**Network-aware privacy risk estimation in online social networks**

**Social Network Analysis and Mining manuscript No.**
(will be inserted by the editor)

# Network-aware Privacy Risk Estimation in Online Social Networks

**Ruggero G. Pensa · Gianpiero Di Blasi · Livio Bioglio**

Received: date / Accepted: date

**Abstract** Online social networks expose their users to privacy leakage risks. To measure the risk, privacy scores can be computed to quantify the users' profile exposure according to their privacy preferences or attitude. However, user privacy can be also influenced by external factors (e.g., the relative risk of the network, the position of the user within the social graph), but state-of-the-art scores do not consider such properties adequately. We define a network-aware privacy score that improve the measurement of user privacy risk according to the characteristics of the network. We assume that users that lie in an unsafe portion of the network are more at risk than users that are mostly surrounded by privacy-aware friends. The effectiveness of our measure is analyzed by means of extensive experiments on two simulated networks and a large graph of real social network users.

**Keywords** privacy measures · online social networks · centrality · simulation · computational social science

## 1 Introduction

Online social networks are permeating most aspects of our life. More than two billions active social accounts are producing petabytes of behavioral and interaction data daily. At the same time, the famous "six degrees of separation" theory has been far exceed in Facebook, where an average degree of 3.57 has been recently observed[1]. This massive interconnection intrinsically exposes social network users to the risk of privacy leakage.

R.G. Pensa, G. Di Blasi, L. Bioglio
University of Turin
Department of Computer Science
Tel.: +39-011-6706798
Fax: +39-011-751603
E-mail: ruggero.pensa@unito.it

[1] `https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/`

If, from one hand, many users are informed about the risks linked to the disclosure of sensitive information (private life events, sexual preferences, diseases, political ideas, among others), on the other hand the awareness of being exposed to privacy breaches each time we disclose information that apparently is not sensitive is still insufficiently widespread. In this regard, daily activities may reveal information that can be used by others in a negative manner. For example, a GPS tag far from home or pictures taken during a journey may alert potential burglars, or the disclosure of family relationships may expose our own or other family members' privacy to criminal offence risks, as well as source of tort liability. Most troubling of all, it has been shown that by leveraging Facebook user's activity it is possible to infer some very private traits of the user's personality (Kosinski et al., 2013). This inference capability has been recently exploited to help propel Donald Trump to victory in the last U.S. presidential elections (González, 2017) and was at the very center of the Facebook–Cambridge Analytica scandal in early 2018. This privacy breach event has multiplied the interests in the protection of human dignity and personal data, and privacy has become a primary concern among social network providers and web/data scientists.

Although social platforms often provide some kind of notification intended to inform their users about the risks of private information disclosure, many people simply overlook the dangers due to the uncontrolled disclosure of their (and others') personal data. Therefore, following the recent scandals, most social media have considerably improved their tools for controlling the privacy settings of the user profile (e.g., Instagram can now limits the visibility of stories to "close-friend"), but such tools are often hidden and not that user-friendly. Consequently, they are barely utilized by most users. Recent machine learning and data mining studies try to go beyond these limitations by proposing some measures of users' profile privacy based on the way they customize their privacy settings (Liu and Terzi, 2010; Wang et al., 2014), or lightening the customization process of the privacy settings by means of guided tools and wizards (Fang and LeFevre, 2010; Song et al., 2018). Privacy measures (Wagner and Eckhoff, 2018), in particular, when associated to popup alerts or other visual components (e.g., gauges or discharging batteries (Talukder et al., 2010)), may enhance user's perception of privacy, according to the principles of *Privacy by Design* specifications (Cavoukian, 2012).

However, the privacy measures proposed so far (Liu and Terzi, 2010; Wang et al., 2014) have a strong limitation. In fact, the privacy risk is not just a matter of users' preferences (i.e. to which friends a user is wishing to disclose each particular action/post); it is also heavily affected by the characteristics of the social network they belong to., i.e., their centrality within the network and the attitude of their friends towards privacy. According to a recent computational science study (Bioglio and Pensa, 2017), even restraining privacy settings are ineffective when the user is located within an unsafe network, i.e., a network where the majority of nodes have little or no awareness about their own and others' privacy. When a user posts something private or sensitive in a subnetwork where the majority of individuals are aware about their own and others'

privacy, the risk that her sensitive information spread in the network is low (this condition is similar to a very well-known phenomenon in epidemiology, called *herd immunity*). Hence, her privacy risk is lower than the risk of a user posting something private in a network where many individuals are not aware about their own or others' privacy. In the latter case, a "like" or a comment under the post in question may trigger the diffusion of private information in the network. Malicious users may leverage this information to cheat or commit some evil acts against the author of the post at issue. To explain the influence of the network on user privacy, let us consider two examples.

*Example 1* Two users $u_1$ and $u_2$ share the same attitude to their own privacy protection. However, user $u_1$ occupies a central position within the social network (her centrality is high (Newman, 2010)), while $u_2$ is a peripheral user (her centrality is low). According to these hypotheses, user $u_1$ should be more exposed to privacy leakage than $u_2$, since $u_1$'s posts have more chance of being diffused than $u_2$'s posts.

*Example 2* Two users $u_1$ and $u_2$ have exactly the same attitude towards the protection of their own privacy. However, user $u_1$ is mostly connected to friends that do not care that much about their own privacy leakage, while $u_2$ is principally surrounded by friends that, instead, care about their own (and others') privacy. According to these hypotheses, user $u_1$ has higher probability of being exposed to privacy leakage than $u_1$. Roughly speaking, there will be some portions of the social network that are weaker than others from the point of view of privacy leakage.

These sample scenarios lead to the intuition that privacy risk in a social network may be modeled similarly as page authority in a hyperlink graph of web pages. According to a well-known theory (Brin and Page, 1998), more authoritative web sites are likely to receive more links from other web sites that are authoritative in their turn. In this paper, we make the hypothesis that the concept of "importance" of a web-page can be transposed into the concept of "privacy risk" of users in a social network as follows: the more an individual is surrounded by friends that are careless about their privacy, the more the privacy of that individual is likely to be exposed to concrete privacy leakage risks.

With the goal of helping users' enhance their privacy awareness in their cybersocial world, in this paper we propose a new network-aware computational method for measuring the privacy risk, inspired by personalized Pagerank (Jeh and Widom, 2003), one of the best known algorithms to rank web pages according to a personalized view of their authority (or importance). Furthermore, with the aim of supporting our claims, we report on a social experiment we performed, which involves more than one hundred Facebook users. Thus, our approach is validated on a sample which is unusually large for this type of experiments (Furini and Tamanini, 2015). Thanks to this experiment, we show the effectiveness of our privacy measure not only on two simulated networks but also on a large network of real Facebook users. Additionally, these

experiments also allow us to infer a practical estimate of the intrinsic risk due to the user attitude towards privacy (i.e., the risk due principally to the users' willingness to disclose their own personal data to other users) using the information carried out by the social graph as an alternative to state-of-the-art policy-based privacy scores (Liu and Terzi, 2010; Wang et al., 2014). Our risk estimation is inferred from the privacy attitude computed for a group of real Facebook users and, additionally, lends itself to an intuitive interpretation based on common social network user experience.

The main contributions of this paper can be summarized as follows: i) we propose a centrality-driven privacy score that takes into account the level of privacy awareness of the subnetwork when estimating the privacy risk of each individual user; ii) starting from a survey conducted on real Facebook users, we identify a practical estimator of the intrinsic privacy risk of each user; iii) we show experimentally that our privacy score captures a more reasonable risk of privacy disclosure and leakage in realistic social networks.

The remainder of the paper is organized as follows: we briefly review the related literature in Section 2; the network-aware privacy score and the intrinsic risk computation are presented in Section 3; Section 4 provides the details and results of our experimental study; finally, in Section 5, we draw some conclusions, discussing both limitations and strengths of our approach, and propose some future research directions.

## 2 Related work

The increasing success of online social networks in early 2010's has also soon highlighted their weaknesses, so that more and more research efforts have been devoted to study privacy protection methods for social profiles. Most research works focus on three main strategies: i) using data anonymization and obfuscation techniques to preserve the identity of social network users; ii) designing privacy protection mechanisms involving access control rules and policy definition; iii) measuring the privacy level of users to provide them with a practical means to assess their actual risk.

*Identity disclosure protection.* In privacy and social network analysis, most research interests have been focused on the identification and formalization of different privacy breaches and on the anonymization of identities in networked data (Rathore et al., 2017). This goal is achieved by modifying the social graph so as to minimize the probability of identifying an individual within the network, by either anonymizing only the network structure or anonymizing both network structure and user attributes (Zheleva and Getoor, 2011). The most relevant contributions address the problem of graph anonymization by applying edge generalization (Hay et al., 2008; Cormode et al., 2009), randomization (Ying and Wu, 2011; Vuokko and Terzi, 2010), modification (Zou et al., 2009; Liu and Terzi, 2008; Zhou and Pei, 2011), or differentially private mechanisms (Hay et al., 2009; Task and Clifton, 2012). The problem of identity disclosure

leveraging user activity in online social networks has been explored as well: Buccafurri et al. (2016), for instance, propose a cryptographic method to protect the identity of the individuals performing "like" actions on social media. All these research works are only marginally related to ours.

