# The Influential Factors for the Variation of Data Sensitivity in Ubiquitous Social Networking

Sapuppo, Antonio

*Published in:*
International Journal of Wireless and Mobile Computing

*DOI (link to publication from Publisher):*
10.1504/IJWMC.2013.054046

*Publication date:*
2013

*Document Version*
Early version, also known as pre-print

# The influential factors for the variation of data sensitivity in ubiquitous social networking

**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:** Ubiquitous social networking services offer new opportunities for developing advantageous relationships by uncovering hidden connections that people share with others nearby. As sharing of personal information is an intrinsic part of ubiquitous social networking, these services are subject to crucial privacy threats. In order to contribute to the design of privacy management systems, we present results of a mixed methods study that investigated the influential factors for the variation of human data sensitivity upon different circumstances. The results indicate that the users' information sensitivity is decreasing inversely proportionally to the relevance of data disclosure for initiation of relationships with others. We suggest privacy designers to take into account the purpose of disclosure and environment as primary indexes for data disclosure. Other influential factors, i.e. activity, mood, location familiarity, number of previous encounters and mutual friends, were as well discovered to influence participants' data disclosure, but as factors of secondary importance.

**Keywords:** privacy; ubiquitous computing; information disclosure; social networking.

**Reference** to this paper should be made as follows: Sapuppo, A. (2013) 'The influential factors for the variation of data sensitivity in ubiquitous social networking', *Int. J. Wireless and Mobile Computing*, Vol. 6, No. 2, pp.115–130

**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 computing originated with many people serving one computer and has gradually evolved into the currently existing possibility for many computers to serve one person anywhere around the world (Weiser, 1996). The latest wave of computing, called ubiquitous computing, shifts the central focus of users' attention away from the computers by embedding many seamless highly specialized devices within people's surroundings (Weiser, 1995). These devices are aware of their current environments and users and, consequently, they are able to improve humans' lives and support their everyday tasks (Weiser and Brown, 1996).

For ubiquitous computing applications to intelligently and naturally support humans who are by nature social beings, it is essential to embody social intelligence, which can be defined as the ability of the environment to acquire and apply users' social context (Dey and Abowd, 2000; Suh and Woo, 2011; Terano, 2001). This led to the development of ubiquitous social computing, where a social dimension has been introduced in order to increase awareness, knowledge and intelligence of ubiquitous computing environments (Youngblood et al., 2006).

The establishment of ubiquitous social computing allows the possibility to transfer online social networking benefits to the physical world, by promoting ubiquitous social networking (in the following referred to as USN) services. These services target at developing possible advantageous relationships such as friendships, partnerships, business relations by uncovering hidden connections that people share with others nearby and thus facilitating initialization of face-to-face interactions between people who do not know each other, but probably should. As a result, the value of social

networking is significantly enhanced and benefits are available immediately upon demand (Eagle and Pentland, 2005; Gupta et al., 2009; Pietiläinen et al., 2009; Tamarit et al., 2009).

When transferring online social networking benefits to the physical world, the privacy threats are indisputably increased due to support of face-to-face interactions between strangers during physical meetings (Sapuppo, 2012b). 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 environments. For example, when USN 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 (Sapuppo, 2012b). To address these privacy concerns, privacy management systems should protect users' personal data privacy as individuals do in ordinary human interactions (Bünnig, 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 and ensure accuracy of selective disclosure of personal information, it is necessary to gain an extensive comprehension of variation of human data sensitivity that affects information disclosure under different circumstances.

In previous studies, the identity of the inquirer was identified as the primary index for selection of data disclosure decisions in ubiquitous computing (Lederer et al., 2003; Davis and Gutwin, 2005; Olson et al., 2005; Jones et al., 2004; Consolvo et al., 2005). On the other hand, USN services advance the attention to other factors as crucial determinants for data disclosure, due to their primary focus on initiation of relationships between strangers (Sapuppo and Sørensen, 2011; Sapuppo and Seet, 2012). Several influential factors that impact personal data disclosure in USN, such as users' current activities and location familiarity, were identified during an empirical investigation, based on predefined data disclosure preferences (Sapuppo, 2012a). This investigation provided statistically significant results, obtained by asking participants to predict their sharing preferences a priori the actual data disclosure. However, there might be a difference between what people say they want to share and what they actually do share in practice (Iachello and Hong, 2007). Furthermore, data disclosure decisions taken at the moment of actual disclosure in ubiquitous social computing environments were found to be more accurate in comparison to predefined privacy preferences, as users might encounter circumstances where data disclosure decisions are not precisely predictable (Jendricke et al., 2002; Bünnig, 2009b,a; Bünnig and Cap, 2009; Lederer et al., 2004). Consequently, it is important to investigate the previously identified influential factors for variation of human data sensitivity, by analyzing these factors based on in situ data disclosure privacy preferences, as well as gain an extensive understanding of people' attitudes and motivations that govern such data disclosure decisions in USN.

In order to achieve these goals, we applied a sequential two-phase mixed methods study for analysis of ad hoc data disclosure preferences in USN with active online social networks users. In the first phase, a quantitative research investigated the relationship between the identified influential factors and ad hoc data disclosure decisions. By exploiting a USN prototype, we collected participants' ad hoc data disclosure decisions and applied the binary logistic regression statistical model for examining whether the selected influential factors can be considered as predictors for users' data disclosure decisions in USN. Information, acquired during the first phase of the study, was explored further in the second phase, where qualitative interviews were used to gain in-depth understanding of different aspects and motivations of users' data disclosure in USN. The results of our mixed methods analysis can provide significant input for the design and development of privacy management systems for USN environments.

The rest of the paper is structured as follows: firstly, we discuss related work regarding influential factors for the disclosure of personal information. In Section 3, we list and define each of the influential factors that might impact on users' data sharing preferences in ubiquitous social networking. In Section 4, we present the design and methodology of the mixed methods study. Further, the information about the participants of our investigation is provided in Section 5. In Section 6, we present the major findings of the analysis. Final conclusions and recommendations for future work are drawn in Section 7.

## 2 Related Work

In past works, different studies have questioned whether the sensitivity of personal data remains unchanged upon different circumstances. In (Lederer et al., 2003), the authors found the sensitivity 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. Other studies provided further insight into the inquirer influential factor by emphasizing that users differentiate choices of disclosure of personal information upon relationships with the inquirer. In fact, they indicated that self-reported closeness was a crucial factor for deciding whether to disclose their personal information to a specific inquirer (Davis and Gutwin,

2005; Wiese et al., 2011). Moreover, other research also highlighted the need to cluster users into manageable categories of inquirers (e.g. friends, family members, co-workers, etc) for taking users' data disclosure decisions, in order to better preserve their data privacy (Olson et al., 2005; Jones et al., 2004).

Even if defining the identity of the inquirer as a crucial parameter, in (Consolvo et al., 2005) the authors investigated other factors that might impact users' personal data disclosure decisions. Firstly, they analyzed 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 assume that it is more useful for the inquirer, rather than for preserving their privacy. Secondly, the authors also indicated users' current mood and activities as relevant factors for personal data disclosure. The former implies that users differentiate their data disclosure upon their humor, e.g. participants were most willing to disclose their personal data when "depressed", in contrast to being "angry". In regard to users' current activity, Consolvo et al discussed that during some activities (e.g. exercising) users were more inclined to share their personal information rather than others, such as studying. Thirdly, the authors indicated that knowing the particular reason for data disclosure would also significantly motivate users to share their personal information (Consolvo et al., 2005). The purpose of disclosure was also researched in other studies, which attempted to ensure that users' personal data is processed for only the intended reason (Byun et al., 2005; Byun and Li, 2008; Petkovic et al., 2011; Tian et al., 2009).

