Privacy Challenges and Solutions in the Social Web

By Grigorios Loukides and Aris Gkoulalas-Divanis

Research related to online social networks has addressed a number of important problems related to the storage, retrieval, and management of social network data. However, privacy concerns stemming from the use of social networks, or the dissemination of social network data, have largely been ignored. And with more than 250 million active Facebook (**http://facebook.com**) users, nearly half of whom log in at least once per day [5], these concerns can't remain unaddressed for long.

Broadly speaking, there are two types of concerns: 1) data access-related issues, where users are allowed access to other users' data while using the social network, and 2) data publication-related issues, or how social network data can be shared with recipients in a privacy-preserving way.

The first type of privacy concerns are raised because social networks contain rich information that users are only willing to share with trusted parties, such as users' real identities, their participation in certain activities or groups, and their multimedia content. Since there are cases in which users lack control of how this information is shared with other parties, privacy may be compromised [3]. For example, conversations in blogs may be recorded and published on the comment-tracking Web site coComment (**http://cocomment.com**) [4] without explicit user notification. Such publication may create privacy concerns, as in a case where a user found messages posted to Citibank published on coComment which questioned the security of Citibank's Web site [12].

Furthermore, users may not realize which of their activities are visible to other users of the social network. For instance, Facebook has different privacy policies for "photo album" and "profile photos," with the former open to more users than the latter.

Ensuring that social networks operate in a privacy-aware manner requires controlling which users' information a certain social network user can access. Although this can be achieved by simple opt-in/opt-out policy, such as Facebook's "edit album privacy" feature, a more careful consideration of the problem involves flexible privacy policy specification and enforcement techniques, or access control. Since such problems are closely related to the area of data security, we do not discuss them further here and refer the reader to Sloman and Lupu [13], and Squicciarini, Shehab, and Paci [14] for details.

Privacy concerns also arise when the owners of social networks disseminate parts of their network data to untrusted recipients, usually marketers, for analyzing or mining. Users may be both emotionally and economically harmed when such analysis reveals their sensitive information. Assume, for example, that a social network owner wants to share data with a marketing researcher aiming to statistically analyze this data in order to identify potential customers based on users' common habits. Releasing the data intact is obviously undesirable because it potentially contains the user's identifying information—personal name, contact information, and phone number—that should not be disclosed.

In addition, users often do not feel comfortable revealing specific activities or affiliations, or sensitive data, such as brand preferences or Web browsing history. This necessitates preventing the inference of such information in released social network data.

Consequently, it is important to identify social network data publication scenarios that may lead to privacy breaches, and develop techniques to prevent them. In what follows, we briefly survey some recent work focusing on attacks related to data dissemination and methodologies to guard against them.

Privacy Attacks on Published Social Network Data

Before discussing specific attacks, it's important to mention that a social network can be modeled as an undirected graph, where each node represents a distinct social network user having a certain *ID* or name, and an edge connecting two nodes corresponds to a relationship, or friendship, between the corresponding users. In addition, we assume that each node carries a label containing a user's attributes. An example of social network is illustrated in Figure 1(a).

Releasing social network data may allow three different types of attacks: identity, link, and content disclosure [11], which will be discussed below. To keep the discussion simple, we do not formally characterize these attacks (see Zhou et al. [18] for a formal presentation).

Identity Disclosure

The identity disclosure problem occurs when published social network data may allow the identity of social network users to be revealed. A straightforward solution to thwart identity disclosure is *de-identification*, that is, removing the user *ID*, or replacing it with a pseudonym.

Unfortunately, this simple process is not sufficient to prevent identity disclosure when the data recipient possesses some form of background knowledge, used in conjunction with the graph structure to discover the identity of individuals. As demonstrated by Hay et al. [6], the risk of identifying a user depends on the structural similarity of the nodes in the released graph, and the attacker's background knowledge.

More specifically, Hay et al. [6] assumed that an attacker knows a user's *ID*, and his or her node's *degree*—the number of this user's friends—or the existence of a specific type of subgraph around him or her. Such information may be available through sources external to the released data or obtained by intruding into the network.

Consider, for example, the graph shown in Figure 1(a) and an attacker knowing that Mary has four friends. Using this knowledge, the attacker is able to infer that Mary is represented using "3" in the published data of Figure 1(b).

