Int. J. Adv. Res. Sci. Technol. Volume 4, Issue 6, 2015, pp.435-439.

IJARST

International Journal of Advanced Research in Science and Technology

journal homepage: www.ijarst.com

ISSN 2319 – 1783 (Print)

ISSN 2320 – 1126 (Online)

Degree Smoothing On Social Networks against Frequent Shared Patterns.

K. Anusha* and K. Venkata Ramana

Dept. of CS&SE, AU College of Engineering, Andhra University, Visakhapatnam, India.

****Corresponding Author's E-mail*: *anusha.konada@gmail.com***

A R T I C L E I N F O

Article history:

Received 19 Sep. 2015

Accepted 05 Oct. 2015

Available online 07 Oct. 2015

Keywords:

Remote Sensing,
GIS,
Habitat Suitability,
LISS and IRS P6.

A B S T R A C T

Advances in technology has made it possible to collect data about individuals and connections between them, such as Email correspondence and friendship. Researchers that have collected such social network data often have a compelling interest in allowing others to analyze the data. However sharing such kind of private information to the public will result in unacceptable disclosure. In this paper we present a framework of, how to provide privacy to individuals in a social network against the adversary from frequent shared patterns. We propose Degree Smoothing method by applying Anonymization and Isomorphism techniques. By taking real world examples we prove that, we can reduce the threat of frequent shared patterns in a social network.

PAPER-QR CODE

IJARST Volume-4, Issue-6

Citation: *Anusha. et al. Degree Smoothing On Social Networks against Frequent Shared Patterns. Int. J. Adv. Res. Sci. Technol. Volume 4, Issue 6, 2015, pp.435-439.*

Introduction:

In recent years, people are very concerned and interested in social network media such as face book, twitter, MySpace etc…without knowing the level of privacy the media is standing on. The data holder of the social network is making more comfortable for the users by providing pop up messages for their online purchases. Individuals think that they are making social network media more flexible and easy way to access things in multitasking way. But the thing is that, when the social network media tie up with the online purchase transactions, and then there is an easy way to the leakage of sensitive information such as name, address, birth etc... In this paper we present some methods to provide privacy to every individual and also provide them from frequent shared patterns. Before generating some new methods to provide privacy we would like to first learn some definitions mentioned in previous papers.

Anonymization:

Anonymization is a technique used to sanitize information whose intent is privacy protection. It is a process of either hiding data or removing personally identifiable information from data sets, so that the people whom the data describe remain anonymous.

Naive Anonymization:

It is a graph G= (V, E) is an isomorphic graph, G_{na}= (V_{na}, E_{na}), defined by a random bijection f: V -> V_{na}. The edges of G_{na} are E_{na} = {(f(x), f(x`) | (x, x`) belongs to E}.

Privacy in social networks:

Before implementing the output, we would like to discuss about a Graph Theory, because all the social network nodes will be shown in the form of graphs only. To make the reader understand the social network in an easy way, we should represent in form of graphs.

Graph Theory: A graph G= (V, E), represents V as a set of nodes and E as set of edges. In a social networking, G represents one particular social networking site i.e, facebook or twitter or g-mail and so on. V represents number of individuals in a social networking and E represents the relationship between two or more individuals.

As we know that social networking sites have been popular in real world and each and every individual has been addicted to it. A Social Networking site is used to

group up many individuals into a block and share their personal information with each other share their posts to the real world in-order to know what's happening in the world and also to gain knowledge.

As it is bringing interest to the individuals, everyone is showing interest to have an account in a social networking site. But after entering, they come to know what kind of privacy they are missing. For example if you take a social network site as facebook, in it when you are creating your profile, it asks you your personal details like name, gender, like's, dislike's and other personal information. So by entering these kind of information, there will be chance to leak your personal details.

Normally it will be fine to fill this kind of data. But once if the adversary is trying to target you in a social network site, then the adversary will have an advantage to know about your personal information. To get a clear understanding about privacy we will give you an example.

In the below graph we can see that the graph has set of nodes and edges. Where V represents set of nodes and E represents the relationship between two or more nodes. Now every individual in the below graph is having the node name, qualification and the items which they have purchased online. When entering into the social networking site with the personal data such as name, qualification and items purchased online will affect the individual nodes if any of the node has been targeted by the adversary.

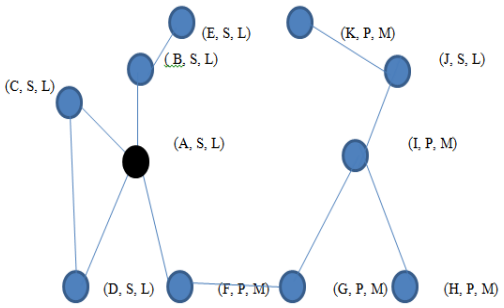

Fig. 1: Social Network Graph

In the below graph if the adversary wants to target node 'A' , then he can easily identify the targeted node because the profile of the node is directly sent into the social networking site without any privacy. To provide privacy for such kind of data we are using a technique called Naive Anonymization. Naive Anonymization is a technique where the original name nodes will be re-named. By doing this way we can provide the targeted node from the adversary. Below graph gives us the example of how Naive Anonymization looks like.

