RESOLVING OF VERTEX DISCLOSURE AND MULTI COMMUNITY DISCLOSURE IN SOCIAL NETWORKS

Gowthamy.R[1*], Uma.P[2]

[*1]M.E.Scholar, Department of Computer Science & Engineering Nandha Engineering College, Erode, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science & Engineering, Nandha Engineering College, Erode, Tamil Nadu, India

ABSTRACT

Social network acts as a platform for sharing different types of information between different typed of users.Social network is modeled as a graph, consisting of vertex and edges. Users involved in sharing data are considered as vertex or node and relationship between different users is mentioned as edges. Main issues likevertex identity disclosure and multicommunity identity disclosurepresent recently in social network. Information about each user identity is vertex identity. Different types of communities where the user involved in social network is defined as multicommunity identity. To overcome these two risks a privacy preservation technique K^W –Structural Diversity Anonimity is introduced. Where K denotes number of nodes in social network and W is a time stamp recording of updation made in the network.

KEYWORDS

Privacy preservation, vertex disclosure and multicommunity disclosure,K^W-structural diversity anonymity technique.

1. INTRODUCTION

Social network is a place of increasing communication between different types of users all over the world. Data sharing continuously increased when the social sites updated incrementally. Each individual user can connect with people at any end of the world [1]. Social network is a representation of a tree structure embedded of nodes and edges. In this graph (Figure 1), vertex denotes number of users participated in social sites and edge denotes the relationship between different users.

According to Knowledge discovery process, social network is organized and data transferred based on user- role based methodology [2]. Users as divided as data provider, data collector, data miner and decision maker. Data provider is a first person who provides information for a data collector, the data send the provider can contain sensitive data, and accordingto the sensitivity leaked he getsreward from the collector. Data collector sends data to miner by hiding or encrypting sensitive data. Data miner mines the useful information from the raw data and transfers to decision maker to decide whether the data is useful or not.

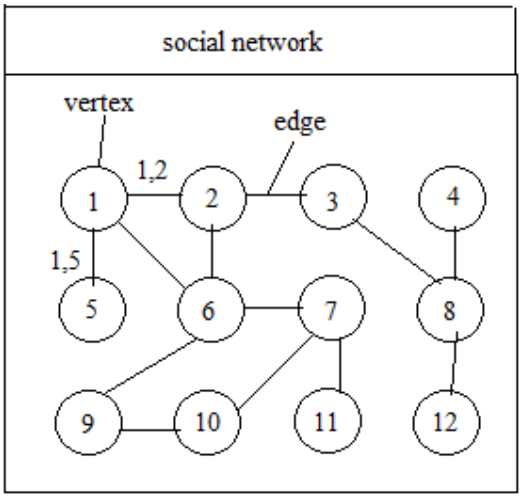

Figure 1.Representation of social network.

Consider a graph of social network with the nodes A, B, C. Where A is a friend ofB and C. The adversary is a person or an attacker who needs to hack and retrieve the sensitive information of the nodes by knowing some information about the particular node. When attacker needs to attack the user C by having some data, he can also gather the information about the nodes A and C through the relationship with A, B and C. An attacker can retrieve information when a particular node is in a community group which is known to the attacker. To overcome this privacy preservation technique has been used to preserve sensitive information of victim user. Attacker can perform different attacks like degree attack, friendship attack, and mutual friendship attack to know sensitive data or to know the identity of a victim.

1.1. Objective

- To analyze the attack model and consider the protection of vertex and community identities of users in the social sites.
- To anonymize each release to satisfy some privacy model before a network is published.
- To prevent the changes to launch attacks made by adversaries due to the lack of consideration in sequential releases
- To avoid adversaries get advantages by gathering victim's information continuously and comparing the multiple releases.
- To show that an adversary can successfully infer a victim's vertex identity and community identity by knowing degrees by calculating number of friend users.

2.PRIVACY PRESERVATION

Data mining [3] plays a vital role in many fields like business, telecommunications, online carts and social networks by providing only needed information from the raw data. Recently, more and more applications have emerged from social networks information. Due to this privacy preservation is identified as a major concern in upcoming development of social network and its applications. Privacy preservation is defined as protecting the each and every information considered as sensitive in the database.

2.1. Privacy preserving data mining (PPDM)

One of the main sub fields in data mining is privacy preserving data publishing (PPDM)[4][5] which provides more development in recent and upcoming years. PPDM is introduced specifically to safeguard sensitive data from unsolicited data for preserving the utility of the particular preserved data. Sensitive raw data and sensitive mining results are the two important phase in which the PPDM consideration is falls.

