ICST Transactions
on Security and Safety

Research Article ICST.ORG

How did you know that about me? Protecting users against unwanted inferences

Sara Motahari*, Julia Mayer, Quentin Jones

New Jersey Institute of Technology, Newark, New Jersey, NJ 07103-3513, USA

Abstract

The widespread adoption of social computing applications is transforming our world. It has changed the way we routinely communicate and navigate our environment and enabled political revolutions. However, despite these applications' ability to support social action, their use puts individual privacy at considerable risk. This is in large part due to the fact that the public sharing of personal information through social computing applications enables potentially unwanted inferences about users' identity, location, or other related personal information. This paper provides a systematic overview of the social inference problem. It highlights the public's and research community's general lack of awareness of the problem and associated risks to user privacy. A *social inference risk prediction framework* is presented and associated empirical studies that attest to its validity. This framework is then used to outline the major research and practical challenges that need to be addressed if we are to deploy effective social inference protection systems. Challenges examined include how to address the computational complexity of social inference risk modeling and designing user interfaces that inform users about social inference opportunities.

Keywords: inference problem, privacy, social computing, ubiquitous computing

Received on 15 February 2011

Copyright © 2011 Motahari *et al.*, licensed to ICST. This is an open access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licences/by/3.0/), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/icst.trans.secsafe.2011.e3

1. Introduction

Social computing applications connect users to each other to support interpersonal communication (e.g. Twitter and Instant Messaging), social navigation (e.g. Facebook), and the sharing of user-generated content (e.g. YouTube and Flickr). While social computing applications gain incredible popularity, their use puts users' privacy at considerable risk. This is because social computing applications are often designed in a manner that impels their users to share personal data as publicly and as widely as possible [1]. Privacy in social computing is complex, multifaceted, and poorly understood, in part because both users' perceptions and the real privacy risks have to be considered when protecting users.

In this paper we discuss a rarely considered but very pervasive and serious privacy threat to users of social computing applications, namely the risk of people making unwanted social inferences. Social inferences occur when unauthorized information is deduced from authorized information about users' identity, location, or other related personal information. Such inferences are often enabled by the public sharing of personal information through social media. Unfortunately, people are often unaware of possible consequences of their personal information sharing, e.g. privacy invasions. This is due to the fact that users have a very limited understanding of the relationship between the information they share intentionally and the information this allows others to infer, e.g. the user's identity, location, or other related personal information (social inferences).

While researchers have mainly addressed the more evident privacy threats related to user access control, the problem of social inferences has not yet been addressed in detail by researchers, particularly in regard to:

(i) measuring social inference risk;

(ii) understanding how social inference risks vary by contexts of use and system types (e.g. mobile vs. online); and

(iii) designing effective social inference risk management systems.

*Corresponding author. Email: Sara.gatmir-motahari@sprint.com

The objective of this paper is to provide a systematic and comprehensive overview of these aspects. We aim at providing a broader understanding of the real-world privacy threats faced by users. This paper will start with a detailed definition and explanations of what social inferences are, when they happen, give examples of real-world incidences and impacts. In this way, we want to draw attention to people's and researchers' lack of understanding and awareness of the problem. We will also examine previous efforts into dealing with inferences as privacy threat and map out their shortcomings as well as their strengths on which we can build. After this, we will present an entropy-based framework to predict the risk followed by findings from studies exploring this method. Finally, we examine the key challenges to the development of effective social inference protection systems that the research community needs to address.

2. The social inference problem

Inferences are generally understood to result from 'the process of arriving at some conclusion that, though it is not logically derivable from the assumed premises possesses some degree of probability relative to the premises' [2]. In the privacy literature *inferences* are understood to result from the process of deducing unrevealed information from authorized information.

In the computing literature inferences have generally been thought of as a confidentiality threat to database security and privacy resulting from database querying or data mining [3, 4, 5, 6, 7]. A well-known example of this general inference problem relates to an organization's database of employees [8], where the relation <Name, Salary> is a secret, but user *u* requests the following two queries: 'List the *rank* and *salary* of all employees' and 'List the *name* and *rank* of all employees'. None of the queries contain the secured <*name; salary*> pair; however, an individual may utilize the known information <Rank, Salary> and <Rank, Name> to infer the private <Name, Salary> information through deductive reasoning. For example, the knowledge that Bob is a manager and all managers earn $x can help one deduce that Bob earns $x.

While it may be possible in some narrow circumstances to make social inferences using database queries or data mining, we will illustrate in this paper that the majority of cases do not result from such deductive reasoning processes. *Social inferences* are the subset of inferences that results from using social applications and can pose serious threats to the privacy of their users. They typically occur through linking authorized information from a social computing application with some background knowledge to infer personal user information, such as identity, location, activities, social relations, and profile information. The underlying logic is:

Background knowledge + Authorized information
→ Social inference opportunity

Background knowledge is defined as any piece of information that is not directly revealed to the users but is available in the outside world or in a system and can be used to infer the attribute in question. It can also be understood as a mental model or world model that provides rules of how to link information together. Such background knowledge is often acquired from real world through experience and observation. Not only may inference require the use of background knowledge, but also the information being inferred (e.g. users' identity at physical appearance granularity) may not be stored in the application database.

In the following sections we will present a more detailed overview of the various ways social inferences can happen and provide examples of real-world incidents as well as user studies which illustrate the impacts and seriousness of social inferences.

2.1. Types of social inferences

First, it is important to understand the variety of ways in which social inferences can occur. In different contexts, different background knowledge or authorized information can yield to a social inference. The way a social inference is made differs depending on:

(i) *Who* makes the inference: person or system;

(ii) *How* the authorized information/background knowledge is acquired (information source);

  (a) over time (historical) or instantaneous and
  (b) from system/web or real world.

(iii) *What* kind of information is used (authorized information/background knowledge) and inferred? (dynamic vs. static, e.g. identity, location, personal information, physical appearance, and social relations).

We refer to social inferences made by people as *human social inferences* and social inferences made by systems as *automated social inferences*. While both types follow the same logic and can be a threat to users' privacy, we focus on the first type, social inferences made by people, in this paper.

The second distinguishing feature of social inferences is the information used to make the inference and how it is acquired. Background knowledge and authorized information might be available only at a certain point in time allowing for an *instantaneous* social inference. On the other hand, when users build up background knowledge over time, e.g. as users interact with applications or observe the world around them over time, which enables a social inference, we refer to it as *historical* social inference.

Both authorized information and background knowledge can come from various different sources, e.g.

