IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 7, September 2014.
www.ijiset.com
ISSN 2348 – 7968

A Study on Leakage of Private Information in Social Networks and its Preventive Measures

Kastoori Sravan Kumar[1] and Dr .G. Venkata Rami Reddy[2]
[1]Computer Science,JNTUH-SIT,Hyderabad,Telangana-500085,India
[2]Associate Professor, JNTUH-SIT,Hyderabad,Telangana-500085,India

ABSTRACT

Social Networks are increasingly utilized by many people.These networks allow users to publish details about themselves.Some of the information provided by the users in the networks mean to be private information.Private information should be protected against unwarranted disclosure and only "Legitimate users" have a right to access it.It can be harmful to data owners and data users if it is misused.Although it is possible to use some types of learning algorithms on released data to predict private information. Inference attacks are initiated using released social networking data to predict private information. Collective inference methods are used to predict sensitive attributes of the data set.Network classification can be carried out with the combination of node details and connecting links in the graph model. Navie bayes classification is used to classify friendship links in a network in which local classifier, a relational classifier, and a collective inference methods are the three components used in the Network classification . Local classifier is a type of learning method and it can be applied in the initial step of collective inference. The relational classifier is a separate type of learning method and it analyzes the link structure and details of the node to identify a model for classification. Collective inference method can be used to increase the classification accuracy from the local and relational data details,which greatly reduces the accuracy of local classifier and give us the maximum accuracy through any combination of classifiers.

Key Terms- *Social Networks,inference attacks,private information*

1.INTRODUCTION

Social networking sites like Facebook, providing actual names and other personal information is published by the site(page known as a 'Profile'). These information consist of birth date, current address, and telephone number,designation and gender. Some sites allow users to publish more information about themselves such as interests,hobbies,favorite books or films, and even relationship status.Today,online social networks have greatly expanded the range of possible interactions, allowing you to share messages, pictures, files and even up to the every second information about what you are doing and where you are. These actions can be performed through the internet.Although these networks can be useful and promote social interaction both online and offline, when we using them you may be making information available to people . Think of a social networking site as being like a huge party. There are people there that you know, as well as some that you don't know at all. Imagine walking through the party with all your personal details.Remember that social networking sites are owned by private businesses, and that they make their money by collecting data about individuals and selling that data on, particularly to third party advertisers. For example, an online retailer that wants to sell its best customers additional

ISSN 2348 – 7968

products could buy details about their social media and mobile habits from advertises to figure out more efficient ways to market to them.Advertisers compile information about you, such as your age, race, sex, weight, height, marital status, education level, political beliefs, buying habits, household health, vacation dreams, and more.

The existing techniques and methods effectively allow to handle direct disclosure of sensitive personal information.However none of the existing techniques handle indirect disclosure of private information.Every individuals connected in social networks often share common attributes. 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. In order to perform privacy inference, we propose an approach to map Bayesian networks to social networks. We discuss prior probability, influence strength and society openness which might affect the inference, and conduct extensive experiments on a real online social network structure.

2.BACKGROUND

Privacy concerns for every individuals in a social network divided into two categories: privacy after data release, and private information leakage.Privacy after data release is related to details of every individuals which are explicitly stated and the details released to others . Private information leakage is that the deatils can be released by the inference relationships.For instance information leakage is a scenario a user, say ram does not enter his gender because of privacy concerns.

Suppose Facebook wishes to release data to electronic arts for their use in advertising games to interested people. However, once electronic arts have this data, they want to identify the political affiliation of users in their data for lobbying efforts. Because they would not only use the names of those individuals who explicitly list their affiliation, but also could determine the affiliation of other users in their data, this would obviously be a privacy violation of hidden details.Facebook allow third party applications on their platform.Those third party applications can access user profile data and can gather lots of information related to user.They may abuse the data.

Sanitization approach involves removing sensitive information and potential linking information that can associate an individual person to the sensitive information.For example,government agencies have to remove the sensitive information from the classified documents before making them available to the public so that the secrecy and privacy are protected.Hospitals may need to sanitize some sensitive information in patient’s medical reports before sharing them with other health care agencies,such as government health department,drugs companies,and research institutes.

In this paper we discuss the problem inference attacks in social network data and we sanitize the data in social networks then examines the effectiveness of sanitization approaches on data set.To protect privacy, we sanitize both details and the underlying link structure of the graph. That is, we delete some information from a user’s profile and remove some links between friends. We also examine the effects of generalizing detail values to more generic values. We then study the effect these methods have on combating possible inference attacks and how they may be used to guide sanitization. We further show that this sanitization still allows the use of other data in the system for further tasks. In addition, we discuss the notion of privacy in social networks and give a formal privacy definition that is applicable to inference attacks discussed in this paper.

ISSN 2348 – 7968

3.NETWORK CLASSIFICATION METHODS ON SOCIAL NETWORKS

Graph represented by a set of homogeneous nodes and a set of homogeneous edges Each node also has a set of Details, one of which is considered private.

