Anomolized Based Security for Private Information Attacks on Social Network

Yokesh Babu Sundaresan¹, Kumaresan P.² and Prabu A.S.³

Second year Computer Science and Engineering and Professors
SCSE, SITE, VIT University, Vellore, Tamil Nadu, India.

Abstract
Online social networks are a place where a large community of people publishes details about themselves and their friends which meant to be kept private. Many social networks provide options and security to user for the purpose of keeping their information private. Yet it is possible to predict private information from the released data using a learning algorithm. In this project, we address the inference attacks issues which uses released social networking data to predict private information. We then devise three possible sanitization techniques that could be used in various situations. Then, we explore the effectiveness of these techniques and attempt to use methods of collective inference to discover sensitive attributes of the data set. We show that we can decrease the effectiveness of both local and relational classification algorithms by using the sanitization methods we described.

Keywords: Inference Attacks, Learning Algorithm, Sanitization techniques, Classification and Collective Inference.

INTRODUCTION
Social networks are online applications that allow their users to connect by means of various link types. As part of their offerings, these networks allow people to list details about themselves that are relevant to the nature of the network. For instance, Facebook is a general-use social network, so individual users list their favorite activities, books, and movies. Conversely, LinkedIn is a professional network; because of this, users specify details which are related to their professional life (i.e., reference letters, previous employment, and so on.) Because these sites gather extensive personal information, social network application providers have a rare opportunity: direct use of this information could be useful to advertisers for direct marketing. However, in practice, privacy concerns can prevent these efforts. This conflict between the desired use of data and individual privacy presents an opportunity for privacy-preserving social network data mining that is, the discovery of information and relationships from social network data without violating privacy.

Privacy concerns of individuals in a social network can be classified into two categories: privacy after data release, and private information leakage. Instances of privacy after data release involve the identification of specific individuals in a data set subsequent to its release to the general public or to paying customers for a specific usage. Perhaps the most illustrative example of this type of privacy breach (and the repercussion thereof) is the AOL search data scandal.

Private information leakage, conversely, is related to details about an individual that are not explicitly stated, but, rather, are inferred through other details released and or relationships to individuals who may express that detail. A trivial example of this type of information leakage is a scenario where a user, say John, does not enter his political affiliation because of privacy concerns. However, it is publicly available that he is a member of the “legalize the same sex marriage.” Using this publicly available information regarding a general group membership, it is easily guessable what John’s political affiliation is. Somewhat less obvious is the favorite movie “The End of the Spear.” We note that this is an issue both in live data (i.e., currently on the server) and in any released data. This paper focuses on the problem of private information leakage for individuals as a direct result of their actions as being part of an online social network.

BACKGROUND
Inside the social networks to encompasses a large volume based on the privacy is defined in the area of privacy. Participation in social networking sites has dramatically increased in recent years. Services such as Friendster, Tribe, or the Facebook allow millions of individuals to create online profiles and share personal information with vast networks of friends-and, often, unknown numbers of strangers. In this paper we study patterns of information revelation in online social networks and their privacy implications.

We analyze the online behavior of more than 4,000 Carnegie Mellon University students who have joined a popular social networking site catered to college.s. We evaluate the amount of information they disclose and study their usage of the site’s privacy settings. We highlight potential attacks on various aspects of their privacy, and we show that only a minimal percentage of users changes the highly permeable privacy preferences.

According to Michael Hay, Gerome Miklau, David Jensen, Philipp Weis, and Siddharth Srivastava was advances in technology have made it possible to collect data about individuals and the connections between them, such as email correspondence and friendships. Agencies and researchers who have collected such social network data often have a compelling interest in allowing others to analyze the data. However, in many cases the data describes relationships that are private (e.g., email correspondence) and sharing the data in full can result in unacceptable disclosures.

Present a framework for assessing the privacy risk of sharing anonymized network data. This includes a model of adversary knowledge, for which we consider several variants and make
connections to known graph theoretical results. On several real-world social networks, we show that simple anonymization techniques are inadequate, resulting in substantial breaches of privacy for even modestly informed adversaries. We propose a novel anonymization technique based on perturbing the network and demonstrate empirically that it leads to substantial reduction of the privacy threat. We also analyze the effect that anonymizing the network has on the utility of the data for social network analysis.

Online 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. An effective, scalable inference attack for released social networking data to infer undisclosed private information about individuals. We then explore the effectiveness of possible sanitization techniques that can be used to combat such an inference attack.

On-line social networks, such as Facebook, are increasingly utilized by many people. These networks allow users to publish details about them-selves and connect to their friends. Some of the information revealed inside these networks is meant to be private. Yet it is possible that corporations could use learning algorithms on released data to predict undisclosed private information. Exploring how to launch inference attacks using released social networking data to infer undisclosed private information about individuals. We then devise three possible sanitization techniques that could be used in various situations.

