Can on-line social network users trust that what they designated as confidential data remains so?

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—Internet users in general and on-line social networks users in particular are becoming more savvy about masking data they consider private. However, some of this masked data may be inferable from other data the user has not masked. Furthermore, even if a user masks all its data, it may still be inferable from the unmasked data of its friends, due to affinities in likes and personal attributes. In contrast to the conventional data mining approach, in which a model is built for all users, we build a rule set which is individualized for each user. In this paper we propose a novel rule induction approach (that incorporates predictive metrics) which enable a user to evaluate the potential risk incurred by unmasked attributes, friends’ attributes and also the risk of befriending new users. We find that all of these risks are quantifiable and a risk ranking of attributes and predictive metrics) which enables a user to evaluate the potential risk incurred by unmasked attributes, friends’ attributes and the risk of befriending new users. We find that all of these risks are quantifiable and a risk ranking of attributes and also the risk of befriending new users. We find that all of these risks are quantifiable and a risk ranking of attributes and also the risk of befriending new users. We find that all of these risks are quantifiable and a risk ranking of attributes and also the risk of befriending new users.

I. INTRODUCTION

Big data technologies have fueled more the debate about privacy. Academics argue on both sides of the issue and governments, legislators and official agencies struggle to find the adequate protection mechanisms and the proper balance. Informed decision-making requires data, and that large volumes of data are now available to be analyzed in order to produce a competitive edge. This is even more appealing given the exponential growth of on-line social networks that have resulted in massive collections of personal information. On-line services, even as simple as search engines, rely heavily on user profiling and data analytics to provide personalization. Such personalization benefits users who receive tailored (and thus, improved) services.

Privacy is now considered to involve both the control of access to data and the practice of such control. In this paper we discuss mechanisms that enable users to evaluate the trade-off in privacy with respect to disclosure. Lack of trust on the secondary use of personal data leads to suspicion from users, who engage in privacy-active behavior (that is, behavior of the user to conceal information). Also, perceived privacy-risk had a strong negative influence on the extent to which respondents participated in online subscription and purchasing. Thus, it is in the interest of web-service providers to protect privacy and increase the level of trust of their users. In this paper, we provide techniques so users can investigate which of their public attributes (including attributes, connections and/or metrics computed out of these) result in informative predictors for classifiers (learned with data mining and machine learning approaches) to disclose what users regard as confidential.

II. RELATED WORK

Naturally, personal data is of interest to the user the data is about. And perhaps the most common principle on what privacy actually represents is that users should at least be able to find out what data is publicly available about them. As a result of this, the European Commission recently forced Google to review its practices regarding management of indexing and storage of personal data. Moreover, on April 15th 2015, Carl Bildt, chair of the Global Commission on Internet Governance (GCIG) called for a new global social contract to protect digital privacy and security with his presentation “Toward a Social Compact for Digital Privacy and Security” at a conference in The Hague. His first key point is that privacy and personal data protection are a fundamental human right, his 4th is that safeguarding online data and consumer awareness is central, while the 5th point deals with the issues with respect to big data and trust.

The core of the dilemma derives from benefits by data analytics to the consumer with tailored services, distribution efficiency, and market identification that reduces the cost of goods and services. Data and its analysis informs policy makers resulting in efficient expenditure of public funds. In fact, data analytics is considered now an essential tool in marketing and retail, since such technology can solve what was considered a very challenging marketing problem. Namely, to “anticipate a major event in a person’s life when their shopping loyalties may change”. The implications for privacy were rapidly apparent when a girl’s pregnancy was predicted before the individual had disclosed the fact to any family members. Nevertheless, such technology is believed to be behind Target’s massively improved revenue from $US 44 billion in 2002 to $USD 67 billion in 2010.

Personalization and user profiling enables better query results, but it is well known that query log analysis has implications for privacy. Your regular use of a search engine
rapidly narrows your age and gender, which enables filtering of query results; and you would find results more relevant. Such service is appreciated by users who get the impression the search engine is better and fulfills their searching needs.

A social network service consists of a representation of each user (often a profile), his or her social links, and a variety of additional services. A user profile includes information-items of several types. Attribute data (gender, religion, marriage status, etc) are direct properties, while usage data (how often you visit, what you click, etc) is derived information, but linkage data (connections to other users and the contact frequency) may be derived from e-mails or on-line social network links. 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. In 2013, using 58,000 volunteers from Facebook in the US, it was demonstrated that Facebook Likes can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religions and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender [9]. It takes just 300 of your likes to create a model that predicts your gender with above 82% accuracy. Distinguishing between Republicans or Democrats can be done with precision above 85% and race (between Africa-American or Caucasian) with higher than 95% precision [9].

