Privacy Tips: Would it be ever possible to empower online social-network users to control the confidentiality of their data?

Vladimir Estivill-Castro
Departament de Tecnologies de la
Informació i les Comunicacions
Universitat Pompeu Fabra
Roc Boronat, 138
Barcelona 08018 Spain
Email: vladimir.estivill@upf.edu

David F. Nettleton
Departament de Tecnologies de la
Informació i les Comunicacions
Universitat Pompeu Fabra
Roc Boronat, 138
Barcelona 08018 Spain
Email: david.nettleton@upf.edu

Abstract—Using the web for communication, purchases, searching information and/or socializing generates data, about ourselves, our connections and our activities, which is collected easily. In online social networks, users volunteer perhaps what is considered more personal information to their selected circles. But each person has personal preferences about what it considers public and what it considers private. The problem is that the information that is public may be used to disclose information that the users expect to remain confidential.

This paper offers a path to provide tips and warnings to each user of an online social network so they can exercise control on the information they consider private not only by not disclosing such information, but by acting on their public information-items that could be informative for those information-items that are private. This is a significant challenge, because most web-applications use personalization to build a context and provide better services. We aim to raise awareness on privacy and to empower users, giving them the possibility to regulate the benefits of personalization with the privacy risks. In this paper we also show that information-items (like relationships) can be chosen as confidential, and that we can provide meaningful warnings on metrics of association and public attributes that are strong predictors of confidential information-items.

I. INTRODUCTION

The revelation in the New York Times by Charles Dehigg that analytics of shopping habits were capable of solving successfully the challenging marketing problem of “anticipating a major event on a person’s life where their shopping loyalties may change” sent shock waves regarding privacy (a girl’s pregnancy was predicted before the individual disclosing the fact). On one other hand, it has been argued that such technology enabled Target to massively improve revenue from \$USD 44 billion in 2002 to \$USD 67 billion in 2010.

Recently, Assist. Prof. J. Vertesi (sociologist at Princeton University) attempted to hide her own pregnancy¹. Whenever online for something related to her pregnancy she used *Tor* [1]². She was very careful to insist with friends, family,

neighbors, and co-workers, to strictly avoid any mention of the pregnancy on Facebook or other online platforms. In the end she concluded the cost was overwhelming and almost impossible, but online social media being the first risk to address, followed by paying for everything with cash.

The biggest risk users face on joining a social network is the dispossession of control on the data about themselves. Users enjoy the benefits of networking, socializing, broadcasting, collaboration and communication on social networks, but with no capacity to regulate the trade-off on restricting to which circles which data is distributed, despite one of the most common definitions of privacy being the degree of control people have on their data. However, the most challenging aspect is to protect users from inferred data: while users may have kept some information-items confidential, the attributes that are made public may enable others to predict the confidential information. Thus the challenge is to make users aware of which non-confidential information about themselves can lead to the discovery of their confidential information.

In this paper, data mining and machine learning techniques will be used to enable users themselves to regulate the usage of their data and we believe that social-network providers would also benefit from offering users this control. We aim to raise awareness on privacy and to empower users, giving them the possibility to regulate the benefits of personalization with the privacy risk. We aim at characterizing risky public predictors of information-items (like relationships) on social networks and to provide meaningful warnings on the associations or connections users make, because such new connections may lead to data disclosure. After reviewing related work, we will use a running example to illustrate the transformation of the social-network data to tables of attribute-value data. The example is also used in the presentation of our methods, but we present experimental results for a real dataset.

II. RELATED WORK

Today, privacy is regarded as having two components. On one hand, determining and CONTROLLING [3] who has access to data. On the other hand, having mechanisms to PRACTICE

¹The authors would like to thank support from project TIN2013-49814-EXP of the MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD, SPAIN.

²See video www.youtube.com/watch?v=GaLpMwdPrYw

³This paper is not about anonymous communication on the Internet

such control [2, Page 403]. Each user has a degree of confidentiality related to how widely available each information-item should be. Also, personal data provides benefits as it enables personalization which increases the quality of service; and thus, personal data is regularly data mined. Thus, providers of web services usually identify valuable marketing opportunities. Web applications also enable capturing how and when (and often where) people use them. Note that recently, public Twitter data was used to identify within a few kilometers (6.3km) the location of over 100 millions US citizens [4] and demonstrated that 89% of Twitter users can be geo-located. This poses challenges to privacy [5] as it is well established that personalization conflicts with users' privacy [6]. "Privacy mechanisms are an important requirement for making systems usable. Users, if they do not feel protected enough from privacy breaches, will stop using them" [7]. The potential socio-economic benefit of the privacy-aware use of social networks is enormous: much more active users with effective valuable data would become available when users can reliably control the levels of disclosure and the privacy risks.

There are at least four books [9], [10], [11], [12] summarizing advances on privacy-preserving data mining. Privacy preserving techniques fall into two categories: *Category 1*) "algorithms for private data publication", and *Category 2*) "algorithms for private query answering". *Privacy in Social Networks* [13] has only brief discussions on evaluating risk. With a few exceptions, the focus of privacy preserving techniques on social-network data has been ACCESS CONTROL [12], [14]. There are three chapters dedicated to SOCIAL NETWORKS [15], [2], [16] in the recent book [12] on PRIVACY-AWARE KNOWLEDGE DISCOVERY. In contrast to all that, our proposal here goes directly to the new issue of PRIVACY PRACTICE [2] that includes both "strategic concealment" and revelation of information, based on a process of negotiation. "Tools should therefore support data concealment and revelation to help individuals practice privacy individually and collectively" [2]. Lederer et al. [17] suggested that privacy settings in Web applications should be enhanced through feedback that helps users understand the privacy implications of system usage under those settings. The feedback tools that exist are PRIVACY MIRRORS (that is, the capability for oneself to explore data about us, but from the eyes of others, and observe what we make visible to others). This is hardly the solution [2]: it is only based on information retrieval, provides little information about a true notion of privacy, has no information on usage data, and more importantly, no ability for negotiation for the attributes that one desires to actually propagate.

