International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-3, Issue-10)

Research Article | **October 2014**

Reactive Label Concealment Shield on Social Network Data

P Archana
Computer Science & Engineering
Prasad Engineering College, India

Dr. K. S. Babu Rao
Computer Science & Engineering
Prasad Engineering College, India

Venkateshwarlu M
Computer Science & Engineering
Prasad Engineering College, India

Abstract—

The information published in the social networks need to be elegant and more individualized. By recognizing this in social networks motivated us, to propose a scheme called privacy protection scheme which prevents the revelation of identities of both users and some selected features in their profiles. Privacy is one of the major concerns when publishing or sharing social network data for social science research and business analysis. Recently, researchers have developed privacy models similar to k-anonymity to prevent node re-identification through structure information. The label-node relationship is not well protected by pure structure methods. Furthermore, existing approaches, which rely on edge editing or node clustering, may significantly alter key graph properties. Each user can pick out the features of his own profile he wishes to hide. In this report, we simulate the users as nodes and the feature as labels in the social networks which are modeled as a graph .Labels in the graph are treated as sensitive or non-sensitive. We treat node labels both as background knowledge an adversary may possess, and as sensitive information that has to be protected. We present privacy protection algorithms that allow for graph data to be published in a form such that an adversary who possesses information about a node's neighbourhood cannot safely infer its identity and its sensitive labels. To this aim, the algorithms transform the original graph into a graph in which nodes are sufficiently indistinguishable. The original graph structure and its properties are also evaluated to find which extent the algorithms preserve privacy. We also demonstrated that the solution we proposed is effective, efficient and scalable than those in anterior research.

Keywords— Privacy, Online Social Network, Privacy protecting in SN, Sensitive information, Sensitive Information

I. INTRODUCTION

The data published in the social networks need to be protected since there is a threat to the sensitive information about users. In social network link Facebook, twitter & many other public network sites uses sensitive information about users like Name, Age, Mobile Number, Email ID, Sex and other information which should be protected. The privacy protection scheme which we proposed prevents the revelation of identities of both users and some selected features in their profiles. Each user can pick out the features of his own profile he wishes to hide. The previous models are only concerned with link revelation and identity of users. The challenge is to devise methods to publish social network data in a form that affords utility without compromising privacy. Previous research has proposed various privacy models with the corresponding protection mechanisms that prevent both inadvertent private information leakage and attacks by malicious adversaries.

We refer to these details and messages as features in the user's profile. We propose a privacy protection scheme that not only prevents the disclosure of identity of users but also the disclosure of selected features in users' profiles. An individual user can select which features of her profile she wishes to conceal. The social networks are modelled as graphs in which users are nodes and features are labels 1. Labels are denoted either as sensitive or as non-sensitive. Figure 1 is a labelled graph representing a small subset of such a social network. Each node in the graph represents a user, and the edge between two nodes represents the fact that the two persons are friends. Labels annotated to the nodes show the locations of users. Each letter represents a city name as a label for each node. Some individuals do not mind their residence being known by the others, but some do, for various reasons. In such case, the privacy of their labels should be protected at data release. Therefore the locations are either sensitive (labels are in red italic in Figure1) or non-sensitive.

The privacy issue arises from the disclosure of sensitive labels. One might suggest that such labels should be simply deleted. Still, such a solution would present an incomplete view of the network and may hide interesting statistical information that does not threaten privacy. A more sophisticated approach consists in releasing information about sensitive labels, while ensuring that the identities of users are protected from privacy threats. We consider such threats as neighbourhood attack, in which an adversary sends out sensitive information based on prior knowledge of the number of neighbours of a target node and the labels of these neighbours. In the example, if an adversary knows that a user has three friends and that these friends are in A (Anil), B (Bhavani) and C (Chandu), respectively, then she can infer that the user is in H (Honey).

We present privacy protection algorithms that allow for graph data to be published in a form such that an adversary cannot safely infer the identity and sensitive labels of users. We consider the case in which the adversary possesses both structural knowledge and label information.

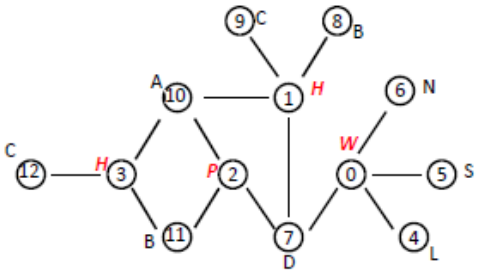
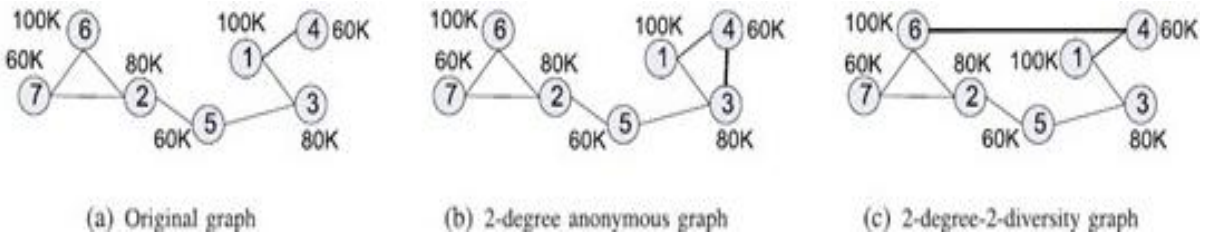


Fig. 1. Labeled graph showing a social network distribution of data

The algorithms that we propose transform the original graph into a graph in which any node with a sensitive label is indistinguishable from at least *l*-1 other nodes. The probability to infer that any node has a certain sensitive label (we call such nodes *sensitive nodes*) is no larger than 1/*l*. For this purpose we design *l*-diversity-like model, where we treat node labels as *both* part of an adversary's background knowledge and as sensitive information that has to be protected.

