Detecting Privacy Preferences from Online Social Footprints: A Literature Review

Taraneh Khazaei[1], Lu Xiao[1,2], Robert E. Mercer[1], Atif Khan[3]
[1]Department of Computer Science, University of Western Ontario
[2]Faculty of Information & Media Studies, University of Western Ontario
[3]InfoTrellis Inc., Toronto, Ontario

Abstract
Providing personalized content can be of great value to both users and vendors. However, effective personalization hinges on collecting large amounts of personal data about users. With the exponential growth of activities in social networking websites, they have become a prominent platform to gather and analyze such information. Even though there exist a considerable number of social media users with publicly available data, previous studies have revealed a dichotomy between privacy-related intentions and behaviours. Users often face difficulties specifying privacy policies that are consistent with their actual privacy concerns and attitudes, and simply follow the default permissive privacy setting. Therefore, despite the availability of data, it is imperative to develop and employ algorithms to automatically predict users' privacy preferences for personalization purposes. In this document, we review prior studies that tackle this challenging task and make use of users' online social footprints to discover their desired privacy settings.
Keywords: Social Privacy, Privacy Preference, Personalization, Machine Learning, User Modeling
doi: 10.9776/16293

Acknowledgements: This project is funded through the MITACS Accelerate program.
Contact: tkhazae@uwo.ca

1 Introduction

With the growing interest in regular communication and information sharing over online social media, privacy has emerged as a serious concern. Prior studies on users' online behaviour indicate that there is a disparity between the privacy-related attitudes of social media users and their actual behaviour in specifying their privacy policies (Acquisti & Gross, 2006; Gross & Acquisti, 2005; Lipford, Besmer, & Watson, 2008; Strater & Lipford, 2008). Even though social media users may be highly concerned about their privacy, they face difficulties managing their privacy policies, so only a small percentage of users change their default privacy settings. This issue may occur due to their misconceptions regarding the visibility of their data (Strater & Lipford, 2008) and the complex and unusable interfaces (Fang & LeFevre, 2010). In addition, psychology research has shown that defaults are often perceived as the recommended course of action (McKenzie, Liersch, & Finkelstein, 2006; Johnson, Bellman, & Lohse, 2002). Additionally, users are rarely reminded to reconsider their privacy policies after their initial profile creation; hence, they often overlook the visibility of their social networking data (Strater & Lipford, 2008).

Meanwhile, as more people engage on social media, they are providing businesses with unprecedented amounts of data that give insight into various facets of customer behaviour. Current social networking platforms allow users to publish their activities, opinions, locations, as well as their social interactions through different forms of communication (e.g., text, image, and video), leading to large *social footprints* (Irani, Webb, Li, & Pu, 2009). The insight into customer behaviour provided by such social footprints affords immense opportunities for businesses to engage audiences with compelling and personalized content and experiences. By harnessing this additional information about individuals, traditional customer databases can be transformed from historical artifacts into powerful business intelligence tools, enabling efficient and effective business decision-making processes. For example, social media data can be used to detect users' upcoming life events and provide them with relevant offers, or to gain insight into users' psychographics and send marketing messages uniquely tailored to them.

In addition to the tremendous potential that social media data can provide for businesses, prior customer studies have shown that customers and clients value personalized content (Kobsa, 2007). In addition,

as customers increase their digital footprints, they expect more personalization (SAS, 2015). However, the effectiveness of this win-win opportunity relies on addressing users' privacy concerns and reconciling the tension between personalization and privacy. The Facebook "Beacon" feature is an alarming example of disregarding users' privacy preferences. Launched in November 2007, "Beacon" allowed third-party websites, such as Coca-Cola, Sony Pictures, and Verizon, to access Facebook profiles and to provide personalized content and services to them and their friends. This feature immediately encountered mass protests and was retracted from Facebook several weeks later.

Users' privacy is violated when information intended for a particular audience (such as one's family and friends) unintentionally becomes available to a broader audience (such as companies and organizations) (Richthammer, Netter, Riesner, Sänger, & Pernul, 2014). Given that users often fail to specify privacy policies that match their actual concerns, it is vital for businesses to take extra precautions when dealing with customer data, even when the underlying data is voluntarily disclosed and is publicly available. Following supplementary privacy-preserving methods provide organizations with a competitive advantage (Preibusch, Kübler, & Beresford, 2013) and allows them to build and maintain customer trust to avoid the negative consequences that may arise from neglecting customers' privacy preferences and to build effective personalization while preserving privacy.