In most of the discussion that follows we will use confidential attribute, as if a user’s wishes were to keep their gender, religion or marriage status confidential; but the point of this paper is that any other property or statement about the other types of information-items in an on-line social network regarding themselves (whether they are connected to another user, for example) could be what the user feels is confidential. The explanation that follows shows that our methods work in an analogous manner in those cases.

III. Method

We will consider only nominal (binary) attributes. There is a technical reason for this that will be apparent when we consider the SAN [11] model for the data; however, there is also a natural reason, and we will argue that considering binary attributes represents no limitation. What the user wants to keep private can be stated as a Boolean predicate. For example,

1) is the gender female/male?
2) is the language spoken among romance languages?
3) is the salary in the range [$50,000, $100,000]?
4) is their shoe size above the Australian average?
5) is user u in regular daily contact with user v?

That is, what users want to conceal is the veracity of such an statement. In some cases, the statement is very likely to be true, or just impossible to obfuscate given the vast amount of data already available about one person. For example, “is the country of residence France” would be very difficult not to disclose for most French people; however, their home address within 100m is probably something most people would prefer not to be determined.

Regardless of the topic or issue we will make the assumption that eventually there is a range for which each user would prefer that the veracity of whether their value is inside or outside such range is almost impossible to establish. In what follows this will be presented in terms of Boolean attributes. Thus numerical attributes, or multi-dimensional attributes (like locations) are to be converted to Boolean attributes. Similarly, nominal values are converted to binary values by an appropriate codification. This is analogous to the NominalToBinary filter of WEKA [12] and as we already alluded to the existence of an edge to an attribute in the SAN [11].

Conceptually, adversaries build classifiers out of the data of all other users and the public data of the current user u, in order to determine whether the Boolean and confidential attribute attrj is true or false. Thus, we assume the entire data set of public-attributes as specified by each users is available to the adversary. The first format we consider for the data is as the common training set in the machine learning literature. Under such model, we consider a data set D where each row corresponds to a user and each column ai corresponds to a Boolean attribute. Naturally, the data is available to the Web-provider. For the moment, the attributes are the result of the on-line activity of each user. However, we wish to assist an individual user u. User u indicates that some columns (say attrj) are confidential, this makes them inaccessible from public view. However, our algorithms will also tip the user of another column attrv, the user is prepared to disclose, unknowing that column attrv is a good predictor of one of the confidential columns.

The earlier work [13] provided warnings about disclosing the value of predictor attribute attrv, for user u using the expected information gain Ig(D, attrv) = \( H(D) - H(D|\text{attr}_v) \); where H denotes the information entropy, which is a measure of uncertainty. The expected information gain measures the change in information entropy from a prior state of knowledge H(D) to a state H(D|\text{attr}_v) where some information \text{attr}_v is given. However, it is important that the information gain be personalized. First, in that it is the user who chooses the attribute (the column) that is determined as confidential. Second, it is not important if attrj is hard to predict in general, it is whether from u’s perspective, the value of attrj in u’s case is easy to predict once u’s public attributes are known.

Moreover, the challenge also has an optimization flavor. We are to minimize the public attributes that require obfuscation. That is, for a user u, we ought to find the smallest set S(u) of columns (attributes), so that any model/classifier/predictor for the truth value of attrj whose accuracy is greater than threshold \( \beta \), uses at least one column in S(u). Then, by obfuscating u’s values in S(u), the confidential attribute-value pair of u becomes hard to predict. Thus, the aim is to find the fewest public-attributes to perturb/hide to reduce the accuracy of predictors of your confidential attribute.

We emphasize here the trade-off is personalized. Again, first, there is a trade-off in what the user would like to reveal. Also, what the user desires to keep public may actually need to be obfuscated (otherwise the confidential information can be predicted and compromised). Second, that although the attribute that u considers confidential may be hard to predict in general, it may be the case that in the light of the social-network data for all other users, and u’s public attributes and connections, the confidential value of u is easy to predict. For example, there have been attempts to develop
systems to predict a person’s profession from social-network data or the images the person posts. In general it is hard to achieve precision above 30% across all professions [14]. Some professions seem hard to predict (like teacher barely reaches 15% accuracy across all methods), but others (like soldier with an accuracy above 60% with almost every method). Thus, people who are soldiers and desire to keep that confidential data or the images the person posts. In general it is hard to predict precision above 30% across all professions [14].