(i) different systems (e.g. through the web);

(ii) real-world observations (e.g. who is nearby at the moment); and

(iii) different time periods, accumulated knowledge resulting from aggregating previously revealed information.

For example, in mobile social applications, social inferences can be the result of either accessing location-based information or the result of social communications, or both. Table 1 provides an overview of different types of background knowledge and authorized information used in the various ways social inferences can occur.

While the focus of this paper is on human social inferences, we briefly describe automated social inferences, which are inferences a system can make about a user based on the system's model of the world. This model consists of explicit and authorized data revealed by the user linked to automatically collected background knowledge (e.g. from other sources like the web, from sensors about the user's context) following certain rules and associations. In this way, automated social inferences can be seen as occurring when systems extend their knowledge by emulating human reasoning, which is why such methods are well known in the field of artificial intelligence, machine learning, and the semantic web. Automated social inferences can be used for both personalization and the collection of unauthorized personal information. A widely used method for inferring identity information is to ask visitors to a website for their date of birth, gender, and zip code, supposedly to allow for personalized services. This information is then used to infer a user's identity typically for additional marketing purposes. Researchers started investigating risks related to automated social inferences made by malicious websites which could find out what groups a user belongs to, and use that information to identify the users [9]. By 'capturing' people's social networking groups from their browser with a trick known as history stealing and then cross-referencing these groups, a user's social-network profile—and therefore his real-life identity—could be inferred 42% of the time. This means that an otherwise anonymous web user could be identified correctly by a malicious site simply because the user visited that site.

While automated social inferences also pose a serious threat to users' privacy, this paper focuses on the risks and challenges associated with human social inferences. Human social inferences are made based on people's model of the world (physical, online, social) combined with what a system reveals about the user. Background knowledge here refers to a user's world model consisting of learned, observed, experienced knowledge combined with rules and relations (e.g. if the person is not a boy, it is very likely that she is a girl; or I can find his home address through a Google search). Due to the limitations of people's memory and associated assessment of social inference risks it is important that we distinguish between instantaneous and historical inferences. An example of an instantaneous social inference is when Alice's cell phone shows her that there is a romantic match nearby, Bob. Since Alice sees only one individual with a similar cell phone nearby, Alice infers this must be Bob, thus increasing her chance of identifying him. An example of a historical social inference is when an anonymous nickname is repeatedly shown through a mobile social computing application as being on the first floor of a gym, where the gym assistant normally sits. Despite a wide variety of other nicknames appearing at the same location, given time this association allows other users to infer that the particular online nick name of the gym assistant.

2.2. Examples and implications

The risk of social computing enabled social inferences is growing both with increasing popularity of social media and with the advent of mobile and location-aware social computing. The subtle nature of the social inference risk is illustrated by an incident that occurred during our deployment of a location-aware campus-based Wiki. *CampusWiki* [10] allowed students to create and edit location-linked content. Editors of pages on CampusWiki could be anonymous or identified, and hide or reveal their physical location. During the first semester CampusWiki was deployed, a student added unpleasant comments about

Table 1. Different ways social inferences can occur.

	Instantaneous	Historical
Human	*Background knowledge*—what user knows about his/her current situation through observing and interacting with the world or the web. *Authorized information*—temporary revealed information by the system, e.g. current location of another user.	*Background knowledge*—what user has learned or acquired over time from experience and observations of the world around or from the web. *Authorized information*—information revealed by the systems over time and collected by user.
Automated	*Background knowledge*—temporarily available information from other sources, e.g. web. *Authorized information*—temporary authorized information, real-time shared information, e.g. current location of user.	*Background knowledge*—algorithms to link or cross-reference, e.g. machine learning. *Authorized information*—system-collected user information over time, e.g. buying behavior, cookies from browsing history contextual data collected (GPS and other sensor data).

a professor. In the process the student kept his name hidden, but revealed the time of his edits, and his location at the time. However, the professor was able to infer his identity by realizing that the comments were added in his classroom when he was teaching. Since only two students were using a laptop during the class in question, he was able to identify the student editor. In this case, the inferrer (professor) used knowledge obtained from physical observations (background knowledge) to ascertain who the student in question was. The result was a confrontation, which led to the student dropping the course.

Unfortunately, the risk of social inferences is underestimated not only by individual users but also by system designers. The release of the Google's BUZZ application raised consumer concerns about how social computing applications protect their users' privacy. Initially, Google automatically allowed BUZZ users to see which other users a user had been frequently chatting or emailing. It soon became obvious that having your communication partners revealed in an uncontrolled fashion could result in leaking of quite sensitive and private information about personal relations and/or activities. The Huffington Post discussed problems of the application which could, e.g. let your current employer know you are engaged in conversations with a competitor [11], which in turn could be used to make an unwanted inference about plans for switching jobs. Google's step of making the contact list sharing in BUZZ an opt-in feature was a good initial step, but users still need more help managing the risk of social inferences as they do not have a thorough understanding of the possible consequence of their actions. These examples suggest that the risk of unwanted social inferences is serious and could jeopardize one's education, career, or personal life.

To explore people's understanding of the privacy threats associated with social computing, we carried out a survey of students at our North Eastern, urban university campus (107 subjects from 17 different majors) [12]. The following seven categories of threats to user privacy in social computing system were introduced to the subjects together with example scenarios:

(i) *Inappropriate Use by Administrators*: The system admin sells personal data without permission [13].

(ii) *Legal Obligations*: The system admin is forced by an organization such as the police to reveal personal data [13].

(iii) *Inadequate Security*: The server is not protected against intrusions or wireless transmission through the air is not secured [14].

(iv) *Designed Invasion* (due to poorly designed features): A cell phone application that reveals location to friends, but does this without informing the user or providing control of this feature [15, 16, 17].

(v) *Instantaneous Social Inferences*: Alice's cell phone shows her that there is a romantic match nearby, Bob. Since Alice sees only one individual with a similar cell phone nearby, Alice infers this must be Bob, thus increasing her chance of identifying him.

(vi) *Historical Social Inferences*: Bob is so often in Alice's office. Their relationship must be romance.

(vii) *Social Leveraging of Privileged Data*: David can't access my location, but Jane can. David asks Jane my location.

