International Journal of Science Engineering and Advance Technology

Method To Prevent Re-Identification Of Individual Nodes By Combining K-Degree Anonymity With L-Diversity

1 Sri M.Vamsi Krishna, 2 Dr.K.V.V.S.Narayana Murthy, 3Ch.Srinu
Dept. of CSE, Chaitanya Institute of Science & Tech., Madhavapatnam, Kakinada E.g.dt, AP, India

Abstract:
A range of privacy models as well as anonymization algorithms have been developed. In tabular micro data some of the no responsive 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 equivalent social relationships. As a result it may be oppressed as a new means to compromise privacy. ITH the rapid growth of social networks such as Face book and LinkedIn more and more researchers establish that it is a great opportunity to get hold of useful information from these social network data such as the user behavior, community growth, disease spreading etc. Though it is supreme that published social network data should not disclose private information of individuals. Therefore how to protect individual's privacy and at the same time protect the utility of social network data becomes a challenging topic. In this paper we believe a graph model where each highest point in the graph is associated with a sensitive label.
Keywords: Social networks, privacy, anonymous.

Introduction:

The privacy preserving aim is to avoid an attacker from reidentifying a user and finding the fact that a certain user has a specific sensitive value. To achieve this objective we define a k-degree-l-diversity (KDLD) model for safely publishing a labelled graph and then expand corresponding graph anonymization algorithms with the slightest deformation to the properties of the original graph such as degrees and distances between nodes. Privacy is one of the chief concerns when publishing or sharing social network data for social science research and business analysis. Recently researchers have developed privacy models comparable to k-anonymity to prevent node reidentification through structure information. Nevertheless even when these privacy models are compulsory an attacker may still be capable to infer one's private information if a group of nodes mainly split the same sensitive labels. In other words the label-node association is not well protected by pure structure anonymization methods. In addition existing approaches which depend on edge editing or node clustering may considerably alter key graph properties. In this paper we define a k-degree-l-diversity anonymity model that believes the protection of structural information as well as sensitive labels of individuals. We additional propose a novel anonymization methodology based on adding noise nodes. We develop a new algorithm by adding noise nodes into the original graph with the deliberation of introducing the least distortion to graph properties. Most significantly we provide a thorough analysis of the theoretical bounds on the number of noise nodes added and their collisions on an important graph property.

Related Work:

The unique patterns such as node degree or sub graph to special nodes can be used to reidentify the nodes. The attack that uses certain surroundings knowledge to reidentify the nodes/links in the published graph is called passive attack. There are two models proposed to issue a privacy preserved graph edge-editing-based model and clustering-based model. The edge editing- based model is to add or delete edges to make the graph convince certain properties according to the privacy requirements. Clustering-based model is to cluster similar nodes mutually to form super nodes. Each super node symbolizes several nodes which are also called a cluster. Then the links between nodes are stand for as the edges between super nodes which is called super edges. Each super edge may represent more than one edge in the original graph. We call the graph that only encloses super nodes and super edges as a clustered graph. Most edge-editing-based graph protection models implement k-anonymity of nodes on dissimilar background knowledge of the attacker.

Existing Method:

Recently a great deal work has been done on anonymizing tabular micro data. A variety of privacy models as well as anonymization algorithms have been developed e.g., kanonymity, l-diversity, t-closeness. In tabular micro data some of the non sensitive attributes called quasi identifiers can be used to reidentify individuals and their responsive 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 concession privacy.

Disadvantages:

The edge-editing means sometimes may modify the distance properties considerably by connecting two faraway nodes jointly or deleting the bridge link between two communities. Mining over these data might get the wrong termination about how the salaries are distributed in the society. Therefore exclusively relying on edge editing may not be a good solution to protect data utility.

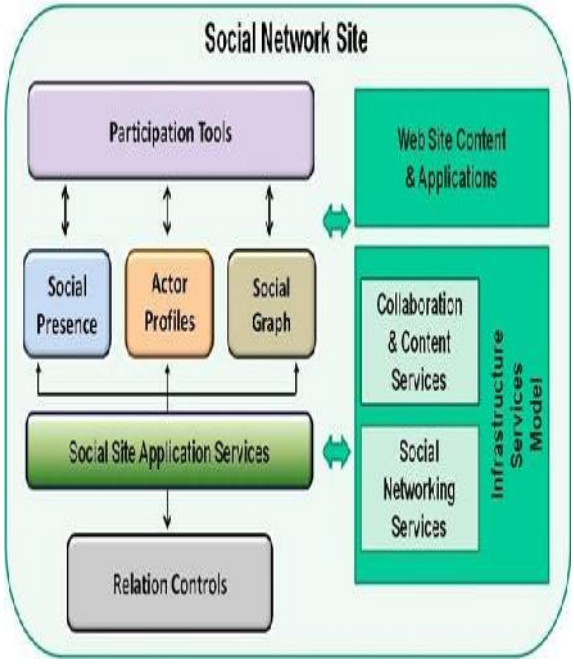
Proposed Method:

