Privacy to Social Network Data

Aniruddha Baitule[#1], Dr. M. B. Chandak[*2]

[#]Department of CSE, RTM Nagpur University M.Tech (4th Sem),WCEM,Nagpur

[1]akbaitule@gmail.com

[*]HOD Department of Computer Science & Engg. , RKNEC, Nagpur, India

[2]mbchandak@gmail.com

Abstract—

We know that there is a chance of privacy threat for publishing private date on social networking websites. So this paper gives you the need of privacy protection mechanism for publishing private data on social network. We propose a mechanism in which user can enter his public & private data on these websites without disclosing private information related to him. Now days use of social networking websites is tremendously increasing. Many social networking websites like Face book, LinkedIn, Whats Up, Twitter etc are present. On Social Networking sites data present in the form of graphs in which nodes represent number of users in network & labels represent information linked to that user or nodes. So while uploading some private information on these websites result in chances of privacy threat. So for that we require a privacy protection mechanism which can keep both private & public data on these websites. Thus we are giving privacy to social network & the advantage is that user can keep his both private & public data on these sites.

For giving the privacy to social network data we are focus on private information about user present on it. Some user don't want to disclose his identity, some don't want to disclose some labels present on his profile. So the aim is to provide protection mechanism to such sensitive labels present on the social networking websites. The adversary first collect background knowledge of network & from that he attack on private information present on network. So the solution is given by our algorithm. Working o algorithm is to insert noisy node within the current network & if the attempt made by an adversary to access private data then we are redirecting them to noisy or false information. In this way we are preserving privacy of the social network data.

I. INTRODUCTION

While publishing data on social network of a particular user there is a possibility of privacy threat. Therefore challenge is to find a method trough which data publishing will be done more securely. Various methods already present which gives us solution on private information leakage & an adversary attack on social data. These solutions are mainly concern with identity & link discloser. The data on social network is present in the form of graph in which node indicates number of user & labels indicate information linked up with user. So this paper gives us solution for need of finer grain & more personalized privacy.

User indicate there age, current location, present employer, occupation etc on social network sites such as twitter, what's up, face book. Some of them are act as public member & some as a private member. A particular user wish to enter which features are public or private. Now losing private information results in failure of model so we devise a method which will keep both public & private data without disclosing identity of user & private labels.

The data on social network is present in the form of graph in which node indicates number of user & labels indicate information

linked up with user. Users can denote each labels as sensitive or insensitive. Figure 1 indicate a social network labelled graph. In which node represents user & an edge between two nodes represent whether two nodes are friend or not. Labels on each node represent location of user i.e. in which city user is staying currently. Red mark on labels indicates location not to be discloses as it is private data.

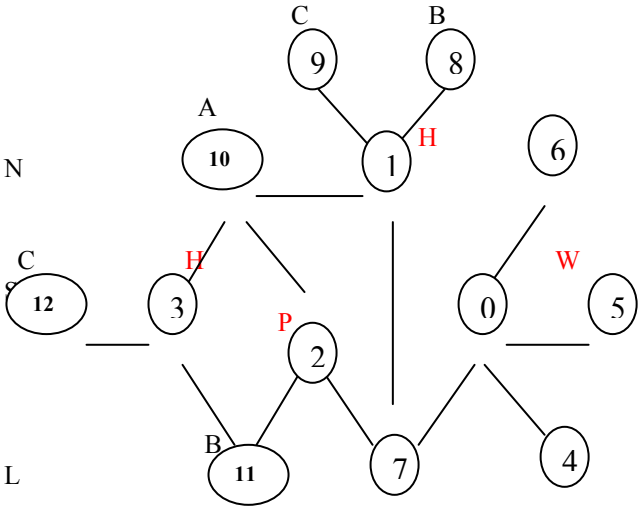


Fig 1. Example of the Labeled Graph representing Social Network

Privacy issue arise when these sensitive labels disclosing. Now the previous solution tells that if the data is sensitive then it should not enter on these social networks but such solution indicate incomplete & poor view of social network. When we are going to delete private labels from these social networks the data linked to that labels also get deleted automatically whether it may private or not. So the new approach is to invent such method which will prevent this information leakage & also ensuring us identities of user is protected. The neighborhood attack in which an adversary collect background information like number of neighbors of a node & labels of these neighbors and later on find the sensitive information. . For 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).

