Review Paper-Social networking with protecting sensitive labels in data Anonymization

Miss.P.S.Kadam[#1], Prof.S.V.Patil[#2], Miss.A.A.Bhosale[#3], Mr.D.H.Dewarde[#4]

[#1]Computer Sci. and Engg.Department,MalajirajeBhosale Technical Campus, Islampur.
[#2]Computer Sci. and Engg.Department, AnnasahebDange College of Engineering, Ashta, Sangli.
[#3]Computer Sci. and Engg.Department,JaywantCollege of Engineering&Management,Killemachindra Gad
[#4]Computer Sci. and Engg.Department,MalajirajeBhosale Technical Campus, Islampur.

E-mail ID: [#1]pallavikadam14@gmail.com, #[2]siddheshwar.patil@gmail.com,
[#3]bhosale.anita11@gmail.com,[#4]dewarde.digambar73@gmail.com

[#1#2#3#4]Shivaji University, Kolhapur, Maharashtra, India.

***Abstract*--**The use of social network sites goes on increasing such as facebook, twitter, linkedin, live journal social network and wiki vote network. By using this, users find that they can obtain more and more useful information such as the user performance, private growth, dispersal of disease etc. It is also important that users private information should not get disclose. Thus, Now a days it is important to protect users privacy and utilization of social network data are challenging. Most of developer developed privacy models such as K-anonymity for protecting node or vertex reidentification in structure information. Users privacy models get forced by other user, if a group of node largely share the same sensitive labels then other users easily find out one's data ,so that structure anonymization method is not purely protected. There are some previous approaches such as edge editing or node clustering .Here structural information as well as sensitive labels of individuals get considered using K-degree l-deversityanonymity model. The new approach in anonymization methodology is adding noise nodes. By considering the least distortion to graph properties,the development of new algorithm using noise nodes into original graph. Most important it will provide an analysis of no.of noise nodes added and their impact on important graph property.

Keywords*-*Anonymization,Noise node,KDLD

*****

I. Introduction

The use of social network sites goes on increasing such as facebook ,twitter and linkedin .By using this, users find that they can obtain more and more useful information such as the user performance, private growth, dispersal of disease etc. It is also important that users private data should not get disclose. Thus, how to protect individual privacy and at the same timepreserve the utility of social network data becomes a challenging.Here consider a graph model where each vertex in the graph is associated with a sensitive label. A variety of privacy models as well as anonymization algorithms havebeen developed (e.g.kanonymity,l-diversity,t-closeness). In tabular microdata, some ofthe nonsensitiveattributes, called quasi identifers, can be used to reidentifyusers data andtheir sensitive attributes or information. When circulating social network data, graph structures are alsoissued with corresponding socialrelationships.

A structure attack is an attack that uses the structure information or data, that is the degree and the subgraph of a node, to recognize the node. To prevent structure attacks,a published graph should fulfill k-anonymity. The aim is to publish a social graph, which always has minimum k candidates in different attack scenarios in order to protect privacy. A k-degree anonymity model is used to prevent degree attacks. A graph is k-degree anonymous if and only if for any node inthis graph, there exist at least k -1 other nodes with thesame degree.

If an opponent knows that one person has three friends in the graph, he can directly know that node 2 is that person and the related attributes of node 2 are discovered. k-degree anonymity can be used to inhibit such structure attacks. Though, in many applications, a social network where each node has sensitive attributes should be circulated. For example, a graph may contain the user salaries which are sensitive label. In this case, only k-degree is not sufficient to prevent the inference of sensitive attributes of individuals. The l-diversity should be adopted for graphs. In this work, selecting the degree-attack, one of the famous attacks methods to show how to design mechanisms of protecting both identities and sensitive labels.

Current approaches for protecting graph privacy can be classified into two categories: clustering[7] and edge editing. The method clustering is to merge a subgraph to form one super node, which is inappropriate for sensitive labeled graphs after theyget merged into one super node, the node-label relations have been vanished. Edge editing methods keep the nodes as it is and only add/delete/swap edges. However, edge editing may largely destroy the characteristics of the graph. The distance characteristics get changed substantially by connecting two faraway nodes or deleting the bridge link between two communities in the edge editing method. Miningover these data might get the wrong conclusion about how the salaries are distributed in the the world. Therefore, solely relying on edge editing may not be a good solution to preserve data utility[1].

