Volume 5, No. 7, July 2014
Journal of Global Research in Computer Science
ISSN-2229-371X
REVIEW ARTICLE
Available Online at www.jgrcs.info

A Study of Privacy Attacks on Social Network Data

*Sri Krishna Adusumalli, Valli Kumari Vatsavayi, Jyothi Vadisala
Department of Computer Science and Systems Engineering, Andhra University
Visakhapatnam, Andhra Pradesh, India -530 003
*srikrishna.au@gmail.com, vallikumari@gmail.com, jyothi.vadisala@gmail.com

Abstract: Online social networks have become an important component of the online activities on the network and one of the most influencing media. Unconstrained by physical spaces, online social networks extend to web users new interesting means to communicate, interact and socialize. While these networks get to frequent data sharing and inter-user communications instantly possible, privacy related issues are their obvious much discussed immediate consequences. Although the impression of privacy may take different forms, the ultimate challenge is how to prevent privacy invasion when much personal information is useable. In this context, we address privacy related issues by resorting to social network analysis. Most of the state-of-art methods focus on vertex re-identification or identity disclosure in a social network. However, in real world social network scenario, vertices are usually associated with sensitive data like disease in health networks. In this paper, we study the literature on privacy in social networks. We formally specify the possible privacy breaches and describe the privacy attacks that have been examined. We represent different categories of privacy disclosures, background knowledge in concert with existing privacy preserving techniques in social network data publishing. We identify a new challenge in sensitive attribute disclosure based on different background knowledge like vertex degree pair of edge and vertex degree.

Keywords: social network, privacy preserving, data publishing, attribute disclosure.

INTRODUCTION

Nowadays, online social media services are growing rapidly day by day and it has given an impact on the way people interact with each other. The Online social networks such as Twitter, Facebook and LinkedIn have become one of the most popular activities on the web [1]. According to the recent study, more than 80% of the university students in America are active members of online social network and spending 30 minutes on average in everyday life [2]. Most of the business owners actively use social network as part of their marketing strategy. These social networks collect huge amount of data about user and their activity and relations. On positive side, this collected data gives great analysis opportunity to data miners/researchers, and on the negative side the data gives a threat to user's data privacy.

The collected data can be used by the researchers to study the disease propagation in health network for analysis purposes and for government institutions in mining social network data for information and security purposes [3]. To study the usefulness of the data, it needs to be shared or published to the public. If this data is directly available to researchers, it will cause the privacy disclosure, which leads us to study how to protect the identity and sensitive information of social network data effectively.

The privacy disclosure in a social network can be grouped to three categories: **1) Identity disclosure**: the identify of an individual who is associated with a vertex is revealed; 2) **Link disclosure**: the sensitive relationships between two individuals are disclosed; 3) **Sensitive attribute disclosure**: the sensitive data associated with each node is compromised e.g., the email message sent/received by the individual in an email communication network. A privacy preservation system over graph and networks should consider all of these issues.

However, most of the privacy preserving data publishing model like *k*-anonymity [4][5], *l*-diversity [6], (α,k)-anonymity [7] are studied extensively can be dealt with relational data only. These methods cannot be applicable to social network data directly. In practice, anonymizing the social network data is much more challenging than relation data due to following issues [8].

First, modeling the background knowledge of the social network is much trickier than relational data. Because in relational data, the attributes in dataset are divided into two categories: quasi-identifier attributes and sensitive attributes. The adversary uses quasi-identifiers as background knowledge and link with external table to identify the individual. However, in social network data, it is much more complicated. Because, the adversary can use any piece of information to identify the vertex or individual such as vertex attributes or properties, vertex sensitive label, vertex degree, neighborhood structure, vertex degree pair of an edge and vertex or edge labels.

Second, developing anonymization method for social network data is much complex than relational data. Because in relational data, the group of tuples can be anonymized without affecting other tuples in the dataset. Relational data researchers extensively follow divide-and-conquer method to anonymize the data. However, in anonymizing social networks data, divide-and-conquer method may not be useful. Because, adding or removing edge may affect other vertices and edges as well as structural properties of social network. Therefore, anonymizing social network data may use some specific method by considering all the properties of the network.