Our mechanism for privacy protection allows the graph data to represent in such a manner that an attacker can't access private labels & can't disclose identity of user. The algorithm transforms the original graph into a graph in which any node with sensitive labels is not distinguishable from other L-1 nodes. The probability to conclude that any node has a sensitive label is not larger than 1/L. For this we design L-diversity-like model, in which we consider node labels as a part of an adversary's background knowledge & as a sensitive information that has to be protected.

This algorithm provides privacy protection by keeping the original graph structure & properties such as degree distribution, density & clustering coefficient. We provide an efficient, reliable & effective solution than previous research work.

The rest of the paper is organized as follows. Section 2 reviews previous works in the area. We define our problem in Section 3 and propose solutions in Section 4. We conclude this work in Section 5.

II. Related Work

The first necessary anonymization technique in both the contexts of micro- and network data consists in removing identification. This naive technique has quickly been recognized as failing to protect privacy. For micro data, Sweeney et al. propose k-anonymity [17] to circumvent possible identity disclosure in naively anonymzed micro data. 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-identification by adversaries with external information that is formalized into structural queries (node refinement queries, sub graph 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 anonymzed 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 modelling social networks as labelled graphs, similarly to what we consider in this paper. To prevent re-identification attacks by adversaries with immediate neighbourhood structural knowledge, Zhou and Pei [25] propose a method that groups nodes and anonymzes the neighbourhoods 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 4 Sensitive Label Privacy Protection on Social Network Data immediate neighbourhood structure with other k-1 nodes. In [26], they improve the privacy guarantee provided by k-anonymity with the idea of l-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, 23] cannot achieve the same privacy guarantee. Moreover, as with the context of micro data, a graph that satisfies a k-anonymity privacy guarantee may still leak sensitive information regarding its labels [13].

III. Problem Defination

We model a network as G(V,E,L^s, L, Γ), where V is a set of nodes, E is s set of edges, L^s is a set of sensitive labels, L is a set of non-sensitive labels and Γ maps nodes to their labels, Γ: V→L^s U 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 neighbourhood information of node v comprises the degree of v and the labels of v's neighbours.

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 neighbourhood information, but attached with different sensitive labels.

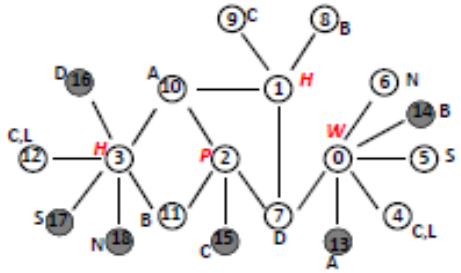


Fig. 1 Privacy-attaining Network Example

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.

IV. PROPOSED WORK

We planned a method which gives the protection to the sensitive labels present on these social networking sites and also if anybody makes an attempt to access these sensitive labels then we will redirect them to the noisy node. When in the network there

are sensitive labels present we inserting some noisy node in that network. So the intruder can't get an exact information because noisy node contain false or misleading information.

ALGORITHM

The main objective of the algorithms 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. We want to group nodes with as similar neighbourhood information as possible so that we can change as few labels as possible and add as few noisy nodes as possible. We propose an algorithm, Global-similarity-based Indirect Noise Node (GINN), that does not attempt to heuristically prune the similarity computation as the other two algorithms, Direct Noisy Node Algorithm (DNN) and Indirect Noisy Node Algorithm (INN) do. Algorithm DNN and INN, which we devise first, sort nodes by degree and compare neighbourhood information of nodes with similar degree. Details about algorithm DNN and INN please refer to [15].

ALGORITHM GINN

The algorithm starts out with group formation, during which all nodes that have not yet been grouped are taken into consideration, in clustering-like fashion. In the first run, two nodes with the maximum similarity of their neighbourhood labels are grouped together. Their neighbour labels are modified to be the same immediately so that nodes in one group always have the same neighbour labels. For two nodes, v1 with neighbourhood label set (LS_{v1}),and v2 with neighbourhood label set (LS_{v2}), we calculate neighbourhood label similarity (NLS) as follows:

$$NLS(v1,v2)= \left| LS_{v1} \cap LS_{v2} \right| / \left| LS_{v1} U\ LS_{v1} \right| \quad (1)$$

Larger value indicates larger similarity of the two neighbourhoods.