Another relevant factor that might impact users' personal data disclosure decisions is anonymity. 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. Having the possibility to remain anonymous would significantly increase data privacy protection and consequently might influence users' personal information disclosure decisions (Langheinrich, 2001). However, in case of necessity for the users' authentication, the pseudoanonymity approach could be applied. Pseudoanonymity is realized through linking the users to IDs, which represent them in specific circumstances and allow them to be recognized as long as they use the same ID (Langheinrich, 2001; Beresford and Stajano, 2003).

All the aforementioned studies, even if acknowledging the existence of other relevant influential factors, commonly indicated the identity of the inquirer as the most crucial determinant for data disclosure in ubiquitous computing environments. On the contrary, in (Sapuppo, 2012a) the author advanced the attention to analysis of other influential factors for data disclosure, due to specific focus on USN and consequent initialization of relationships between strangers. The findings strongly encourage privacy designers of USN to take into account the purpose of data disclosure factor as the primary index for decisions about data disclosure to strangers. Moreover, it strongly recommends to consider the access & control influential factor, which implies the right for users to be able to influence other people's access to one's personal data, even after the actual disclosure. Further, it also suggests to consider other influential factors for the disclosure of personal information, however as indexes of secondary importance, i.e. familiarity with the current location, current activity and other information about inquirer, such as the number of previous encounters and mutual friends (Sapuppo, 2012a). The mutual friends influential factor was also confirmed to significantly impact personal data disclosure decisions in online social networks, where users were proven to be much more likely to disclose sensitive information to strangers if they have a friend in common (Nagle and Singh, 2009).

The mixed methods study, presented in this paper, is based on the findings of the past work, discussed in this section. Especially, the results presented in this article target at complementing the outcomes of the empirical analysis in USN, described in (Sapuppo, 2012a). However, differently from (Sapuppo, 2012a), in this investigation we focus on analyzing ad hoc data disclosure decisions, which were proven to be more accurate for preserving users' data privacy in USN (Jendricke et al., 2002; Bünnig, 2009b,a; Bünnig and Cap, 2009). Moreover, we quantitatively analyze additional influential factors that were not taken into consideration in the previous empirical analysis (Sapuppo, 2012a). We refer to research on the current mood and different aspects of the users' current location: type of current environment (e.g. work, social, holiday) and location familiarity, evaluated according to the amount of time that the users usually spend in a specific location (e.g. daily, monthly, first time in this location, etc). Finally, in this mixed methods study, we also supplemented empirical results with findings of qualitative interviews in order to further explore the participants' attitudes towards the impact of influential factors on their data disclosure decisions.

## 3 The influential factors

In this section we present the influential factors that might impact users' data disclosure decisions in USN. Importantly, some of the factors, introduced in Section 2, were not taken into consideration. We refer to anonymity, granularity of disclosed information and identity of the inquirer. Even if acknowledging the importance of these influential factors for users' overall data disclosure, we did not find them to be relevant in USN for the reasons described in the following.

Firstly, the granularity of disclosed information is often applied in disclosure of social locations among acquaintances and refers to extent of details of current geo location, e.g. country, city or neighborhood to the exact address where the user is (Iachello et al., 2005; Smith et al., 2005; Barkhuus et al., 2008). However, USN services commonly exploit opportunistic networks to promote social networking between strangers during physical meetings (Eagle and Pentland, 2005; Sapuppo, 2012b; Tamarit et al., 2009). In opportunistic networks, the data exchange is restricted to the range of the adopted wireless technology and thus users are aware about each others' location only when they are in the proximity (Sapuppo, 2012b). Consequently, the granularity of the disclosed information was not considered as a relevant factor for data disclosure in USN. Secondly, the anonymity and pseudoanonymity influential factors were not also included in this research because it would cause significant losses of potential networking opportunities. In USN users must be identifiable, as they must allow others to link their profiles to *real* people in order to have the possibility to gain USN benefits. Thirdly, we did not consider the identity of the inquirer to be a relevant influential factor in USN, due to its focus on initiation of relationships between strangers. Instead, we took into account other information that the users have in common with the inquirer, such as number of mutual friends and previous encounters.

The selected influential factors for data disclosure in USN were clustered into three different groups: contextual information, interrelated attributes and design properties. The first group regards influential factors related to the current contextual circumstances of the users' encounters in USN environments, e.g. where is the user, what he is doing, etc. The second group of influential factors consists of information that the user has in common with the inquirer, e.g. similar music preferences or number of mutual friends. Finally, the last group corresponds to design solutions that should be taken into consideration when implementing USN services. A definition of each of the selected influential factors is following provided.

*Contextual data:*

- Environment: is considered to be the current location of the users, grouped according to their ordinary activities in that location, e.g. work environments, social environments, work trips, etc.
- Location familiarity: is considered to be the users' familiarity with their current location evaluated according to the amount of time that users usually spend in a specific location, e.g. daily, monthly, first time in this location, etc.
- Activity: is considered to be the current action of the user, e.g. working, relaxing, etc.
- Mood: is considered to be the users' current status of emotion, e.g. depressed, happy, sad, angry, etc.

*Interrelated attributes:*

- Familiar strangers: is considered to be the number of times that the users have already encountered the inquirer, e.g. 120 times in the last 3 weeks, etc. Notably, encountering does not necessarily imply interaction - they may have just passed by each other without noticing.
- Mutual friends: is considered to be the number of mutual friends that the users have with the inquirer, e.g. 6 common friends, etc.
- Purpose of disclosure: is considered to be the reason why a specific personal information is disclosed, e.g. potential networking benefits are foreseen because users have related interests or career abilities and expectations.

*Design properties:*

- Access and control: is considered to be the right to add, remove or modify any information disclosed at any time. This design solution enables users to control other people's access to their personal data even after actual disclosure.

Importantly, even if acknowledging the importance of access & control as a relevant influential factor for data disclosure in USN, this design property will not be further discussed in this paper. We preferred focusing on contextual data and interrelated attributes influential factors in our mixed method study, because of our target to in-depth analyze the variation of human data sensitivity under different circumstances. The design properties are a complex problem worth investigating, however they were found to influence participants' data disclosure, due to perceived increase of comfort with data disclosure and better usability of USN services, rather than shaping data sensitivity under different circumstances (Sapuppo, 2012a,b).

## 4 Investigation methodology and design

In this section we present the methodology and design of our investigation that aims at analyzing the influential factors, introduced in Section 3. In order to ensure the validity of answers, we helped participants to get more familiar with the USN concept during an introductory meeting, at beginning of the study. We introduced to the participants the existing USN prototype Spiderweb (Sapuppo, 2010) as well as its services (presented also in this video[1]) and other USN applications, already available in the market, i.e. Sonar[2] and Aka-Aki[3]. We also illustrated how potential networking benefits can be gained through USN, as shown in this video about

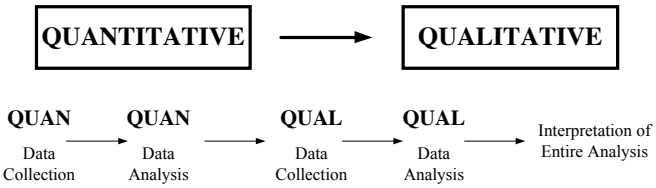



**Figure 1** Sequential explanatory design of the mixed methods study

Aka-Aki[4]. Further, we presented different scenarios from everyday lives, where these services might be applied, such as professional areas, dating and big events, e.g. conferences and exhibitions, as described in (Sapuppo, 2012b; Eagle and Pentland, 2005). Finally, we discussed with the participants the potential networking benefits in the identified application areas as well as possible privacy threats that might arise as a result of the information disclosure in USN, e.g. potential undesired face-to-face interactions (Sapuppo, 2012b).