Carminati et al [15, and other citations within] also insist in providing participants in Web applications (and in particular on a social network) with more and more control over their data. The proposal of Carminati et al [15] consists of enabling users to describe their access policies through a rich Semantic Web logic language derived from OWL. This proposal is also followed by others [18], [19]. But we follow a different approach as the work of Gürses and Berendt [2] has shown that, even the simplest logic-based models result in inconsistencies and challenging privacy scenarios when two co-owners define separately the rules of access for their information objects (as well as other reasons).

Liu and Terzi [20] define privacy scores, extended to in-

clude networks structure and usage data. However, our measure and definition of sensitivity is different, based on inference and predictability by classifiers. We consider that sensitivity of an attribute is not only evaluated by how many users want to keep this private, but how much it is a predictor. Also, we do not make assumptions about the independence between independent variables, the independence between users and the independence between items and users.

We extend the mechanisms suggested earlier [8]. In this paper we explore information-items from social networks derived from the relationships (the links among users).

III. PERSONALIZED PROTECTIVE ANALYTICS

While there is research on identifying the most informative non-sensitive attributes for a confidential attribute [21], such research aims at producing sanitized public data sets. Thus, in such a case, one authority globally decides which attributes are sensitive and which ones are not, and records of all individuals are treated the same. In contrast to that, we follow here the approach [8] that enables each user to tailor their data, first by each user's selection of which attributes are sensitive and which are not, and also each user selecting which non-sensitive attributes to obfuscate and how. However, we will be dealing with the graph structure of social networks for the first time.

Thus, we review briefly the earlier personalized approach [8]. Our research is also tailored to the individual. We take into consideration that, what is public for one person may be confidential to another, and also, how precise the predictions and inferences by adversaries are, is user dependent. Most people would find innocuous that the city-field of their address is known; however, within what radius their home-location should be disclosed varies from person to person. Whatever a user considers confidential, we will codify as a Boolean predicate P . Examples include whether the user is female or male, whether the user speaks a language that belongs to romance languages, whether the user salary is in a range such as [\$50,000.\$100,000], whether their height is above average or whether it maintains regular contact with another user V . We would manipulate the size of ranges or sets, and the issue is then to what certainty can P versus $\neg P$ be predicted. This enables us to codify numeric or nominal attributes as nominal-binary attributes. The utility of this will be apparent later³.

Thus, consider a data set D where each row corresponds to a user and each column a_i corresponds to an attribute. Naturally, the data is available to the Web-provider. For the moment, the attributes are the result of the online activity of each user. However, we wish to assist an individual user c . User c indicates that some columns (say a_p) are confidential, this makes them inaccessible from public view. However, our algorithms will also tip the user of another column a_v , the user is prepared to disclose, but column a_v is a good predictor of one of the confidential columns.

The earlier work [8] provided warnings about disclosing the value v_v of predictor attribute a_v for user c using the expected information gain $IG(D, \vec{a}) = H(D) - H(D|\vec{a})$; where H denotes the information entropy, which is a measure of uncertainty. The expected information gain measures the change in

³This is analogous to the NominalToBinary filter of WEKA [22] or the existence or not of an edge to an attribute in the SAN [23].

information entropy from a prior state of knowledge $H(D)$ to a state $H(D|\vec{a})$ where some information $\vec{a}$ is given. However, it is important that the information gain is *personalized*.

That is, if user c wants to conceal the attribute-value pair $a_p = v_c$, it is the value v_c that must be obfuscated. If by analysing D , one finds that another attribute a_v is not a good predictor of the attribute a_p in general, that may offer no protection if the value for user c for a_v is a good predictor of $a_p = v_c$. As a running example, consider a deck of poker-playing cards. Suppose that c holds secretly two cards (like in the game of Texas Hold'em) whose rank is confidential. The sum of the ranks of the two cards (where picture cards and aces have rank above 10) is not a good predictor in general of what arbitrary players hold as their two cards, and we may incorrectly assume it is safe to release the value of the sum. However, if we know that for c the sum of the cards is 4, then it must be the case that both cards have rank 2.

Thus, from the individual perspective of user c , what we need to evaluate is how good is a_v as a predictor between two choices $a_p = v_c$ versus $a_p \neq v_c$, when we have the information for c ; namely when we know the value of attribute a_v for user c . Thus, we assess the information gain on the choices $a_p = v_c$ versus $a_p \neq v_c$ when we do not know c 's value for a_v and if we know it. We can generalize from a single predictor a_v , and a single column a_p as the confidential attribute. Let A be the total set of attributes, and A_c those attributes user c considers confidential. That is, c has not provided the values for the columns in A_c ; nevertheless, his/her values for the remaining attributes in $A \setminus A_c$ are disclosed. The problem is that a set of columns in $A \setminus A_c$ with the corresponding values for user c could be a very good predictor of $a_p = v_c$ versus $a_p \neq v_c$ for some $a_p \in A_c$. Since the data set D provides already some information, user c shall also produce a personal threshold θ_c of risk. Then, the challenge is to find $\mathcal{A}$, the family of all subsets $A_d \subseteq A \setminus A_c$ where $H_{a_p=v_c}(D) - H_{a_p=v_c}(D|a_c, A_d)$ is above θ_c for some confidential attribute $a_p \in A_c$.

The second challenge is to find the smallest set A_s of attributes in $A \setminus A_c$ that intersects every set in $\mathcal{A}$. Note that the second challenge is the well-known NP-complete problem named HITTING SET. It is not surprising that an early approach used heuristics [8] to attempt this challenge, even though they were dealing with just the attributes and not the relationships in a social network.

IV. PROTECTING ATTRIBUTE AND CONNECTION INFERENCE

Inference from the data in an online social network is based on principles like social influence [24], which suggest users mimic other users they are connected to, and common attributes are shared by members of communities (highly connected individuals with similar profiles). The inverse also happens [25], users with common attributes are likely to become connected [26], [27].