A. From tabular model of the data to the SAN model

Inference from the data in an on-line social network is based on principles like social influence [15], which suggest that social neighbors connected to user $u$ is the set of all social neighborhoods, or $u$’s immediate friends. However, if $u$ is an attribute node $\Gamma_{u,v}(u)$ are all users that have this attribute. If $u$ is a user node, $\Gamma_{u,v}(u)$ is all attribute neighbors of user $u$; that is the known attribute-value pairs of user $u$. If $u$ is an attribute-node $\Gamma_{a,v}(u) = \emptyset$ always.

Now, $\Gamma_{u,v}(u) = \Gamma_{s,v}(u) \cup \Gamma_{a,v}(u)$. The fact that there is no universally accepted ratio of to what extent attributes (as opposed to connections) exercise more influence is generic modelled by allowing user-defined weights on $G$ [19]; 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.

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

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

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

The second predictor is called Adamic-Adar (AA), and the conventional statistic (for social nodes $u$ and $v$) is defined as follows [20, 19].

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

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

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

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

Note that there does not need to be an edge in the SAN between user $u$ and attribute $a$ to compute $m_{AA}(u,a)$; 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 $m_{AA}(u,a)$. Also, for each user $u$, we can create predictors with respect to a social connection to another user $v$ using the 4 statistics $m_{CN}(u,v)$, $m_{CN-SAN}(u,v)$, $m_{AA}(u,v)$, and $m_{AA-SAN}(u,v)$.
The forest of trees we extract classification rules. For each of these rules \( R_j \) appears among the sensitive rules. The following criteria evaluate the risk of a rule \( R_j \).

- **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.

\[ m_{\mathbb{A}}(u,v), \text{ and } m_{\mathbb{A}-\text{num}}(u,v). \] 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, Å) 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 statical \( m_{\mathbb{A}}((10, £), \text{hearts}) \).

**B. 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 \[21\] to convert numeric attributes to nominal attributes as this is the default in WEKA \[12\] which is the package we use for our implementation. Now the dataset \( D \) has columns that are all nominal attributes.

User \( c \) provides a confidence value \( \beta \) for which the predictability of edge in the SAN should not be larger. A forest of trees is built using the algorithm for building decision trees. 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 each tree in the forest, the user \( p \), versus \( c \) 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 \( \beta \) which are at a depth less that 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 \( \beta \) on the 3 top levels of the tree.

From the forest of trees we extract classification rules. For each tree in the forest, the user \( c \) itself, lands on only one leaf. The rule is the conjunction of attributes from the path of the tree to the leaf. That is, if \( l_{ap}=v_c \) is the label for the leaf on the path \( x_{i_1} = a_{11}, x_{i_2} = a_{21}, \ldots x_{i_p} = a_{p1} \), then the logic rule \( (x_{i_1} = a_{11}) \land (x_{i_2} = a_{21}) \land \ldots \land (x_{i_p} = a_{p1}) \Rightarrow a_p = v_c \) is the classification rule for this leaf. All the rules of the forest provides a list \( R(b_u, \beta) = \{R_1, R_2, \ldots, R_m \} \) of classification rules.\[3\] For each of these rules \( R_j \in R(b_u, \beta) \), the following criteria evaluate the risk of a rule \( R_j \).

- 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.

---

1 Because the \( \land \) (Boolean AND operator) commutes, a rule may appear in more than one of the trees of the forest, but only one is kept.
Towards Identifying Risky Connections

"Trust is the glue of life. It’s the most essential ingredient in effective communication. It’s the foundational principle that holds all relationships."

Stephen Covey

<table>
<thead>
<tr>
<th>ID</th>
<th>Risk Ranking (one user) by \text{CUM_SENSITIVITY}.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First</td>
</tr>
<tr>
<td>h₀</td>
<td>m_{AA}(u, h₀)</td>
</tr>
<tr>
<td>h₁</td>
<td>m_{AA}(u, h₁)</td>
</tr>
<tr>
<td>h₂</td>
<td>m_{AA}(u, h₂)</td>
</tr>
<tr>
<td>h₃</td>
<td>m_{CN_SAN}(u, h₃)</td>
</tr>
<tr>
<td>h₄</td>
<td>m_{AA}(u, h₄)</td>
</tr>
</tbody>
</table>

Table I: SAN predictors of connections.

