Security and Privacy in Social Networks

Barbara Carminati
Università degli Studi dell'Insubria, Italy

Elena Ferrari
Università degli Studi dell'Insubria, Italy

Andrea Perego
Università degli Studi dell'Insubria, Italy

INTRODUCTION

Web-based social networks (WBSNs) are online communities that allow users to publish resources (e.g., personal data, annotations, blogs) and to establish relationships, possibly of a different type ("friend," "colleague," etc.) for purposes that may concern business, entertainment, religion, dating, and so forth. In the last few years, the usage and diffusion of WBSNs has been increasing, with about 300 Web sites collecting the information of more than 400 million registered users. As a result, the "net model" is today used more and more to communicate, share information, make decisions, and 'do business' by companies and organizations (Staab et al., 2005).

Regardless of the purpose of a WBSN, one of the main reasons for participating in social networking is to share and exchange information with other users. Recently, thanks to the adoption of Semantic Web technologies such as FOAF and other RDF-based vocabularies (Brickley & Miller, 2005; Davis & Vitiello, 2005; Golbeck, 2004), accessing and disseminating information over multiple WBSNs has been made simpler (Ding, Zhou, Finin, & Joshi, 2005). If this has been quite a relevant improvement towards an easier sharing of information, it makes more urgent that content owners have control over information access. In fact, making available possibly sensitive and private data and resources implies that they can be used by third parties for purposes different from the intended ones. As a matter of fact, users' personal data and resources are regularly exploited not only by companies for marketing purposes, but also by governments and institutions for tracking persons' behaviors and opinions, and in the worst case, by online predators (Barnes, 2006).

It is then a challenging issue to devise security mechanisms for social networks, able to protect private information and regulate access to shared resources. In this article, besides providing an overview of the characteristics of the WBSN environment and its protection requirements, we illustrate the current approaches and future trends to social network security, with particular attention paid to the emerging technologies related to the so-called Web 2.0.

BACKGROUND

Usually, a social network is defined as a *small-world network* (Watts, 2003), consisting of a set of individuals (persons, groups, organizations) connected by personal, work, or trust relationships. Social networking is then a quite broad and generic notion, which in the Web context might be applied to any kind of virtual community. For instance, users registered to a Web service, such as Web mail, online journals, or newspapers requiring a subscription, can be considered as a social network. In the following, we adopt the definition provided by Golbeck (2005), according to which an online community's Web site can be considered a Web-based social network only if it satisfies the following conditions:

- Relationships are explicitly specified by its members, and not inferred from existing interactions (e.g., a mailing list can be used to infer implicit relationships).
- Relationships are stored and managed by using technologies, such as database management systems, allowing relationship analysis and regulating access and retrieval of relationship data.
- Members are able to access relationship information, at least partially.

Born in the late 1990s, in the last few years WBSNs gained increasing interest and diffusion. Although the first and most successful ones, such as MySpace, Friendster, and Facebook, were formerly designed for entertainment and socialization purposes, they are currently establishing themselves as a business model, through which institutions and organizations can set up a collaborative environment for specific purposes, and where it is possible to share resources at an intra- and inter-organizational level. Due to the great amount of collected data, WBSNs are currently the subject of great interest for statistical analysis (Wasserman & Faust, 1994; Freeman, 2004), since they may provide useful information not only to social researchers, but also for marketing purposes.

WBSNs may provide different kinds of services, ranging from information and contact sharing, to collaborative

rating, collaborative work environments, and so on. However, independently from the specific purposes of a WBSN, members' relationships are the core information on which all the provided services are based. In fact, they can be used not only to create connections among people sharing similar interests, but also to customize WBSN services themselves. This is particularly true in WBSNs supporting collaborative rating: in such a context, ratings may be given different weights, depending on the relationships existing between WBSN members. For instance, it may be the case that a given WBSN member m_1 considers more relevant (or trustworthy) the opinions of member m_2 than, say, those of member m_3. For this purpose, some WBSNs allow their members not only to specify personal relationships (e.g., "friend of," "colleague of") but also to establish *trust* relationships, which express how much they trust the other members either with respect to a specific topic (*topical trust*) or in general (*absolute trust*). For a thorough discussion on trust relationships and how they can be used, we refer the reader to the work by Golbeck and Hendler (2006).

As far as security is concerned, current WBSNs enforce simple protection mechanisms, which only allow their members to label given information as public or private, or to make it available to WBSN members with whom there exists a direct relationship of a given type (friend, colleague, etc.). However, these solutions on one hand may dramatically reduce the possibility of sharing information, which is the basic function of a WBSN, and on the other hand, they do not necessarily grant the required protection to personal information. In fact, giving to WBSN members just the choice of stating whether a given resource is public or private may result in hiding a huge amount of information. Moreover, it may frequently happen that WBSN members make publicly available resources that are accessed by people different from the ones they intended—the most typical case is a student publishing photos or blogs in recreational WBSNs, without considering that they can be accessed by his or her teachers.

Additionally, personal information and relationships among WBSN members must be protected when WBSN data are analyzed by data mining tools, that is, tools capable of analyzing massive datasets of personal information with the purpose of extracting models of social and commercial interest.

SECURITY AND PRIVACY REQUIREMENTS IN SOCIAL NETWORKS

In this section we consider the security and privacy issues related to WBSNs from two different points of view. First, we discuss the privacy-preserving techniques adopted to allow statistical analysis on social network data without compromising WBSN members' privacy, and then we illustrate the current approaches aimed at enforcing privacy protection when performing access control.

Privacy-Preserving Social Network Analysis

Data collected by WBSNs are an important source for social and marketing analysis, which may provide useful information on the evolution of a social community, collaborative problem solving, information distribution, and so on. Additionally, they can also be used to optimize social network services and customize them with respect to users' preferences and interests. However, when analyzing WBSN data for statistical purposes, it is necessary to avoid as much as possible disclosing private information about WBSN members.

So far, this issue has been addressed by anonymizing the network graph according to two main strategies, namely, *node anonymization* and *edge perturbation*. The former strategy aims at hiding members' identities by labeling the corresponding network nodes with random identifiers (naïve anonymization). In case nodes are associated with attributes which can be used to identify the corresponding user, the possibility of using techniques based on k-anonymity (Sweeney, 2002) has been discussed−see, e.g., Zheleva and Getoor (2007). By contrast, edge perturbation performs a set of random edge deletions and insertions, which prevent an attacker from inferring the identity of network nodes based on the existing relationships but, at the same time, preserve the utility of the graph for network analysis.

It has been noticed that the proposed solutions to node anonymization do not grant total privacy protection. In particular, Backstrom, Dwork, and Kleinberg (2007) carried out an extensive analysis of the possible attacks, and argued that the most effective strategies for privacy protection are those based on *interactive* techniques. According to this approach, the anonymized network graph is not disclosed; rather it is analyzed by the social network management system itself upon submission of a query, and then the result is perturbed by adding noise to the real answer.

By contrast, edge perturbation, when combined with node anonymization, grants a greater degree of protection. Examples of how such techniques are applied are provided by Frikken and Golle (2006), Hay, Miklau, Jensen, Weis, and Srivastava (2007), and Zheleva and Getoor (2007). In particular, Hay et al. (2007) report experimental results which show that random edge deletions and insertions grant graph anonymity when the perturbation affects a percentage of graph edges ranging from 5% to 10%. By contrast, a perturbation rate greater than 10% dramatically increases information loss, thus making useless the results obtained by