International Journal of Security and Its Applications
Vol. 10, No. 1 (2016), pp.95-102
http://dx.doi.org/10.14257/ijsia.2016.10.1.10

A Combined Model of Clustering and Classification Methods for Preserving Privacy in Social Networks against Inference and Neighborhood Attacks

Ali Zaghian[1] and Ayoub Bagheri[2,*]

[1]*Department of Mathematics and Cryptography, Malek Ashtar University of Technology, Isfahan, Iran, P. Box: 83145/115*
[2]*Department of Electrical and Computer Engineering, university of Kashan, Isfahan, Iran*
a_zaghian@mut-es.ac.ir, a.bagheri@kashanu.ac.ir

Abstract

In the last decade online social networks has gained remarkable attention. Facebook or Google+, are example social network services which allow people to create online profiles and share personal information with their friends. These networks publish details about users while some of the information revealed inside is private. In order to address privacy concerns, many social networks allow users to hide their private or sensitive information in their profiles from the public. In this paper, we focus on the problem of information revelation in online social networks by preserving the privacy of sensitive information in their data using machine learning and data mining algorithms. We show how an adversary can launch an inference or neighborhood attack to exploit an online social network using released data and structure of the network to predict the private information and attributes of users. For this purpose, we propose a new data mining based model that uses neighborhood information and attributes details of a user to infer private attributes of user profiles. The proposed model consists of two main parts: a clustering approach to ensure the k-anonymity and a classification algorithm to preserve the privacy against inference attacks. Finally we explore the effectiveness of some sanitization techniques that can be used to combat such inference attacks, and we show experimentally the success of different neighborhood re-identification strategies. Our experimental results reveal that using combination of data mining algorithm can notably help to preserve private and sensitive information in social network data.

Keywords: *Social network analysis, privacy preserving, inference, neighborhood, friendship attack, data mining*

1. Introduction

In recent years with the increasing popularity of online social networks, more services such as Facebook, Google+, Instagram and Twitter have emerged. These social networks are online applications that allow users to publish their personal attributes and relationships with friends [1]. For instance, Twitter is an online social networking service that enables users to send and read short 140-character messages called "tweets". Registered users can read and post tweets, but unregistered users can only read them. Likewise, Facebook is an online social networking service which users can create a user profile, add other users as friends, exchange messages, post status updates and photos, share videos and receive notifications when others update their profiles. Additionally, in Facebook users may join common-interest user groups, organized by workplace, school or

Ayoub Bagheri is the corresponding author.

ISSN: 1738-9976 IJSIA

college, or other characteristics, and categorize their friends into lists such as "People From Work" or "Close Friends". Facebook had over 1.44 billion monthly active users as of March 2015 [1-3].

As social networks grow rapidly, by their high popularity, interesting opportunities can be mined from the data provided. For example, social network data could be used for advertising or marketing for the customers. Simultaneously, in these opportunities, people's information privacy has become one of the most urgent issues. In recent years many research studies have been developed in the area of privacy preserving of social networks [4-22]. In these researches, one of the most privacy concerns of individuals in a social network is private information leakage. Private information leakage is about personal details of a user or an individual that is not explicitly stated (*e.g.*, political or religious interest, gender, occupation), while it is inferred through other details released and/or relationships (friendships or neighborhood) to individuals who may express that attribute [6, 17]. In the context of social networks, revealing private information can be done by inference or neighborhood (friendship) attacks. Generally, an inference attack uses details of attributes of user profiles, while neighborhood attack utilizes structure of graph data to infer the private information. In neighborhood attacks, ensuring k-anonymity meaning if every user is indistinguishable from at least k-1 other people, is the main approach of preserving sensitive data [7, 14-17].

In this paper, we focus on the problem of user private information leakage in an online social network service by ensuring k-anonymity and also by predicting private attributes that a user is not willing to publish and propose data sanitization approaches on preventing such private information leakage. In this work, we propose a privacy preserving model based on data mining for detecting private attributes of users. More precisely, by using a machine learning algorithm we introduce a combination model of clustering and classification to infer private attributes using real-life social network data including friendship information. Finally to protect privacy, we sanitize link and attribute details by proposing some information deleting against possible inference attacks.

