Anonymization Methodology for Sensitive Labels Protection in Social Network

Ms.B.Rashmi[1], Mrs.J.Suganya[2]

[1]**Post -Graduate Student, Department of Computer Science, SCAD Engineering College, Cheranmahadevi**

[2]**Assistant Professor, Department of Computer Science, SCAD Engineering College, Cheranmahadevi**

Abstract

Privacy preservation is the major problem when sharing data's in social networks. Various Privacy models are developed avoid node reidentification. Yet an attacker may still hack users private information, if nodes largely share the same type of sensitive labels. In this paper, we propose a scheme namely K-degree-L-diversity model useful for preserving social network datas.Ouranonymization methodology based on addition of noise nodes into the original social graph which reduces error rate and perform better results.

***IndexTerms :** Socialnetworks, privacy preservation, Anonymization*

1. Introduction

Data mining refers to Knowledge Discovery in Databases. The data mining process is the extraction of information from various data sets and transform to an understandable manner. A social network is a social graph made up of actors such as individuals or organizations and connections. A social network service consists of a representation of each users, social links and variety of additional services. Most social network services provide means for users to interact over the Internet, such as e-mail and messaging. Social network sites are varied and they incorporate new information and communication tools. The major drawbacks of social networks opens up the possibility of hackers to commit fraud and increases the risk of people falling prey to outline scams resulting in data or identity theft and potentially results in lost productivity. It refers to confidentiality of employer trade secrets and their private information. In order to provide security to social network users our algorithms issue anonymized views of the graph with significantly smaller information losses and analyze their privacy and communication complexity. Formally, in this social networks always represented as a graph, which we refer to as the social graph. The node of such a graph represents an actor and the edges represent ties between those actors.

II. Surveyon Anonymization Techniques

Anonymization techniques prevents node replication attacks. The goal is to publish a social graph, in order to protect privacy. Recent approaches for protecting social graph privacy are edge editing and clustering method. Edge-editing method [4],[6],[10] keep the nodes unchanged and by increasing or decreasing or by swapping edges of original graph. The edge editing method substantially changes the distance of node properties by linking faraway nodes together or breaking the bridge link of two communities. Clustering method which often merges a sub graphs to super nodes. Otherwise grouping of "similar" nodes as super nodes. The super node represents a "cluster."Then linking between nodes are refers as the edges in between super nodes called "super edges." Each super edge describes more edges in the original graph.

Graph based anomaly detection which introduces a model for calculation of regularity graph with rigorous applications to anomaly detection. [9]Graph based anomaly detection refers anomalous substructure detection looks unusual substructures within a entire graph. Secondly anomalous sub graph detection, in which graph is partitioned into set of vertices and it is tested for unusual patterns. Anomalous sub graph and anomalous substructure detection is more important for detecting unusual patterns. In addition conditional entropy measures both regularity and anomaly detection.

The novel technique anatomy which releases sensitive data.[12].This approach describes protection of privacy and various correlation of micro data. In addition they introduce linear time algorithm for generating anatomized tables and by decreasing the error reconstruction of the micro data. This innovative technique developed anatomy, preserves privacy and correlation in Microdata.

[3]This paper describes notion of Distance-based outliers. Specifically, Outlier detection can be efficiently created for large datasets and it is very important knowledge discovery task. These algorithms always support datasets with many more number of attributes. Secondly they performs optimized cell based algorithm that has a linear complexity. Cell based algorithms made for large resident datasets and also made a challenge that no data page is read more number of times and this scheme turns unsatisfied because there is no limit on size of the datasets or various numbers of dimensions in social network graphs.

Simple attacks of a K-anonymized dataset have few subtle, but major privacy problems. [7]Firstly they tell an attacker can discover sensitive attributes values. Secondly attackers often have background knowledge. Meanwhile it does not guarantee privacy against hackers. This Framework provides stronger privacy guarantees and describes well represented values for sensitive attributes.