Third, measuring information loss in anonymized social network data is much challenging than that in relational data. In relational data, the information loss is measured either tuple by tuple or record by record. For a given tuple in

the original data set and the corresponding tuple in the anonymized table the distance is used to measure the information loss at tuple level. The sum of information loss at each tuple is used to measure the information loss at table level. However, social network data consist of set of vertices and set of edges. It is hard to compare two social networks by comparing the vertices and edges individually. Social network properties like connectivity, diameter and betweenness of the social network will not be same even if two social networks have the same number of vertices and edges. So there are different ways to measure the quality and information loss in social network.

To combat these challenges, several researchers have recently proposed different types of privacy models, adversaries, and graph modification algorithms. Unfortunately, none of the work is linked to solve all the problems in one shot. Protecting against each kind of privacy disclosures may require different methods or combination of them.

BACKGROUND KNOWLEDGE MODEL

Background knowledge is the piece of information that is known to the adversary and can used by the adversary to infer the privacy of an individual. In social network data, the information that can be used as background knowledge to intrude into user privacy are personal attributes and structural attributes. With these two, an adversary may conduct different types of attacks against social network privacy. Therefore, background knowledge plays an important role in modeling privacy attacks on social network data and developing different anonymization methods to protect privacy. In privacy preservation in publishing social networks, due to the complex structures of graph data, the background knowledge of adversaries is modeled in following ways:

- ***a. Identifying attributes of vertices:*** A vertex may be linked uniquely to an individual by a set of attributes, where the set of identifying attributes play a role similar to a quasi-identifier in the re-identification attacks on relational data. Vertex attributes are often modeled as labels in a social network. An adversary may know a few attribute values of some victims or target individuals. Such background knowledge may be misused for privacy attacks [9].
- ***b. Vertex degrees:*** The degree of a vertex in the social network captures how many direct edges or relationships the corresponding individual is connected to others in the social network. This information does not carry any sensitive information but is potentially used as background knowledge to find the individual in a social network. This information can be easily collected by the adversary. For example in Facebook, Twitter and LinkedIn social networks, an adversary can acquire the number of friends of an individual by viewing the individual profile [10] [11] [12] [13].
- ***c. Neighborhood:*** This refers to a set of neighboring individuals that has direct social links to a target individual which they also have mutual link between them. This neighborhood information of some individuals known by the adversary may lead to privacy breach. For instance, if an adversary knows that the target individual has four best friends who also have link/edge to each other, the adversary could use this information to query the data to map individual that has neighborhood contain a clique size of four [8][14].
- ***d. Sub-graph:*** This refers to a set of relationships in which the target individual is connected to a graph which is a subset of the whole graph. For example, an adversary may create a set of dummy profiles and create social edge/link between these profiles with specific patterns. The adversary then uses those dummy profiles to set up a social edge to target individual. The social link can be established as easy as adding target to the friend list. Another way is, the adversary naively constructs a combination with other friends which form a small individually identifiable sub-graph [15].
- ***e. Vertex degree Pair of an edge:*** The vertex degree pair of an edge in social network contains vertex degrees of two individuals and their friendship relation. The adversary use this as background knowledge to re-identify the target individual and his friend as well as associated personal sensitive information like disease and other activities in the social network. This type of attack is called as friendship attack [16].
- ***f. Link relationship:*** An adversary may know that there are some specific link relationships between some target individuals. Link disclosure occurs when sensitive link structure information is leaked as a result of social network data publication, or inferred by compromised social network users. For example, in a social network, edges may carry labels recording the channels individual use to communicate with each other such as phone, email or messaging. An adversary may try to use the background that a target individual uses only phone to contact her friends in the network to link the target individual to vertices in the social network [9] [13] [17].

Backstrom et al considered two different types of attacks [15]. The first, called an active attack, involves creating new user accounts and establishing relationships with existing users. This allows a malicious data recipient, or attacker, to identify the "fake" users that were created and their relationships to other users in the published data. When some of these relations are sensitive, link disclosure may occur as a result.

The second type of attack is called passive and involves an attacker who has not tampered with the network data prior to its publication as in the active attack, but is able to locate himself or herself in the published data, as well as sensitive relations of users that he or she is related to.