As soon as Naive Anonymization is done we think that we have provided privacy to the nodes. But we got a problem; if the adversary is having external information about the targeted node, then it is hard to provide privacy to the nodes.

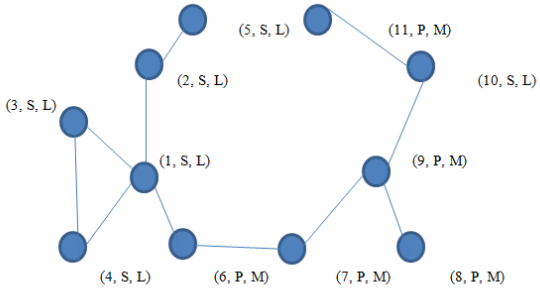

Fig. 2: Naive Anonymization

For example, if the targeted node is '1' and if the adversary knows that node '1' is having four friends and two of the friends are very close. Then the adversary can easily find out the target node is '1' because in the above graph only node '1' is having four friends and in that nodes '4' and '3' are close to each other.

This problem that is been occurred. This problem is known as neighborhood pair attack. With the help of neighbors we can find the targeted node easily. To overcome this problem we should make the graph Anonymous and Isomorphic. This technique is known as Degree Smoothing.

Degree Smoothing is a technique where we should see that atleast two or more nodes degee in a graph should be equal and the sub graph around one graph should be equal and same as that of the other graph.

For the above graph we are applying Degree Smoothing. We are using this method because, our graph consists of set of {V, E, L} where V represents nodes, E represents edges and L represents labels in a graph. Our main part is that we should provide privacy to the nodes based on the frequent shared patterns. Now for example, in the above graph we can see that node '1' is a student and it purchased an item called laptop online. Node '2' is also a student and purchased an item called laptop online. By gathering such kind of information the adversary will separate all the nodes who purchased same items online and will put them a side. By doing this way he can easily identify the target node. To avoid this kind of attack from the adversary we need to add duplicate nodes to the graph. We should see that the duplicate nodes should not affect any of the neighbours and the nodes. When we add edges to the nodes we should see that no two items purchased nodes should be added.

This kind of approach will be solved by Degree Smoothing. This technique should satisfy two conditions:

1. It should be Anonymous
2. It should be Isomorphic

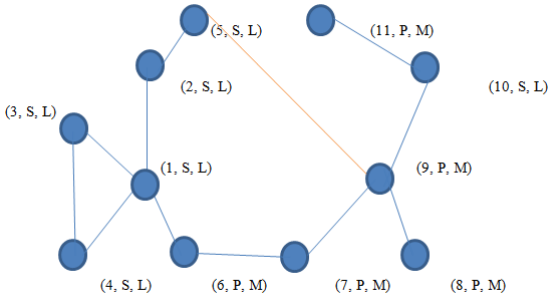

Fig. 3: Anonymous Graph

In the above graph we added an edge between nodes '9' and '5'. Before adding the edges, we should apply degree smoothing technique. Degree Smoothing is a method which should satisfy anonymization and isomorphic techniques.

Now in the above example we added an edge between 9 and 5 because the items purchased for node 9 is different from node 5. First we should make the graph anonymous. To make the graph anonymous we should apply degree smoothing. Arrange all the nodes in descending order based on the degree. Now take the anonymity value K>= 2 and add edges and see that no two nodes should have same item sets.

Steps for applying Degree Smoothing:

1. Arrange the nodes in descending order based on the degree

2. Now take K>=2 where K is the anonymization value

3. Add edges to make the graph anonymize and no two graphs items should be same

4. After making the graph Anonymize now make the graph Isomorphic

After making the graph Anonymize and Isomorphic we will get the final output as the following graph.

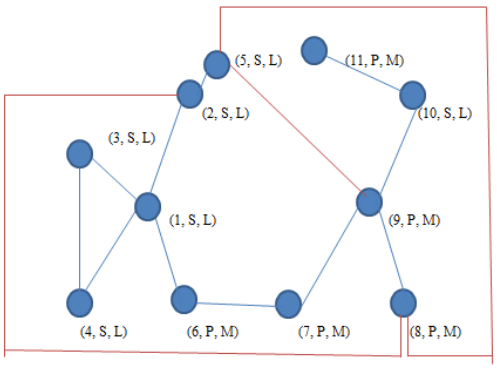

Fig. 3: Anonymize and Isomorphic Graph

The above graph is the final graph which satisfies Degree Smoothing. Now after adding edges to the graph we should see whether the graph is Isomorphic and Anonymize. For checking this process we use BFS rule.

