Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Privacy Preserving and Overcoming Information Inference Attacks on Social Networks

Drishya T T, Shyja N K
CSE Department, MIT Anjarakandy,
Kannur University, Kerala, India

Abstract— Social networks are online applications which allows its users to connect by various link types. Online social networks, such as Facebook, Linkedln are progressively utilized by many people. Facebook is a general-use social network, which allows their individual users list their favorite activities, and so on. But some of the information revealed are meant to be private. However, privacy concerns can be used to prevent these efforts .This leads for privacy-preserving social network data mining, which is the discovery of information and relationships from social network data without violating privacy. This paper explores how to launch inference attacks using the released social networking data to predict private information that a user is not willing to disclose and explore the effect of possible data sanitization approaches on preventing such private information leakage.

Keywords— Data mining, Data sanitization, Privacy preserving, Social network analysis, Social network privacy

I. INTRODUCTION

Social networks are online applications that allow their users to connect with their friends. As part of their offerings, these networks allow people to list details about themselves that are relevant to the nature of the network. Facebook is a popular general-use social network where individual users list their favorite activities, books, and movies. Because these sites gather extensive private personal information, so the social network application providers have a rare opportunity. That is direct use of this information could be useful to advertisers for direct marketing. However, privacy concerns can be used to prevent these efforts [1]. 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. Examples 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 deals with the details about an individual that are not specifically stated, but, rather, are inferred through other details released and/ or relationships to individuals who may express that detail. 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.

This paper explores how the online social network data could be used to predict some individual private information that a user is not willing to disclose and explore the effect of possible data sanitization approaches on preventing such private information exposure, also letting the receiver of the sanitized data to do inference on the non-exclusive details. We 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 affiliation, sexual orientation) and explore the effect of possible data sanitization approaches on preventing such private information leakage, while letting the receiver of the sanitized data to do inference on non private details.

This problem of private information leakage could be an important issue in some cases. Recently, both ABC News [2] and the Boston Globe [3] published reports indicating that it is possible to determine a user's sexual orientation by obtaining a relatively small subgraph from Facebook that includes only the user's gender, the gender they are interested in, and their friends in that subgraph. Guessing an individual's sexual orientation or some other personal detail may seem like insignificant, but in some cases, it may create a negative repercussions (e.g., discrimination, and so on.). For example, by using the disclosed social network data (e.g., family history, life style habits, and so on.), guessing an individual's likelihood of getting Alzheimer disease for health insurance and employment purposes could be problematic.

The remainder of this paper is organized as follows: In Section 2, we describe some previous work in the area of social network anonymization. In section 3, we define the problem. In section 4, we describe the techniques to generalize the data. The section 4, describes the expected results.

II. RELATED WORK

The area of privacy inside a social network encompasses a large breadth, based on how privacy is defined. In [4], Backstrom et al. consider an attack against an anonymized network. In their model, the network consists of only nodes

and edges. Detail values are not included. The goal of the attacker is simply to identify people. In systems which include e-mail and messaging networks, one needs to set up the data to protect the privacy of individual users while preserving the global network properties. This is done through anonymization, a simple procedure in which each individual's "name" e.g., e-mail address, phone number, or actual name is replaced by a random user ID, but the connections between the people are revealed. Anonymization is intended to exactly preserve the pure unannotated structure of the communication graph while suppressing the "who" information. In anonymized social networks, passive attacks are carried out by individuals who try to learn the identities of nodes only after the anonymized network has been released. In contrast, an active attack tries to compromise privacy by strategically creating new user accounts and links before the anonymized network is released, so that these new nodes and edges will then be present in the anonymized network.

The active attacks will make use of the following two types of operations. First, an individual can create a new user account on the system. This adds a new node to *G*. Second, a node *u* can decide to communicate with a node *v*, this adds the undirected edge *(u, v)* to *G*. The goal of the attack is to take an arbitrary set of targeted users *w1, . . . ,wb*, and for each pair of them, to use the anonymized copy of *G* to learn whether the edge *(wi, wj)* in fact exists. This is the sense in which the privacy of these users will be compromised. In a passive attack, regular users are able to discover their locations in *G* using their knowledge of the local structure of the network around them. While there are a number of different types of passive attacks that could be implemented, here we imagine that a small coalition of passive attackers collude to discover their location. By doing so, they compromise the privacy of some of their neighbors those connected to a unique subset of the coalition, and hence unambiguously recognizable once the coalition is found. This work is not directly relevant to all settings in which social network data is used. This paper ignores details and do not consider the effect of the existence of details on privacy.