CATEGORIES OF ANONYMIZATION METHODS

Generalization, suppression and perturbation are three well noted anonymization schemes for relation data. Although privacy preserving social network data is a new challenge, several privacy models and methods have been developed. In social network data publication different methods are proposed to model the different background knowledge, because one method cannot solve all the problems in one shot. We categorize the existing anonymization methods on

social network data publication into three categories such as 1) Identity preserving methods, 2) Link preserving methods, and 3) Sensitive attribute preserving methods. The proposed models in each of these categories are briefly discussed in following sections.

Identity Preserving Methods:

Identity preserving model deals with protecting individual identity from being re-identified. Formally, the problem can be defined as: Given a published social network data, if an adversary can identify the vertex of a target individual by analyzing topological features of the vertex based on his background knowledge about the individual from the social network, then the identity of target individual is disclosed.

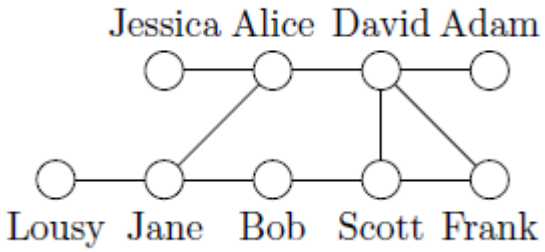

Figure 1. Original social network G.

Table I. Original Graph Data

Id	***Name***	***Degree***	***Disease***
1	Jessica	1	HIV
2	Alice	3	HIV
3	David	4	Flu
4	Adam	1	HIV
5	Lousy	1	Cancer
6	Jane	3	Flu
7	Bob	2	Cancer
8	Scott	3	Cancer
9	Frank	2	Cancer

A simple way to protect individual from being re-identified, the data is anonymized by removing the individual identifiable information such as Name so as to de-associate the vertices from the specific real-world individual. This conventional way is known as Naive Anonymization as shown in Figure 2. However, even after removing individual identifiable attribute, it is not sufficient to guarantee privacy [15], because the structure of the graph is retained unmodified.

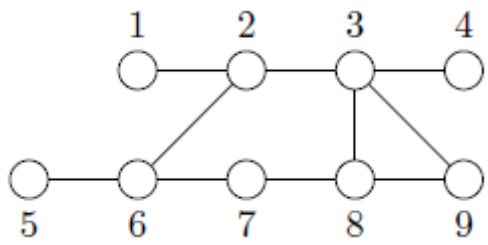

Figure 2. Naive Anonymized graph G*.

For Example, if the adversary wants to re-identify the vertex of David and he also knows that David has four friends in the graph then the adversary identifies that the vertex labeled 3 corresponds to David as this is the only node with degree '4' (see Figure 2).

Recently several anonymization methods are developed to prevent vertex re-identification attacks based on graph features [10] [12] [14] [16] [18]. Many of these methods are also inspired by the *k*-anonymity concept in relational data. In the following sub section, we discuss the identity preserving models in terms of graph structural features they use.

K-*Candidate Anonymity:*

Hay et al. studied the problem of re-identifying a target individual in the naively-anonymized social network [11]. They noted that the structural similarity of the vertices in the social graph and the background knowledge an adversary obtains jointly determines the extent to which an individual can be distinguished. For instance, if an adversary knows that target individual in social network has exactly 6 friends, then the adversary locate all the vertices in the network with degree 6. If there are very limited vertices satisfying this, then the target individual might be uniquely identified.

To overcome this, Hay et al. proposed a privacy model named as *k*-candidate anonymity for social networks, which is based on the notation of *k*-anonymity [19]. A social network or graph satisfies *k*-candidate anonymity with respect to a structural query if the number of the matching candidate vertices is at least *k*. The query evaluates the existence of the neighbors of a vertex or the structure of the sub-graph in the neighborhood of a vertex. It implicitly models the background knowledge of an adversary using the following two types of queries.

Vertex refinement queries: This group of queries, with increasing adversary knowledge, models the local neighborhood structure of a vertex in the social network.

Sub-graph queries: This class of queries verifies the existence of a subgraph around the target vertex. The descriptive power of a subgraph query is measured by the number of edges in the subgraph.