Afterwards, the participants engaged in a mixed methods study, composed of quantitative and qualitative investigations. We preferred to run a mixed methods study, in comparison to carrying out only one of the two selected investigations, because this approach allows to gain a broad understanding of the research problem as well as ensures greater overall validity of results (Creswell and Clark, 2007). As illustrated in Figure 1, this study followed the sequential explanatory strategy characterized by collection and analysis of quantitative data in the first phase of research, which then provides input for the subsequent qualitative investigation. We selected a sequential explanatory strategy in comparison to others, e.g. concurrent triangulation or transformative designs, because of its straightforward nature that enables to gain in-depth understanding of obtained findings and especially pay particular attention to unexpected results, arising from the quantitative study (Morse, 1991). In the following we explain in details the two phases of the study.

### *4.1 Phase 1: Quantitative investigation*

The first phase of the study comprises a quantitative investigation that analyzes the statistical relationships between the selected influential factors and participants' in situ data disclosure decisions. In the following, firstly, we introduce the techniques utilized for collection of data about users' personal data disclosure decisions and afterwards we describe the statistical methods chosen for analyzing the acquired information.

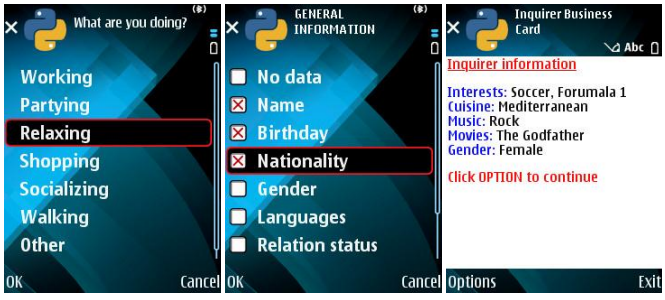
#### *4.1.1 Data collection*

In order to collect data about ad hoc information disclosure decisions, participants were asked to utilize a mobile application that simulates the USN behavior. We preferred to provide a new mobile application, designed specifically for this investigation, rather than utilizing the Spiderweb mobile social network or other existing USN applications, due to two reasons that are explained in the following. First, these applications are not widely spread yet and participants would probably encounter difficulties in finding opportunities to disclose their personal information to other *real* users. Second, the provided USN prototype was explicitly designed to collect data about participants' information disclosure decisions for further analysis, based on the selected influential factors.

Several times a day, the USN prototype was randomly asking participants to specify their current circumstances and their related ad hoc data disclosure decisions. Three screenshots of the USN prototype are shown in Figure 2. Firstly, participants were inquired to specify their current circumstances, as illustrated by an example in Figure 2-A:

- Which is your current mood? The participants could choose between the following range of answers: happy, angry, excited, stressed, sad, worried or other, which implied unrestricted description of their current mood;
- Where are you? The participants could choose between the following range of answers: work environment, social environment, holiday, work trip, on the move or other, which implied unrestricted description of their current location;
- How often are you usually here? The participants could choose between the following range of answers: daily, weekly, monthly, few times per year, first time here;
- What are you doing? The participants could choose between the following range of answers: working, partying, relaxing, shopping, socializing, walking or other, which implied unrestricted description of their current activity.

After the participants provided information about the current circumstances, the USN prototype was asking them to express their ad hoc data disclosure preferences, as shown by an example in Figure 2-B. Participants were aware that potential networking benefits would be directly proportional to the amount of shared




**A** **B** **C**

**Figure 2** Three screenshots of mobile prototype simulating the ubiquitous social networking behavior

information, thus their ad hoc data disclosure decisions were representing a compromise between privacy risks and potential networking benefits. The selection of data types to be disclosed was provided in accordance to data categorization in popular online social networks sites (e.g. gender, age and favorite music). This categorization was already used in previous investigations about disclosure of personal information in USN and the detailed description of the provided data types can be found in (Sapuppo and Seet, 2012).

When participants had expressed their ad hoc data disclosure preferences according to the current circumstances, the USN prototype presented them a business card of a hypothetical inquirer, composed of some interrelated attributes between the participants and the inquirer, as illustrated by an example in Figure 2-C. Three different kinds of attributes were randomly selected by the application:

1. Participants were informed about the number of times they had encountered the hypothetical inquirer during the last 3 months under similar circumstances. This number was randomly ranging from 2 to 900.

2. Participants were informed about a number of mutual friends with the hypothetical inquirer. This number was randomly ranging from 2 to 80.

3. Participants were informed about personal information of the hypothetical inquirer, randomly related either to work or social activities. The presented inquirer's personal information was purposely matching the one of the participants, e.g. shared tastes in music, movies, food or career skills, abilities and expectations. Notably, we were capable of finding these interrelated attributes between the participants and hypothetical inquirers, because we had previously collected participants' personal information about their work and social activities and preferences, during the introductory meeting at the beginning of the study.

Finally, after highlighting the interrelated attributes with the hypothetical inquirer, the USN prototype asked the participants whether they would like to extend their ad hoc data disclosure decisions with any other personal information, which was previously preferred to be kept private. Both initial and extended ad hoc data disclosure decisions were stored in the local memory of the provided mobile phones together with the respective circumstances in order to be applied for further statistical analysis.

### *4.1.2 Data analysis*

The collected data disclosure decisions were analyzed by applying the logistic regression method. We selected this approach because it does not require strict assumptions as other statistical methods like ordinary least squares regression or linear discriminant function analysis (Peng et al., 2002). In contrast to the other two mentioned methods, the logistic regression does not assume linearity between independent and dependent variables nor normality or equal variance within each group of the independent variables (Efron, 1975; Tabachnick et al., 2001; Press and Wilson, 1978; Burns and Burns, 2008). Moreover, we decided to run a binary logistic regression, instead of other kinds of logistic regression methods, such as multiple or ordinary, because our dependent variable was dichotomous, i.e. either disclose the information or not, and the categorical typology of our independent variables, e.g. environment, activity, mood, data type. The research hypothesis posed to the data was that the likelihood of a USN user to disclose specific personal information is dependent on the investigated influential factors. Thus, the variables were defined as follows:

- Dependent variable: whether specific users' personal information, e.g. music taste, is disclosed (1 = yes, 0 = no);

- Independent variables (or predictors): data type (e.g. name, music taste or career skills) and selected influential factors introduced in section 3, e.g. environment, current activities, etc.