Proposal of a novel idea to protect important graph properties such as distances between nodes by adding certain noise nodes into a graph. In proposed system privacy preserving goal is to avoid an attacker from reidentifying a user and finding the fact that a certain user has a detailed sensitive value. To achieve this goal we define a k-degree-l-diversity (KDLD) model for safely publishing a labeled graph and then develop consequent graph anonymization algorithms with the least distortion to the properties of the original graph such as degrees and distances between nodes.

ADVANTAGES:

We merge k-degree anonymity with l-diversity to avoid not only the reidentification of individual nodes but also the exposure of a sensitive attribute connected with each node. We propose a novel graph construction system which makes use of noise nodes to save utilities of the original graph. Two key properties are considered attach as few noise edges as possible and adjust the distance between nodes as less as possible. We present systematic results to show the relationship between the number of noise nodes added and their impacts on an important graph property.

System Architecture:



User Pane:

User Pane is a block of information about a given user like those normally found on a forum post but can be used in other places as well. From core it accumulates the user picture, name, join date, online status, contact link and profile information. In addition any component or subject can feed it more information via the preprocess system. All of this information is then get together and displayed using a template file.

User Relationships:

User Relationship module allow users to generate named relationships between each other. It is the basic building block for a social networking site or any site where users are conscious of one another and communicate.

Content Access:

Content access permits you to supervise permissions for content types by role and author. It allows you to specify custom view, edit and delete permissions for each content type. Optionally you can facilitate per content access settings so you can modify the access for each content node.

Protection Of Structural Information:

The privacy preserving goal is to avoid an attacker from re-identifying a user and finding the fact that a certain user has a precise sensitive value.

Algorithm Used:

Algorithm Uinn

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. For two nodes, v_1 with neighborhood label set (LS_{v1}), and v_2 with neighborhood label set (LS_{v2}), we calculate neighborhood label similarity (NLS) as follows:

$$\mathrm{NLS}(v_1, v_2) = \frac{|LS_{v_1} \cap LS_{v_2}|}{|LS_{v_1} \cup LS_{v_2}|}$$

Larger value indicates larger similarity of the two neighborhoods.

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.

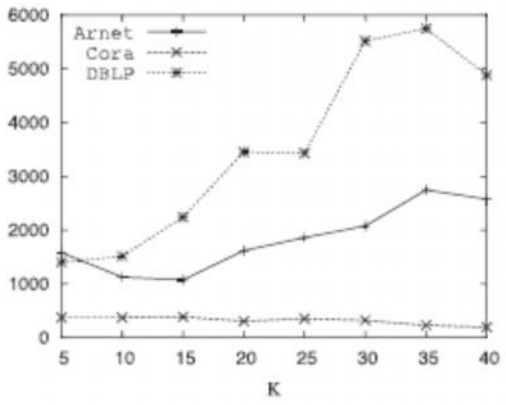
After having formed these groups, we need to ensure that each group's members are indistinguishable in terms of neighborhood information. Thus, neighborhood labels are modified after every grouping operation, so that

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 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 neighborhood 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 neighbors' labels. We consider the unification of two nodes' neighborhood labels as an example. One node may need a noisy node to be added as its immediate neighbor since it does not have a neighbor with certain label that the other node has; such a label on the other node may not be modifiable, as it is already connected to another sensitive node, which prevents the re-modification on existing modified groups.

```
1  P = the sensitive degree sequence of the original graph G;
2  Set R = {};
3  while |P| > 0 do
4      Group C = {P[0]};
5      int d = P[0].d;
6      Remove P[0] out of P;
7      while ¬(C satisfies the Safety Grouping Condition) do
8          for i = 0; i < |P|; i++ do
9              if P[i].d ≡ d then
10                 C = C ∪ {P[i]};
11                 Remove P[i] out of P;
12                 break;
13             else
14                 if P[i].s is not in the top l − 1 appearance label set of C then
15                     C = C ∪ {P[i]};
16                     Remove P[i] out of P;
17                     break;
18         if i == |P| then
19             R = R ∪ C;
20             break;
21     Set the target degrees of elements in C as corresponding nodes' mean degree;
22     Copy C into P^new if C satisfies SG condition;
23 Assign the elements in R to existing groups;
```

Experimental Results:



A strong supposition that the attacker knows the precise description of noise nodes and uses these descriptions to filter noise nodes out. We want to test whether the characteristics of noise nodes are particular evaluating to original nodes. We first presume the attacker already knows the degrees of all the noise nodes and uses this information to sieve all the nodes with these degrees out. From the result we can see our algorithm is very resourceful and the largest running time is less than 6,000 ms.

