Paper ID: CSEIT18

SOCIAL NETWORKING WITH PROTECTING SENSITIVE LABELS USING ANONYMIZATION METHODOLOGY

Ms.Pallavi S.Kadam
Computer Sci. and Engg. Dept. Annasaheb Dange College of Engg. Ashta ,Sangli ,India
Siddheshwar V. Patil
Information Technology Dept Annasaheb Dange College of Engg. Ashta ,Sangli ,India
Anita A. Bhosale
Computer Sci. and Engg. Dept. Annasaheb Dange College of Engg. Ashta ,Sangli ,India

Abstract--The use of social network sites goes on increasing day by day e.g. wiki vote, live journal social network, twitter, LinkedIn network. By victimization these social networks, users get a lot of helpful data of alternative user's like the user performance, non-public growth, spread of sickness, salaries etc. it's additionally vital that user's non-public data shouldn't get reveal. Thus, nowadays it's essential to safeguard user's privacy and utilization of social network information. the bulk of developer developed privacy models like K-anonymity for protecting node or vertex reidentification in structure data. User's privacy models get forced by alternative user, if a bunch of node for the most part shares similar sensitive labels then new users simply establish recent user's information, so solely structure anonymization technique isn't entirely protected or helpful. There are many previous approaches like edge editing or clustering of nodes during these paper lots of focuses on structural information furthermore as sensitive labels of individuals gets thought-about K-degree l-diversity anonymity model. The advance technology in anonymization methodology is adding noise nodes. By considering the minimum changes to original graph properties, the event of recent algorithmic rule adding noise nodes into original graph. Most important it will provide associate analysis of no. of noise nodes a lot of and their impact on very important graph property

Keywords- Anonymization, Noise node, KDLD

I. INTRODUCION

The use of social network sites goes on increasing such as LinkedIn, twitter and face book .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 time preserve the utility of social network data becomes a challenging. Here think about a graph model wherever every vertex within the graph is related to a sensitive label. a range of privacy models moreover as anonymization algorithms are developed (e.g. k-anonymity, l-diversity, t-closeness). In tabular small knowledge, some typically non sensitive attributes, referred to as similar identifiers, want to reidentify user's knowledge and their sensitive attributes or data. Once current social network knowledge, graph structures are issued with corresponding social relationships.

A structure attack is AN attack that uses the structure info or knowledge that's the degree and therefore the sub graph of a node, to acknowledge the node. to stop structure attacks revealed graph ought to 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 namelessness model is employed to stop degree attacks. A graph is k-degree anonymous if and only if for any node in this graph, there is at least k -1 other node with the same 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 several applications, a social network wherever every node has sensitive attributes ought to 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 and edge editing. The method clustering means to merge a sub graph to form one super node, which is unsuitable for sensitive labeled graphs after that they get merged into one super node, the node-label relations have been vanished. Edge editing methods carry on the nodes as it is and only add or delete or swap edges. Whereas, 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. Mining over these data might get the wrong conclusion about how the salaries are distributed in the the world. Therefore, only 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 reidentify for that purpose widely used Power Law distribution to satisfy social networks. By adding noise nodes, some graph properties will be better maintained than 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 securely issuing 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 among the key considerations once commercial enterprise or sharing social network knowledge for scientific discipline analysis and business analysis.
- Privacy models like k-anonymity to stop node reidentification through structure data. However, even once these privacy models area unit enforced , Associate in Nursing assaulter should be able to infer different personal data if a bunch of nodes for the most part share a similar sensitive labels.
- Proposed approach defines the k-degree-l-diversity anonymity model that considers the protection of structural data further as sensitive labels of people.
- Proposed technique can turn out anonymization methodology supported adding noise nodes. It develops a brand new algorithmic rule by adding noise nodes into the initial graph with the thought of introducing the smallest amount distortion to graph properties [1].