Thus, a series of techniques exist to discover two types of information-items from the data in a social network. These two types are the attribute-value pair of a user, and the existence or imminent emergence of a link between two users. However, the recent proposal of the Social-Attribute Network (SAN) [23] provides an overarching framework where

Fig. 1: Data organization for predictors in online social networks using the SAN.

both types of inference can be encapsulated as one generic inference challenge (i.e. predicting an edge), and also many of the formulations and methods to perform the inference are also generically presented. Note that in our alerts for privacy this provides a generic framework for how adversaries would attempt to discover the confidential information about a user. We also emphasize here that what a user considers confidential information may include the fact that they are connected to some other user in the social network.

The SAN is a graph model to predict links and infer attributes. The nodes $G(V)$ of the graph G are of two types (but it is not a bipartite graph): each of the n users is a node (so called *social nodes*), but also each of the m attribute-values is a node (so called *attribute nodes*). The connections of the social network are edges in the graph G between the corresponding social nodes; that is all social relationships are edges in the set $G(E)$ of edges in G . But, one has to explain actually in more detail the attribute nodes. As a running examples we again use a deck of poker playing cards. In this case, each suit (spades, hearts, diamonds, clubs) is a node, and so are other attributes (like being a *red card*, or a *black card*, being a picture card, or being a numeral). The framed section of Fig. 1 shows a section of the SAN where $(K, \spadesuit)$ and $(10, \heartsuit)$ are connected in the social network, and just two attributes (*colour* and *suit*). The SAN and the early discussion can now be interpreted together as follows. If user u wants to keep private a relationship to user v , then u wants to keep the attribute $a_p = \text{connected to } v$ difficult to predict. In terms of the SAN, we are predicting the link between user u and v . Our challenge is to identify what attributes (attributes and metrics) are the predictors of the links between the nodes u and v .

Once this data-model is built, the unsupervised-link prediction algorithms [28] and the supervised methods for link and attribute prediction can be formulated [23]. We just need some notation. First, $\Gamma_{s+}(u)$ is the set of all **social** neighbors connected to u ; this is the 1-hop neighborhood of user u in the social network, or u 's immediate friends. However, if u is an attribute node $\Gamma_{s+}(u)$ are all users that have this attribute. If u is a user node, $\Gamma_{a+}(u)$ is all **attribute** neighbors of user u ; that is the known attribute-value pairs of user u . If u is an

attribute-node $\Gamma_{a+}(u) = \emptyset$ always.

Now, $\Gamma_+(u) = \Gamma_{s+}(u) \cup \Gamma_{a+}(u)$. The fact that there is no universally accepted ratio of to what extent attributes (as opposed to connections) exercise more influence is generically modelled by allowing user-defined weights on G [28]; that is $w(v)$ is the weight of any node $v \in G$ and $w(v, v')$ is the weight of any edge $(v, v') \in G(E)$. Unless explicitly specified, all weights will be 1. Now we can formulate in *SAN* terms the two versions of the *Common Neighbour* (CN) statistic used as a predictor. The so called *conventional* CN only considers social neighbors.

$$m_{\text{CN}}(u, v) = \sum_{t \in \Gamma_{s+}(u) \cap \Gamma_{s+}(v)} w(t). \quad (1)$$

While CN-*SAN* includes attributes (i.e. common attributes even if u and v are not connected, the score adds up).

$$m_{\text{CN-SAN}}(u, v) = \sum_{t \in \Gamma_+(u) \cap \Gamma_+(v)} w(t). \quad (2)$$

The second predictor is called *Adamic-Adar* (AA), and the *conventional* statistic (for social nodes u and v) is defined as follows [29], [28].

$$m_{\text{AA}}(u, v) = \sum_{t \in \Gamma_{s+}(u) \cap \Gamma_{s+}(v)} \frac{w(t)}{\log |\Gamma_{s+}(t)|}. \quad (3)$$

Here, the importance of a common neighbour is scaled by the inverse of the social degree of that neighbour. There is a AA-*SAN* version.

$$m_{\text{AA-SAN}}(u, v) = \sum_{t \in \Gamma_+(u) \cap \Gamma_+(v)} \frac{w(t)}{\log |\Gamma_+(t)|}. \quad (4)$$

In the *SAN*, the *Adamic-Adar* statistic is also defined for a social node u and an attribute node a

$$m_{\text{AA}}(u, a) = \sum_{t \in \Gamma_{s+}(u) \cap \Gamma_{s+}(a)} \frac{w(t)}{\log |\Gamma_+(t)|}. \quad (5)$$

Note that there does not need to be an edge in the *SAN* between user u and attribute a to compute $m_{\text{AA}}(u, a)$; this will be important when discussing the prediction of this edge.

In general, for each user u and each attribute, we can add a column to D as $m_{\text{AA}}(u, a)$. Also, for each user u , we can create predictors which respect to social connection to another user v using the 4 statistics $m_{\text{CN}}(u, v)$, $m_{\text{CN-SAN}}(u, v)$, $m_{\text{AA}}(u, v)$, and $m_{\text{AA-SAN}}(u, v)$. Fig. 1 illustrates this with two tables. The first table shows that each attribute has now a new numeric column. The second table shows that for each two users, we have 4 additional columns. Again, in Fig. 1 (**K, ♠**) and (**10, ♥**) are socially connected. We have highlighted a cell in the tables of Fig. 1 shows that user (**10, ♥**) would like to keep the fact of being a heart confidential. All other attributes about the user can be the predictors, including the statistic $m_{\text{AA}}((\mathbf{10}, \heartsuit), \heartsuit)$.

A. The general method

Assume user c wants to keep an attribute or connection in the *SAN* confidential. Earlier, this fact had the form $a_p = v_c$ in the table D that represents all the data. With the *SAN*, the prediction of a nominal attribute is the prediction of an edge. Since all other attributes have already been converted to nominal attributes, the first step of our new generic method is the discretization of the attributes derived from the *SAN*'s scores. The approach we take is to use the Fayyad & Irani's MDL method [30] to convert numeric attributes to nominal attributes as this is the default in WEKA [22] which is the package we use for our implementation. Now the dataset D has columns that are all nominal attributes.

