ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

Term Frequency Based Sequence Generation Algorithm for Graph Based Data Anonymization

S.Charanyaa[1], K.Sangeetha[2]

M.Tech. Student, Dept of Information Technology, S.N.S. College of Technology, Coimbatore, TamilNadu, India[1]

Assistant Professor, Dept of Information Technology, S.N.S. College of Technology, Coimbatore, TamilNadu, India[2]

ABSTRACT: A big deal of research has been performed in the area of graphical data anonymization. Because of the wide range of application of graphical data from social network data to large warehouse data and knowledge engineering domains. Notion of k-anonymity has been proposed in literature, which is a framework for protecting privacy, emphasizing the lemma that a database to be k-anonymous, every tuple should be different from at least k-1 other tuples in accordance with their quasi-identifiers(QIDs). Inspite of the existence of k-anonymity framework, malicious users and misfeasers may get authorization to the sensitive information if a set of nodes exhibit alike attributes. In this paper we make a systematic analysis on structure anonymization mechanisms and models projected in the literature. Also we discuss the simulation analysis of KDLD model creation and construction. We propose a Term Frequency Based Sequence Generation Algorithm (TFSGA) which creates node sequence based on term frequency of tuples with minimal distortion. We experimentally show the efficiency of the proposed algorithm under varying cluster sizes.

KEYWORDS: Data Anonymization; Graphical Data; Sensitive information; k-anonymity; l-diversity; Database Privacy; Cluster

I. INTRODUCTION

Recent days have seen a steep rise in the usage of social networking sites such as Facebook and LinkedIn. This rise has led an intruder to gain constructive information such as behavioural pattern of potential client, rate of growth of community, wide spread of a specific syndrome in a ecological environment. Key challenge appears in guaranteeing privacy and utility while preserving private sensitive information of individuals. Lot of works done on anonymizing a relational database has been reported in literature. k-anonymity approach, proposed by L.Sweeney et.al., (2001) [26] , is a model for protecting privacy, which emphasises the lemma that a database to be k-anonymous, each record should be indistinguishable from at least k-1 other records with respect to their quasi-identifiers. Quasi-Identifiers are attributes whose values when taken together as a group can potentially identify an individual. Beacuse k-anonymity failed to secure the attribute disclosure, and is susceptible to homogeneity attack and background knowledge attack A.Machanavajjhala et.al.,[20] (2006) introduced a new privacy notation called 'l-diversity'. A.Machanavajjhala et.al., terms an equivalence class to possess l-diversity if there exist atleast 'l' well represented values for the sensitive attribute. A table is said to have l-diversity if every equivalence class of the table exhibits l-diversity. Privacy is measured by the information gain of an observer. Before seeing the released table the observer may think that something might happen to the sensitive attribute value of a single person. After seeing the released table the observer may have the details about the sensitive attributes. In 2010, Ningui Li et.al., proposed a concept called t-closeness, which poses the restriction on the distance between the class and the whole table should be no more than a threshold 't'. Graph structures are also published hand-in-hand when publishing social network data as it may be exploited to compromise privacy. The degree and subgraph of a node could be used to identify a node. It is observed from literature that inorder to prevent structure attacks the graph is enforced to satisfy k-anonymity.

The remainder of the paper is organized as follows. Section 2 describes the basic definition and primitives of privacy preserving databases. Section 3 portrays a broad yet in-depth literature survey on applications where node with sensitive attributes should be published and approaches for protecting graph privacy. Section 4 describes the significance of K-Degree L-Diversity model creation. The construction of KDLD model is dealt in Section 5. The three major steps involved in construction are discussed in stepwise manner. Section 6 gives a detailed investigations on results obtained and the corresponding discussion. Section 7 concludes the paper and outlines the future work.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

II. BASIC DEFINITION AND PRIMITIVES

Data refers to organized personal information in the form of rows and columns. Row [37] refers to individual tuple or record and column refers to the field. Tuple that forms a part of a single table are not necessarily unique. Column of a table is referred to as attribute that refers to the field of information, thereby an attribute can be concluded as domain. It is necessary that attribute that forms a part of the table should be unique. According to L.Sweeney et.al., (2001) [26] each row in a table is an ordered n-tuple of values $<d_1,d_2,....d_n>$ such that each value d_j forms a part of the domain of j^{th} column for j=1,2,...n where 'n' denoted the number of columns.

