PRIVACY PROTECTION SCHEME FOR PREVENTING THE DISCLOSURE INFORMATION OF USERS PROFILES

Koya Yogitha[1], N Shirisha[2]
[1]*M.Tech, CSE Dept, MLRIT, Hyderabad*
[2]*M.Tech, Asst. Professor, MLRIT, Dundigal, Ranga Reddy*

***Abstract*— this paper is motivated by the recognition of the need for a finer grain and more personalized privacy in data publication of social networks. We propose a privacy protection scheme that not only prevents the disclosure of identity of users but also the disclosure of selected features in users' profiles. An individual user can select which features of her profile she wishes to conceal. The social networks are modeled as graphs in which users are nodes and features are labels. Labels are denoted either as sensitive or as non-sensitive. We treat node labels both as background knowledge an adversary may possess, and as sensitive information that has to be protected. We present privacy protection algorithms that allow for graph data to be published in a form such that an adversary who possesses information about a node's neighborhood cannot safely infer its identity and its sensitive labels. To this aim, the algorithms transform the original graph into a graph in which nodes are sufficiently indistinguishable. The algorithms are designed to do so while losing as little information and while preserving as much utility as possible. We evaluate empirically the extent to which the algorithms preserve the original graph's structure and properties. We show that our solution is effective, efficient and scalable while offering stronger privacy guarantees than those in previous research.**

***Index Terms*- social networks, link analysis, anonymization, privacy in data mining.**

I. INTRODUCTION

Anonymized Social Networks. Digital traces of human social interactions can now be found in a wide variety of on-line settings, and this has made them rich sources of data for large-scale studies of social networks. While a number of these on-line data sources are based on

publicly crawl able blogging and social networking sites [6, 12] where users have explicitly chosen to publish their links to others, many of the most promising opportunities for the study of social networks are emerging from data on domains where users have strong expectations of privacy these include e-mail and messaging networks, as well as the link structure of closed (i.e. “members-only”) on-line communities [1, 2]. As a useful working example, consider a “communication graph,” in which nodes are e-mail addresses, and there is a directed edge (u, v) if u has sent at least a certain number of e-mail messages or instant messages to v, or if v is included in u’s address book. Here we will be considering the “purest” form of social network data, in which there are simply nodes corresponding to individuals and edges indicating social interaction, without any further annotation such as time-stamps or textual data.

The present work: Attacks on anonymized social networks. In this paper we 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, and with a surprisingly small investment of effort by the attacker.

We describe active attacks in which an adversary chooses an arbitrary set of users whose privacy it wishes to violate, creates a small number of new user accounts with edges to these targeted users, and creates a pattern of links among the new accounts with the goal of making it stand out in the anonymized graph structure. The adversary then efficiently finds these new accounts together with the targeted users in the anonymized network that is released.

At a theoretical level, the creation of O(plog n) nodes by the attacker in an n-node network can begin compromising the privacy of arbitrary targeted nodes, with high probability for any network; in experiments, we find that on a 4.4-million-node social network, the creation of 7 nodes by an attacker (with degrees comparable to those of typical nodes in the network) can compromise the privacy of roughly 2400 edge relations on average.

The privacy issue arises from the disclosure of sensitive labels. One might suggest that such labels should be simply deleted. Still, such a solution would present an incomplete view of the network and may hide interesting statistical information that does not threaten privacy. A more sophisticated approach consists in releasing information about sensitive labels, while ensuring that the identities of users are protected from privacy threats. We consider such threats as neighborhood attack, in which an adversary finds out sensitive information based on prior knowledge of the number of neighbors of a target node and the labels of these neighbors. In the example, if an adversary knows that a user has three friends and that these friends are in A (Alexandria), B (Berlin) and C (Copenhagen), respectively, then she can infer that the user is in H (Helsinki). We present privacy protection algorithms that allow for graph data to be published in a form such that an adversary cannot safely infer the identity and sensitive labels of users. We consider the case in which the adversary possesses both structural knowledge and label information.

The algorithms that we propose transform the original graph into a graph in which any node with a sensitive label is indistinguishable from at least 1 other node. The probability to infer that any node has a certain sensitive label (we call such nodes sensitive nodes) is no larger than 1=`. For this purpose we design `-diversity-like model, where we treat node labels as both part of an adversary's background knowledge and as sensitive information that has to be protected.

