Alemany-Bordera, J.; Del Val Noguera, E.; Alberola Oltra, JM.; García-Fornes, A. (2018). Estimation of privacy risk through centrality metrics. Future Generation Computer Systems. 82:63-76. https://doi.org/10.1016/j.future.2017.12.030

Estimation of privacy risk through centrality metrics

J. Alemany, E. del Val, J. Alberola, A. García-Fornes[1]

{*jalemany1,edelval,jalberola,agarcia*}*@dsic.upv.es*
Universidad Politècnica de València,
Camino de Vera s/n, Valencia (Spain)

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.

Keywords: Privacy, Social Networks, Information Sharing

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, 4, 5]. Some works try to address these issues with the automation of privacy settings [6, 7, 8, 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, 11, 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, 15, 16, 17]. Several works have studied spreading dynamics and influential or relevant individuals in these processes based on structural properties [18, 19, 20, 21, 22]. From the point of view of determining the privacy risk associated to a user's sharing action, 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, 27, 28, 29]

have been used to detect these relevant users in networks [30, 21, 31].

Not all users have the same perception of risk [32, 33, 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 labelling 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 reconfigure 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

	Type of information			User intervention	Privacy risk estimation
	profile items	actions			
		audience	reachability		
Liu et al. [3]	✓				✓
Nepali et al. [4]	✓				✓
Shehab et al. [5]	✓				
Fang et al. [6]		✓		✓	
Vidyalakshmi et al. [7]				✓	✓
Bilogrevic et al. [8]		✓		✓	
Calikli et al. [10]		✓			
Kafali et al. [11]		✓		✓	
Mester et al. [12]				✓	
Yang et al. [36]					✓
Our work			✓		✓

Table 1: Overview of approaches related to privacy in social networks. We considered three main features: (i) the type of information considered to evaluate the user's privacy risk (i.e., the user's profile items or actions). In the case that the approach considers actions, the goal can be to determine if the information shared was received by the intended audience or to estimate the reachability of the information; (ii) if the approach requires user intervention as input for the privacy risk estimation; and (iii) if the approach provides a privacy risk metric to the 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, 38, 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.

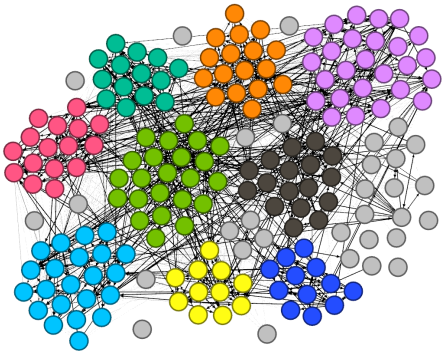
In this paper, we deal with this problem with the proposal of a Privacy Risk Score (PRS) that is focused on the risk of potential re-sharing actions from the expected and unexpected audience that might receive the message. The main contributions of this work are the following: (i) the proposed PRS metric considers the paths that information follows as a result of sharing actions without the user's intervention; (ii) the calculation of the PRS metric for different users' risk perceptions; (iii) we provide and evaluate a set of centrality metrics to estimate PRS values in scenarios where there is a lack of a global view of the network and/or data about the users' sharing activity.

3. Privacy risk scenario

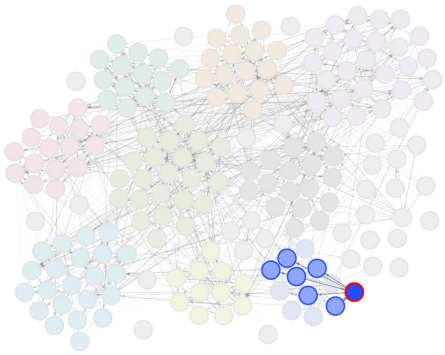
Privacy risk not only concerns the problem that information might reach people who were initially not expected to receive it, but it also involves the problem of losing control over the scope of the information. In Figure 1, we describe this privacy risk problem in online social networks.

The social network is structured into nine communities (see Figure 1a). Nodes represent users and the node color corresponds to a community. Gray nodes represent isolated users (i.e., they do not belong to any community). In Figure 1b, the user represented by the node encircled in red shares a message on his/her wall. The user determines the audience depending on his/her selected privacy policy (e.g., *friends*). Therefore, only their friends can see the message (see Figure 1c, nodes encircled in green). If a node encircled in green performs a sharing action, the message could reach other communities causing a privacy problem.

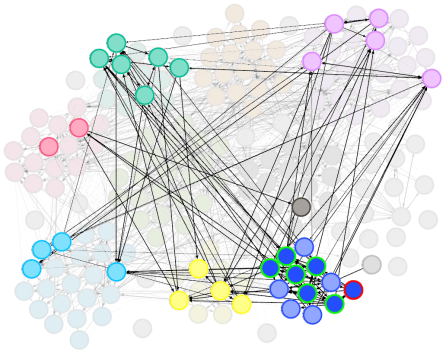
The Privacy Risk Score metric proposed in this paper deals with this problem by providing information about the potential privacy risk of an action. The PRS aims to increase the users' awareness about the reachability of their publications in the social network even though they have restricted the visibility of their publications. Figure 2 shows the workflow phases for calculating the PRS. First, the activity in the social network is monitorized (specifically, the path followed



(a) A social network structured into communities.