A. *Attributes*

Consider a relation $R(a_1, a_2,... a_n)$ with finite set of tuples. Then the finite set of attributes of R are $\{a_1, a_2,... a_n\}$, provided a table $R(a_1, a_2,... a_n)$, $\{a_1, a_2,... a_i\} \subseteq \{a_1, a_2,... a_n\}$ and a tuple $l \in R$, $l[a_i,... a_n]$ corresponds to ordered set of values $v_i,...v_j$ of $a_i... a_j$ in l. R $[a_i... a_j]$ corresponds to projection of attribute values $a_1, a_2,...,a_n$ in R, thereby maintaining tuple duplicates.

According to Ningui Li, Tiancheng Li et.al., [17] (2010), attributes among itself can be divided into 3 categories namely

1. Explicit identifiers- Attributes that clearly identifies individuals. For eg, Social Security Number for a US citizen.
2. Quasi identifiers- Attributes whose values when taken together can potentially identify an individual. Eg., postal code, age, sex of a person. Combination of these can lead to disclosure of personal information.
3. Sensitive identifiers- That are attributes needed to be supplied for researchers keeping the identifiers anonymous. For eg, 'disease' attribute in a hospital database, 'salary' attribute in an employee database.

TABLE 1:
MICRODATA DATABASE CONTAINING SENSITIVE INFORMATION

Race	Birth	Gender	Zipcode	Disease (Sensitive Information)
Black	1965	M	0213	Shortbreath
Black	1965	M	0213	Shortbreath
Black	1965	M	0214	Hypertension
White	1964	F	0213	Obesity
White	1965	F	0214	Chestpain
White	1967	M	0213	Shortbreath
White	1964	M	0214	Chestpain

B. *Quasi- Identifiers*

As proposed by L.Sweeney et.al., (2001) [26], A single attribute or a set of attributes that, in combination with some outside world information that can identify a single individual tuple in a relation is termed as quasi-identifier. Given a set of entities E, and a table $B(a_1,\ldots,a_n)$, $f_a: E \rightarrow B$ and $f_b: B \rightarrow E'$, where $E \rightarrow E'$. A quasi-identifier of B, written as U_E, is a set of attributes $\{a_i,\ldots,a_j\} \rightarrow \{a_1,\ldots,a_n\}$ where: $\exists s_i \in U$ such that $f_a(f_b(s_i)[U_E]) = s_i$.

C. *k-Anonymity*

Let RT(A1,A2,….An) be a table and QI_{RT} be the Quasi identifier. RT is said to be k-anonymous [26] if and only if each sequence of values in $RT[QI_{RT}]$ appears atleast k-times in $RT[QI_{RT}]$. In short, the Quasi identifier must appear atleast 'k' times in RT, where k=1,2,3,... where 'k' is termed to be the anonymity of the table.

D. *l-diversity*

Since k-anonymity failed to secure the attribute disclosure, and is susceptible to homogeneity attack and background knowledge attack A.Machanavajjhala et.al, (2006) [38] introduced a new privacy notation called 'l-diversity'[20]. An equivalence class is said to possess l-diversity if there are atleast 'l' well represented values for the sensitive attribute. A table is said to have l-diversity if every equivalence class of the table has l-diversity.

