PROVIDING PRIVACY PROTECTION FOR SENSITIVE LABELS IN SOCIAL NETWORKS

Pushpa Hosamani[1], Parameshwarappa C M[2]

[1]PG Student, [2]Prof. and Head of the Dept.

[1]Dept. of Computer Science and Engineering, STJIT College, Ranebennur, Karnataka, (India)

[2]Dept. of Computer Science and Engineering, Visvesvaraya Technological University, Belgaum

ABSTRACT

As everyone needs privacy while uploading or sharing social network data which helps in social science research and in the business analysis. Hence, this paper is motivated by the recognition of the need for a finer grain and more personalized privacy in data publication of social networks. To gain privacy 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 his/her profiles he/she wishes to hide. The social networks are modelled 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 which has to be protected. We provide privacy protection algorithms which allow that graph data to be published in such a way such that an adversary who possesses information about a node's neighbourhood 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.

Keywords: Social Networks; Privacy Protection Scheme; Sensitive Labels; Data Mining

I. INTRODUCTION

Social network data which can be published of particular individual should always wants to gain its privacy. Means sensitive information of users/individuals among the social networks should be protected. The main challenge is to plan correct methods to publish social network data in a form that provides utility without compromising privacy. Previous research has proposed various privacy models with the corresponding protection mechanisms that prevent both accidental private information leakage and attacks by malicious adversaries. These early privacy models are mostly concerned with identity and link disclosure. The social networks are modelled as graphs in which users are nodes and social connections are edges.

Users entrust social networks such as Facebook, LinkedIn, etc., with a wealth of personal information such as their age, address, current location or political orientation. We treat these details and messages as features in the user's profile. 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 his/her profile he/she wishes to hide.

The social networks are modelled as graphs in which users are nodes and features are labels. Labels are considered either as sensitive or as non-sensitive. Figure1 is a labelled graph representing a small subset of such a social network. Each node in the graph represents a user, and the edge between two nodes represents the fact

that the two persons are friends. Labels presented to the nodes show the locations of users. Each letter represents a city name as a label for each node. Some individuals do not mind their address being known by the others, but some do, for various reasons. In such case, the privacy of their labels should be protected at data release. Therefore the locations are either sensitive (labels are in red italic in Figure 1) or non-sensitive.

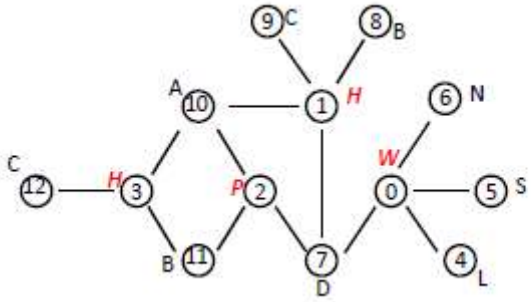

Fig.1. Example of the Labelled Graph Representing a Social Network

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 neighbourhood attack, in which an adversary finds out sensitive information based on prior knowledge of the number of neighbours of a target node and the labels of these neighbours. 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 he/ she can guess that the user is in H (Helsinki).

We present privacy protection algorithms that allow the data to be published in a way 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 *l*-1 other nodes. The probability to infer that any node has a certain sensitive label (we call such nodes sensitive nodes) is no larger than 1/*l*. For this purpose we design *l*-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.

II. BACKGROUND

Problem Definition: We model a network as G (V, E, L^s, L, and Γ), where V is a set of nodes, E is a set of edges, L^s is a set of sensitive labels, and L is a set of non-sensitive labels. Γ maps nodes to their labels, Γ: V → $L^s \cup$ L. Then we propose a privacy model, l-sensitive-label-diversity; in this model, we treat node labels both as part of an adversary's background knowledge, and as sensitive information that has to be protected. These concepts are clarified by the following definitions:

Definition 1. The neighborhood information of node v comprises the degree of v and the labels of v's neighbors.

Definition 2. (L-sensitive-label-diversity) For each node v that associates with a sensitive label, there must be at least l-1 other nodes with the same neighborhood information, but attached with different sensitive labels.

