ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Distributed Privacy Protection on Social Network Data Using Sensitive Label Technique

P.Sai Ramya
M.Tech. Student,
Dept of CSE,
Indur Institute of Engineering and Technology,
Telangana, India.

A.Bhagya
Assistant Professor,
Dept of CSE,
Indur Institute of Engineering and Technology,
Telangana , India.

ABSTRACT:

Privacy is one of the major concerns when publishing or sharing social network data for social science research and business analysis. Recently, researchers have developed privacy models similar to k-anonymity to prevent node reidentification through structure information. However, even when these privacy models are enforced, an attacker may still be able to infer one's private information if a group of nodes largely share the same sensitive labels (i.e., attributes). In other words, the label-node relationship is not well protected by pure structure anonymization methods. Furthermore, existing approaches, which rely on edge editing or node clustering, may significantly alter key graph properties. In this paper, we define a k-degree-l-diversity anonymity model that considers the protection of structural information as well as sensitive labels of individuals. We had seen a novel anonymization methodology based on adding noise nodes. We implemented that algorithm by adding noise nodes into the original graph with the consideration of introducing the least distortion to graph properties. We here propose novel approach to reduce number of noise node so that decrease the complexity within networks.

Index Terms:

Privacy, Online Social Network, Privacy protecting in SN, Sensitive information.

INTRODUCTION:

With the rapid growth of social networks, such as Facebook and Linkedin, more and more researchersfound that it is a great opportunity to obtain useful information from these social network data, such as the user behavior, community growth, disease spreading, etc.

However, it is paramount that published social network data should not reveal private information of individuals. Thus, how to protect individual's privacy and at the same time preserve the utility of social network data becomes a challenging topic. In this paper, a graph model where each vertex in the graph is associated with a sensitive label. Recently, much work has been done on anonymizing tabular microdata. A variety of privacy models as well as anonymization algorithms have been developed. In tabular microdata, some of the nonsensitive attributes, called quasi identifiers, can be used to reidentify individuals and their sensitive attributes. When publishing social network data, graph structures are also published with corresponding social relationships.

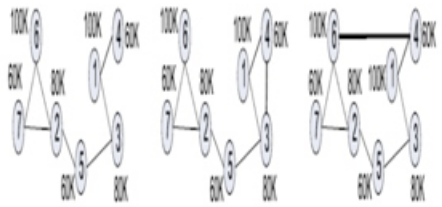
As a result, it may be exploited as a new means to compromise privacy. A structure attack refers to an attack that uses the structure information, such as the degree and the subgraph of a node, to identify the node. To prevent structure attacks, a published graph should satisfy k-anonymity. The goal is to publish a social graph, which always has at least k candidates in different attack scenarios in order to protect privacy. Liu and Terzi did pioneer work in this direction that defined a k-degree anonymity model to prevent degree attacks (Attacks use the degree of a node).Both vertices and edges of a social network graph can be associated with personal information.

A vertex can have properties of a single person, for example, the identity, medical record, shopping transactions, locations, and employment history. An edge can have properties of relationships between two persons. There are two edge properties, namely the type and the weight of a relationship. The type defines the nature of the relationship, for example, an edge can represent a friendship, a recommendation, an email, or a phone call between two members.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The weight defines a quantitative measure of the relationship, for example, a weight can describe the degree of a friendship, the trustworthiness (Das et al., 2010) of a recommendation, and the frequency of a communication. Notice that, even without any label, an edge alone can indicate the existence of a relationship. Privacy can be breached via a social network graph if the vertex of the person or edges involving that vertex leaks some private information. For example, suppose Bill from New York City is a member of an online social network and sets his privacy control not to publicly release his current city. If the owner of the social network publishes the social graph to a third party, say for application development, Bill will be represented by a vertex in the social graph.