To test the research hypothesis, a six-predictor logistic model was applied for data collected from each participant. We preferred to consider each of the participants as a separate test case, rather than evaluating all the participants' data disclosure decisions collectively. This choice was motivated by our focus on ensuring user's personal privacy, i.e. the process where an individual selectively shares his/her own personal information, such as email address, career skills and abilities, to others (Lederer et al., 2004). In order to separately analyze participants' data disclosure decisions, we aimed at collecting enough data sharing preferences to run the logistic regression statistical method per each of the participants. Consequently, we asked participants to utilize the USN prototype for 11 days. At least three times per day, participants provided their initial and extended data disclosure decisions that were composed of 25 data types each, as introduced in Section 4.1.1. The total number of collected data disclosure decisions per each of the participants during the time of the test was the following:

*11 (days) * 3 (times per day) * 2 (initial and extended data disclosure decisions) * 25 (data types) = 1650,*

which considerably exceeded the minimum recommended sample size, e.g. at least 50 cases per predictor, as suggested in (Burns and Burns, 2008; Peng et al., 2002).

The logistic regression analysis was carried out in SPSS version 19 in the Windows 7 environment and addressed:

- Overall evaluation of the model: we present results of the chi-square statistic in order to evaluate the overall significance of the model. When the significance level of the chi-square statistic is lower than *.050*, we can determine that the overall model is statistically significant. In such cases, there is not a high probability to obtain the presented chi-square statistic value under the condition that the data types and influential factors, taken together, do not have impact on users' data disclosure decisions;

- Goodness-of-fit-statistics: we present results of the Hosmer-Lemeshow (H-L) test, which assesses how accurately the model's estimates fit to the actual data. The accuracy is considered to be acceptable when the H-L significance is greater than *.050*. Additionally to the H-L statistic, we also present results of the Nagelkerke index ($R^2$) that investigates the strength of relationship between the dependent and independent variables. This index ranges from 0 to 1, with the value 1 representing the strongest relationship between the variables;

- Assessment of predicted probabilities: we present information about the overall proportion of cases that the model classified correctly. In the ideal model, this proportion would amount to 100%;

- Statistical tests of individual predictors: we show results of the Wald chi square statistic, which provides an index of the significance of each analyzed independent variable. The predictors are considered to be relevant when the corresponding significance value is less than *.050*. In this case, it is possible to determine that the analyzed independent variable has a significant impact on the users' data disclosure decisions.

## *4.2 Phase 2: Qualitative investigation*

The second phase of the study comprises a qualitative investigation that was conducted to better understand the impact of the influential factors on participants' personal data disclosure decisions as well as research on subjective motivations causing the quantitative results. In the following, we introduce the techniques utilized for collection of qualitative data, followed by the strategy for data analysis.

### *4.2.1 Data collection*

Qualitative interviews were preferred alternatively to other investigation methods, such as handing out questionnaires or establishing a focus group interview. This method was chosen because of the following two reasons: (i) lack of participants' extensive experience in utilizing USN services and (ii) potential misinterpretation of the research questions due to their complexity and ambiguity. Moreover, we decided to run semi-structured interviews to better understand the motivation behind the participants responses and ensure that general areas of information are collected from each participant, however still allowing adaptability of the interview process (McNamara, 1999; Creswell, 2009; Kvale, 2004).

Questions were related to the selected influential factors and respective statistical results, obtained during the first phase of the study. Specifically, per each influential factor, we asked the participants to reflect on how important each of the factors was for their personal data disclosure decisions and to elaborate on the reasons. Moreover, after showing the statistical results of the quantitative investigation to the participants, we asked them whether they could confirm these results and comment on any surprising outcomes, obtained in the quantitative investigation.

### *4.2.2 Data analysis*

The strategy utilized for analysis of the information, collected during the qualitative interviews, follows a hierarchical approach, illustrated in Figure 3. At beginning we transcribed the qualitative interviews (step 1) and reviewed them in order to gain a generic understanding of the participants' attitudes towards influential factors that might impact their data disclosure decisions in USN (step 2). In step 3, we organized the transcribed answers in different segments based on the influential factors and, in the next step, we generated the description of participants and themes. The former regards information about the participants (e.g. gender, privacy clusters), while the latter refers to the categories of the major research findings. In step 5, we interrelated the themes and description data categories and, in the last step, we interpreted the qualitative data while taking into

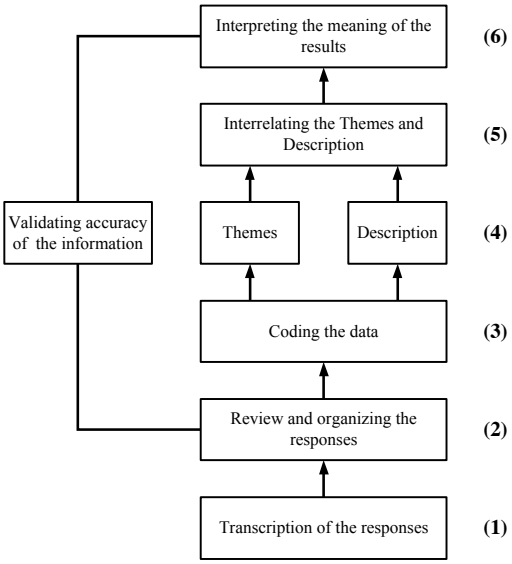



**Figure 3** Qualitative data analysis of the mixed methods study

account, when relevant, the interconnection between themes and description.

In order to ensure accuracy of the findings, two different techniques were applied during the analysis of the qualitative data, as shown in Figure 3. The first one was the triangulation of different data sources, carried out during the second step of our analysis. Particularly, we provided to participants a questionnaire for classifying the influential factors according to the impact that they had on their personal data disclosure decisions. Consequently, we were capable of understanding whether our first review of the transcripts resulted in correct assumptions. Afterwards, during the last step of our analysis, we applied the second technique, i.e. member checking, which determined the interpretations' accuracy of the collected qualitative responses. Specifically, we sent out the final findings of the qualitative investigation back to the participants in order to get feedback on the accuracy of interpretation. When needed, follow-up interviews with the participants were conducted to give them the opportunity to additionally comment on the findings.

## 5 Participants

The participants were randomly selected by sending out email invitations to take part in this study. The selection was limited to online social networks users. We determined this category to be the most relevant because of their advanced experience in social networks, even if the perception towards the services might vary between virtual and physical worlds.

Respondents were asked to provide information about their demographic characteristics. Particularly, we focused on three demographic features, namely gender, age and occupation, which were further applied for clustering purpose. Participants were also asked to indicate their privacy preferences on visibility of their own personal data (e.g. user profile, pictures, posts) in their main OSN site. Based on these answers, we were able to observe patterns among data disclosure attitudes. Consequently, we also 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: these participants also cared about loss of privacy due to 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.

When recruiting the participants, we aimed to achieve stratification between participants' privacy clusters to ensure that specific characteristics of individuals are represented in the sample in accordance to the proportion in the entire population (Floyd and Fowler, 2002). Consequently, in this study, we target at obtaining similar proportions of participants' privacy clusters in reference to our latest empirical investigation where a random sample was selected (Sapuppo, 2012a). In total we recruited 13 participants with the following privacy and demographic characteristics:

- Gender: 8 of the participants were male, while 5 of them were females;
- Age: 6 of the respondents were between 26 and 35 years old, 5 of them were younger than 26 years and 2 participants were older than 35 years;
- Occupation: 7 of the participants were studying and 6 of them were working at the time of the investigation;
- Privacy: 7 of the respondents were pragmatists, 3 of them were fundamentalists and 3 of the participants were unconcerned.

The detailed demographic and privacy characteristics of each participant are illustrated in Table 1.