E. *t-closeness*

Privacy is measured by the information gain of an observer. Before seeing the released table the observer may think that something might happen to the sensitive attribute value of a single person. After seeing the released table the observer may have the details about the sensitive attributes. t-closeness [17] should have the distance between the class and the whole table is no more than a threshold t, Ningui Li et.al., (2010)[17].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

III. LITERARY REVIEW

A. Campan, T.M. Truta, and N. Cooper (2010) [5] have proposed a new approach for privacy preserving, where requirements on the quantity of deformation allowed on the initial data are forced in order to preserve its usefulness. Their approach consists of specifying quasiidentifiers' generalization constraints, and achieving p-sensitive k-anonymity within the imposed constraints. According to their point of view, limiting the amount of allowed generalization when masking microdata is essential for real life datasets. They formulated an algorithm for generating constrained p-sensitive k-anonymous microdata and named it as constrained p-sensitive k-anonymity model, and proved that the algorithm is in par with other similar algorithms existing in the literature in terms of quality of result. K.

Clustering method [37] is carried out by merging a subgraph to one super node, which is unsuitable for sensitive labeled graphs, because when a group of nodes are merged into a single super node, the node-label relations have been lost. G. Cormode, D. Srivastava, T. Yu, and Q. Zhang (2008) [7] introduced a new family of anonymizations, for bipartite graph data, called (k,l)-groupings. Authors identified a class of "safe" (k,l)-groupings that have provable guarantees to resist a variety of attacks, and show how to find such safe groupings. E. Zheleva and L. Getoor (2007)[30] threw light on the problem of preserving the privacy of sensitive relationships in graph data. Their experimental investigation revealed the victory of several re-identification strategies under varying structural characteristics of the data. A. Campan and T.M. Truta (2008) [4] contributed in the development of a greedy privacy algorithm for anonymizing a social network and the introduction of a structural information loss measure that quantifies the amount of information lost due to edge generalization in the anonymization process. The authors proposed SaNGreeA (Social Network Greedy Anonymization) algorithm, which performs a greedy clustering processing to generate a k-anonymous masked social network and quantified the generalization information loss and structural information loss. A clustered graph is a graph which contains only super nodes and super edges (2013) [35].

Edge-editing methods [37] keep the nodes in the original graph unchanged and only add/delete/swap edges. K.B. Frikken and P. Golle (2006) [10] proposed a method to reconstructing the whole graph privately, i.e., in a way that hides the correspondence between the nodes and edges in the graph and the real-life entities and relationships that they represent to assuage these privacy concerns. Also the authors propose protocols to privately assemble the pieces of a graph in ways that diminish these threats. These protocols substantially restrict the ability of adversaries to compromise the privacy of truthful entities. Also mining over these data might lead to erroneous conclusion about how the salaries are distributed in the society. Hence, exclusively relying on edge editing may not always be a solution to preserve data utility. Another novel idea is proposed by Mingxuan Yuan, Lei Chen et.al., (2013) [35] to preserve important graph properties, such as distances between nodes by appending some "noise" nodes into a graph. The core idea behind this is that many social networks satisfy the Power Law distribution [2], i.e., there exist a huge number of low degree vertices in the graph which could be used to hide appended noise nodes from being reidentified. Some graph nodes could be preserved much better by appending noise nodes than the existing pure edge-editing method.

IV. K-DEGREE L-DIVERSITY MODEL CREATION

Despite the k-anonymity model, an intruder may gain access the sensitive information if a set of nodes share similar attributes. Also we make a detailed study on k-degree-l-diversity anonymity model, which takes into consideration the structural information and sensitive labels of individuals as well. Also the study the algorithmic impact of adding noise nodes to original graph and the rigorous analyses is also important. Privacy is one of the major concerns when publishing or sharing social network data for social science research and business analysis. In the process of finding the degree of nodes, if two or three nodes which have same degree and same information such as salary, year and then they assume the node which has three degree also have the same information.

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. They did pioneer work in this direction that defined a k-degree anonymity model to prevent degree attacks (attacks use the degree of a node). A graph is k-degree anonymous if and only if for any node in this graph, there exist at least k-other nodes with the same degree.