*Access control and policy definition.* Although it is well established that user privacy protection in social media involve multiple complex factors (Litt, 2013), the existing privacy controls for online social networking sites are barely utilized in practice, since they are not fully socially aware (Misra and Such, 2016). This statement is confirmed by (Liu et al., 2011), who show that 36% of Facebook content is shared without modifying the default privacy settings and, consequently, it is exposed to more users than expected. Privacy fatigue, (i.e., the tendency of online users to disclose greater information over time due to increasingly complex and less usable privacy controls) is another factor that has been recognized to play a significant role in favoring behaviors which endanger information privacy (Choi et al., 2018). Thus, many research efforts aim at supporting people to recognize and prevent privacy issues in social media. Squicciarini et al. (2014) propose an ontology-based mechanism for privacy protection, which supports semi-automated generation of access rules for users' profile items. Fang and LeFevre (2010), instead, propose a prototype of social networking wizard leveraging active learning. Their tool build a classification model of a user's friends based on their profile to assign them access privileges on the user's profile items. The model is improved incrementally by asking the user to allow or deny the visibility of profile items to friends selected according to an uncertainty sampling criterion. More recently, Song et al. (2018) have proposed a taxonomy-guided learning model that predicts which personal aspects are uncovered by the posts and constructs standard guidelines to regularize users' actions for preventing their privacy leakage. Such and Rovatsos (2016) and Such and Criado (2016) suggest to negotiate (and possibly merge) conflicting privacy preferences of multiple users on any individual item by using a computational mechanism. Other very recent works focus on recommendation of privacy settings for images representing sensitive content (Yu et al., 2017) and on solving possible privacy leakage conflicts in photos containing individuals others than the user who posted them (Xu et al., 2017). Finally, in Pensa and Blasi (2017), the authors propose a profile control framework to adjust personal privacy settings according to a measured risk of leakage.

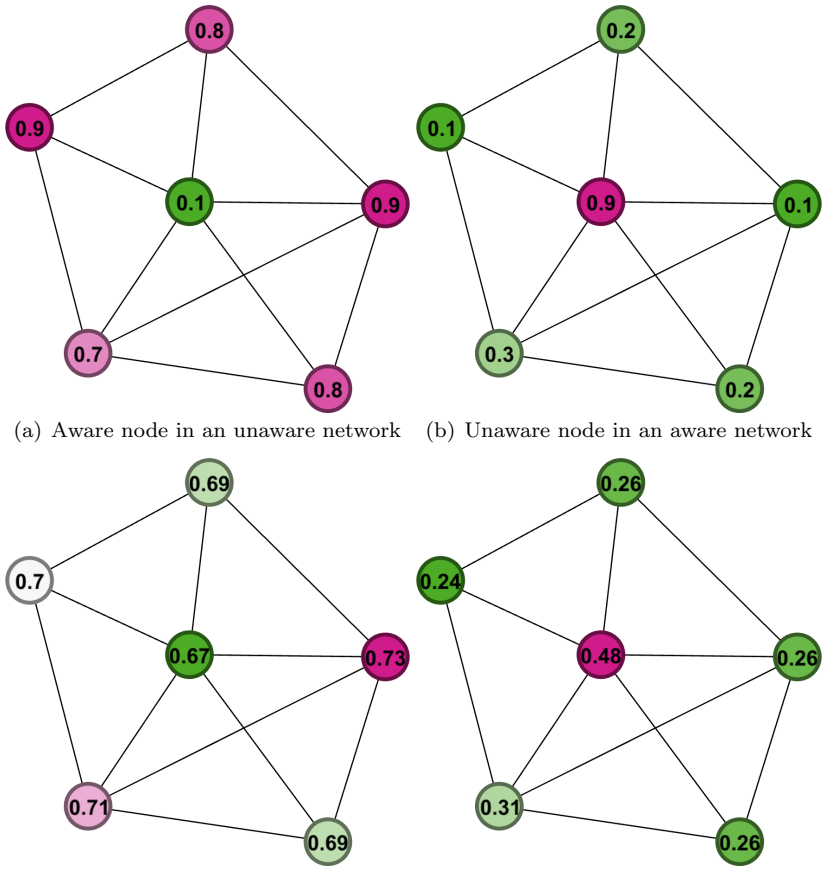
*Tools and metrics for privacy assessment.* In most cases, disclosing information on the web is the result of a voluntary activity. Hence, a common opinion is that users should care about their privacy during their interaction with other social network users. Thus, the problem of measuring and improving risk perception has gained popularity among researchers. Cetto et al. (2014), for instance, present an online game that allows Facebook users to test their knowledge of the visibility of their actual shared personal items and provides them with some recommendations aimed at improving their privacy settings.

Akcora et al. (2012a,b) propose a measure that quantify of how much it might be risky to have interactions with friends, in terms of disclosure of private information. As Fang and LeFevre (2010), they also use an active learning approach to estimate user risk from few required user interactions. The authors Talukder et al. (2010) present a privacy protection tool that measures the inference probability of sensitive attributes from friendship links. In addition, they suggest self-sanitization actions to regulate the amount of leakage. Becker and Chen (2009) show that a majority of users' personal attributes can be inferred from social circles and present a tool to detect unintentional information loss in online social networks by quantifying the privacy risk attributed to friend relationships in Facebook. Privacy metrics, whose goal is to measure the degree of privacy enjoyed by users in a system and the amount of protection offered by privacy-enhancing technologies (Wagner and Eckhoff, 2018), have attracted the interest of several important studies. Liu and Terzi (2010), for instance, propose a framework to compute a privacy score measuring the user's potential privacy risk. This score increases with the sensitivity of information items and their visibility, e.g., the number of users knowing about each item, and leverages the item response theory (Keller and Schweid, 2011) as theoretical basis for the mathematical formulation of the score. Similarly, Wang et al. (2014) measure the user privacy exposure in a social network using a privacy index. Differently from Liu and Terzi (2010), however, this index requires predefined sensitivity values for users' items and requires the availability of user privacy settings. Several privacy scores are also proposed in a recent social network simulation game designed to help teachers educate school children on privacy issues (Bioglio et al., 2018). However, these scores are intended as a reward mechanism, rather than risk assessment measures.

The positioning of our work is, in fact, in the last branch of research. However, differently from the above mentioned papers, our proposal takes into account the context of the user subnetwork. In detail, we propose a privacy score that, compared to those presented in the closest related works (Liu and Terzi, 2010; Wang et al., 2014), takes into account the privacy attitude of the network as well as the centrality of the user.

## 3 Computing privacy scores

When measuring users' privacy in online social network, their preferences are only one side of the coin; in fact, user leakage risk is also affected by the context in which an individual is immersed. Besides users' own attitude on disclosing very private facts, the attitude of their friends towards privacy plays an important role too: users that likes or share friends' posts more often than the others, contribute most to the rapid spread of information (Bioglio and Pensa, 2017). Another factor influencing users risk is constituted by their position within the network (marginal users are certainly less exposed than very central users). In this section we present a score that quantify the privacy leakage of users considering the risks due not only to their attitude towards privacy




(a) Aware node in an unaware network (b) Unaware node in an aware network

(c) Contex-aware privacy scores for network (a) (d) Contex-aware privacy scores for network (b)

Fig. 1: Privacy risk and network-aware privacy score in two differently aware networks.

but also to the attitude of their subnetwork. Before addressing the technical details of our proposal and formalize the problem, we briefly introduce some basic necessary mathematical notation.

We consider a set of $n$ users participating in a social network, here represented as a directed graph $G(V, E)$, where $V$ is a set of $n$ vertices $\{v_1, \ldots, v_n\}$ such that each vertex $v_i \in V$ represents a user and $E$ is a set of directed edges $E = \{(v_i, v_j)\}$. Given a pair of users $v_i, v_j \in V$, $(v_i, v_j) \in E$ iif there exists a link from $v_i$ to $v_j$ (e.g., users $v_j$ is in the friend list/circle of $v_i$ or $v_i$ follows $v_j$). Each user is characterized by an *intrinsic privacy risk* $\rho_p(v_i)$, which is defined as the user propensity to privacy leakage. Two examples of social graphs with two different privacy risk values are given in Figure 1(a) and 1(b). The assumption is that some users are more prone to disclose their

personal data than others. This propensity is reflected in the way users configure their privacy settings. In the following, we first provide the definition of our network-aware privacy score independently from the specific choice of a reliable intrinsic privacy risk measure. Then, we instantiate $\rho_p(v_i)$ according to an established and reliable definition of privacy score (Liu and Terzi, 2010).

