Generating Risk Reduction Recommendations to Decrease Vulnerability of Public Online Profiles

Janet Zhu, Sicong Zhang, Lisa Singh, Grace Hui Yang, Micah Sherr
Georgetown University
Department of Computer Science
Washington, DC, 20057 USA

***Abstract*—Preserving online privacy is becoming increasingly challenging due in large part to the continued growth of social media. Those who choose to share their information publicly may not realize what features of their profiles make their public data more identifiable and potentially vulnerable to cross-site record linkage. This paper proposes a risk reduction recommendation method that suggests removal or modification of a small number of attributes to make a profile less unique, thereby reducing the identifiability and vulnerability of the user. Empirical results on data collected from Google+, LinkedIn, and Foursquare show that users' vulnerability in terms of identifiability and data exposure level can be significantly reduced while public profile utility can be maintained using our proposed approach.**

I. Introduction

While methods and tools for obtaining public information traces of a user's online profile are emerging [5, 7, 8, 10, 11], users lack the ability to determine which attributes make their public profiles more distinguishing and potentially more linkable across sites. (We use the term *public profile* to refer to the attribute values that a user makes available to the public.) Similarly, we currently lack methods for adequately assessing a user's vulnerability compared to others in the population. Unfortunately, what users do not know has profound implications for their privacy: (1) users do not know how vulnerable they are compared to other online users, and therefore cannot determine their relative exposure; and (2) they do not know how to best adjust their public profiles to reduce their vulnerability, and must therefore make *uninformed* decisions when attempting to safeguard their privacy. This paper aims to address these problems by providing *risk reduction recommendations* of changes that allow users to better hide within a crowd, thereby reducing their identifiability. Some online social network services make privacy concerning suggestions based on the information posted on its own site [6]. However, we have not seen suggestions to edit user profiles based on the vulnerability from cross-site attacks.

Our approach to generating recommendations can be viewed as a macro-strategy that makes profile-level suggestions for removing or changing attributes to ensure that the profile 'falls within a crowd'. More specifically, the risk reduction recommendations focus on suggesting modifications to a user's public profile to directly match a *persona*, where a *persona* is a set of attribute-value pairs that occur together in a population with a frequency above a predefined threshold. A profile is deemed safe if it matches a persona because that profile's particular set of attributes occurs enough times in the population to allow it to blend into a homogeneous group. The concept of blending into a crowd is similar to the idea of k-anonymity [15], but we extend it to the realm of public social network data containing shared attributes across websites.

When generating recommendations, we have competing objectives. On one hand, we want to decrease the distinctiveness of each profile. On the other hand, we still want to share some information. Removing all attributes minimizes the vulnerability associated with sharing potentially distinguishing information, but it is an unappealing option to users who want to maintain a public profile. The *profile utility = 0*. Keeping all attributes preserves profile utility at the potential expense of maintaining a reasonable level of privacy. Therefore, we are concerned with striking a balance between these two goals. Our empirical evaluation using data from three different social networking sites shows the strengths and weaknesses of our approach. We will show that persona-based recommendations are effective in allowing a user to reduce his/her vulnerability risk while still maintaining profile utility.

II. Related Work

This paper investigates risk reduction recommendations for individuals concerned about web privacy. To the best of our knowledge, this is a novel problem that has not been previously explored, but there are a number of areas of relevant research. Much work exists on privacy in social networks specific to cross-site social identity linkage [5, 8, 10, 11, 13, 14, 16]. These works support our claim that inference across sites is possible and in some cases, straightforward. None of this literature discusses methods for reducing the likelihood of inference or decreasing the vulnerability of user public profiles.

Many methods have been proposed for recommendation systems [3, 9, 12, 17] that recommend products or attempt to improve the quality of service. For example, Google News makes use of users' click histories to personalize their news experience [3] and Amazon uses item-to-item collaborative filtering and other algorithms to recommend products to online shoppers [9]. The focus of these recommendation systems is to use social links and/or relevant trends and attribute values to recommend products. In contrast, we focus on risk reduction recommendations. These recommendations are based on identifying features that make individuals more unique. To the best

of our knowledge, this paper is the first to propose developing individualized recommendations to improve one's web privacy.