Conclusion:

We put into practice both distinct l-diversity and recursive assortment. In order to attain the requirement of k-degree-l-diversity we design a noise node adding algorithm to build a new graph from the original graph with the restraint of initiate fewer distortions to the original graph. We give a thorough analysis of the theoretical bounds on the number of noise nodes further and their impacts on an important graph property. Our widespread experimental results show that the noise node adding algorithms can realize a better result than the previous work using edge editing only. It is an attractive course to study clever algorithms which can decrease the number of noise nodes if the noise nodes donate to both anonymization and diversity. Another interesting direction is to believe how to implement this protection replica in a distributed environment where dissimilar publishers publish their data autonomously and their data are overlapping. In a distributed environment even though the data published by each publisher gratify certain privacy requirements an attacker can still break user's solitude by combining the data published by different publishers together.

References:

[1] L. Backstrom, C. Dwork, and J.M. Kleinberg, "Wherefore Art Thou r3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography," Proc. Int'l Conf. World Wide Web (WWW), pp. 181-190, 2007.

[2] A.-L. Baraba´si and R. Albert, "Emergence of Scaling in Random Networks," Science, vol. 286, pp. 509-512, 1999.
[3] 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.
[4] A. Campan 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.
[5] A. Campan, T.M. Truta, and N. Cooper, "P-Sensitive K-Anonymity with Generalization Constraints," Trans. Data Privacy, vol. 2, pp. 65-89, 2010.
[6] J. Cheng, A.W.-c. Fu, and J. Liu, "K-Isomorphism: Privacy Preserving Network Publication against Structural Attacks," Proc. Int'l Conf. Management of Data, pp. 459-470, 2010.
[7] G. Cormode, D. Srivastava, T. Yu, and Q. Zhang, "Anonymizing Bipartite Graph Data Using Safe Groupings," Proc. VLDB Endowment, vol. 1, pp. 833-844, 2008.
[8] S. Das, O. Egecioglu, and A.E. Abbadi, "Privacy Preserving in Weighted Social Network," Proc. Int'l Conf. Data Eng. (ICDE '10), pp. 904-907, 2010.
[9] W. Eberle and L. Holder, "Discovering Structural Anomalies in Graph-Based Data," Proc. IEEE Seventh Int'l Conf. Data Mining Workshops (ICDM '07), pp. 393-398, 2007.
[10] K.B. Frikken and P. Golle, "Private Social Network Analysis: How to Assemble Pieces of a Graph Privately," Proc. Fifth ACM Workshop Privacy in Electronic Soc. (WPES '06), pp. 89-98, 2006.
[11] S.R. Ganta, S. Kasiviswanathan, and A. Smith, "CompositionAttacks and Auxiliary Information in Data Privacy," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 265-273, 2008.
[12] G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "Fast Data Anonymization with Low Information Loss," Proc. 33rd Int'l Conf. Very Large Data Bases (VLDB '07), pp. 758-769, 2007.
[13] G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "A Framework for Efficient Data Anonymization Under Privacy and Accuracy Constraints," ACM Trans. Database Systems, vol. 34, pp. 9:1-9:47, July 2009.
[14] J. Han, Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, Inc., 2005.
[15] 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.
[16] 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.

Authors:

Sri. M. Vamsi krishna, well Known Author and excellent teacher Received M.Tech (AI &R), M.Tech (CS) from Andhra university is working as Professor and HOD, Department of CSE, Chaitanya Institute Science and Technology. He has 13 years of teaching & research experience. He has 20 publications of both national and international conferences /journals. His area of Interest includes AI, Computer Networks, information security, flavors of Unix Operating systems and other advances in computer Applications.

Dr.K.V.V.S.Narayana Murthy M.Tech, Ph.D. Professor of CSE department in Chaitanya Institute of Science & technology. He has 15 years of teaching & research experience to his credit number of publications both national and international conferences /journals. His area of interest includes compiler design, formal languages and automata theory, computer organization.

Ch.Srinu is a student of Chaitanya Institute of science & technology Madhavapatnam, Kakinada Presently he is pursuing M.Tech[Computer science Engineering] from this college and he received B.Tech from Sri Sai Aditya Institute of Science & Technology affiliated to JNT University, Kakinada in the year 2012. His area of interest includes Computer Networks and data mining and warehousing, all current trends and techniques in Computer Science.