2.2. Privacy preserving data Publishing (PPDP)

There are four types of users divided according to user role based methodology like data provider, data collector, and data miner and decision maker. In these four users different data transformation process is done. The Privacy preserving data publishing [6][7] technique has been carried out at the process of data collector. Data collector modifies sensitive information send by the data provider. While modifying, data needs to be preserved and maintains its utility of the data simultaneously. There are two phases present in Privacy preserving data publishing they are data collecting and data publishing. The process of recording the incoming data present in the database is data collection and the process of transforming the collected data by the data holder to the data miner is data publishing. The received person performs mining tasks.

3. TARGET OF PROTECTION

Backstrom et al [8] defines, there are two types of attack like active attack and passive attack carried out by the attacker to hack the sensitive data. Active attack means the attacker plants plenty of malicious sub graphs inside the network before its release. When the sub graph grows the, information about the nodes and their relationship has been automatically taken by the attacker. The next attack is passive attack where a small unique subgraph is planted to identify the vertices in the social network.

Privacy preservation is helpful in protecting the graph data against these two common types of attacks [9]. Protection in graph is to be given to two types of sensitive information like nodes or vertices and edges of the graph.

3.1. Vertex or node information

Personal information that attached with the particular user is defined as the vertex information. For example consider an email sent from one user to another user, here mail id, name of the id holder and mobile number is considered as vertex information.

3.2. Multicommunity information

Each user is connected with other different users called friend users. The relationship among them is mentioned as edges in graphs. There is large number of community groups involved in social networks. The information about the different groups in with the user or node participate are known as multicommunity information.

4. DISCLOSURE MODELS

There are large amount of privacy risks present in the social network sites. In that some of the disclosures are mentioned as mentioned below.

4.1. Identity disclosure

Identity defines personal information (address, mobile number, email id, disease, bank details) of a particular user node in the social network graph [10][11]. When an attacker uses some attack model to reveal the personal data is denoted as identity disclosure.

4.2. Link disclosure

Link is a relationship between one user with other users (friend user). In the graph structure edges are mentioned as links [12][13]. Revealing of community details of a friend user by knowing the link details of the victim is known as link disclosure.

4.3. Sensitive attributes disclosure

Backstrom et al.[14]he data that are sent from and to recipient are the sensitive attributes enclosed within the social network. Each node is composed of different attributes like email id, disease details, bank information of a particular user etc.

5. K^W- STRUCTURAL DIVERSITY ANONIMITY

To reduce or solve the privacy risks of vertex and multicommunity disclosure in social network a new privacy model[15] that suite for the dynamic network is introduced known as K^W- Structural Diversity Anonimity.Where K denotes the number of users connected with the particular user known as edges. W denotes the timestamp of each action done at the social network. This technique is completely for safeguarding the private data of all users engaged in the social network. K^W- Structural Diversity Anonimity is a technique purely extractedbased on the concept of k-anonymity. To increase and maintain the utility of data present in social network a table is created known as CS- Table.

5.1. Graph construction

As the first step, anundirected graph should be created according to the users and their relationship with others. In this graph vertex or node represents the user involved in the social network and their relationship along with other members is denoted as edges. Once the owner of the site creates a graph construction of CS-table started.

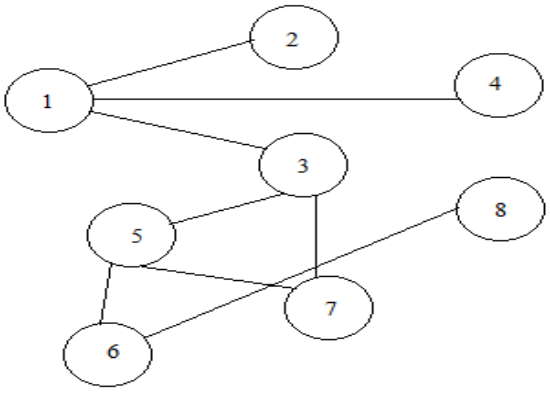

Figure 2.Construction of social network.

From figure 2, construction of social network in the form of graph is demonstrated. Here nodes are given with unique identification with numbers. Each node connected with other nodes is considered as friend users of particular node. In the above graph node 1 is connected with {2,3,4} therefore node 1 friend user is {2,3,4}. Likewise other nodes are connected with their friend user.

5.2. CS-Table

A three columned table named as Cluster sequence table. The table consists of three attributes like vertex, degree and multicommunity attribute. Vertex denotes the personal information a user like unique identification. Degree of the vertex denotes number of friend user connected with the particular user. Multicommunity denotes the groups in which the user involved. User has no limitations to participate in more than one community group. A user can participate in more than one community groups. There are two processes in snapshotting the actions done at the social network. Attributes like vertex information, degree of the vertex and information about community groups are added in the CS-Table once the graph is created. Once the table is created it is automatically updated at each release in the social network.