The method requires that user c provides a confidence value β for which the predictability of edge in the *SAN* should not be larger. A forest of trees is built. The *information gain* is the heuristic to choose the next node of the decision tree. The standard algorithm for decision trees is used, but again personalization applies. We build the tree for the classes $a_p = v_c$ versus $a_p \neq v_c$. Note this analysis depends on who is the user c wishing to conceal something and what attribute the user wishes to conceal. Moreover, every time a node is to be expanded, all attributes that have information gain larger than β which are at a depth less than 3 are kept in a queue to build alternate trees. Thus, the forest of trees consists of all decision trees to classify the classes $a_p = v_c$ versus $a_p \neq v_c$ using attributes with information gain larger than β on the 3 top levels of a tree.

From the forest of trees we extract classification rules. For each tree in the forest, a rule is the conjunction of attributes from the path of the tree to the leaf where the user c itself lands. That is, if $l_{a_p=v_c}$ is the label for the leaf on the path $x_{i_1} = a_{11}, x_{i_2} = a_{21}, \dots, x_{i_p} = a_{p1}$, then the logic rule $(x_{i_1} = a_{11}) \wedge (x_{i_2} = a_{21}) \wedge \dots \wedge (x_{i_p} = a_{p1}) \rightarrow a_p = v_c$ is the classification rule for this leaf. All the rules of all the paths for in the forest provides a list $R(b_u, \beta) = \{R_1, R_2, \dots, R_m\}$ of classification rules⁴. For each of these rules $R_j \in R(b_u, \beta)$, the following criteria evaluate the risk of a rule [8].

- The **support** s_j (the ratio of the number $|R_j|$ of records that fall into the leaf of the rule to the total $|T|$ of records in the data set); that is, $s_j = |R_j|/|T|$.
- The **confidence** c_j of the rule R_j is the ratio of the number $|R_j^+|$ of records correctly classified to the number $|R_j|$ of records in the leaf.
- The **safety** $-\log(c_j) - \log(s_j)$ is the number of bits needed by the adversary to disclose the confidential attribute pair after the rule R_j is applied.
- Alternatively, we can define the **sensitivity** S_j of a rule R_j as $S_j = c_j + s_j = |R_j|/|T| + |R_j^+|/|R_j|$.

Now we can rank [8] the attributes the user should consider most dangerous by two metrics.

- The cumulative sensitivity CUM_SENSITIVITY of an attribute x_i is the sum of the sensitivities of the rules the

⁴Because the $\wedge$ (Boolean AND operator) commutes, a rule may appear in more than one of the trees of the forest, but only one is kept.

ID	binary attribute
74	<i>bangalore institute of technology</i>
61	<i>aspiag service srl</i>
27	<i>fh aachen</i>
542	<i>\u6578\u5b78\u7cfb</i>
322	<i>1180</i>
311	<i>general studies</i>
564	<i>post graduate in marketing</i>

TABLE I: The 7 most frequent attributes in infer-attrib/SEP4.txt.

attribute is an antecedent for. Namely

$$\begin{aligned} \text{CUM_SENSITIVITY}(x_i) \\ = \sum_{R_j \in R(b_u, \beta) \wedge x_i \text{ is an antecedent of } R_j} S_j. \end{aligned}$$

- **TOTAL_COUNT** counts the number of times an attribute x_i ($i \in \{1, \dots, d\}$) appears among the sensitive rules.

$$\begin{aligned} \text{TOTAL_COUNT}(x_i) \\ = \|\{R_j \in R(b_u, \beta) \mid x_i \text{ is an antecedent in } R_j\}\|. \end{aligned}$$

B. An illustration

For the purpose of illustration of the above statistics we will create an artificial *SAN* based on the famous Southern Women dataset [31], [32]⁵. There are 18 women who may or may not have attended one of 14 events. However, we will create artificial attributes using a standard deck of 52 poker-playing cards. A social node will be a $(\text{person_name}, \text{numeral_event})$. The first attribute of a $(\text{person_name}, \text{numeral_event})$ node is the *suit*. This would be chosen as the highest count for the size of the intersection of the set of letters in suit name and the set of letters in the name of the person. The rank will be assigned as *ace* for the first and 14th event, while $\langle 2, 3, \dots, 10 \rangle$ for the second to 10th event. The 11th event will get J, the 12th will get Q, the 13th a K and the 14th will get *ace* again.

The social edges in our artificial *SAN* will be derived from the following rules.

- Every two nodes with the same *person_name* have a link.
- Every two nodes with the same *numeral_event* have a link.

The attributes of each social node $(\text{person_name}, \text{numeral_event})$ will be completely derived from their card assignment and would be

- 1) *card_color* with possible values $\{\text{red}, \text{black}\}$,
- 2) *card_rank* with possible values $\{2, 3, \dots, 10, J, Q, K, A\}$,
- 3) *is_face_card* with attributes $\{\text{picture_card}, \text{numeral}\}$,
- 4) *had_rounded* with value *yes* for $\spadesuit$, $\clubsuit$ and $\heartsuit$, but no for $\diamondsuit$.

Naturally, in this artificially generated *SAN* there will be many dependencies between attributes and between the attributes of social nodes. Nevertheless, because the data set is so small, results would be easy to reproduce⁶.

In our first experiment we made the suit the confidential attribute, and we investigated the predictability over 7 different

ID	degree
$h_0=2727$	45
$h_1=2038$	53
$h_2=16$	36
$h_3=4292$	37
$h_4=971$	40

TABLE II: Nodes with largest number of connections in infer-attrib/SEP4.txt

users u with this attribute equal to *hearts*. Remarkably, the metric $m_{AA}(u, \text{hearts})$ is a strong predictor, and both (the **CUM_SENSITIVITY** and the **TOTAL_COUNT**) identify this metric as a very high predictor. It is very pleasing to see that we can find when these social-network metrics can be identified as high predictors and the users can be warned about the risk of their use to predict an attribute the user desires to keep confidential (i.e. the adversary does not have access to the attribute value pair *suit=hearts*, just the relationships and the data of others; i.e. the *SAN*).

The **CUM_SENSITIVITY** and the **TOTAL_COUNT** identify also another 6 attributes of equal high risk. This again demonstrates the merits of our approach, we require that each user be warned of all the potential predictors. As opposed to the adversary, which requires just one accurate model.