The solutions proposed to address social network privacy issues include studies that present a set of privacy-enhancing principles and guidelines to design personalization systems (Kobsa, 2007; Toch, Wang, & Cranor, 2012), the works that suggest usable interfaces and visualization tools for specifying privacy policies (Mazzia, LeFevre, & Adar, 2012; Lipford et al., 2008; Anwar, Fong, Yang, & Hamilton, 2010; Gao & Berendt, 2013), as well as automated policy prediction models and frameworks (Squicciarini, Karumanchi, Lin, & DeSisto, 2014; Fang & LeFevre, 2010). In this review, we focus on the later direction. Specifically, we review the approaches that propose automated methods and utilize large social footprints available in online social networks to predict desired privacy settings. We conceptualize that an online social network is a virtual place in which individuals are allowed to create profiles and share their personal attributes, preferences, and opinions. In addition, they can connect to each other through different types of relationships and establish and maintain rich interactions with their peers on the network.

The remainder of this document is as follows: Section 2 describes the concept of privacy in the context of social networks. Then the major approaches on privacy preference inference in the context of social networks are provided in Section 3. Section 4 discusses a set of gaps based on the literature review. Finally, the manuscript concludes in Section 5

2 Privacy in Social Networks

Social networks are typically represented as a graph $G = \langle V, E \rangle$, where each user corresponds to a node $i \in V$. An edge $(i, j) \in E$ in the graph indicates some sort of social connection between the two users i and j. The labeling function F can be defined as $F : V \rightarrow R$, where V is the set registered users and $R = \{r_1, r_2, ..., r_n\}$ is the finite set of the possible relationships connecting the users. A relationship r_k can be either bidirectional (e.g., friend relation in Facebook) or unidirectional (e.g., follower relation in Twitter). In addition, each user can have a set of properties and profile items $P = \{p_1, p_2, ..., p_n\}$ that indicates who a user is in the social network, such as their identity and personal information. Users may also be associated with a set of contents $C = \{c_1, c_2, ..., c_n\}$ that describes what a user has exposed in the social network, such as uploaded text, images, videos, and other data items created through various activities in the network.

In the majority of the current social networks, privacy settings are described in terms of access control for the shared profile and content items. Most of the popular social networking sites such as Facebook and Google+ allow users to specify their privacy settings by controlling "who sees what" of their data. Varying granularity degrees of privacy specification are typically supported for both "who" and "what" variables. For example, users can set the information visibility as either public or private; or they can assign various specifications for different groups of their social contacts, or even different social contacts individually. Likewise, users may be allowed to specify privacy attributes for all of their published items at once, different categories of items, or each piece of shared items individually. In addition, the information access control can be either specified as a binary value (e.g., *allow* and *deny*) or on a nominal scale (e.g., *view*, *comment*, and *re-share*).

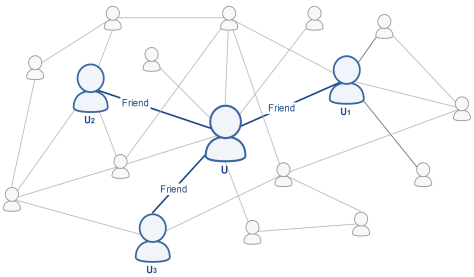

Figure 1: An example of a social network with focal user U.

	u_1	u_2	u_3
p_1	allow	deny	deny
c_1	allow	allow	allow
c_2	deny	allow	deny

Table 1: An example privacy specification in a social network.

For instance, consider the partial network of a user presented in Figure 1, where the focal user U is connected with three social contacts u_1, u_2, u_3 through a similar relation type (i.e., friendship). Suppose this user has one published profile item $p = \{p_1\}$ and two shared content items $c = \{c_1, c_2\}$. Table 1 represents a possible privacy specification for user U in a social network, where access control is specified at a binary level for each of the social contacts and each of the published items separately. As discussed earlier, managing such privacy settings can be a cumbersome and a tedious task for hundreds of social contacts and shared items and so is often overlooked by users. As such, some researchers have utilized online social footprints and have proposed semi-automatic or automatic techniques to derive privacy policies that are similar to what current social networks provide as their privacy settings (e.g., (Danezis, 2009)). Instead of focusing on privacy configurations that are specific to a particular social networking site, a set of studies attempted to use such footprints to characterize users' general privacy preferences. In these works, privacy preferences of users can be determined by mapping users to a binary, numeric, or an ordinal scale of desired privacy (e.g., (Caliskan Islam, Walsh, & Greenstadt, 2014)). Here, we attempt to provide a comprehensive review of the studies that take either of the approaches to infer privacy attributes.

3 Literature Review

Due to the growing interest in regular communication and information sharing over online social media, research on these platforms gained great attention in the last few decades. The rich information available in social media can greatly benefit individuals, communities and societies, businesses, politicians and governments, as well as scholars. The prior works on automatic detection of privacy preferences are categorized based on the type of the data used to make the prediction. Therefore, Section 3.1 explains the studies that have relied on the potential links between personal characteristics of the users and their privacy preferences to infer the privacy attributes. Section 3.2 presents the algorithms that are primarily focused on the users' social context and ties. In addition, some researchers have used the content published by users to derive their desired privacy features. These content-based approaches are reviewed in Section 3.3.