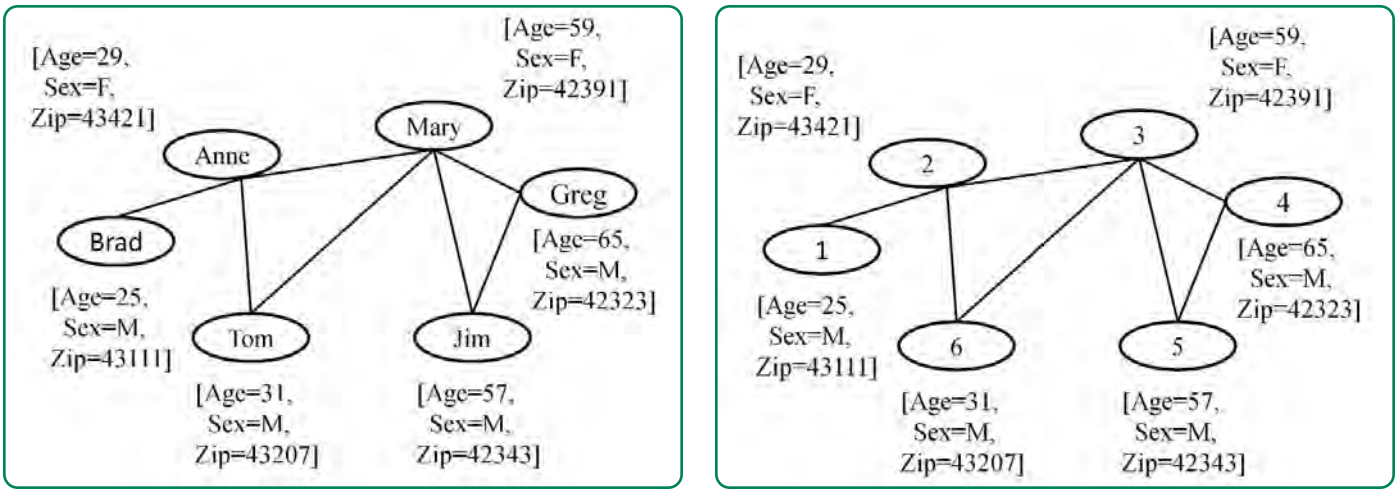

Figure 1: An example of a social network and its de-identified counterpart where (a), on the left, shows an original social network and (b), on the right, a de-identified social network.

Link Disclosure

As mentioned previously, some relationships, or links, between social network users may also convey sensitive information. 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.

In the first context, 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 [17].

Backstrom et al. [2] considered two different types of attacks. 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.

> **"Ensuring that social networks operate in a privacy-aware manner requires controlling which users' information a certain social network user can access. Although this can be achieved by simple opt-in/opt-out policy, a more careful consideration of the problem involves flexible privacy policy specification and enforcement techniques, or access control."**

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. Experiments conducted using a social network data consisting of 4.4 million users extracted from LiveJournal.com (**http://livejournal.com**), a popular blogging site, demonstrated that creating only seven new user accounts results in compromising approximately 2,400 links on average.

As an example of an active link disclosure attack, consider the network data of Figure 1(a), in which every user has knowledge of his or her immediate neighbors only, and an attacker using Greg and Jim as "fake" users, both of whom are related to Mary. When this data is published, as in Figure 1(b), the attacker can identify Mary in the released data, and disclose her potentially sensitive links to Anne and Tom. Such an attack is realistic in networks where a user can interact with other users without their consent, for example, via email or instant messages. Similarly, when Brad is able to identify himself in the data of Figure 1(b) (as he is the only user having one neighbor), he can issue a passive attack, disclosing Anne's links.

Now let's consider what happens when link structure is inferred from unpublished data. Different from the previous network data publishing scenario, there are cases in which the data owner is not willing to reveal the links that exist among users, since these links can be an asset used to maximize profits (through advertising products to users, for example). However, it has been shown that an adversary can still reveal sensitive relationships by exploiting the link structure of a number of non-anonymous nodes, which are bribed to reveal their links to the adversary.

Specifically, Korolova et al. [8] showed that the number of compromised nodes needed to reveal links decreases exponentially with an increase of the maximum distance in which the incident nodes remain visible to a compromised node. Using social network data extracted from LiveJournal.com, the authors also showed that, for realistic values of maximum distance (3), bribing a very small percentage of users (36 out of 572,949 users) suffices to disclose the relationships of 80 percent of the nodes to their immediate neighbors.

To exemplify this attack scenario, assume now that the data of Figure 1(a) is extracted from LinkedIn (**http://linkedin.com**), a social network where each user has knowledge of his or her friends and their immediate friends, and an attacker who has bribed Mary. By doing so, the attacker is able to reveal the link structure of the entire network.