To protect against these types of attacks, Hay et.al proposed a random perturbation technique that modifies the graph through sequence of random edge deletion followed by insertions. This method can potentially reduce the risk of re-identification but it does not guarantee that the modified or anonymized graph satisfies *k*-candidate anonymity. Also it does not guarantee the utility of the original graph can be well preserved.

K-Degree Anonymity:

Liu and Terzi identified the problem of identity re-identification based on degree structural properties [12]. For instance, given the naive anonymized version of social network in Figure 2 in which the identify attributes has been removed. Assume that the target individual is David. If the adversary knows that David has four friends (four degree), by using vertex refinement queries adversary could uniquely identify David has vertex 4 in the naive anonymized graph.

To prevent vertex re-identification through vertex degree as background knowledge, the authors proposed the model of *k*-degree anonymity. The *k*-degree anonymity states that for every vertex, there should be other *k* similar vertices that are indistinguishable base on the number of degree. The *k*-degree anonymity is achieved using following two steps.

a. First, starting from the original graph degree sequence, construct a new degree sequence that is *k*-anonymous in such a way that the anonymized cost is minimized.
b. Second step involves constructing new *k*-anonymous graph according to the new degree sequence. This step adopts dynamic programming to get optimal degree sequence and greedy-based edge swapping for graph transformation strategy.

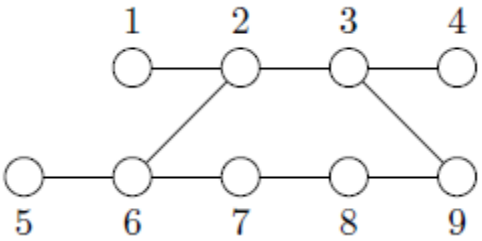

Figure 3. 2-degree anonymized graph G*.

For example, consider the Figure 3 is 2-degree anonymous because each cluster contains at least 2 vertices with respect to vertex degree.

K-Neighborhood Anonymity:

Zhou and Pei studied the subgraph constructed by the immediate neighbors of a target vertex/individual [8] [14]. The assumption is that the unique structure of the neighborhood subgraph can be used by the adversary to differentiate the target individual from the others in the social network. This assumption is closely related to subgraph knowledge queries. Based on this assumption, the authors proposed a new notation of the anonymity on social network named as *k*-neighborhood anonymity. The *k*-neighborhood anonymity states that a graph is k-anonymous, if for every vertex there exists at least *k*-1 other vertices that share isomorphic neighborhoods. The *k*-neighborhood anonymity is achieved using following three steps.

a. First, it marks all the nodes as “unanonymized" and sorts them in descending order of their neighborhood size. Here neighborhood size is defined as the number of edges and nodes of the subgraph constructed immediate neighbors of a vertex. Then the algorithm picks up the first “unanonymized" vertex from the sorted list, find the top (*k*-1) other nodes from the list whose neighborhood subgraphs are most similar to that vertex.
b. Second, iteratively considers every pair of vertices and for each pair the algorithm modifies their neighborhood subgraph to make them isomorphic to each other. This modification is achieved by adding extra edges while keeping the vertices intact.
c. Third, after all the neighborhood subgraphs of these *k* vertices are pair-wise isomorphic, the algorithm marks these *k* vertices as “anonymized". This process is continued until all the vertices in the graph are “anonymized".

K-Automorphism Anonymity:

Zou et.al [18] proposed the method of *k*-automorphism based on the assumption that the adversary may study and know the sub-graph around the target individual. If such sub-graph is unique in the anonymized social network graph, then the target vertex *v* in the sub-graph still has the treat of identity disclosure. The aim of *k*-automorphism model is to construct a new graph so that any sub-graph around a vertex *v*, there are at least *k* similar sub-graphs isomorphic to *v*. To achieve *k*-automorphism the authors proposed k-match algorithm that employs heuristic approach in anonymization process. This anonymization process is done through a series of graph alignment and edge copy which introduces new edges within and among partition groups. To maximize the utility, edge addition should be performed minimally which implies that sub-graphs within one group should be very similar to each other.

k^2-Degree Anonymity:

Tai et.al identify a new type of attack, called a friendship attack, based on the vertex degree pair of an edge [16]. Using the vertex degree of two individuals and their friendship relation, the adversary can issue a friendship attack on the published social network to re-identify the vertices corresponding to an adversary. For instance consider the Figure 1, the adversary knows that Adam and David are friends and (1,4) is the vertex degree pair of Adam and David as background knowledge. The adversary use this information and then identify the target individuals uniquely from naive anonymous graph shown in Figure 2 with vertex degree pair (1,4).