If an adversary is able to derive that Bill is in New York City, we say that Bill's privacy is breached. The adversary can find the city of Bill if she can identify Bill's vertex in the graph and observe the city property of the vertex. Up to now, we only considered social graphs representing online social networks. In fact, social graphs have been used to represent data from many off-line sources include databases of hospitals, telephone companies, law enforcement and other government agencies. There can be a variety of attacks that use social network graphs to breach privacy. For social network graphs from off-line sources, the privacy attacks often use a published social graph. However, for social network graphs from on-line sources, there are more varieties of attacks. We will discuss these attacks in more detail in the subsequent sections. The rest of the Chapter is organized as follows.Here we discuss privacy attacks targeted to vertices and methods that protect privacy from these attacks, we introduce some issues and methods of edge anonymization.We discuss a specific edge anonymity method based on vertex-degree as an example to examine some technical details.



a)Original graph SN b)2-degree anonymous graph c)2degree 2-diversity graph SN.

Fig. 1a shows an example of a possible structure attack using degree collect the information. If an adversary knows that one person has three friends in a graph, he can know that node 2 is that person and the related attributes of node 2 are revealed. K-degree anonymity can be used to prevent such structured attacks in SN. However, in many applications in, a social network where each node has sensitive attributes should be published. For example, a graph may contain the user salaries which are sensitive. In this case, k-degree alone is not sufficient to prevent the inference of sensitive attributes of individuals. Fig. 1b shows a graph that satisfies 2-degree anonymity but node labels are not consider in a graph. In it, nodes 2 and 3 have the same degree 3, but they both have the label "80K." If an attacker knows someone has three friends in the social networks, he can conclude that this person's salary is 80K without exactly re-identify the node. Therefore, when sensitive labels are considered, the l-diversity should be adopted for graphs. Again, the l-diversity concept here has the same meaning as that defined over tabular data.

Related Work:

The first necessary anonymization technique in both the contexts of micro- and network data consists in removing identification. This nave 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 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, subgraph knowledge queries). Recognizing the problem, several workspropose 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, 3]. To our knowledge, Zhou and Pei [25, 26] and Yuan et al.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[23] were the first to consider modeling social networks as labeled graphs, similarly to what we con- sider 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 anonymizes the neighbour hoods 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.

Scope of project:

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

1)The label-node relationship is not well protected by pure structure anonymization methods.

2) k-degree-l-diversity anonymity model that considers the protection of structural information as well as sensitive labels of individuals.

3) Adding noise nodes into the original graph with the consideration of introducing the least distortion to graph properties.

Existing system:

Edge-Editing –Based Model:

The edge editing- based model is to add or delete edges to make the graph satisfy certain properties according to the privacy requirements. Most edge-editing-based graph protection models implement k-anonymity of nodes on different background knowledge of the attacker. Liu and Terzi defined and implemented k-degree-anonymous model on

network structure that is for published network, for any node, there exists at least other k-1 nodes have the same degree as this node. Zhou and Pei considered k-neighborhood anonymous model: for every node, there exist at least other k-1 nodes sharing isomorphic neighborhoods.

Clustering-Based Model:

Clustering-based model is to cluster "similar" nodes together to form super nodes. Each super node represents several nodes which are also called a "cluster." Then, the links between nodes are represented as the edges between super nodes which is called "super edges." Each super edge may represent more than one edge in the original graph. The graph that only contains super nodes and super edges are called as clustered graph.

Disadvantages:

Simply removing the identifiers in social networks does not guarantee privacy

Proposed system:

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.

Advantages:

1)It helps publishers publish a unified data together to guarantee the privacy.

2)Low overhead.

3)Preserve social Distance.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Algorithm:

The main objective of the algorithms that we propose is to make suitable group- ing of nodes, and appropriate modification of neighbors' labels of nodes of each group to satisfy the l-sensitive-label-diversity requirement. We want to group nodes with as similar neighborhood 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 neighborhood 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 neighborhood labels are grouped together. Their neighbor labels are modified to be the same immediately so that nodes in one group always have the same neighbor labels.