Content Disclosure

In content disclosure, the sensitive data associated with each node is compromised when social network data is published. This is because published data may contain users' personal information, such as demographics, which have been shown to allow uncovering the identities of users when linked to publicly available data sources [15]. Consider, for example, the network data of Figure 1(b). Although, this data has been de-identified, knowing the age of a particular user allows an attacker to identify this user in the graph.

Methods for Privacy-Preserving Publication of Social Network Data

In this section, we briefly present some indicative methodologies that have been proposed to prevent identity, link and content disclosure, thus enabling social network data to be published in a privacy-preserving manner. All these methods are applied on de-identified social network data.

Identity Disclosure

Several algorithms that attempt to prevent identity disclosure by limiting the probability of identifying nodes of the social network in the published data have been developed. These algorithms fall into three categories with respect to the data modification techniques they employ.

The first category of algorithms is based on random perturbation. Algorithms of this category perform local modifications to the structure of the graph representing social network data to achieve privacy. The main idea behind these algorithms is that attackers will not be able to use their background knowledge to exclude nodes that do not match the individual they are trying to re-identify, since all the graphs that could have been modified using perturbation to produce the anonymized graph need to be considered instead.

The first algorithm of this kind was proposed by Hay et al. [7] and works in two steps. It first constructs an anonymized graph by deleting m edges of the original graph selected uniformly at random. This creates an interim graph from which m edges, selected uniformly at random from the set of non-existent edges in this graph, are subsequently inserted to it.

Consider, for example, an attacker knowing that Anne has two friends (i.e., a degree of 2) and applying this algorithm on the de-identified version of the graph shown in Figure 2 (left side) using m=1. After deleting the edge between Anne and Tom, the algorithm inserts an edge between Brad and Tom, constructing the graph of Figure 2 (right side) as a (possible) result. Notice that the attacker cannot identify the node corresponding to Anne using the latter graph, even when the attacker has knowledge of the type of algorithm applied for protection and the parameter m. In other words, this attacker is not certain about which of the three nodes corresponds to Anne. In fact, the graph illustrated on the right side of Figure 2 could have been created from either of the graphs shown in Figure 3. However, this algorithm does not bound the probability of identifying nodes in the social network, nor does it minimize data distortion.

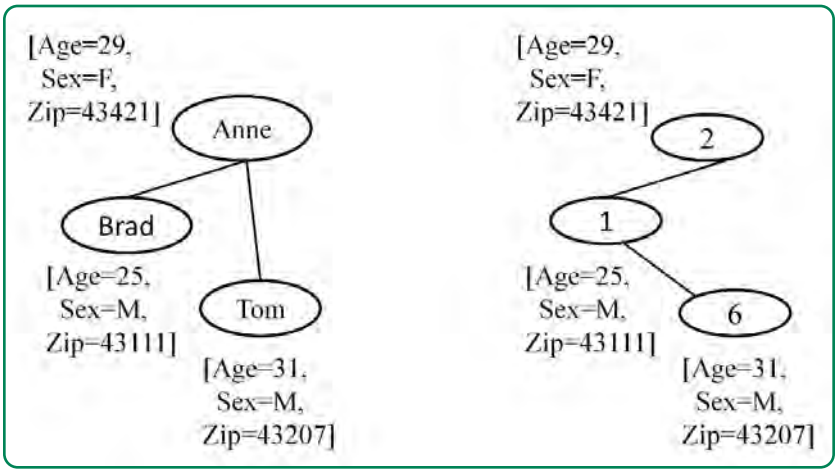

Figure 2: An example of a social network and its anonymized counterpart. The original social network before de-identification is shown on the left, and the anonymized social network using [7] *with* m*=1 is shown on the right.*

Liu and Terzi have also proposed an algorithm based on random perturbation [10]. This algorithm modifies the graph so as to ensure that there are at least k nodes having the same degree in the anonymized graph. Thus, it can prevent identity disclosure when an attacker knows the degree of some nodes of the graph, because this attacker will need to distinguish the node corresponding to a target individual among k nodes sharing the same degree. In addition, this algorithm optimizes a simple data utility function based on the number of edges added to the graph.