While considering the above problem, in this work the basic idea is to maintain important graph properties, like distances between nodes by adding certain "noise" nodes into a graph. According to noise adding concept will concern the following observation.Small degree vertices in the graph are used to hide added noise nodes from being reidentified for that purpose widely used Power Law distribution to satisfy social networks.By adding noise nodes, some graph properties will be better maintainthan edge-editing method. In this privacy preserving goal is to prevent an attacker from reidentifying a user and finding the fact that a certain user has a specific delicate value. After considering above observations, k-degree-l-diversity (KDLD) model for securelyissuing a labeled graph, and then develop corresponding graph anonymization algorithms with the least distortion to the properties of the original graph, such as degrees and distances between nodes[2].

Scope-

- Privacy is one of the major concerns when publishing or sharing social network data forsocial science research and business analysis.

- Privacy models similar to k-anonymity to prevent node reidentification through structureinformation. However, even when these privacy models are enforced, an attacker maystill be able to infer other private information if a group of nodes largely share the samesensitive labels.
- Proposed approach definethe k-degree-l-diversity anonymity model that considers the protectionof structural information as well as sensitive labels of individuals.

- Proposed method will produce anonymization methodology based on adding noise nodes.It develop a new algorithm by adding noise nodes into the original graph with the considerationof introducing the least distortion to graph properties[1].

II. System architecture

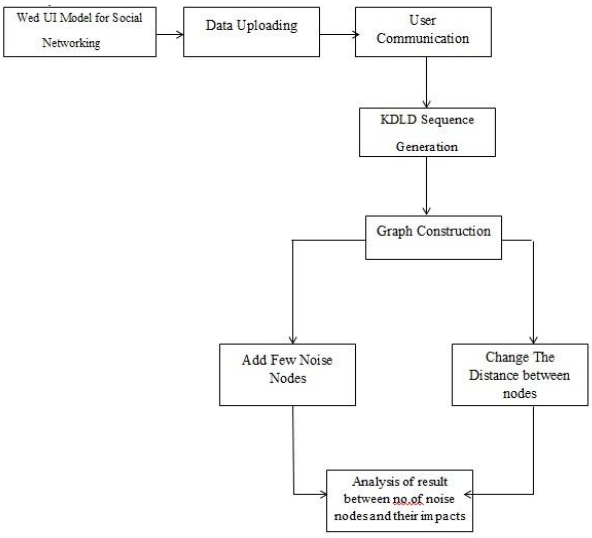
Wed UI Model for Social Networking
Data Uploading
User Communication
KDLD Sequence Generation
Graph Construction
Add Few Noise Nodes
Change The Distance between nodes
Analysis of result between no.of noise nodes and their impacts

Fig. shows the system architecture. K-degree anonymity with l-diversity to prevent not only the reidentification of individualnodes but also the revelation of a sensitive attribute associated with each node. If the k-degreel-diversity constraint satisfies create KDLD graph. A KDLD graph protects two aspects of eachuser when an attacker uses degree information to attack a novel graph construction techniquewhich makes use of noise nodes to preserve utilities of the original graph. Two key propertiesare considered: Add as few noise edges as possible. Change the distance between nodes as lessas possible. The noise edges/nodes added should connect nodes that are close with respect tothe social distance. There exist a large number of low degree vertices in the graph which couldbe used to hide addednoise nodes from being re-identified. By carefully inserting noise nodes,some graph properties could be better preserved than a pure edge-editing method[3][4].

III. Methodology

1.Web UI Module for Social Networking

It is an web user interface module. It contains all the user related information. It is themodule through which user has connection with each other. In this module the employeedata is collected.In this module, Users are having authentication and security to access thedetail which is presented in the ontology system. Before accessing or searching the detailsuser should have the account in that otherwise they should register first.

2.Data Uploading and user communication

Each employee has unique Id, Name and Sensitive Label Salary. It contains uploadingof user information such as their unique Id, name, sensitive attributes, images, own profileinformation etc. This modules collect all the information of user and loaded to the system database. Based on the employee data construct the SocialNetwork Graph. In this modulethere is communication between various user. Number of user can communicate with eachother by sharing there personal information. Some of them can upload image or give thecomment on that status[5].

3.KDLD sequence generation

KDLD sequence is generated by combining k-anonymity and l-diversity anonymizationtechniques. The preprocessed metadata is k degree anonymized in which every tupleshould be different from at least k-1 other tuples in accordance with their quasi-identifiers(QIDs). The k-anonymized dataset is again anonymized by applying l-diversity techniqueto provide diversification in the equivalence class[6].

4.Graph Construction

By using following two perspectives graph is constructed based on the new KDLD sequence generation.