Then subjects were asked to rate their awareness of each threat as well as their privacy concerns depending on the type of information, such as identity, location, status, profile information, and social ties. Respondents' answers highlighted the expected issue that people have comparatively little awareness or concerns about the risk of social inferences. Subjects expressed less awareness over inferences (categories v, vi, and vii, where categories v and vi represent the social inference problem) and more awareness of the first four categories (Friedman, $\chi^2 = 299$, $df = 6, n = 102, p < 0.001$). Combining the inference categories into one new variable and the first three categories into another one also shows a statistically significant difference between them (Wilcoxon's Signed Rank $u = -7.91$, $n = 102, p < 0.001$). Furthermore, subjects were generally more concerned over the threats they were more aware of. They are most concerned over being hacked, which is not surprising considering the ability of hacking stories to make the news headlines. *Inappropriate Use by Administrators* (i) and *Designed Invasion* (iv) come next and the other categories were least worrisome for them (Friedman, $\chi^2 = 64.4$, $df = 2, n = 99, p < 0.001$).

To determine if people's general lack of concern about social inferences is due to their ability to manage such risk, or due to a lack of understanding of the extent of such risks, we conducted an experiment [12, 18]. Two hundred ninety-two individuals (146 pairs) engaged in anonymous chat with each other. Subjects were asked questions about their desired level of anonymity, if they think they had maintained their desired level of anonymity, and if they could identify who they chatted with. The level of desired anonymity varied greatly, however 72% of the subjects who had anonymity concerns did not want to be exactly identified by their name or face and 6.3% of them did not want to be narrowed down to two people or less. It was surprising to us just how poorly subjects were able to assess what their chat partner knew about their identity. Subject's estimated-degree-of-anonymity was smaller than maintained-degree-of-anonymity in 20% of the cases, which means at least 20% of the subjects revealed identifiers that put them at the risk of *unwanted* identity inference. This illustrates that while people are

often able to make social inferences, they are unable to estimate social inference risks and are therefore unable to maintain their desired degree of anonymity.

The above examples, survey, and experiment illustrate the mismatch between real privacy risks and user perceptions and suggest that at present users of social computing applications are unable to routinely maintain their desired and expected level of personal privacy. Even when users are in complete control of the information they reveal they are not able to maintain their desired degree of anonymity because individuals are unable to correctly judge inference risks. The research community is yet to come to grips with the unique privacy challenges associated with social inferences. This calls for a more thorough examination of the social inference problem as a privacy threat in order to build effective social inference protection systems.

To address this fundamental gap in the literature, below we present a social inference risk prediction framework and associated confirmatory studies, and then outline some of the major research challenges facing those that wish to build effective social inference protection systems.

3. Determining the social inference risk

In order to protect individuals from unwanted social inferences, we need to be able to measure the extent of the social inference risk in various contexts. This determination can be achieved through the steps outlined in our social inference risk prediction framework. We start this section with an overview of previous attempts to predict social inference risk. Then we will introduce our social inference prediction framework and present findings from validating studies.

3.1. Previous work on the inference problem

To date, most of the research into inferences has addressed the general inference problem in databases, which focuses on the problem of detecting and removing unwanted inference channels. An inference channel in a database is a means by which one can deduce unauthorized data from authorized data. In order to detect an inference channel in a database, the inference risk has to be predicted. Two types of techniques have been previously proposed. One makes use of semantic data modeling methods to predict chances of inferences and locate inference channels in the database design, in order to redesign the database for the removal of these channels [19]. The other one evaluates database queries to understand whether they lead to unauthorized inferences [20]. These techniques have been studied for statistical databases [20], multilevel secure databases [21, 22], and general-purpose databases [8, 23].

In addition, deductive inferences can also result from analysis of data sets generated for data mining. Although such data sets are usually cleaned up from sensitive information, common data mining techniques could lead to leakage of some previously eliminated sensitive information. Therefore, a few researchers have attempted to prevent the inferences in data sets generated for data mining, such as medical records [7].

However, as noted in the Section 1, social inferences are often made about personal user information that is not stored in the application database and leveraging background knowledge that is also not stored in the database. As a result, systems cannot protect users from social inferences by applying traditional database inference protection techniques. An alternative way is to use an entropy-based approach. An early attempt to use entropy measurements in anonymity protection came from Serjantov and Danezis [24]. They suggested that the anonymity level of networking nodes (transmission and routing systems) could be measured using an information theory approach, i.e. entropy measurements. They proposed measuring the degree of anonymity of geographically fixed nodes (such as desktops) assuming that network attackers have partial information about the topology of the network (which can be considered the attacker's background knowledge).

Denning and Morgenstern, pioneers in calculating the partial inference risk, employed classical information theory to measure the inference chance [3, 25]. Given two data items x and y, let $H(y)$ denote the entropy of y and $H_x(y)$ denote the entropy of y given x, where entropy is as defined in information theory. Then, the reduction in uncertainty of y given x is defined as follows:

$$Infer(x \rightarrow y) = \frac{H(y) - H_x(y)}{H(y)}.$$

The value of $Infer\ (x \rightarrow y)$ is between 0 and 1, representing how likely it is to derive y given x. If the value is 1, then y can be definitely inferred given x. Denning and Morgenstern did not know how to use this formulation in real situations and they mention the serious drawbacks of using this technique [3]. Firstly, it is difficult, if not impossible, to determine the value of $H_x(y)$; secondly, the computational complexity that is required to draw the inference is ignored [3]. Nevertheless, this formulation has the advantage of presenting the probabilistic nature of inference (i.e. inference is a relative not an absolute concept).

Another approach addressing the problem of identity inferences based on usernames was proposed by Lemos [26]. They produce an analytical model that estimates the uniqueness of a username and then assign a probability that a single username from two different online services refers to the same user. Probability estimates

are calculated using language models and Markov chain techniques. Their results show that entropy measures of usernames can be used to link accounts and to identify users.

The above overview shows that previous entropy-based research into predicting the impact of background knowledge on inference risks has made important steps toward privacy protection and provides significant insights, such as the use of entropy-based modeling approaches. However, this research has failed to take into account a number of key considerations. The above methods are based on the assumption that all data used by the inferrer are inside a self-contained system. Therefore, the presented solutions can only protect from a narrow range of social inferences because, as seen above, social inference opportunities are enabled through use of information from outside the system in question (background information). In addition, the proposed solutions assume that the information that is at risk of being inferred is stored inside the system in question, which is, e.g. usually not the case for users' identity at physical appearance granularity. Therefore, if we are to help users protect themselves from unwanted social inferences, mechanisms will need to be developed that take into account inferrer's potential background knowledge. They also do not deal with the dynamic nature of personal and contextual information in social computing applications and they make unverified assumptions about the inferrer's background knowledge. We address these drawbacks in our framework by

(i) considering the probabilistic nature of social inferences by using an attribute's information entropy to measure the level of its uncertainty;

(ii) taking the inferrer's background knowledge and historical data into account by calculating the conditional entropies conditioned on both revealed data and background knowledge; and

(iii) dynamically updating the level of entropy.