colours/shading. In general, the communities correspond to the top users which act as ‘hubs’. The community and hub structure are key challenges to data privacy because users tend to “flock together” due to affinities derived from common characteristics and interests, which are assigned and recorded as the attribute-values we use in our information gain calculation. The top 5 users are in good correspondence to the communities. In general, for 4 of these hubs, there is correlation between being a neighbor and having at least one attribute in common (with 2727 as an exception). We illustrate one of the heuristics in action in the case the information-item (that user \( u \) designates as confidential) is a connection. Then, for each 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, a row correspond to a popular user \( h_i \) 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 \( i \in \{0, \ldots, 4\} \); resulting in 20 additional columns. The metric values are discretised into 10 possible categories in steps of 0.10, giving \( 0.00-0.10, 0.11-0.20, \ldots, 0.91-1.00 \). We investigate our method to identify which other attribute \( a_p \). The earlier heuristics \( \text{CUM\_SENSITIVITY} \) and \( \text{TOTAL\_COUNT} \) both indicate that the set \( S \) is indeed \( \{\text{political orientation}, \text{religion}\} \) as these are the top predictors for \( a_p \). Hence, we decide to mask the values of these attributes in our own profile. However, note that users \( c_3 \) and \( c_4 \) have both of these attributes unmasked. They do not consider this information confidential. In order to quantify the increased risk of our subgroup \( N \) of friends with respect to all the users \( D \), we can create a subset of users \( \{u, D\} \) comprised of ourselves and our friends. Algorithm 1 first processes the subset \( N \) to quantify the predictability \( P_s \) of our class-value. Secondly, it processes the complete dataset \( D \) to all users to quantify the predictability \( P_a \) of the class-value. The difference between \( P_s \) and \( P_a \) will be the additional risk generated by our friends attribute-values, with respect to the complete dataset \( D \). We can try eliminating the highest risk users and recalculating the risk value, until it falls below a predetermined threshold.

However, Algorithm 1 processes the neighbors as a complete set and does not give an individual risk value for each neighbour. We will see how to do this later in Algorithm 3.

\text{Algorithm 1 Neighbour Collective Risk (subset based)}

\textbf{Input:} User Instance \( u \), Neighbour Instance set \( N \).
\textbf{Output:} Risk for neighbour instance set \( N \) for all risk attributes \( r_j \in A \) do
\begin{align*}
P_s &\leftarrow \text{classification precision}(N, r_j) \\
P_a &\leftarrow \text{classification precision}(D, r_j) \\
P_{s\_a} &\leftarrow \|P_s - P_a\| \\
\end{align*}
end for
\textbf{return} \( P_{s\_a} \)

<table>
<thead>
<tr>
<th>user</th>
<th>political orientation</th>
<th>religion</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td>x</td>
<td>Muslim</td>
</tr>
<tr>
<td>c₂</td>
<td>x</td>
<td>Christian</td>
</tr>
<tr>
<td>c₃</td>
<td>extreme left</td>
<td>Muslim</td>
</tr>
<tr>
<td>c₄</td>
<td>x</td>
<td>Christian</td>
</tr>
</tbody>
</table>

Table II: The projection of 4 users on 2 attributes (\textit{political orientation} and \textit{religion}).
Following on from the description of Algorithm 1, we observe that it is possible the estimate of $P_c$ is inaccurate because the immediate neighborhood is too small and the rule induction method does not have enough examples. As an alternative, we can create flag attributes (similar to discretized SAN metrics), which supply information about what proportion of our friends have matching values for the two attributes we consider predictors of the sensitive attribute $a_p$ (in this illustration, the attributes religion and political orientation are predictors). Then, we can include these flag attributes in the attribute set for the personalized information gain analysis.

In terms of the SAN, Algorithm 2 computes $PM_{r_j}(u) = \|\Gamma s+(r_j) \cup \Gamma s+(u)\|/\|\Gamma s+(u)\|$. Algorithm 2 illustrates how we process the flag attributes.

Algorithm 2 Neighbour Collective Risk flag based

Input: User Instance $u$, Neighbour Instance set $N$
Output: Risk for neighbour instance set $N$
Initialize vector $PM$ of proportion matching risk attributes to zero.
for all risk attributes $r_j \in A$
do for all neighbours $n_j \in N$
do if $r_{ij}$ equals $u_j$ then
$PM_j ← PM_j + 1$
end if
end for
$PM_j ← PM_j/\|N\|$
end for
return $PM$

The privacy preferences among users may vary. Promiscuous users are likely to divulge information and have a low level of confidentiality for information about themselves or information about others (their neighbors). Thus, befriending a promiscuous user is risky and thus, the challenge is to identify such promiscuous users, which may involve a dynamic aspect as promiscuity may require assessing the history of behavior of users regarding their information. In the context of social networks, the privacy risks are even larger as we consider Trojan horse approaches. These are “false” users, or “collusions” among users in order to extract information of genuine users by establishing relationships with the genuine user. The challenge here then is also to alert genuine users of potential Trojan horse users (which have the malicious intent of collecting information to later pass on to third parties).