3.1 Personal and Profile Attributes

There is a large body of research linking demographic information as well as personal traits to privacy preferences. For example, various surveys have established a positive relation of age, education, and income linked with privacy concerns (Kobsa, 2007). In the context of social media, various studies have been conducted to find possible connections between gender and age and privacy behaviour (Fogel & Nehmad, 2009; Lewis, Kaufman, & Christakis, 2008; Boyd & Hargittai, 2010; Dey, Jelveh, & Ross, 2012). For gender, the

results are inconclusive as some of the studies found no gender difference to privacy settings, while some found female users to be more private. Similarity, even though Dey et al. (Dey et al., 2012) have shown that adults tend to be more private in social media, the research conducted by Christofides et al. (Christofides, Muise, & Desmarais, 2011) has shown that adults and adolescents exhibit similar privacy behaviour. Similar attempts have been also made to study the possible connections between location (Dey et al., 2012; Christofides et al., 2011) as well as ethnicity (Minkus & Memon, 2014) with privacy attributes.

In addition, personality attributes of people have been shown to be associated with their privacy attitudes and behaviour. For instance, in the context of location-based services, Junglas et al. (Junglas, Johnson, & Spitzmüller, 2008) have researched the possible connections of the so-called Big Five personality traits (i.e., agreeableness, extraversion, emotional stability, openness to experience, and conscientiousness) and privacy concerns. They found that people who scored high on agreeableness, conscientiousness, and openness expressed lower levels of privacy concerns.

Despite the large body of works linking demographics and personality features to privacy concerns, to the best of our knowledge, there exist only a few works that have utilized this type of information to recommend privacy features. By using an online survey, Minkus and Memon (Minkus & Memon, 2014) examined the privacy settings of users on Facebook and related their choices to demographic and personality features. The survey results are later used to build and deploy an online application, called MyPrivacy, that automatically recommends privacy settings. Their survey results provide evidence for earlier studies, indicating that personality traits and demographics are linked with privacy behaviour. In particular, they found that neuroticism, age, ethnicity, and the self-reported concern for privacy are related to the customized privacy settings of users on Facebook. Therefore, MyPrivacy first asks multiple questions from users to determine these attributes and then uses a supervised learning algorithm to recommended privacy settings. The evaluation of MyPrivacy showed positive subjective opinions of real Facebook users toward the tool. To recommend privacy settings for a particular shared item, (Naini Djafari, Altingovde, Kawase, Herder, & Niederée, 2015) proposed a supervised method as well. Their algorithm is built on a set of demographic features including age, gender, and location; along with a set of metadata associated with the shared item.

The lack of approaches focused on personal attributes can be due to the inconsistent and inconclusive results obtained from the studies that analyze the connections between such attributes and privacy preferences. These conflicting empirical differences may stem from the differences in what they measure as privacy preference. While some researchers may measure privacy behaviours to indicate privacy preference, some may be focused on privacy attitudes. Besides, in these works, the data collection process is often limited to specific and often rather small participant pools, such as people living in New York City (Dey et al., 2012) or college students (Lewis et al., 2008). Further studies with large and diverse participant sets may lead to consistent results that can be used reliably in automatic prediction tools. In addition, demographics and personal attributes may not be directly accessible through social media profiles, leading to the availability of a sparse set of attributes. Even though successful attempts have been made to extract this information from users' activities in their social network (Adali & Golbeck, 2012; Golbeck, Robles, & Turner, 2011; Rao, Yarowsky, Shreevats, & Gupta, 2010), these approaches are often complex and require extra computational resources.

3.2 Social Context

Compared to the use of personal attributes, a large set of studies have focused on the social context of the focal user to analyze and predict privacy-related features. These studies can be categorized into two primary groups. The first set of works mainly focuses on privacy in terms of information visibility to different groups of social contacts, often referred to as social circles. Hence, they propose approaches to assist users in creating and maintaining such social circles and their corresponding privacy policies.

While the aforementioned group of works are focused on partitioning and clustering users' social contacts, they do not make use of the valuable information hidden in their social context. Hence, another set of researchers has adapted techniques from the area of collaborative filtering to assign privacy policies to a user based on the preferences of other users. One approach to determine this set of users is to select them from within the social contacts of the focal user (e.g., friends in Facebook or followers in Twitter). This method follows the principle of homophily, which refers to the tendency of people to associate with similar individuals and has been observered in the context of online social networks (McPherson, Smith-Lovin, & Cook, 2001). As an alternative to the use of social contacts, a set of researchers has developed and used a set

of similarity measures to select users with similar backgrounds and characteristics with the focal user.

3.2.1 Social Circle Management and Labeling

Given that users have on average hundreds of friends[1,2], specifying a policy that manages access to various information items is a difficult and a tedious task even for privacy-conscious users. As a result, with the aim of easing the process of privacy policy management, there have been attempts to automatically categorize users' social contacts into meaningful social circles. Some studies have moved beyond clustering and proposed techniques to infer user's preferred privacy settings for the created circles of contacts.

