iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 Issue: III Month of publication: March 2014

DOI:

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Prediction of Concealed Information from Social Networks

R. Shiny Jenita[#1], J. A. M. Rexie [*2],

[#]*PG student, * Assistant Professor at Department of Computer Science and Engineering, Karunya University, Coimbatore*

***Abstract*— The online applications like social networks that allow the users to connect each other by different links. The users can update their profile with information of their friends and their private information. Some of their private information has higher possibilities to predict the private information from the user's information using some learning algorithms. To reduce accuracy of the profile, three refinement techniques are created which is used in various situations and the effectiveness of these techniques are explored. These techniques remove the details and friendship links together. This is the best way to reduce classifier accuracy. This method is probably infeasible in maintaining the use of social networks. Naive Bayes algorithm is used to gives the maximum accuracy that is able to find the classifier of a profile. The objective is to reduce the classifier accuracy, while the details and the friendship links are removed together.**

***Keywords*— Privacy attacks, anonymization network, data mining, social network analysis**

I. Introduction

Data Mining is known as Knowledge-Discovery in Databases (KDD). It is used for automatically searching large volumes of data for patterns and also explaining the past and predicting the future data. Research issues in data mining are mining complex knowledge from complex data, in a network setting, process-related problems and security, privacy and data integrity [5].

The online applications allow their users to connect by means of various link types. For example, Facebook is a general-use social network, so individual users list their favourite activities, books, and movies. Likewise, LinkedIn is a professional network; because, users specify details which are related to their professional life [5].

The main objective is to develop a technique to prevent the inference attacks on privacy information in Social Networks. Social networks are an online application that allow the users to connect by various links and also allows the users to share the details to their friends in the network. The user can bring out the information of them in the networks and some of their informations are private. These private information has higher possibilities to predict the private information from the user's information using some learning algorithms [1].

An inference attack is a data mining technique performed by analysing data for criminally gain knowledge about a subject or database. This gained knowledge may lead to predict the user's private information. So the privacy information is leaked [1].

In this process, the collective inference does not improve on using a simple local classification method to identify nodes. Removing details and friendship links together is the best way to reduce classifier accuracy. The public information has higher possibilities to predict the private information from the user's information using some learning algorithms. This project explored the effect of removing details and links in preventing sensitive information leakage [5].

II. ANONYMIZATION TECHNIQUES

The graph anonymization is the anonymization techniques which is 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 [8].

NODE ANONYMIZATION

Assume that the nodes have been anonymized with one of the techniques introduced for single table data. This anonymization provides a clustering of the nodes into m equivalence classes (C1, . . . , Cm) such that each node is indistinguishable in its quasi-identifying attributes from some minimum number of other nodes [5]. Using the notation C(vi) = Ck to specify that a node vi belongs to equivalence class Ck. The anonymization of nodes creates equivalent classes of nodes. Note, however, that these equivalent classes are based on node attributes only, there may be nodes with different identifying structural properties and edges [6].

EDGE ANONYMIZATION

For the relational part of the graph five possible anonymization approaches are described. The range from one which removes the least amount of information to a very restrictive one, which removes the greatest amount of relational data [5].

III. PRIVACY ATTACKS USING LINKS

Link-based privacy attacks take advantage of autocorrelation, so that the property of that the attribute values of linked objects are correlated. Example of this autocorrelation is that people who are friends often share common characteristics each other [5].

In addition to friendship or link information, the social networks offer a very rich structure through the group memberships of users. Every individual users in a group are bound together by some observed or hidden interest(s) that they share, and every individuals often belong to more than one group. Likewise groups offer a broad perspective on a person, and it may be possible to use them for sensitive attribute inference[7]. This problem becomes more complex, and their distributions suggest different values for the sensitive attribute. It is possible to construct a method which uses both links and groups to predict the sensitive attributes of users.

Use a simple method which combines the flat-link and the group-based classification models into one: LINK-GROUP. It uses all links and groups as features Thus utilizing the full power of available data. Like LINK and GROUP, LINK-GROUP can used in any traditional classifier[7]. The advantage of this method is able to discover the sensitive attribute values of some users with surprisingly high accuracy on the real-world social-media datasets [5].