In our running example above, despite we only consider $a_p$ as the confidential attribute, the fact that “religion and political orientation are predictors of $a_p$” would make us review our settings and conceal them, but also we may review these predictors for our own connections. With respect to Table[11] user $c_3$ is very discrete and has masked both predictor attributes. However, person $c_2$ has masked one out of the two and users $c_3$ and $c_4$ have not masked any. In the case of user $c_3$, we could say that the data is innocuous because (for the social network we model) the attribute-values of $c_3$ represent the modal values of the complete dataset. However, $c_2$ has attribute values which are good predictors (a discussion of the social network model will follow) to define the class $a_p = YES$. We also see that the attribute values for religion in two of the unmasked cases is Muslim. It is reasonable to assume that our attribute-values have a greater than average affinity with those of our friends [25]. Hence, there is a greater than average probability that our religion is Muslim. With respect to admitting new friends, we would therefore recommend to exclude users $c_2$ and $c_4$. We may also exclude user $c_3$ although the specific attribute values are not the predictor ones. On the other hand, we would accept user $c_1$ who has all the sensitive attributes masked. If we already have these four users as friends, we would recommend eliminating users $c_2$, $c_4$ and maybe $c_3$, from our list of friends.

In Algorithm 3 we see an embodiment of the evaluation of the risk for each individual neighbour. In Algorithm 3 the risk is simply equal to the number of sensitive attributes whose values are the same as user $u$. We could make the scheme more elaborate by awarding a value of 0.25 to neighbors who do not have the sensitive attribute masked although it is not the same as user $u$. Furthermore, we could use an ontology to calculate a distance between the categorical values. For example, in the case of religion, Christian and Jewish are closer together than Christian and Muslim. In the case of political orientation, a distance measure is made easier because the category is ordinal. For example, Centre and Far Left are at distance 3 (see Table III row ID 6).

Algorithm 3 Neighbour Individual Risk

Input: User Instance $u$, Neighbour Instance set $N$, complete Instance set $C$
Output: Top risk neighbour $N_{top}$ for instance set $N$
Initialize vector $NR$ of neighbour risk to zero
for all neighbour $n_i \in N$ do
for all risk attribute $r_j \in A$ do
if $r_{ij}$ equals $u_j$ then
$NR_i ← NR_i + 1$
end if
end for
end for
return $max_j NR_j$

A. Experimentation with simulated data

We initially explore these issues using synthetic online social network data. A sophisticated program [26] was used to produce realistic mixed data type attribute values populated into a topology generated by RMat [27]. The attributes that populate the social network are given in Table III. The attribute $a_p$ corresponds to the classes class$_s$ and class$_p$ with value “YES” and are indicated with *. We note that attributes 8 and 9 are a categorization of the output of Algorithm 2.

The following rules are used to generate the class flags for religion and political orientation, respectively.

Rule 1: if religion==Muslim then reassign class$_s ← “Yes”$
else reassign class$_s ← “No”$
Rule 2: if Political orientation==“Left” or Political orientation==“Far Left”
then reassign class$_p ← “Yes”$
else reassign class$_p ← “No”$