Adu-Oppong et al. (Adu-Oppong, Gardine, Kapadia, & Tsang, 2008) proposed that the clustering algorithm presented in (Mishra, Schreiber, Stanton, & Tarjan, 2007) can be used to effectively create social circles of densely and closely connected contacts in unidirectional networks. Following this approach, (α, β)-clusters will be formed so that any node in a cluster is adjacent to at least a β-fraction of the cluster and any node outside of a cluster is adjacent to at most an α-fraction of the cluster. In a somewhat similar approach, Danezis (Danezis, 2009) proposed an algorithm to cluster one's social contacts into circles that are closely related to each other and have many links within themselves, while having fewer links with those who are not in the circle. In (Squicciarini et al., 2014), a large number of unique characteristics such as educational background, hobbies, and age are taken into account for clustering social contacts. A modified version of the apriori algorithms (Agrawal & Srikant, 1994) is used to dynamically select clustering features based on the attributes of the social contacts of the focal user.

Jones and O'Neill (Jones & O'Neill, 2010) conducted user studies and interviews to understand user rationales when grouping their social contacts for the purpose of privacy management. As a result of this experiment, a set of six criteria commonly considered by users was identified. Since these criteria are related to the relationships between users, a network clustering algorithm, called SCAN (Xu, Yuruk, Feng, & Schweiger, 2007) is used to group one's social network into various circles.

Some researchers have proposed supplementary techniques to clustering to recommend privacy settings for the created clusters. In (Shehab, Cheek, Touati, Squicciarini, & Cheng, 2010), for instance, after the clusters of contacts are formed, the user is asked to label a number of randomly selected contacts from each cluster. Through the labeling process, the user indicates his/her willingness to share a specific item with them. A classifier is trained on the profile attributes as well as the network attributes of the labeled contacts to predict the privacy preferences of the user for unlabeled contacts relative to a specific object information item. They achieved an accuracy of 83% with 20% training.

Fang and LeFevre (Fang & LeFevre, 2010) built a privacy wizard that iteratively asks the user to label carefully-selected informative contacts. In these questions, the user specifies his/her willingness to share a specific piece of profile information with a social contact. To automatically label other contacts, these labeled information is utilized in a classifier, wherein contacts are represented as feature vectors that encompass users' community structure and profile features such as age, gender, and education.

A classifier is developed in (Li, Li, Wang, & Ginjala, 2011) to decide whether a data item should be visible to a contact of a user. Based on the assumption that the privacy labels that have been explicitly assigned to friends are correct, users' current privacy settings is used as the labeled data. Similar to (Fang & LeFevre, 2010), friends are then represented as a feature vector that includes their community attributes and their personal features. A classifier is then built on this feature set to assign privacy labels to unlabeled contacts.

3.2.2 Collaborative Filtering for Privacy Inference

Instead of categorization and labeling of the social context, some researchers have proposed methods to identify privacy preferences based on the privacy characteristics of the social context. Users' information sharing behaviour has shown to be extensively influenced by an inner circle of close friends (Caliskan Islam et al., 2014). For instance, the amount of private information shared by a user has shown to be correlated with the amount of private information shared by friends. Similarily, people with similar backgrounds tend to have similar privacy concerns (Squicciarini, Lin, Sundareswaran, & Wede, 2015). These findings has

[1] http://www.theguardian.com/news/datablog/2014/feb/04/facebook-in-numbers-statistics
[2] http://news.yahoo.com/twitter-statistics-by-the-numbers-153151584.html

motivated researchers to adapt collaborative filtering methods and determine one's privacy preferences from attributes of his/her network. Collaborative filtering uses the known preferences of a group of users to make recommendations of the unknown preferences of other users and is mainly utilized in the context of recommendation systems (Su & Khoshgoftaar, 2009).

Squicciarini et al. (Squicciarini et al., 2014; Squicciarini, Karumanchi, Lin, & DeSisto, 2012) provide an algorithm to form social circles based on users' characteristics such as their gender, hobbies, and occupation. These circles are further utilized to recommend privacy policies for newly added objects (i.e., added contacts or uploaded data items). When a new object is uploaded, the system first seeks the social circles that is most likely to deal with the object in a similar way as the user. Then the privacy policies used by the selected circle is the basis for predicting the privacy policy for the newly added object. Similar idea is applied to user-uploaded images (Squicciarini et al., 2015), in which a policy prediction algorithm assigns a policy to a newly uploaded image based on the information captured from social circles.