### 3.1 Network-aware privacy score

By definition, the *intrinsic privacy risk* $\rho_p(v_i)$ does not consider the context in which an individual is immersed. However, the actual privacy leakage risk of users is crucially affected by the properties of the social network they belong to: two users sharing the same attitude towards their own privacy protection are not necessarily subject to the same risk. If a user is mostly surrounded by friends that do not care that much about privacy, then she should be more exposed to privacy leakage than a user who is principally connected to friends that care about their own (and others') privacy. This consideration leads to the intuition that privacy risk in a social network may be modeled similarly as page authority in a hyperlink graph of web pages. Hence, we transpose the concept of "importance" of a web-page into the concept of "privacy risk" of users in a social network as follows: the more an individual is surrounded by friends that are careless about their privacy, the less the individual her/himself is likely to be protected from privacy leakage. One of the most popular algorithms to rank web pages based on their centrality (or authority) is Pagerank (Brin and Page, 1998). For a given directed graph $G(V, E)$, where $V$ is a set of $n$ vertices $\{v_1, \ldots, v_n\}$ and $E$ is a set of directed edges $E = \{(v_i, v_j)\}$, Pagerank is defined as the distribution that satisfies:

$$\boldsymbol{P} = d\boldsymbol{A}^\top \boldsymbol{P} + \frac{(1-d)}{n}\mathbf{1} \tag{1}$$

where $\boldsymbol{P} = [p(v_1), \ldots, p(v_n)]^\top$ is the Pagerank vector ($p(v_i)$ being the Pagerank associated to vertex $v_i$), $d = [0, 1]$ is the damping factor (the $1-d$ quantity is also known as restart probability), $\mathbf{1}$ is a vector of $n$ ones, and $\boldsymbol{A}$ is a $n \times n$ matrix such that each element $a_{ij} = 1/deg^+(v_i)$ ($deg^+(v_i)$ being the outdegree of $v_i$) if $(v_i, v_j) \in E$ ($a_{ij} = 0$ otherwise).

The computation of Pagerank values can be done using the well known power iteration method (Golub and van der Vorst, 2000), whose complexity is $O(I \times |E|)$, $I$ being the number of iterations (Bianchini et al., 2005). The algorithm is reported to converge quickly even for graphs containing millions of nodes (Brin and Page, 1998), thus the effective complexity is linear in the number of edges. Nevertheless, many research efforts have been devoted to speeding-up Pagerank computation (Chen et al., 2004; Kamvar et al., 2003; McSherry, 2005).

In our specific problem, each user $v_i \in V$ has an associated intrinsic privacy score $\rho_p(v_i)$. Consequently, instead of considering a uniform constant vector for the computation of Pagerank, we will use a non-uniform vector where the

component corresponding to node $v_i$ is equal to $\rho_p(v_i)/\sum_{k=1}^{n}\rho_p(v_k)$. This setting is similar to the definition of *personalized Pagerank* (Jeh and Widom, 2003), used to create a personalized view of the relative importance of the nodes. We can now introduce our network-aware privacy score (called *NetP-Score*), defined by the following distribution:

$$\boldsymbol{P} = d\boldsymbol{A}^{\top}\boldsymbol{P} + \frac{(1-d)}{\sum_{k=1}^{n}\rho_p(v_k)}\boldsymbol{\rho} \tag{2}$$

where $\boldsymbol{\rho} = [\rho_p(v_1), \ldots, \rho_p(v_n)]^{\top}$.

In those settings where the link between two users is always reciprocal (if there is a link from $v_i$ to $v_j$ then there is also a link from $v_j$ to $v_i$), the social network is represented as an undirected graph $G(V, E)$, where $E$ is such that if $(v_i, v_j) \in E$, then $(v_j, v_i) \in E$. In this case, $\boldsymbol{A}$ is symmetric and each element $a_{ij} = a_{ji} = 1/deg(v_i)$ ($deg(v_i)$ being the degree of $v_i$) if $(v_i, v_j) \in E$ ($a_{ij} = 0$ otherwise).

Equation 2 provides a set of values that are not necessarily in the same scale as the intrinsic risk. Hence, the final values of the privacy score, namely NetP-Score($v_i$), requires the execution of the following re-scaling operation:

$$\text{NetP-Score}(v_i) = p(v_i) \cdot \frac{range(\rho_p)}{range(p)} \tag{3}$$

where $p(v_i)$ is the network-aware privacy score value for node $v_i$ and NetP-Score($v_i$) denotes the recomputed privacy score value. Moreover, $range(\rho_p) = \max\{\rho_p(v_j)\} - \min\{\rho_p(v_j)\}$, $range(p) = \max\{p(v_j)\} - \min\{p(v_j)\}$ are the overall range of the values of the intrinsic risk and of the network-aware score respectively. In practice, the network-aware privacy score is adjusted so as to have the same range as the intrinsic risk.

An example of network-aware score computation is given in Figure 1. In Figure 1(a), we provide an example of graph where an aware user (the central one) is surrounded by unaware users (i.e., users with high intrinsic risk). Figure 1(c) represents the same network with the computed *NetP-Scores*: the score value of the central user is adjusted according to the network and it is sensibly higher than in Figure 1(a). Instead, in Figure 1(b), we provide a network with the same topology but different intrinsic risks. In particular the unaware central user (with high risk) is surrounded by rather aware users (with low privacy risk). In this case, our measure for the central user is revised downward (see Figure 1(d)), according to a context in which all other users form a kind of barrier protecting the privacy of the central users. In this little toy example we use an intermediate damping factor value ($d = 0.5$).

### *3.1.1 Choice of a reliable damping factor*

The choice of damping factor $d$ is not trivial in our setting. A common assumption in information retrieval is that $d$ should be set to 0.85 (Brin and Page, 1998), which gives much more importance to inbound links. However, in our

| User | Age | Job | Politics |
|---|---|---|---|
| $u_1$ | 2 | 2 | 0 |
| $u_2$ | 3 | 2 | 1 |
| $u_3$ | 3 | 3 | 1 |
| $u_4$ | 4 | 4 | 4 |

(a) A sample response matrix $\boldsymbol{R}$

| User | $\overline{\rho}_p$(Age) | $\overline{\rho}_p$(Job) | $\overline{\rho}_p$(Politics) | $\rho_p$ |
|---|---|---|---|---|
| $u_1$ | 0.0556 | 0.2222 | 0.0000 | 0.2778 |
| $u_2$ | 0.3611 | 0.3194 | 0.1111 | 0.7917 |
| $u_3$ | 0.6667 | 0.4167 | 0.1111 | 1.1944 |
| $u_4$ | 1.0000 | 1.0000 | 1.0000 | 3.0000 |

(b) Intrinsic privacy risks for the response matrix $\boldsymbol{R}$

Fig. 2: A response matrix of a small set of users and topics (left) and the resulting intrinsic privacy risk (right).

case, this may depend on the particular type of social network involved. For instance, some social networking platforms heavily promote sharing actions. In this case a high value of $d$ may provide a better estimate of the real privacy risk for users that, despite their restrictive privacy settings, are located in a relatively unsafe subnetwork. On the contrary, in those scenarios where sharing actions of users are not that visible to other users, small values of $d$ provide more realistic privacy leakage estimates.

### 3.2 Policy-based intrinsic privacy risk

We have defined the *intrinsic privacy risk* as the user propensity to privacy leakage. Assuming that users' activity in a social network is known, measuring their intrinsic privacy risk is not trivial. Our choice is based on the privacy score defined by Liu and Terzi (2010). Each user in $V$ may disclose information related to a set of $m$ topics $T = \{t_1, \ldots, t_m\}$, corresponding, for instance, to personal aspects such as religion, workplace, political views, health status, birthplace, gender, age, vacations and so on[2]. An $n \times m$ response matrix $\boldsymbol{R}$ is associated to the set of $n$ users $V$ and the set of $m$ topics $T$. Each element $r_{ij}$ of $\boldsymbol{R}$ corresponds to a privacy degree encoding the willingness of user $v_i$ to make information associated with topic $t_j$ visible to other social network users. Here we adopt the multinomial case, where entries in $\boldsymbol{R}$ take any non-negative integer values in $\{0, 1, \ldots, \ell\}$, where $r_{ij} = h$ (with $h \in \{0, 1, \ldots, \ell\}$) means that user $v_i$ discloses information related to topic $t_j$ to users that are at most $h$ links away in the social network $G$ (e.g., if $r_{ij} = 0$ user $u_i$ wants to keep information about $t_j$ private, if $r_{ij} = 1$ user $v_i$ is willing to make information related to $t_j$ available to all friends, if $r_{ij} = 2$ user $v_i$ is willing to let the friends of her or his friends access information about $t_j$, and so on). An example of response matrix for four users and three topics is given in Figure 2(a).

Thanks to the response matrix $\boldsymbol{R}$, the two main components of the privacy score can be computed, namely, the sensitivity $\sigma_j$ of a topic $t_j$, and the visibility $v_{ij}$ of a topic $t_j$ due to $v_i$. Liu and Terzi (2010) use a mathematical model based on polytomous item response theory (a well known theory in psy-

[2] In this work, we refer to $T$ as a fixed set of user-decided topics/aspects. It is out of the scope of this paper to consider automatic topic/aspect inference of each user's action or posted item.

chometrics, (Keller and Schweid, 2011)) to compute sensitivity and visibility. Differently from Liu and Terzi (2010), we adopt a simpler but still effective formulation that, additionally, is computationally less expensive.

According to this formulation, for any visibility degree $h = \{1, \ldots, \ell - 1\}$, sensitivity is calculated as follows:

$$\sigma_{jh} = \frac{1}{2}\left(\frac{n - \sum_{i=1}^{n} \mathbb{1}_{(r_{ij} \geq h)}}{n} + \frac{n - \sum_{i=1}^{n} \mathbb{1}_{(r_{ij} \geq h+1)}}{n}\right) \tag{4}$$

where $\mathbb{1}_A$ is the function returning 1 when condition $A$ is true and 0 when $A$ is false. When $h$ equals one of the two extreme values ($h = 0$ or $h = \ell$), the sensitivity values are computed differently. In detail,

$$\sigma_{j0} = \frac{n - \sum_{i=1}^{n} \mathbb{1}_{(r_{ij} \geq 1)}}{n} \tag{5}$$

for $h = 0$, and

$$\sigma_{j\ell} = \frac{n - \sum_{i=1}^{n} \mathbb{1}_{(r_{ij} \geq \ell)}}{n} \tag{6}$$

when $h = \ell$.

