International Journal of Research (IJR) Vol-1, Issue-9, October 2014 **ISSN 2348-6848**

Avoiding Implication Attacks on Confidential Information

U.Mahender[1] & Burri Naresh[2]

1. Asst. Professor,
TKR College of Engineering and Technology, Hyderabad, Telangana, India

2. PG Scholar,
TKR College of Engineering and Technology,
Hyderabad, Telangana, India

***ABSTRACT*:** *Online social networks are used by various public. These Social networks allow their users to connect by means of various link types in which the network gives an opportunity for public to list details about themselves that are relevant to the nature of the network. Here there is a chance of implication when user released some personal information in the network. Social network is represented as graph structure in which nodes and edges denotes user's of network and relationship links with friends. In this paper, the social network data has been confidential with the help of cooperative arrangement (both node and link arrangement) method. Using the cooperative arrangement method the system could infer more sensitive information from the network with high correctness. In cooperative arrangement method, it involves three components called local classifier, relational classifier and cooperative Implication. From this experiments conducted in this research work, it is observed that the future work provide better Arrangement correctness due to the application of cooperative arrangement method in link analysis.*

.

KEYWORDS: Social Network Analysis, Data Mining, Inference, machine learning methods, Cooperative Arrangement Algorithm.

I.INTRODUCTION

Social networking used to connect and share information with friends. Public may use social networking services for different reasons: to network with new contacts, reconnect with previous friends, maintain present relationships, construct or promote a business or project, participate in planning about a confident topic, or just have fun meeting and interacting with other users.

Facebook and Twitter, have a broad series of users. LinkedIn has positioned itself as professional networking sites profiles include continue information, and groups are formed to share questions and ideas with peers in similar fields. Unlike traditional personal homepages, public in these societies publish not only their personal attribute, but also their relationships with friends. It may cause the seclusion violation in social networks. Information seclusion is needed for users. Existing techniques are used to prevent direct disclosure of sensitive personal information.

This paper focuses on social network data arseriesment and inferring the individual's private information. More private information is inferred by applying cooperative arseriesment 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 (e.g. political or religious affiliation).

For instance, in an office, public

connect to each other because of similar professions. Therefore, it is probable that one may be able to infer someone's attribute from the attribute of his/her friends. In such cases, seclusion is indirectly disclosed by their social relations rather than from the owner directly. This is called personal information escape from Implication.

II. RELATED WORK

Lars Backstrom, Cynthia Dwork and Jon Kleinberg 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 only to identify people. Backstrom and Kleinberg consider a "statement graph," in which nodes are e-mail addresses, and there is a directed edge (u, v) if u has sent at least a confident number of e-mail messages or instant messages to v, or if v is included in up's address book.

Here they will be considering the "purest" form of social network data, in which there are only nodes corresponding to individuals and edges indicating social interaction, without any further annotation such as time-stamps or textual data.

Michael Hay, Gerome Miklau, David Jensen, Philipp Weis, and Siddharth Srivastava consider several ways of anonymizing social networks. Advances in technology have made it probable to collect data about individuals and the connections between them, such as email messages and friendships. Agencies and researchers who have collected such social network data often have a compelling interest in allowing others to analyze the data.

Hay et al. and Liu and Terzi consider several ways of anonymizing social networks. Our work focuses on inferring details from nodes in the network,not individually identify individuals. 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 group, LiveJournal, they use hypothetical attribute to analyze their learning algorithm.

Compared to Jianming He approch, provide techniques that can help with choosing the most effective details or links that need to be removed for protecting seclusion. Sen and Getoor compare various methods of link-based arseriesment including loopy belief propagation, mean field relaxation labeling, and iterative arseriesment. They rate each algorithm in terms of its robustness to noise, both in attribute values and correlations across links. And also compare the performance of these arseriesment methods &various types of correlations across links.

Zheleva and Getoor attempt to predict the private attribute of users in four real-world data sets: Facebook, Flickr, Dogster, and BibSonomy. They do not effort to actually anonymize or sanitize any graph data. Zheleva and Getoor work provides a substantial motivation for the need of the solution future in our work.

Talukder et al. propose a method of measuring the amount of information that a user reveals to the outside world and which automatically determines which information (on a per-user basis) should be removed to increase the seclusion of an individual.

III. PROPOSED SYSTEM

The proposed system use cooperative arseriesment algorithm for classifying the social network data. It has three components: local classifier, relational classifier and cooperative Implication. Relaxation labeling is used as cooperative Implication method. By applying the cooperative arseriesment method the system could infer (indirect disclosure) the user private information using the released network data.