In (Shehab & Touati, 2012), active learning and the properties of the social graph are first used to detect a set of the most informative contacts to be labeled as training samples. In the labeling process, the user specifies whether he/she is willing to share a specific data item with the selected contact. Then an iterative semi-supervised approach is followed to label the other contacts of the user, where labels are propagated from labeled instances to unlabeled instances in the social graph. This propagation is guided by the user similarity metric that is represented through edge weights. The similarity computation is based on profile information of contacts, their networks metrics, as well as the community structure. The evaluation results of this approach provided higher accuracy and precision compared to a supervised learning and a random walk based approach.

In the context of a location-based social network, Toch et al. (Toch, Sadeh, & Hong, 2010) provide users with recommended privacy policies that similar users have previously selected. A large set of privacy policies is first clustered based on their location, time, and social group properties. Policies within each cluster are then ranked according to the number of policies they are similar to and their similarity degree. To recommend privacy policies, clusters that are relevant to the current user are selected based on the policies chosen by similar users (e,.g., users that are within the same Facebook network). Finally, top-ranked policies from the selected clusters are presented as recommendations.

Collaborative filtering is also followed in (Ghazinour, Matwin, & Sokolova, 2013), where the authors take advantage of a set of profile features, user's interests implied in their social media, and their privacy configurations to find a set of users similar to the user of focus. In their approach, users are first characterized according to their privacy preference as either privacy fundamentalist, privacy pragmatist, or privacy unconcerned. Users' privacy decisions and settings regarding their photo albums is considered as an indication of their privacy preference. In particular, users are assigned to these three categories based on the number of their public, customized, and private photo albums. Then K-nearest neighbour algorithm is used to determine which privacy categorization the focus user belongs to. Based on the features of the assigned category, the system then recommends privacy settings.

3.3 Published Content

A frequent user activity on social networks is to publish and share content such as status messages, comments, images, and videos. All instances of shared data types can be used to draw inferences about the users' personality and preferences. In particular, natural language has been shown to be a reflection and a mediator of internal states (Pennebaker, Mehl, & Niederhoffer, 2003). Our words can reveal personality, emotional states and feelings, attention patterns, thought, and social situations (Pennebaker et al., 2003; Gill, Vasalou, Papoutsi, & Joinson, 2011). Therefore, a variety of automated content analysis techniques have been developed to measure such psychometric metrics from natural language. These methods range from the use of predefined dictionaries and taxonomies such as Language Inquiry and Word Count (LIWC) to more sophisticated computational algorithms that utilize complex data mining and machine learning based techniques.

In the context of privacy, Gill et al. (Gill et al., 2011) provide a set of privacy-related categories, each of which is associated with a number of words that are relevant in the semantic analysis of the privacy domain. The dictionary consists of 388 privacy related words that are grouped into eight high-level theoretically sound categories based on their semantic similarity. LIWC contains a large number of semantic categories with possible relevance to privacy features. Therefore, LIWC is used as the baseline for the evaluation of the

privacy dictionary. The evaluation results indicate that the privacy dictionary is capable of capturing unique linguistic features in privacy language and is more reliable in detecting privacy-oriented content.

Caliskan-Islam et al. (Caliskan Islam et al., 2014) used the privacy dictionary, along with a variety of methods and tools including topic modeling, named entity recognition, and sentiment analysis to automatically deem if a tweet contains private information. Annotated data were collected from Amazon Mechanical Turk (AMT), wherein AMT workers were asked to label collected tweets according to privacy categories. Then users are given privacy scores based on the amount of private information they published in their Twitter timeline. The timelines of these labeled users are then utilized in a supervised machine learning technique to assign privacy scores to unlabeled users based on their shared textual content.

The prediction model proposed in (Naini Djafari et al., 2015) follows a supervised machine learning approach to recommend privacy settings for a given post in Facebook. Besides the demographic features (as explained in Section 3.1), they used a set of content-based features associated with the post. The sentiment score of the post is included, along with some topical attributes. In addition, the entire bag-of-word representation of the content is taken into account, where only a set of words with a high tf-idf score is considered. Some contextual metadata elements are also used, such as the time of the day the post is shared.

Given an unstructured linguistic content published by a user, (Srivastava & Geethakumari, 2013) first detects sensitive information such as phone number, address, and location from the text. Then the model proposed in (Liu & Terzi, 2010) is adopted to quantify the privacy risk of the user, wherein the identified sensitive parts are treated the same way as information items in (Liu & Terzi, 2010). In (Liu & Terzi, 2010), a mathematically sound model is developed, taking into account the sensitivity and visibility of the shared items. The proposed model provides users with a privacy score that quantifies the potential privacy risk of the user. The premise behind their model is that the more sensitive information the user discloses, the higher his or her privacy risk. In addition, the more visible the shared information becomes in the network, the higher the privacy risk.