Several methods have been proposed for assessing an individual's level of exposure within online social networks [2, 4, 6]. However, because the goals of the papers are different, the proposed scoring methods are not applicable to this work.

III. Background and Notation

In this section, we first introduce the notation and definitions that will be used throughout the paper. We then more formally define our risk reduction task.

Definitions and notation. Let $\mathcal{D}$ represent a data set of individuals in a population. This data set contains attribute-value pairs from multiple sites and can be viewed as a collection of user profiles. Let $U = \{A_1, A_2, \ldots, A_j\}$ represent a particular individual or user in $\mathcal{D}$, where $A_i = \langle \alpha_i, v_i \rangle$ is an attribute-value pair consisting of an attribute α_i and a specific value v_i. Examples include $\langle$gender, `female`$\rangle$ and $\langle$country, `USA`$\rangle$. For simplicity, we assume that all the attribute-value pairs associated with each user U correctly describe the user. We refer to the set of attribute-value pairs associated with U as U's *public profile*. This public profile may contain attribute-values pairs from a single site or from multiple sites. By definition, U wants to maintain some attribute value pairs in his/her public profile. In other words, he or she wants to have a public presence on one or more social media sites. One measure of U's exposure relative to others is to compute a *data exposure level* δ, where $\delta = |U|/|U_{avg}|$. Here $|U|$ is the number of attributes U publicly shares and $|U_{avg}|$ is the average number of attributes shared by individuals in $\mathcal{D}$. A $\delta > 1$ indicates that a user shares more than the average person in $\mathcal{D}$.

We define an adversary to be an individual that knows a small amount of information about U (e.g. name) and looks at public data found on different sites in $\mathcal{D}$ to learn more attribute value pairs available in U's public profile. An adversary uses record linkage techniques to discover more attribute values in U's public profile by linking and aggregating attribute value pairs across different social media sites, keeping only those pairs that have a high confidence of being true[1]. The *identifiability* $\mathcal{I}$ of a profile is a measure of how many attributes an adversary can infer with high confidence. The set of attribute-value pairs known to the adversary is referred to as his/her set of beliefs, $\mathcal{B} = \mathcal{B}_{\text{initial}} + \mathcal{B}_{\text{learned}}$, where $\mathcal{B}_{\text{initial}}$ are the adversary's initial beliefs or knowledge about U and $\mathcal{B}_{\text{learned}}$ are the beliefs determined by record linkage across different sites. Table I shows an example of an adversary's set of learned beliefs.[2] It is the adversary's goal to learn all of the attributes for a target user U, i.e. for $\mathcal{B} = U$.

Finally, let $\mathcal{R}$ represent a recommendation system that gives users risk reduction recommendations. $\mathcal{R}$ does not have access to record linkage strategies used by an adversary to create $\mathcal{B}$. Instead, $\mathcal{R}$ is given $\mathcal{B}_{\text{learned}}$ as input and must return a set of recommendations consisting of attribute-value pairs that should be removed or modified in order to reduce U's vulnerability, where vulnerability can be measured using identifiability $\mathcal{I}$ or data exposure level δ. We also do not want to recommend removing or changing all attribute-value pairs since users want to maintain an online presence. Notice that we consider attributes in $\mathcal{B}_{\text{initial}}$ to be non-removable since these are known values to the adversary and are also assumed to be desirable by U to be maintained publicly.

$\mathcal{B}_{\text{learned}}$	Confidence
$\langle$Gender, Female$\rangle$	0.9
$\langle$Language, French$\rangle$	0.75
$\langle$College, UVA$\rangle$	0.8
$\langle$Occupation, Teacher$\rangle$	0.9
$\langle$Industry, Design$\rangle$	0.6
$\langle$State, Wyoming$\rangle$	0.75

TABLE I: Example: adversary's starting beliefs & confidences.