II. SYSTEM ARCHITECTURE

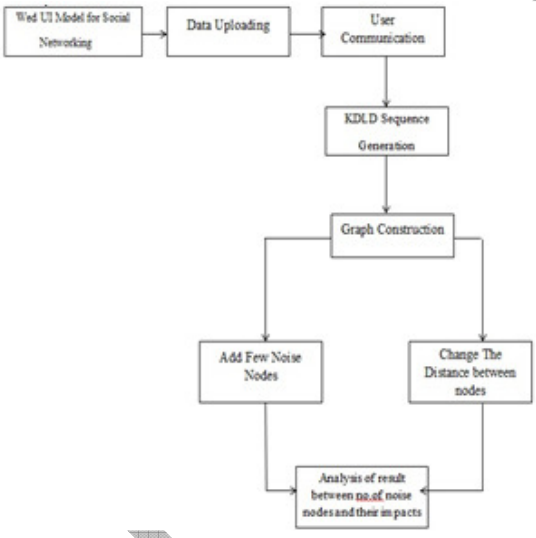

Fig.1 System Architecture

Fig. shows the system architecture. K-degree anonymity with l-diversity to prevent not only the reidentification of individual nodes but also the revelation of a sensitive attribute associated with each node. If the k-degree l-diversity constraint satisfies create KDLD graph. A KDLD graph protects two aspects of each user when an attacker uses degree information to attack a novel graph construction technique which makes use of noise nodes to preserve utilities of the original graph. Two key properties are considered: Add as few noise edges as possible. Change the distance between nodes as less as possible. The noise edges/nodes added should connect nodes that are close with respect to the social distance. There exist a large number of low degree vertices in the graph which could be used to hide added noise 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

A. Web UI Module for Social Networking

It is an web user interface module. It contains all the user related information. It is the module through which user has connection with each other. In this module the employee data is collected. In this 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.

B. Data Uploading and user communication

Each employee has unique Id, Name and Sensitive Label Salary. It contains uploading of user information such as their unique Id, name, sensitive attributes, images, own profile information etc. This module collects all the information of user and loaded to the system database. Based on the employee data construct the Social Network Graph. In this module there is communication between various users. Number of user can communicate with each other by sharing their personal information. Some of them can upload image or give the comment on that status [5].

C. KDLD sequence generation

KDLD sequence is generated by combining k-anonymity and l-diversity anonymization techniques. The preprocessed metadata is k degree anonymized in which every tuple should be different from at least k-1 other tuples in accordance with their quasi-identifiers (QIDs). First we create the sensitive degree sequence e.g. [(2,4,80), (3,4,100) (1,2,100) (3,2,80) (4,2,60) (5,2,60) (7,2,60)] it is 2D2D sequence. All the corresponding nodes in same group should be adjusted to have same degree. To generate KDLD sequence K-L Based algorithm is used. The algorithm tends to put same degree into same group to reduce the degree change [6].

D. 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.

1 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.

There are three cases according to conditions, algorithm proceeds. Following are steps of algorithm.

Algorithm 1- Neighborhood Edge Editing

For each node u need to increase degree do

d=u's degree ;

d'=u'target degree;

for (i=0; i<d'-d; i++) do

find v,w has link (u,v),(u,w) and v need to decrease degree;
if v,w exist then
Remove link(v,w)
Add link (u,w);
Else
Break;
For each node u need to increase degree do
For each node v need to increase degree do
If u,v are 2 hop neighbor then
If u,v do not have link then
Add link(u,v)
For each node u need to decrease degree do
For each node v need to decrease degree do
If u,v have link then
Rmove link (u,v);
If~(u,v are 2 hop neighbour) then
Add link (u,v);

2 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.

Algorithm 2-Adding node Decrease Degree

For every node that need to decrease degree do
d=u's degree;
target =target degree of u;
while true do
Select a sensitive value s from u's original hop neighbour ;
Create a new node n with sensitive value s;
d'=1;
target_new=Select_Closest_Degree_InGroup(d+2-(target degree));
Connect node u with n;
d=d+1;
While true do
Random select a link (u,v) while it is in G
Delete link (u, v) Create link (n, v)
d'=d'+1;
d= d-1;
If d'= =target_new V d = =target then
Break;
If d= = target then
Break;

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

Algorithm 3-Adding node Increase Degree

For each node u need to increase degree do
For i=0; i<increase_num; i++
Create a new node with n;
Connect node u with n;
For every node v that is one or two hop neighbor of node u do
If v needs to increase its degree then
Connect node v with n
While n's degree is not in sensitive degree sequence of published graph and n's degree >min group degree do
Remove the last connection created to n ;
i=i-1;