**Table 1** Information about the participants

| User | Gender | Age | Occupation | Privacy |
|---|---|---|---|---|
| 1 | Male | $> 35$ | Empl. | Prag. |
| 2 | Male | 26-35 | Empl. | Unco. |
| 3 | Male | 26-35 | Empl. | Fund. |
| 4 | Male | 26-35 | Empl. | Prag. |
| 5 | Fema. | 26-35 | Empl. | Fund. |
| 6 | Male | 26-35 | Stud. | Fund. |
| 7 | Fema. | $< 26$ | Stud. | Prag. |
| 8 | Male | $> 35$ | Empl. | Unco. |
| 9 | Fema. | 26-35 | Stud. | Prag. |
| 10 | Fema. | $< 26$ | Stud. | Unco. |
| 11 | Male | $< 26$ | Stud. | Prag. |
| 12 | Male | $< 26$ | Stud. | Prag. |
| 13 | Fema. | $< 26$ | Stud. | Prag. |

## 6 Investigation results

In this section we present the results of our mixed methods study that investigates whether users' personal

data disclosure decisions in USN are impacted by the influential factors, defined in Section 3. In the following, we present the quantitative results followed by the outcomes, obtained during the qualitative investigation.

### *6.1 Quantitative results*

Table 2 presents the results of the binary logistic analysis that was conducted to predict participants' data disclosure decisions using the data type of the disclosed information and the selected influential factors as predictors. Importantly, we did not consider all the 13 participants sharing preferences as a collective sample, instead we run the model per each of the participants as a separate test case, due to focus on personal privacy (refer to Section 4.1.2 for more details).

The overall evaluation of the model was found to be statistically significant for each of the participants, as the significance value of the Chi-square statistics was always observed to be lower than *.000*. Further, in the majority of the cases the model was also found to accurately fit to the actual data, because the H-L significance value was observed to be more than *.050* in 12 out of 13 cases. As well, the Nagerlkerke $R^2$ index was observed to be higher than *.750* for the majority of the participants, which indicated moderately strong (75% or more) relationships between predicted outcomes and predictors. The overall success of predictions ranged from 84.6% to 93.9% with a mean value of 90%.

In order to analyze whether the selected influential factors could be considered as relevant predictors for users' data disclosure in USN, we present results of the Wald chi square statistic applied to 6 predictors: data type, environment, location familiarity, mood, activity and interrelated attributes. The data type was found to be a statistically significant predictor, because the *p-values* for all participants were observed to be less than *.000*, as shown in Table 2. In fact, the information disclosure decisions were strongly influenced by the kind of shared data, e.g. participants might have decided to disclose music tastes, but not home address. Moreover, the other investigated predictors, related to the contextual data and interrelated attributes, are discussed in the following subsections.

#### *6.1.1 Contextual data influential factors*

In this section we describe the Wald statistic results in regard to the contextual data influential factors, i.e. environment, location familiarity, activity and mood. As shown in Table 2, the Wald criterion demonstrated that the environment influential factor was found to be a statistically significant predictor for participants' data disclosure. The *p-values* for all the participants were observed to be less than *.050*. Notably, contradicting results were discovered in regard to the other three contextual data influential factors, i.e. activity, location familiarity and mood. The current activity was found to be statistically significant for 9 out of 13 participants, while the mood and location familiarity were both found to be statistically significant for 6 out of 13 participants. In Table 3, we present the impact of activity, location familiarity and mood influential factors on different participants' clusters.

In regard to the privacy clusters, the most relevant results can be observed among the fundamentalists where none of the participants was found to be influenced by the location familiarity, while all of them together with the unconcerned participants were strongly affected by the activity influential factor. No relevant differences were noted among the gender clusters. In relation to the age groups, 4 out of 5 participants younger than 26 years were found to be impacted by the location familiarity in contrast to other clusters, while participants from 26 to 35 years old presented significant results in regard to the activity influential factor. Finally, among the occupation

**Table 2** Results of the binary logistic regression analysis

| User | Prediction | Model's overall evaluation | | | Goodness-of-fit | | Individual predictors | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | correct | Chi-square | Dif | Sig. | H-L Sig. | $R^2$ | Type | Env | Mood | Fam | Act | Int |
| 1 | 93.4% | 895.310 | 44 | .000 | .977 | .886 | .000 | .000 | .000 | .002 | **.060** | .000 |
| 2 | 88.4% | 1268.333 | 43 | .000 | .000 | .663 | .000 | .000 | **.800** | .004 | .000 | .000 |
| 3 | 92.1% | 589.700 | 47 | .000 | .978 | .563 | .000 | .002 | **.269** | **.145** | .000 | .000 |
| 4 | 92.6% | 1380.669 | 43 | .000 | .115 | .827 | .000 | .000 | .000 | **.538** | .000 | .000 |
| 5 | 90.4% | 1169.623 | 47 | .000 | .075 | .764 | .000 | .004 | **.262** | **.832** | .004 | .000 |
| 6 | 87.5% | 1202.384 | 48 | .000 | .200 | .702 | .000 | .001 | .000 | **.267** | .000 | .000 |
| 7 | 84.6% | 1093.781 | 48 | .000 | .079 | .652 | .000 | .013 | .000 | **.238** | .000 | .000 |
| 8 | 91.8% | 1123.657 | 39 | .000 | .673 | .808 | .000 | .031 | **.297** | **.765** | .010 | .000 |
| 9 | 93.9% | 991.477 | 44 | .000 | .377 | .785 | .000 | .049 | .000 | **.236** | **.122** | .000 |
| 10 | 85.3% | 1339.721 | 47 | .000 | .113 | .744 | .000 | .000 | .000 | .017 | .000 | .000 |
| 11 | 88.1% | 1413.007 | 47 | .000 | .695 | .767 | .000 | .003 | **.108** | .033 | **.074** | .000 |
| 12 | 92.1% | 1533.357 | 41 | .000 | .840 | .837 | .000 | .008 | **.163** | .007 | .000 | .000 |
| 13 | 88.1% | 1185.852 | 44 | .000 | .386 | .772 | .000 | .000 | **.795** | .000 | **.201** | .000 |

**Type**: Data type; **Env**: Environment; **Fam**: Location familiarity; **Act**: Activity; **Int**: Interrelated attributes.

**Table 3** Impact of mood, location familiarity and activity influential factors on different clusters

| | | N | Mood | LF | Act |
|---|---|---|---|---|---|
| Privacy | Fund. | 3 | 1 | 0 | **3** |
| | Prag. | 7 | 4 | 4 | 3 |
| | Unco. | 3 | 1 | 2 | **3** |
| Gender | Male | 8 | 3 | 4 | 6 |
| | Fema. | 5 | 3 | 2 | 3 |
| Age | $< 26$ | 5 | 2 | **4** | 3 |
| | 26-35 | 6 | 3 | 1 | **5** |
| | $> 35$ | 2 | 1 | 1 | 1 |
| Occupation | Stud. | 7 | 4 | 4 | 4 |
| | Empl. | 6 | 2 | 2 | **5** |

**N**: Total number of participants; **LF**: Location Familiarity; **Act**: Activity.

clusters, the employed participants were observed to be highly impacted only by the activity influential factor.

### *6.1.2 Interrelated attributes influential factors*

As shown in Table 2, the Wald criterion demonstrated that the interrelated attributes predictor was found to be statistically significant for personal data disclosure in USN, as the *p-values* for all the participants were observed to be less than *.000*. Consequently, in this section, we get insight into each of the interrelated attributes influential factors by separately analyzing whether informing the USN users about varying numbers of mutual friends and previous encounters as well as diverse profiles similarities might differently influence participants' data disclosure preferences.