The edge-editing method sometimes may change the distance proper-ties substantially by connecting two faraway nodes together or deleting the bridge link between two communities. This phenomenon is not preferred. Mining over these data might get the wrong conclusion about how the salaries are distributed in the society. Therefore, solely relying on edge editing may not be a good solution to preserve data utility. To address this issue, a novel idea is

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

proposed to preserve important graph properties, such as distances between nodes by adding certain "noise" nodes into a graph. This idea is based on the following key observation. Most social networks satisfy the Power Law distribution i.e., there exist a large number of low degree vertices in the graph which could be used to hide added noise nodes from being reidentified. By carefully inserting noise nodes, some graph properties could be better preserved than a pure edge-editing method. The distances between the original nodes are mostly preserved. The k-anonymity model prevents the re-identification of nodes and prevents the structural attacks. L-diversity model prevents the attacks (Homogeneity attack and background knowledge attack) and adding of noise nodes to change the degree of the nodes that was done through l-distinct labels of the specified connecting pairs of nodes. The k-degree l-diversity (KDLD) model is a combination of two models K-degree and L-diversity. This model not only prevents the node re-identification, but also exposes the labels for each node when publishing. For this a new technique of graph construction technique proposed, which uses the noise nodes to be safeguarding the original graph usage by adding less number of noise nodes and changing the distance between the pair of nodes.

V. KDLD MODEL CONSTRUCTION

KDLD model is a combination of two models K-degree and L-diversity. This model not only prevents the node re-identification, but also exposes the labels for each node when publishing. For this a new technique of graph construction technique proposed, which uses the noise nodes to be safeguarding the original graph usage by adding less number of noise nodes and changing the distance between the pair of nodes.

A. *Data Uploading and Pre-processing*

Data uploading module is to upload the real dataset into the execution environment. Once the data is uploaded, it needs to be preprocessed. Preprocessing the data means truncating the most unwanted symbols from the dataset. This preprocessed data is used for KDLD sequence generation.

B. *KDLD Sequence Generation*

KDLD sequence is a combination of k-anonymity and l-diversity anonymization techniques. Consider a social network graph whose degree sequence is depicted as P. The degree sequence P holds n triples i.e., (id, d, s) where id is to denote the node, d is the degree of the node and s defines the sensitive labels associated with that particular node. k-l based or l-k based algorithm is applied to the degree sequence P to generate new degree sequence P^{new}. P^{new} is constructed in such a way that the node whose degree changes should be very small. Nodes with the similar degrees are grouped. C_{new} is the cost of creating new group and C_{merge} is the cost of merging the node to the same group. The module tends to put the node with similar degree as a group and the mean degree change is computed. Based on mean value noise node is added.

Given an unprocessed input dataset D; cost C_{new} and C_{merge}, term frequency T_{tf}, the term frequency based KDLD sequence generation algorithm that will create a sequence P_{new} with minimal distortion is given below:

Algorithm 1: Term Frequency Based KDLD Sequence Generation Algorithm

Input: An unprocessed input dataset D; cost C_{new} and C_{merge} term frequency T_{tf}

1. Preprocess (D);
2. **Repeat** until no tuple is ungrouped
3. **Group** Dataset -> $G_1, G_2, \ldots, G_{n-1}$, G_n w.r.t. term frequency T_{tf} **;**
4. **cluster: Form Cluster** on the basis of cost value C_{new} and C_{merge};
5. **Repeat** for all nodes
6. **Calculate** Degree(Node);
7. **Calculate** Mean_Degree(AllNodes);
8. **goto** cluster;

Finally a new KDLD sequence (P_{new}) will be created based on term frequency of tuples with least distortion of each nodes.