Equations 4, 5 and 6 have the following meaning: the less users adopt at least privacy degree $h$ for topic $t_j$, the more sensitive $t_j$ is w.r.t. degree $h$. Instead, when $h$ takes intermediate values (i.e., $h = \{1, \ldots, \ell - 1\}$), sensitivity is computed according to both degrees $h$ and $h + 1$. This guarantees that $\sigma_{j0} < \sigma_{j1} < \ldots < \sigma_{j\ell}$.

*Example 3 (Sensitivity computation)* Given the response matrix depicted in Figure 2(a) and $\ell = 4$, the sensitivity values for the topic *Job* are computed as follows:

- $\sigma_{\text{Job0}} = \frac{4-4}{4} = 0$ (all users have set the visibility level to at least 1),
- $\sigma_{\text{Job1}} = \frac{1}{2}\left(\frac{4-4}{4} + \frac{4-4}{4}\right) = 0$ (all users have set the visibility to at least 2),
- $\sigma_{\text{Job2}} = \frac{1}{2}\left(\frac{4-4}{4} + \frac{4-2}{4}\right) = \frac{1}{4}$ (only users $u_3$ and $u_4$ have set the visibility level to at least 3),
- $\sigma_{\text{Job3}} = \frac{1}{2}\left(\frac{4-2}{4} + \frac{4-1}{4}\right) = \frac{5}{8}$ (only users $u_3$ and $u_4$ have set the visibility level to at least 3, and $u_4$ is the only user having set the visibility to 4),
- $\sigma_{\text{Job4}} = \frac{4-1}{4} = \frac{3}{4}$ ($u_4$ is the only user having set the visibility level to 4).

In the simplified formulation, the visibility, for any degree $h = \{0, \ldots, \ell\}$ is calculated as follows:

$$v_{ijh} = Pr(r_{ij} = h) \times h \tag{7}$$

where $Pr(r_{ij} = h)$ is the probability that $r_{ij}$ equals $h$. Under the assumption that topics and users are mutually independent, probability $Pr(r_{ij} = h)$ can be computed using the following formula:

$$Pr(r_{ij} = h) = \frac{\sum_{i=1}^{n} \mathbb{1}_{(r_{ij}=h)}}{n} \times \frac{\sum_{j=1}^{m} \mathbb{1}_{(r_{ij}=h)}}{m} \tag{8}$$

The intuitive interpretation of Equation 7 is that visibility $v_{ijh}$ is higher when the sensitivity of topic $t_j$ is low and user $v_j$ has a low attitude towards her own privacy protection, regardless of the topic.

*Example 4 (Visibility computation)* Given the response matrix depicted in Figure 2(a) and $\ell = 4$, the visibility values $v_{2\text{Job}h}$ for the topic *Job* and user $u_2$ are computed as follows:

- $v_{2\text{Job}0} = \frac{0}{4} \cdot \frac{0}{3} \cdot 0 = 0$,
- $v_{2\text{Job}1} = \frac{0}{4} \cdot \frac{1}{3} \cdot 1 = 0$,
- $v_{2\text{Job}2} = \frac{2}{4} \cdot \frac{1}{3} \cdot 2 = \frac{1}{3}$,
- $v_{2\text{Job}3} = \frac{1}{4} \cdot \frac{1}{3} \cdot 3 = \frac{1}{4}$,
- $v_{2\text{Job}4} = \frac{1}{4} \cdot \frac{0}{3} \cdot 4 = 0$.

To compute the intrinsic privacy risk $\overline{\rho}_p(v_i, t_j)$ for a given user $v_i$ and a given topic $t_j$, we use the following formula:

$$\overline{\rho}_p(v_i, t_j) = \frac{\rho_p(v_i, t_j)}{\max\limits_{v_k \in V} \rho_p(v_k, t_j)} \tag{9}$$

where

$$\rho_p(v_i, t_j) = \sum_{h=0}^{\ell} \sigma_{jh} \times v_{ijh}. \tag{10}$$

and $\max\limits_{v_k \in V} \rho_p(v_k, t_j)$ is the maximum value of Equation 10 among all users. Although normalization is not strictly required, we use it to unify the scale of the intrinsic risk.

Finally, the overall intrinsic privacy risk $\rho_p(v_i)$ for any given user $v_i$ can be computed as follows:

$$\rho_p(v_i) = \sum_{j=1}^{m} \overline{\rho}_p(v_i, t_j). \tag{11}$$

The intuitive interpretation of Equation 9, 10 and 11 is as follows: $\rho_p(v_i) = 0$ means that either the topic $t_j$ is not sensitive at all (i.e., in each element of the summation, $\sigma_{jh} = 0$), or the information on topic $t_j$ is kept private (i.e., in each element of the summation, $v_{ijh} = 0$). Conversely, the privacy risk is maximum when a user $v_i$ makes all sensitive information ($\sigma_{jh} = 1$) visible

to all her or his friends ($v_{ijh} = 1$). As a consequence, users that have the tendency to make information about sensitive topics visible to a larger public are more susceptible to privacy leakage.

*Example 5 (Intrinsic risk computation)* Given the response matrix depicted in Figure 2(a) and $\ell = 4$, the (non normalized) intrinsic risk $\rho_p(u_2, \text{Job})$ for the topic *Job* and user $u_2$ is computed as follows:

$$\rho_p(u_2, \text{Job}) = 0 \cdot 0 + 0 \cdot 0 + \frac{1}{4} \cdot \frac{1}{3} + \frac{5}{8} \cdot \frac{1}{4} + \frac{3}{4} \cdot 0 = 0.2396.$$

The values of the intrinsic privacy risk for all users and topics is given in Figure 2(b), together with the overall intrinsic privacy risks (rightmost column).

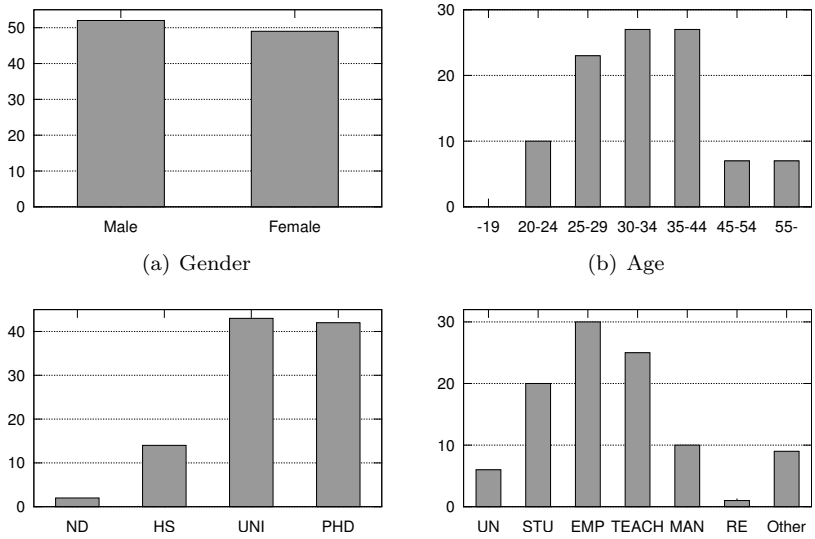
### 3.3 Theoretical complexity

Here we investigate the theoretical time complexity for computing our network-aware privacy score in realistic scenarios. Let $n$ be the total number of users in the social network and $m$ the overall number of topics. Computing the response matrix $\boldsymbol{R}$ requires $O(n \times m)$ operation. Hence, the computation of the intrinsic privacy risk for all users requires $O(n \times m \times \ell)$ operations for obtaining all sensitivity values $\sigma_{jh}$ ($\ell$ being the overall number of visibility degrees in $\boldsymbol{R}$), the same cost for obtaining all visibility values $v_{ijh}$ as well as the final value of the risk. The overall time complexity for computing the privacy risk values for all users and topics is then in $O(n \times m \times \ell)$. However, it is straightforward to suppose that, in a real-world scenario, $\ell << m$ and $m << n$. According to these realistic assumptions, $n$ dominates all other terms and the overall time complexity of the intrinsic privacy risk computation is in $O(n)$. We recall that the power iteration method, needed to compute the values of the network-aware privacy score requires $O(I \times |E|)$ operations (where $I$ is the number of iterations and $|E|$ is the total number of edges in the network). We can easily assume that $n << |E|$ (for instance, the number of edges in the Facebook social network is 95 times the number of nodes (Ugander et al., 2011), although, according to a more recent survey[3], the ratio is even larger) and $I << |E|$ (Brin and Page, 1998), so we conclude that the overall time complexity for computing all network-aware privacy score in a social network with $|E|$ friendship links is in $O(|E|)$.