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

Algorithm 4- New Node Degree Setting

Select pair of noise nodes that each pair of nodes are within 3 hop to each other ;
Build a link of each pair;
For every node n has even degree do
Select an even degree target_new (n.d<target_new) in KDLD seq.
For every node n has odd degree do
Select an odd degree target_new (n.d<target_new) in KDLD seq.
For each noise node n do
While n.d !=target_new do
Find a link(u,v) with minimum dis(u,n)+dis(v,n)/2 in current graph where u and v are not connected to n;
Remove link(u,v);
Add link(n,u);
Add link (n,v);

5 New Node Label Setting: The last step is to assign sensitive labels to noise nodes to 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 is obviously the new added noise nodes can have any sensitive label. Use the following way to sensitive label to a noise node n: suppose u is the original node in G which n is created for. Then randomly find a label from the direct neighbors of u in the original graph G [8].

Algorithm 5- New Node Label Setting

For every noise node n do
u is the node G which n is created for ;
Randomly select a node v where there exists link (u,v) in G
Set n's sensitive label as v's sensitive label;

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

This module represents 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. In this 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 influential users (RRTI) will be calculated. ACSPL: In order to measure the connections between any two sensitive labels (including the same label), we define average path length between any two labelsl1 and l2 as:

$$ACSPL_{G,G'} = \frac{\sum_{\forall l1,l2} Abs\left(APL_{G,(l1,l2)} \quad APL_{G',(l1,l2)}\right)}{\binom{M}{2} + M}$$

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. EXPERIMENTS AND RESULTS

Following fig. shows the results. For testing purpose we are taking only 10 nodes. It also has edges connected to each other.

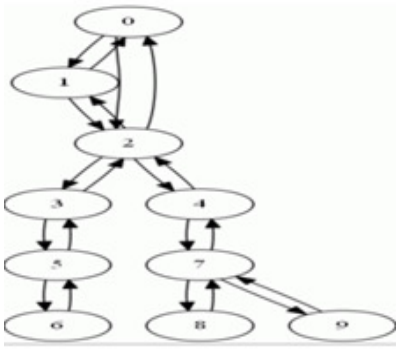

Fig.2 Original graph

Fig.2 shows the original graph, there are 10 nodes and edges connected to each other. The dataset is shown in fig.3

Fig.3 Generation of Sequence and KDLD degree

Fig.3 shows dataset table .Firstly calculates Sensitive degree sequence that means the corresponding nodes in same group should be adjusted to have same degree. Second calculate KDLD Sequence by using KL algorithm, Here assume value as K=2and L=2.

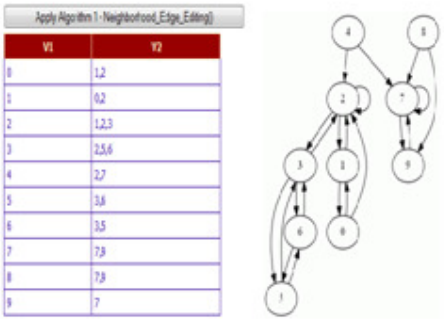

Fig.4 Result of Algorithm 1

Above fig.4 shows result of algorithm 1 i.e. Neighborhood Edge Editing. These algorithms checks each node and its neighbor, if there is need then it will add new edge otherwise ignore.

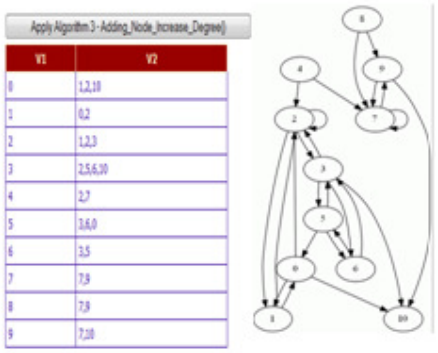

Fig.5 Result of algorithm 3

Above fig.5 shows result of algorithm 3 i.e. Adding node decrease degree. Here algorithm adds the new node that is called as noise node .Previously it has 0 to 9 nodes by applying algorithm 3 it will add node 10 as noise node. Here some node's degree gets changed. This is done because attacker cannot find its original degree

There are some more results, shown by following figures. Fig 7 shows APL result of live journal social network which has 95 edges and 4000 nodes respectively. In terms of changing k values using different graph construction algorithm. The straight line shows the value for original graph and second shows the values for anonymized graph. That means the algorithm which are used in this paper performs much better.