Problem statement More formally, the problem can be defined as follows: U has a public profile that is a set of attribute-value pairs across a set of sites. Given an initial set of beliefs, $\mathcal{B}_{\text{initial}}$, an adversary discovers a set of beliefs with high confidence $\mathcal{B}_{\text{learned}}$ about an individual U. We need to recommend a minimal set of attribute modifications $\mathcal{M}$ to U's public profile that when implemented maximally reduces the vulnerability $\mathcal{V}$ of U while still maintaining utility by requiring that at least τ attributes remain in U's public profile.

IV. Persona-Based Recommendations

There are three levels of information that can be considered when making risk reduction recommendations: the attributes α, particular attribute-value pairs $A_i \in U$, or the entire profile of user U. Algorithms that focus on the distinguishing power of individual attributes or attribute value pairs are considered micro-strategies for risk reduction recommendations. To use these strategies, we need insight into the record linkage methods used. However, our recommendation system does not have that insight. Algorithms that consider the entire profile of the user can be viewed as macro-strategies. These anonymity strategies attempt to ensure that a profile falls within a crowd of other people with similar profiles. *Persona-based recommendations* fall into this category and are the focus of this section.

Here we consider how various combinations of $A_i \in U$ result in different levels of identifiability. The idea is that if a profile resembles the profiles of many others in the population, it would "blend into a crowd" and thus become less identifiable. In persona-based recommendations, we want to modify $\mathcal{B}_{\text{learned}}$ to match U to a pre-computed set of personas, where a persona P_m is an algorithmically-generated set of attribute-value pairs A that appear together frequently in $\mathcal{D}$. $P_m = \{A_1, A_2, \ldots, A_m\}$ represents a persona of size m, or a *m-persona*, consisting of attribute-value pairs. We say that a persona P_m has q-anonymity if at least q individuals in $\mathcal{D}$ contain the values in P_m. Notice that q-anonymity is not the

[1] As discussed in §II, a number of different algorithms have been proposed for record linkage across different sites [5, 8, 10, 11, 13, 14, 16].

[2] In this work we conservatively assume an adversary initially infers accurate beliefs. If the beliefs are not accurate, the user U is less identifiable.

same notion as k-anonymity since it includes all the attributes that can be learned, not just sensitive ones.

One way to generate personas is to use frequent itemset mining [1]. We refer to the list of personas generated using this method as $\mathcal{P}_{\text{list}}$. We consider an m-persona *large* if it occurs at least $min\_support$ times in the population $\mathcal{D}$. Instead of maintaining all the large itemsets, we make modifications that only maintain those that are at least size τ, i.e. the minimum number of beliefs required to maintain some utility. Algorithm 1 shows our persona-based recommendations. It takes as input an adversary's learned beliefs $\mathcal{B}_{\text{learned}}$, the precomputed persona list $\mathcal{P}_{\text{list}}$. It outputs a list of modifications $\mathcal{M}$ to the set of beliefs that when implemented increase the user's anonymity.

The algorithm goes through each of the personas in $\mathcal{P}_{\text{list}}$ and computes the edit distance d between the user profile $\mathcal{B}_{\text{learned}}$ and a persona $P_i \in P_{\text{list}}$ (line 6). Different types of edit operations are allowed to match a user profile to a persona, including removals, additions, or changes. A removal is defined as a complete discarding of an attribute A_i. An addition is an insertion of a new attribute A_i that was not present in the profile previously. A change takes an existing A_i and modifies the value v. The P_i with the lowest d is added to the matched list, $matchedList.\ make\_mods$ returns the set of modifications necessary for $\mathcal{B}_{\text{learned}}$ to match the best selected persona. In the event that multiple personas match with the same edit distance, all of the possible sets of modifications are returned. Of course, we may want to favor different types of modifications. For example, it may be more desirable to a user to remove an attribute value instead of adding a 'fake' attribute value.This is what we do in our evaluation – we set the cost for a removal less than that of an addition or change.

Algorithm 1 Persona-Based Recommendations.