BFS is the process where we start searching the nodes in Breadth First Search. Now for the above graph we start applying BFS rule. The node to be checked is node 5 then 11, 2, 10, 3, 1, 9, 4, 6, 7, 8 like this all the nodes will be checked whether the graph is Anonymous and Isomorphic.

Algorithm:

Degree Smoothing

Input: Arrang the nodes in descending order

Output: Making the graph anonymous and isomorphic

1. for all v $ V, find the degree of each node v

2. sort the nodes in descending order

3. for all nodes(v $ V) do the following

4. compare (vi and vj)

5. find the difference in their degree

6. add the edges between nodes having least degree's such that L values are different

7. end

From the above algorithm we say that by adding noise to the graph we can make the graph isomorphic and anonymous by using degree smoothing technique. So, when we apply the same process to the entire node we will get the final output.

Related Works:

Advances in technology have made it possible to collect data about individuals and the connections between them, such as email correspondence and friendships. Agencies and researchers who have collected such social network data often have a compelling interest in allowing others to analyze the data. However, in many cases the data describes relationships that are private (e.g., email correspondence) and sharing the data in full can result in unacceptable disclosures. Michael Hay proposed a framework for assessing the privacy risk of sharing anonymized [1] network data. This includes a model of adversary knowledge, for which we consider several variants and make connections to known graph theoretical results.

On several real-world social networks, we show that simple anonymization techniques are inadequate, resulting in substantial breaches of privacy for even modestly informed adversaries. We propose a novel anonymization technique based on perturbing the network and demonstrate empirically that it leads to substantial reduction of the privacy threat. We also analyze the effect that anonymizing the network has on the utility of the data for social network analysis.

Vertex re-identification is one of the significant and challenging problems in social network. Mohd Izuan Hafez Ninggal shows a new type of vertex re-identification attack [2] called neighbourhood-pair

attack. This attack utilizes the neighbourhood topologies of two connected vertices. We show both theoretically and empirically that this attack is possible on anonymized social network and has higher re-identification rate than the existing structural attacks.

One of the most important concerns in publishing social network data for social science research and business analysis is to balance between the individual's privacy protection and data utility. Recently, researchers have developed lots of privacy models and anonymous techniques to prevent re-identifying of relevant information of nodes through structure information of social networks, but most of the existing methods did not cater for the individuals' personalized privacy requirements and did not take full advantage of distributed characteristics of the social network nodes. Jia Jiao and Peng Liu specifies three types of privacy attributes for various individuals and develop a personalized k-degree-ldiversity (PKDLD) [3] anonymity model. Furthermore, we design and implement a graph anonymization algorithm with less distortion to the properties of the original graph. Finally, we conduct experiments on some real-world datasets to evaluate the practical efficiency of our methods, and the experimental results show that our algorithm reduces the anonymous cost efficiently and improves the data utility.

Social network data provide valuable information for companies to better understand the characteristics of their potential customers with respect to their communities. Yet, sharing social network data in its raw form raises serious privacy concerns because a successful privacy attack not only compromises the sensitive information of the target victim but also the relationship with his/her friends or even their private information. In recent years, several anonymization techniques have been proposed to solve these issues. Most of them focus on how to achieve a given privacy model but fail to preserve the data mining knowledge required for data recipients. Benjamin C. M. Fung, Yan'an Jin and Jiaming Li proposed a method to k-anonymize a social network dataset with the goal of preserving frequent sharing patterns, one of the most important kinds of knowledge required for marketing and consumer behaviour analysis. Experimental results on real-life data illustrate the trade-off between privacy and utility loss with respect to the preservation of frequent sharing patterns.

Experimental Analysis:

In the Smoothing Degree algorithm, the heuristic function first selects the vertices with the lowest degree and then computes the impact on spatterns. The computational complexity of the algorithm is O(knlogn), where k is the anonymization threshold, n is the number of the vertices with the lowest degree. In the Make Isomorphic algorithm, we use BFS code to encode the 1-neighborhood a given vertex. We also need to find those affected vertices and put them into V List again. Considering the worst case, the computational complexity of the algorithm is O(k3 × |V | + k × |V |2), where k is the anonymization threshold and |V | is the number of vertices inmN1(vi), where vi ∈ TopK.