The greedy privacy algorithm introduces that structural information loss measure quantifies loss of information because of edge generalization process. [2] Greedy privacy algorithm can be user balanced that are monitored by other data holders. The Main drawback of this approach denotes that when nodes of similar groups are merged into super node and so relations of various nodes have been lost.

Perturbation strategies which refer to Gaussian randomization multiplication and Greedy perturbation algorithm focused on graph theory. [5]The perturbation strategies having same nearest paths and lengths often close to original graph of social network. Perturbation strategies which does not perturb edges in which social network structure changes over time.

Neighborhood attacks often having back ground knowledge of target victim and connection among neighbors. The victim re-identified even victim's sensitive information is preserved by various anonymization techniques. [14]Neighborhood attacks are now in practice. However social networks can answer various aggregate queries. In this paper they are referring only one neighborhood and it is much better to focus on d neighborhoods (d>1) are protected.

Eigen values of networks are intimately linked to many topological features. [12]In this paper, they organized one spectrum randomization approach which automatically improves the graph randomization methods. The utility of graph preserves more privacy protection. Graph spectrum relates many real graphs and provides a perspective edge randomization.

A novel approach for anonymizing social network data models offers network structure by allowing samples from that model. [5] This scheme guarantees anonymity by preserving estimation of a wide variety of social network measures. Generalized graph splits the nodes and showed that wide range of graph analysis can be measured accurately and by protecting against risk reidentification.

Graph based anonymity notion prevents the identity disclosure of users profiles in which an attacker often have certain prior knowledge. However this method unsuits for social network graphs because in a graph nodes and edges are being related to each other. [4]A single modification of an edge and node can spread to entire social network. Finally, measuring the utility of a graph becomes more typical. They are not aware of effective metrics of various information loss occurred by the modified nodes and edges.

A new set of various techniques for anonymizing social network data based on merging the entities into classes and by mapping the entities and nodes often denoted them in anonymized target graph. [1]Anonymization techniques are being a challenge to attackers with larger background knowledge information. Yet turned unsatisfied because it has lower utility and less graph structure is here revealed.

The framework presentation for analyzing privacy preservation develops a new re-identification algorithm for target of various anonymized social network graphs. [8] Our Deanonymizing algorithm is based on network topology, does not contain creation of more number of dummy "Sybil" nodes. Existing defenses works between the overlapping of target network and adversary's information. A generic reidentification algorithm showed that it can successfully monitors and de-anonymize lot of users in anonymous social network graph. Since human names has not unique identity, this algorithm having overlap problem in membership.

Random link attacks (RLAs) performs multiple false identities and creates interactions among various users profiles to attack regular users of social networks. We have showed that RLA attackers can be splitted by their spectral characteristics [11]. Random link attack is a special collaborative attack. The malicious user has complete control by breaking nodes and captures them to attack a more number of randomly chosen victim nodes. Our spectrum detection approach works when hackers choose random victims or by attacking few victims while performing their collaborative attacks.

The state of analyzation of privacy protection in social network graphs describes effective anonymization attacks to protect from hackers. [10]In this paper, starclique, a minimal graph required k-anonymity, where user is identified for all possible contributions of data objects. The identification of social intersection attack can compromise users to identify shared objects relying on social graph topology.

III.ProposedWork

Finally in our proposed approach, we are developing KDLD sequence for target node creation of social network graphs. Given a graph G and its degree sequence consists of triplet namely node position, degree and sensitive labels.

KDLD SEQUENCE GENERATION: Given the sensitive degree sequence P and two integers k and L, computes a KDLD sequence. To obtain a new KDLD sequence, same group nodes are needed to be modified for next graph construction process. We further employ two algorithms:

- K-L BASED
- L-K BASED

The algorithms keeping the nodes of similar degrees to same group to reduce node reidentification process. Algorithm K -L-BASED chooses firstly K elements in original social graph and by monitoring the next element into current group until L-diversity constraint is satisfied.