## 4 Experimental results

In this section we report and discuss the results of the experiments that we conducted on two simulated networks and a Facebook graph generated from the ego-networks of real Facebook users. The main objectives of our experiments are: i) to analyze the relationship between users' attitude towards privacy

[3] `http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/`




(a) Gender (b) Age

(c) Education (d) Job position

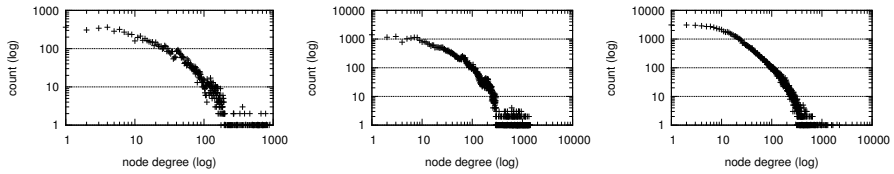
Fig. 3: Demographic statistics of the volunteers that participated in our online experiment. In the graph concerning education, ND stands for "No diploma", HS stands for "High school", UNI for "University/College degree", and PHD for "Ph.D. or other postgraduate degree". In the graph representing job positions, UN stands for "Unemployed", STU for "Student", TEACH for "Professor/Teacher/Researcher", MAN for "Manager/Professional", and RE stands for "Retired". The $y$-axis represents the number of respondents.

self-protection and the value of the network-aware privacy score; ii) to show the relationship between users' privacy score and their centrality in the social network; iii) to study the relationship between the scores and the effects of information propagation in the network.

The section is organized as follows: first, we describe the data and how we gathered them; then we analyze the behavior of our network-aware privacy score; finally, we study the relationships between information propagation and users' privacy. All experiments are performed on a server equipped with 2 Intel Xeon E5-2643 quad-core CPU's, 128GB RAM, running Arch Linux (kernel release: 4.5.1).

### 4.1 Datasets

In our experiments we use two simulated networks and a snapshot of the Facebook graph consisting on the ego-networks of real Facebook users. The two simulated networks (**SN10K** and **SN50K**) are generated using LDBC–




(a) SN10K (b) SN50K (c) FB75K

Fig. 4: Degree distribution in the three networks

SNB Data Generator[4] which produces graphs that mimic the characteristics of real social networks (Erling et al., 2015). The two graphs have 10,000 and 50,000 nodes respectively and are generated using the default configuration which tries to model the degree distribution as that observed in Facebook.

The third network (**FB75K**) is a snapshot of the real Facebook graph that has been generated leveraging an online experiment. This experiment was conducted as follows. We promoted an online experiment aimed at obtaining the required data to infer an approximate distribution of the privacy risk depending on the node degree. The online experiment was conducted in two phases. In the first phase we promoted the web page of the experiment[5] where people could voluntarily grant us access to their friends network. The participants were perfectly aware about the data we asked for and the purpose of our experiment. Moreover, all data were collected, stored and processed according to all EU regulations in force at the time[6]. In this first phase, data were gathered through version 1.0 of Facebook Graph API, used in a Facebook application written in Java. During spring 2015, we collected the anonymous ego-networks of 185 volunteers. We also asked them for some demographic data. From the related statistics reported in Figure 3 it turns out that there is not an emergent category of participants, even though, in general, they are highly-educated. The social network of all participants plus their friends consists of 75,193 nodes and 1,377,672 edges. Although the overall social network has been obtained by merging the participants' anonymous ego networks, the largest connected component contains 73,050 nodes (i.e., 97.15% of the overall social graph) and 1,333,276 edges (i.e., 96.78% of the number of edges in the overall graph). To achieve this goal, we foster the virality or our Facebook application by allowing it to propose the publication of a special post inviting all the participants' friends to join the experiment. Note that, once collected, the nodes identities were entirely replaced by anonymous IDs.

Some network statistics (number of edges and nodes, average clustering coefficient, average degree) about the datasets are reported in Table 1, while

[4] `https://github.com/ldbc/ldbc_snb_datagen`

[5] `http://kdd.di.unito.it/privacyawareness`

[6] The data collection/storage and processing protocols have been approved by the Law Office of our institution.

Table 1: Some datasets characteristics

| Dataset | #nodes | #edges | Avg. deg. | Avg. CC |
|---|---|---|---|---|
| SN10K | 9,226 | 183,004 | 39.671 | 0.160 |
| SN50K | 42,969 | 1,233,281 | 57.403 | 0.113 |
| FB75K | 75,193 | 1,377,672 | 36.644 | 0.613 |

Figure 4 presents their degree distribution. All graphs used in the experiments are considered as undirected.

### 4.2 Intrinsic risk estimation

The intrinsic privacy risk defined in Section 3.2 requires the availability of the privacy policy settings decided by each user or, alternatively, it can be inferred by the actual social activity of the users. However, this information is often missing (some social networking platforms implements very basic privacy policies), unreliable (e.g., some users often set privacy settings lazily) or simply not available (not provided by social platform API's). This is a major issue for security and complex network scientists as well: the privacy settings of users strongly influences the way information is propagated across the network. Moreover, knowing the attitude of users toward privacy is crucial to understand whether (and to whom) a user is willing to share a post/link. Last but not the least, privacy attitude is the first factor contributing to determining the privacy risk of users.

As a first experiment, we try to estimate the intrinsic privacy risk using the only local information often largely available in social networks: the degree of a user node (i.e., the number of friends in Facebook or followers in asymmetric social networks such as Twitter or Instagram). It is worth noting that there is no need to access the entire social network in order to know the degree of a node. To this purpose all the participants in the online experiments described in the previous section were contacted for an online survey. The participants had to indicate to which extent they were willing to disclose five different topics (job, relationship status, vacations, political views, personal life). Possible answers were: visible to no one, to close friends only, to friends except acquaintances, to all friends, to all friends of friends, visible to everyone on Facebook. The topics were proposed in form of direct questions (see Table 2) with different degree of sensitivity. In winter 2015/2016, 101 out of 185 participants answered all questions of the survey. We used the answers to fill the $101 \times 5$ response matrix $\boldsymbol{R}$ (see Section 3.2). Entries in $\boldsymbol{R}$ take values in $\{0, \ldots, 5\}$, where $r_{ij} = 0$ means that participant $v_i$ does not want to disclose information about question $Q_j$, $r_{ij} = 1$ means that participant $v_i$ is willing to make information related to $Q_j$ available to close friends and so on. Finally, we computed the intrinsic privacy risk of all 101 participants according to Equation 11 described in Section 3.2. Consequently, for each of the 101 volunteers, we have the node degree and the true privacy score computed

Table 2: The five questions (and related topics) of our online survey

| Question | Question text | Topic |
|---|---|---|
| **Q1** | Which people would you like to tell that you have just changed job? | job |
| **Q2** | If your relationship status changed, which friends would you like to tell? | relationship status |
| **Q3** | After a nice holiday, which friends would you share your photos with? | vacations |
| **Q4** | With whom would you like to share a comment on current affairs/politics? | political views |
| **Q5** | With whom would you like to share your mood or something personal that happened to you? | personal life |

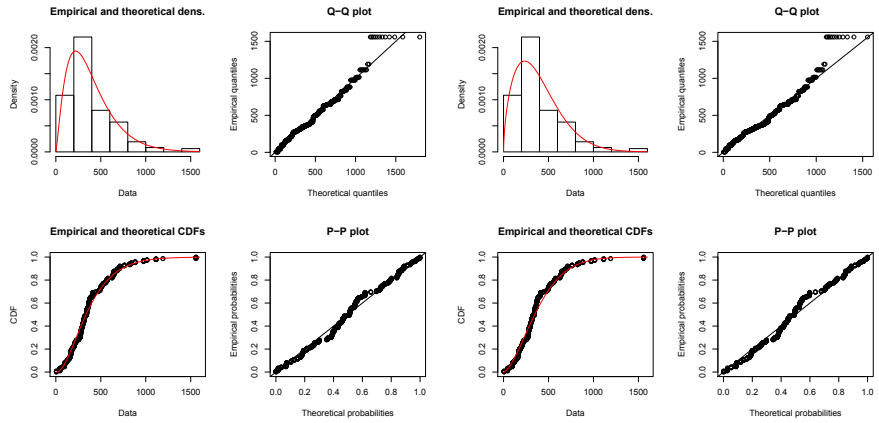
according to (Liu and Terzi, 2010). Since we want to model the behavior of the intrinsic risk according to the degree, we build graph such that the $x$-axis represents the node degrees, while the $y$-axis is proportional to the intrinsic risk associated to each node degree. Consequently, for each participant $v_i$, we associate a point $(x_i, y_i)$ in the graph, where $x_i = deg(v_i)$ and $y_i = \frac{\rho_p(v_i)}{\sum_{v_i} \rho_p(v_i)}$, where $\rho_p(v_i)$ is the intrinsic risk associated to user $v_i$ and computed according to Equation 11.

In order to infer the correct distribution function fitting the set of $(x_i, y_i)$ points, we analyze the Skewness-Kurtosis plot (Cullen and Frey, 1999) using the approach described in (Delignette-Muller and Dutang, 2015) on our sample. Skewness is a measure of symmetry, or more precisely, the lack of symmetry, while Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. The obtained plot suggests that the best two candidate distribution families are Gamma and Weibull. Hence, we use the maximum likelihood estimation method to fit our data to both Gamma and Weibull distribution and analyze the resulting empirical and theoretical PDF (probability distribution function), CDF (cumulative distribution function), P-P (Probability-Probability) and Q-Q (quantile-quantile) plots (see Figure 5). The two sets of plots are almost identical, but the Gamma distribution seems to fit slightly better our data than Weibull distribution. This is also confirmed by the values obtained by applying the Akaike information criterion (Akaike, 1974) (respectively 12982.52 and 12996.07 for Gamma and Weibull). Consequently, we retain the $\Gamma(k, \theta)$ (Gamma) distribution with the shape and scale parameters estimated by the maximum likelihood method ($k$=2.2989 and $\theta$=169.9461).