In Table IV we see the ranking of the attributes with respect to the class attribute class$_s$, in descending order of overall ranking score. We recall that the class attribute has been set, for testing purposes, according to RULE 1. From Table IV we see that political orientation and flag-religion are the top two ranked attributes, which is to be expected due to RULE 1, and given that we have created correlations in the synthetic data between religion, political orientation, residence, profession and Likes. We note that the religion attribute itself was not
TABLE III: The attributes of the simulated social network.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Domain of values</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: age</td>
<td>[18-25], [26-35], [36-45], [46-55], [56-63], [64-75]</td>
</tr>
<tr>
<td>1: gender</td>
<td>[male, female]</td>
</tr>
<tr>
<td>2: residence</td>
<td>{ Palo Alto, Santa Barbara, Boca Raton, Boston, Norfolk, San Jose }</td>
</tr>
<tr>
<td>3: religion</td>
<td>{ Buddhist, Christian, Hindu, Jewish, Muslim, Sikh, Traditional Spirituality, Other Religions }</td>
</tr>
<tr>
<td>4: marital status</td>
<td>{ Single, Married, Divorced, Widowed }</td>
</tr>
<tr>
<td>5: profession</td>
<td>{ Manager, Professional, Service, Sales and office, Natural resources consultation and maintenance, Production, Transportation and material moving, Student }</td>
</tr>
<tr>
<td>6: political orientation</td>
<td>{ Far Left, Left, Center Left, Center Right, Right, Far Right }</td>
</tr>
<tr>
<td>7: sexual orientation</td>
<td>{ Attractive, Bisexual, Heterosexual, Homosexual }</td>
</tr>
<tr>
<td>+8: % friends with same religion</td>
<td>{ 0.0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, 0.8-1.0 }</td>
</tr>
<tr>
<td>+9: % friends with same politics</td>
<td>{ 0.0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, 0.8-1.0 }</td>
</tr>
<tr>
<td>11: Likes 2</td>
<td>{ }</td>
</tr>
<tr>
<td>12: Likes 3</td>
<td>}</td>
</tr>
<tr>
<td>13: class_p</td>
<td>{ ‘YES’, ‘NO’ }</td>
</tr>
<tr>
<td>14: class_p</td>
<td>{ ‘YES’, ‘NO’ }</td>
</tr>
</tbody>
</table>

TABLE IV: Predictors for class_p ranked by overall frequency (212 users).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Ranking score</th>
</tr>
</thead>
<tbody>
<tr>
<td>10: (likes 1)</td>
<td>453</td>
</tr>
<tr>
<td>2: (residence)</td>
<td>451</td>
</tr>
<tr>
<td>11: (likes 2)</td>
<td>418</td>
</tr>
<tr>
<td>9: (flag-politics)</td>
<td>292</td>
</tr>
<tr>
<td>12: (likes 3)</td>
<td>198</td>
</tr>
<tr>
<td>0: (age)</td>
<td>171</td>
</tr>
<tr>
<td>5: (profession)</td>
<td>154</td>
</tr>
<tr>
<td>8: (flag-religion)</td>
<td>147</td>
</tr>
<tr>
<td>4: (marital status)</td>
<td>65</td>
</tr>
<tr>
<td>7: sexual orientation</td>
<td>5</td>
</tr>
<tr>
<td>3: (gender)</td>
<td>1</td>
</tr>
</tbody>
</table>

TABLE V: Predictors for class_p ranked by overall frequency (157 users).

B. How to assist the user and gain trust

In Figure 3 we see the scenario in which a user decides to hide some information about him/herself, the ‘political orientation’. However, we see that other attributes which remain visible are good predictors for ‘political orientation’. Our method identifies which other attributes have to be hidden, ranking them by CUM_SENSITIVITY with respect to ‘political orientation’. In Figure 4 we see another scenario where although a user has successively masked the attributes, the users friends also act as predictors for the private attribute ‘political orientation’. Again, our method identifies which users are the top predictors for ‘political orientation’ by ranking them by predictive sensitivity for this attribute. For the high predictors,
Fig. 5: Masked attributes and candidate friends.

we can also calculate the affinity between our friend and us, thus giving an additional risk factor. Finally in Figure 5 we see a similar scenario for the case of potential friends. That is, before we establish a friendship link, we can evaluate if this new friend will compromise the privacy of our hidden attribute-value, ‘political orientation’.

VI. CONCLUSIONS

As opposed to data anonymization, data sanitization, or statistical disclosure control where a data administrator decides globally what attributes are confidential, we have presented a novel approach for personal online data privacy customized for individual users which enables them to evaluate the privacy risks of their personal data. Our approach is radically new because each user decides what is shown as is, what is blocked and what gets generalized or obfuscated about his/her own data. This can be evaluated in the context of a user’s own data attributes, those of its friends and those of candidate friends. Using a sophisticated synthetic data generator we are able to simulate different scenarios for attribute-value combinations and friend affinities. A key innovation is that a rule model is built for each user, rather than the conventional approach of creating a general model for all users. We have confirmed that the risk attributes and users are clearly identifiable through our implementation of a SAN data structure and metrics, together with the Information Gain ranking. Effectively, the risk attributes defined in the synthetic data and the top ranked hits by Information Gain coincide as the top privacy risks.

REFERENCES