This data set is small, so we also enumerated all possible sets of 3 predictors (metrics or attributes). Since the **CUM_SENSITIVITY** and the **TOTAL_COUNT** identify also another 6 attributes of equal high risk, we checked how often these top predictors include a predictor in a highly predictive 3-set. We evaluate the goodness of the recommendation made by the **CUM_SENSITIVITY** or the **TOTAL_COUNT** as a solution to the Hitting-Set problem for this instance. There are 603 3-sets of predictors with information gain above 0.99 for each of the 7 different user vectors which attempt to conceal the attribute pair *suit=hearts*. In all 7 cases, the **CUM_SENSITIVITY** and the **TOTAL_COUNT** hit all 603 of them.

C. A real (larger) data set

In our next set of experiments we used the freely distributed dataset infer-attrib/SEP4.txt previously used [23] to infer attributes of users and links between users in a social network. The data set is about applicants and their background in areas of information technology. This data set has over 5,000 users and there are around 10,000 attributes possibly present for a user; although on average, each user has about 4 attributes. Data cleansing of this data set consists of filtering by two key attributes: education and employment [23]. Each one of these attributes can have many values, thus in the corresponding *SAN* we still have many attribute-nodes.

1) *When the information-item (that user u designates as confidential) is an attribute.*: We demonstrate now how our method identifies risk when a user would like to keep confidential an attribute-value pair regarding an attribute. We identified 7 attribute-nodes in the corresponding *SAN* whose total count is 92 or more in the complete dataset. These 7 attributes are presented in Table I. We will experiment with these. For each user, and each attribute, we took the position that the user wants to keep confidential whether they have or not the attribute. Then, we use the rest of the data set to generate

⁵Downloaded from toreopsahl.com/datasets/.

⁶We can make the dataset available upon request.

Risk Ranking (one user)					
ID	First	Second	Third	Fourth	Fifth
74	$m_{AA}(u, 546)$	$m_{AA}(u, 27)$	$m_{AA}(u, 74)$	542	$m_{AA}(u, 322)$
61	542	$m_{AA}(u, 322)$	$m_{AA}(u, 74)$	$m_{AA}(u, 311)$	546
27	$m_{AA}(u, 546)$	$m_{AA}(u, 322)$	$m_{AA}(u, 61)$	322	$m_{AA}(u, 74)$
542	$m_{AA}(u, 322)$	61	$m_{AA}(u, 546)$	$m_{AA}(u, 27)$	$m_{AA}(u, 74)$
322	546	74	311	$m_{AA}(u, 61)$	$m_{AA}(u, 322)$
311	546	322	74	27	$m_{AA}(u, 542)$
546	74	322	$m_{AA}(u, 546)$	311	542

TABLE III: SAN predictors ranked by CUM_SENSITIVITY.

Risk Ranking (one user)					
ID	First	Second	Third	Fourth	Fifth
74	542	546	$m_{AA}(u, 546)$	$m_{AA}(u, 74)$	$m_{AA}(u, 27)$
61	542	$m_{AA}(u, 322)$	546	$m_{AA}(u, 74)$	$m_{AA}(u, 27)$
27	$m_{AA}(u, 542)$	$m_{AA}(u, 322)$	546	$m_{AA}(u, 74)$	322
542	$m_{AA}(u, 322)$	67	$m_{AA}(u, 322)$	546	$m_{AA}(u, 74)$
322	546	74	311	542	$m_{AA}(u, 542)$
311	546	74	322	61	$m_{AA}(u, 542)$
546	74	542	322	$m_{AA}(u, 546)$	311

TABLE IV: SAN predictors ranked by TOTAL_COUNT.

predictors. In table format, the rows are the users, and the first 7 columns are binary attributes indicating whether the user has or not the attribute. We add 7 columns more to this table by adding the value $m_{AA}(u, a_i)$ where a_i is the i -th ID of the corresponding frequent attributes. We note that the metric values were normalized between 0.0 and 1.0 and then discretised into 10 possible categories in steps of 0.10, giving $\{0.00-0.10, 0.11-0.20, \dots, 0.91-1.00\}$.

Table III shows results for a randomly chosen user u using CUM_SENSITIVITY while Table IV shows results for the TOTAL_COUNT. From these tables we see that attribute-pairs can be at risk by metrics calculated out of the social-network data quite effectively, which probably justifies the interest of data analysts for online social networks. Our approach enables us to alert users appropriately, since for some attributes (like 311 and 546) other attributes are the most informative predictors; while for others (like 71, 27 and 542), the metric $m_{AA}(u, \cdot)$ of some other attributes is among the highly informative predictors. Note that in general the top highly ranked predictors coincide between CUM_SENSITIVITY and TOTAL_COUNT, for the same confidential attribute, although the order may change.

The rankings in Table III and Table IV for the risky predictors for an attribute change if we have a different user. There are 21 different users (by the 7 attributes and 7 values of the metric $m_{AA}(u, \cdot)$). We ran the method on all the different users, and the personalization aspect of our approach results in a different assessment of the most risky predictors for each individual user when they want to keep confidential one of the 7 attributes. One could also obtain global recommendations by awarding 5 points every time a predictor ranks first and decrease the award with the ranking position, giving only one point to when a predictor ranks 5th. Adding all points across all users gives an assessment of the risk of a predictors across the users. Table V and Table VI summarize these results.

The AA-SAN is an example of a *topological feature* that can be used to predict whether a user u has attribute a [23] (even if such an edge is missing in the SAN for the data). In

Risk Ranking (21 users)					
ID	First	Second	Third	Fourth	Fifth
74	546	$m_{AA}(u, 546)$	27	542	$m_{AA}(u, 27)$
61	542	$m_{AA}(u, 322)$	546	$m_{AA}(u, 74)$	311
27	546	$m_{AA}(u, 546)$	$m_{AA}(u, 322)$	$m_{AA}(u, 61)$	$m_{AA}(u, 74)$
542	$m_{AA}(u, 322)$	61	546	$m_{AA}(u, 27)$	$m_{AA}(u, 74)$
322	546	74	$m_{AA}(u, 61)$	311	$m_{AA}(u, 322)$
311	27	546	322	$m_{AA}(u, 61)$	$m_{AA}(u, 74)$
546	322	$m_{AA}(u, 546)$	74	311	542

TABLE V: SAN predictors ranked by CUM_SENSITIVITY.