IV. ATTACKS ON ANONYMIZED SOCIAL NETWORK

In this method present both active and passive attacks on anonymized social networks, showing that both types of attacks can be used to reveal the true identities of targeted users, even from just a single anonymized copy of the network [5].

The active attacks will make use of the following two types of operations. In the first operation, 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 [3].

The passive attack is based on the observation that most nodes in real social network data already belong to a small uniquely identifiable subgraph. If a user u is able to collide with a coalition of k − 1 friends after the release of the network, he or she will be able to identify additional nodes that are connected to this coalition, and thereby learn the edge relations among them [3].

IV. LEARNING METHODS

Data Sanitization is the technique which is used to disguising sensitive information and developed databases by overwriting it with looking realistic but false data of a similar type. The data in testing environments should be sanitized in order to protect valuable business information [2]. Basically there are two types of security. The first type is concerned data integrity. In this type the modification of the records is strictly controlled. The second type of security is the protection of the information from inappropriate visibility. For example of datas are Names, addresses, phone numbers and credit card details [2] [5].

V. NAIVE BAYES ALGORITHM

Bayesian classification provides practical learning algorithms and prior knowledge and observed data can be combined. It provides a useful perspective for understanding and evaluating many learning algorithms. And also calculates explicit probabilities for hypothesis and it is robust to noise in input data. Bayesian reasoning is particularly suited when the dimensionality of the inputs is high. It is applied to decision making and inferential statistics that deals with probability inference, used to predict future events. Naive Bayes models parameter estimation uses the method of maximum likelihood. It requires a small amount of training data to estimate the parameters is the advantage [5].

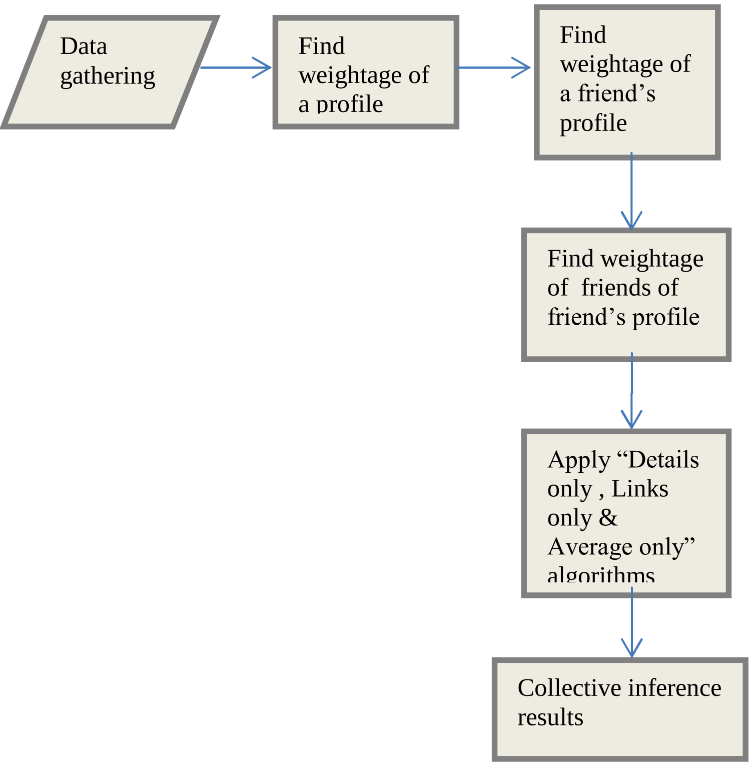
Data gathering
Find weightage of a profile
Find weightage of a friend's profile
Find weightage of friends of friend's profile
Apply "Details only , Links only & Average only" algorithms
Collective inference results

Fig.1. Prediction of Information

VI. MODULE DESCRIPTION

The modules used for implementation are as follows:

1. Data Gathering
2. Find weightage of user profile
3. Removing friendship links
4. Removing friends of friend's links

For implementation, entire project is divided into four modules. They are data gathering module, find weightage module, removing friend's links module and removing friends of friend's links module. These are used to reduce classifier accuracy.

DATA GATHERING MODULE

In this module the user's data are collected. The users login their profile and update their information. But the new users register their names before login. This information are stored automatically in the database. The user's friend's details are also stored in their profile.

