On Link Privacy in Randomizing Social Networks

Xiaowei Ying and Xintao Wu*

University of North Carolina at Charlotte
{xying, xwu}@uncc.edu

Abstract. Many applications of social networks require relationship anonymity due to the sensitive, stigmatizing, or confidential nature of relationship. Recent work showed that the simple technique of anonymizing graphs by replacing the identifying information of the nodes with random ids does not guarantee privacy since the identification of the nodes can be seriously jeopardized by applying subgraph queries. In this paper, we investigate how well an edge based graph randomization approach can protect sensitive links. We show via theoretical studies and empirical evaluations that various similarity measures can be exploited by attackers to significantly improve their confidence and accuracy of predicted sensitive links between nodes with high similarity values.

1 Introduction

Social networks are of significant importance in various application domains such as marketing, psychology, epidemiology and homeland security. Many applications of social networks such as anonymous Web browsing require relationship anonymity due to the sensitive, stigmatizing, or confidential nature of relationship. For example, most people prefer to conceal the truth regarding their illegal or unethical behaviors which are customarily disapproved of by society.

One natural approach is to publishing a node-anonymized version of the network that permits useful analysis without disclosing the identity of the individuals represented by the nodes. The recent work [1, 4] pointed out that this simple technique of anonymizing graphs by replacing the identifying information of the nodes with random ids does not guarantee privacy since the identification of the vertices can be seriously jeopardized by applying subgraph queries. Another approach is to randomizing edges to protect sensitive links [3, 4, 6, 8, 11]. For example, we can remove some true edges and/or add some false edges. After the randomization, the randomized graph is expected to be different from the original one. As a result, the true sensitive or confidential relationship will not be much disclosed even if the identification of the vertices is achieved by attackers.

We will explore how well the edge randomization can protect those sensitive links. In [8], Ying and Wu preliminarily investigated the relationship between the amount of randomization and the attacker's ability to infer the presence of a link and presented a randomization strategy that can preserve the spectral properties (and utility) of the graph. However, the effect on privacy due to randomization was quantified by considering only the magnitude information of randomization. It has been well known that

* This work was supported in part by U.S. National Science Foundation IIS-0546027 and CNS-0831204.

T. Theeramunkong et al. (Eds.): PAKDD 2009, LNAI 5476, pp. 28–39, 2009.

graph topological features have close relations with the existence of links and various proximity measures have been exploited to predict the existence of a future link [5]. In this paper, we will investigate formally how attackers may exploit proximity measure values (derived from the released randomized graph) to breach link privacy. Privacy of a sensitive link is jeopardized if attackers' confidence of prediction is higher than some tolerated threshold or is significantly greater than the a-priori belief (without the exploit of the released randomized data). Hence it is of great importance for data owners to be aware of potential attacks and quantify the magnitude of perturbation to better protect sensitive links.

2 Related Work

Social network analysis has increasing interest in the database, data mining, and theory communities. The current state of the art is that there has been little work dedicated to privacy preserving social network analysis with the exception of some very recent work [1–4, 6, 8–11].

In [1], Backstrom and et al. described a family of attacks such that an adversary can learn whether edges exist or not between specific targeted pairs of nodes from node-anonymized social networks. Similarly in [4], Hay and et al. further observed that the structure of the graph itself (e.g., the degree of the nodes or the degree of the node's neighbors) determines the extent to which an individual in the network can be distinguished.

In [6], Liu and Terzi investigated how to modify a graph via a set of edge addition (or deletion) operations in order to construct a new k-degree anonymous graph, in which every node has the same degree with at least $k-1$ other nodes. In [11], Zhou and Pei anonymized the graph by generalizing node labels and inserting edges until each neighborhood is indistinguishable to at least $k-1$ others. In [2, 10], authors applied a structural anonymization approach called *edge generalization* that consists of collapsing clusters together with their component nodes' structure, rather than add or delete edges from the social network dataset. Although the above proposed approaches would preserve privacy, however, it is not clear how useful the anonymized graph is since many topological features may be lost.

The problems of how to generate a synthetic graph preserving various topological features of a real social network and how attackers may exploit the topological features of the released graph to breach link privacy were recently studied in [9]. However, the attacking model in [9] was based on the probability of existence of a link across all possible graphs in the graph space. In this paper, the attacking model is to exploit the relationship between existence of a link and the similarity measure values of node pairs in one released randomized graph.

We would point out that our problem of attacking methods on a randomized graph is different from the classic link prediction problem investigated in [5]. The classic link prediction focuses on network evolution models and is to predict the existence of a future link between two nodes given a snapshot of a current social network. The change due to randomization is different with that due to network evolutions. Nevertheless, various graph proximity measures used in the classic link prediction could be used by attackers.