The paper is organized as follows. In Section 2, we briefly introduce the background and related work. In Section 3, we explain the proposed model for social network anonymization. In Section 4, we describe an empirical evaluation and discuss the major experimental results. Finally we conclude with a summary and some future research directions in Section 5.

2. Related Work

In the research area of social network attacks many studies has been conducted [4-22]. For the neighborhood attacks, Zhou and Pei [7] proposed a solution to combat the adversary's 1-neighborhood attacks. They proposed an anonymization technique for social networks to prevent the neighborhood attacks by using depth-first search. Hay *et al.* [8] presented a framework to model the adversaries' background knowledge as vertex requirement structural queries, and proposed a privacy requirement k-candidate anonymity. Cheng *et al.* [9] proposed a k-anonymity model, which disconnects the original graph into k-isomorphic subgraph. Zou *et al.* [10] proposed the k-automorphism model, which converts the original network into a k-anonymized network, while it does not prevent the neighborhood attack. Thomson and Yao [11] have presented two clustering algorithms for clustering undirected graphs that group similar graph nodes into clusters. They have developed an inter-cluster matching method for anonymizing social networks by adding and removing edges based on the social role of the nodes.

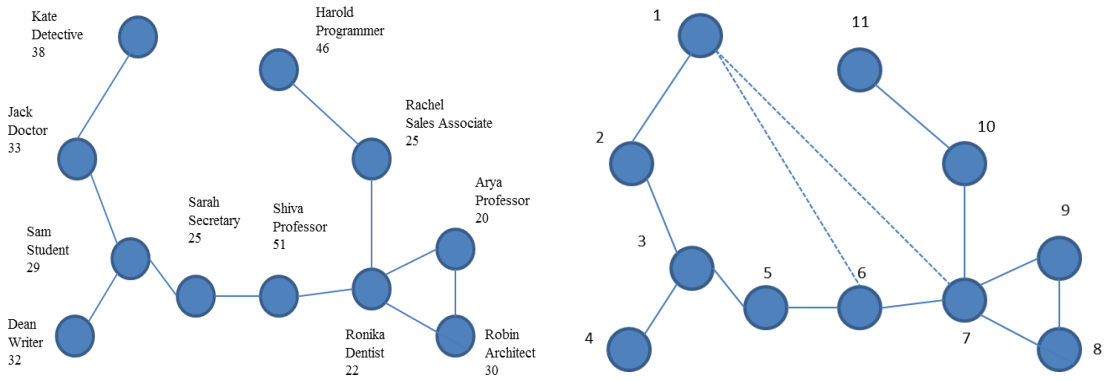
For the inference attacks some attackers are try to simply identify people, while in others rather than identifying individuals, try to infer private information inside social networks [14-22]. The first category considers an attack against an

anonymized network, which consists of only nodes and edges to identify individuals, while the second category considers attribute details and focuses on inferring details from nodes in the network, not only identifying individuals. He *et al.* [15] consider ways to infer private information via friendship links by creating a Bayesian Network from the links inside a social network. They use a real social network and user attributes to analyze their learning algorithm.

3. Proposed Model for Social Network Anonymization

3.1. Model structure

We model social networks as graphs that are undirected and is defined as a tuple G = (V, E) where V is a set of nodes, and E a set of edges. In this configuration of social network, each node represents a person in the graph, and each link represents a friendship. Each node n_i in the graph, has a set of attributes or details, $\{a_{1,i}, \ldots, a_{N,i}\}$. Figure 1 shows a raw social network along with its anonymization example.



(a)The Social Network (b) 2-Anonymous Social Network

Figure 1. Social Network Example and its Anonymization

A social network graph is k-anonymous if every node in it is indistinguishable from at least k-1 other nodes [5]. For example, the social network in Figure 1(b) satisfies 2-anonymity.

To prevent from inference and neighborhood attacks, in this paper, we propose a data mining model to first by utilizing a clustering approach provides k-anonymity of user individuals in the network against neighborhood attacks and second by using a classification algorithm infers private attributes of user profiles to sanitize the network against inference attacks. Figure 2 shows the structure of the proposed model. In the following, we discuss each of the steps in the proposed model.

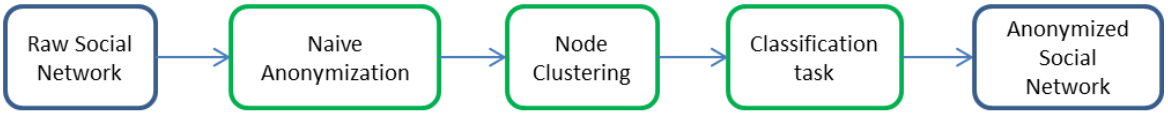


Figure 2. Structure of the Proposed Model for Social Network Anonymization

3.2 Naive Anonymization

In the first step of the proposed model, all the node identifiers of the raw social network data are removed and replaced by temporary identities *e.g.* numbers. By using this step in the model, an adversary with no prior knowledge on the network cannot re-identify any targeted user or its friend's identities.

3.3 Node Clustering

In the step of node clustering, a clustering approach is introduced for grouping similar nodes by a metric. In the proposed model, first Louvain Method [22] is executed to partition the complete graph into sub-graphs. Then a node aggregation method based on k-means clustering has been used which groups the nodes by using cosine similarity function into super-nodes each of which contains at least k and at most 2k-1 of the initial nodes. In other words, the clustering approach ensures that the k-anonymity of social network graph is achieved by obtaining that every node is incorporated into a cluster within which there are at least k-1 other nodes. Cosine similarity metric has been chosen which calculates a distance based on the degree of the reference node and number of edges in the sub-graph.

3.4 Classification Task

Classification task step is used to prevent from inference attacks on finding private attributes of user. Some examples of a sensitive private attribute can be gender, political interest or occupation. For this task we used Naïve Bayes algorithm to detect the private attribute as the class label from the training dataset. Naive Bayes (NB) algorithm is a kind of important classification algorithm because it has a high speed and is easy to implement [23-25]. We used the MAP (Maximum a Posteriori) Naive Bayes algorithm in our experiments as a classifier the anonymization model. In a classification problem, training and test dataset have to be labelled by a human expert and the classifier predicts the class of each data in test dataset. Naive Bayes algorithm assigns a new review document with a class with the maximum probability. This maximum value can be calculated by Equation (1):

$$Naive\ Bayes\ Classifier: c_{NB} = argmax_{c_j \in C} P(c_j) \prod_i P(a_i|c_j) \quad (1)$$

Where c_{NB} is the assigned class or output of Naive Bayes algorithm, c_j shows the class jth, $P(c_j)$ is prior probability of class j in the set of all classes C and $P(a_i|c_j)$ shows conditional probability of attribute i in class j. Output of the Naive Bayes algorithm is the maximum probability between classes.

3.5 Sanitization Techniques

After predicting the private attributes or link information, sanitization techniques can be used to anonymize the network. In other words, the goal of sanitization is to add, modify or remove attributes that may prevent algorithms from being able to infer a user's private information. Because of the structure of social network data adding extra false attributes or modifying existing data may be incorrect and unacceptable by users. Therefore, while the removing of an attribute does not introduce any misleading information in the graph data, we focus on sanitizing a social network by removing attributes rather than by adding false information. For this purpose, we use Naïve Bayes probabilities to choose which attributes to remove. Equation 2, shows that the attribute x is the most highly indicative of a class, hence the Naïve Bayes algorithm suggest attribute x to remove.

$$x = argmax_x[\forall c_j \in C: P(a_i|c_j)] \quad (2)$$

In this equation, x is the candidate attribute, C is the collection of classes and ai shows an attribute.

4. Experimental Evaluation and Results

In this section we evaluate the performance of the proposed model with respect to the the anonymity of social network data. The experiments were conducted on a computer with Core i3 2.13GHz CPU with 4GB memory running on Windows 7. In the following, we first introduce the datasets we used in the experiments and then evaluation metrics and the important experimental results will be discussed.

4.1 Datasets

We conducted the experiments on three famous datasets, namely Gnutella05, Gnutella08 [26, 27], and Adult. Gnutella05 and Gnutella08 are snapshots of the Gnutella peer-to-peer file sharing network in August 2002. In both datasets, the graph is directed, and vertices represent host computers and the edges represent the connections between the Gnutella hosts. Gnutella05 has 8,846 vertices and 31,839 edges. Gnutella08 has 6,301 vertices and 20,777 edges. As in the research study by Fung et. al. [20], while the two datasets have no labels, we used the Adult dataset to synthesize the vertex labels. Adult has 45,222 records on 8 categorical attributes. As the numbers of records in Adult are different from the number of vertices in Gnutella05 and Gnutella08, we associated each record in adult with Gnutella05 and Gnutella08 based on the order given in the raw datasets.

The Gnutella05 and Gnutella08 datasets were taken directly from the Stanford Large Network Dataset Collection (SNAP) website (available at http://snap.stanford.edu/data/). In the case of the Adult dataset, the data was taken from UCI machine learning repository available at http://archive.ics.uci.edu/ml/datasets/Adult.

4.2 Evaluation Metrics

Two metrics are used in order to evaluate information loss and performance of the proposed approaches in the model. These metrics are basic graph statistics namely degree and clustering coefficient [12, 20]. For the degree and clustering coefficient the distribution of each variable in the original data file is correlated with that of the same variable in the anonymized file, and the deviation from 1 is the information loss.

If G is the original graph, G' the anonymized graph, and m1 and m1' their degree values respectively, then the information loss will be:

$$\text{Information Loss}(G, G', m1) = 1 - \text{correlation}(m1, m1') \quad (3)$$

where correlation is a mapping correlation function. The information loss for clustering coefficient metric (m2) would follow in a similar manner.

4.3. Experimental Results

To obtain the performance of the proposed model, we evaluated clustering performance by calculating information loss with degree and clustering coefficient measures.

In the clustering algorithm for ensuring k-anonymity the parameter k is given to produces a graph consisting of super-nodes which contain a minimum of k and a maximum of 2k-1 basic nodes. In our experiments we calculated the degree and

clustering coefficient metrics on the two of our social network dataset for k=0, k=2, k=4, k=8 and k=16. In these experiments, if a super-node reaches to the maximum size of nodes, it will be divided into two super-nodes, each containing k nodes. Nodes are grouped based on similarity using the cosine similarity metric so that an adversary will be unable to distinguish between the nodes in a group. As mentioned in the evaluation metric section, the difference between the correlation of each experiment of k and k=0 is interpreted as the information loss. Figure 3 shows the information loss by increasing k in each of datasets.

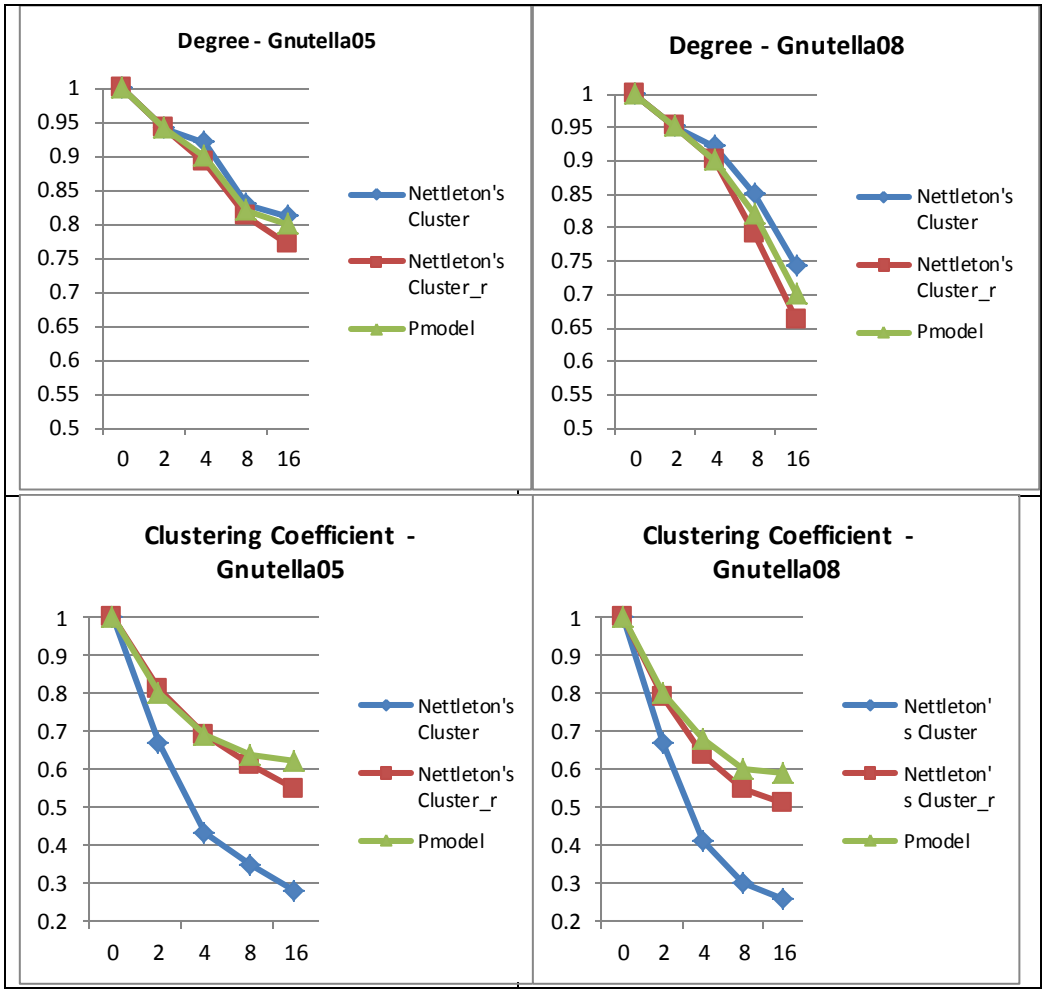


Figure 3. Information Loss for Anonymization in Different Datasets

In Figure 3, we compared the results of the proposed model (PModel) for clustering to the methods proposed by Nettleton et.al. [12]. Nettleton et.al proposed two methods of clustering social network data by using neighborhood sub-graph matching method as a similarity calculation method. The first one, named Cluster, has no constraints so it can choose nodes to match anywhere in the graph. And the second one is constrained by the community which is called Cluster_r. As can be seen in Figure 3 in the degree metric results, the Nettleton's cluster method outperforms PModel and Nettleton's cluster methods. This is because the Nettleton's cluster method globally checks for node matching. In the clustering coefficient metric our method outperforms other approaches, because the distance function is based on cosine similarity.

5. Conclusions

Social networks are online applications that allow users to publish their personal attributes and relationships with friends. As social networks grow rapidly, by their high popularity, interesting opportunities can be mined from the data provided. Simultaneously, in these opportunities, people's information privacy has become one of the most urgent issues. In this paper we have studied a model anonymization of social network analysis. The proposed model tries to preserve privacy against some kind of inference and neighborhood attacks. We examined the model on real-life dataset of social networks and demonstrate that our model can preserve much of the data and characteristics of social networks. In the future work, we will continue to propose more advanced neighborhood and inference attacks with more complex attack scenarios.

References

[1] L. Adamic, E. Adar, How to search a social network, Social Networks, vol. 27, no. 3, **(2005)**, pp. 187–203.
[2] G. Kossinets, D. J. Watts, Empirical analysis of an evolving social network, Science, vol. 311, no. 5757, **(2006)**, pp. 88–90.
[3] R. Kumar, J. Novak, A. Tomkins, Structure and evolution of online social networks, In Link mining: models, algorithms, and applications, Springer New York, **(2010)**, pp. 337-357.
[4] B. K. Tripathy, G. K. Panda, A new approach to manage security against neigborhood attacks in social networks, International Conference on Advances in Social Networks Analysis and Mining, **(2010)**, pp. 264-269.
[5] L. Sweeney, K-anonymity: a model for protecting privacy, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, **(2002)**, pp. 557–570.
[6] A. Machanavajjhala, D. Kifer, J. Gehrke, M. Venkitasubramaniam, l-diversity: Privacy beyond k-anonymity, ACM Transactions on Knowledge Discovery from Data (TKDD) 1, no. 1 **(2007)** 3.
[7] B. Zhou, J. Pei, Preserving privacy in social networks against neighborhood attacks, In IEEE 24th International Conference on Data Engineering, **(2008)** 506-515.
[8] M. Hay, G. Miklau, D. Jensen, D. Towsley, P. Weis, Resisting structural re-identification in anonymized social networks, The VLDB Journal, vol. 19, no. 6, **(2008)**, pp. 797–823.
[9] J. Cheng, A. W. Fu, J. Liu, K-isomorphism: privacy preserving network publication against structural attacks, In Proceedings of SIGMOD, Indianapolis, **(2010)**, pp. 459–470.
[10] L. Zou, L. Chen, M. T. O¨ zsu, K-automorphism: a general framework for privacy preserving network publication, Journal Proceedings of the VLDB Endowment, vol. 2, no. 1, **(2009)**, pp. 946–957.
[11] B. Thompson, D. Yao, The union-split algorithm and cluster-based anonymization of social networks, In Proceedings of the 4th International Symposium on Information, Computer, and Communications Security, **(2009)** , pp.218-227.
[12] D. F. Nettleton, V. Torra, A. Dries, A comparison of clustering and modification based graph anonymization methods with constraints, International Journal of Computer Applications, **(2014)** 95(20).
[13] C. H. Tai, P. S. Yu, D. N. Yang, M. S. Chen, Privacy-preserving social network publication against friendship attacks, In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, **(2011)** pp. 1262-1270.
[14] C. Sun, P. S. Yu, X. Kong, Y. Fu, Privacy preserving social network publication against mutual friend attacks, In 13th IEEE International Conference on Data Mining, **(2013)**, pp. 883-890.
[15] J. He, W. Chu, V. Liu. Inferring privacy information from social networks, In Intelligence and Security Informatics, **(2006)**, pp. 154-165.
[16] R. Gross, A. Acquisti, J. H. Heinz, Information revelation and privacy in online social networks, In WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society, **(2005)**, pp. 71–80.
[17] J. Lindamood, R. Heatherly, M. Kantarcioglu, B. Thuraisingham, Inferring private information using social network data, In Proceedings of the 18th ACM international conference on World wide web, **(2009)**, pp. 1145-1146.
[18] E. Zheleva, L. Getoor, Preserving the privacy of sensitive relationships in graph data, Privacy, security, and trust in KDD. Springer Berlin Heidelberg, **(2008)**, pp. 153-171.
[19] E. Zheleva, L. Getoor, To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles, In proceedings of the 18th international ACM conference on World wide web, **(2009)**

[20] B. Fung, Y. A. Jin, J. Li, Preserving privacy and frequent sharing patterns for social network data publishing, In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining**, (2013)** , pp. 479-485.
[21] N. Mohammed, B. C. M. Fung, and M. Debbabi, Anonymity meets game theory: secure data integration with malicious participants, Very Large Data Bases Journal (VLDBJ), vol. 20, no. 4, **(2011)**, pp. 567–588.
[22] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and Etienne Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008, no. 10 **(2008)** 1000.
[23] A. Bagheri, M. Saraee, F. de Jong, Sentiment classification in persian: Introducing a mutual information-based method for feature selection, In Electrical Engineering (ICEE), 2013 21st Iranian Conference on, **(2013)**, pp. 1-6.
[24] T. Mitchell, T., Machine Learning, second edition, McGraw Hill, **(1997)**.
[25] J. Han, M. Kamber, J, Pei, Data Mining: Concepts and techniques, Morgan kaufmann, **(2006)**.
[26] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters. ACM Transactions on Knowledge Discovery from Data (TKDD), vol.1, no. 1, **(2007)** 2.
[27] M. Ripeanu and I. Foster. Mapping the Gnutella Network - Macroscopic Properties of Large-scale P2P Networks and Implications for System Design. In Internet Computing Journal 6(1), **(2002)** 1.