Risk Ranking (21 users)					
ID	First	Second	Third	Fourth	Fifth
74	542	27	546	$m_{AA}(u, 546)$	322
61	542	546	$m_{AA}(u, 7542)$	$m_{AA}(u, 74)$	$m_{AA}(u, 322)$
27	$m_{AA}(u, 542)$	546	542	$m_{AA}(u, 74)$	$m_{AA}(u, 322)$
542	$m_{AA}(u, 322)$	546	61	74	$m_{AA}(u, 542)$
322	546	74	542	27	311
311	27	546	322	74	$m_{AA}(u, 542)$
546	74	542	322	311	$m_{AA}(u, 546)$

TABLE VI: SAN predictors ranked by TOTAL_COUNT.

the literature a Support Vector Machine (SVM) on a SAN from the training set is used in prediction to whether u has attribute a or not. Typically no more than nine topological features are used [23]. This in a sense summarizes the supervised approaches the adversary may take to infer the confidential attribute. Here we use them to identify the risk; and in that sense also provide users with an alert or a warning. Also, we use them as ground truth to the findings in Table V and Table VI. Thus, we learn (for each of the 7 attributes) a decision tree using the `j48` algorithm from WEKA [22] and also the `SMO SVM` learner from WEKA [22] to emulate the adversary's inference mechanism in the literature [23]. While these learning algorithms do not provide the depth of our method (up to 5 risky predictors), they indicate what globally are the strongest predictors for the seven attributes.

In the case of SVM the most highly weighted predictor is 546 for attribute 74 and 61, 27, 542, 322 and 311. For 546, the strongest predictor is 74. In the case of C4.5, the root attribute used in the decision tree (and thus the strongest predictor) was 546 for 74, 27, 542 and 322. Decision trees indicate no predictor attribute for 61 and for attribute 311. While the predictor for 546 is also 74. One can look in levels further down the decision tree or the ranking in the SVM, however, this approach is not as informative of the risk represented by the metrics computed out of the social networks. There is remarkable correspondence on the predictors between our general method and the predictors detected with isolated classifiers.

2) *When the information-item (that user u designates as confidential) is a connection.*: To illustrate our approach in this case we also use the same data set, but we identify the top 5 users by degree; that is, we obtain the 5 users with the highest number of neighbors. We will designate this set of users as $H = \{h_0, h_1, h_2, h_3, h_4\}$ (see Table II). Then, for each other user u in the dataset, we investigate the methods when u wants to hide whether or not u has a connection to each of those highly popular users. In table format, rows correspond to each other user u and the first five columns indicate if u has h_i as neighbour (YES) or not (NO). We can now calculate the metrics $m_{CN}(u, h_i)$, $m_{CN-SAN}(u, h_i)$, $m_{AA}(u, h_i)$, and

$m_{AA-SAN}(u, h_i)$, for $i \in \{0, \dots, 4\}$; resulting in 20 additional columns. The metric values are discretised into 10 possible categories in steps of 0.10, giving $\{0.00-0.10, 0.11-0.20, \dots, 0.91-1.00\}$. We investigate our method to identify which other of the 24 columns (predictors) constitute high risk predictors for each of the first 5 columns. Table VII and Table VIII show the rankings of our heuristic general methods for highlighting the predictors of a connection. These results are for a random user u , and for example, the first row of Table VII shows that if user u would hide the connection to user h_0 , the metrics $m_{AA}(u, h_0)$ and $m_{AA-SAN}(u, h_0)$ would remain high predictors to infer such connection. This would give a warning to user u that hiding attributes in common and connections in common with h_0 is also necessary to inhibit this predictor.

We also processed this data with the $j48$ and the *SVM* provided by WEKA [22], which enables us to rank the potential predictors, although not as detailed as our general method (again, $j48$ would only give some attribute at the root of the decision tree, because the interest there is just to obtain a strong classifier). We observe significant correspondence in our results. For example, for predicting the friendships with h_4 , the *SVM* model uses as highest predictors $m_{CN-SAN}(u, h_4)$, $m_{AA}(u, h_4)$, and $m_{AA-SAN}(u, h_4)$. The correspondence is not that strong with $j48$, but $m_{AA}(u, h_4)$ is at the top of the tree.

We emphasize again that results are personalized, that is, Table VII and Table VIII would change if we consider another user. That is, our methods are personalized and the recommendations of what are the highest predictors to what the user designates as confidential depend on the attributes (and connections) the user leaves public and their values. Nevertheless, it is possible to create an aggregation over all users (see Table IX and Table X). Moreover, when one creates this aggregate, one observes more agreement with the attributes ranked high by individual classifiers like the $j48$ decision tree algorithm and the *SVM* provided by WEKA [22]. In general, we observe that our generic method not only identifies attributes at the very top of the tree, but also attributes that appear several times lower down in the decision tree. So these predictors identified with our heuristics are still significant and risk a confidentiality breach, but are used more specifically by the decision tree. That is, a predictor appearing low in a decision tree does not imply it is safe to disclose it. Our methods actually identify these potential risky predictors.

V. CONCLUSIONS

In this paper we have presented a novel approach that differs from current work on privacy-preserving data mining in that users and attributes are not treated homogeneously. Each user decides what attributes they consider confidential, but then, we alert them of what metrics (that are derived from their public attributes or public associations) and what other of their public attributes enable accurate prediction of their confidential values. Such information is personalized 1) to the user and 2) to the attribute-value pair or connection the user wants to keep confidential. Thus, users can obfuscate, perturb, delete, and/or generalize the values of attributes or of the associations that enable the building of predictors for classifiers. We have illustrated the methods that enable us to investigate the sets of predictors and prioritize them as per their predictive power: that is, the potential they have to raise

the information gain of an adversary that knows the data of the other users in the network plus the public information of this user. We have shown that our heuristic search makes it possible to identify not just one predictor, but to approximate closely the smallest set of attributes and metrics the user needs to act upon in order to affect all sets of high informants.