5.3. Constructing CS-Table

Once a graph based structure is formed according to the information gathered from social network, Cluster Sequence table is created. Table is constructed from three columns like vertex information of a user node, and their degree sequence, which is calculated according to number of friends connected with each user. Lastly details about community groups evolved within the social network and the user participation records.

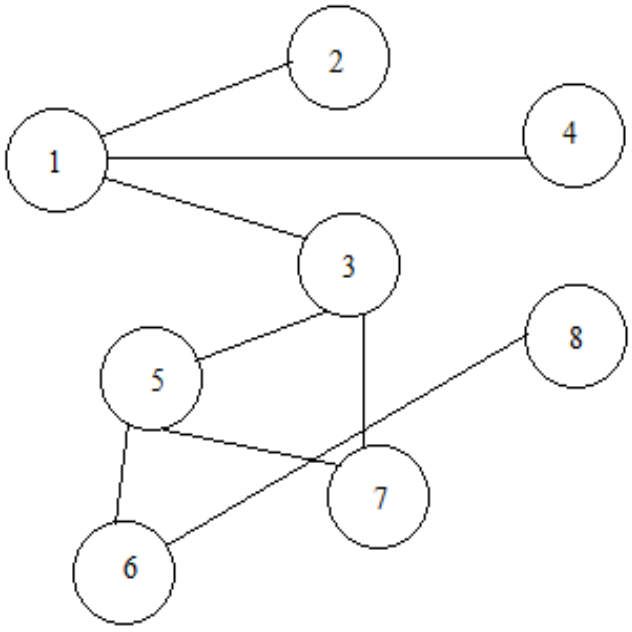

Figure 3.Constructing social network consist of nodes and edges.

Vertex	Degree	Community
1	0-3	{1}-{1}
2	0-1	{1}-{1,2}
3	0-3	{1}-{1}
4	0-1	{1}-{1,2,3}
5	0-3	{1}-{1}
6	0-2	{1}-{1}
7	0-2	{1}-{1}
8	0-1	{1}-{1,2}

Table 1. Constructing CS-Table from undirected graph

From figure 2, the social network structure is entered as entries in Cluster sequence table. At particular time periodsnapshot is taken form the network and marked as a graph. For node 1 its degree is denoted as 0-3 because node 1 have no friends and joined with other three nodes like (2, 3, 4) therefore it is mentioned as 0-3. Next is community identity, if a vertex or user involved in one community then in the table id of that particular community group is entered.

5.4. Updating CS-Table

Once Cluster Sequence Table is created it is automatically update at each time stamp when any actions done at the social network. At each time instance only one change can be accepted.

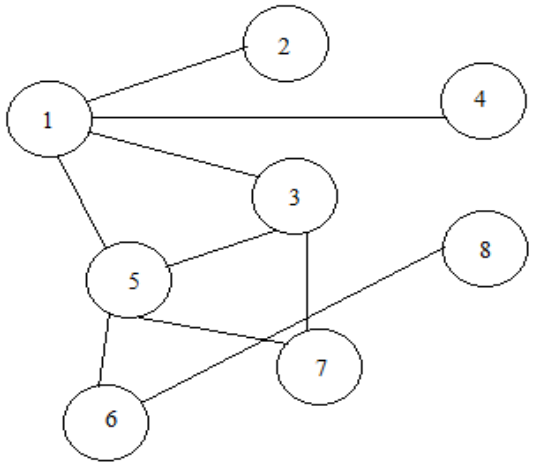

Figure 4.Updation of social network from above graph.

Vertex	Degree	Community
1	3-4	{1}-{1,3}
2	0-1	{1}-{1,2}
3	0-3	{1}-{1}
4	0-1	{1}-{1,2,3}
5	0-3	{1}-{1}
6	0-2	{1}-{1}
7	0-2	{1}-{1}
8	0-1	{1}-{1,2}

Table 2. Updating CS-Table from undirected graph

As shown in above table 2, once any updations done at the social network(figure 3) it is taken as a snapshot and recorded in cluster sequence table and saved as a new snapshot in the database. Here node 1 is connected with another friend node therefore degree of the vertex 1 is modified from 3 to 4. Then coming up to multicommunity details node 1 participate in another community which is denoted with id 3. Therefore vertex 1is changed from {1} to {1,3}.

5.5. Anonymization process

Once all the entries in the cluster sequence table is filled, next process is to anonymizes the original data into fake one. Because when an attacker hacks the network and try to get sensitive data, he provided with the fake data by adding fake vertex at the anonymized table. Grouping all vetex which is involved in same group is also an anonymization process for safeguarding original data.