According to this choice, the estimated intrinsic privacy risk of a user $v_i$ is given by:

$$\gamma_p(v_i, k, \theta) = \gamma\left(deg(v_i); k, \theta\right) = \frac{1}{\Gamma(k)\theta^2} \times deg(v_i)^{k-1} \times e^{-\frac{deg(v_i)}{\theta}} \qquad (12)$$

where $deg(v_i)$ is the degree of node/user $v_i$, $\gamma(deg(v_i); k, \theta)$ is the probability density function, and $\Gamma(k)$ is the gamma function $\Gamma(z)$ for $z = k$. Since the mode of a generic distribution $\Gamma(k, \theta)$ is given by $(k-1)\theta$, the maximum value




(a) Gamma distribution plots (b) Weibull distribution plots

Fig. 5: Distribution plots showing the agreement between the real distribution and two possible estimations

is known. Hence, our measure (that we name $\Gamma$-Score) can be easily normalized and is given by

$$\Gamma\text{-Score}(v_i) = \overline{\gamma}_p(v_i, k, \theta) = \frac{\gamma_p(v_i, k, \theta)}{\gamma((k-1)\theta; k, \theta)} \tag{13}$$

Notice that, if the social graph were directed, one should take into account the indegree $deg^-(v_i)$ of each node $v_i$.

We can try to provide a practical explanation to the particular shape (given by parameter $k$) of the distribution (see the top-left plot in Figure 5(a)). The key of the interpretation comes from the typical user experience in social networks. Let us consider a user who has just joined a social platform. At the very beginning of her cybersocial experience, she will probably have very few followers/friends and a weak social activity (few published posts/pictures and other content). The more she adds new connections, the more she gets involved in the cybersocial environment and is eager to publish new (possibly sensitive) content, being more exposed to privacy leakage. However, when the number of friends becomes large, according to well established sociological and anthropological theories (Roberts et al., 2009; Dunbar, 2016), many links are likely to be weak (i.e., they represent sporadic online and offline interactions) and the user starts to be more conscious about the leakage risks concerning her and her friends' privacy. When the number of user's connection is very high (Facebook allows a maximum of 5,000 friends), privacy is unlikely to be an issue: people with such large ego-networks are popular personalities or celebrities that usually publish posts of general interest or like/share content from other popular social profiles.

Additionally, the particular estimation of parameter $\theta$ may also be explained intuitively. First, notice that lower values of $\theta$ result in lower values of the mode and a shrunk bell shape[7], i.e., the maximum value of the $\Gamma$-Score is reached with a low number of friends. Conversely, for higher values of $\theta$, the maximum value of the $\Gamma$-Score is reached with a higher number of friends, and the bell is larger. In our experimental study, we obtain $\theta = 169.9461$, and it can be observed that this value is not far from 150, a salient and well-known number in sociology. It corresponds to the Dunbar's number (Roberts et al., 2009), i.e., the cognitive limit to the number of individuals with whom one can maintain stable social relationships. Even more interestingly, our $\theta$ is very close to half the average number of Facebook friends, which is reported to be 338 for adult users[8]. These findings seem to confirm that our estimation is founded on solid bases, despite its simplicity.

#### *4.2.1 Effects of the biased sampling*

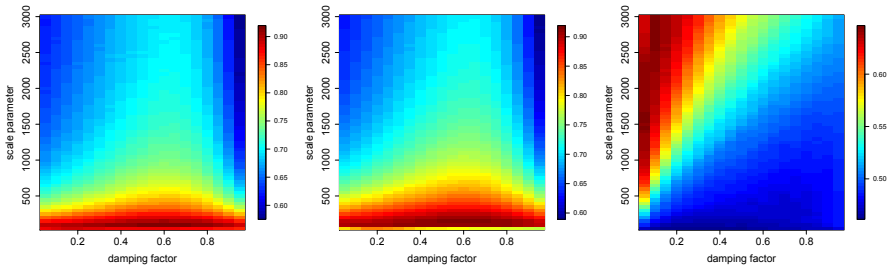
As a results of our particular recruitment campaign, our sample is biased towards highly educated people. In fact, from Figure 3(c) it turns out that most respondents own a University diploma or a higher degree. To assess the impact of this bias on our experiment, we first measure the mean and standard deviation of the intrinsic privacy risk (computed according to Equation 11) in each of the following three sub-populations: people owning at most a high school diploma (HS), people with a University or College degree (UNI), and people owning a Ph.D. or another postgraduate degree (PHD). The computed means show slight differences among the three groups: $2.6110 \pm 0.6069$ for the HS group, $2.1970 \pm 0.8482$ for the UNI group, and $2.3466 \pm 0.8710$ for the PHD group. Thus, we observe that there exists no proportionality between education degree and intrinsic privacy risk. In addition, to assess the statistical significance of these differences, we perform an unpaired two-tailed $t$-test for unequal sample sizes. We used the Benjamini–Hochberg correction procedure to controls the false discovery rate (Benjamini and Hochberg, 1995). The results clearly indicate ($p > 0.1$ in all tests) that the null hypothesis that the P-Scores are drawn from the same distribution cannot be rejected: the differences are not statistically significant. Consequently, we can reasonably conclude that the level of education is unlikely to constitute a major bias in our experiment.

### 4.3 Results on simulated networks

In Section 3.1 we have introduced our score that measures the privacy risk of users according to the characteristics of their subnetworks. Here, we investigate experimentally to what extent it is a good estimate of the objective privacy risk of the users. To this purpose, we conducted several experiments involving the

[7] The typical shape of the Gamma distribution is a skewed bell.

[8] `http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/`




(a) SN10K (b) SN50K (c) FB75K

Fig. 6: Quality index $\mathcal{I}$ based on Spearman's correlation computed once between the network-aware privacy score and the intrinsic privacy risk and once between the network-aware privacy score and the eigenvector centrality of nodes.

two simulated networks (SN10K and SN50K) as follows. First we compute the intrinsic privacy risk of each node using the $\Gamma$-Score (see Section 4.2) according to the following strategy. First, the intrinsic risk of each node $v_i$ is generated randomly from a Gaussian distribution $N(\mu, \sigma^2)$ with $\mu = \overline{\gamma}_p(v_i, k, \theta)$ and $\sigma^2 = 1.0$, where $\theta \in [50, 3000]$ and $k = 2.2989$ (as computed in Section 4.2). Then, for each experimental setting we compute the network-aware privacy score using the power-iteration method (Golub and van der Vorst, 2000) to solve Equation 2 (see Section 3.1). The number of iterations is set to 100. We repeat the experiments for varying values of the damping factor $d$ in the interval $[0.05, 0.95]$. We then measured the Spearman's rank correlation coefficient (Spearman, 1904) between the intrinsic privacy risk and the network-aware privacy score (**NetP-Score**).

The Spearman's rank coefficient assesses monotonic relationships between two series of values. Given a set of $n$ objects $X = x_i \dots x_n$ and two functions $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$, the Spearman's coefficient is computed as:

$$\rho = 1 - 6 \times \frac{\sum_{i=1}^{n} \left(rank_f(x_i) - rank_g(x_i)\right)^2}{n(n^2 - 1)} \tag{14}$$

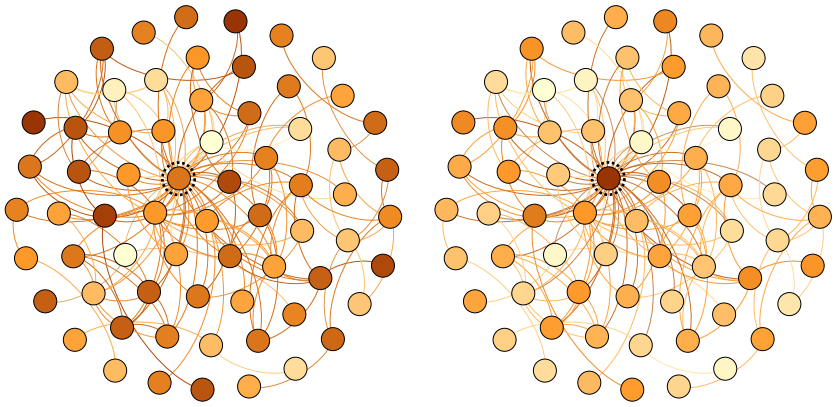
where $rank_f(x_i)$ and $rank_g(x_i)$ are the rank of object $x_i$ in the two series of function values computed for $X$. It measures the correlation between the two rankings of the same set of objects and its values range between $-1$ (when the rankings are maximally inversely correlated) and $+1$ (for the maximum positive rank correlation). The significance of the rank correlation can be assessed by verifying whether the null hypothesis (i.e., that $\rho$ is not significantly different from zero) can be rejected. Since the quantity $t = \rho\sqrt{(n-2)/(1-\rho^2)}$ is distributed approximately as the Student t-distribution with $n - 2$ degrees of freedom, the null hypothesis can be verified by performing the well-known two-tailed t-test.