(a) Add Few Noise nodes

(b) Change the distance between nodes

Graph construction module includes the following steps.

(A)Neighborhood Edge Editing: It is the concept of adding new edges between the nodes.Neighborhood rule is followed in this approach i.e., to add edge between two neighbors,so that the path the nodes would be short as possible.

(B) Adding Node Decrease Degree: For any node whose degree is larger than its targetdegree in Pnew, then decrease its degree to the target degree by making using of noise nodes.

(C) Adding Node Increase Degree: For any node whose degree is smaller than its targetdegree in Pnew, then increase its degree to the target degree by making using of noise nodes.

(D) New Node Degree Setting: For any noise node, if its degree does not appear in Pnew,does some adjustment to make it has a degree in Pnew. Then, the noise nodes areadded into the same degree groups in Pnew[7].

(E) New Node Label Setting: The last step is to assign sensitive labels to noise nodesto make all the same degree group still satisfy the requirement of distinct l-diversity.Since in each same degree group, there are already l distinct sensitive labels in it, it isobviously the new added noise nodes can have any sensitive label. Use the followingway to sensitive label to a noise node n: suppose u is the original node in G which n iscreated for. Then randomly find a label from the direct neighbors of u in the originalgraph G[8].

5 Analysis of result between no. of noise nodes and their impacts

This module represent the analytical results to show the relationship between the number of noise nodes added and their impacts on an important graph property. This work will be compared with the noise node adding algorithms with previous work using edge editing only. Inthis work different datasets will be considered i.e. Live journal social network or Wiki vote network. Another interesting direction is to consider how to implement this protection model, where different publishers publish their data independently and their data are overlapping. Average Change of Sensitive Label Path Length(ACSPL) and Remaining ratio of top influentialusers(RRTI) will be calculated. ACSPL: In order to measure the connections between any twosensitive labels (including the same label), we define average path length between any two labelsl1 and l2 as:

$$ACSPL_{G,G'} = \frac{\sum_{\forall l_1,l_2} Abs(APL_{G,(l_1,l_2)} - APL_{G',(l_1,l_2)})}{\binom{M}{2} + M},$$

RRTI: One important data mining task on a graph are to find the top influential users (experts) in it.The larger RRTI is, the better the published graph preserves the information in the

$$RRTI = \frac{|INF_G \cap INF_{G'}|}{|INF_G|}.$$

IV. Conclusion

From above discussion, here conclusion is that, it must to hide the sensitive data from thirty party applications . Propose a privacy protection scheme that not only prevents the revelation of identity of users but also the disclosure of selected features in users' profiles. In this k-degree-l-diversity model forprivacy preserving social network data publishing.The main difference between previous and this system is mainly focus on noise node addingalgorithm to construct a new graph from the original graphwith the constraint of introducing fewer distortion to theoriginal graph. Protocols should be designed to help thesepublishers publish a unified data together to guarantee theprivacy.

References

[1] Mingxuan Yuan, Lei Chen, "Protecting Sensitive Labels in Social Network DataAnonymization ",IEEE transaction on knowledge and data engineering, Vol. 25, No. 3,March 2013.

[2] Chongjing Sun, Philip S. Yuz, Xiangnan Kong and Yan Fu, "Privacy Preserving SocialNetwork Publication Against Mutual Friend Attacks," arXiv:1401.3201v1 [cs.DB] 11 Oct2013

[3] Mr. A.Stalin Irudhaya Raj, Ms. N.Radhika, "Securing Sensitive Information in Social NetworkData Anonymization ," A.Stalin Irudhaya Raj et al, International Journal of ComputerScience and Mobile Applications, Vol.2 Issue. 1, January- 2012.

[4] S.Charanyaa, K.Sangeetha , "Strategies for Knowledge Based Attack Detection in GraphicalData Anonymization International Journal of Advanced Research in Computer andCommunication Engineering Vol. 3, Issue 2, February 2012.

[5] Lijie Zhang andWeining Zhang , "Privacy Protection of Social Network Graphs," 2010.

[6] J. Cheng, A.W.c. Fu, and J. Liu , "K-Isomorphism: Privacy Preserving Network Publicationagainst Structural Attacks" 2010.

[7] Xiaoyong Liu,W. Bruce Croft, "Cluster-Based Retrieval Using Language Models ," 2007.

[8] W. Eberle and L. Holder,, "Discovering Structural Anomalies in Graph-Based Data," Proc.IEEE Seventh Intl Conf. Data Mining Workshops, 2007.