1: **Input:** $B_{\text{learned}}, P_{\text{list}}$
2: **Output:** $\mathcal{M}$
3: $matchedList = null$
4: $min\_dist = \infty$
5: **for** each $P_i \in \mathcal{P}_{\text{list}}$ **do**
6: $\quad d = compute\_distance(\mathcal{B}_{\text{learned}}, P_i)$
7: $\quad$ **if** $d < min\_dist$ **then**
8: $\quad\quad min\_dist = d$
9: $\quad\quad update\_matched\_list(matchedList, P_i, d)$
10: $\quad$ **end if**
11: **end for**
12: $\mathcal{M} = make\_mods(\mathcal{B}_{learned}, matchedList)$
13: **return** $\mathcal{M}$

When considering the privacy guarantees, suppose we define a user to be *protected* if q other users share the same attribute value pairs with him/her. Recall that P_{list} is generated using frequent itemset mining with a support level of ϵ controlling the size and content of P_{list}. To be specific, a persona $P_m = \{A_1, A_2, ..., A_m\}$ will be included in P_{list} if and only if there exists at least $\epsilon \times |D|$ different user profiles $U_j \in D$, such that $P \subseteq U_j$. In our persona based recommendation algorithm, each user profile is modified to a persona from P_{list} that is the most similar to the user's original profile. Hence, this modified user profile (persona) will have at least $\epsilon \times |D|$ other user profiles who share the same attribute value pairs. Therefore, by definition, if a user adjusts his/her public profile to one that matches a persona, the user is protected.

V. Analysis of Recommendation Method

A recommendation system $\mathcal{R}$ may modify any subset of attribute-value pairs in order to decrease the user's vulnerability $\mathcal{V}$. While the actual decrease in vulnerability depends on the specific data associated with U, we can estimate the decrease in vulnerability by considering different distributions for attributes in D in the context of $\mathcal{B}_{\text{initial}}$ and $\mathcal{B}_{\text{learned}}$. We now present different cases to better understand this decrease.

Case 1: The attributes in $\mathcal{B}_{\text{initial}}$, i.e. the non-removable attribute-value pairs in U, are unique. For example, an individual may have a unique name. This means that record linkage can be done trivially using $\mathcal{D}$. In this case, the adversary will be guaranteed to find all attribute values of U that are publicly available. The only way that U can reduce his/her vulnerability is to remove public data from $\mathcal{D}$. For every attribute removed, the user's vulnerability ($\mathcal{I}$ or δ) is reduced by the number removed. No additional reduction can be guaranteed.

Case 2: The attributes in $\mathcal{B}_{\text{initial}}$, i.e. the non-removable attribute-value pairs in U are not unique. Assume there are K profiles other than U in $\mathcal{D}$ that have the same attribute-value pair, where K is a sufficiently high number of profiles to keep U hidden. In this case, the adversary can not use these values to learn other values with high confidence. Instead, the inference is based on the uniqueness of other attributes of U for record linkage. By recommending that U update values to match a persona, by definition of persona, enough other individuals in $\mathcal{D}$ share the set of publicly accessible attribute-value pairs an adversary may try to learn. We can, therefore, guarantee that these attributes cannot be determined with high confidence by the adversary.

VI. Experiments

We now empirically evaluate the persona recommendations approach to risk reduction. Because this is a new task, it cannot easily be compared to a baseline. However, we do evaluate it against other possible strategies in the context of identifiability.

A. Data Sets

We collected profile information from 29,915 users on about.me, a site that allows users to share information across social media accounts. We used the set of profiles from about.me as the verified ground truth profiles for every user.

To compute identifiability, $\mathcal{I}$, we need to simulate data collection by an adversary. We collect all of the individuals profiles from Google+, LinkedIn, and Foursquare using their respective APIs that have the same first and last name as the about.me profiles. Using this approach, we collect 49 different attribute types. Focusing on the structured attributes that users have control of, gives us 16 attributes, e.g. company, occupation, group, education, location, age range, relationship status, college, gender, etc. These attributes have between 4

Attribute-Value Pair	Frequency
⟨Gender, Male⟩	415,329
⟨Gender, Female⟩	287,079
⟨State, AR⟩	49,798
⟨Location, New York, NY⟩	30,598
⟨Graduation Year, 2012⟩	4,861
⟨College, Unity College⟩	4,821
⟨Country, Belgium⟩	101
⟨Company, Sun Microsystems⟩	91
⟨Occupation, Translator⟩	64
...	...