In the context of image sharing, (Squicciarini et al., 2015) uses the previous images by users and their corresponding privacy policies to assign a privacy policy to a new image. Image clustering and policy association mining are used for privacy generation. However, if the user is new or there have been significant changes to the user's privacy trends, users' social context is used to predict the policy as explained in Section 3.2. To detect images with private content, Zerr et al. (Zerr, Siersdorfer, Hare, & Demidova, 2012) used a variety of visual features, such as the occurrence of faces, in a supervised learning algorithm. They also utilized the textual metadata associated with images and found correlations between topics and the content of private images. For instance, topics used to describe personal concepts, emotions and sentiment, and human body were shown to be mostly used for private images. On the other hand, topics related to nature, architecture, and inanimate objects have been mostly found in non-private images.

4 Discussion

Automatic privacy inference has received relatively little attention. In addition, our analysis of the related literature suggests that the research efforts are mainly focused on the prediction and recommendation of privacy settings that are specific to the underlying social network. On the other hand, attempts have been made to quantify users' privacy risk (Liu & Terzi, 2010; Becker & Chen, 2009) and users' current privacy level (Caliskan Islam et al., 2014). Ghazinour et al. (Ghazinour et al., 2013) (discussed in 3.2.2) characterized individuals by classifying them into different levels of privacy concerns based on their privacy settings on their photo albums. This lack of work may be attributed to the fact that privacy is context-dependent issue (Nissenbaum, 2009), making it challenging to develop generic methods. However, a recent study in the context of mobile applications has revealed that despite the diversity of privacy preferences, users can be clustered into a set of meaningful privacy profiles that effectively captures their desired privacy (Lin, Liu, Sadeh, & Hong, 2014). These studies imply the potential of characterizing users according to their platform-independent privacy preferences.

Another research gap we identified is the limited set of data types that prior studies have focused on. These data types are often the users' profile features, social and network attributes, and the content of communications. Many other data types are left unexplored in the literature (Richthammer et al., 2014; Beye et al., 2012). For instance, *ratings/interests* of users can be of value in gaining insight into one's privacy

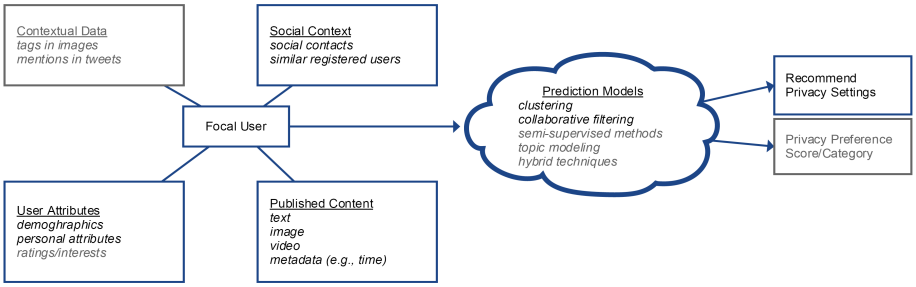

Figure 2: An overview of privacy prediction approaches in terms of their input, techniques, and goals.

preferences and latent attributes, and are often readily accessible in users' profiles. However, to the best of our knowledge, only (Ghazinour et al., 2013) has used it to infer privacy. Another example is *contextual data*. This data type refers to the property of an item that is made explicit and is provided with semantics, such as the tags provided in Facebook images and status messages or mentions in tweets. Although users' privacy preferences may be revealed from these contextual data, researchers have not used them in detecting privacy features.

By analyzing the approaches used, it can be seen that the existing literature lacks a study of hybrid techniques and of mixed data types. However, in similar areas such as recommendation systems, hybrid approaches have been shown to be very effective and can offset limitations of either approach and improve the prediction performance (Su & Khoshgoftaar, 2009).

Many of the reviewed studies have utilized supervised methods to classify and predict privacy attributes. However, they require labeled input, and may not seem feasible in the context of social media, where the labeled information normally constitutes a very small portion of the available data. In such a context, unsupervised and semi-supervised techniques can be of great interest. In particular, semi-supervised techniques, which have the advantage of utilizing fewer labeled data to achieve better predictions, can be a potential research avenue to explore further. A graph-based semi-supervised method has been proposed to effectively capture privacy preference (Shehab & Touati, 2012), and other methods such as Expectation and Maximization (EM), topic modeling, and co-training need to be investigated further. For instance, co-training has been successfully used to detect users' latent personal attributes in social networks (Mo, Wang, Li, Hong, & King, 2010).

Figure 2 provides an overview of our reviewed studies in terms of their input, proposed techniques, as well as their goals and purposes. Examples of each of these elements are also provided. The figure also indicates the research gaps we discussed above with gray boxes. The discussed limitations of prior studies call for further attempts to deeply analyze how different facets of large online social footprints can be utilized to effectively characterize users' privacy preferences.

