Fisher, K. E. (2008). Mobile social networking: An information grounds perspective. In *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, page 153. IEEE.

Davis, S. and Gutwin, C. (2005). Using relationship to control disclosure in awareness servers. In *Proceedings of Graphics Interface 2005*, pages 145–152. Canadian Human-Computer Communications Society.

Directive, E. (1995). 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the EC*, 23.

Eagle, N. and Pentland, A. (2005). Social serendipity: Mobilizing social software. *IEEE Pervasive Computing*, pages 28–34.

Gregg, R. E. (1975). The privacy act of 1974. *Army Law.*, page 25.

Hong, J. I., Ng, J. D., Lederer, S., and Landay, J. A. (2004). Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 91–100. ACM.

Iachello, G., Smith, I., Consolvo, S., Chen, M., and Abowd, G. D. (2005). Developing privacy guidelines for social location disclosure applications and services. In *Proceedings of the 2005 symposium on Usable privacy and security*, pages 65–76. ACM.

Jendricke, U., Kreutzer, M., and Zugenmaier, A. (2002). Pervasive privacy with identity management. In *Proceedings of the Workshop on Security in Ubiquitous Computing, Ubicomp*, volume 2002. Citeseer.

Jones, Q., Grandhi, S. A., Whittaker, S., Chivakula, K., and Terveen, L. (2004). Putting systems into place: a qualitative study of design requirements for location-aware community systems. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 202–211. ACM.

Kapadia, A., Henderson, T., Fielding, J., and Kotz, D. (2007). Virtual walls: Protecting digital privacy in pervasive environments. *Pervasive Computing*, pages 162–179.

Langheinrich, M. (2001). Privacy by design - principles of privacy-aware ubiquitous systems. In *Ubicomp 2001: Ubiquitous Computing*, pages 273–291. Springer.

Lederer, S., Hong, J. I., Dey, A. K., and Landay, J. A. (2004). Personal privacy through understanding and action: five pitfalls for designers. *Personal and Ubiquitous Computing*, 8(6):440–454.

Lederer, S., Mankoff, J., and Dey, A. K. (2003a). Who wants to know what when? privacy preference determinants in ubiquitous computing. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 724–725. ACM.

Lederer, S., Mankoff, J., Dey, A. K., and Beckmann, C. (2003b). *Managing personal information disclosure in ubiquitous computing environments*. Citeseer.

Lilien, L., Kamal, Z. H., Bhuse, V., and Gupta, A. (2006). Opportunistic networks: the concept and research challenges in privacy and security. *Proc.of the WSPWN*, pages 134–147.

Massacci, F., Mylopoulos, J., and Zannone, N. (2006). Hierarchical hippocratic databases with minimal disclosure for virtual organizations. *The VLDB journal*, 15(4):370–387.

Milgram, S. (1977). The familiar stranger: An aspect of urban anonymity. *The individual in a social world*, pages 51–53.

Moore, D. S. and McCabe, G. P. (2005). *Introduction to the Practice of Statistics Chapters 14-17*. WH Freeman & Co.

Olson, J. S., Grudin, J., and Horvitz, E. (2005). A study of preferences for sharing and privacy. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1985–1988. ACM.

Palen, L. and Dourish, P. (2003). Unpacking privacy for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–136. ACM.

Sapuppo, A. (2010). Spiderweb: a social mobile network. In *Wireless Conference (EW), 2010 European*, pages 475–481.

Sapuppo, A. and Sørensen, L. T. (2011). Local social networks. In *International Proceedings of Computer Science and Information Technology - Computer Communication and Management*, volume 5, pages 15–22.

Smith, I., Consolvo, S., Lamarca, A., Hightower, J., Scott, J., Sohn, T., Hughes, J., Iachello, G., and Abowd, G. D. (2005). Social disclosure of place: From location technology to communication practices. *Pervasive Computing*, pages 134–151.

Tian, Y., Song, B., and Huh, E. N. (2009). A privacy-aware system using threat-based evaluation and feedback method in untrusted ubiquitous environments. *Security Technology*, pages 193–200.

Weisstein, E. W. (2004). Bonferroni correction. *MathWorldA Wolfram Web Resource*.

Westin, A. F. (1991). Harris-equifax consumer privacy survey 1991. *Atlanta, GA: Equifax Inc.*

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.

Wright, D., Gutwirth, S., Friedewald, M., Hert, P. D., Langheinrich, M., and Moscibroda, A. (2009). Privacy, trust and policy-making: Challenges and responses. *Computer Law & Security Report*, 25(1):69–83.

Yee, G. (2010). Using privacy policies to protect privacy in ubicomp. In *Proceedings of the IEEE 19th International Conference on Advanced Information Networking and Applications (AINA 2005Volume II)*.

Ziv, N. D. and Mulloth, B. (2007). An exploration on mobile social networking: Dodgeball as a case in point. In *Mobile Business, 2006. ICMB'06. International Conference on*, page 21. IEEE.