In order to achieve this goal, we selected the initial ad hoc data disclosure decisions, when no information about the inquirer was provided, as baseline reference category for the binary logistic regression model. Afterwards, the baseline reference category was compared to the respective extended ad hoc data disclosure decisions, categorized according to various scenarios of the three interrelated attributes influential factors, described in the following.

For familiar strangers, firstly, we took into consideration answers about data disclosure, when few previous meetings, ranging from 2 to 9, were indicated. Afterwards, as second scenario, we analyzed only answers based on encounters, ranging from 50 to 100. As a third scenario we only considered data disclosure decisions, where previous meetings with the inquirer were indicated to range from 700 to 900. Three different scenarios were also applied for the mutual friends influential factor, with the first scenario ranging from 2 to 4 mutual friends, second scenario ranging from 10 to 23 and the last scenario ranging from 50 to 80 common friends with the inquirer. In regard to the purpose of disclosure, firstly, we considered answers that were only related to social profile similarities and afterwards we evaluated only disclosure preferences, based on work profile similarities. The individual results of the Wald statistics for familiar strangers, mutual friends and purpose of disclosure factors are presented in Table 4 and Figure 4. In Table 4, we show the significance values for each scenario, while Figure 4 presents the $Exp(B)$ mean values, which indicate the average change in probability of disclosing personal data, caused by providing information about the inquirer.

As shown in Table 4, for familiar strangers, the majority of participants was not influenced by this predictor when few meetings with the inquirer were known before any actual data disclosure, i.e 2-9 previous encounters. When this number was increased, many more participants were impacted by being familiar strangers with the inquirer. In fact, in the second scenario, 11 out 13 participants were influenced and, in the last scenario, only one participant was not impacted by this factor. These outcomes were also confirmed in the results presented in Figure 4, where the probability of disclosing personal information significantly increased

**Table 4** Wald statistic significance values for different scenarios of the interrelated attributes influential factors

| User | Familiar Strangers | | | Mutual Friends | | | Purpose of Disclosure | |
|---|---|---|---|---|---|---|---|---|
| ID | (2,9) | (50,100) | (700,900) | (2,4) | (10,23) | (50,80) | Social | Work |
| 1 | .029 | .009 | .000 | **.063** | .028 | **.201** | .005 | **.646** |
| 2 | **.439** | .006 | .023 | .000 | **.052** | .001 | .000 | .000 |
| 3 | .044 | .000 | **.315** | **.322** | .001 | .001 | **.748** | .001 |
| 4 | **.675** | .000 | .000 | .001 | .000 | .000 | .000 | .000 |
| 5 | **.315** | .024 | .000 | .022 | .000 | .000 | .016 | .000 |
| 6 | .006 | .000 | .000 | **.144** | .000 | .008 | .008 | .000 |
| 7 | **.228** | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 8 | **.312** | .013 | .000 | .023 | **.637** | **.099** | .009 | .000 |
| 9 | **.188** | .000 | .000 | **.134** | .000 | .000 | .000 | .000 |
| 10 | .016 | **.065** | .000 | **.103** | **.165** | .003 | .028 | **.236** |
| 11 | .000 | .000 | .000 | .001 | .000 | .000 | .000 | .000 |
| 12 | **.090** | **.130** | .000 | .004 | **.954** | .001 | .004 | **.174** |
| 13 | **.097** | .000 | .000 | .017 | .033 | **.079** | **.162** | .000 |

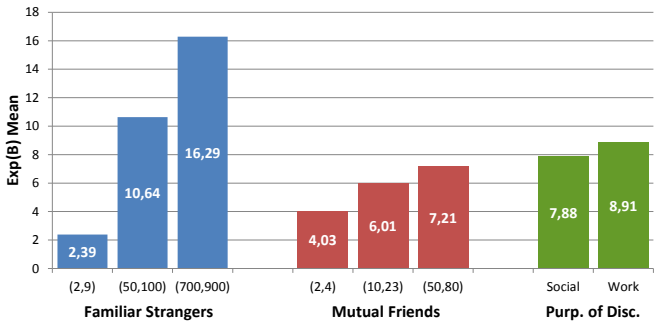



**Figure 4** Change of probability of data disclosure under different scenarios

proportionally to the number of previous meetings with the inquirers. Participants were approximately 16 times more likely to disclose their personal information when being aware about a large number of previous encounters, i.e. between 700 and 900. In regard to the mutual friends influential factor, we did not observe a significant difference between the scenarios, despite the varying numbers of mutual friends. As shown in Table 4, 8 out 13 participants were influenced by having 2-4 common friends with the inquirers, while in the last scenario (i.e 50-80 common friends) only 2 additional participants were found to be impacted by this influential factor. As well, in Figure 4, we can still observe a relevant, but not significant, increase of probability to disclose personal information, proportional to the number of mutual friends. Finally, the purpose of disclosure was found to be a very strong predictor, as we observed that all the participants were influenced by this determinant in at least one of the two scenarios, i.e. either work or social. As shown in Table 4, 10 out 13 participants were found to be influenced by this factor in the work scenario and 11 out 13 of them were observed to be impacted by knowing beforehand to have social similarities with the inquirer. Both scenarios also presented relevant changes of probability to disclose personal information with slightly higher results when relevant professional networking benefits could be foreseen, as illustrated in Figure 4.

### *6.2 Qualitative results*

In this section we present the outcomes of the second phase of the study. Firstly, we describe the results of the qualitative investigation about the contextual data influential factors, i.e. mood, environment, location familiarity and activity, followed by the ones regarding the interrelated attributes influential factors, i.e. familiar strangers, mutual friends and purpose of disclosure.

#### *6.2.1 Contextual data influential factors*

During the qualitative interviews, initially, we inquired the participants about the environment influential factor, as it was found to be statistically significant for all them during the quantitative investigation. As shown in Figure 5, all the participants confirmed the importance of the current environment, because it significantly shaped the relevancy of certain data types to be disclosed. Participants emphasized that they utilized different strategies for differentiation of data disclosure, based on different environments, as one of them noted:

> *"I prefer to split my work and social lives, because I don't want that my lifestyle would be known at work. Thus, even if I did not consider some of the personal data to be sensitive, such as interests or hobbies, I wished to keep it private in my ordinary work environment, as I do not consider them relevant for those situations"*

Similarly to the quantitative outcomes, during the qualitative investigations, we also discovered contradicting results in regard to the other three contextual data influential factors, i.e. current mood, activity and location familiarity. When comparing the qualitative answers with quantitative results, the most significant difference was observed in regard to the mood influential factor, because 8 out of 13 participants claimed that their data disclosure is affected by their current humor, as shown in Figure 5. In fact, two participants, who did not present statistically significant results for the mood factor in the quantitative investigation, stated that their data disclosure is affected by their current humor. The participants commonly agreed that this influential factor would determine their acceptance to exploit USN services, rather than shaping the extent of their data disclosure, as one of them said:

> *"When I was stressed, tired or irritated, I did not disclose any of my personal information, because I did not want to engage in any new social interaction, even if I probably lost relevant networking benefits. Actually, when I am in those moods, I would prefer to switch off these services"*

In regard to the current activity, during the qualitative interviews, 8 out of 13 participants claimed that their data disclosure decisions were influenced by this factor. The majority of the participants noted a relationship between the current environment and current activity for their data disclosure in USN. They