REFERENCES

- [1] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," *13th Conf. on USENIX Security Symposium - V. 13*, ser. SSYM'04. Berkeley, CA, USENIX Assoc., pp. 21–21.
- [2] G. Gürses and B. Berendt, "The social web and privacy: Practices, reciprocity and conflict detection in social networks," in *Privacy-Aware Knowledge Discovery, Novel Applications and New Techniques*. CRC Press, 2010, pp. 395–429.
- [3] R. Wishart, R. Corapi, A. Madhavapeddy, and M. Sloman, "Privacy butler: A personal privacy rights manager for online presence," in *8th Annual IEEE Int. Conf. on Pervasive Computing and Communications, PerCom*. Mannheim: IEEE, 2010, pp. 672–677.
- [4] R. Compton, D. Jurgens, and D. Allen, "Geotagging one hundred million twitter accounts with total variation minimization," in *2014 IEEE Int. Conf. on Big Data 2014*, Washington, DC, USA: IEEE, 2014, pp. 393–401.
- [5] J. M. Such, A. García-Fornes, A. Espinosa, and J. Bellver, "Magentix2: A privacy-enhancing agent platform," *Eng. Appl. of AI*, vol. 26, no. 1, pp. 96–109, 2013.
- [6] N. Ramakrishnan, B. Keller, B. J. Mirza, A. Grama, and G. Karypis, "Privacy risks in recommender systems," *IEEE Internet Computing*, vol. 5, no. 6, pp. 54–62, 2001.
- [7] R. Baraglia, C. Lucchese, R. Orlando, S. Perego, and F. Silvestri, "Preserving privacy in web recommender systems," in *Privacy-Aware Knowledge Discovery, Novel Applications and New Techniques*. CRC Press, 2010, pp. 369–391.
- [8] V. Estivill-Castro, P. Hough, and M. Islam, "Empowering users of social networks to assess their privacy risks," in *2014 IEEE Int. Conf. on Big Data*, Washington, DC, USA: IEEE, 2014, pp. 644–649.
- [9] J. Vaidya, Y. Zhu, and C. W. Clifton, *Privacy Preserving Data Mining*, ser. Advances in Information Security. Springer, 2006, vol. 19.
- [10] C. C. Aggarwal and P. S. Yu, Eds., *Privacy-Preserving Data Mining - Models and Algorithms*, ser. Advances in Database Systems. Springer, 2008, vol. 34.
- [11] F. Giannotti and D. Pedreschi, Eds., *Mobility, Data Mining and Privacy - Geographic Knowledge Discovery*. Springer, 2008.
- [12] F. Bonchi and E. Ferrari, Eds., *Privacy-Aware Knowledge Discovery - Novel Applications and techniques*, ser. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, 2010.
- [13] E. Zheleva, E. Terzi, and L. Getoor, *Privacy in Social Networks*, ser. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool, 2012.
- [14] B. Könings, D. Piendl, F. Schaub, and M. Weber, "PrivacyJudge: Effective privacy controls for online published information," in *Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third Int. Conf. on Social Computing (SocialCom)*, Boston, MA, USA, 2011, pp. 935–941.
- [15] B. Carminati, E. Ferrari, M. Kantarcioglu, and B. Thuraisingham, "Privacy protection of personal data in social networks," in *Privacy-Aware Knowledge Discovery, Novel Applications and New Techniques*. CRC Press, 2010, pp. 433–458.
- [16] M. Hay, G. Miklau, and D. Jensen, "Analyzing private network data," in *Privacy-Aware Knowledge Discovery, Novel Applications and New Techniques*. CRC Press, 2010, pp. 459–498.
- [17] S. Lederer, H. J. I. A. K. Dey, and J. A. Landay, "Personal privacy through understanding and action: five pitfalls for designers," *Personal and Ubiquitous Computing*, vol. 8, no. 6, pp. 440–454, 2004.
- [18] J. Domingo-Ferrer, A. Viejo, F. Sebé, and U. González-Nicolás, "Privacy homomorphisms for social networks with private relationships," *Computer Networks*, vol. 52, no. 15, pp. 3007–3016, 2008.

Risk Ranking (one user)					
ID	First	Second	Third	Fourth	Fifth
h_0	$m_{AA}(u, h_0)$	$m_{AA-SAN}(u, h_0)$	h_4	$m_{AA}(u, h_4)$	$m_{AA-SAN}(u, h_4)$
h_1	$m_{AA}(u, h_1)$	$m_{AA-SAN}(u, h_1)$	$m_{AA}(u, h_0)$	$m_{AA-SAN}(u, h_0)$	$m_{CN}(u, h_0)$
h_2	$m_{AA}(u, h_3)$	$m_{AA-SAN}(u, h_3)$	h_0	h_4	$m_{AA}(u, h_1)$
h_3	$m_{CN-SAN}(u, h_1)$	$m_{CN-SAN}(u, h_3)$	$m_{CN-SAN}(u, h_2)$	h_0	$m_{CN-SAN}(u, h_4)$
h_4	$m_{AA}(u, h_4)$	$m_{AA-SAN}(u, h_4)$	$m_{CN}(u, h_4)$	$m_{CN-SAN}(u, h_4)$	h_0

TABLE VII: SAN predictors of connections ranked by CUM_SENSITIVITY.

Risk Ranking (one user)					
ID	First	Second	Third	Fourth	Fifth
h_0	h_4	$m_{CN}(u, h_0)$	$m_{CN-SAN}(u, h_2)$	h_2	$m_{CN-SAN}(u, h_0)$
h_1	$m_{AA}(u, h_1)$	$m_{AA-SAN}(u, h_1)$	$m_{CN-SAN}(u, h_4)$	$m_{CN-SAN}(u, h_2)$	$m_{CN}(u, h_4)$
h_2	h_3	$m_{CN-SAN}(u, h_4)$	h_0	$m_{CN-SAN}(u, h_2)$	h_4
h_3	$m_{CN}(u, h_3)$	$m_{CN-SAN}(u, h_3)$	$m_{CN-SAN}(u, h_2)$	h_0	$m_{CN-SAN}(u, h_1)$
h_4	$m_{AA}(u, h_4)$	$m_{AA-SAN}(u, h_4)$	$m_{CN-SAN}(u, h_1)$	$m_{AA}(u, h_1)$	$m_{AA-SAN}(u, h_1)$

TABLE VIII: SAN predictors of connections ranked by TOTAL_COUNT.