3 Link Privacy Analysis

A network $G(n, m)$ is a set of n nodes connected by a set of m links. The network considered here is binary, symmetric, connected, and without self-loops. Let $A = (a_{ij})_{n\times n}$ be its adjacency matrix, $a_{ij} = 1$ if node i and j are connected and $a_{ij} = 0$ otherwise. $\widetilde{G}$ is the randomized graph obtained by randomly adding k false edges followed by deleting k true edges. This strategy keeps the total number of edges in the original graph unchanged. We denote $\widetilde{A} = (\tilde{a}_{ij})_{n\times n}$ be the adjacency matrix of $\widetilde{G}$.

When it comes to link privacy, it is usually $a_{ij} = 1$ that people want to hide, not $a_{ij} = 0$ and attackers are capable of calculating posterior probabilities. Formally, we use $P(a_{ij} = 1)$ to denote the users' prior belief about the event of $a_{ij} = 1$ and use $P(a_{ij} = 1|\widetilde{G})$ to denote its posterior belief about $a_{ij} = 1$. The released graph $\widetilde{G}$ is regarded as jeopardizing the privacy if $P(a_{ij} = 1|\widetilde{G}) > P(a_{ij} = 1)$.

In [8], we preliminarily investigated the relationship between the amount of randomization and the attacker's ability to infer the presence of a link. The results are shown as follows. When the attacker knows only parameter m and n, the prior belief is

$$P(a_{ij} = 1) = \frac{2m}{n(n-1)}. \tag{1}$$

With the released graph and perturbation parameter k, the posterior belief is

$$P(a_{ij} = 1|\tilde{a}_{ij} = 1) = \frac{m-k}{m}, \quad P(a_{ij} = 1|\tilde{a}_{ij} = 0) = \frac{k}{\binom{n}{2} - m} \tag{2}$$

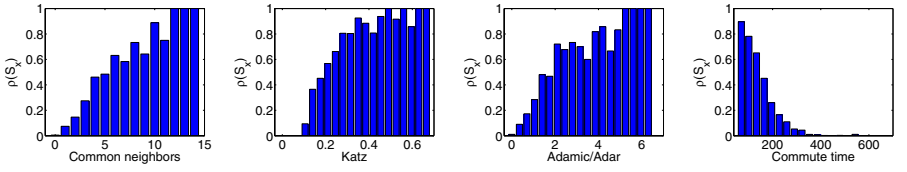
Equation 2 is based on the Addition/Deletion without replacement[1].

In this paper, we further investigate whether topological features of the released network can be exploited by attackers to breach the link privacy. More specifically, we focus on to what extent a given sensitive relationship can be breached by attackers who exploit proximity measure values of node pairs. Proximity measures have been shown to be effective in the classic link prediction problem (i.e., predicting the future existence of links among nodes given a snapshot of a current graph). However, link prediction in our context is to predict the likelihood of existence of original links from the randomized graph. This is challenging since the proximity measure values calculated from the randomized graph can be varied from those of the original graph. In section 3.1, we empirically show the close relationship between various similarity measures of node pairs and probability of link existence between them. In section 3.2, we conduct theoretical studies and quantify how much the posterior belief can be enhanced by exploiting those similarity measures.

3.1 Existence of a Link vs. Similarity Measure

Let m_{ij} be a similarity measure on node pair (i, j) in graph G (a larger value of m_{ij} indicates that nodes i and j are more similar). We apply four similarity measures in

[1] Refer to [8] for the Addition/Deletion with replacement. For large graphs, the difference between the above is small.



(a) Common neighbors (b) Katz (c) Adamic/Adar (d) Commute time

Fig. 1. Similarity measure vs. the prob. of true edges in the original graph ($\rho(S_x)$) for polbooks

this paper. The first one is the number of common neighbors: $CN_{ij} = \sum_{k=1}^{n} a_{ik}a_{kj}$. The second one is the Adamic/Adar measure , which is the weighted number of common neighbors. The weights are assigned based on the information theory: $Ad_{ij} = \sum_{k=1}^{n} \frac{1}{\log d_k} a_{ik}a_{kj}$, where d_k is the degree of node k. The third one is the Katz measure, which is a weighted sum of the number of paths in the graph that connect two nodes, with shorter paths being given the larger weight: $K_{ij} = \sum_{k=1}^{\infty} \beta^k P_{ij}^{(k)}$, where $P_{ij}^{(k)}$ denotes the number of paths from i to j with length equal to k while β is a damping factor. In this paper, we take $\beta = 0.1$. The fourth one is the commute time CT_{ij}, which is the expected steps of random walks from i to j and back to i. The commute time is a distance measure: more similar nodes have smaller CT values.