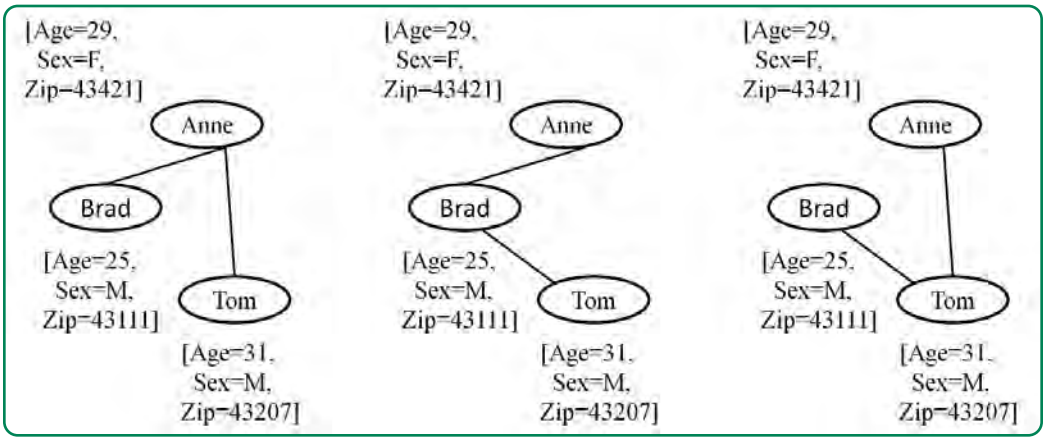

Figure 3: Possible graphs that could have been created by applying random perturbation methods on the social network of Figure 2.

The second type of method is based on grouping nodes and edges together in an attempt to prevent an attacker from distinguishing between nodes belonging in the same group [18]. An algorithm based on this idea was proposed in [6]. This algorithm works by placing the nodes of the original network data into groups of a controlled size, called "supernodes,"

and the edges by "superedges," which report the density of edges in the original graph. This process ensures that the probability of identifying a node in the published data, which is related to the of the supernode it belongs, is limited. Furthermore, it attempts to preserve utility by finding the partitioning that best retains structural characteristics of the original graph using a fitting model based on a maximum likelihood approach.

Finally, Zhou and Pei [18] have developed an algorithm which combines random perturbation and data generalization to prevent identity disclosure. Different from the previous two types of methods, their algorithm takes into account the labels of nodes (containing user attributes), and performs anonymization by generalizing these attributes and by inserting edges. However, it considers a limited class of attackers with knowledge of the induced subgraph of a node and its immediate neighbors and quantifies information loss using a heuristic measure based on the number of edges and nodes added to the anonymized neighborhoods and the way labels are generalized.

Link Disclosure

Perturbation of social network data has also been applied to thwart link disclosure attacks by Ying and Wu [16]. The authors considered how a graph can be published in a form that conceals the sensitive links, while preserving the spectrum of the original graph, which reflects topological properties including diameter, long paths, and the cluster structure. Two methods based on edge modification were also proposed. 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 way that the degree distribution of the original graph is not affected.

Instead of modifying the graph structure, Lescovec and Faloutsos [9] proposed generating a graph with a different structure than the original, but with similar topological properties, such as degree distribution, diameter, or spectrum. 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 developed. The underlying assumption of this algorithm is that the employed graph fitting model is able to generate graphs that obey many of the patterns found in real graphs. The effectiveness of this approach was experimentally verified using real-world data.

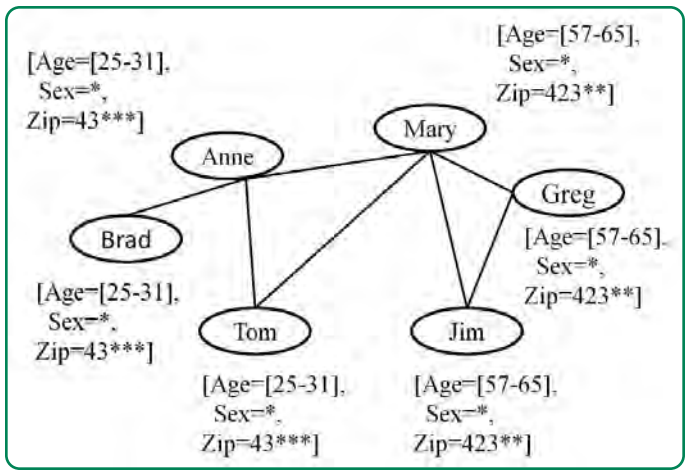

Figure 4: A 2-anonymization of the social network of Figure 1a is shown. (The asterisk denotes a character that has been removed, and age values are transformed into intervals.)

Content Disclosure

Preventing content disclosure may be achieved by applying perturbation and anonymization approaches originally developed for relational data [1], after representing users' identity and attributes contained in the label as a relational table. An example of preventing content disclosure through the application of 2-anonymity [15], is shown in Figure 4. Notice that in this data no user can be identified based on the set of attributes {Age, Sex, Zip} with a probability of more than 1/2.

Staying Connected, Privately

Privacy-preserving publication of social networks is a very promising, challenging, and rapidly evolving research area. Ensuring privacy in this context still requires a number of issues to be carefully addressed.