The advantage of the system:

Cooperative arseriesment used to improve the classifier correctness. The cooperative Implication method (relaxation labeling) runs 99 iterations for classifying the network data. It uses local classifier as first iteration and set as a prior, and relational classifier as second iteration for trying more combinations with nodes and links to gain more user attribute which is used to infer the personal information.

SYSTEM ARCHITECTURE

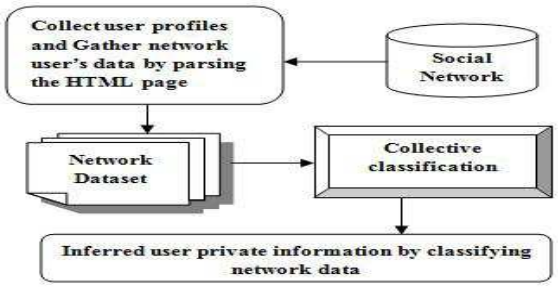

Fig 3.1 System Architecture Diagram

Crawl the Social (Ex.Facebook) network to gather data for experiments. Here the crawler overloaded a profile, parsed the details away of the HTML, and stored the details inside a My SQL database. Then, the crawler loaded all friends of the present profile and stored the friends inside the database both as friend-ship links and as probable profiles to later crawl.

By crawling the profile the dataset has been collected for the research. From the dataset, the user profiles and links are converted into the graph structure. Then use the cooperative arseriesment method on social network user data to infer the user's private information

SOCIAL NETWORK DATA GATHERING

For future work the details have been collect as follows. Username and password details of users in social network such as Facebook are collect. Log in to user accounts and download their profiles as .html files. Now apply html parser to that parses HTML files and collects attribute values of user profiles. Store the results in database. The report in database are exported into .csv format file for network arseriesment. Model the dataset file as network graph

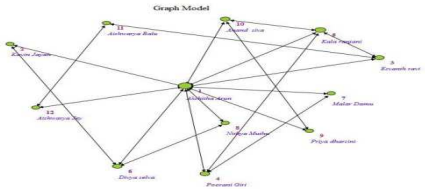

Fig 3.2 Social network graph structure

A Social network is represented a graph structure. The graph representation contains vertex, edges and details, where each node represents a unique user of the social network. The set of edges in the diagram, which are the links defined in the social network and the links used to establish the connection between the friends in the network

NETWORK ARRANGEMENT

Cooperative Implication is a method of classifying social network data using a combination of node details and connecting links in the social graph. Each of these classifiers consists of three components: a local classifier, a relational classifier, and a cooperative Implication algorithm.

Local classifiers are a type of learning method that are applied in the initial step of cooperative Implication. Naive bayes algorithm is used as a local classsifier. This classifier construct s a model based on the details of nodes in the training set. It then applies this model to nodes.

The relational classifier is a separate type of learning algorithm that looks at the link structure of the graph, and uses the labels of nodes. Four relational classifiers: class-distribution relational neighbor (cdRN), weighted-vote relational neighbor (wvRN), network-only Bayes classifier (nBC), and network-only link-

based arseriesment (nLB).

Local classifiers think about only the details of the node it is classifying. And relational classifiers consider only the link structure of a node. Cooperative Implication uses both node and links in the network to improve the classifier correctness. By using a local classifier in the first iteration, cooperative Implication ensures that every node will have an initial probabilistic arseriesment, 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 cooperative Implication method also controls the length of time the algorithm runs.

For cooperative Implication, relaxation labeling was best when there are few known labels. For relational arseriesment, the link-based classifier clearly was preferable when various labels were known. The lower-variance methods (wvRN and cdRN) dominated when fewer labels were known. Relaxation Labeling - repeatedly estimate class distributions on all unknowns, based on present estimates.
Steps involved in Cooperative arrangement:

Step 1: Assign initial label using local classifier. Use naïve bayes algorithm as local classifier.

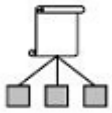

Fig 3.3 Node in a Network

Step 2: In first iteration the Naïve Bayes classifier selects the most likely arrangement V_{nb} given the attribute value $a_1, a_2, \ldots, a_n$. This result in,

$$V_{nb} = \text{argmax}_{v_j \in V} P(v_j) \prod P(a_i | v_j)$$

Generally estimate P (a_i |V_j) using m-estimates:

$$P(a_i|v_j) = \frac{n_c + mp}{n + m}$$

Where, n = the number of training examples for which v = v_j ; n_c = number of examples for which v = v_j and a= a_i ; p = a priori estimate for P(a_i | v_J) ; m = the equivalent sample size

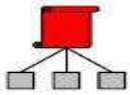

Fig 3.4 Assign Initial Label

Step 3: Assign Initial Label which has high probability. Set initial label as prior. Start the second iteration using relational classifier as weighted vote Relational Neighbor.