Then nodes having the maximum similarity with any node in the group are clustered into the group till the group has ` nodes with different sensitive labels. Thereafter, the algorithm proceeds to create the next group. If fewer than ` nodes are left after the last group's formation, these remainder nodes are clustered into existing groups according to the similarities between nodes and groups. After having formed these groups, we need to ensure that each group's members are indistinguishable in terms of neighbourhood information. Thus, neighbourhood labels are modified after every grouping operation, so that labels of nodes can be accordingly updated immediately for the next grouping operation.

This modification process ensures that all nodes in a group have the same neighbourhood information. The objective is achieved by a series of modification operations. To modify graph with as low information loss as possible, we devise three modification operations: label union, edge insertion and noise node addition. Label union and edge insertion among nearby nodes are preferred to node addition, as they incur less alteration to the overall graph structure.

Edge insertion is to complement for both a missing label and insufficient degree value. A node is linked to an existing nearby (two-hop away) node with that label. Label union adds the missing label values by creating super-values 6 Sensitive Label Privacy Protection on Social Network Data shared among labels of nodes. The labels of two or more nodes coalesce their values to a single super-label value, being the union of their values. This approach maintains data integrity, in the sense that the true label of node is included among the values of its label super-value. After such edge insertion and label union operations, if there are nodes in a group still having different neighbourhood information, noise nodes with non-sensitive labels are added into the graph so as to render the nodes in group indistinguishable in terms of their neighbours' labels. We consider the unification of two nodes' neighbourhood

labels as an example. One node may need a noisy node to be added as its immediate neighbour since it does not have a neighbour with certain label that the other node has; such a label on the other node may not be modifiable, as its is already connected to another sensitive node, which prevents the re-modification on existing modified groups. Sensitive node, which prevents the re-modification on existing modified groups.

Algorithm 1: Global-Similarity-based Indirect Noisy Node Algorithm
Input: graph $G(V,E,L,L^S)$, parameter l;
Result: Modified Graph G'

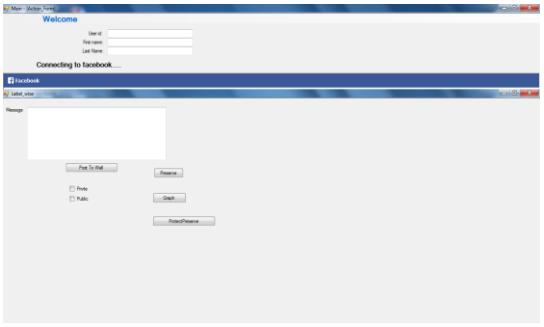
```
1  while V_left > 0 do
2      if | Vleft | ≥ l then
3          compute pairwise node similarities;
4          group G ←v1, v2 with Max_similarity;
5          Modify neighbors of G;
6          while |G| < l do
7              dissimilarity (V_left,G);
8              group G ←v with Max_similarity;
9              Modify neighbors of G without actually adding
               noisy nodes ;
10   else if | Vleft | < l then
11       for each v ϵ V_left do
12           similarity(v, G_s);
13           G_Max_similarity ←v;
14       Modify neighbors of G_Max similarity without actually
         adding noisy nodes;
15 Add expected noisy nodes;
16 Return G'(V`,E',L');
```

In this algorithm, noise node addition operation that is expected to make the nodes inside each group satisfy L-sensitive-label-diversity are recorded, but not performed right away. Only after all the preliminary grouping operations are performed, the algorithm proceeds to process the expected node addition operation at the final step. Then, if two nodes are expected to have the same labels of neighbours and are within two hops (having common neighbours), only one node is added. In other words, we merge some noisy nodes with the same label, thus resulting in fewer noisy nodes.