First, although there has been much progress in terms of designing algorithms to protect privacy in relational data, these algorithms are generally not applicable in the context of social network data due to its significant complex structure. This calls for effective and efficient algorithms to protect social network data from identity and link disclosure. On the positive side, principles and methodologies originally developed for relational data, such as *k*-anonymity and generalization [15] can provide the basis for designing such algorithms.

Second, since a social network is an environment in which millions of users interact with one another on a daily basis, both the dynamic nature of the data and the privacy expectations of users need to be taken into account to design useful methodologies for publishing social network data.

Biographies

*Grigorios Loukides (**grigorios.loukides@vanderbilt.edu**) is currently a postdoctoral research fellow in the Department of Biomedical Informatics at Vanderbilt University. He holds a BSc from the University of Crete, Greece, and a PhD from Cardiff University, both in computer science. His research interests lie broadly in the fields of privacy and trust in data management and emerging database applications.*

*Aris Gkoulalas-Divanis (**aris.gkoulalas@vanderbilt.edu**) has a BS from the University of Ioannina, an MS from the University of Minnesota, and a PhD from the University of Thessaly in computer science. He is currently a postdoctoral research fellow in the Department of Biomedical Informatics at Vanderbilt University. He has served as a research assistant in both the University of Minnesota (2003-2005) and the University of Manchester (2006). His research interests are in the areas of privacy preserving data mining, privacy in medical records and privacy in location-based services.*

References

1. Aggarwal, C. C. and Yu, P. S. 2008. Privacy-preserving data mining: Models and algorithms. In *Advances in Database Systems*. Springer.

2. Backstrom, L., Dwork, C., and Kleinberg, J. 2007. Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the International World Wide Web Conference*. 181-190.

3. Chew, M., Balfanz, D., and Laurie, B. 2008. (Under)mining privacy in social networks. W2SP: Web 2.0 Security and Privacy, W2SP website.

4. CoComment: Comment tracking service. 2009. **http://www.cocomment.com** (accessed 9/17/2009).
5. Facebook usage statistics. 2009. **http://www.facebook.com/press/info.php?statistics** (accessed 9/17/2009).
6. Hay, M., Miklau, G., Jensen, D., Towsley, D.F., and Weis, P. 2008. Resisting structural re-identification in anonymized social networks. *Proc. VLDB 1*, 1. 102-114.
7. Hay, M., Miklau, G., Jensen, D., Weis, P., and Srivastava, S. 2007. Anonymizing social networks. Tech. rep. TR07-19. University of Massachusetts Amherst.
8. Korolova, A., Motwani, R., Nabar, S. U., and Xu, Y. 2008. Link privacy in social networks. In *Proceedings of the ACM Conference on International and Data Management* (*CIKM*). 289-298.
9. Leskovec, J. and Faloutsos, C. 2007. Scalable modeling of real graphs using Kronecker multiplication. In *Proceedings of the International Conference on Machine Learning* (*ICML*). 497-504.
10. Liu, K. and Terzi, E. 2008. Towards identity anonymization on graphs. *ACM SIGMOD*. 93-106.
11. Liu, K., Das, K., Grandison, T., and Kargupta H. 2008. Privacy-preserving data analysis on graphs and social networks. In *Next Generation of Data Mining* Kargupta, H., Han, J., Yu, P., Motwani, R., and Kumar, V., Eds. CRC Press. Chapter 21, 419-437.
12. Ratcliffe-Lee, J. 2007. Huge Security hole in Citibank's online account center. **http://jratlee.tumblr.com/post/189652** (March 2007).
13. Sloman, M. and Lupu, E. 2002. Security and management policy specification. *IEEE Network 16*, 2. 10-19.
14. Squicciarini, A. C., Shehab, M., and Paci, F. 2009. Collective privacy management in social networks. In *Proceedings of the International World Wide Web Conference*. 521-530.
15. Sweeney, L. 2002. Achieving *k*-anonymity privacy protection using generalization and suppression. *Int. J. Uncertainty, Fuzziness and Knowl.-Based Syst. 10*, 5. 571-588.
16. Ying, X. and Wu, X. 2008. Randomizing social networks: A spectrum preserving approach. *Secure Data Management Workshop*. 739-750.
17. Zheleva, E. and Getoor, L. 2007. Preserving the privacy of sensitive relationships in graph data. *Privacy, Security, and Trust in KDD*. Springer. 153-171.
18. Zhou, B., Pei, J. and Luk, W. 2008. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *ACM SIGKDD Explorations 10*, 2. 12-22.