Risk Ranking (all users)					
ID	First	Second	Third	Fourth	Fifth
h_0	h_4	$m_{CN}(u, h_0)$	$m_{AA}(u, h_0)$	$m_{AA-SAN}(u, h_4)$	$m_{AA}(u, h_1)$
h_1	$m_{AA}(u, h_1)$	$m_{AA}(u, h_4)$	$m_{AA}(u, h_1)$	$m_{AA-SAN}(u, h_2)$	$m_{AA}(u, h_2)$
h_2	$m_{AA-SAN}(u, h_2)$	$m_{AA}(u, h_2)$	$m_{AA-SAN}(u, h_4)$	$m_{AA-SAN}(u, h_1)$	$m_{CN}(u, h_2)$
h_3	h_2	$m_{CN-SAN}(u, h_3)$	h_0	$m_{CN}(u, h_1)$	$m_{AA}(u, h_1)$
h_4	$m_{AA}(u, h_4)$	$m_{AA-SAN}(u, h_4)$	$m_{CN}(u, h_4)$	$m_{CN-SAN}(u, h_4)$	$m_{AA}(u, h_2)$

TABLE IX: SAN predictors of connections ranked by CUM_SENSITIVITY.

Risk Ranking (all users)					
ID	First	Second	Third	Fourth	Fifth
h_0	h_4	$m_{CN}(u, h_0)$	$m_{AA}(u, h_2)$	$m_{AA-SAN}(u, h_1)$	$m_{AA-SAN}(u, h_4)$
h_1	$m_{AA}(u, h_1)$	$m_{AA-SAN}(u, h_1)$	$m_{AA}(u, h_4)$	$m_{AA-SAN}(u, h_2)$	$m_{AA-SAN}(u, h_4)$
h_2	$m_{AA-SAN}(u, h_2)$	$m_{CN}(u, h_2)$	$m_{AA}(u, h_2)$	$m_{AA}(u, h_1)$	$m_{CN-SAN}(u, h_4)$
h_3	h_2	h_0	$m_{CN-SAN}(u, h_3)$	$m_{CN}(u, h_3)$	$m_{AA-SAN}(u, h_1)$
h_4	$m_{CN}(u, h_4)$	$m_{CN-SAN}(u, h_4)$	$m_{AA-SAN}(u, h_4)$	$m_{AA}(u, h_4)$	$m_{AA-SAN}(u, h_2)$

TABLE X: SAN predictors of connections ranked by TOTAL_COUNT.

- [19] J. Domingo-Ferrer, "A public-key protocol for social networks with private relationships," in *Modeling Decisions for Artificial Intelligence, 4th Int. Conf., MDAI*, ser. LNCS, vol. 4617. Kitakyushu, Japan: Springer, 2007, pp. 373–379.
- [20] K. Liu and E. Terzi, "A framework for computing the privacy scores of users in online social networks," *ACM Trans. Knowl. Discov. Data*, vol. 5, no. 1, pp. 6:1–6:30, 2010.
- [21] Y. Jafer, S. Matwin, and M. Sokolova, "Privacy-aware filter-based feature selection," in *2014 IEEE Int. Conf. on Big Data, Big Data 2014*, J. Lin, et al. Eds. Washington, DC, USA: IEEE, 2014, pp. 1–5.
- [22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.
- [23] N. Z. Gong, A. Talwalkar, L. W. Mackey, L. Huang, E. C. R. Shin, E. Stefanov, E. Shi, and D. Song, "Joint link prediction and attribute inference using a social-attribute network," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 2, p. 27, 2014.
- [24] T. La Fond and J. Neville, "Randomization tests for distinguishing social influence and homophily effects," in *19th Int. Conf. on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 601–610.
- [25] G. Kossinets and D. Watts, "Empirical analysis of an evolving social network," *Science*, vol. 311, no. 5757, pp. 88–90, 2006.
- [26] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins, "Structure and evolution of blogspace," *Commun. ACM*, vol. 47, pp. 35–39, 2004.
- [27] M. Kim and J. Leskovec, "Modeling social networks with node attributes using the multiplicative attribute graph model," in *In UAI*, 2011.
- [28] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," in *Twelfth Int. Conf. on Information and Knowledge Management*, ser. CIKM '03. ACM, 2003, pp. 556–559.
- [29] L. Adamic and E. Adar, "Friends and neighbors on the web," *Social Networks*, vol. 25, pp. 211–230, 2001.
- [30] U. M. Fayyad and K. B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," in *13th Int. Joint Conf. on Artificial Intelligence*, R. Bajcsy, Ed. Chambéry, France: Morgan Kaufmann, 3 1993, pp. 1022–1029.
- [31] A. Davis, B. B. Gardner, and M. R. Gardner, *Deep South: a social anthropological study of caste and class*. Chicago, IL, USA: University of Chicago Press, 1941.
- [32] L. Freeman, "Finding social groups: A meta-analysis of the southern women data," in *Dynamic Social Network Modeling and Analysis. The National Academies*, R. Breiger, K. Carley, and P. Pattison, Eds. Washington, DC.: National Academies Press, 2003, pp. 39–97.
- [33] M. J. J. and J. Leskovec, "Learning to discover social circles in ego networks," in *Advances in Neural Information Processing Systems 25: 26th Annual Conf. on Neural Information Processing Systems 2012*, P. Bartlett, et al. Eds., Lake Tahoe, Nevada, USA, 2012, pp. 548–556.
- [34] A. Kokkos and T. Tzouramanis, "A robust gender inference model for online social networks and its application to linkedin and twitter," *First Monday*, vol. 19, no. 9, 2014.
- [35] D. Chakrabarti, Y. Zhan, and C. Faloutsos, "R-MAT: A recursive model for graph mining," in *Fourth SIAM Int. Conf. on Data Mining*, M. W. Berry, et al. Eds. Lake Buena Vista, Florida, USA: SIAM, 2004, pp. 442–446.