Future Generation Computer Systems 82 (2018) 63–76

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Estimation of privacy risk through centrality metrics

J. Alemany *, E. del Val, J. Alberola, A. García-Fornes

Universitat Politècnica de València, Camino de Vera s/n, Valencia, Spain

HIGHLIGHTS

- PRS estimates the reachability of users' posts without requiring users' intervention.
- PRS is based on the distance between users and the potential audience.
- PRS offers values based on friendship levels to deal with different risk perceptions.
- We provide and evaluate a set of centrality metrics to estimate the PRS values.
- Centrality metrics are good estimators of PRS in scenarios with a lack of knowledge.

ARTICLE INFO

Article history:
Received 26 May 2017
Received in revised form 17 November 2017
Accepted 22 December 2017
Available online 27 December 2017

Keywords:
Privacy
Social networks
Information sharing

ABSTRACT

Users are not often aware of privacy risks and disclose information in online social networks. They do not consider the audience that will have access to it or the risk that the information continues to spread and may reach an unexpected audience. Moreover, not all users have the same perception of risk. To overcome these issues, we propose a Privacy Risk Score (PRS) that: (1) estimates the reachability of an user's sharing action based on the distance between the user and the potential audience; (2) is described in levels to adjust to the risk perception of individuals; (3) does not require the explicit interaction of individuals since it considers information flows; and (4) can be approximated by centrality metrics for scenarios where there is no access to data about information flows. In this case, if there is access to the network structure, the results show that global metrics such as closeness have a high degree of correlation with PRS. Otherwise, local and social centrality metrics based on ego-networks provide a suitable approximation to PRS. The results in real social networks confirm that local and social centrality metrics based on degree perform well in estimating the privacy risk of users.

1. Introduction

The popularity of mobile devices and applications that are related to online social networking has changed the way we communicate. People now share their opinions, ideas, photos, etc. in online social networks (OSN) [1,2]. When sharing information, users are not often aware of who will or will not have access to what they have just published. This uncertainty creates a risk in the privacy of the user, which in some cases may have negative consequences if the scope of the publication reaches people who were not in the original audience. Applications related to OSN offer the possibility to configure options that are related to the privacy profile of users. However, this is often a tedious task and is usually focused on protecting the information related to the user profile and not to the privacy of the user's publications [3–5]. Some works try to address these issues with the automation of privacy settings [6–9]. However, these proposals usually require an initial intervention by the user and do not solve the problem of increasing privacy awareness. Other approaches deal with the improvement of the awareness of users regarding the misalignment of users' expected audience with the actual audience [10–12]. However, these approaches do not deal with the problem that a publication might produce if the expected audience performs sharing actions among their contacts. Assuming this scenario, there is still a potential privacy risk that should be considered.

The topological location of a user in a network is one of the main factors that influences the scope that a certain sharing action can reach [13]. The scope of a sharing action can be seen as the effect of a diffusion process. In the area of Complex Networks, spreading processes such as epidemics or information diffusion have been analyzed [14–17]. Several works have studied spreading dynamics and influential or relevant individuals in these processes based on structural properties [18–22]. From the point of view of determining the privacy risk associated to a user's sharing action,

* Corresponding author.
E-mail addresses: jalemany1@dsic.upv.es (J. Alemany), edelval@dsic.upv.es (E. del Val), jalberola@dsic.upv.es (J. Alberola), agarcia@dsic.upv.es (A. García-Fornes).

https://doi.org/10.1016/j.future.2017.12.030
0167-739X/© 2017 Elsevier B.V. All rights reserved.

it is interesting to determine if there are influential users in the path that information follows who increase the privacy risk score if they perform a re-sharing action. Influential users can initiate and conduct the dissemination of a sharing action more efficiently than "normal" users. Therefore, influential users in networks are normally more responsible for large cascades of information diffusion and contribute to increasing the privacy risk. Traditionally, centrality metrics such as degree [23], pagerank [20], k-core [24,18], closeness [25], or betweenness [26–29] have been used to detect these relevant users in networks [30,21,31].

Not all users have the same perception of risk [32–34]. On one hand, there are some users who are more comfortable with the possibility that their information can be seen by others and are even interested in achieving that effect. On the other hand, there are users that have greater privacy concerns and prefer not to disclose information that could be seen by users beyond their direct friends [35]. Depending on the users' concerns, different levels of risk perception should be considered.