To prevent friendship attacks, Tai et.al proposed the notation of k^2-degree anonymity, which ensures that the probability of a vertex identity being revealed is not greater than 1/*k* even if an adversary knows a certain degree pair of two vertices. For achieving *k*-degree anonymity the authors proposed two approaches i) Integer Programming formulation to find optimal solutions for small datasets. ii) Scalable Heuristic approach for anonymizing large scale social network data.

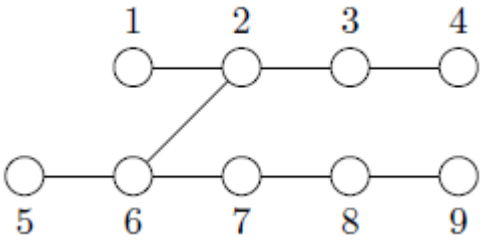

Figure 4. 2^2-degree anonymized graph G*.

Link Preserving Methods:

Link disclosure occurs when sensitive link structure information is leaked as a result of social network data publication, or inferred by compromised social network users. Inferring link structure from anonymized data, the social network owner wants to publish the social network to untrusted recipients for analysis purposes in a way that sensitive relationships between users cannot be inferred from the published data, for example, through the use of graph mining techniques. Some of the techniques are described int this section to solve the problems.

Link Re-identification:

Zheleva and Getoor [17] identified the problem of link re-identification, which they define as inferred sensitive relationships from anonymized graph Data. Entities are represented as Graph nodes and the edges are the relationships among the nodes. Edges are classified as either sensitive or observed and addressed the main problem to downplay the chance of predicting sensitive edges based on observed edges. Utility is measured by the number of observational edges removed. If removal of observations is high, then lower is the overall utility. This is accomplished by using one of the five approaches outlined in the report. Their first approach is called Intact edges, in which it contains only observational edges. The second approach, called Partial edge removal, deletes observational edges that may lead to the inference of a sensitive relationship. In the first two approaches, the number of nodes in the graph was unchanged and the edges constructed as links between their versions. In the cluster- edge approach, all the nodes are broken up into a single node (per cluster) and a decision is reached on which edges to include in the collapsed graph. The Cluster-edge with constraints approach uses a more restrictive technique for the observed edges, by creating edges between equivalence classes if and only if the equivalence class nodes have the same constraints as any two nodes in the original graph. The final approach, called removed edges, deletes all relationships/edges from the graph. They recognize that the potency of the approaches depends on the morphological and statistical characteristics of the underlying graph. In short, Zheleva and Getoor concentrated on an often aspect of link disclosure - mitigating the danger of link re-identification.

Random Perturbation for Private Relationship Protection:

Perturbation of social network data has also been applied to thwart link disclosure attacks by Ying and Wu [13]. The authors studied how a graph can be printed in a form that conceals the sensitive connections, while preserving the spectrum of the original graph, which reflects topological properties including diameter, long ways, and the cluster structure. They proposed two methods on edge modification. The first one repeatedly adds a random edge, and subsequently deletes another edge, so as to preserve the number of edges. The second approach swaps pairs of edges in a mode that the degree distribution of the original graph is not impressed.

Synthetic Graph Generation:

Instead of modifying the graph structure, Lescovec and Faloutsos [22] proposed to generate a graph with a different structure than the original, but with similar topological properties, such as degree distribution, diameter, or spectrum instead of modifying the structure of a graph. The intuition behind this approach is that the resultant graph would still protect privacy, while allowing graph analysis in practice. An efficient, linear algorithm based on a graph fitting model was also produced. The fundamental premise of this algorithm is that the employed graph fitting model is capable to generate graphs that obey many of the patterns found in real graphs.