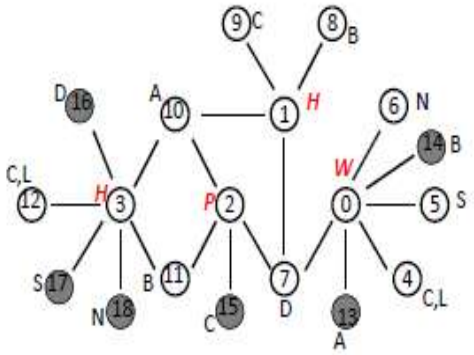

Fig.2. Privacy- Attaining Network Example

In example 1, nodes 0, 1, 2, and 3 have sensitive labels. The neighborhood information of node 0 includes its degree, which is 4, and the labels on nodes 4, 5, 6, and 7, which are L, S, N, and D, respectively. For node 2, the neighborhood information includes degree 3 and the labels on nodes 7, 10, and 11, which are D, A, and B. The graph in Figure 2 satisfies 2-sensitive-label-diversity; that is because, in this graph, nodes 0 and 3 are indistinguishable, having six neighbors with label A, B, {C,L}, D, S, N separately; likewise, nodes 1 and 2 are indistinguishable, as they both have four neighbors with labels A, B, C, D separately.

III. RELATED WORK

The first necessary anonymization technique in both the contexts of micro and network data involves in removing identification. This 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. l-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-identifcation 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, and 3].

To our knowledge, Zhou and Pei [25, 26] and Yuan et al. [23] were the first to consider modelling social networks as labelled 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 neighbourhood 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 immediate neighborhood structure with other k-1 nodes. In [26], they improve the privacy guarantee provided by k-anonymity with the idea of `-diversity, to protect labels on nodes as well. Yuan et al. [23] try to be more practical by considering users' different privacy concerns. They divide privacy requirements into three levels, and suggest methods to generalize labels and modify structure corresponding to every privacy

demand. Nevertheless, neither Zhou and Pei, nor Yuan et al. consider labels as a part of the background knowledge. However, in case adversaries hold label information, the methods of [25, 26, and 23] cannot achieve the same privacy guarantee. Moreover, as with the context of microdata, a graph that satisfies a k-anonymity privacy guarantee may still leak sensitive information regarding its labels [13].

IV. SCOPE OF THE PROJECT

The objectives of this project are as follows:

- The main objective of the algorithm that we propose is to make suitable grouping of nodes, and appropriate modification of neighbours' labels of nodes of each group to satisfy the *l*-sensitive-label-diversity requirement.
- It helps to investigate the protection of private label information in social network data publication.
- It experiments on both real and synthetic data sets confirm the effectiveness, efficiency and scalability of that approach in maintaining critical graph properties while providing a comprehensible privacy guarantee.
- We can publish the Non-sensitive data to every-one in social Network.
- We can post sensitive data to particular people and in the same way we can post non-sensitive data to everyone like ads or job posts.

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.

VI. ACKNOWLEDGMENT

I consider it is a privilege to express my gratitude and respect to all those who guiding me in the progress of my paper.

I wish my grateful thanks to **Prof. Parameshwarappa C M M Tech (Soft. Engg),** project guide and HOD, for his invaluable support and guidance. **Pushpa Hosamani**

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-identification 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,

[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. l-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.

[19] O. Tore, A. Filip, and S. John. Node centrality in weighted networks: generalizing degree and shortest paths. Social Networks, 32(3), 2010.

[20] W. Wu, Y. Xiao, W. Wang, Z. He, and Z. Wang. K-symmetry model for identity anonymization in social networks. In EDBT, 2010.

[21] X. Ying and X.Wu. Randomizing social networks: a spectrum perserving approach. In SDM, 2008.

[22] X. Ying and X. Wu. On link privacy in randomizing social networks. In PAKDD, 2009.

[23] M. Yuan, L. Chen, and P. S. Yu. Personalized privacy protection in social networks. PVLDB, 4(2), 2010.

[24] L. Zhang and W. Zhang. Edge anonymity in social network graphs. In CSE, 2009.

[25] B. Zhou and J. Pei. Preserving privacy in social networks against neighborhood attacks. In ICDE, 2008.

[26] B. Zhou and J. Pei. The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks. Knowledge and Information Systems, 28(1), 2010.

[27] L. Zou, L. Chen, and M. T. Ozsu. K-automorphism: a general framework for privacy-preserving network publication. PVLDB, 2(1), 2009

BIOGRAPHY

Pushpa Hosamani is a student pursuing her Master degree in Computer Science and Engineering department at STJ Institute of Technology, Ranebennur, Karnataka, India. Her research interests are Computer Science related aspects such as Data Mining technology, Java programming language.

Prof. Parameshwarappa C M, is HOD in the department of Computer Science and Engineering at STJ Institute of Technology, Ranebennur, Karnataka, India. He received his Master degree in Software Engineering. His research interests are related to Software Engineering.