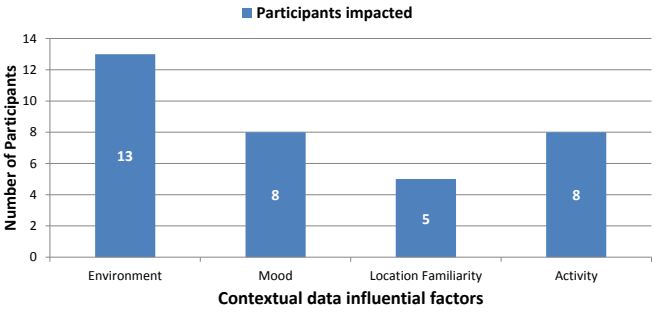



**Figure 5** Impact of the context data influential factors on participants

discussed that when deciding their sharing preferences, the current environment had higher influence than the current activity. In fact, the participants' data disclosure decisions were essentially based on the impact of the current environment, but refined by taking into consideration the current activity. However, one of the respondents emphasized that the impact of the current activity influential factor might increase proportionally to the duration of the activity:

> *"When I had quick coffee breaks with my work colleagues, I did not relevantly change my data disclosure preferences, but during a barbecue event at my work, I additionally disclosed some of my personal information related to social activities"*

Finally, 5 out 13 participants confirmed the importance of the location familiarity influential factor for their data disclosure decisions. Such inclination can be explained by the fact that some people develop an unconscious trust in more familiar places. This led them to also share more personal information, which would have been detained otherwise. For instance, one of the respondents claimed:

> *"When I went to a very familiar cafe in the city center, I disclosed personal information that I usually share in all leisure places, but I additionally shared other data, e.g. my political views, which I usually kept private in other social environments. My political views is sensitive information, but I knew that cafe very well and the kind of people that go there, so I believed that most of them were very open-minded. I did not feel that my political views were so sensitive anymore and I decided to share it, as it was relevant in that case"*

However, all the other participants, i.e. 7 out of 13, did not provide similar comments. They believed that a more familiar location does not necessarily lead to a more trustable environment, as in such places there is still no control over other people surrounding the user.

### *6.2.2 Interrelated attributes influential factors*

The first interrelated attribute that we discussed with the participants was familiar strangers. As shown in Figure 6, 11 out of 13 participants confirmed the relevance of being familiar strangers with the inquirer for their data disclosure decisions. Participants discussed that being familiar stranger with the encountered person might mean that they have been many times at the same locations. Thus, participants expected to have common interests related to that particular location, as for example one of them noted:

> *"I was aware that we had something in common: we lived in the same neighborhood, we often went to the same poker club, etc. In such cases, I was additionally sharing personal information, previously preferred to be kept private, which was specifically relevant for those circumstances"*

Further, some of the participants as well highlighted that knowing the number of previous encounters also provided them a feeling of increased comfort with data disclosure. They were aware that these users were not malicious, i.e. people only interested in retrieving other users' personal information. A few respondents also felt that it was worth disclosing personal data to these particular people, because they were active users, often exploiting these services. As a result, participants were expecting to have higher probability of receiving potential networking benefits in exchange to their information disclosure.

In regard to the mutual friends, all the participants confirmed the statistical results, as 11 out 13 of them claimed to be impacted by this factor. They emphasized that this information unconsciously increased a feeling of curiosity about the encountered users and, consequently, motivated them to share more personal data, in order to easily initiate a face-to-face interaction. The respondents also confirmed the statistic results, which did not indicate a significant proportional relation between the increasing number of mutual friends and the probability of disclosing personal information. Instead, all of them emphasized that their data disclosure decisions would probably be impacted by knowing about the identity of the mutual friends, even if this feature was not tested during the quantitative phase of this study. They provided comments similar to the following:

> *"If the friend that I have in common with the inquirer was a close friend of mine, then I would definitely like to share more of my personal information. On the contrary, if the mutual friend was a person that I do not like or someone who had a strong influence on me (e.g. my boss), then I would probably not disclose some of my personal information"*

As shown in Figure 6, the last interrelated attribute, purpose of disclosure, was found to be relevant for all the participants, because after knowing about similarities

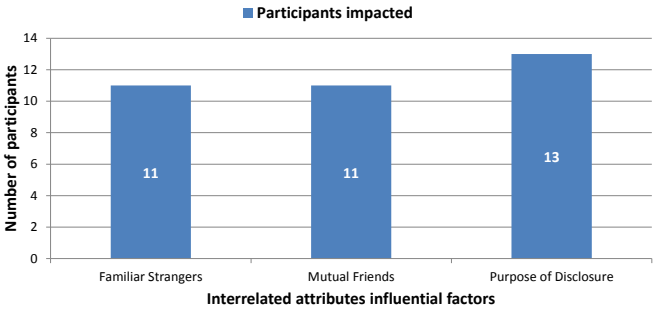



**Figure 6** Impact of the interrelated attributes influential factors on participants

with the inquirers, the participants had a reason for sharing their personal information, which would be kept hidden otherwise. Moreover, participants discussed that they were highly motivated to share the same data types as the ones, disclosed by the inquirers. Participants thought that sharing this data might be relevant for initiating potential face-to-face interactions. For example, one of the respondents claimed:

> *"When I was utilizing the USN prototype, I received a business card from another user who was from Senegal as me and I also learned that we were both studying at the same university. This information strongly motivated me to share my personal data because I really wanted to know her. Naturally, after receiving her business card, I decided to share the matching data types (i.e. nationality and university) as well as other data types that she had disclosed to me, even if her preferences were not matching mine (e.g. favorite books, movies, etc). I did so, because I assumed that she wanted to know this information about other users, as she was sharing it herself"*

Moreover, after receiving the inquirers' personal information, in many cases, participants significantly changed their data disclosure decisions, which were previously based only on the contextual data influential factors (e.g. location, activity). Participants explained that they were motivated to change their sharing preferences, because they could foresee the relevance for disclosing other personal data. For instance, when being at a social environment, they usually did not include data related to work activities. However, after knowing that encountered users were working in their same professional area, respondents felt motivated to share also data related to work activities, because they expected to receive relevant professional networking benefits in exchange. Finally, the purpose of disclosure influential factor as well encouraged participants to disclose personal information that was usually considered to be too sensitive to be shared, as one of the participants noted:

> *"I am usually very cautious about disclosing my political views or religion information, because I don't know how other people might react to it. However, when utilizing the provided USN application, after receiving information that the inquirer had matching political views or religion, I did not have anymore concerns about disclosing this information and I felt that it was relevant to do it"*

## 7 Discussion

In this paper we describe a mixed methods study, which investigated the influential factors for variation of human data sensitivity upon different circumstances in order to contribute to the design of privacy management systems of ubiquitous social networking. The results of this investigation showed that users prefer to share different subsets of their profiles under different situations. The disclosed personal information was selected by compromising between perceptions of data sensitivity for the current circumstances and evaluations of data relevance for gaining potential networking benefits. We found that participants' data sensitivity was decreasing inversely proportionally to the relevance of information disclosure for initiation of networking.

The current environment contextual data influential factor was considered as a crucial determinant for data disclosure, because it primarily guided the participants in evaluation of their data sensitivity and relevance for exploiting ubiquitous social networking services. Similarly to the current environment, the purpose of disclosure interrelated attribute as well significantly guided the participants in taking their data disclosure decisions and in some cases, when potential significant networking benefits could be clearly foreseen, this factor was found to motivate participants to alter their data disclosure decisions, based on the contextual data influential factors (e.g. environment, activity).