Cnew: The costofdeveloping a newgroup forthenextk elements.

Cmerge:Thecostofmergingthenextelementintothecurrent group.

In this way, target node generation can be created and after that graph construction process is to be generated as follows.

Graph construction:

Neighborhood_Edge_Editing(): Neighborhood operation describes by adding or by deleting the nodes and edges in the KDLD sequence generation. By doing this modification sensitive labels are being protected from hackers.

Adding_Node_Decrease_Degree(): If the node degree is larger than target KDLD sequence generation node, we need to decrease the degree of node by breaking the links between two hop neighbors and by making a direct links to noise nodes.

Adding_Node_Increase_Degree(): If the node degree is smaller than target KDLD sequence generation node, we need to increase the degree of node by connecting the links between two hop neighbors and by breaking a direct links to noise nodes.

New_Node_Degree_Setting(): This operation describes by assigning degrees to noise nodes. Suppose whose noise node degree is an even number, we select an even degree or if it is odd degree we have to assign odd degree for target nodes.

New_Node_Label_Setting(): The final step is to assign labels to newly modified social network graphs. By doing this it is more helpful for preserving distances between labels and remaining labels in social network graphs.

IV.Experimental Results

Original Graph of Social Network

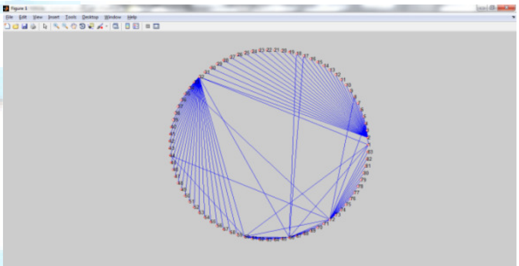

Figure 4.1Graph of Social Network

The above figure describes the data sets of social network. In this Original graph we are going to implement the proposed anonymization algorithms.

Graph Construction

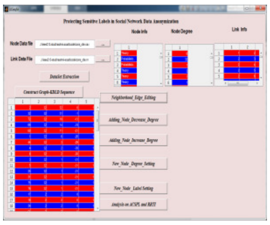
Protecting Sensitive Labels in Social Network Data Anonymization
Node Data file
Link Data File
Node Info
Node Degree
Link Info
Detailes Extraction
Construct Graph-KDLD Sequence
Neighborhood_Edge_Editing
Adding_Node_Decrease_Degree
Adding_Node_Increase_Degree
New_Node_Degree_Setting
New_Node_Label Setting
Analysis on ACSPL and RRTI

Figure 4.2 Graph construction

In this graph construction, each node and labels details are clearly predicted and by doing adjustments like adding node increase degree, adding node decrease degree, new node label setting and new node degree setting and by doing operations and making use of noise nodes.

New Graph

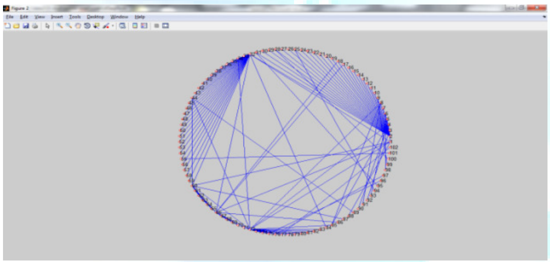

Figure 4.3New Graph

After doing above modifications and by adding noise nodes, the following graph will be generated like this.

V.Performance Analysis

The following figure describes the comparisons of existing approach and proposed approach.