3.2. Social inference risk prediction framework

We developed a framework [27, 28] to model and reliably predict social inference risk. As seen before, social inferences happen as a result of low information entropy. The framework is based on this logic: as individuals or applications collect more information about a user, such as his/her current situation, our uncertainty about other attributes, such as his/her identity, may be reduced, thus increasing the opportunity of a social inference. This uncertainty can be measured by *information entropy*. *Information* [29], as used in information theory for telecommunications, is a measure of the decrease of uncertainty in a signal value at the receiver site. Here we use the fact that the more uncertain or random an event (outcome) is, the higher the *entropy* it possesses. If an event is very likely or very unlikely to happen, it will not be highly random and will have low entropy. Therefore, entropy is influenced by the probability of possible outcomes. It also depends on the number of possible events, because more possible outcomes make the result more uncertain. In our context the probability of an event is the probability that an attribute (such as a user's name) takes a specific value. As the inferrer collects more information, the number of entities that match her/his collected information decreases, resulting in fewer possible values for the attribute and lower information entropy.

The scenario below illustrates the logic of our framework. It describes the actual behavior of a pair of subjects in our experimental examination of the social inference risk prediction framework [18]:

> Bob engages in an online chat with Alice. At the start of communication, Bob does not know anything about his chat partner. He is not told the name of the chat partner or anything else about her, so all users are equally likely to be his partner. After they start chatting, Alice's language and chat style help Bob guess her gender and that she is Hispanic (and Alice confirms his guess during the course of conversation). After a while, Alice reveals that she plays for the university's women's soccer team. Bob, who has prior knowledge of this soccer team, knows that it has only one Hispanic member. This allows Bob to then infer Alice's identity.

Here, as Bob combines his background knowledge of the female Hispanic soccer players on campus with what Alice reveals, his uncertainty about his chat partner's identity decreases, thus increasing the opportunity of a social inference. This uncertainty can be measured by *information entropy*. In the case of the above scenario Bob's background knowledge of the ethnic makeup of soccer players on his campus is all that is necessary for an identifying social inference. In order to calculate the information entropy of an attribute, the background knowledge of a user has to be modeled. This is needed in order to identify:

(i) What attributes, if revealed, can help the inferrer to reduce the identity entropy of a user and how they change conditional probabilities.

(ii) What attributes, even if not revealed, can help the inferrer reduce the identity entropy of a user and how they change conditional probabilities (such as guessing Alice's gender from her chat style and using that to infer her identity).

However, as Jajodia and Meadows [21] say, 'we have no way of controlling what data is learned outside of the database, and our abilities to predict it will be limited'. Thus, even the best model can give us only an approximate idea

of how safe a database is from illegal inferences. Nevertheless, the results of our studies suggest that considering the context and community of users enables systems to effectively model background knowledge.

Background knowledge can be estimated using the following different methods listed here with increasing levels of accuracy [11, 28].

(i) *Method 1*. Simply assume that the inferrer can link the information in the application database to the outside world, thus being able to estimate the number of matching users and their probabilities based on the existing database. *Weakness*: Some of the attributes in the database are not usually known to the inferrer while some parts of the inferrer's background knowledge may not exist in the database.

(ii) *Method 2*. Hypothesize about the inferrer's likely background knowledge taking the context of the application into consideration.

(iii) *Method 3*. Utilize the data from user studies designed to capture the users' background knowledge. The advantage of this method is a reliable modeling of background knowledge.

(iv) *Method 4*. Extension of the latter two methods with application usage data that allow for continuous monitoring of an inferrer's knowledge.

Two user studies provided comparative value and practicality of the second and third method. The results suggested that Method 2 was almost as accurate as Method 3 in the realm of computer-mediated communication and proximity-based applications [11, 28]. This means that context and community of users allows to effectively model background knowledge for a specific context.

After estimating all the significant information (including background knowledge) available to the inferrer, Q, the conditional instantaneous information entropy of attribute Φ, is defined as

$$H_c = H(\Phi|Q) = \sum_{i=1}^{V} P_c(i).\log_2 P_c(i),$$

where V is the number of possible values for attribute Φ, $P_c(i)$ is the probability that the i^{th} possible value is thought to be the true value by the inferrer. $P_c(i)$ is the posterior probability of each value given Q. Q includes the inferrer's background knowledge as well as the information currently being revealed by the system. In the case of Bob and Alice mentioned above, at the beginning of the chat, V equals the number of students and $P_c(i)$ is uniformly distributed over V. As the revelations continue, V decreases and also $P_c(i)$ deviates from a uniform distribution until entropy is lower than a risky threshold. Historical information entropy is defined in the same way, but by including previously revealed information in Q.

More details on calculating the conditional probabilities and information entropy can be found in [11, 28] where the authors also set thresholds for information entropy based on users' preferences. If the information entropy is lower than its threshold, there is a high inference risk. Therefore, instantaneous and historical inference functions are based on instantaneous and historical information entropies [11, 28].

3.3. Testing the risk prediction framework

To the best of our knowledge, the above risk prediction framework is the only method proposed to predict the social inference risk. The following empirical studies showed that it is able to strongly predict the risk of social inferences [11, 18, 28].

The information entropy modeling approach was tested in two different areas of social computing:

(i) for anonymous computer-mediated communication (Anon-CMC), through a laboratory chat experiment between unknown chat partners [18] described in part earlier in this paper; and

(ii) for proximity-based social applications, through a mobile phone field study that lasted for 4 weeks and explored patterns of co-location and anonymity of the subjects [11, 28].

Moreover, large-scale simulations were done for both areas to verify the entropy-based social inference risk modeling framework and to investigate the problem and appropriate actions on a larger scale for various situations.

The laboratory experiment involved 532 collected user profiles, out of which 292 subjects completed a chat experiment and the post-experiment survey [18]. Subjects participated in a study consisting of three phases:

(i) Phase I—*Online Personal Profile Entry*: Subjects were filling out an extensive online profile (67 individual profile items clustered into five broad categories: basic information, personal information, education information, contact information, and interests).