Sensitive Attribute Preserving Methods:

In Social network even if the graph is *k*-anonymous using any of the above identity preserving models there is a possibility of privacy leak. The adversary use any of the structural background knowledge to identify the sensitive value of an individual if cluster of vertices anonymized together share same sensitive information or the probability of particular sensitive value in the cluster is larger than remaining sensitive values, even if we can't find out which vertex is associated with a particular individual. In this context very limited work has been done on sensitive attribute disclosure, and there has been a growing interest in it. The existing work looks at the identifying structural properties of the graph vertices or considers relations to the attributes of vertex.

Zhou and Pei extended their work in [14] and introduced *l*-diversity into social network anonymization [8]. In this case, each vertex is associated with some attributes including identifier, quasi-identifier and sensitive attributes. If an adversary can re-identify the sensitive attribute values of one target individual with a high confidence, the privacy of that individual is breached. An *l*-diverse graph makes sure that the adversary cannot infer the sensitive attribute value with a confidence over 1/*l*. The authors extend the *k*-anonymity method developed against the 1-neighborhood attack to handle the *l*-diversity problem.

Yu et al proposed a graphic *l* -diversity anonymous model for preserving privacy in social network data, which could protect vertex re-identification as well as vertex sensitive attribute re-identification based on *k*-degree anonymous [20]. The author proposed heuristic algorithm which could transform the original graph to an *l* -diversity via the three anonymous strategies namely Adjust Group, Redirect Edges and Assign Residue respectively.

Regarding social network data [8] [20] [21] have studied *l*-diversity, but there are certain shortcomings. Tai et.al does not restrict the frequency of sensitive attribute, so it could not protect probability inference attack. Zhou and Pei [8] do not protect sensitive attributes against other background knowledge based attacks except neighborhood attack. Yu et.al [20] protects sensitive attributes based on *k*-degree anonymity concept and the authors do not model against the other background knowledge.

Unfortunately, none of the works are linked to solve all the sensitive attribute disclosure problems in one shot. Protecting against each kind of privacy background knowledge or attack may require different methods or combination of them. The following two examples demonstrate the sensitive attribute disclosure based on vertex degree pair of an edge and vertex degree as background knowledge.

Example 1: Consider social network graph in Figure 1. In this each vertex in the graph is associated with sensitive attributes shown in Table 1. Figure 3 is a 2-degree anonymous of Figure 2. Assume the adversary knows the vertex degree of bob is 2 as background knowledge. The adversary cannot identify the vertex of Bob directly in Figure 2 because the graph is 2-degree anonymous. However, the vertices Bob, Scott and Frank have same vertex degree and no other vertex has the degree, the adversary is sure that Bob must be in one of the three

vertices. However, the adversary identifies the Bob's sensitive information because all these three vertices share the same sensitive information.

Example 2: Figure 4 is 22-degree anonymous of Figure 1. If the adversary knows that Bob and Scott are friends and (2,2) is the vertex degree pair of Bob and Scott as background knowledge. The adversary cannot identify the vertex of Bob directly in Figure 2 because the graph is 2- degree and 22-degree anonymous. However, the vertices Bob and Scott have same vertex degree pair (2,2) and no one else has the same degree pair. However, the adversary identifies the bob sensitive information with 100% because both these vertices share same sensitive information.

The above two examples clearly demonstrate that a *k*-degree or k^2-degree anonymized social network may still disclose sensitive information due to lack of diversity.

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, Link and Sensitive attribute disclosure methods. We identified a problem on sensitive attribute disclosure based on different background knowledge such as vertex degree and vertex degree pair edge.

REFERENCES

[1] Alexa. The top 500 sites on the web. http://www.alexa.com/topsites, 2011.

[2] Joseph Bonneau and Sren Preibusch. The Privacy Jungle: On the Market for Data Protection in Social Networks. In Tyler Moore, David Pym, and Christos Ioannidis, editors, Economics of Information Security and Privacy, pages 121-167. Springer US, 2010.

[3] D. Rosenblum. What Anyone Can Know: The Privacy Risks of Social Networking Sites. Security Privacy, IEEE, 5(3):40-49, May 2007

[4] 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.

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

[6] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam *l*-diversity: Privacy Beyond *k*-anonymity. ACM Trans. Knowl. Discov. Data, 1(1), March 2007.

[7] 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.

[8] 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.

[9] Alina Campan and Traian Marius Truta. A Clustering Approach for Data and Structural Anonymity in Social Networks. In Privacy, Security, and Trust in KDD Workshop PinKDD, 2008.