5 Summary and Conclusion

Mining the treasure trove that exists in social media has tremendous potential for companies to improve the customer experience through personalization and targeted marketing. However, customers may not be willing to be profiled online due to their privacy concerns. On the other hand, users' privacy concerns are often not well translated into their social network privacy configuration, resulting in generating data that is accessible to the public. It is a dilemma for companies whether to use one's publicly available data to provide valuable personalized content or not to use such data to avoid disregarding privacy preferences. One potential direction to manage this dilemma is to develop novel algorithms and techniques that take advantage of users' social footprints and characterize their privacy behaviour and attitude. In this document, we reviewed the existing literature on automatic privacy preference inference in the context of social networks.

We categorized and reviewed the existing studies on privacy preference inference according to the data type of focus, namely demographics and profile features, social context and network features, as well as the shared content. The potential and limitations of the approaches are further discussed, where a set of

gaps are identified in the literature. Based on our study of the literature, we call for more studies of general user modeling and characterization according to their privacy preference. In addition, researchers studying privacy detection are encouraged to use a wider range of data types available in social media as well as hybrid techniques to make the predictions.

References

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Proceedings of the conference on privacy enhancing technologies* (pp. 36–58).

Adali, S., & Golbeck, J. (2012). Predicting personality with social behavior. In *Proceedings of the IEEE/ACM conference on advances in social networks analysis and mining* (p. 302-309).

Adu-Oppong, F., Gardine, C., Kapadia, A., & Tsang, P. (2008). Social circles: Tracking privacy in social networks. In *Proceedings of the symposium on usable privacy and security.*

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of the conference on very large databases* (p. 487-499).

Anwar, M., Fong, P. W., Yang, X.-D., & Hamilton, H. (2010). Visualizing privacy implications of access control policies in social network systems. In J. Garcia-Alfaro, G. Navarro-Arribas, N. Cuppens-Boulahia, & Y. Roudier (Eds.), *Data privacy management and autonomous spontaneous security* (Vol. 5939, p. 106-120). Springer Berlin Heidelberg.

Becker, J. L., & Chen, H. (2009). Measuring privacy risk in online social networks. In *Web 2.0 security and privacy workshop.*

Beye, M., Jeckmans, A., Erkin, Z., Hartel, P., Lagendijk, R., & Tang, Q. (2012). Privacy in online social networks. In *Computational social networks: security and privacy* (p. 87-113).

Boyd, D., & Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday*, *15*(8), 1-24.

Caliskan Islam, A., Walsh, J., & Greenstadt, R. (2014). Privacy detective: Detecting private information and collective privacy behavior in a large social network. In *Proceedings of the workshop on privacy in the electronic society* (pp. 35–46).

Christofides, E., Muise, A., & Desmarais, S. (2011). Hey mom, what's on your facebook? Comparing facebook disclosure and privacy in adolescents and adults. *Social Psychological and Personality Science*, *3*.

Danezis, G. (2009). Inferring privacy policies for social networking services. In *Proceedings of the acm workshop on security and artificial intelligence* (p. 5-10).

Dey, R., Jelveh, Z., & Ross, K. (2012). Facebook users have become much more private: A large-scale study. In *Proceedings of the conference on pervasive computing and communications workshops* (p. 346-352).

Fang, L., & LeFevre, K. (2010). Privacy wizards for social networking sites. In *Proceedings of the international conference on world wide web* (pp. 351–360).

Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, *25*(1), 153-160.

Gao, B., & Berendt, B. (2013). Circles, posts and privacy in egocentric social networks: An exploratory visualization approach. In *Proceedings of the IEEE/ACM conference on advances in social networks analysis and mining* (pp. 792–796).

Ghazinour, K., Matwin, S., & Sokolova, M. (2013). Monitoring and recommending privacy settings in social networks. In *Proceedings of the joint EDBT/ICDT workshops* (p. 164-168).

Gill, A. J., Vasalou, A., Papoutsi, C., & Joinson, A. N. (2011). Privacy dictionary: A linguistic taxonomy of privacy for content analysis. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3227–3236).

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In *CHI extended abstracts on human factors in computing systems* (pp. 253–262).

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the ACM workshop on privacy in the electronic society* (pp. 71–80).

Irani, D., Webb, S., Li, K., & Pu, C. (2009). Large online social footprints–An emerging threat. In *Proceedings of the conference on computational science and engineering* (Vol. 3, p. 271-276).

Johnson, E., Bellman, S., & Lohse, G. (2002). Defaults, framing and privacy: Why opting in-opting out. *Marketing Letters*, *13*(1), 5-15.

Jones, S., & O'Neill, E. (2010). Feasibility of structural network clustering for group-based privacy control in social networks. In *Proceedings of the symposium on usable privacy and security* (pp. 9:1–9:13).

Junglas, I. A., Johnson, N. A., & Spitzmüller, C. (2008). Personality traits and concern for privacy: An empirical study in the context of location-based services. *European Journal of Information Systems*, *17*(4), 387–402.

Kobsa, A. (2007). Privacy-enhanced personalization. *Communications of ACM*, *50*(8), 24–33.

Lewis, K., Kaufman, J., & Christakis, N. (2008). The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, *14*(1), 79-100.