(ii) Phase II—*Introductory Chat Experiment*: Subjects were chatting with a randomly assigned unknown chat partner using a custom-developed software application designed to aid in communication and exchange of personal profile information. The user interface (UI) of the introductory chat experiment is shown in Figure 1. During the chat, subjects were able to see their own profile on the left side of the screen, a chat box for typing in the center, and their chat partner's profile (all information hidden by default) on the right side. Throughout the conversation subjects could decide to reveal parts of the profile or request their

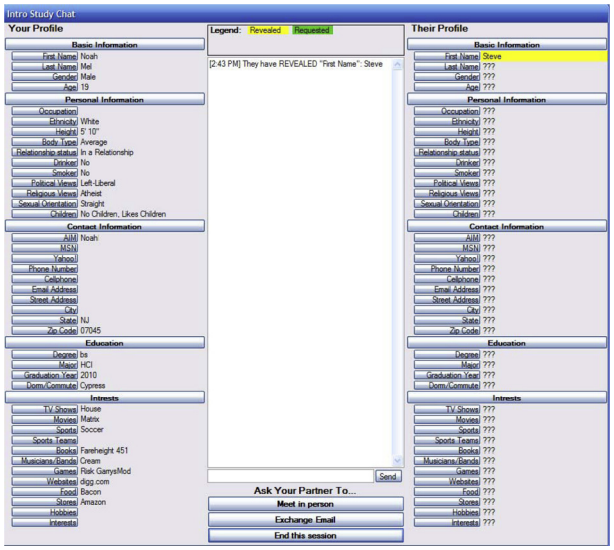

Figure 1. User interface of introductory chat study between anonymous chat partners.

chat partner to reveal a certain field. Revealed fields were highlighted yellow while requested fields were highlighted green.

(iii) Phase III—*Post-Chat Survey*: Subjects were asked if they could guess their chat partner's identity, or other attributes (physical characteristics), and how they made the guesses, as well as how anonymous they wanted to be and what they revealed about themselves.

Phases I and III served to inform us as to how people perceived and understood social inference (detailed results can be found in [18]) while Phase II allowed us to test the entropy-based risk prediction and validate our framework. Many variables were measured from this laboratory experiment such as the duration of the chat, the number of revealed profile items, subjects' intended level of anonymity, and subjects' demographics. Results of a binary logistic regression showed that among all measured variables, information entropy was the only statistically significant predictor of the inference risk (Wald's $\chi^2 = 6.018$, $\exp(\beta) = 0.705$, $p = 0.014$). Background knowledge of the users was investigated and taken into account in calculating the information entropy.

The proximity-based social application field study was carried out in two stages. The first stage lasted for 3 weeks and 180 subjects participated in it. The second stage lasted for 4 weeks. One hundred sixty-five subjects started the field study, out of which 129 subjects completed 4 weeks of the study. The subjects used a proximity-based social application that showed user the nicknames of users in their vicinity (as shown in Figure 2). Every time subjects changed their location and stayed in a new location for 5 min or when they had not answered a questionnaire

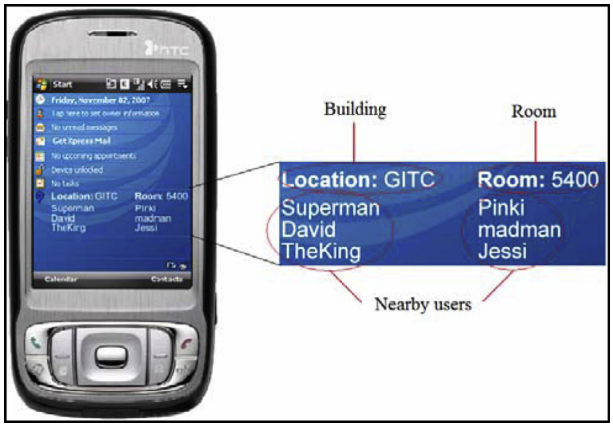

Figure 2. Proximity-based social application study.

for at least 2 hours, they answered pop-up questionnaires asking what they could guess about the identity of nickname owners and if they could map them to people in their vicinity.

Various place-, time-, proximity-, and subject-related parameters were measured. A binary stepwise logistic regression analysis was performed on the identity-inference-incident (dependent variable) and all the independent variables listed above. The only variables left in the analysis were the instantaneous inference function (Wald's $\chi^2 = 5.818$, $\exp(\beta) = 0.970$, $p = 0.012$) and historical inferences function (Wald's $\chi^2 = 53.001$, $\exp(\beta) = 1.084$, $p < 0.001$). Users' background knowledge in this context was investigated and modeled for calculating information entropy [11, 12]. In both studies, entropy-based inference modeling was again the strongest predictor of social inference opportunities.

Experimental results show that social inferences are not rare and that they are more common in CMC than in proximity-based application. We used the experimental data from the above studies to investigate the extent of the risk of identity inference for both applications on a larger scale. Parameters for the simulation models, such as the diversity of profile items and their statistical distribution as well as the probability of revealing profile items and statistical distribution of nearby users, were derived from the experimental data from previously explained user studies to approximate real-world deployments. Additional information such as the number of courses, statistical distribution of the number of students in a class, and enrolment statistics were obtained from university admission statistics. Entropy thresholds were calculated based on the desired degree of anonymity (a desired degree of anonymity of u means that the users wished to be indistinguishable from $u - 1$ other users).

For the Anon-CMC simulation online interactions of the users were simulated and information entropy was calculated for each simulated chat based on their revealed profile information. Results shown in Figure 3 indicate

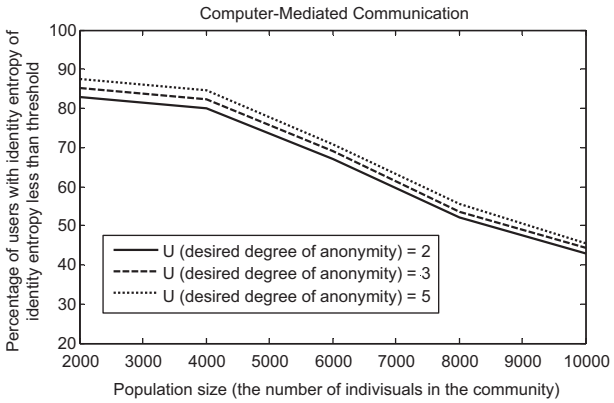

Figure 3. Risk of identity inference for computer-mediated communication.

that users reveal information that 50% of the time could lead to the invasion of their desired degree of anonymity (failure to maintain their desired degree of anonymity), which shows that identity inferences can be quite prevalent in CMC.