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 neighborhood information. Thus, neighborhood 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 neighborhood 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.

LITERATURE REVIEW:

Online social Networks have always been an important component of our daily life, but currently that more and more people are connected to the Internet, and their online counterpart is satisfying an increasingly vital role. Consider a graph model where each vertex in the graph is associated with as the sensitive label or (private information). According to survey privacy related issues in social networking is very important. Since this work explores the Preserving privacy in publishing social network data becomes an important concern. With some local knowledge about individuals in a social network, an adversary may attack the privacy of some victimseasily.

Unfortunately, most of the previous studies on privacy preservation data publishing can deal with relational data only, and cannot be applied to social network data. In this paper, we take an initiative towards preserving privacy in social network data. Specifically, we identify an essential type of privacy attacks: neighborhood attacks.

If an adversary has some knowledge about the neighbors of a target victim and the relationship among the neighbors, the victim may be reidentified from a social network even if the victim's identity is preserved using the conventional anonymization techniques. To protect privacy against neighborhood attacks, we extend the conventional k-anonymity and l-diversity models from relational data to social network data. We show that the problems of computing optimal k-anonymous and ldiverse social networks are NP-hard. We develop practical solutions to the problems. The empirical study indicates that the anonymized social network data by our methods can still be used to answer aggregate network queries with high accuracy. The increasing popularity of social networks has initiated a fertile research area in information extraction and data mining.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Although such analysis can facilitate better understanding of sociological, behavioral, and other interesting phenomena, there is growing concern about personal privacy being breached, thereby requiring effective anonymization techniques.

PROBLEM DISCRIPTION:

The publication of social network data entails a privacy threat for their users. Sensitive information about users of the social networks should be protected. The challenge is to devise methods to publish social network data in a form that affords utility without compromising privacy. Previous research has pro-posed various privacy models with the corresponding protection mechanisms that prevent both inadvertent private information leakage and attacks by malicious adversaries.

The social networks are modeled as graphs in which users are nodes and features are labels. Labels are denoted either as sensitive or as non-sensitive. Each node in the graph represents a use. Main challenge is how it works in distributed environment.

Objective:

To develop a new technique to provide privacy and security of social network data in distributed environment with the help of graph property.

The objectives of project are as follows:

1)We can publish the Non sensitive data to every-one in social Network.
2)Add minimum no of noise and improve anonymization technique.
3)Increase and decrease the graph with edge editing.
4)Assign sensitive label to noise node.
5)Protecting sensitive data of each individual user.
6)Security in distributed architecture.

IMPLEMENTATION DETAILS:

Anonymization is a clustering problem one or more nodes are connected each other in various graph in social network and sharing information and resources in social networking business as well as personal relations.

Mathematical Model:

Social Network Graph: a social network graph is a four tuple G (V, E, σ,) where V is a set of vertices, and each vertex represents a node in the social network. E V × V is the set of edges between vertices, _ is a set of labels that vertices have.: V -> σ maps vertices to their labels. Since each noise node connects with at least one noise edge, "Low Overhead" also limits the number of noise nodes that can be added.

PROBLEM DISCRIPTION:

The publication of social network data entails a privacy threat for their users. Sensitive information about users of the social networks should be protected. The challenge is to devise methods to publish social network data in a form that affords utility without compromising privacy. Previous research has pro-posed various privacy models with the corresponding protection mechanisms that prevent both inadvertent private information leakage and attacks by malicious adversaries. The social networks are modeled as graphs in which users are nodes and features are labels. Labels are denoted either as sensitive or as non-sensitive. Each node in the graph represents a use. Main challenge is how it works in distributed environment.

Objective:

To develop a new technique to provide privacy and security of social network data in distributed environment with the help of graph property.
The objectives of project are as follows:

1)We can publish the Non sensitive data to every-one in social Network.