In [5], He et al. consider ways to infer private information via friendship links by creating a Bayesian network from the links inside a social network. While they crawl a real social network, LiveJournal, they use hypothetical attributes to analyze their learning algorithm. This work focuses on social network data classification and inferring the individual's private information. More private information is inferred by applying collective classification algorithm. The system explores how the online social network data could be used to predict some individual private trait that a user is not willing to disclose. For instance, in an office, people connect to each other because of similar professions. Therefore, it is possible that one may be able to infer someone's attribute from the attributes of his/her friends. In such cases, privacy is indirectly disclosed by their social relations rather than from the owner directly. This is called personal information leakage from inference. This system uses a collective classification algorithm for classifying the social network data. It has three components: local classifier, relational classifier and collective inference. Relaxation labeling is used as collective inference method. By applying the collective classification method the system could infer (indirect disclosure) the user private information using the released network data.

The system showed that, user's private information can be inferred via social relations and release of personal information in the social network. To protect the individual's private information leakage in social networks, the system either hide our friendship relations or ask our friends to hide their attributes. Compared to this work, we provide techniques that can help with choosing the most effective details or links that need to be removed for protecting privacy. Finally, we explore the effect of collective inference techniques in possible inference attacks.

In [6], Zheleva and Getoor proposed several methods of social graph anonymization, focusing mainly on the idea that by anonymizing both the nodes in the group and the link structure, that one thereby anonymizes the graph as a whole. The challenge of anonymizing graph data lies in understanding dependencies and removing sensitive information which can be inferred by direct or indirect means.

This work concentrates on hiding the identity of entities, considering the case where relationships between entities are to be kept private. In social network data, based on the friendship relationships of a person and the public preferences of the friends such as political affiliation, it may be possible to infer the personal preferences of the person in question as well. The process of anonymization involves taking the unanonymized graph data, making some modifications, and constructing a new released graph which will be made available to the adversary. The modifications include changes to both the nodes and edges of the graph. The anonymization of nodes creates equivalent classes of nodes. For the edge data, five different anonymization strategies were used.

This work proposed several methods of social graph anonymization, focusing mainly on the idea that by anonymizing both the nodes in the group and the link structure, that one thereby anonymizes the graph as a whole. However, their methods all focus on anonymity in the structure itself. Also, much of the uniqueness in the data may be lost. Through method of anonymity preservation in the proposed model, we can maintain the full uniqueness in each node, which allows more information in the data post release.

Online social networks, such as Facebook, are increasingly utilized by many people. These networks allow users to publish details about themselves and to connect to their friends. Some of the information revealed inside these networks is meant to be private. Yet it is possible to use learning algorithms on released data to predict private information. [7] explored how to launch inference attacks using released social networking data to predict private information. And then devise three possible sanitization techniques that could be used in various situations. Then, the effectiveness of these techniques and attempt to use methods of collective inference to discover sensitive attributes of the data set. This approach can decrease the effectiveness of both local and relational classification algorithms by using the sanitization methods described.

III. PROBLEM IDENTIFICATION

The existing systems only prevent the direct disclosure of the private information. But we introduce new inference attacks by predicting the private information from the released social network data. We sanitize both details and the underlying link structure of the graph. This is done by deleting some information from a user's profile and removing some links between friends in the graph. We also focus on generalizing detail values to more generic values. Then devise techniques that could be used in protection of the predicted private data. We can then show that sanitization still allows the use of other data in the system for further tasks.

IV. PROPOSED MODEL

A social network is represented as a graph, G = {V,E,D} where V is the set of nodes in the graph, where each node ni represents a unique user of the social network. The set of edges in the graph is represented bu E, that are the links defined in the social network. For any given friendship link F(i,j) between user ni and user nj, we assume that both F(i,j) Є E and F(j,i) Є E. D is the set of details from the social network. To evaluate the effect that changing a person's details has on their privacy, first we need to create a learning method that could predict a person's private details (for example, we assume that political affiliation is unspecified for some subset of our population). Our goal is to understand the feasibility of possible inference attacks and the effectiveness of various sanitization techniques combating against those attacks. So, initially a simple naive Bayes classifier is used. Using naive Bayes as learning algorithm allows us to easily scale the implementation to the large size and diverseness of the social network data set. It also has the advantage of allowing simple selection techniques to remove detail and link information when trying to hide the class of a network node.

4.1. Network Classification

Collective inference is a method for classifying the social network data using node details and connecting links in the social graph. Classifiers consists of three components: a local classifier, a relational classifier, and a collective inference algorithm. The local classifier is a type of learning method that are applied in the initial step of collective inference. Usually, it is a classification technique that examines the details of a node and then constructs a classification scheme based on the details that it finds there. A model is created based on the details of nodes in the training set. This is then applied to model to nodes in the testing set to classify them. The relational classifier is another type of learning algorithm that looks at the link structure of the graph, and then use the labels of nodes in the training set to develop a model which it uses to classify the nodes in the test set. Collective inference attempts to make up for these deficiencies by using both local and relational classifiers in a precise manner to attempt to increase the classification accuracy of nodes in the network. The collective inference method also controls the length of time the algorithm runs. As we remove details from the network, the set of nodes similar to any given node will also change. This can result in the decrease in accuracy of the links classifier.