Figure 4 shows the probability that a user is at the risk of instantaneous identity inference in a proximity-based application. The *y*-axis shows the percentage of users whose identity entropy was lower than its threshold. Entropy threshold was calculated based on their desired degree of anonymity, Φ. The *x*-axis represents the desired degree of anonymity. Each curve depicts the risk for a different mean of nearby population density. As expected, more crowded environments have a lower chance of being at the identity-inference risk. Simulation of the risk of historical inferences and experimental results show that for a given population density, historical inferences happen less frequently than instantaneous inferences. Therefore, identity-inference risks happen less frequently for proximity-based applications than for CMC.

To conclude, these experiments and simulations verified the entropy-based social inference risk prediction framework and showed that we are able to predict social

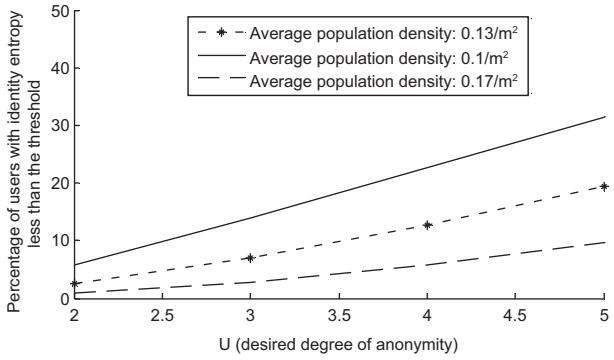

Figure 4. Risk of identity inference for proximity-based applications.

inference risk. In order to protect individuals from unwanted social inferences, systems will need to be deployed that systematically reduce this risk.

In the next section we propose several features of social inference protection systems and point out open key challenges for future research.

4. Social inference protection systems

Historically, inference protection has been thought of as an access control issue, where the main challenge is to make sure that potential privacy invaders cannot get the results of dangerous queries that enable inferences. As we have shown social inferences cannot be protected in this way because the relevant data are not stored in the user-application database. As a result, new enhanced techniques and systems will need to be developed if we are to protect users from unwanted social inferences. As shown above, we are able to reliably predict the social inference risk using an entropy-based approach, but this alone provides little protection for the user. Instead, social inference protection systems need to be developed as shown in Figure 5 that:

(i) determine users' privacy preferences;

(ii) monitor users and their environments;

(iii) use stored and contextual data collected to calculate users' current social inference risks; and

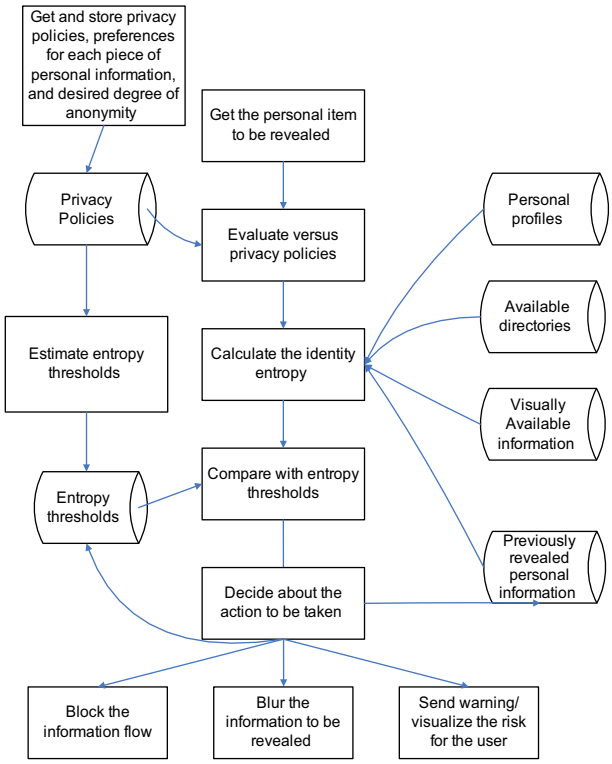

Figure 5. Components and processes in a social inference protection system instantiation.

(iv) utilize this knowledge to reduce the dangers by either

  (a) blocking information exchange;
  (b) blurring information exchanged; or
  (c) providing visualizations/warnings to users that support their taking appropriate actions.

Unfortunately, there is limited knowledge about how to build such a system. The key areas where further research is needed can be divided into risk prediction and risk management challenges.

4.1. Challenges in risk prediction

Information entropy as used in our presented framework showed to be a reliable predictor for social inference risk. The next steps to be taken are to explore other sources of background knowledge and improve modeling of background knowledge as well as historical information.

Modeling background knowledge. Background knowledge can be acquired from a variety of sources. For example, users can socially leverage privileged data and obtain knowledge from their friends and social ties that could inherently become part of their background knowledge. Such sources were insignificant in the above studies. However, in the case of finding such knowledge or other distributed types of background knowledge, such information sources have to be appropriately rated based on their reliability and access probability. Such effects should be parameterized in order to tailor the conditional probability equations used to estimate the information entropy. Studies of background knowledge can be merged with initial studies of the application, such as usability studies, so that the estimation can be obtained with a low cost. Finding efficient and inexpensive ways of background knowledge modeling for various applications of ubiquitous and social computing is a challenge yet to be addressed.

Modeling the historical information. Moreover, modeling of historical information has to be improved. Historical information enables social inferences, thus achieving a precise calculation of information entropy needs a model of the information previously presented to a potential inferrer. In the few solutions suggested for the information aggregation problems in databases, historical information of an adversary is usually considered to include all of the previously revealed information. However, a study that focused on the modeling of historical data for social inference risk prediction shows that the exact proximity history of the users determines the inference risk and that the optimum history length was 2 weeks [30]. In fact, the users do not have a perfect memory of their past visual observations.

It should be noted that previous work had a number of limitations. It derived the best values of time-weighting parameters and the history length from a simple proximity-based social application that only reveals nicknames. Furthermore, this study was carried out for 4 weeks, whereas a real application may be used for years. Gaining data over a longer period of time brings up new issues and calls for improved models. Therefore, researchers have to look more into this problem of historical information modeling.