We denote the Spearman $\rho$ between the intrinsic privacy risk and the network-aware privacy score with $\rho_s$. To achieve significant results, we run each experiment 30 times. Moreover, to investigate the ability of considering the network topology in the computation of the privacy score, we measured the Spearman's correlation between the network-aware privacy score and the *eigenvector centrality* (Newman, 2010) of each node. We use $\rho_c$ to denote this measure. Finally, we compute a quality index taking into account both $\rho_s$ and $\rho_c$ as $\mathcal{I} = [(1+\rho_s)(1+\rho_c)]/4$. This index takes values in $[0,1]$ and is maximum (resp. minimum) when both Spearman coefficient are equal to 1 (resp. -1). The goal of this index is to identify a range of parameters values enabling the computation of network-aware privacy scores that exhibit a strong correlation with both the intrinsic privacy risk and the centrality of the nodes.

The average results are reported in Figure 6(a) and 6(b). Noticeably, the $\mathcal{I}$ index values are always high ($\mathcal{I} > 0.55$), even though in the region defined by the two intervals $0.5 \leq d \leq 0.7$ and $100 \leq \theta \leq 300$ the quality index reach its maximum values (around 0.90, meaning that **NetP-Score**'s are strongly positively correlated with both the intrinsic risk and the eigenvector centrality). It is worth noting that in these experiments the p-values are always very small ($p < 10^{-15}$), thus indicating that the results are statistically significant.

### 4.4 Results on the Facebook graph

We set up a slightly different experiment for the social network retrieved from Facebook (FB75K). The main difference is that, instead of computing the $\Gamma$-Score for all nodes, for the nodes corresponding to the participants in our survey we use the P-Score (Liu and Terzi, 2010) computed from the response matrix $\boldsymbol{R}$ obtained by processing their answers. All other $75,193 - 101$ nodes are handled as described in Section 4.3. The remainder of the experimental setup is the same as described in Section 4.3. The results are reported in Figure 6(c), but, differently from the previous setting, the Spearman's coefficient is computed only on the set of 101 participants. Contrary to the simulated setting, in this case the network-aware privacy score (NetP-Score) exhibit slightly smaller quality index w.r.t. the P-Score. It can be observed that, overall, the quality index $\mathcal{I}$ ranges between 0.45 and 0.65. In detail, the average Spearman's correlation between the intrinsic privacy risk and the network-aware privacy score is 0.1292. This result probably means that, for some users, the privacy score defined by Liu and Terzi (2010) is not always a good estimate of their objective privacy risk: the low value of the Spearman's coefficient shows that there is a gap between the privacy leakage risk computed by only leveraging users' privacy preferences and the real privacy risk which takes into account the weakness of the network surrounding them. On the other hand, the network-aware privacy scores computed on the 101 participants are always positively correlated with their eigenvector centrality, despite their intrinsic risk. These results confirm our initial claim: to measure the objective privacy



(a) Intrinsic risk (P-Score) (b) Network-aware score (NetP-Score)

Fig. 7: A snapshot of a small portion of the Facebook graph (FB75K) consisting of a connected component with 67 participants and their computed intrinsic risk (left) and network-aware privacy scores (right). Darker nodes have higher scores (best viewed in color version).

risk, any privacy metric should be contextualized within the social graph by considering its influence on each user.

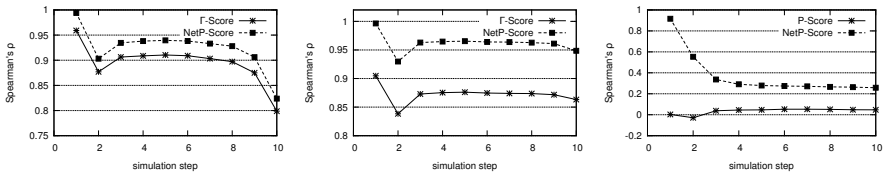
### *4.4.1 Visual inspection of the Facebook graph*

In addition to the previous analysis, we also report here an in-depth visual inspection of a small portion of the Facebook graph (FB75K). To obtain this portion, we first extract the subgraph consisting of all participants in the experiment described in Section 4.1 that are directly connected to at least another participant. 74 out of 101 participants are selected in this phase. Then, we consider the subgraph induced by these 74 nodes, which, in its turn, consists of 163 undirected edges, with an average degree of 4.405 and 4 connected components. By using the intrinsic privacy risk (P-Score) obtained as specified in Section 4.2, we compute the network-aware privacy score (NetP-Score) using Equation 3. Note that, since we only consider a small portion of the network, the results are different from those obtained in Section 4.4, which are computed on a more realistic snapshot. However, by recomputing the scores for this small subgraph, we are able to show some interesting insights. In Figure 7 we report the largest connected component of this subgraph, consisting

of 67 nodes and 159 edges. In Figure 7(a) nodes are colored according to their intrinsic risk (darker nodes corresponds to higher scores). In Figure 7(b), instead, node darkness is proportional to our network-aware privacy score. It is interesting to notice that, due to the low degree of almost all nodes and their rather "peripheral" position in the network, the NetP-Scores are lower than the P-Scores. There are only few exceptions, the most evident of which is represented by the highlighted node in the two pictures: it consists of a very central node (its degree is 43 and its eigenvector centrality is the maximum among all 67 nodes), surrounded by several nodes that exhibit a higher intrinsic score. Although real social networks are more complex than that presented in Figure 7, the outcomes of this visual analysis confirm the importance of considering the impact of the overall network on individual nodes.

### 4.5 Reliability of the privacy scores

As a concluding experiment, we study the relationship between the different privacy score definitions and the effects of information propagation across the network. A good privacy score should take into account the amount of nodes that may potentially access and diffuse some information coming from other nodes in the same network. For this reason, we perform several Monte Carlo simulations of an information propagation scenario within the two synthetic networks (SN10K and SN50K) and our snapshot of Facebook (FB75K). In particular, we adopted the Susceptible-Infectious-Recovered (SIR) epidemic model, a well-studied model that describes the transmission of a disease through a population. At each step an individual may be susceptible (S) to the disease, infectious (I) or recovered (R), that is immune to the epidemic. An infectious individual may infect a susceptible one with an infection probability $\beta$, and convert him into an infectious individual, or recover from the disease with a recovery probability $\gamma$, becoming recovered. The SIR model has been also applied for modeling the diffusion of information in social networks (Gruhl et al., 2004). In our experiments, for all nodes we set an infection probability $\beta = 0.5$ and a recovery probability $\gamma = 0.3$. Then, we select $N$ seed nodes that, in turn, are considered as the individuals that start the infection (i.e., information diffusion process) and measure the number of nodes (called *prevalence rate*) that are either infected (I) or recovered (R) after each step of the simulation. For datasets SN10K and SN50K we select $N = 100$ random nodes, while for FB75K the seed nodes are the 101 Facebook users that participated in our online experiment. Finally, for each simulation step we compute the Spearman's $\rho$ coefficient between the prevalence rate and the privacy scores ($\Gamma$-Score and NetP-Score for SN10K and SN50K, P-Score (Liu and Terzi, 2010) and NetP-Score for FB75K). Parameters $d$ and $\theta$ for the simulated networks are set according to the best results obtained in the previous experiments (see Section 4.3), while for the Facebook networks $\theta$ is set according to Section 4.2 (the damping factor is the same as for the simulated networks). The results are reported in Figure 8 (parameter settings are given in the captions).




(a) SN10K ($d$=0.6, $\theta$=100) (b) SN50K ($d$=0.6, $\theta$=150) (c) FB75K ($d$=0.6, $\theta$=170)

Fig. 8: Spearman's correlation computed once between the intrinsic privacy risk (P-Score or $\Gamma$-Score) and the total number of infected nodes (prevalence) and once between the network-aware privacy score (NetP-Score) and the total number of infected nodes (prevalence). In all the experiments $p < 0.005$ (worst value) except for the P-Score (Liu and Terzi, 2010) in FB75K ($p > 0.3$).

As expected, our NetP-Score shows a better privacy leakage estimation than the $\Gamma$-Score (see Figure 8(a) and 8(b)). However, the most interesting result concerns the Facebook network (Figure 8(c)): in this case, in fact, the gap between the P-Score's $\rho$ and our network-aware score's $\rho$ in the very first iterations is significantly large. Assuming that the P-Score correctly measures the privacy risk based on users' privacy preferences, a possible explanation is that the users underestimate their centrality within the network. Undoubtedly, by construction, our Facebook snapshot cannot be considered a statistically valid sample of the entire Facebook graph, but the huge difference in terms of correlation with the prevalence rate confirms that privacy leakage metrics should not ignore the context in which the users operate within the social network.

## 5 Conclusions and future work

With the final goal of supporting users' privacy awareness in online social networks, we have proposed a context-aware definition of privacy score, inspired by Pagerank. This measure, as shown in our experiments, is a good estimate of the objective privacy risk of the users. Moreover we have also inferred experimentally a new intrinsic privacy risk score that estimates well the real user attitude towards privacy. The results highlight the necessity of incorporating privacy measure computation within any domain-specific or general-purpose social media and networking platforms. Additionally, the low computational requirements of our measure would not affect the responsiveness of social platforms, and can be of inspiration for the design of privacy-enhancing social networking components, in compliance with the principles of *Privacy by Design* (Cavoukian, 2012).