6. CONCLUSIONS

For data collector, his privacy-preserving objective is to release useful data to data miners without disclosing data provider's identities and sensitive information about them. To achieve this privacy level, the userwants to develop proper privacy models to quantify the possible loss of privacy under different attacks, and apply anonymization techniques to the data. Most of the privacy preservation technique relies on static network for preserving privacy breaches, therefore privacy breaches easily occurs in dynamic networks. Mainly identity disclosure, community disclosure occurs frequently in dynamic network. Therefore future work focus on providing privacy to dynamic network, using K^w – structural diversity model which provides better performance to prevent the disclosures. Where K denotes number of users connects with the particular user and W denotes timestamp of actions done at the network.

REFERENCES

[1] L. Backstrom, D.P. Huttenlocher, J.M. Kleinberg, and X. Lan, "Group Formation in Large Networks: Membership, Growth, and Evolution," Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), 2006.

[2] Lei Xu, Chunxiao Jiang, (Member, IEEE), Jian Wang, (Member, IEEE),Jian Yuan, (Member, IEEE), And Yong Ren, (Member, IEEE) "Information Security in Big Data: Privacy and Data Mining", Received September 21, 2014, accepted October 4, 2014, date of publication October 9, 2014, date of current version October 20, 2014.

[3] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques.San Mateo, CA, USA: Morgan Kaufmann, 2006.

[4] R. Agrawal and R. Srikant, ``Privacy-preserving data mining," ACM SIGMOD Rec., vol. 29, no. 2, pp. 439_450, 2000.

[5] Y. Lindell and B. Pinkas, ``Privacy preserving data mining," in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 2000, pp. 36_54.

[6] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, ``Privacy-preserving data publishing: A survey of recent developments," ACM Comput. Surv., vol. 42, no. 4, Jun. 2010, Art. ID 14.

[7] R. C.-W. Wong and A. W.-C. Fu, ``Privacy-preserving data publishing: An overview," Synthesis Lectures Data Manage., vol. 2, no. 1, pp. 1_138, 2010.

[8] L. Backstrom, C. Dwork, and J. M. Kleinberg. Wherefore art thou r3579x?:anonymized social networks, hidden patterns, and structural steganography. In WWW, pages 181–190, 2007.

[9] James Cheng, Ada Wai-Chee Fu, Jia Liu "K-Isomorphism: Privacy Preserving Network Publication against Structural Attacks", SIGMOD'10, Indianapolis, Indiana, USA, June 6–11, 2010

[10] L. Sweeney. Achieving k-Anonymity Privacy Protection Using Generalization and Suppression. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10(5):571-588, 2002.

[11] L. Sweeney. k-anonymity: a model for protecting privacy. International Journal on Uncertainty Fuzziness and Knowledge-based Systems, 10(5):557-570, 2002.

[12] R.CW. Wong, J. Li, A.WC. Fu, and Ke. Wang. (α, k)-Anonymity: An Enhanced k-Anonymity Model for Privacy Preserving Data Publishing. In Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining, pages 754-759, Philadelphia, PA, 2006.

[13] Bin Zhou and Jian Pei. The k-anonymity and l-diversity Approaches for Privacy Preservation in Social Networks Against Neighborhood Attacks. Knowl. Inf. Syst., 28(1):47-77, July 2011.

[14] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore Art Thou R3579x?:Anonymized Social Networks, Hidden Patterns, and Structural Steganography. In Proceedings of the 16th International Conference on World Wide Web,WWW'07, pages 181-190, New York, NY, USA, 2007. ACM.

[15] Chih-Hua Tai, Peng-Jui Tseng, Philip S. Yu, Fellow, IEEE, and Ming-Syan Chen, Fellow, IEEE,"Identity Protection in Sequential Releases of Dynamic Networks",IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 3, March 2014.

AUTHORS

R.Gowthamy completed her B.E. degree inComputer Science and Engineering from Velalar College Of Engineering AndTechnology Erode, Indian 2014. She is currentlydoing her M.E(Computer Science andEngineering) in NandhaEngineering College(Autonomous), Erode, India.

P.Uma completed her B.Sc degree in Computer Science from Erode Arts college for Women, Erode, India in 2000. She completed her M.Sc degree in Computer Science from Navarasam Arts And Science College For Women,Arachalur,India in 2002. She completed M.E. degree in Computer Science and engineering from Kongu Engineering college, Perundurai, India in 2008. Presently she is working as Assistant Professor in Computer Science and Engineering Department in Nandha Engineering College (Autonomous), Erode, India.