2)Add minimum no of noise and improve anonymization technique.

3)Increase and decrease the graph with edge editing.

4)Assign sensitive label to noise node.

5)Protecting sensitive data of each individual user.

6)Security in distributed architecture.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

IMPLEMENTATION DETAILS:

Anonymization is a clustering problem one or more nodes are connected each other in various graph in social network and sharing information and resources in social networking business as well as personal relations.

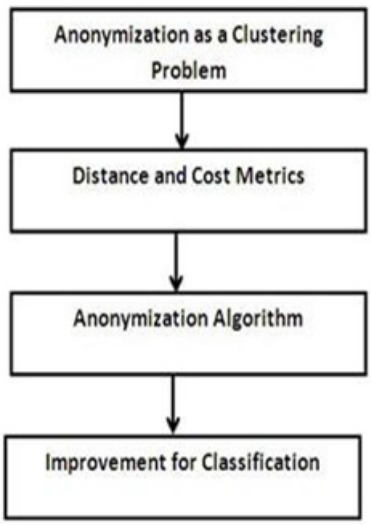
Mathematical Model:

Social Network Graph: a social network graph is a four tuple G (V, E, σ,) where V is a set of vertices, and each vertex represents a node in the social network. E V × V is the set of edges between vertices, _ is a set of labels that vertices have: V -> σ maps vertices to their labels. Since each noise node connects with at least one noise edge, "Low Overhead" also limits the number of noise nodes that can be added.

SYSTEM ARCHITUCTURE:

Fig 5.1 shows the system architecture Anonymizaion techniques uses, distance between nodes & edges are measured, anonymization algorithm (Noise node adding) are for privacy preserving of graph. In system architecture following components are important.

1)Anonymization as a clustering problem.
2)Distance and cost of graph are measured.
3)Applying Edge editing algorithm.
4)Assigning Sensitive labels and sequence generation.
5)Improve anonymization technique.
6)Protecting privacy in distributed environment.



set (Nodes and Edges), and DBLP data set (Nodes and Edges) Details of these data set can be found in online supplement material and results.

CONCLUSION:

In this paper, k-degree-l-diversity model has implemented for privacy preserving social network data publishing. Implementation of both distinct l-diversity and recursive (c, l)-diversity also happened. In order to achieve the requirement of k-degree-l-diversity, a noise node adding algorithm to construct a new graph from the original graph with the constraint of introducing fewer distortions to the original graph. Rigorous analysis of the theoretical bounds on the number of noise nodes added and their impacts on an important graph property. 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 an interesting direction to study clever algorithms which can reduce the number of noise nodes if the noise nodes contribute to both anonymization and diversity. Another interesting direction is to consider how to implement this protection model in a distributed environment, where different publishers publish their data independently and their data are overlapping. In a distributed environment, although the data published by each publisher satisfy certain privacy requirements, an attacker can still break user's privacy by combining the data published by different publishers together. Protocols should be designed to help these publishers publish a unified data together to guarantee the privacy.

REFERENCES:

[1] Lars Backstrom, Cynthia Dwork, Jon Kleinberg, "Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography".

[2] Xiaokui Xiao,Yufei Tao , "Anatomy: Simple and Effective Privacy Preservation"

[3] James Cheng, Ada Wai-Chee Fu, Jia Liu, "K-Isomorphism: Privacy Preserving Network Publication against Structural Attacks"

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[4] Keith B. Frikken, Philippe Golle "Private Social Network Analysis: How to Assemble Pieces of a Graph Privately".

[5]SrivatsavaRanjitGanta,ShivaKasiviswanathan,Adam Smith "Composition Attacks and Auxiliary Information in Data Privacy".

[6] Michael Hay,Gerome Miklau,David Jensen,Don Towsley,Philipp Weis "Resisting Structural Reidentification Anonymized Social Networks".

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, 2008.