As future work we plan to better define our intrinsic privacy score by conducting an extensive experimental campaign involving more online social network users. Furthermore, the policy-based definition that we adopted to compute the intrinsic privacy risk is limited to a well-defined set of topics or aspects (e.g. work status, photo albums, relationship status). However, topic

detection techniques can be applied to natural language posts or pictures to understand their contents, as proposed by Song et al. (2018). A further refinement of this work, will consists in directly inferring the sensitivity of posted items by leveraging topic modeling, natural language processing techniques, and text categorization algorithms. Finally, an interesting research question deserving further investigation is whether, in general, aware users are mostly surrounded by aware users and vice-versa. Although some effects related to *homophily* — the theory according to which similar nodes may be more likely to attach to each other than dissimilar ones (McPherson et al., 2001) — may exist, the role of privacy attitude in determining social network ties is probably involved in more complex phenomena.

**Acknowledgements** This work is supported by Fondazione CRT (grant numbers 2015-1638 and 2017-2323). The authors wish to thank the anonymous reviewers for their valuable comments and all the volunteers who participated in the survey.

## References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19(6):716–723

Akcora CG, Carminati B, Ferrari E (2012a) Privacy in social networks: How risky is your social graph? In: Proceedings of IEEE ICDE 2012, IEEE Computer Society, pp 9–19

Akcora CG, Carminati B, Ferrari E (2012b) Risks of friendships on social networks. In: Proceedings of IEEE ICDM 2012, IEEE Computer Society, pp 810–815

Becker J, Chen H (2009) Measuring privacy risk in online social networks. In: Proceedings of Web 2.0 Security and Privacy (W2SP) 2009

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological) 57(1):289–300

Bianchini M, Gori M, Scarselli F (2005) Inside pagerank. ACM Trans Internet Techn 5(1):92–128

Bioglio L, Pensa RG (2017) Impact of neighbors on the privacy of individuals in online social networks. In: Proceedings of the International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland, Elsevier, Procedia Computer Science, vol 108, pp 28–37

Bioglio L, Capecchi S, Peiretti F, Sayed D, Torasso A, Pensa RG (2018) A Social Network Simulation Game to Raise Awareness of Privacy among School Children. IEEE Transactions on Learning Technologies pp 1–14, available online

Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. Computer Networks 30(1-7):107–117

Buccafurri F, Fotia L, Lax G, Saraswat V (2016) Analysis-preserving protection of user privacy against information leakage of social-network likes. Inf Sci 328:340–358

Cavoukian A (2012) Privacy by design [leading edge]. IEEE Technol Soc Mag 31(4):18–19

Cetto A, Netter M, Pernul G, Richthammer C, Riesner M, Roth C, Sänger J (2014) Friend inspector: A serious game to enhance privacy awareness in social networks. In: Proceedings of IDGEI 2014

Chen Y, Gan Q, Suel T (2004) Local methods for estimating pagerank values. In: Proceedings of ACM CIKM 2004, ACM, pp 381–389

Choi H, Park J, Jung Y (2018) The role of privacy fatigue in online privacy behavior. Computers in Human Behavior 81:42–51

Cormode G, Srivastava D, Bhagat S, Krishnamurthy B (2009) Class-based graph anonymization for social network data. PVLDB 2(1):766–777

Cullen AC, Frey HC (1999) Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press, New York

Delignette-Muller ML, Dutang C (2015) fitdistrplus: An R package for fitting distributions. Journal of Statistical Software 64(4):1–34, URL `http://www.jstatsoft.org/v64/i04/`

Dunbar RIM (2016) Do online social media cut through the constraints that limit the size of offline social networks? Royal Society Open Science 3(1)

Erling O, Averbuch A, Larriba-Pey J, Chafi H, Gubichev A, Prat-Pérez A, Pham M, Boncz PA (2015) The LDBC social network benchmark: Interactive workload. In: Proceedings of ACM SIGMOD 2015, ACM, pp 619–630

Fang L, LeFevre K (2010) Privacy wizards for social networking sites. In: Proceedings of WWW 2010, ACM, pp 351–360

Furini M, Tamanini V (2015) Location privacy and public metadata in social media platforms: attitudes, behaviors and opinions. Multimedia Tools Appl 74(21):9795–9825

Golub GH, van der Vorst HA (2000) Eigenvalue computation in the 20th century. Journal of Computational and Applied Mathematics 123(1–2):35–65

González RJ (2017) Hacking the citizenry?: Personality profiling, "big data" and the election of donald trump. Anthropology Today 33(3):9–12

Gruhl D, Liben-Nowell D, Guha RV, Tomkins A (2004) Information diffusion through blogspace. SIGKDD Explorations 6(2):43–52

Hay M, Miklau G, Jensen D, Towsley DF, Weis P (2008) Resisting structural re-identification in anonymized social networks. PVLDB 1(1):102–114

Hay M, Li C, Miklau G, Jensen D (2009) Accurate estimation of the degree distribution of private networks. In: Proceedings of ICDM 2009, IEEE, pp 169–178

Jeh G, Widom J (2003) Scaling personalized web search. In: Proceedings of WWW 2003, ACM, pp 271–279

Kamvar SD, Haveliwala TH, Manning CD, Golub GH (2003) Extrapolation methods for accelerating pagerank computations. In: Proceedings of WWW

2003, ACM, pp 261–270
Keller LA, Schweid JA (2011) Handbook of polytomous item response theory models edited by michael l. nering and remo ostini. Journal of Educational Measurement 48(1):98–100
Kosinski M, Stillwell D, Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. PNAS 110(15):5802–5805
Litt E (2013) Understanding social network site users' privacy tool use. Computers in Human Behavior 29(4):1649–1656
Liu K, Terzi E (2008) Towards identity anonymization on graphs. In: Proceedings of ACM SIGMOD 2008, ACM, pp 93–106
Liu K, Terzi E (2010) A framework for computing the privacy scores of users in online social networks. TKDD 5(1):6:1–6:30
Liu Y, Gummadi PK, Krishnamurthy B, Mislove A (2011) Analyzing facebook privacy settings: user expectations vs. reality. In: Proceedings of ACM SIGCOMM IMC '11, ACM, pp 61–70
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology 27(1):415–444
McSherry F (2005) A uniform approach to accelerated pagerank computation. In: Proceedings of WWW 2005, ACM, pp 575–582
Misra G, Such JM (2016) How socially aware are social media privacy controls? IEEE Computer 49(3):96–99
Newman M (2010) Networks: An Introduction. Oxford University Press, Inc., New York, NY, USA
Pensa RG, Blasi GD (2017) A privacy self-assessment framework for online social networks. Expert Syst Appl 86:18–31
Rathore S, Sharma PK, Loia V, Jeong YS, Park JH (2017) Social network security: Issues, challenges, threats, and solutions. Information Sciences 421:43 – 69
Roberts SGB, Dunbar RIM, Pollet TV, Kuppens T (2009) Exploring variation in active network size: Constraints and ego characteristics. Social Networks 31(2):138–146
Song X, Wang X, Nie L, He X, Chen Z, Liu W (2018) A personal privacy preserving framework: I let you know who can see what. In: Proceedings of ACM SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018, ACM, pp 295–304
Spearman C (1904) The proof and measurement of association between two things. The American Journal of Psychology 15(1):72–101
Squicciarini AC, Paci F, Sundareswaran S (2014) Prima: a comprehensive approach to privacy protection in social network sites. Annales des Télécommunications 69(1-2):21–36
Such JM, Criado N (2016) Resolving multi-party privacy conflicts in social media. IEEE Trans Knowl Data Eng 28(7):1851–1863
Such JM, Rovatsos M (2016) Privacy policy negotiation in social media. TAAS 11(1):4:1–4:29

Talukder N, Ouzzani M, Elmagarmid AK, Elmeleegy H, Yakout M (2010) Privometer: Privacy protection in social networks. In: Proceedings of M3SN'10, IEEE, pp 266–269

Task C, Clifton C (2012) A guide to differential privacy theory in social network analysis. In: Proceedings of ASONAM 2012, IEEE, pp 411–417

Ugander J, Karrer B, Backstrom L, Marlow C (2011) The anatomy of the facebook social graph. CoRR abs/1111.4503

Vuokko N, Terzi E (2010) Reconstructing randomized social networks. In: Proceedings of SIAM SDM 2010, SIAM, pp 49–59

Wagner I, Eckhoff D (2018) Technical privacy metrics: A systematic survey. ACM Comput Surv 51(3):57:1–57:38

Wang Y, Nepali RK, Nikolai J (2014) Social network privacy measurement and simulation. In: Proceedings of ICNC 2014, IEEE, pp 802–806

Xu K, Guo Y, Guo L, Fang Y, Li X (2017) My privacy my decision: Control of photo sharing on online social networks. IEEE Trans Dependable Sec Comput 14(2):199–210

Ying X, Wu X (2011) On link privacy in randomizing social networks. Knowl Inf Syst 28(3):645–663

Yu J, Zhang B, Kuang Z, Lin D, Fan J (2017) iprivacy: Image privacy protection by identifying sensitive objects via deep multi-task learning. IEEE Trans Information Forensics and Security 12(5):1005–1016

Zheleva E, Getoor L (2011) Privacy in social networks: A survey. In: Social Network Data Analytics, Springer US, pp 277–306

Zhou B, Pei J (2011) The *k*-anonymity and *l*-diversity approaches for privacy preservation in social networks against neighborhood attacks. Knowl Inf Syst 28(1):47–77

Zou L, Chen L, Özsu MT (2009) K-automorphism: A general framework for privacy preserving network publication. PVLDB 2(1):946–957