Then, explore the effectiveness of these techniques by implementing them on a dataset obtained from the Dallas/Fort Worth, Texas network of the Facebook social networking application and attempting to use methods of collective inference to discover sensitive attributes of the data set. We show that we can decrease the effectiveness of both local and relational classification algorithms by using the sanitization methods we described. Further, we discover a problem domain where collective inference degrades the performance of classification algorithms for determining private attributes. Privacy preserving analysis of a social network aims at a better understanding of the network and its behavior, while at the same time protecting the privacy of its individuals. We propose an anonymization method for weighted graphs, i.e., for social networks where the strengths of links are important. This is in contrast with many previous studies which only consider unweighted graphs. Weights can be essential for social network analysis, but they pose new challenges to privacy preserving network analysis.

We mainly consider prevention of identity disclosure, but we also touch on edge and edge weight disclosure in weighted graphs. We propose a method that provides k-anonymity of nodes against attacks where the adversary has information about the structure of the network, including its edge weights. The method is efficient, and it has been evaluated in terms of privacy and utility on real word datasets.

**EXISTING SYSTEM**

Existing privacy definitions such as k-anonymity diversity and so on are defined for relational data only. K-anonymity tries to make sure that an individual cannot be identified from the data but does not consider inference attacks that can be launched to infer private information. Differential privacy algorithm basically guarantees that the results are very similar with or without the data of any single user. In other words, differentially privacy guarantees that the change in one record, does not change the result too much.

**EXISTING SYSTEM DISADVANTAGES**

1. K-anonymity does not try to protect against inference attacks directly.
2. Differential privacy algorithm definition does not protect against the building of an accurate data mining model that can predict sensitive information.
3. Differential privacy definition is not directly applicable for preventing sensitive detail disclosure.

**PROPOSED WORK**

Our proposed system we develop methods for anonymization and classification task to preventing data from inference attack. A social graph is formed based on the user as a node and friendship links as their edges along with their detail types and detail values. Each individual have different private detail type so when a user do not specify a detail we consider that detail as private. First we use simple naive Bayes classifier as learning method that could predict a person’s private details. Finally we remove the private details for hiding the privacy of the users.

**ADVANTAGES:**

1. Much of the uniqueness in the data may be lost by using quasi identifiers but our method of anonymity preservation maintain the full uniqueness in each node, which allows more information in the data post release.
2. This system will keep the user information securely in the network.
3. It provides privacy for the user profiles so that other person will not able to view the user’s profiles if he is not a user’ friend.
4. We can share photos or messages to our friends privately which means send to particular friend or to a group.

**ARCHITECTURE ON SOCIAL NETWORKS**

In Online Social Networks Service, Usually are sharing the information to the group of people, individuals they create a profile and to give information about themselves. Using some personal information may be it identifiable information like social security numbers, like name and phone number which it uniquely identify a person. Some personal or sensitive
In Other Situations, the data needs to be published and shared with others. It usually contain the valuable information to be enable better social targeting of advertisements. The social networks sites, are famous from of social media are the applications that enable participants to connect by creating personal information profiles, inviting friends and colleagues to have access those profiles and sending e-mails and private messages between each others. Facebook is the world’s most popular application of this kind where it currently has more than 900 million active users and they spend over 1000 billion minutes per month of using the websites.

Privacy concerns of individuals in a social network can be classified into two categories: privacy after data release, and private information leakage. Instances of privacy after data release involve the identification of specific individuals in a data set subsequent to its release to the general public or to paying customers for a specific usage. Private information leakage, conversely, is related to details about an individual that are not explicitly stated, rather, are inferred through other details released or relationships to individuals who may express that detail. Social network data could be used to predict some individual private detail that a user is not list their affiliation, but also through inference could determine the affiliation of other users in their data, this would obviously be a privacy violation of hidden details.

Explore how the online social network data could be used to predict some individual private detail that a user is not willing to disclose (e.g., political or religious affiliation) and explore the effect of possible data sanitization approaches on preventing such private information leakage, while allowing the recipient of the sanitized data to do inference on non-private details.
Specifically, $RVx(a)$ is an average value for how often a node of class $Cx$ has a link to a node of class $Ca$. To classify node $ni$, the algorithm builds a class vector, $CVi$, where $CVi(a)$ is a count of how often $ni$ has a link to a node of class $Ca$. The class probabilities are calculated by comparing $CVi$ to $RVx$ for all classes $Cx$. The nBC classifier uses Bayes Theorem to classify based only on the link structure of a node.

That is, it defines

$$P(n_i=C_x|N)=\frac{P(N|n_i=C_x)XP(n_i=C_x)}{\prod_{nj \in Ni} P(n_j=C_x)XP(n_i=C_x)}$$

SOCIAL GRAPH CONSTRUCTION

This module provides the basic functionality required by a social networking website. The functionalities include registration, login form, adding friends, and search friends, adding post and commenting. This module helps us by creating the social data by extracting the user’s profile and friendship link within this social network because it is hardly possible for us to get user details from the other social networks.