Li, Q., Li, J., Wang, H., & Ginjala, A. (2011). Semantics-enhanced privacy recommendation for social networking sites. In *Proceedings of the conference on trust, security and privacy in computing and communications* (p. 226-233).

Lin, J., Liu, B., Sadeh, N., & Hong, J. I. (2014). Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Symposium on usable privacy and security.*

Lipford, H. R., Besmer, A., & Watson, J. (2008). Understanding privacy settings in Facebook with an audience view. In *Proceedings of the conference on usability, psychology, and security* (p. 2:1-2:8).

Liu, K., & Terzi, E. (2010). A framework for computing the privacy scores of users in online social networks. *ACM Transactions on Knowledge Discovery from Data*, *5*(1), 6:1–6:30.

Mazzia, A., LeFevre, K., & Adar, E. (2012). The PViz comprehension tool for social network privacy settings. In *Proceedings of the symposium on usable privacy and security* (pp. 13:1–13:12).

McKenzie, C. R., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, *17*(5), 414-420.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415-444.

Minkus, T., & Memon, N. (2014). Leveraging Personalization To Facilitate Privacy. *ArXiv e-prints.*

Mishra, N., Schreiber, R., Stanton, I., & Tarjan, R. (2007). Clustering social networks. In A. Bonato & F. R. Chung (Eds.), *Algorithms and models for the web-graph* (Vol. 4863, p. 56-67). Springer Berlin Heidelberg.

Mo, M., Wang, D., Li, B., Hong, D., & King, I. (2010). Exploit of online social networks with semi-supervised learning. In *Proceedings of the joint conference on neural networks* (p. 1-8).

Naini Djafari, K., Altingovde, I., Kawase, R., Herder, E., & Niederée, C. (2015). Analyzing and predicting privacy settings in the social web. In F. Ricci, K. Bontcheva, O. Conlan, & S. Lawless (Eds.), *User modeling, adaptation and personalization* (Vol. 9146, p. 104-117). Springer International Publishing.

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life.* Stanford University Press.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, *54*(1), 547–577.

Preibusch, S., Kübler, D., & Beresford, A. (2013). Price versus privacy: An experiment into the competitive advantage of collecting less personal information. *Electronic Commerce Research*, *13*(4), 423-455.

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in twitter. In *Proceedings of the international workshop on search and mining user-generated contents* (p. 37-44).

Richthammer, C., Netter, M., Riesner, M., Sänger, J., & Pernul, G. (2014). Taxonomy of social network data types. *EURASIP Journal on Information Security*, *2014*(1).

SAS. (2015). Finding the right balance between personalization and privacy. *SAS Report.*

Shehab, M., Cheek, G., Touati, H., Squicciarini, A., & Cheng, P. (2010). User centric policy management in online social networks. In *Proceedings of the IEEE symposium on policies for distributed systems and networks* (p. 9-13).

Shehab, M., & Touati, H. (2012). Semi-supervised policy recommendation for online social networks. In *Proceedings of the conference on advances in social networks analysis and mining* (p. 360-367).

Squicciarini, A., Karumanchi, S., Lin, D., & DeSisto, N. (2012). Automatic social group organization and privacy management. In *Proceedings of the conference on collaborative computing: Networking, applications and worksharing* (p. 89-96).

Squicciarini, A., Karumanchi, S., Lin, D., & DeSisto, N. (2014). Identifying hidden social circles for advanced privacy configuration. *Computers & Security*, *41*, 40 - 51.

Squicciarini, A., Lin, D., Sundareswaran, S., & Wede, J. (2015). Privacy policy inference of user-uploaded images on content sharing sites. *IEEE Transactions on Knowledge and Data Engineering*, *27*(1), 193-206.

Srivastava, A., & Geethakumari, G. (2013). Measuring privacy leaks in online social networks. In *Proceedings of the conference on advances in computing, communications and informatics* (p. 2095-2100).

Strater, K., & Lipford, H. R. (2008). Strategies and struggles with privacy in an online social networking community. In *Proceedings of the British HCI group annual conference on people and computers: Culture, creativity, interaction* (pp. 111–119).

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 1-19.

Toch, E., Sadeh, N. M., & Hong, J. (2010). Generating default privacy policies for online social networks. In *Extended abstracts on human factors in computing systems* (p. 4243-4248).

Toch, E., Wang, Y., & Cranor, L. (2012). Personalization and privacy: A survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction*, *22*(1-2), 203-220.

Xu, X., Yuruk, N., Feng, Z., & Schweiger, T. A. (2007). SCAN: A structural clustering algorithm for networks. In *Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining* (pp. 824–833).

Zerr, S., Siersdorfer, S., Hare, J., & Demidova, E. (2012). Privacy-aware image classification and search. In *Proceedings of the international ACM SIGIR conference on research and development in information retrieval* (pp. 35–44).