The algorithms are designed to provide privacy protection while losing as little information and while preserving as much utility as possible. In view of the trade-off between data privacy and utility [16], we evaluate empirically the extent to which the algorithms preserve the original graph's structure and properties such as density, degree distribution and clustering coefficient. We show that our solution is effective, efficient and scalable while offering stronger privacy guarantees than those in previous research, and that our algorithms scale well as data size grows.

II. FIGURES



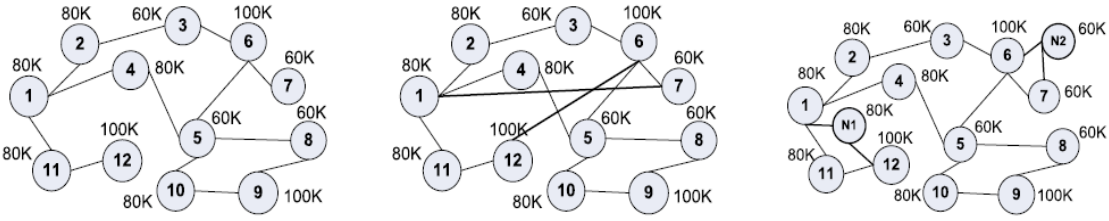


Fig. 2. Example for adding noise node.

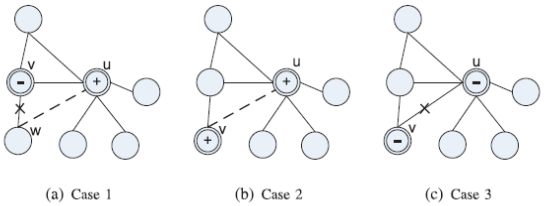


Fig. 3. Edge editing with neighborhood rule.

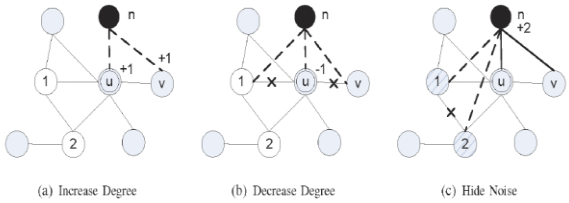


Fig. 4. Strategies to adding noise nodes.

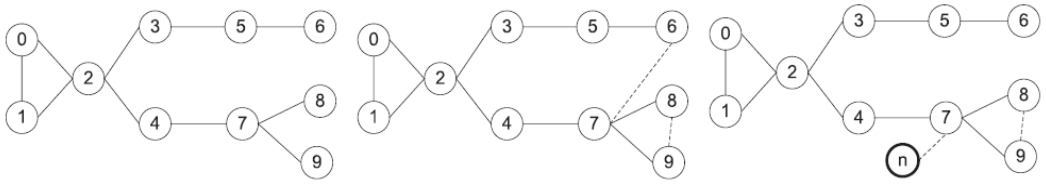


Fig. 5. Publish a graph with neighborhood graph anonymity.

III. CONCLUSIONS

The personal data published in the social networks is protected and investigated in this paper. We investigated the protection of private label information in social network data publication. Graphs are rich label information, which are categorized to be either sensitive or non-sensitive. We assume that adversaries possess prior knowledge about a node's degree and the labels of its neighbours, and can use that to infer the sensitive labels of targets. To infer the sensitive labels of targets the rivals use the prior knowledge about node's degree and labels of its neighbours. Both rivals background knowledge and sensitive information of node labels take part in attaining privacy while publishing the data through our model. We accompany our model with algorithms that transform a network graph before publication, so as to limit adversaries' confidence about sensitive label data. We guaranteed a clear privacy with experiments on real and synthetic data sets which bear out the scalability, effectiveness and efficiency. In our approach we also maintain critical graph properties to provide guaranteed privacy.

ACKNOWLEDGMENT

Our thanks go out to number of people without whose valuable guidance and contributions, the project would have remained in the conception stage for eternity.

We are grateful to the principal of our college, **Dr. K.S. Babu Rao**, who is the guiding lamp of the institution.

The project entrusted to us, is proposed to be implemented in the college practically. We understood the repercussions of even small errors in such projects. In this regard we profusely thank **M Venkateshwarlu**; Dept. of CSE for reposing faith in us by awarding such a crucial project to us.

Immense thanks to our internal guide, **Dr. K.S. Babu Rao**, who was a great source of encouragement for us. He had a fine balance of hands-on & hands-off approaches, which brought out the best in us and in our project.

Archana P 126L1D5801

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. Barabási 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.