3.1NAIVE BAYES CLASSIFICATION

Bayes classification is a graphical representation of joint probability distribution over a set of variables.It consists of network structure and collection of conditional tables.Classification based only on specified attributes in the node.

$$Argmax_x[P(C_x|D_{1,i,..}Dm,i)] = argmax_x[P(C_x)*P(D_{1,i}|C_x)*.....*P(D_{m,i}|C_x)/P(D_{1,i.....,}D_{m,i})]$$

3.2NAIVE BAYES ON FRIENDSHIP LINKS

Rather than calculate the probability from person nx to ny we calculate the probability of a link from nx to a person with ny‘s traits.

$$P(C_x|F_{i,j}) = P(C_x|L_1,L_2.....L_m) = P(C_x)*P(L_1|C_x)*....*P(L_m|C_x) \mid P(L_1.....,L_m)$$

3.3LINK WEIGHTS

Links also have associated weights.Represents how ‘close’ a friendship is suspected to be using the following formula:

$$W_{A,B} = |(D_{1,a}.....T_{N,a}) \Omega (T_{1,b},.....,T_{N,b})| / (|D_{*,a}|)$$

3.4LOCAL CLASSIFIERS

Local classifier is a type of learning method it can be applied in the initial step of collective inference.It is a classification technique that examines node details and constructs a classification network based on the details of the node. For instance, the naive Bayes classifier is one of the standard example of Bayes classification. This classifier construct a model based on the details of nodes in the training set.

3.5RELATIONAL CLASSIFIERS

The relational classifier is a type of learning method it analyze the link structure of the graph, and uses the details of the nodes in the training set to construct a model which it uses to classify the nodes in the test set.

3.6COLLECTIVE INFERENCE METHODS

Unfortunately, there are some issues with local classifier and relational classifier. Local classifiers analyze only the details of the node for classification.Relational classifiers analyze only the link structure of a node for its classification.Major problem with the relational classifier is that we might divide the labeled test sets so every node connect to at least one node in the training set ,in real time it may not satisfy the requirement.If this requirement is not met, then relational classification will be unable to classify nodes which have no neighbors in the training set.

Here,Collective inference attempts to combine up for these problems by using both local and relational classifiers in a precise manner to increase the classification accuracy of nodes in the network. By using a local classifier in the first iteration, collective inference ensures that every node will have an initial probabilistic classification, referred to as a prior. The algorithm then uses a relational classifier to reclassify nodes. At each of these steps $i > 2$, the relational classifier uses the fully labeled graph from step i - 1 to classify each node in the graph.

The collective inference method also controls the length of time the algorithm runs. Some algorithms specify a number of iterations to run, while others converge after a general length of time. That is, at each step i, the algorithm uses the probability estimates, not a single classified label, from step i - 1 to calculate new probability estimates. Further, to weight of each subsequent iteration compared to the previous iterations. As such, after we perform our sanitization techniques, to classify the nodes to examine the effectiveness of our approaches.

3.7REMOVING DETAILS

Ensures that no ‘false’ information is added to

ISSN 2348 – 7968

the network, all details in the released graph were entered by the user.Details that have the highest global probability of indicating political affiliation removed from the network.

3.8REMOVING LINKS

Ensures that the link structure of the released graph is a subset of the original graph.Removes links from each node that are the most like the current node.

3.9REMOVING DETAILS AND LINKS

Sanitize both details and links structure of the graph then delete the information form user's profile and remove some friendship links between friends.

4.CONCLUSION AND FUTURE WORK

Our survey provides novel contributions to the Networks.Consider more number of attributes which are not limit in the data set for the research.For instance,employers are inferencely attacked and consequently address challenging problems such as data disambiguation and named entity recognition.The different methods reduce information loss,and suggest users to reduce their privacy risk. When the results are combined from the collective inference methods with the individual results, are 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.Project greatly reduce the accuracy of local classifiers, which give us the maximum accuracy that are able to achieve through any combination of classifiers.Efficient graph management mechanism is provided in the social network for user profile classification system. This classification system supports incremental mining model with privacy ensured classification scheme.Future work can be conducted in identifying key nodes of the graph structure to see if removing or altering these nodes can decrease the private information leakage.

5.REFERENCES

[1] R. Heatherly, J. Lindamood, B. Thuraisingham and M. Kantarcioglu "Inferring Private Information Using Social Network Data," Proc. 18th Int'lConf. World Wide Web (WWW), 2009.

[2]Data Mining: Concepts and techniques,Second Edition,Micheline kamber and JaiWei Han.

[3] A.K. Elmagarmid, H. Elmeleegy, N. Talukder, M. Ouzzani and M. Yakout, "Privometer: Privacy Protection in Social Networks," Proc. IEEE 26th Int'l Conf. Data Eng. Workshops (ICDE '10), pp. 266-269, 2010.

[4]Database Modeling and Design:Logical Design,Fourth Edition,Thomas P.Nadeau,Sam S.LightStone and Toby J.Teorey.

AUTHOR PROFILE

Kastoori Sravan Kumar is currently pursuing his M.Tech(Computer Science) in School of IT,JNTU-Hyderabad. He did his B.Tech in Information Technology from J.B.Institute of Engineering and Technology, Hyderabad. His research area interest includes data mining and computer networks**.**

Dr.G.Venkata Rami Reddy is currently working as Associate Professor in School of IT,JNTU Hyderabad.He has more than ten years of experience. His articles and publications are published all over the world.His area of interest includes Web Technologies, Data Base and Management Systems.