FIND WEIGHTAGE MODULE

In this module the users profile weight is calculated by Naïve Bayes method. It is used to find the probability of a particular profile. Using this probability value easily identify the private information of a particular user.

REMOVING FRIEND'S LINKS MODULE

In this module, reduce the weight of the particular profile. The users are connecting with their friends links also. So first remove all the links which are not needed.

REMOVING FRIENDS OF FRIEND'S LINKS MODULE

In this module, to reduce the weight of the profile, remove the friends of friend's link in a particular profile. This is probably infeasible in maintaining the use of social networks. However, by removing only details, greatly reduce the accuracy of local classifiers. It gives us the maximum accuracy that is able to achieve through any combination of classifiers.

VII. PERFORMANCE METRICS

This project implemented four algorithms to predict the religion affiliation of each user. The first algorithm is called “Details Only”. This algorithm uses to predict political affiliation and ignores friendship links. The second algorithm is called “Links Only”. This algorithm uses to predict political affiliation using friendship links and does not consider the details of a person. The third algorithm is called “Average”. The Average algorithm predicts a node’s class value based on the following equation:

$$\mathbf{P_A(C^i_a) = 0.5 * P_D(C^i_a) + 0.5 * P_L(C^i_a)}$$

where $\mathbf{P_D}$ and $\mathbf{P_L}$ are the numerical probabilities assigned by the Details Only and Links Only algorithms, respectively [4][5].

To estimate social user’s private information, in future exploit the underlying social network structure to design an iterative algorithm. This algorithm derives private information estimates based on friends’, friends of friend’s and so on. If removing or altering these nodes can decrease information leakage. This technique would greatly reduce privacy leakages in its service.

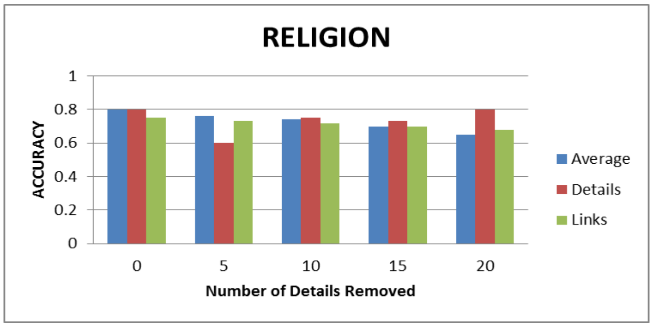

Fig.2. Result analysis

VIII.CONCLUSION

The proposed system uses learning methods in anonymization and classification tasks for hiding private information. After removal of details, test the removed details as an anonymization technique by using variety of different classification algorithms to test the effectiveness of proposed method. The effect of removing details and links is preventing sensitive information leakage. Compared with most previous studies, the proposed system overcomes the drawbacks of previous techniques. In the process, combine the results fromthe collective inference implications with the individual results, 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, removing friends of friend’s links reduces the accuracy of local classifiers.

REFERENCES

1. L. Backstrom,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.
2. *“Data Sanitization Techniques”*, A Net 2000 Ltd. White Paper.
3. M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava,*“Anonymizing Social Networks,”* Technical Report 07-19, Univ.of Massachusetts Amherst, 2007.
4. IEEE Transactions On Knowledge And Data Engineering, VOL. 25, NO. 8, AUGUST 2013 “*Preventing Private Information Inference Attacks On Social Networks”*,Raymond Heatherly, Murat Kantarcioglu, And Bhavani Thuraisingham, Fellow, IEEE
5. R.Shiny Jenita,J.A.M.Rexie “Security in prediction of private information on social network” International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 12, December - 2013
6. Van Eecke, Maarten Truyens *“Privacy and social networks”.*
7. D.J. Watts K. Liu and E. Terzi, *“Towards Identity Anonymization onGraphs,”* Proc. ACM SIGMOD Int’l Conf. Management of Data(SIGMOD ’08), pp. 93-106, 2008.
8. Zhou, Jian Pei, WoShun Luk School of Computing Science Simon Fraser University,Canada woshun@cs.sfu.ca,” *A Brief Survey on Anonymization Techniques for Privacy Preserving Publishing of Social Network Data”.*