TABLE II: Frequencies for a sample of attribute-value pairs.

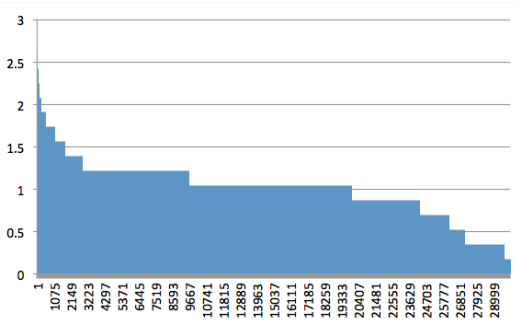


Fig. 1: Attribute distribution among users

(Gender) and over 275,000 (Company) distinct values. All the user profiles in our data set have between 1 and 15 attributes. Table II shows a small sample of attribute-value pairs and their frequency. There are over 2.5 million attribute-value pairs in the population.

To maximize the potential for linking across sites, our experiments focus on a subset of approximately 3000 individuals who had ground truth attributes across all three websites. We collected all attributes that could be obtained from the three sites using first name and last name as the initial search query and then built $\mathcal{B}_{\text{learned}}$. There were typically numerous profiles that matched a particular name from a single site. In our data set 74% of ground truth users have 3 or more candidate profiles on Google+, 62% on Linkedin and 45% on Foursquare. Because of API rate limitations, Google+ commonly returned a max of 50 profiles, Linkedin 25 and Foursquare 100. For the ground truth set of individuals, we computed the data exposure level δ (defined in Section III) of these individuals within the larger population. Figure 1 shows these results. The x-axis shows a sorted list of users and the y-axis shows the δ value for the user. We see that two thirds of these individuals have a δ value above 1, indicating that they are sharing more attributes than the average in the larger population. This metric is a simple way for users to understand how their data sharing practices compare to others who are using the same social media sites.

B. Experimental Design

Here we describe the decisions encountered when setting up this experiment, including the adversarial attack details, the persona construction details, and the edit distance metric used to understand the amount of change required for a user to "hide" his/her public data in a crowd.

1) Generating Adversarial Profiles: The information in $\mathcal{B}_{\text{learned}}$ was populated by generating user profiles for each individual by identifying attributes that occurred above a certain threshold within a single site and also across sites using the record linking algorithm proposed by Singh et al. [14]. Note that while we used the algorithm in Singh et al., any record linkage algorithm would work. Also, profiles can exist on only a single site as well. In these cases, we can consider only single site inference approaches. The key take away is that the recommendation system does not have access to a specific algorithm for computing a public profile. Instead, once it has a reliable public profile, it uses knowledge about the population to determine what should be removed.

The algorithm presented in Singh et al. [14] can be summarized as follows. Given a person, the algorithm identifies all profiles on a particular site that have the same name. It then looks for common attribute values. If the probability that an attribute appears on the profile for a given website is above a tunable site threshold, that attribute is added to the set of beliefs. This is done again across websites to see what attribute values are the same. For example, if ⟨Gender, `female`⟩ appears on both LinkedIn and Google+ for profiles with the same initial beliefs, $B_{initial}$, and the probability is higher than the threshold, we add ⟨Gender, `female`⟩ to the set of beliefs learned by the adversary, $\mathcal{B}_{\text{learned}}$. This process is repeated until no additional attribute value pairs can be inferred. We pause to mention that Singh et al. [14] discusses threshold tuning and parameter setting in general. We refer you to that paper for me details about those aspects of the record linking algorithm.

We calculate $\mathcal{I}$ by finding the normalized sum of confidences C_i across all $\mathcal{B}_i \in \mathcal{B}_{\text{learned}}$. Recall, one of our goals is to reduce $\mathcal{I}$ through a set of recommended attribute modifications $\mathcal{M}$ resulting from the persona-based algorithm. Using $\mathcal{M}$, we determine the new public profile that the adversary can learn, $\mathcal{B}_{\text{new}}$, and compute a new identifiability $\mathcal{I}_{\text{new}}$ score. The new identifiability $\mathcal{I}_{\text{new}}$ is then computed for $\mathcal{B}_{\text{new}}$ in the same way that $\mathcal{I}$ was computed for $\mathcal{B}_{\text{learned}}$. To maintain utility, we set the minimum number of attributes in $\mathcal{B}_{\text{new}}$ to be 2 ($\tau = 2$). We also analyze the effect of allowing for different modification strategies (removal, edit, addition) in the next subsection.