V. RESULTS

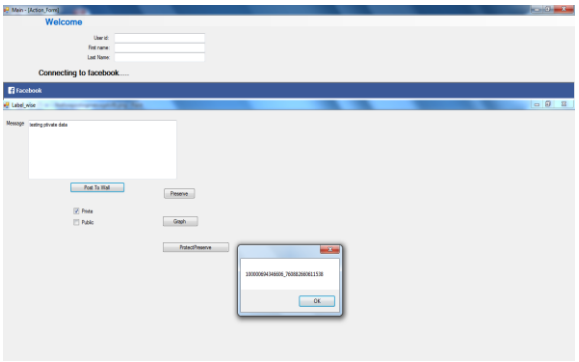
A. *Window for Posting messages on social networking websites*

A window which is use to post messages on social networking websites. After posting data will appear on sites. If somebody is getting our social networking site login-id & password & if he can delete our private labels then we can give them security by using preserve button.



B. *Window showing that message successfully appears on social networking websites*

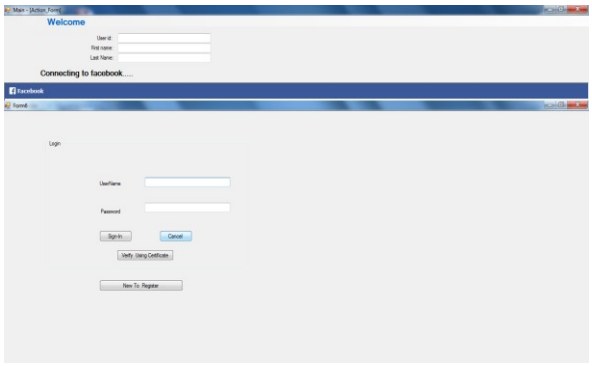
A window showing that by pressing post to wall button data is successfully posted on our social networking websites.



C. *Window showing how we can protect an attack on node by an intruder*

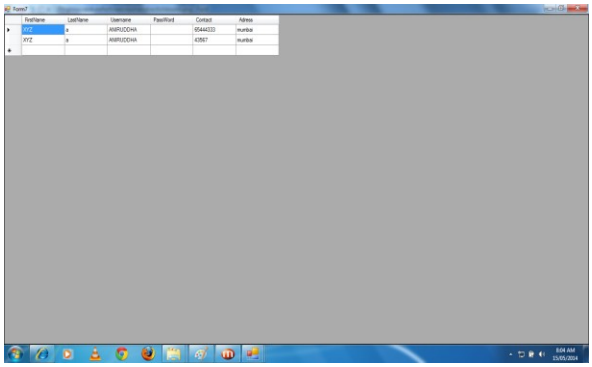
Another window which can use to protect attack on nodes by an intruder. Through this we can protect nodes or user.For that each user have to register first as its original data & fake data that will display if an intruder is trying to attack on node. In this login window we are providing facility for login by entering username & password or by entering username, password & certificate. In the first part user have to

enter username & password only. In second part user have to enter username & password along with certificate which is generated in c drive in certificate folder.



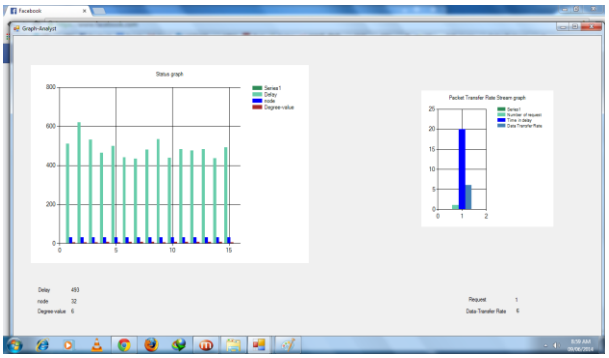
D. *Window showing how to redirect an intruder to noisy of false data*

If the user fail to enter correct username, password or username, password & certificate then we will redirect it to fake or noisy data. Figure shown below represent fake or noisy data.



E. *Graph Analysis*

Graph analysis shows the delay, node & degree value parameter in which if the delay is increasing tremendously then it will show that an intruder is making an attempt of attack on your social networking profile.



CONCLUSIONS

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.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered.

Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

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.Sensitive Label Privacy Protection on Social Network Data 9

[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, 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 anonymzed 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. `l-diversity: privacy beyond k-anonymity. In ICDE, 2006.

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