Computational complexity. Another practical challenge in implementing social inference prediction algorithms is the computation complexity. The framework explained in this section can estimate the level of anonymity in any situation where personal attributes are shared, especially in social computing. However, the computational complexity of calculating parameters such as V (the number of relevant entities) and the probability of each one might raise concerns over the practicality of building a social inference protection system for synchronous communications. In synchronous online communications, a social inference protection system should be able to estimate the risk in acceptable time. An algorithm with acceptable delay and computational complexity was proposed in [27] that used basic properties of information entropy for this purpose. The aim of this work was not to find an optimal solution to this problem, but to find an algorithm that reduces the complexity substantially, as compared to the brute-force algorithm. It also considered the worst case where one server was used for anonymity estimation. Nevertheless, in an application with a very large number of users, many linkable attributes, and highly dynamic profiles, distributed servers or even more efficient algorithms may be needed. In distributed systems, each server may not be fully aware of a user's history or other users' profiles. In that case, decision making under incomplete knowledge may be inevitable and risk prediction algorithms need to make an optimal decision based on the locally known data. For example, the user's previous revelations or the values of certain attributes of other users will also have a probabilistic nature. Entropy thresholds may be fuzzy rather than exact preset values. Therefore, future research into this problem is needed.

4.2. Challenges in risk management

After a social inference protection system detects a situation with high inference risk (with entropy below a certain threshold), the system has to take proper action to reduce this risk. This section presents two different levels of system action which could be taken:

(i) automated risk management; and

(ii) semi-automated risk management.

Moreover, UI approached for risk management features in social inference protection systems is proposed and discussed.

Automated inference protection systems directly intervene in risky information exchanges. Users, designers, or system administrators could set the desired anonymity

level on a privacy control setting page and the system would then act accordingly to prevent users from unwanted social inferences. Possible methods to be used in such systems are, e.g. blocking the exchange of specific items of information, or blurring the exchanged information automatically when the predicted social inference risk is too high. Lowering the granularity of revealed information, e.g. revealing the location at floor precision instead of room precision, or showing an anonymous name instead of a nickname can lower the social inference risk. However, the practicality and comparative utility of blurring or blocking the information exchanged (e.g. providing a user's home state as opposed to home city or age range as opposed to birth date) need to be further examined.

However, the main purpose of social computing is to allow people communicating with each other and therefore automated protection by simply blocking the information exchange might defeat this purpose. Our simulation results imply that automatic management of the risk, such as blocking the information flow, can severely degrade the utility of CMC applications, because such interferences with the flow of communication can happen quite frequently. Instead, more sophisticated and dynamic solutions for risk management are needed.

The semi-automated approach detects the risk automatically, but the action is left to the user. Instead of specifying general privacy preferences, users are informed about their social inference risk and given control about the reduction method at all time. To enable users to take appropriate risk reduction actions, they need to be made aware of the risk.

User interfaces need to present users with understandable social inference risk visualizations and controls to allow them adjust their information sharing accordingly. Many social computing applications already provide UIs intended to inform users as to the personal information they are sharing. They aim at supporting user impression and privacy management, providing users with the ability to manage their personal privacy, as a communication aid, and to inform users about the reasoning behind a software personalization. Generally, users are able to view how their profile is seen by other users. However, current privacy management interfaces fail to provide users with an understanding of the privacy risks they face or suitable options for truly controlling the information being shared. Simple rule-based privacy settings do not cope well with the dynamic and context-dependent nature of people's privacy preferences and information needs. Users do not wish to constantly set rules to manage what they are sharing, and at the same time they do not want to have sensitive information put in danger as a result of sharing fairly innocuous information. Moreover, a user may need to know not only what is directly being shared with other users but also what aspects of their profile can be inferred.

Semi-automated risk management interfaces need to inform users about their current social inference risk (risk visualizations) and allow users to control their privacy preferences accordingly (control interfaces). Risk visualizations aim at supporting users' awareness of social inference risks by providing a status of the current social inference risk the user is exposed to, while control interfaces provide users the means to adjust both their general desired anonymity level as well as their current personal information sharing based on potential social inference risks rather than static information sharing rules.

The systems can, for example, show users how uniquely they have specified themselves so far, or send a warning message when revealing a piece of information would enable their partner to invade their desired degree of anonymity.

Alternative UI approaches to inform users about their current social inference risk can include

(i) *pop-up warnings* during application use, which highlight the risks associated with taking the decision to reveal particular items of info- rmation;

(ii) *risk status lookup* UI where users can check now and then their overall current social inference risk; and

(iii) *awareness displays* that permanently show users their social inference risk in an unobtrusive way.

These alternative approaches need to be further explored in terms of their utility and usability. User studies need to investigate their usefulness and effectiveness in various contexts to understand how to best design systems that truly help the user manage their social inference risks.

To decide whether semi-automatic or fully automatic methods should be used in social computing application, the seriousness and frequency of identity inferences in the domains of computer-mediated communication and proximity-based social applications were explored in a previous study [28]. The results suggested that automatic control of information exchanges in computer-mediated interpersonal communication can overly interrupt the information exchange because social computing applications are designed to exchange information. Automated control of information exchange can degrade system usability or be frustrating for the user. Therefore, the semi-automated methods seem to be more appropriate solutions to reduce social inference risk. However, the automated method appeared to work fine in proximity-based social applications.

5. Discussion and conclusion

Social computing applications are becoming an essential part of our lives and as a result are fundamentally changing the ways in which we need to think about privacy and

privacy management. We illustrated that the serious threat social inferences pose to user privacy is real and that they happen frequently. Users are often unaware of the possible negative consequences of their personal information sharing. Furthermore, many social computing applications currently misrepresent the privacy protection they provide to users by implying that a user's anonymity can be effectively protected by simply allowing them to select not to publicly share a subset of personal information.

The comprehensive overview of the social inference problem in this paper illustrated that the social inferences problem is a serious threat to user privacy in social computing applications. Unfortunately, this has not resulted in researchers looking holistically at the social inference problem. Nevertheless, the social inference problem becomes more and more prominent and seems to be expanding along with the ever-increasing adoption of mobile social computing systems that merge online social interactions with context-aware computing. The problem is further complicated by social computing application users sharing different personal information with different potential inferrers, and by sensitivity of user information as well as user's privacy preferences often being highly dynamic depending on changing user location and context. Users may be willing to compromise their privacy settings to have more meaningful and productive communication. Social inference protection systems have to take into account the need for adjustments of privacy preferences based on context, social inference risks, and user needs.