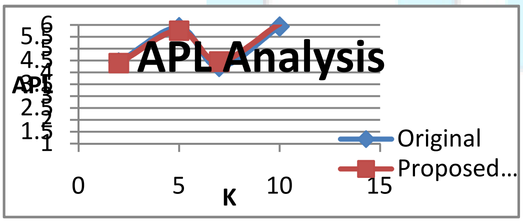

Figure 4.4Graph of Social Network

V.Conclusions and Future Work

In this paper, we proposes a scheme namely k-degree-L-diversity model for privacy preservation of social network data. The requirement can be achieved by noise node adding algorithm to a newly generated graph from the original graph with the constraint of introducing distortions to the original social network graphs. Extensive experimental results demonstrate that the noise node adding algorithms can performs a better result than the previous work of edge editing method. In a distributed environment, data publication satisfy certain privacy requirements, an hacker can still collapse privacy by connecting the data by different users. Similar Protocols should be designed to help the data publishers to guarantee the Privacy preservation.

References

[1]S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava,"Class-Based Graph Anonymization for Social Network Data,"Proc. VLDB Endowment, vol. 2, pp. 766-777, 2009.

[2]A. Cam pan and T.M. Truta, "A Clustering Approach for Data and Structural Anonymity in Social Networks," Proc. Second ACM SIGKDD Int'l Workshop Privacy, Security, and Trust in KDD (PinKDD '08), 2008.

[3] E.M. Knorr, R.T. Ng, and V. Tucakov, "Distance-Based Outliers: Algorithms and Applications," The VLDB J., vol. 8, pp. 237-253, Feb. 2000.

[4] K. Liu and E. Terzi, "Towards Identity Anonymization on Graphs," SIGMOD '08: Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 93-106, 2008

[5]M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis, "Resisting Structural Re-Identification in Anonymized Social Networks," Proc. VLDB Endowment, vol. 1, pp. 102-114, 2008

[6]L. Liu, J. Wang, J. Liu, and J. Zhang, "Privacy Preserving in Social Networks against Sensitive Edge Disclosure," Technical Report CMIDA-HiPSCCS 006-08, 2008

[7]A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-Diversity: Privacy Beyond K Anonymity," ACM Trans. Knowledge Discovery Data, vol. 1, article 3, Mar. 200139

[8] A. Narayanan and V. Shmatikov, "De-Anonymizing Social Networks," Proc. IEEE 30th Symp. Security and Privacy, pp. 173-187, 2009.

[9]C.C. Noble and D.J. Cook, "Graph-Based Anomaly Detection,"Proc. Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '03), pp. 631-636, 2003

[10] K.P. Puttaswamy, A. Sala, and B.Y. Zhao, "Starclique: Guaranteeing User Privacy in Social

Networks against Intersection Attacks," Proc. Fifth Int'l Conf. Emerging Networking Experiments and Technologies (Context '09), pp. 157-168,

[11]X. Xiao and Y. Tao, "Anatomy: Simple and Effective Privacy Preservation," Proc. 32nd Int'l Conf. Very Large Databases (VLDB '06), pp. 139-150, 2006

[12]X. Ying and X. Wu, "Randomizing Social Networks: A Spectrum Preserving Approach," Proc. Eighth SIAM Conf. Data Mining (SDM '08), 2008.

[13] X. Ying, X. Wu, and D. Barbara, "Spectrum Based Fraud Detection in Social Networks," Proc. IEEE 27th Int'l Conf. Very Large Databases (VLDB '11), 2011

[14]B. Zhou and J. Pei, "Preserving Privacy in Social Networks Against Neighborhood Attacks," Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE '08), pp. 506-515, 2008.

Authors Profile

Ms.B.Rashmi1,Pursuing her M.E in Computer Science and engineering from the department of computer science in SCAD engineering ssCollege, Cheranmahadevi and she received her B.E from the department of Computer Science and Engineering in 2010 from SCAD College Of Engineering and Technology under Anna University, Chennai.

Mrs.J.Suganya2,Assistant Professor from the department of computer science in SCAD engineering College, Cheranmahadevi and received her M.E degree in SCAD College Of Engineering and Technology Under Anna University, Chennai and she received her B.E degree from the department of Computer Science in Noorul Islam College Of Engineering Under MS University, Tirunelveli.