Step 4: In the wvRN relational classifier, to classify a node n_i, each of its neighbors, n_j, is given a weight. The probability of n_i being in class C_x is the weighted mean of the class probabilities of n_i's neighbors.
That is,

$$P(C_x^i|\mathcal{N}_i) = \frac{1}{Z} \sum_{n_j \in \mathcal{N}_i} [w_{i,j} \times P(C_x^j)],$$

Where N_i is the set of neighbors of n_i and $w_{i,j}$ is a link weight parameter given to the wvRN classifier. Assume that all link weights are 1.

Step 5: Learn a classifier from the labels or/and attributes of its neighbors to the label of one node. Here the network information is used.

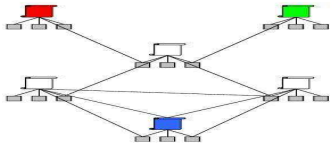

Fig.3.5 Use the attributes of related objects

Step 6: Apply relational classifier to each node iteratively and reclassify the labels.
Step 7: Relaxation labeling is used to assign the number of iterations to run and Iterate until the inconsistency between neighboring labels is minimized.

IV. RESULTS AND DISCUSSION

When classifying the social network data by cooperative arrangement method, it improves the classifier correctness. By doing this, the future system could infer user private information with high correctness. Consider the details and correctness of the classifiers when infer the private information with various arrangement methods

Correctness Private Data	Local Classifier Only	Relational Classifier Only	Cooperative Arrangement
Gender	0.7214	0.1672	0.8621
Religion	0.5134	0.4751	0.9519
Political Views	0.5541	0.2151	0.6273
Sexual Orientation	0.4023	0.2543	0.6979

Table 4.1

Classifier Correctness

In future system, local classifier uses the naïve bayes algorithm. Naïve bayes classifies the user nodes in the network and it finds the probability based on the node attribute. wvRN algorithm is used for relational arseriesment. It used to infer the details from the friendship links. Both the algorithms are infer the data from node/links. In this the system first it Classifies the node attribute and set as prior. So here some class labels are known. For cooperative Implication, relaxation labeling and wvRN was the best when there are few known labels. Relational classifier is used as relational classifier and reassigns the class labels based on the link details. The table 5.1 shows that the calculation of various classifier correctness

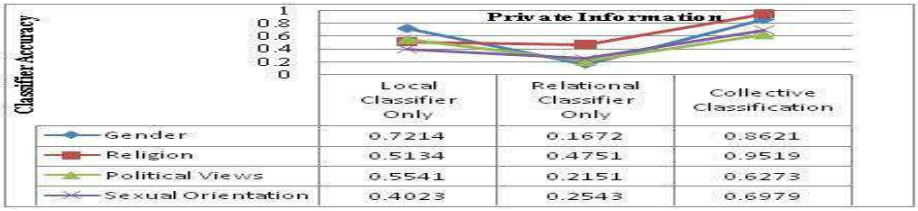

Fig.4.1 Calculating Classifier Correctness

From this experiments conducted in this research work, it is observed that the future work provide better arrangement correctness due to the application of cooperative arrangement method in link analysis

V. CONCLUSION AND FUTURE **WORK**

Here cooperative arrangement method used to infer the private information from the user nodes and related links. 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 systems either hide our friendship relations or ask our friends to hide their attributes. For protecting the user's private information perform the sanitization process and suppression techniques on the network data. When sanitize the network data it reduces the chance of inferring the individuals private information.

REFERENCES

[1] Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham, "Preventing Private Information Inference Attacks on Social Networks," IEEE Trans. Knowledge and Data Engineering, vol. 25, no. 8, Aug 2013, pp.1849-1861.

[2] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore Art Thou r3579x?: Anonymized Social Networks, Hidden Patterns, and Structural

[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] K. Liu and E. Terzi, "Towards Identity Anonymization on Graphs," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '08), pp. 93-106, 2008.

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

[6] P. Sen and L. Getoor, "Link-Based Arrangement," Technical Report CS-TR-4858, Univ. of Maryland, Feb. 2007.

[7] S.A. Macskassy and F. Provost, "Arrangement in Networked Data: A Toolkit and a Univariate Case Study," J. Machine Learning Research, vol. 8, pp. 935-983, 2007.

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

[9] E. Zheleva and L. Getoor, "To Join or Not to Join: The Illusion of Seclusion in Social Networks with Mixed Public and Private user Profiles," Technical Report CS-TR-4926, Univ. of Maryland, College Park, July 2008.

Authors Profile:

1.

Asst. Professor, Department of Computer Science & Engineering,TKR College of Engineering and Technology, Hyderabad,Telangana, India.

2.

PG Scholar,Department of Computer Science & Engineering,TKR College of Engineering and Technology, Hyderabad, Telangana, India.