Our empirically validated social inference risk prediction framework suggests mechanisms by which many of the current inadequacies could be addressed. We believe that this is a significant first step toward providing individuals with tools for managing their social inference risks so that privacy needs are better met and more importantly people's awareness of the possible consequences of their information sharing choices made apparent. However, our knowledge regarding the social inference problem in other circumstances than the ones examined in this paper is lacking. We see future research progressing along several trajectories. First, we expect to see more work on background knowledge estimation and modeling. Second, there is potential for a thorough empirical examination of how to model historical information to better predict social inference risks. Third, entropy estimation algorithms need to be optimized for computational complexity. Finally, research into alternative visualization approaches to inform users about social inference risks is needed that can provide design guidelines.

References

[1] Hoadley, M.C., Xu, H., Lee, J. and Rosson, M.B. (2009) Privacy as information access and illusory control: the case of the facebook news feed privacy outcry. *Electron. Com. Res. Appl. (Special Issue on Social Networks and Web 2.0)* **9**(1): 50–60.

[2] Mifflin, H.e. (2004) *The American Heritage Dictionary of the English Language* (New York, NY: Houghton Mifflin).

[3] Denning, D.E. and Morgenstern, M. (1986) Military database technology study: AI techniques for security and reliability. *SRI Technical Report.*

[4] Machanavajjhala, A., Gehrke, J. and Kifer, D. (2006) l-diversity: privacy beyond *k*-anonymity. In *Proceedings of the 22nd IEEE International Conference on Data Engineering (ICDE 2006).*

[5] O'Leary, D.E. (1995) Some privacy issues in knowledge discovery: The OECD personal privacy guidelines. *IEEE Expert: Intell. Syst. Appl.* **10**(2): 48–52.

[6] Sweeney, L. (2002) Achieving *k*-anonymity privacy protection using generalization and suppression. *Int. J. Uncertainty Fuzziness Knowledge Based Syst.* **10**(5): 571–588.

[7] Zhan, J. and Matwin, S. (2006) A crypto-based approach to privacy-preserving collaborative data mining. In *Sixth IEEE International Conference on Data Mining Workshops*, 546–550.

[8] Brodsky, A., Farkas, C. and Jajodia, S. (2000) Secure databases: constraints, inference channels, and monitoring disclosures. *IEEE Trans. Knowl. Data Eng.* **2**(6): 900–919.

[9] Wonderacek, G., Holz, T., Kirda, E. and Kruegel, C. (2010) A practical attack to de-anonymize social network users. In *IEEE Symposium on Security and Privacy.*

[10] Schuler, R.P., Laws, N., Bajaj, S., Grandhi, S.A. and Jones, Q. (2007) Finding your way with CampusWiki: a location-aware Wiki to support community building. In *The ACMs Conference on Human Factors in Computing Systems CHI2007* (San Jose, CA).

[11] Magid, L. (2010) EPIC complains while Google tries to fix buzz privacy snafus. *The Huffungton Post.*

[12] Motahari, S., Manikopoulos, C., Hiltz, R. and Jones, Q. (2007) Seven privacy worries in ubiquitous social computing. In *ACM International Conference Proceeding Series; Proceedings of the 3rd Symposium on Usable Privacy and Security* (ACM), 171–172.

[13] Langheinrich, M. (2001) Privacy by design—principles of privacy-aware ubiquitous systems. In *Third International Conference on Ubiquitous Computing (UbiComp 2001)* (ACM), 273–291.

[14] Stallings, W. (1999) *Cryptography and Network Security Principles and Practices* (New Jerssey, NJ: Pearson Prentice Hall).

[15] Cornwell, J., Fette, I., Hsieh, G., Prabaker, M., Rao, J., Tang, K., Vaniea, K., *et al.* (2007) User-controllable security and privacy for pervasive computing. In *Proceedings of the 8th IEEE Workshop on Mobile Computing Systems & Applications.*

[16] Hong, J.I. and Landay, J.A. (2004) An architecture for privacy-sensitive ubiquitous computing. In *2nd International Conference on Mobile Systems, Applications, and Services* (ACM), 177–189.

[17] Samarati, P. and De Capitani di Vimercati, S. (2001) Access control: policies, models, and mechanisms. In Focardi, R. and Gorrieri, R. [eds.] *Foundations of Security Analysis and Design* (Berlin, Heidelberg, New York: Springer-Verlag), 137–196.

[18] Motahari, S., Ziavras, S., Schular, R. and Jones, Q. (2008) Identity inference as a privacy risk in computer-mediated communication. *IEEE Hawaii International Conference on System Sciences (HICSS-42)*, 1–23.

[19] Chen, Y. and Chu, W. (2010) Database security protection via inference detection. In *IEEE International Conference on Intelligence and Security Informatics*.

[20] Lunt, T.F. (1991) Current issues in statistical database security. *IFIP Transactions, Results of the IFIP WG 11.3 Workshop on Database Security V: Status and Prospects A-6*, 381–385.

[21] Jajodia, S. and Meadows, C. (1995) *Inference Problems in Multilevel Secure Database Management Systems* (Los Alamitos, CA: IEEE Computer Society Press).

[22] Stachour, P.D. and Thuraisingham, B. (1990) Design of LDV: a multilevel secure relational database management. *IEEE Trans. Knowl. Data Eng.* **2**(2): 190–209.

[23] Dawson, S., Capitani, S.D. and Samarati, d.V.P. (1999) Specification and enforcement of classification and inference constraints. In *IEEE Symposium on Security and Privacy*, 181–195.

[24] Serjantov, A. and Danezis, G. (2002) Towards an information theoretic metric for anonymity. In *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)* (ACM).

[25] Morgenstern, M. (1987) Security and inference in multilevel database and knowledge-base systems. In *Proceedings of the 1987 ACM SIGMOD International Conference on Management of Data* (ACM).

[26] Lemos, R. (2011) How Your Username May Betray You, http://www.technologyreview.com/web/32326/?p1=MstRcnt&a=f, 2011.

[27] Motahari, S., Ziavras, S. and Jones, Q. (2010) Online anonymity protection in computer-mediated communication. *IEEE Trans. Inf. Forensics Secur.* **5**(3): 570–580.

[28] Motahari, S., Ziavras, S., Naaman, M., Ismail, M. and Jones, Q. (2009) Social inference risk modeling in mobile and social applications. *IEEE International Conference on Information Privacy, Security, Risk and Trust (PASSAT)*, 125–132.

[29] Shannon, C.E. (1950) Prediction and entropy of printed English. *Bell Syst. Tech. J.* **30**: 50–64.

[30] Motahari, S. (2010) *Inference Prevention in Ubiquitous Social Computing*. (Newark, NJ: Electrical and Computer Engineering Department, Institute of Technology).