GENERALIZATION

This constructs a social graph from the extracted information from the profile and posts. The graph has each user as a node and their friendship relationships as their edges. We also extract the details types and detail values of each user and form a vector model by partially identifying the private attributes. We generalize the graph to reduce the complexity of the graph and add probability values to each detail types.

CLASSIFICATION

This module uses a naïve Bayes classifier which uses Bayes theorem for classifying the private detail types. The Bayes classifier is first applied on details generate weights to each type and then it is applied on the link and generate another weight, finally both weights are combined to find the details types of individual.

HIDING PRIVATE DETAILS

This is used to hide the private details which have been identified by the above classifier by applying an anomaly technique. The process gets the private detail type from the classifier and hides the value of the details type. The data is not visible to the outsiders after the execution of this module.

Anonymity, Ano: is the degree to which an individual cannot be identified from a given set of individuals. Usually, an anonymity set contains similar information of individual such as age, height, occupation.

Pseudonymity, Pse: is the process of identifying an individual by some kind of pseudonym such as an ID number, Employee number, nicknames, etc. This has to ensure that the data does not give away any links of personal identification information to its pseudonomic name.

Unobservability, Unob: is the degree to where it is feasibly impossible for an attacker to distinguish between real data and noisy data in an environment when there is random noise injected to protect privacy.

Unlinkability, Unli: is the process of individual sequence of actions not being connected to each other. For example, a person can travel to two different places in one day. But each individual action cannot be linked together and associated with that person. Even though the true identity of the person is not given away, the sequence of actions must not be linked together as well.

Identity management, Id Mgmt: is based on deriving identities based on certain context and attributes of the individual from a set of attributes. From a certain set of attributes, one can derive different aspects of identities about a person.

EXPERIMENTAL RESULTS

Asp.net: This is Based on the Microsoft Visual Studio 2012 Project. This Web is an easy to use environment for developing dynamic web application. It features a Streamlined interface that focuses on providing the tools that you need for creating web application. Compatible with any browser or mobile device ASP.NET web page automatically renders the correct browser. It is any language supported by the .NET common language runtime, such as Microsoft Visual Basic And Microsoft Visual C# and Lot of features of ASP.NET Web forms like Server controls, Master Pages, Working With Data, Membership, Client script and Frameworks, Security, Performance, Internationalization, Debugging and Error Handling and Development and Hosting. Creating Database within the Visual Studio of MYSQL Database. And Connecting the database to store the Dataset. In this article is assumed that you have already familiar with MYSQL dot net connector with minimum knowledge that you are able to perform the four basic operation SELECT, INSERT, UPDATE, DELETE and JOIN OPERATION.
After finish your registration then you can login in the web Page. A login is the entering of identifier information into as system by a user in order to access that system.

After Entering the login page you can check your activities, friends and you can manage your profile details by own setting and it would be automatically saved in your database.

If You were login your admin you can see lot of stuffs, like Raw Profile, Raw Activities and so on. First thing you can see the Raw dataset to all the basic information of the User.

The implementation of this project. Finding from domain generalization of simply using prediction and guess the most populated class from background knowledge, method of does not decrease the accuracy of classification on the data set to reduce the complexity of the graph and add probability values to each detail types.

The Administration Panel can be accessed. Using the first thing you will see is a prompt to login to access the administration panel. This username and Password is different from your sites login details. After you Login, you will be taken to the dashboard which will show you statistics of your site.

Data Classification is the process of organizing the data into categories for its most effective and efficient use. This can be particular importance for risk management. We Classify data and specify the roles and responsibilities of user data. Once a data classification scheme has been created, security standards that define the data’s lifecycle requirements should be
addressed. We try to sensitive internal data that if disclosed could negatively affect operations. So internal data that is not meant for public disclosure.

The Most of anonymous graph structure that has the lowest predictive accuracy is achieved when remove both details and links from the graph. To prevent effect of sanitization attack techniques further test the removal of details as an anonymization technique by using a variety of different classification algorithms to test the effectiveness of our method. Using Bayesian classifier to perform sanitization techniques make its easier to identify the individual details that make a class label more likely, decrease the accuracy of a far larger set of classifiers and hides the values of the details type.

FUTURE WORK
In future work all the above process should be in Gaussian Naïve Bayes in that graph Structure will be generated to identify inference attack to each users home page.

CONCLUSION
We addressed various issues related to private information leakage in social networks. We show that using both friendship links and details together gives better predictability than details alone. In addition, we explored the effect of removing details and links in preventing sensitive information leakage. In the process, we discovered situations in which collective inferencing does not improve on using a simple local classification method to identify nodes. When we combine the results from the collective inference implications with the individual results, we begin to see that removing details and friendship links together is the best way to reduce classifier accuracy. This is probably infeasible in maintaining the use of social networks. However, we also show that by removing only details, we greatly reduce the accuracy of local classifiers, which give us the maximum accuracy that we were able to achieve through any combination of classifiers.

REFERENCES