Following the results of this mixed methods study, we suggest designers of privacy management systems of ubiquitous social networking to take into consideration the other two contextual data influential factors, i.e. current activities and location familiarity, however as indexes of secondary importance if compared to the current environment. In fact, these two influential factors motivated participants to refine grained selection of disclosed personal information, rather than being significant primary predictors for personal information disclosure. The current activity refined data disclosure decisions for the majority of the participants, while the location familiarity presented contradictory results where only a few of the participants were influenced. The last contextual data influential factor, i.e. mood, was found to have impact on overall acceptance to exploit ubiquitous social networking services, rather than shaping the participants' data sensitivity. Thus, we suggest privacy designers to utilize information about user's current humor as a trigger to interrupt their participation in ubiquitous social networking environments.

Among the interrelated attributes, familiar strangers and mutual friends can be considered as relevant predictors for data disclosure in ubiquitous social networking, as they were found to be statistically significant during the quantitative investigation. However, we suggest privacy designers to consider them as indexes of secondary importance, when compared to the purpose of disclosure. In fact, these factors provided

a feeling of increased comfort with data disclosure as well as motivated curiosity to start an interaction with other users, rather than guiding the participants in evaluating data relevance for better exploiting these services.

During the qualitative interviews, participants highlighted many aspects of the investigated influential factors that still need further attention. Firstly, participants emphasized that the duration of the current activity might have a different influence on their data disclosure decisions. Especially in case of activities with very long duration, it is suggested to analyze whether the current activity might impact the evaluation of data disclosure relevance more than the current environment influential factor. Further, it is also important to statistically investigate whether knowing the identity of the mutual friends would influence users' data disclosure decisions. Two relevant aspects are suggested to be taken into consideration: self-reported closeness and clustering of users' friends into manageable categories (e.g. co-workers). Lastly, additional analysis with a large scale of participants is required to confirm the results of the pilot test, presented in this paper. As well, due to contradictory results, gained when investigating the location familiarity influential factor, further research is needed to in-depth analyze its influence for the variation of data sensitivity in ubiquitous social networking.

## Acknowledgements

This work is supported by Nokia and developed as a part of the Converged Advanced Mobile Media Platforms (CAMMP) project[5], funded by the Danish Advanced Technology Foundation. The author is extremely grateful to the participants of the investigation who took the time to take part in this study. Without their participation and feedback, this work would not have been possible. Finally, the author thanks Lene Sørensen and the anonymous reviewers for their valuable comments on the paper.

## References

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.

Beresford, A. R. and Stajano, F. (2003). Location privacy in pervasive computing. *Pervasive Computing, IEEE*, 2(1):46–55.

Bünnig, C. (2009a). Simulation and analysis of ad hoc privacy control in smart environments. *Intelligent Interactive Assistance and Mobile Multimedia Computing*, 53:307–318.

Bünnig, C. (2009b). Smart privacy management in ubiquitous computing environments. *Human Interface and the Management of Information.Information and Interaction*, 5618:131–139.

Bünnig, 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.

Burns, R. B. and Burns, R. A. (2008). *Business research methods and statistics using SPSS*. SAGE Publications Ltd.

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.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications, Inc.

Creswell, J. W. and Clark, V. L. P. (2007). *Designing and conducting mixed methods research*. Thousand Oaks (California): Sage Publications.

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.

Dey, A. K. and Abowd, G. D. (2000). Towards a better understanding of context and context-awareness. In *CHI 2000 workshop on the what, who, where, when, and how of context-awareness*, volume 4, pages 1–6. Citeseer.

Eagle, N. and Pentland, A. (2005). Social serendipity: Mobilizing social software. *IEEE Pervasive Computing*, 4(2):28–34.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association*, pages 892–898.

Floyd, J. and Fowler, J. (2002). *Survey research methods*. Thousands Oaks, Sage.

Gupta, A., Kalra, A., Boston, D., and Borcea, C. (2009). Mobisoc: a middleware for mobile social computing applications. *Mobile Networks and Applications*, 14(1):35–52.

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. and Hong, J. (2007). End-user privacy in human-computer interaction. *Foundations and Trends in Human-Computer Interaction*, 1(1):1–137.

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*. ACM Press.

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.

Kvale, S. (2004). Interviews: An introduction to qualitative research interviewing. *Evaluation and program planning*, 20(3):287–288.

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. (2003). 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.

McNamara, C. (1999). General guidelines for conducting interviews. *Retrieved December*, 20:2003.

Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40(1):120–123.

Nagle, F. and Singh, L. (2009). Can friends be trusted? exploring privacy in online social networks. In *Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in*, pages 312–315. IEEE.

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.

Peng, C. Y. J., Lee, K. L., and Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1):3–14.

Petkovic, M., Prandi, D., and Zannone, N. (2011). Purpose control: did you process the data for the intended purpose? *Secure Data Management*, pages 145–168.

Pietiläinen, A. K., Oliver, E., LeBrun, J., Varghese, G., and Diot, C. (2009). Mobiclique: middleware for mobile social networking. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 49–54. ACM.

Press, S. J. and Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, pages 699–705.

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

Sapuppo, A. (2012a). Privacy analysis in mobile social networks: the influential factors for disclosure of personal data. *International Journal of Wireless and Mobile Computing*, 5(4):315–326.

Sapuppo, A. (2012b). Ubiquitous social networking: Concept and evaluation. *Sensor Letters*, 10(8):1632–1644.

Sapuppo, A. and Seet, B. C. (2012). An empirical investigation of disclosure of personal information in ubiquitous social computing. *International Journal of Computer Theory and Engineering*, 4(3):373–378.

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.

Suh, J. and Woo, C. (2011). Design and development of a social intelligence based context-aware middleware using blackboard. In *Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on*, pages 908–910. IEEE.

Tabachnick, B. G., Fidell, L. S., and Osterlind, S. J. (2001). *Using multivariate statistics*. Allyn and Bacon Boston.

Tamarit, P., Calafate, C. T., Cano, J. C., and Manzoni, P. (2009). Bluefriend: Using bluetooth technology for mobile social networking. In *Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous, 2009. MobiQuitous' 09. 6th Annual International*, pages 1–2. IEEE.

Terano, T. (2001). *New frontiers in artificial intelligence: joint JSAI 2001 workshop post-proceedings*, volume 2253. Springer Verlag.

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.

Weiser, M. (1995). The computer for the 21st century. *Scientific American*, 272(3):78–89.

Weiser, M. (1996). Open house. *Rank Xerox PARC*.

Weiser, M. and Brown, J. S. (1996). Designing calm technology. In *PowerGrid Journal*. Citeseer.

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

Wiese, J., Kelley, P. G., Cranor, L. F., Dabbish, L., Hong, J. I., and Zimmerman, J. (2011). Are you close with me? are you nearby? investigating social groups, closeness, and willingness to share. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 197–206. ACM.

Youngblood, M., Cook, D. J., and Holder, L. B. (2006). Seamlessly engineering a smart environment. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 1, pages 548–553. IEEE.

## Note

[1]http://www.youtube.com/watch?v=DgeVNv10CIM
[2]http://www.sonar.me
[3]http://www.aka-aki.com/
[4]http://www.youtube.com/watch?v=mvRgtT4LawU
[5]http://www.cammp.dk