2) Persona generation: To generate personas, we need to identify combinations of profile attributes that occur frequently together. We use a modified version of Apriori [1], an offline frequent itemset mining algorithm, to compute these personas. For these experiments, the minimum number of attributes required to be in the public profile is 2 ($\tau = 2$).

A critical decision in generating personas using frequent itemset mining is the selection of the support level. In order to find a support level that generates the most effective number of $P \in \mathcal{P}_{\text{list}}$, we conducted a sensitivity analysis using a range of $min\_support$ values between 0.0007 and 0.0025. We found an exponential relationship between $|\mathcal{P}_{\text{list}}|$ and $min\_support$, illustrated in Figure 2a. The number of personas ranged from approximately 500 to 6,000. A support level of 0.001 is the point at which the slope of the curve decreases most drastically. Supports greater than 0.001 produced fewer personas,

$\mathcal{B}_{\text{learned}}$	P_1	P_2	P_3
⟨Gender, Female⟩ ⟨State, Wyoming⟩ ⟨College, UVA⟩ ⟨Language, French⟩ ⟨Occupation, Teacher⟩ ⟨Industry, Design⟩	⟨Gender, Male⟩ ⟨Language, French⟩ ⟨Country, France⟩	⟨Gender, Female⟩ ⟨State, Wyoming⟩ ⟨Language, Spanish⟩ ⟨Industry, Design⟩	⟨Gender, Female⟩ ⟨College, UVA⟩
Weighted distance	14	7	4
Unweighted distance	6	3	4

TABLE III: Edit distance cost of mapping profiles to personas.

with more people corresponding to each persona, or a larger crowd. However, profiles required too many modifications in order to match a given persona. Supports less than 0.001 generated more personas, and thus a smaller crowd, with no improvement in the number of modifications. Therefore, 0.001 was selected for these experiments. This generated approximately 2,500 personas from a population of 29,915 individuals, where each persona mapped to at least 30 profiles. Personas ranged from containing two attributes to six attributes. We pause to mention that if this methodology was used on a larger population, the level of anonymity of the personas would be even higher. For example, given a population of 1,000,000 people, a support of 0.001 would generate personas that map to 1,000 people.

To demonstrate that this strategy is easily scalable to a larger population, we conducted a small experiment to show a linear relationship between persona length and population size. We took random samples of 5000, 10000, 15000, 20000, and 25000 users and generated personas for each of these samples. We did 5 runs for each sample size for every support level between 0.0007 and 0.001. The linear relationship between sample size (x-axis) and the number of personas generated (y-axis) is shown in Figure 2b.

3) Edit distance metric: Recall that when we map a user profile to a persona, we compute the edit distance between each persona in P_{list} and the adversary's learned beliefs $\mathcal{B}_{learned}$, searching for the persona(s) with the smallest distance. Because we allow for different types of modifications to the user profile, i.e., deletions, updates, and additions, we need to understand the impact of having the same cost for each modification type (unweighted) vs. having different ones (weighted). Table III shows an example of both unweighted and weighted edit distance. The first column contains the learned beliefs $\mathcal{B}_{\text{learned}}$. The rightmost three columns show three different candidate personas. The last two rows display the calculation for both weighted and unweighted edit distance. In the weighted computation, additions and changes are assigned a cost of 5 and removals a cost of 1, whereas in the unweighted computation, each modification has the same cost of 1. This example illustrates the trade-off between modifying too much information and electing to share false information. For example, if $\mathcal{B}_{\text{learned}}$ were to match to P_2, there would only be 3 modifications that had to happen (2 removals and a value change), but the change increase the distance in the weighted version since false information decreases profile utility more.