Let $\rho(\Omega)$ denote the proportion of true edges in the set of node pairs Ω:

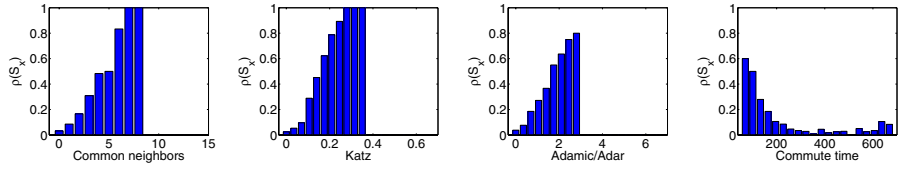
$$\rho(\Omega) = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} a_{ij},$$

where $|\Omega|$ denotes the number of elements in set Ω. Let $S_x = \{(i,j) : m_{ij} = x\}$ denote the set of all node pairs with the similarity measure $m_{ij} = x$. Hence $\rho(S_x)$ denotes the proportion of true edges in the S_x, which can be considered as the probability of existence of a link between node pair (i,j) in S_x. Next, we empirically show how $\rho(S_x)$ varies with x in real social networks.

Figure 1 shows how the proportions of true edges in S_x are varied with similarity measure values x in terms of four measures (Common neighbors, Katz, Adamic/Adar, and Commute time) in the US political books network (polbooks). The polbooks network[2] contains 105 nodes and 441 edges, and nodes represent books about US politics sold by the online bookseller Amazon.com while edges represent frequent co-purchasing of books by the same buyers on Amazon. We can observe that $\rho(S_x)$ increases with x. In other words, the probability that $a_{ij} = 1$ is highly correlated with similarity measure m_{ij}: the larger m_{ij} is, the more likely a_{ij} is equal to 1.

We then perturbed the polbooks network by adding 200 false edges and deleting 200 true edges. From the perturbed graph $\widetilde{G}$, we define $\widetilde{S}_x = \{(i,j) : \tilde{m}_{ij} = x\}$ as the set of node pairs with similarity measure $\tilde{m}_{ij} = x$. Figure 2 shows how the proportions of true edges in $\widetilde{S}_x$ (i.e., the probability of existence of a link) are varied with similarity measure values x in terms of four measures in the randomized polbooks network. We

[2] (http://www-personal.umich.edu/~mejn/netdata/)



(a) Common neighbors (b) Katz (c) Adamic/Adar (d) Commute time

Fig. 2. Similarity measure vs. the prob. of true edges in the randomized graph ($\rho(\widetilde{S}_x)$) for polbooks

can observe that the same pattern still holds even if the randomized graph itself is quite different from the original one (200 false edges out of 441 edges). In the next section, we will show how attackers exploit $\tilde{m}_{ij}$ in the perturbed graph $\widetilde{G}$ to improve their posterior belief on existence of a true link between nodes (i, j) in the original graph.

3.2 Link Prediction by Exploiting Similarity Measure

In this section, we quantify how much the posterior belief can be enhanced by exploiting similarity measure between two node (i, j) in the randomized graph. We present our quantification in a series of results and leave detailed proofs in Appendix.

Recall the randomization strategy is to randomly add k false edges followed by deleting k true edges. In other words, every true link is to be deleted independently with probability p_1 and every non-existing link is to be added independently with probability p_2. We can easily derive $p_1 = k/m$ and $p_2 = k/[\binom{n}{2} - m]$.

Let $\tilde{m}_{ij}$ denote the similarity measure of node i and j in $\widetilde{G}$. We define $\widetilde{S}_x = \{(i, j) : \tilde{m}_{ij} = x\}$ as the set of node pairs with $\tilde{m}_{ij} = x$ in the perturbed graph. Then we have $P(a_{ij} = 1|\tilde{m}_{ij} = x) = \rho(\widetilde{S}_x)$, and $P(a_{ij} = 0|\tilde{m}_{ij} = x) = 1 - \rho(\widetilde{S}_x)$. Recall that $\rho(\widetilde{S}_x)$ denotes the proportion of true edges in the set $\widetilde{S}_x$ derived from the perturbed graph. Also notice that $P(\tilde{a}_{ij} = 1|a_{ij} = 1) = 1 - p_1$ and $P(\tilde{a}_{ij} = 1|a_{ij} = 0) = p_2$. With the Bayes' theorem, the posterior belief is then given by

$$P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x) = \frac{(1 - p_1)\rho(\widetilde{S}_x)}{(1 - p_1)\rho(\widetilde{S}_x) + p_2[1 - \rho(\widetilde{S}_x)]}, \tag{3}$$

$$P(a_{ij} = 1|\tilde{a}_{ij} = 0, \tilde{m}_{ij} = x) = \frac{p_1\rho(\widetilde{S}_x)}{p_1\rho(\widetilde{S}_x) + (1 - p_2)[1 - \rho(\widetilde{S}_x)]}. \tag{4}$$

Equation 3 (Equation 4) shows the enhanced posterior belief that an observed (missing) edge (i, j) in the $\tilde{G}$ is a true edge in G. The following property shows that the event of an observed link $\tilde{a}_{ij} = 1$ usually has more indications to be a true link than that of $\tilde{a}_{ij} = 0$.