In this article, we propose a Privacy Risk Score (PRS) for measuring the privacy in social networks, which provides the following major contributions:

- The privacy is oriented to the reachability of a user-sharing action instead of being focused on the misalignment of the users' expected audience with the actual audience.
- The measure provided is not only global, but it is also adjustable to the risk perception of each individual.
- The PRS does not require the user to provide information explicitly since it takes into account the paths that the publications follow in the social network.
- We provide an estimation of this measure for those scenarios in which information related to flow paths is not available. This estimation is based on an analysis of the relationship between global, local, and social centrality metrics and the proposed measure.

The rest of the paper is structured as follows. Section 2 presents previous approaches that are related to privacy score metrics. Section 3 exposes the privacy risks in social networks with an example of scenario and proposes a solution. Section 4 describes the concept of friendship level and presents the PRS. Section 5 describes a set of global, local, and social centrality metrics to estimate the PRS. Section 6, presents a set of experiments that were performed to evaluate the suitability of centrality metrics to estimate the PRS in synthetic and real network topologies. Finally, Section 7 presents conclusions.

2. Related work

In the literature, there are works that try to tackle the problem of improving the awareness of the effect of communicative actions from different perspectives. Table 1 provides an overview of relevant contributions in this area, which are classified according to the dimensions of focus.

There are approaches that provide wizards to facilitate the management of privacy profile settings. Liu et al. [3] propose a mathematical model to estimate both the sensitivity and the visibility of information items. The model computes the privacy score as a combination of the partial privacy scores of each one of the user's profile items. The privacy score considers the privacy settings of users with respect to their profile items as well as their positions. A similar approach is presented by Nepali et al. [4]. They propose a social network model, SONET, for privacy monitoring and ranking. The authors consider a privacy risk indicator that is used to describe an entity's privacy exposure factor based on the known attributes (the sensitivity and visibility of the attribute). Shehab et al. [5] present a privacy policy recommendation approach that is based on the idea that nearby users should have similar labels (permissions). The approach requires users to label a small set of their friends. These labels are propagated over the social network to provide users with privacy policy recommendations. Fang et al. [6] present a privacy wizard that considers previous labeling processes of friends as the input for their classifier. The wizard then infers labels for the other remaining friends. Vidyalakshmi et al. [7] present a framework for calculating a privacy score metric considering users' personal attitude towards privacy and communication information. Bilogrevic et al. [8] propose an information-sharing system that decides (semi-)automatically whether to share information with others. They consider a vector that encodes whether or not the information is shared based on user decisions, and then a logistic classifier makes the remaining decisions. These approaches require user intervention and assume that users are privacy aware of the consequences of their decisions. They are focused on a local view of the social network and do not evaluate other collateral effects such as information diffusion processes.

Some approaches focus on providing information about which people have or may have received information that was not addressed to them initially. These works help them to increase their privacy risk awareness and better define their social groups more carefully. Calikli et al. [10] propose an adaptive architecture that provides sharing recommendations to users as well as assisting them to re-configure the users' groups. Their proposal is based on social contexts and conflicts. This approach depends on the provision of accurate user's social contexts and conflict rules. Kafali et al. [11] provide an approach that is based on model checking that checks whether certain properties hold. The system uses as input privacy agreements of the users, user relations, the content they upload as well as some inference rules. The system specifies whether the property of interest can or cannot be violated in a given social network. Mester et al. [12] developed a platform where agents interact to reach a consensus on a post to be published. The agent is aware of the user's privacy concerns, expectations, and the user's friends. When a user is about to post new content, the agent reasons on behalf of the user to decide which other users would be affected by the post and contacts those users' agents. However, the privacy concerns of a user should be predefined. Yang et al. [36] present a privacy metric of user i sharing information with a neighbor j as a trade-off between user i's concerns and incentives of sharing information with j. They present privacy risk as an individual metric, without considering other potential users that might re-share information.

From our point of view, privacy risk does not only concern the problem that information might reach people who were initially not expected to receive it. Assuming that people who received the information are part of the target audience, it must also be taken into account that there is still a problem if one user of this intended audience re-shares the information. Then, the original user loses control over the scope of the information. For this reason, it is important to consider the privacy problem from a network perspective instead of individuals alone. The audience that is allowed see the information that a user publishes is influenced by the structure of the social network. Network models that mimic the patterns of connection in real networks (i.e., Erdös–Rényi [37–39], Barabási–Albert [40,41], and Watts–Strogatz [42,43]) facilitate the analysis of the implications of those patterns [44]. Small-world, Scale-free, and Random models are very common structures in social networks. The Small-world model is characterized by the transitivity in strong social ties and the ability of weak ties to reach across clusters. The Scale-free model exhibits a power-law degree distribution where there is a small set of vertices with a degree that greatly exceeds the average. The random model assigns equal probability to all graphs with exactly the same number of edges.