Data Utility on Frequent Spatterns

The first experiment is to evaluate the impact of anonymization on frequent spatterns. The utility loss is calculated by FSLoss = A−B/B , where B and A denotes the number of frequent spatterns extracted before and after anonymization, respectively. The value of FSLoss is non-negative. The higher value of FSLoss means the higher number of false positive frequent spatterns, implying higher utility loss. Figures below depicts the utility loss on frequent spatterns with anonymization threshold $5 \leq k \leq 20$, and minimum support MinSup = 8%, 12%, 16%, 20% on p2p-Gnutella08 and p2p-Gnutella05. For example, at MinSup = 16%, FSLoss = 21.3%, 22.7%, 26.7%, 28% for $5 \leq k \leq 20$, respectively. This result suggests that as k increase, more fake edges have to be added in order to achieve the k-anonymity requirement, resulting in higher FSLoss. Yet, the impact of anonymization on FSLoss is mild.

Privacy threats on social network data can be summarized into three types, namely identity disclosure, attribute disclosure, and link re-identification, depending on the adversary's background knowledge on the social network data. For identity disclosure, the attack goal is to identify the vertex that represents a target victim. For attribute disclosure, the attack goal is to identify or infer some sensitive information about a target victim. For link re-identification, the attack goal is to identify sensitive relationships of a target victim. We briefly review the related works on thwarting identity disclosures with anonymization technique in the enhanced model, which represents the social network data as a graph in which labelled vertices denote participants and their associated information such as jobs and purchased items and edges denote the relationships between participants. Existing anonymization techniques for thwarting identity disclosure on social networks are primarily classified into three categories: adding or removing edges, generalization, and randomization.

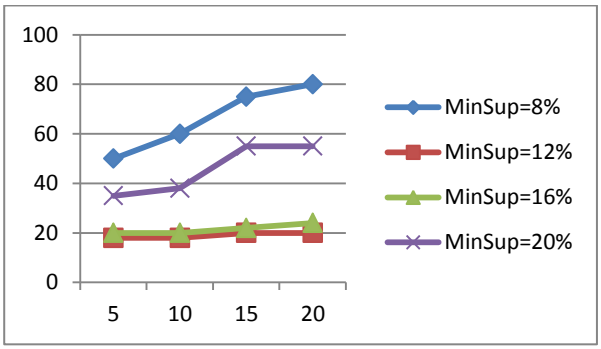

Fig. 4: participants and their associated information

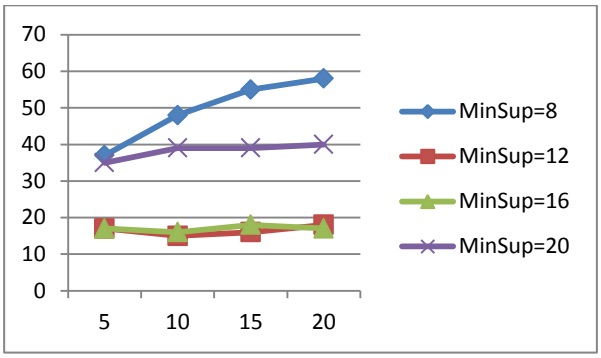

Fig. 4: participants and their associated information

Conclusion:

In this paper, few recent anonymization techniques for privacy preserving publishing of social network data are studied. We discussed the issues in privacy preservation in social network data comparing to the relational data, and examined the possible problem formulation in three important aspects: privacy, background knowledge, and data utility. We reviewed the anonymization methods for privacy preserving in three categories: Identity, Links and labels. We identified a problem on labels based on different background knowledge such as vertex degree and vertex degree pair edge.

References:

1. Hay, M., Miklau, G., Jensen, D., Weis, P., Srivastava, S:" Anonymizing social networks". Computer Science Department Faculty Publication Series, pp. 180, 2007
2. Aleksandra Korolova, Rajeev Motwani, Shubha U. Nabar, Ying Xu:" Link Privacy in Social Networks". In Proceedings of the 17th ACM conference on Information and knowledge management, pp. 289-198, 2008
3. Zhou, B., Pei, J.:" Preserving privacy in social networks against neighborhood attacks". In proceedings of IEEE 24th International Conference on Data Engineering, ICDE 2008, pp. 506–515. IEEE, 2008
4. A. Narayanan, V Shmatikov.: "De-anonymizing Social Networks". In proceedings of Security and Privacy, 30th IEEE Symposium, pp. 173-187, 2009
5. Jia Jiao, Peng Liu, and Xianxian Li.: "A Personalized Privacy Preserving method for Publishing Social Network Data". In proceedings of 11th Annual Conference, TAMC 2014-Springer, pp. 141-157, 2014
6. Mohd Izuan Hafez Ninggal and Jemal H. Abawajy.: "Neighborhood-Pair attack in Social Network Data Publishing". In proceedings of 10th International Conference, MOBIQUITOUS 2013-Springer, pp. 726-731, 2014
7. Benjamin C. M. Fung, Yan'an Jin, Jiaming Li.: "Preserving Privacy and Frequent Sharing Patterns for Social Networking Data Publishing". 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining