Inferring Private Information Using Social Network Data

Jack Lindamood
Facebook
cep221@gmail.com

Raymond Heatherly
University of Texas at Dallas
rdh061000@utdallas.edu

Murat Kantarcioglu
University of Texas at Dallas
muratk@utdallas.edu

Bhavani Thuraisingham
University of Texas at Dallas
bxt043000@utdallas.edu

ABSTRACT

On-line social networks, such as Facebook, are increasingly utilized by many users. These networks allow people to publish details about themselves and connect to their friends. Some of the information revealed inside these networks is private and it is possible that corporations could use learning algorithms on the released data to predict undisclosed private information. In this paper, we explore how to launch inference attacks using released social networking data to predict undisclosed private information about individuals. We then explore the effectiveness of possible sanitization techniques that can be used to combat such inference attacks under different scenarios.

Categories and Subject Descriptors

I.5.1 [**Pattern Recognition**]: Models; I.2.6 [**Artificial Intelligence**]: Learning

General Terms

Algorithms, Experimentation

Keywords

Social networks, privacy, inference

1. INTRODUCTION

Social networks are platforms that allow people to publish details about themselves and to connect to other members of the network through friendship links. Recently, the popularity of such on-line social networks is increasing significantly. For example, Facebook now claims to have more than 110 million active users.[1] The existence of on-line social networks that can be easily mined for various reasons creates both interesting opportunities and challenges. For example, social network data could be used for marketing products to the right customers. At the same time, privacy concerns can prevent such efforts in practice [1]. Therefore, for future social network applications, privacy emerges as an important concern.

In this paper, we focus on the problem of individual private information leakage due to being part of an on-line social network. More specifically, we explore how the on-line social network data could be used to predict some individual private trait that a user is not willing to disclose (e.g., political or religious affiliation) and explore the effect of possible data sanitization alternatives on preventing such private information leakage.

To our knowledge this is the first comprehensive paper that discusses the problem of inferring private traits using real-life social network data and possible sanitization approaches to prevent such inference. First, we present a modification of Naïve Bayes classification that is suitable for classifying large amount of social network data. Our modified Naïve Bayes algorithm predicts privacy sensitive trait information using both node traits and link structure. We compare the accuracy of our learning method based on link structure against the accuracy of our learning method based on node traits. Please see extended version of this paper [3] for further details of our modified Naive Bayes classifier.

In order to protect privacy, we sanitize both trait (e.g., deleting some information from a user's on-line profile) and link details (e.g., deleting links between friends) and explore the effect they have on combating possible inference attacks. Our initial results indicate that just sanitizing trait information or link information may not be enough to prevent inference attacks and comprehensive sanitization techniques that involve both aspects are needed in practice.

Similar to our paper, in [2], authors consider ways to infer private information via friendship links by creating a Bayesian Network from the links inside a social network. A similar privacy problem for online social networks is discussed in [4]. Compared to [2] and [4], we provide techniques that help in choosing the most effective traits or links that need to be removed for protecting privacy.

2. EXPERIMENTS

We wrote a program to crawl the Facebook network to gather data for our research. Because of the size of Facebook's social network, we limited crawling to profiles inside the Dallas/Forth Worth (DFW) network. This means that if two people share a common friend that is outside the DFW network, this is not reflected inside the database. Also, some people have enabled privacy restrictions on their profile and prevented the crawler from seeing their profile details. [2] Our total crawl resulted in over 167,000 profiles, almost 4.5 million profile details, and over 3 million friendship links. All but 22 of the people crawled were inside one, large component of diameter 16.

[1] http://www.facebook.com/press/info.php?statistics

[2] The default privacy setting for Facebook users is to have all profile information revealed to others inside their network.

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

Classifier	0t, 0l	0t, 10l	10t, 0l	10t, 10l
Naïve Bayes	0.7533	0.7157	0.6838	0.6790
Details Only	0.7942	0.7942	0.7003	0.7003
Links Only	0.7163	0.5855	0.6977	0.6066
Average	0.7970	0.7799	0.7184	0.7069

Table 1: Comparison of local classification methods

For our experiments, we consider only the subset of the graph for which we know the expressed political affiliation as either "Conservative" or "Liberal". This reduces our overall set size from approximately 160,000 to approximately 35,000 nodes.

To compare our methods to a traditional Naïve Bayes classifier, we implemented our own version of a traditional Naïve Bayes classifier. Then, we use the ideas discussed in [3] to create a list of the most representative traits in the graph, which we use to remove the 10 most predictive traits from the graph. That is, when we say that we remove K traits, we calculate which K traits are globally the most likely to reveal your true political affiliation and then remove those traits from every node that originally had them. Similarly, we use the ideas discussed in [3] to remove the 10 most telling links from every node in the graph. Unlike removing traits, which is done globally, removal of links is done locally. Finally, we combine the two methods and generate test sets with both 10 traits and 10 links removed from the graph. We refer to these sets as 0t, 0l; 10t, 0l; 0t, 10l; 10t, 10l removed, respectively. Following this, we randomly divide our nodes to form sets of 50% of the nodes in the training and 50% in the test sets. We repeated the previous process five times, and run each experiment independently. We then take the average of each of these five runs as the overall accuracy.

Our results, as shown in Table 1, indicate that the Average algorithm substantially outperformed traditional Naïve Bayes and the Links algorithm. Additionally, the Average algorithm generally performed better than the Details Only algorithm with the exception of the (0 traits, 10 links) experiments. An examination of the Links results for that experiment shows that the drop in Average accuracy can be accounted for by the exceptionally low performance of the Links classifier and the consistent Details Only performance for that point.

Also, as a verification of expected results, the Details classification accuracy only decreased when we removed traits from nodes, and the (0t, *) accuracies are approximately equivalent. Similarly, the Links accuracies were mostly affected by the removal of links between nodes, and the (*, 0l) points of interest are approximately equal. The difference of in accuracy between (0t, 0l) and (10t, 0l) can be accounted for by the weighting portion of the Links calculations, that depend on the similarity between two nodes.

Next, we examine the specific affects of removing traits. We first test the local classification accuracies after removing K traits, where $K \in [0, 10]$. After removing the K traits, we randomize our collection of nodes and create a test set of 50% of the nodes in the training and test sets. We then test the accuracy of the local classifier on this test set. We repeat 5 times and average the results for the overall accuracy for K, at each classifier. The results of this are shown in Figure 1. As evident from the results, after removing one trait, the classification accuracy immediately decreases significantly.

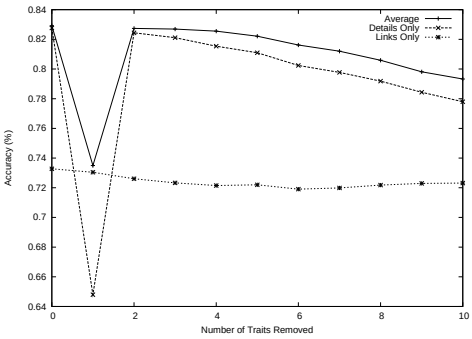

Figure 1: Local Classification accuracy by number of traits removed

After removing an additional trait, the classification returns to its prior accuracy, and for each subsequent trait removed we see a slight downward trend in classification accuracies.

The sudden downward spike can be easily explained by looking at the trait removal lists. The highest-ranked trait is evidence for the trait value of "Liberal". Removing this trait makes the probability of being "Conservative" outweigh the probability of a trait being "Liberal". This is why the Details accuracy is approximately the same as merely guessing the majority class for each node. However, when we remove the second trait, which is representative of being "Conservative" the probabilities again balance. None of the remaining traits are as highly indicative as the initial two, so we instead see a gradual decrease in the accuracy over the tested parameters. Unsurprisingly, the Links Only classifier is only slightly affected by the removal of traits.

In [3], we report additional experimental results that show the impact of link removal, collective inference and varying labeled vs unlabeled nodes ratios.

3. CONCLUSION AND FUTURE WORK

We addressed various issues related to private information leakage in social networks. Especially, we explored the effect of removing traits and links in preventing sensitive information leakage. Our results indicate that removing trait details and friendship links together is the best way to reduce classifier accuracy, but this is probably infeasible in maintaining the use of social networks. However, we also show that by removing only traits, we greatly reduce the accuracy of local classifiers, which is the maximum accuracy that we were able to achieve through any combination of methods.

4. REFERENCES

[1] Facebook Beacon, 2007. http://blog.facebook.com/blog.php?post=7584397130.

[2] J. He, W. Chu, and V. Liu. Inferring privacy information from social networks. In Mehrotra, editor, *Proceedings of Intelligence and Security Informatics*, volume LNCS 3975, 2006.

[3] R. Heatherly, M. Kantarcioglu, J. Lindamood, and B. Thuraisingham. Preventing private information inference attacks on social networks. Technical Report UTDCS-03-09, University of Texas at Dallas, 2009.

[4] E. Zheleva and L. Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *WWW*, 2009.