4.2. Hiding Private Information

Our goal is to release rich social network data set while preventing sensitive detail disclosure using some data mining techniques. To formalize a privacy definition in our context, we have to address two issues regarding an inference attack. Initially, we need to have some understanding of the potential prior information (i.e., background knowledge) the adversary can use to launch an inference attack. Next, we need to analyze the potential success of inference attack given the adversary's background information. To label the first issue, we may try to come up with a privacy definition that is successful against all possible background information. It may not be possible to stop inference attacks against all background information. To label the second issue, we need to estimate the performance of the best classifier that can be built by using the released social network data and the adversary's background knowledge.

4.3. Data Generalization

To overcome the inference attacks on privacy, we try to provide detail anonymization for social networks. If a user inputs a favorite activity as the Boston Celtics, we could have, as an example, Boston Celtics →NBA →Basketball. This means that to completely anonymize the entry of "Boston Celtics" in a user's details, we replace it with "basketball". We also have the option of maintaining a bit more specificity by replacing it instead with "NBA". An outline of generalization algorithm is as follows:

Generalize(Δ,*G*)
1: *G'*←*G*
2: while Classify(*G*) - Classify(*G'*) ≤ Δ do
3: *S* ← all details that can be further generalized
4: *s* ←getHighestInfoGainAttrib(S)
5: Gen(*s, G'*)
6: end while
7: return *G'*

At each step, we generalize each detail type by one level [Lines 3-5] by determining which attributes can be further generalized without complete removal and keep a list of the accuracy of this generalization. At the end of each round, we

"permanently" store the individual detail type that provides the greatest privacy savings [Line 4]. When the changed graph, G'', meets the chosen privacy requirement, we consider it ready for release.

4.4. SVM as Classification Technique

Support Vector Machines (SVMs) are supervised learning methods that are used for classification and regression tasks that originated from statistical learning theory. SVM is a global classification model that generates non-overlapping partitions and usually employs all attributes. The aim is to solve only the problem of interest without solving a more difficult problem as an intermediate step. SVMs have been gaining popularity due to many attractive features and promising empirical performance. A classification task involves training and test sets which is composed of data instances. Each data instance in the training set contains some target value (class label) and several attributes (features). The goal of a classifier is to produce a model able to predict target values of data instances in the testing set, only for the known attributes. They are capable of delivering higher performance in terms of classification accuracy than the other data classification algorithms. Classifier accuracy is much more increased.

V. EXPECTED RESULTS

An online social networking site is created. Users can register and login into the site. The user can find friends, share data, pictures etc. One can post data on other's wall. New groups can be created. Users can upload photos by tagging their friends and also the can set the privacy for the photos being shared. Privacy can be provided as public or private. Data shared in private will be viewed only by the user's friends. Different inference attacks are launched on the data released in this social network site. The data is usually achieved through the private information leakage by sharing with friends or by other ways. These attacks can be prevented by removing certain details from the node as well as removing some links between the nodes. Classifier technique used will provide higher accuracy in the classification process compared to other techniques. That is mainly by data generalization. More distant the friendship link between two nodes, the more generalized data will be viewed by the friend.

VI. CONCLUSION

In this paper, various issues related to the private information leakage were discussed. Also shows that using both friendship links and details together gives better predictability than details alone. In addition, we also discussed the effect of removing details and links in preventing sensitive information leakage. By removing only details, we could reduce the accuracy of local classifiers, which give us the maximum accuracy that we were able to achieve through any combination of classifiers.

REFERENCES

[1] Facebook Beacon, 2007.

[2] K.M. Heussner, "'Gaydar' n Facebook: Can Your Friends Reveal Sexual Orientation?" ABC News, http://abcnews.go.com/Technology/gaydar-facebook-friends/story?id=8633224#. UZ939UqheOs, Sept. 2009.

[3] C. Johnson, "*Project Gaydar*," The Boston Globe, Sept. 2009.

[4] L. Backstrom, C. Dwork, and J. Kleinberg, "*Wherefore Art Thou r3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography*," Proc. 16th Int'l Conf. World Wide Web (WWW '07), pp. 181-190, 2007.

[5] J. He, W. Chu, and V. Liu, *"Inferring Privacy Information from Social Networks*," Proc. Intelligence and Security Informatics, 2006.

[6] E. Zheleva and L. Getoor, "*Preserving the Privacy of Sensitive Relationships in Graph Data*," Proc. First ACM SIGKDD Int'l Conf. Privacy, Security, and Trust in KDD, pp. 153-171, 2008.

[7] Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham, "*Preventing Private Information Inference Attacks on Social Networks*", 2013