Property 1. *Let r denote the sparse ratio of the graph, $r = m/\binom{n}{2}$. If $k \leq (1 - r)m$, given a fixed x, we have the following inequality stands:*

$$P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x) \geq P(a_{ij} = 1|\tilde{a}_{ij} = 0, \tilde{m}_{ij} = x). \tag{5}$$

Many real-world social networks are very sparse ($r \approx 0$). Hence $k \leq (1-r)m$ is usually satisfied. We thus focus on the risk of the released links, $P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x)$.

One issue here is that attackers cannot know the proportion of true edges in $\widetilde{S}_x$ from the perturbed graph. What they can know actually is the proportion of observed edges in $\widetilde{S}_x$. Our next result shows the maximum likelihood estimate of $\rho(\widetilde{S}_x)$ can be derived from the proportion of observed edges in $\widetilde{S}_x$.

Result 1. *Given the perturbed graph and a fixed x, define $\widetilde{S}_x^1 = \widetilde{S}_x \cap \widetilde{E} = \{(i,j) : \tilde{a}_{ij} = 1, \tilde{m}_{ij} = x\}$. Assume $p_1 + p_2 \neq 1$, then the maximum likelihood estimator (MLE) of $\rho(\widetilde{S}_x)$ is given by*

$$\hat{\rho}\left(\widetilde{S}_x\right) = \frac{|\widetilde{S}_x^1|/|\widetilde{S}_x| - p_2}{1 - p_1 - p_2}, \tag{6}$$

and the MLE is unbiased.

By replacing $\rho(\widetilde{S}_x)$ in Equation 3 with $\hat{\rho}(\widetilde{S}_x)$ (shown in Equation 6), we have derived our enhanced posterior belief $P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x)$. Attackers may simply calculate the posterior belief of all node pairs in the perturbed graph and choose top-t node pairs as predicted candidate links.

For those similarity measures with continuous ranges (e.g., commute time), the number of node pairs with similarity measure equal exactly to x is usually small. In practice, we can apply histogram approximation or use the kernel estimator to smooth the estimation.

We would emphasize that our enhanced posterior belief $P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x)$ more accurately reflect the existence of a true link than the posterior belief $P(a_{ij} = 1|\tilde{a}_{ij} = 1)$ without exploiting the similarity measure derived in previous work [8]. We can see that $P(a_{ij} = 1|\tilde{a}_{ij} = 1)$ (shown in Equation 2) is the same for all observed links. On the contrary, our enhanced posterior belief $P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x)$ tends to be larger for those observed links with higher similarity values, and tends to be smaller for links with lower similarity values. Hence, it can more accurately reflect the existence of true links. We show our theoretical explanations in Results 2 and 3 and will compare the precisions of top-t predicted links derived from these two posterior beliefs in our empirical evaluations.

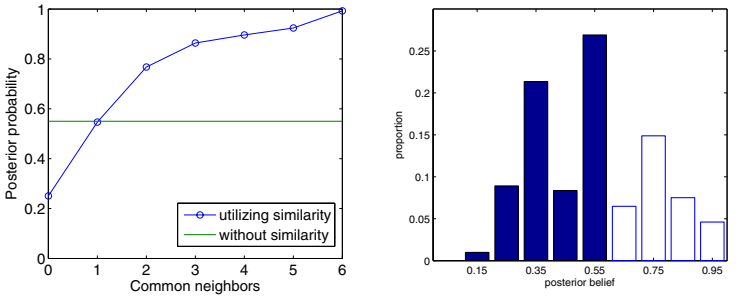
Result 2. *$P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x)$ is an increasing function of $\rho(\widetilde{S}_x)$, and when $\rho(\widetilde{S}_x) \geq \frac{p_2}{p_1+p_2}$, we have the following inequality stands:*

$$P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x) \geq P(a_{ij} = 1|\tilde{a}_{ij} = 1). \tag{7}$$

Our next result shows more clearly the relationship between a-priori belief (Equation 1), posterior belief without exploiting similarity measures (Equation 2), and our enhanced posterior belief with exploiting similarity measures (Equations 3 4).

Result 3. *Both the sum of a-priori belief over all node pairs and the sum of posterior belief (without exploiting similarity measures) overall all node pairs are equal to the number of edges:*

$$\textstyle\sum_{i<j} P(a_{ij} = 1) = \sum_{i<j} P(a_{ij} = 1|\tilde{a}_{ij}) = m.$$



(a) Posterior belief vs. Common neighbors

(b) Posterior belief distribution

Fig. 3. Posterior belief for polbooks network

The sum of our enhanced posterior belief (with exploiting similarity measures) also approaches to the number of edges:

$$\sum_{i<j} P(a_{ij} = 1|\tilde{a}_{ij}, \tilde{m}_{ij}) \to m \quad as \quad n \to \infty.$$

Figure 3 shows the relationship between the two posterior beliefs and the common neighbors for the polbooks data. We set $k = 200$. We can observe that the posterior belief without exploiting the similarity measure, $P(a_{ij} = 1|\tilde{a}_{ij} = 1)$, is 0.55 for all observed links. However, our enhanced posterior belief $P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij})$ are greater than 0.55 for those links with more than 2 common neighbors as shown in Figure 3(a). Figure 3(b) shows the distribution of the calculated posterior belief values. We can observe that 33.5% of released links have their posterior beliefs enhanced with similarity measures.

3.3 Privacy Protection Measure

In the privacy preserving data mining, one natural question from data owner is how many perturbations we need such that we can guarantee the protection for all sensitive individual edges are above some tolerated threshold. When attackers utilize the similarity measure, the absolute measure of protection for an individual link (i, j) can be defined as

$$\tau_a(i, j) = 1 - \max_x \left\{ \max_{t=0,1} P(a_{ij} = 1|\tilde{a}_{ij} = t, \tilde{m}_{ij} = x) \right\} \tag{8}$$

where the second term denotes the maximal suspicion of existing $a_{ij} = 1$. Compared with the protection under the attack without exploiting similarity measures, we define the relative measure of protection as

$$\tau_r(i, j) = \frac{\tau_a(i, j)}{1 - \max_{t=0,1} P(a_{ij} = 1|\tilde{a}_{ij} = t)}.$$

The measures of protection (τ_a and τ_r) are defined in terms of one individual edge. In the privacy preserving data mining, one natural question is how many perturbations we need such that we can guarantee the protection for all individual edges are above the threshold. Our next result shows the formula of the minimum number of perturbations to achieve the protection of all individual links. It is of great importance to evaluate the relationship between the required minimum number of perturbations and the utility loss of the perturbed graph. Due to space limitations, we leave this as our future work.

Result 4. *In the original graph, let* $S_x = \{(i,j) : m_{ij} = x\}$, $\rho_{\max} = \max_x \rho(S_x)$, *and sparse ratio* $r = m/\binom{n}{2}$. *When the protection threshold* $\epsilon < \frac{1-\rho_{\max}}{1-r}$, *there exists the minimum* k *such that* $\tau_r(i,j) \geq \epsilon$ *stands for all the node pair* (i,j) *is given by:*

$$k_{\min} = \frac{[(1-r)\epsilon\rho_{\max} - r(1-\rho_{\max})]m}{\epsilon(\rho_{\max} - r)}. \tag{9}$$

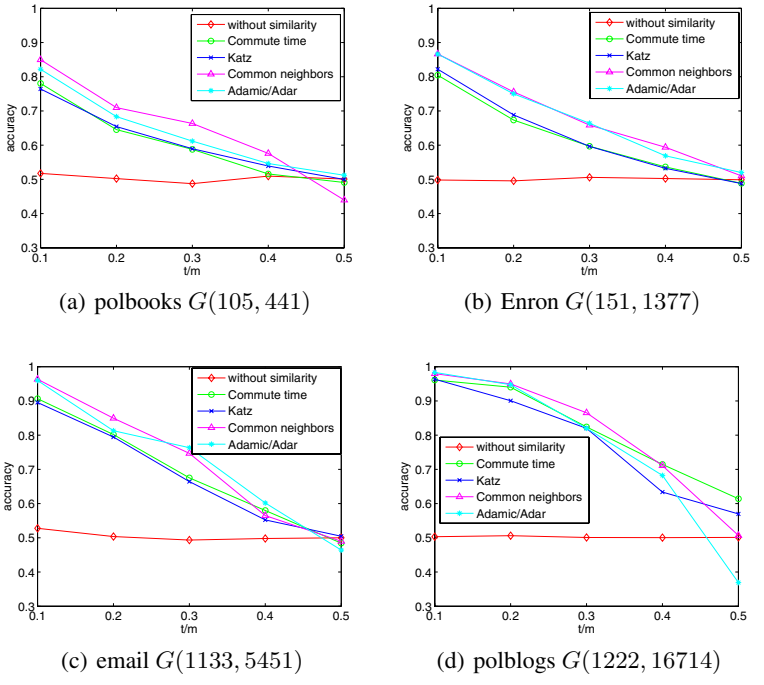
4 Empirical Evaluation

We used four network data sets (*polbooks, Enron, email, polblogs*) in our evaluation. The *Enron* network was built from email corpus of a real organization over the course covering a 3 years period. We used a pre-processed version of the dataset provided by [7]. This dataset contains 252,759 emails from 151 Enron employees, mainly senior managers. The *email* graph is the network of e-mail interchanges between members of the Univeristy Rovira i Virgili (Tarragona)[3]. The *polblogs* compiles the data on the links among US political blogs, containing over 1,000 vertices and 15,000 edges, which is based on incoming and outgoing links and posts around the time of the 2004 presidential election[4].

For each graph G, we randomly add k false edges and delete k true edges. We set $k = 0.5m$ in this paper, which corresponds to a relatively large perturbation. We also conducted evaluations with other k values and skip their results due to space limitations. We applied four similarity measures (Common neighbors, Katz, Adamic/Adar, Commute time) to predict top-t candidate links. We varied t values from $0.1m$ to $0.5m$ for all four data sets.

For each t, we calculated the precision of prediction links with different similarity measures. We also calculated the precision of prediction links using the posterior belief without exploiting the similarity measure. Figure 4 plots our results on four data sets. We can observe that for all four data sets we can achieve very high accuracy (greater than 0.8) by using our enhanced posterior belief for a subset (top $0.1m$) of released links, which indicates severe privacy disclosures for those sensitive links. We can also see that our enhanced posterior belief achieve higher precisions than the previous posterior belief without exploiting similarity measures for most links ($0.5m$) with high similarity measure values, indicating that the network topology does indeed contain latent information from which to infer interactions. From Figure 4, we can also observe that we achieve different precisions using different similarity measures: one measure which

[3] `http://deim.urv.cat/~aarenas/data/welcome.htm`
[4] `http://www-personal.umich.edu/~mejn/netdata/`



(a) polbooks $G(105, 441)$ (b) Enron $G(151, 1377)$

(c) email $G(1133, 5451)$ (d) polblogs $G(1222, 16714)$

Fig. 4. Precision of top t predictions by the posterior belief w/o similarity measures for four data sets

achieves the highest precision for one data set is not necessarily the one for another data set. It is of great significance to explore what similarity measures can be exploited by attackers to achieve the highest privacy disclosure for a given social network. We will investigate this in our future work.

5 Conclusion and Future Work

In this paper, we have investigated how well the edge randomization approach via addition/deletion can protect privacy of sensitive links. We have conducted theoretical analysis and empirical evaluations to show that node proximity measures can be exploited by attackers to enhance the posterior belief and prediction accuracy of the existence of sensitive links among nodes with high similarity values.

There are some other aspects of this work that merit further research. Among them, we will continue the line of this research by investigating other edge randomization approaches (e.g., edge switches). We will also investigate how the edge based randomization affects the graph's utility. We are interested in comparing theoretically and empirically the edge based randomization with the k-degree anonymization approaches [6, 11] in terms of the privacy vs. utility tradeoff.

References

1. Backstrom, L., Dwork, C., Kleinberg, J.: Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In: WWW 2007: Proceedings of the 16th international conference on World Wide Web, pp. 181–190. ACM Press, New York (2007)
2. Campan, A., Truta, T. M.: A clustering approach for data and structural anonymity in social networks. In: PinKDD (2008)
3. Hay, M., Miklau, G., Jensen, D., Towsely, D., Weis, P.: Resisting structural re-identification in anonymized social networks. In: VLDB (2008)
4. Hay, M., Miklau, G., Jensen, D., Weis, P., Srivastava, S.: Anonymizing social networks. University of Massachusetts Technical Report, 07-19 (2007)
5. Liben-Nowell, D., Kleinberg, J.: The link prediction problem for social networks. In: CIKM 2003: Proceedings of the twelfth international conference on Information and knowledge management, pp. 556–559. ACM, New York (2003)
6. Liu, K., Terzi, E.: Towards identity anonymization on graphs. In: Proceedings of the ACM SIGMOD Conference, Vancouver, Canada. ACM Press, New York (2008)
7. Shetty, J., Adibi, J.: The Enron email dataset database schema and brief statistical report. Information Sciences Institute Technical Report, University of Southern California (2004)
8. Ying, X., Wu, X.: Randomizing social networks: a spectrum preserving approach. In: Proc. of the 8th SIAM Conference on Data Mining (April 2008)
9. Ying, X., Wu, X.: Graph generation with prescribed feature constraints. In: Proc. of the 9th SIAM Conference on Data Mining (2009)
10. Zheleva, E., Getoor, L.: Preserving the privacy of sensitive relationships in graph data. In: Bonchi, F., Ferrari, E., Malin, B., Saygın, Y. (eds.) PInKDD 2007. LNCS, vol. 4890, pp. 153–171. Springer, Heidelberg (2008)
11. Zhou, B., Pei, J.: Preserving Privacy in Social Networks Against Neighborhood Attacks. Data Engineering, 2008. In: ICDE 2008. IEEE 24th International Conference on, pp. 506–515 (2008)

A Proofs

Proof of Property 1
It is easy to verify that inequality 5 stands if and only if $(1-p_1-p_2)[1-\rho(\widetilde{S}_x)] \geq 0$. We need only guarantee that $1-p_1-p_2 \geq 0$. Notice that $p_1 = k/m$ and $p_2 = k/[\binom{n}{2} - m]$, then $1 - p_1 - p_2 \geq 0$ if and only if $k \leq \left[1 - m/\binom{n}{2}\right] m = (1-r)m$. □

Proof of Result 1
Let $N = |\widetilde{S}_x|$, $N_1 = |\widetilde{S}_x^1|$ and $\rho = \rho(\widetilde{S}_x)$. Then, for a randomly selected node pair (i, j), $\tilde{a}_{ij}$ is a Bernoulli random variable:

$$P(\tilde{a}_{ij} = 1 | \tilde{m}_{ij} = x) = (1-p_1)\rho + p_2(1-\rho)$$
$$P(\tilde{a}_{ij} = 0 | \tilde{m}_{ij} = x) = p_1\rho + (1-p_2)(1-\rho)$$

Then the likelihood function of $\widetilde{S}_x$ is

$$L = [(1-p_1)\rho + p_2(1-\rho)]^{N_1} [p_1\rho + (1-p_2)(1-\rho)]^{N-N_1} .$$

Take derivative to $\ln L$ with respect of ρ, we have

$$\frac{d\ln L}{d\rho} = \frac{N_1(1-p_1-p_2)}{(1-p_1)\rho + p_2(1-\rho)} - \frac{(N-N_1)(1-p_1-p_2)}{p_1\rho + (1-p_2)(1-\rho)}.$$

Set $\frac{d\ln L}{d\rho} = 0$, we have $\hat{\rho} = \frac{N_1/N - p_2}{1-p_1-p_2}$, and the unbiasedness is then obvious. □

Proof of Result 2
Notice that $P(a_{ij} = 1|\tilde{a}_{ij} = 1) = \frac{m-k}{m} = 1 - p_1$, and with Equation 3, it is easy to verify this result. □

Proof of Result 3
$\sum_{i<j} P(a_{ij} = 1) = m$ is obvious. Notice that the number of edges does not change along the perturbation, then we have

$$\begin{aligned}\sum_{i<j} P(a_{ij} = 1|\tilde{a}_{ij}) &= \sum_{(i,j)\in\widetilde{E}} P(a_{ij} = 1|\tilde{a}_{ij} = 1) + \sum_{(i,j)\notin\widetilde{E}} P(a_{ij} = 1|\tilde{a}_{ij} = 0) \\ &= m \cdot \tfrac{m-k}{m} + \left[\tbinom{n}{2} - m\right] \cdot k/\left[\tbinom{n}{2} - m\right] = m. \end{aligned} \tag{10}$$

When attackers utilize the similarity measures with MLE, we first show

$$E[\textstyle\sum_{i<j} P(a_{ij} = 1|\tilde{a}_{ij}, \tilde{m}_{ij})] = m.$$

$$\begin{aligned} E\left[\sum_{i<j} P(a_{ij} = 1|\tilde{a}_{ij}, \tilde{m}_{ij})\right] = \sum_x \Bigg\{ &\sum_{(i,j)\in\widetilde{S}_x^1} E[P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x)] \\ &+ \sum_{(i,j)\in\widetilde{S}_x - \widetilde{S}_x^1} E[P(a_{ij} = 1|\tilde{a}_{ij} = 0, \tilde{m}_{ij} = x)] \Bigg\} \end{aligned} \tag{11}$$

With the MLE in Equation 6, we have

$$\begin{aligned} &\textstyle\sum_{(i,j)\in\widetilde{S}_x^1} E[P(a_{ij} = 1|\tilde{a}_{ij} = 1, \tilde{m}_{ij} = x)] \\ =&\frac{(1-p_1)\hat{\rho}_1(\widetilde{S}_x)}{(1-p_1)\hat{\rho}_1(\widetilde{S}_x) + p_2[1-\hat{\rho}_1(\widetilde{S}_x)]}|\widetilde{S}_x^1| \\ =&(1-p_1)|\widetilde{S}_x|E[\hat{\rho}_1(\widetilde{S}_x)] \quad \text{(substitute Equation 6)} \\ =&(1-p_1)|\widetilde{S}_x|\rho(\widetilde{S}_x) \\ =&(1-p_1)\textstyle\sum_{(i,j)\in\widetilde{S}_x} a_{ij} \quad \text{(by the definition of } \rho(\cdot)) \end{aligned} \tag{12}$$

Similarly, we have

$$\textstyle\sum_{(i,j)\in\widetilde{S}_x - \widetilde{S}_x^1} E[P(a_{ij} = 1|\tilde{a}_{ij} = 0, \tilde{m}_{ij} = x)] = p_1 \sum_{(i,j)\in\widetilde{S}_x} a_{ij} \tag{13}$$

Combining Equation 11, 12 and 13 together, we have

$$E\left[\sum_{i<j} P(a_{ij}=1|\tilde{a}_{ij},\tilde{m}_{ij})\right] = \sum_x \sum_{(i,j)\in\widetilde{S}_x} a_{ij} = \sum_{i,j} a_{ij} = m.$$

Then, due to the law of large number, we can conclude that

$$\sum_{i<j} P(a_{ij}=1|\tilde{a}_{ij},\tilde{m}_{ij}) \to m \quad \text{as} \quad n\to\infty,$$

and we prove the result. □

Proof of Result 4

When $k \le (1-r)m$, with Result 1 and 2, we have that

$$\max_x \{\max_{t=0,1} P(a_{ij}=1|\tilde{a}_{ij}=t,\tilde{m}_{ij}=x)\} = P(a_{ij}=1|\tilde{a}_{ij}=1,\tilde{m}_{ij}=x_0),$$

where x_0 is the value such that $\rho(\widetilde{S}_x)$ is maximized: $\rho(\widetilde{S}_{x_0}) = \max_x \rho(\widetilde{S}_x)$. Let $\tilde{\rho}_{\max} = \rho(\widetilde{S}_{x_0})$. Meanwhile,we can also conclude

$$\max_{t=0,1} P(a_{ij}=1|\tilde{a}_{ij}=t) = P(a_{ij}=1|\tilde{a}_{ij}=1).$$

Then we have

$$\tau_r(i,j) = \frac{p_2[1-\tilde{\rho}_{\max}]}{p_1[(1-p_1)\tilde{\rho}_{\max} + p_2(1-\tilde{\rho}_{\max})]}. \tag{14}$$

Substitute $p_1 = \frac{k}{m} = \frac{k}{rN}$ and $p_2 = \frac{k}{N-m} = \frac{k}{(1-r)N}$ into Equation 14, we can verify that $\tau_r(i,j)$ is an increasing function of k, and the maximum value is $\frac{1-\tilde{\rho}_{\max}}{1-r}$ when $k=(1-r)m$.

When $k \ge (1-r)m$, we similarly have the following:

$$\max_x \{\max_{t=0,1} P(a_{ij}=1|\tilde{a}_{ij}=t,\tilde{m}_{ij}=x)\} = P(a_{ij}=1|\tilde{a}_{ij}=0,\tilde{m}_{ij}=x_0),$$
$$\max_{t=0,1} P(a_{ij}=1|\tilde{a}_{ij}=t) = P(a_{ij}=1|\tilde{a}_{ij}=0).$$

In this case, $\tau_r(i,j)$ is a decreasing function of k, and the maximum is also $\frac{1-\tilde{\rho}_{\max}}{1-r}$ when $k=(1-r)m$.

Therefore, $k_{\min}$ exists if and only if $\epsilon \le \frac{1-\tilde{\rho}_{\max}}{1-r}$, and $k_{\min} < (1-r)m$. Then, $\tau_r(i,j)$ is given by Equation 14. Solving the inequality $\tau_r(i,j) \ge \epsilon$, we have that

$$k \ge \frac{[(1-r)\epsilon\tilde{\rho}_{\max} - r(1-\tilde{\rho}_{\max})]m}{\epsilon(\tilde{\rho}_{\max}-r)}.$$

However, $\tilde{\rho}_{\max} = \max_x \rho(\widetilde{S}_x)$ varies from time to time due to the perturbation, and data owner can substitute it with the true maximum value $\rho_{\max} = \max_x \rho(S_x)$, then we get the result. □