The algorithms are designed to provide privacy protection while losing as little information and while preserving as much utility as possible. In view of the tradeo_ between data privacy and utility [16], we evaluate empirically the extent to which the algorithms preserve the original graph's structure and properties such as density, degree distribution and clustering coefficient. We show that our solution is effective, efficient and scalable while offering stronger privacy guarantees than those in previous research, and that our algorithms scale well as data size grows.

II. Problem Statement

The current trend in the Social Network it not giving the privacy about user profile views. The method of data sharing or (Posting) has taking more time and not under the certain condition of displaying sensitive and non-sensitive data.

Problems on existing system:

There is no way to publish the Non sensitive data to all in social Network.

It's not providing privacy about user profiles.

Some mechanisms that prevent both inadvertent private information leakage and attacks by malicious adversaries.

Here, we extend the existing definitions of modules and we introduced the sensitive or non-sensitive label concept in our project. We overcome the existing system disadvantages in our project.

Advantages:

1. We can publish the Non sensitive data to every-one in social Network.

2. Its providing privacy for the user profiles so that unwanted persons not able to view your profiles.

3. We can post sensitive data to particular peoples and same way we can post non-sensitive data to everyone like ads or job posts.

III. System Development

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

User Module:

Users are having authentication and security to access the detail which is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first.

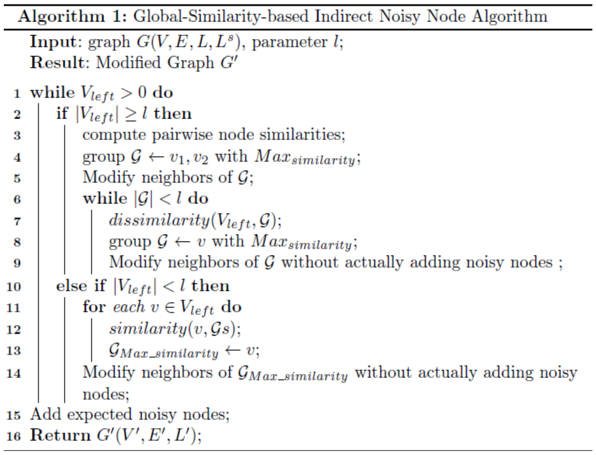
Information Loss:

We aim to keep information loss low. In- formation loss in this case contains both structure information loss and label information loss. There are some non sensitive data's are Loss due to Privacy making so we can't send out full information to the public.

Sensitive Label Privacy Protection:

There are who post the image to the online social network if allow the people for showing the image it will display to his requesters it make as the sensitive to that user. Thesis is very useful to make sensitive data for the public.

Algorithm:



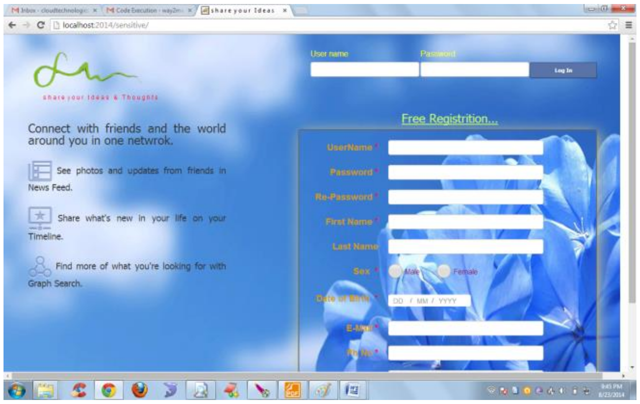
Algorithm 1: Global-Similarity-based Indirect Noisy Node Algorithm

```
Input: graph $G(V, E, L, L^s)$, parameter $l$;
Result: Modified Graph $G'$

while $V_{left} > 0$ do
    if $|V_{left}| \geq l$ then
        compute pairwise node similarities;
        group $\mathcal{G} \leftarrow v_1, v_2$ with $Max_{similarity}$;
        Modify neighbors of $\mathcal{G}$;
        while $|\mathcal{G}| < l$ do
            $dissimilarity(V_{left}, \mathcal{G})$;
            group $\mathcal{G} \leftarrow v$ with $Max_{similarity}$;
            Modify neighbors of $\mathcal{G}$ without actually adding noisy nodes ;
    else if $|V_{left}| < l$ then
        for each $v \in V_{left}$ do
            $similarity(v, \mathcal{G}s)$;
            $\mathcal{G}_{Max\_similarity} \leftarrow v$;
        Modify neighbors of $\mathcal{G}_{Max\_similarity}$ without actually adding noisy
        nodes;
Add expected noisy nodes;
Return $G'(V', E', L')$;
```

Experimental Results:

We compare the data utilities we preserve from the original graphs, in view of measurements on degree distribution, label distribution, degree centrality [9], clustering coefficient, average path length, graph density, and radius.