[10] Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis. Resisting Structural Re-identification Anonymized Social Networks. Proc. VLDB Endow., 1(1):102-114, August 2008.

[11] Michael Hay, Gerome Miklau, David Jensen, Philipp Weis, and Siddharth Srivastava. Anonymizing social networks. Technical Report, University of Massachusetts Amherst, pages 07-19, March 2007.

[12] Kun Liu and Evimaria Terzi. Towards Identity Anonymization on Graphs. In Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, SIGMOD '08, pages 93-106, New York, NY, USA, 2008. ACM.

[13] Xiaowei Ying and Xintao Wu. Randomizing Social Networks: A Spectrum Preserving Approach. In SDM, pages 739-750. SIAM, 2008.

[14] Bin Zhou and Jian Pei. Preserving Privacy in Social Networks Against Neighborhood Attacks. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, ICDE '08, pages 506-515, Washington, DC, USA, 2008. IEEE Computer Society.

[15] 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.

[16] Chih-Hua Tai, Philip S. Yu, De-Nian Yang, and Ming-Syan Chen. Privacy-preserving Social Network Publication Against Friendship Attacks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11, pages 1262-1270, New York, NY, USA, 2011. ACM.

[17] Elena Zheleva and Lise Getoor. Preserving the Privacy of Sensitive Relationships in Graph Data. In Proceedings of the 1st ACM SIGKDD International Conference on Privacy, Security, and Trust in KDD, PinKDD'07, pages 153-171, Berlin, Heidelberg, 2008.Springer-Verlag.

[18] Lei Zou, Lei Chen, and M. Tamer Ozsu. K-automorphism: A General Framework for Privacy Preserving Network Publication. Proc. VLDB Endow., 2(1):946-957, August 2009.

[19] P. Samarati and L. Sweeney. Generalizing Data To Provide Anonymity When Disclosing Information. In Proceedings of the 17th ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems, PODS'98, page 188, Seattle, Washington, USA,1998.

[20] Liangwen Yu, Jiawei Zhu, Zhengang Wu, Tao Yang, Jianbin Hu, and Zhong Chen. Privacy Protection in Social Networks Using *l*-Diversity. In ICICS, pages 435-444. Springer Berlin Heidelberg, 2012.

[21] Chih-Hua Tai, Philip S. Yu, De-Nian Yang, and Ming-Syan Chen. Structural Diversity for Privacy in Publishing Social Networks. In SDM, pages 35-46. SIAM / Omnipress, 2011.

[22] Jure Leskovec and Christos Faloutsos. Scalable modeling of real graphs using Kronecker multiplication. In Proceedings of 2007 International Conference on Machine Learning (ICML'07), pages 497–504, Corvallis, OR, June 2007.

Short Bio Data for the Authors

 **A. Sri Krishna** received B.Tech degree in Information Technology from JNTU Hyderabad, Andhra Pradesh, India in 2007. He received his M.Tech in 2009 from Andhra University and is currently pursuing his Ph.D. in Computer Science and Systems Engineering at Andhra University. He also worked as a Senior Research Fellow (SRF) in DST funded project by Department of Science and Technology, Ministry of Science and Technology, Government of India. His research interests include Data Privacy, Social Networks, and Big Data Analysis. He is a student member of IEEE.

 **Dr. V. Valli Kumari** is currently a Professor in Computer Science and Systems Engineering department and is also Honorary Director of Andhra University Computer Centre. She has over twenty two years of teaching experience. She was awarded a gold medal for the best research in 2008 by Andhra University. Her research areas include Web Mining, Data and Security Engineering and have 90 publications in various conferences and journals of international and national repute. She is an active member of IEEE, ACM, CRSI and CSI. She is also the founder vice-chair for IEEE Vizag bay Sub-Section and a fellow of IETE.

 **V. Jyothi** received B.Tech degree in Information Technology from JNTU Hyderabad, Andhra Pradesh, India in 2005. He received his M.Tech in 2010 from Andhra University and is currently pursuing his Ph.D. in Computer Science and Systems Engineering at Andhra University. Her research interests include Data Privacy, Social Networks, and Image Processing.