C. Persona Matching Results and Discussion

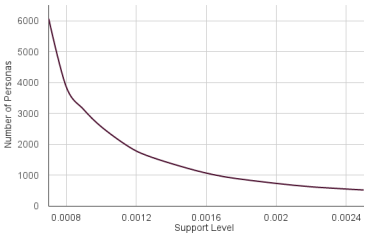
For this experiment, we compute the original identifiability of each user, make the modifications to the profiles based on the lowest weighted edit distance, determine the new set of beliefs the adversary can learn using the modified profile data, and recompute the identifiability for the user. For this set of users, the final improvement to the identifiability score for the Persona-Matching algorithm was calculated to be 24.35%. For over 50% of the profiles, only one or two attributes were removed. We also considered a few other strategies for attribute removal as a comparison to understand whether or not 24% is a reasonable reduction. Specifically, we consider strategies that remove attributes randomly (RANDOM), remove attributes based on the most distinguishing attribute measured by number of distinct values in the attribute domain (DOMAIN), and remove attributes based on uniqueness of attribute-value pairs in the population (VALUE). Table IV shows the reduction in identifiability using these approaches. We see that none of these methods are as effective as persona-matching.

When considering weighted edit distance, using the selected support level of 0.001, over 72% of the modifications were removals, meaning that this approach only required users to display erroneous information on their profiles less than 30% of the time. Figure 3 displays the ordered distribution of the three modifications. The x-axis represents the different profiles. Each profile is represented by a point on the graph. The metric on the y-axis is the frequency of each modification. According to this graph, about one-third of users did not have any modifications made to their profiles, meaning they matched perfectly to a persona. This is the best case scenario - no edits and high utility is maintained. About three-quarters of profiles did not have any additions or changes, which is a positive result because we want to minimize recommendations that introduce erroneous information and reduce utility. The largest number of removals made on any profile was 11, but this happened extremely infrequently. Only 14 out of 1600 profiles needed 8 or more removals. For those profiles, a large amount of utility is lost to maintain a high level of anonymity.

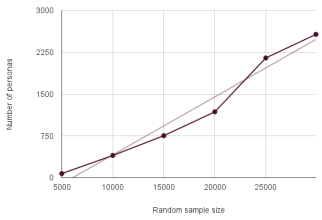
The modifications by attribute type are broken down in Figure 4. According to the figure, Location, Occupation, and Education are all commonly removed attributes using the persona-based algorithm. This is also consistent to removals recommended by the attribute-based algorithm, which adhered to expectations. Surprisingly, Gender was the fourth most common attribute to appear in modifications. The reason behind this is that Gender was the most frequently added attribute, but was not often recommended for removal or modifications. This is an interesting finding, signifying that adding Gender to a profile in fact decreased its identifiability.

Method	Remove 1	Remove 2	Remove 3
RANDOM	-5.12%	-11.05%	-14.04%
DOMAIN	-9.24%	-13.20%	-17.55%
VALUE	-10.76%	-16.43%	-20.84%

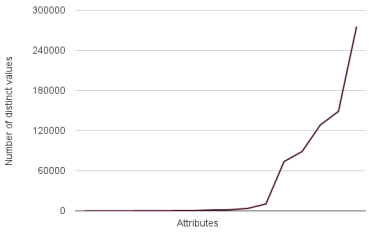
TABLE IV: Percentage decrease of identifiability ($\mathcal{I}$) when different numbers of attributes are removed.



(a) Relationship between support level (x-axis) and number of personas (y-axis).



(b) Relationship between population size and number of personas. $R^2 = 0.964$.



(c) Frequency of counts for attribute values

Fig. 2: Experimental design parameter analyzes

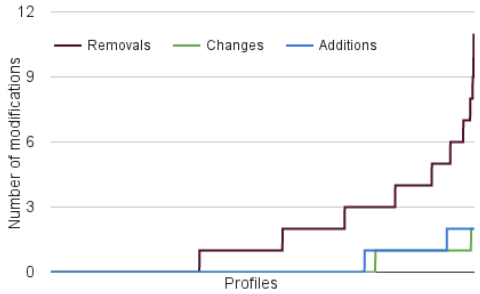


Fig. 3: Distribution of the three types of modifications.

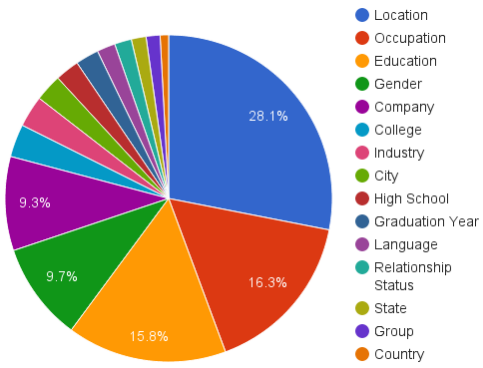


Fig. 4: Breakdown of modifications by attribute type.

VII. Conclusions

This paper explores a new methodology for reducing identifiability of public profile information, as well as giving users a way to assess their personal vulnerability based on the amount of data they choose to share publicly. Our approach produced recommendations of changes that significantly decreased the identifiability of a profile by considering commonly occurring attribute combinations in a population, denoted as personas. Modifying a profile to match an existing persona simulates blending into a crowd of similar individuals. The benefits of the persona-based recommendation system include its flexibility, the largest reduction in identifiability, and the decoupling from the construction of the belief set. A possible future direction would be to provide customized recommendations based on the level of privacy users want to maintain and consider different weights for attributes that are more sensitive. Finally, although 29,915 individuals were sufficient as a proof of concept, it would be ideal to demonstrate how these algorithms fare on more realistically-sized populations.

Acknowledgement

This work was supported in part by NSF grant #CNS-1223825. Any opinions, findings, conclusions, and recommendations expressed in this work are those of the authors and do not necessarily reflect the views of NSF.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *VLDB*, 1994.

[2] S. Alim, D. Neagu, and M. Ridley. A vulnerability evaluation framework for online social network profiles: Axioms & propositions. *Int. J. Intern. Technol. Secur. Syst.*, 4(2/3), July 2012.

[3] A. Das, M. Datar, and A. Garg. Google news personalization: Scalable online collaborative filtering. In *WWW*, 2007.

[4] J. Ferro, L. Singh, and M. Sherr. Identifying individual vulnerability based on public data. In *PST*, 2013.

[5] O. Goga, H. Lei, S. H. K. Parthasarathi, G. Friedland, R. Sommer, and R. Teixeira. Exploiting innocuous activity for correlating users across sites. In *WWW*, 2013.

[6] P. Gundecha, G. Barbier, and H. Liu. Exploiting vulnerability to secure user privacy on a social network site. In *KDD*, 2011.

[7] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff. Identifying users across social tagging systems. In *AAAI*, 2011.

[8] P. Jain, P. Kumaraguru, and A. Joshi. @i seek 'fb.me': Identifying users across multiple online social networks. In *WWW Companion*, 2013.

[9] G. Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, Jan 2003.

[10] J. Liu, F. Zhang, X. Song, Y.-I. Song, C.-Y. Lin, and H.-W. Hon. What's in a name?: An unsupervised approach to link users across communities. In *WSDM*, 2013.

[11] S. Liu, S. Wang, F. Zhu, J. Zhang, and R. Krishnan. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In *SIGMOD*, 2014.

[12] X. Liu and K. Aberer. Soco: A social network aided context-aware recommender system. In *WWW*, 2013.

[13] A. Ramachandran, L. Singh, E. Porter, and F. Nagle. Exploring re-identification risks in public domains. In *PST*, July 2012.

[14] L. Singh, H. Yang, M. Sherr, Y. Wei, A. Hian-Cheong, K. Tian, J. Zhu, S. Zhang, T. Vaidya, and E. Asgarli. Helping users understand their web footprints. In *WWW 2015 Companion*.

[15] L. Sweeney. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, 2002.

[16] R. Zafarani and H. Liu. Connecting users across social media sites: A behavioral-modeling approach. In *KDD*, 2013.

[17] H. Zhu, H. Xiong, Y. Ge, and E. Chen. Mobile app recommendations with security and privacy awareness. In *KDD*, 2014.