C. *Graph Construction*

Based on the new KDLD sequence (P^{new}), a graph is constructed using graph construction module. Edge

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

editing is an approach to protect graph privacy.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 neighbor nodes, so that the path the nodes would be as short as possible. Assigning additional node is to amplify the degree for increased anonymization. Consider nodes u and v whose degree needs to be increased and are within two hop distance. In this case a noise node(n) is created for u which increases u's degree. This noise node n is again connected to v to increase its own degree. This process continues until the degree of noise node is equal to mean degree of P^{new}. The process of adding noise nodes may also diminish the degree of a particular node. Consider a node u whose degree needs to be decreased. In this case a noise node is created for u and now its degree is increased. To decrease the degree of u, go for randomly deleting edge between u and v and connecting noise nodes n to v. As a result the path between u and v is reduced to one hop but not a remarkable reduction. Thus the degree is reduced with the help of noise nodes.

VI. RESULTS AND DISCUSSIONS

An proposed to preserve important graph properties, such as distances between nodes by adding certain "noise" nodes into a graph. In proposed system, privacy preserving is to prevent an attacker from reidentifying a user and finding the fact that a certain user has a specific sensitive value. K-degree anonymity with l-diversity is combined to prevent not only the reidentification of individual nodes but also the revelation of a sensitive attribute associated with each node. The algorithm is implemented and tested in a Dell Laptop with Intel Core i5 Pentium IV CPU with 6 GB RAM and 64-bit Windows XP OS. The algorithm is implemented in Netbeans IDE 6.9.1 with MySQL as backend.

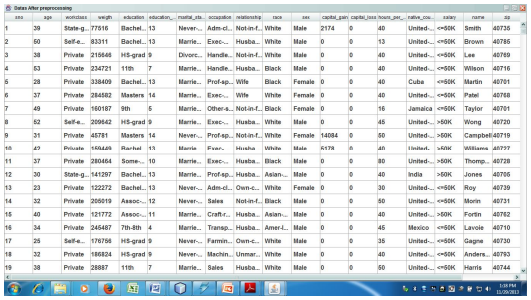

Fig 1. Viewing Dataset after Preprocessing

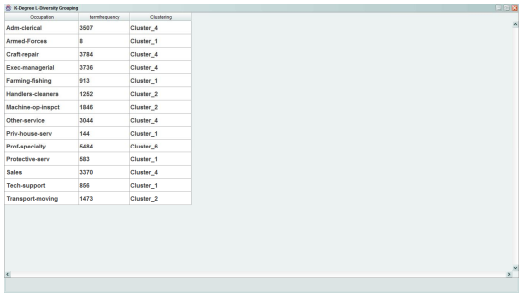

Fig 2 Clustered data after KDLD Sequence Generation based on term frequency

The second step in the project is KDLD Sequence generation. KDLD sequence is a combination of k-anonymity and l-diversity anonymization techniques. Nodes with the similar degrees are grouped (i.e., Clustered). C_{new} is the cost of creating new group and C_{merge} is the cost of merging the node to the same group. The KDLD sequence generation and clustering is illusrated in Fig. 2

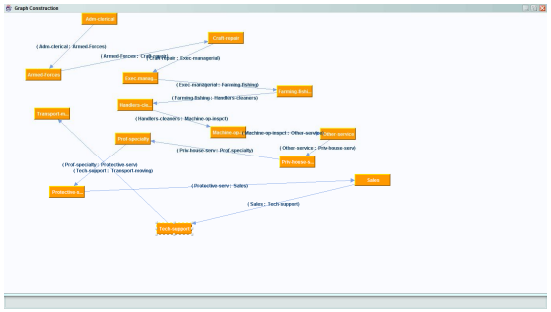

Fig 3 Graph Construction with Sensitive Labels

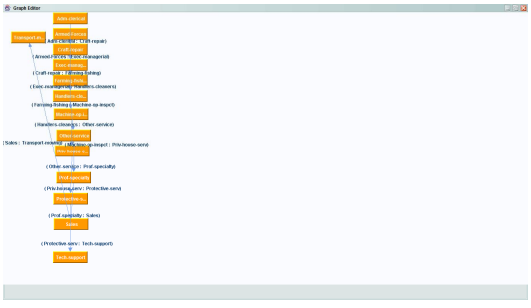

Fig 4 Graph Construction Edge Editing Step

Based on the new KDLD sequence (P_{new}), a graph is constructed using graph construction module. This module will check the mean degree of a group and constructs a graph with minimum degree changes by using the sub