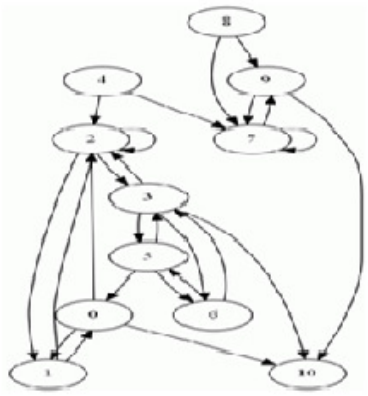

Fig.6 Anonymize Graph

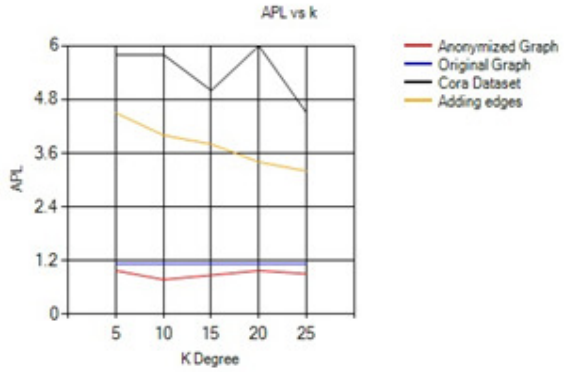

Fig. 7 Average Path length for different k

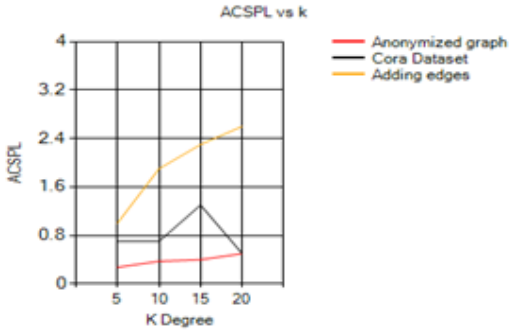

Fig.8Average Change of sensitive label path length

Fig.8 shows the ACPSL results. The less the ACSPL value is, better a graph construction algorithm works.

Fig.9 shows percentage of average label distribution change. This change includes the labels that only appear a small number of times in original graph.

Here record the running time of algorithm for different k in Fig. 10.From the result it is observed that running time is less than 7000 ms.

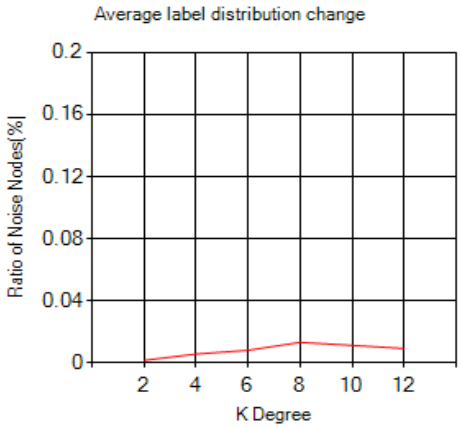

Fig.9 Change of label distribution

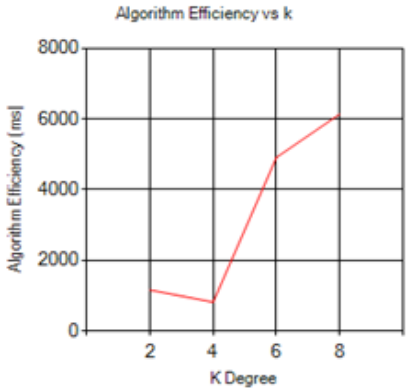

Fig.10 Algorithm efficiency

V. CONCLUSION

In this, a work proposed on k-degree-l-diversity model for protecting sensitive label in social network. In order to accomplish the requirement of k-degree-l-diversity, a work is done on designing a noise node adding algorithm to construct a new graph from old graph i.e. anonymized graph with the constraint of introducing fewer distortions to the original graph. The main difference between previously worked system and this system is that it focuses on noise node adding algorithm. Our extensive experimental results demonstrate that the noise node adding algorithms can achieve a better result than the previous work using edge editing only. It is a motivating way to study clever algorithms which can reduce the number of noise nodes if the noise nodes contribute to both anonymization and diversity.

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.