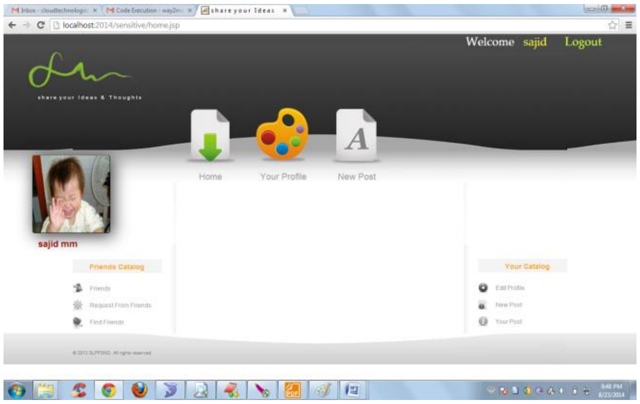
Connect with friends and the world around you in one netwrok.

See photos and updates from friends in News Feed.

Share what's new in your life on your Timeline.

Find more of what you're looking for with Graph Search.

In view of utility of released data, we aim to keep information loss low. Information loss in this case contains both structure information loss and label information loss.



IV. RELATED WORK

Here, we imagine that a coalition X of size k is initiated by one user who recruits k −1 of his or her neighbors to join the coalition. (Other structures could lead to analogous attacks.) We assume that the users in the coalition know the edges amongst themselves the internal structure of H = G[X], using the terminology from the active attack. We also assume that they know the names of their neighbors outside X. This latter assumption is reasonable in many cases: for example, if G is an undirected graph built from messages sent and received, then each user in X knows its incident edges. Other scenarios imply different levels of information: for example, if an undirected released network G is obtained from a directed graph where (u, v) indicates that v is in u's address book, then a node u does not necessarily know all its inbound edges, and hence doesn't know its full neighbor set in the undirected graph G. However, in the comparably plausible variant in which the directed version of an address book network is released, the nodes in X will have all the information they need for the passive attack.

To prove the correctness and efficiency of the attack, we show two things: with high probability the construction produces a unique copy of H in G, and with high probability, the search tree T in the recovery algorithm does not grow too large. It is important to stress that although these proofs are somewhat intricate; this complexity is an aspect of the analysis, not of the algorithms themselves. The construction of H and the recovery algorithm have already been fully specified in the previous subsection, and they are quite simple to implement. In keeping with this, we have structured this subsection and the next (on computational experiments) so they can be read essentially independently of each other.

The first necessary anonymization technique in both the contexts of micro- and network data consists in removing identification. This nave technique has quickly been recognized as failing to protect privacy. For microdata, Sweeney et al. propose k-anonymity [17] to circumvent possible identity disclosure in naively anonymized microdata. `-diversity is proposed in [13] in order to further prevent attribute disclosure.

Similarly for network data, Backstrom et al., in [2], show that naive anonymization is insufficient as the structure of the released graph may reveal the identity of the individuals corresponding to the nodes. Hay et al. [9] emphasize this problem and quantify the risk of re-identification by adversaries with external information that is formalized into structural queries (node refinement queries, subgraph knowledge queries). Recognizing the problem, several works [5, 11, 18, 20{22, 24, 27, 8, 4, 6] propose techniques that can be applied to the naive anonymized graph, further modifying the graph in order to provide certain privacy guarantee. Some works are based on graph models other than simple graph [12, 7, 10, 3]. To our knowledge, Zhou and Pei [25, 26] and Yuan et al. [23] were the first to consider modeling social networks as labeled graphs, similarly to what we consider in this paper. To prevent re-identification attacks by adversaries with immediate neighborhood structural knowledge, Zhou and

Pei [25] propose a method that groups nodes and anonymizes the neighborhoods of nodes in the same group

by generalizing node labels and adding edges. They enforce a k-anonymity privacy constraint on the graph, each node of which is guaranteed to have the same

V. CONCLUSION

In this paper we have investigated the protection of private label information in social network data publication. We consider graphs with rich label information, which are categorized to be either sensitive or non-sensitive. We assume that adversaries possess prior knowledge about a node's degree and the labels of its neighbors, and can use that to infer the sensitive labels of targets. We suggested a model for attaining privacy while publishing the data, in which node labels are both part of adversaries' background knowledge and sensitive information that has to be protected. We accompany our model with algorithms that transform a network graph before publication, so as to limit adversaries' confidence about sensitive label data. Our experiments on both real and synthetic data sets confirm the effectiveness, efficiency and scalability of our approach in maintaining critical graph properties while providing a comprehensible privacy guarantee.

REFERENCES

[1] L. A. Adamic and N. Glance. The political blogosphere and the 2004 U.S. election: divided they blog. In LinkKDD, 2005.

[2] L. Backstrom, C. Dwork, and J. M. Kleinberg. Wherefore art thou R3579X?: anonymized social networks, hidden patterns, and structural steganography. Commun. ACM, 54(12), 2011.

[3] S. Bhagat, G. Cormode, B. Krishnamurthy, and D. S. and. Class-based graph anonymization for social network data. PVLDB, 2(1), 2009.

[4] A. Campan and T. M. Truta. A clustering approach for data and structural anonymity in social networks. In PinKDD, 2008.

[5] J. Cheng, A. W.-C. Fu, and J. Liu. K-isomorphism: privacy-preserving network publication against structural attacks. In SIGMOD, 2010.

[6] G. Cormode, D. Srivastava, T. Yu, and Q. Zhang. Anonymizing bipartite graph data using safe groupings. PVLDB, 19(1), 2010.

[7] S. Das, • O. Egecioglu, and A. E. Abbadi. Anonymizing weighted social network graphs. In ICDE, 2010.

[8] A. G. Francesco Bonchi and T. Tassa. Identity obfuscation in graphs through the information theoretic lens. In ICDE, 2011.

[9] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis. Resisting structural re-identi_cation in anonymized social networks. PVLDB, 1(1), 2008.

[10] Y. Li and H. Shen. Anonymizing graphs against weight-based attacks. In ICDM Workshops, 2010.

[11] K. Liu and E. Terzi. Towards identity anonymization on graphs. In SIGMOD, 2008.

[12] L. Liu, J.Wang, J. Liu, and J. Zhang. Privacy preserving in social networks against sensitive edge disclosure. In SIAM International Conference on Data Mining, 2009.

[13] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. `-diversity: privacy beyond k-anonymity. In ICDE, 2006.

[14] MPI. http://socialnetworks.mpi-sws.org/.

[15] Y. Song, P. Karras, Q. Xiao, and S. Bressan. Sensitive label privacy protection on social network data. Technical report TRD3/12, 2012.

[16] Y. Song, S. Nobari, X. Lu, P. Karras, and S. Bressan. On the privacy and utility of anonymized social networks. In iiWAS, pages 246{253, 2011.

[17] L. Sweeney. K-anonymity: a model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5), 2002.

[18] C.-H. Tai, P. S. Yu, D.-N. Yang, and M.-S. Chen. Privacy-preserving social network publication against friendship attacks. In SIGKDD, 2011.

AUTHOR DETAILS :

First Author: ***Koya Yogitha*** received B.Tech Degree in Computer Science and Engineering from Swarna Bharathi College of Engineering in the year 2012. She is currently M.Tech student in the Computer Science and Engineering from MLR Institute of Technology. And her research interested areas in the field of Data mining, Cloud Computing and Mobile Computing.

Second Author: ***N Shirisha*** working as an Asst. Professor in MLR Institute of Technology, Dundigal, Ranga reddy.. He has completed his M.Tech CSE and he has 3 years of teaching experience. Her research interested area Networking.

Third Author: G Kiran Kumar is working as Associate Professor & HOD-CSE in MLR Institute of technology. He did M.Tech from Osmania University, Hyderabad, and submitted Ph.D from Nagarjuna University. His research areas include Data Mining, Spatial data mining, Software Engineering.