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

modules Edge Editing, Adding noise node to increase degree and Adding noise node to decrease degree as illustrated in Fig. 3. Edge editing is an approach to protect graph privacy. Neighbourhood rule is followed in this approach i.e., to add edge between two neighbour nodes, so that the path the nodes would be as short as possible as illustrated in Fig 4

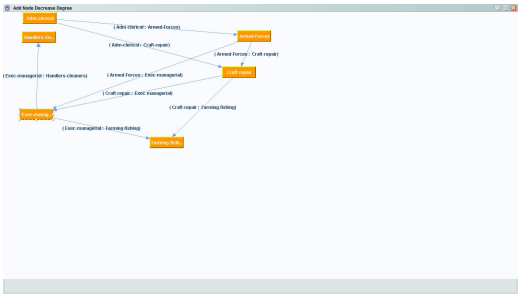

Fig 5 Graph Construction Adding Node Decrease Degree Step

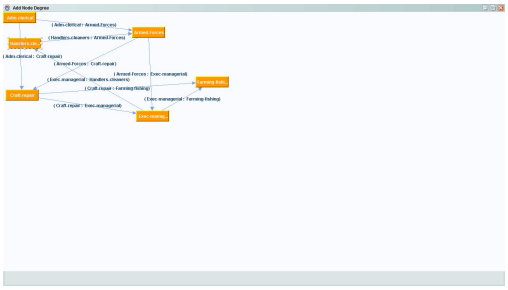

Fig 6 Graph Construction Adding Node Increase Degree Step

This sub module is to increase the degree of the node. Consider nodes u and v whose degree needs to be increased and are within two hop distance. In this case a noise node(n) is created for u which increases u's degree. This noise node n is again connected to v to increase its own degree. This process continues until the degree of noise node is equal to mean degree of P_{new}. as illustrated in Fig.5. Consider a node u whose degree needs to be decreased. In this case a noise node is created for u and now its degree is increased. To decrease the degree of u, random deleting edge between u and v is made and noise nodes n to v is connected. As a result the path between u and v is reduced to one hop but not a remarkable reduction. Thus the degree is reduced with the help of noise nodes as illustrated in Fig.6

VII. POTENTIAL LIMITATIONS KDLD MODEL

Existing model includes the following limitations

(1) The existing model is a combination of two models k-degree and l-diversity mechanisms, the model does not consider the attacks caused by intruders for the graph based data.
(2) The model prevents the node re-identification and exposes the labels for each node when publishing, thereby there is a serious scenario for intruders to attack and fetch sensitive information.
(3) The edge-editing method occasionally may alter the distance properties considerably by connecting two remote nodes together or removing the bridge link between two societies.
(4) Mining over the data might lead to a wrong conclusion about how the salaries are distributed in the society. Therefore, solely relying on edge editing may not be a good ideology to preserve data utility.

VIII. CONCLUSION AND FUTURE WORK

To improve high degree of anonymization in graph structure, k-degree l-diversity (KDLD) is implemented with noise nodes. Increasing or decreasing the noise node in the graphical structure is done based on algorithmic steps. The purpose of adding noise node is to perplex the intruders. To achieve the property of k-degree-l-diversity, we implemented a noise node appending algorithm inorder to construct a new node-appended graph from the original graph with the limitation of inserting minimal distortions to the original graph. It is evident that the algorithm to append noise node achieve good anonymity than working with edge editing only. The term frequency based KDLD model exhibits the limitation as the model does not take into consideration of attacks caused by intruders for this graph based data. Also, the model prevents the node re-identification and exposes the labels for each node when publishing, thereby there is a serious scenario for intruders to attack and fetch sensitive information.

In future we planned to enhance the algorithm for significant improvement in terms of algorithm efficiency, percentage of noise nodes and in terms of several other metrics. Also we have planned to propose attack detection scenarios against homogeneity attack, background knowledge attack and similarity attack on graph based anonymized data.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

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, "Composition Attacks 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.
17. N. Li and T. Li, "T-Closeness: Privacy Beyond K-Anonymity and L-Diversity," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE '07), pp. 106-115, 2007.
18. 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.
19. 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.
20. 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. 2007.
21. A. Narayanan and V. Shmatikov, "De-Anonymizing Social Networks," Proc. IEEE 30th Symp. Security and Privacy, pp. 173-187, 2009.
22. 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.
23. L. Page, S. Brin, R. Motwani, and T. Winograd, "The Pagerank Citation Ranking: Bringing Order to the Web," Proc. World Wide Web Conf. Series, 1998.
24. 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 (CoNEXT '09), pp. 157-168, 2009.
25. N. Shrivastava, A. Majumder, and R. Rastogi, "Mining (Social) Network Graphs to Detect Random Link Attacks," Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE '08), pp. 486-495, 2008.
26. L. Sweeney, "K-Anonymity: A Model for Protecting Privacy," Int'l J. Uncertain. Fuzziness Knowledge-Based Systems, vol. 10, pp. 557-570, 2002.
27. 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.
28. 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.
29. X. Ying and X. Wu, "Randomizing Social Networks: A Spectrum Preserving Approach," Proc. Eighth SIAM Conf. Data Mining (SDM'08), 2008.
30. E. Zheleva and L. Getoor, "Preserving the Privacy of Sensitive Relationships in Graph Data," Proc. First SIGKDD Int'l Workshop Privacy, Security, and Trust in KDD (PinKDD '07), pp. 153-171, 2007.
31. E. Zheleva and L. Getoor, "To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles," Proc. 18th Int'l Conf. World Wide Web (WWW '09), pp. 531-540, 2009.
32. 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.
33. B. Zhou and J. Pei, "The K-Anonymity and L-Diversity Approaches for Privacy Preservation in Social Networks against Neighborhood Attacks," Knowledge and Information Systems, vol. 28, pp. 47-77, 2011.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

34. L. Zou, L. Chen, and M.T. O¨zsu, "K-Automorphism: A General Framework for Privacy Preserving Network Publication," Proc. VLDB Endowment, vol. 2, pp. 946-957, 2009.
35. Mingxuan Yuan, Lei Chen, Philip S. Yu, Ting Yu, "Protecting Sensitive Labels in Social Network Data Anonymization", IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 3, pp.633-647, March 2013
36. S.Balamurugan, P.Visalakshi, "Modified Partitioning Algorithm for Privacy Preservation in Microdata Publishing with Full Functional Dependencies", Australian Journal of Basic and Applied Sciences, 7(8): pp.316-323, July 2013
37. S.Charanyaa, T.Shanmugapriya, "Certain Investigations on Approaches for Protecting Graph Privacy in Data Anonymization", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 8, October 2013
38. S.Charanyaa, T.Shanmugapriya, "A Survey on Attack Prevention and Handling Strategies in Graph Based Data Anonymization" International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 10, October 2013

BIOGRAPHY

S.Charanyaa obtained her B.Tech degree in Information Technology from Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India. She is currently pursuing her M.Tech degree in Information Technology at S.N.S. College of Technology, Coimbatore, Tamilnadu, India. She has published 2 International Journals in the research domain of Database Privacy. Her areas of research interest accumulate in the areas of Database Privacy, Object Modeling Techniques, and Software Engineering.

Prof.K.Sangeetha is currently working as Assistant Professor in the Department of Information Technology at S.N.S. College of Technology, Coimbatore, Tamilnadu, India. She has 5 years of teaching experience. She has published a number of research papers which include 3 International Journals, 5 National Conferences and 3 International Conferences. Her areas of research interest accumulate in the area of Computer Networks.