(b) Sharing action initiated by the node encircled in red.



(c) Potential audience in level 2.

Figure 1: Example of a potential privacy risk in online social networks.

by user messages). This information is used to establish the reachability of the actions performed by each user and to calculate the PRS value. Then, when a user is going to post a message, the PRS values analyzed until that moment are shown to the user. The PRS of a user would provide him/her with an estimation of the visibility of an action at different levels of friendship or in general. By taking into account their privacy risk perception and their PRS, users could make better decisions about sharing or not sharing a message on their walls.

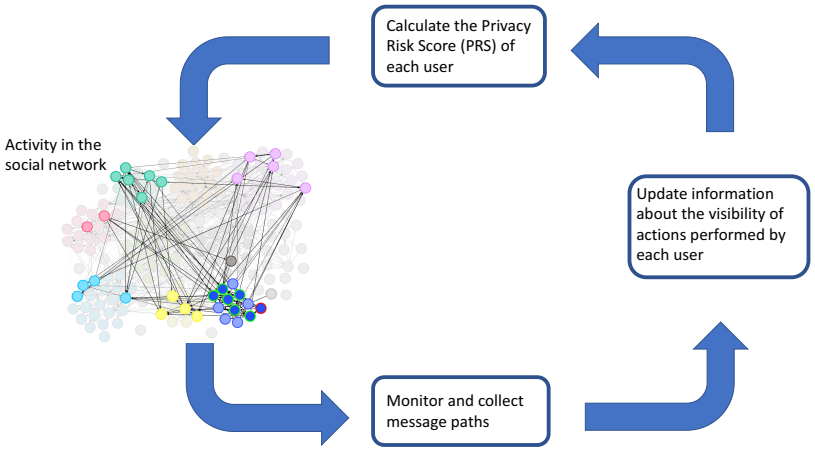


Figure 2: Flow chart of the phases for calculating the PRS in a social network.

4. Privacy Risk Score (PRS)

To define how our proposed PRS metric works, first we are going to explain some important concepts. We assume that there is a social network $\mathcal{G}$ that consists of N nodes, where every node $a_i \in \{a_1, ..., a_n\}$ represents an agent (i.e., a user of the social network). Agents are connected through links that represent friendship relationships and correspond to the edges $E \subseteq N \times N$ of $\mathcal{G}$. We assume that friendship links are bidirectional, and, therefore, the social network is undirected. We define the adjacency matrix $\mathbf{A}$ to represent these links. Given two agents a_i and a_j, if there is a link between these agents, we represent this as $\mathbf{A}_{a_i,a_j} = 1$ and $\mathbf{A}_{a_i,a_j} = 0$ if there is not a link. Considering an agent a_i, we define a level L as the subset of agents whose shortest distance to a_i is l:

$$L_{a_i}(l) \subseteq N, \forall a_j \in L_{a_i}(l) : d(a_i, a_j) = l \wedge \nexists d'(a_i, a_j) < d(a_i, a_j)$$

We define the Privacy Risk Score (PRS) for an agent a_i that performs a message diffusion action (i.e., publishes a message m on its wall, comments on an existing post, shares a post, etc.) as an indicator of the potential risk of this

message to be diffused over the social network (i.e., potential visibility). The higher the PRS value, the higher the threat to agent a_i's privacy.

4.1. Calculation of the PRS metric

In a social network $\mathcal{G}$, there is a set of paths that messages follow more frequently than others [45, 46]. If an agent is in these paths and performs a diffusion action, it has a higher privacy risk than another agent that is out of these paths. Therefore, an agent's position in the network is relevant to the privacy risk. Furthermore, not all users have the same view of risk when sharing information. As an example, some users may consider that sharing information with friends of friends might be risky, while others may consider that the true risk is at the next level of friendship. Therefore, the estimation of the PRS for an agent a_i should be provided in friendship levels in order to deal with different levels of risk perception.

In addition, according to the information diffusion model SIR (Susceptible, Infected, and Removed) [19], the time instant in which a diffusion action of a message is performed is also important for measuring the privacy risk. This model states that the privacy risk related to the diffusion of a message is higher during the initial stages than when the message has already been diffused through the social network. In other words, the diffusion risk of a message is higher when an agent diffuses a new message since no other agents have viewed it yet. Therefore, the calculation of the PRS also includes the stage of the message in which an agent a_i interacts as a diffusion action. To represent this, we define $T = \{1, 2, \ldots, n\}$ as the stages of the message, which are the product of the diffusion process of the message. This variable is represented for each message and indicates the number of steps from its creation. The value of the variable T (and also of the variable L) is limited by the network diameter. Therefore, if its value is not too high, the network diameter is a good approximation of T and L. For the sake of simplicity, we assume that an agent can carry out a single message diffusion action (i.e., re-share a message, comment on a message, etc.), allowing other agents to see this message at that time instant.

Considering the above two factors (friendship level and risk of initial stages), we define a $T \times N$ reachability matrix $\gamma_{\mathbf{i}}$ associated to each agent a_i to represent the number of messages that an agent a_i has diffused in a certain stage t and have been seen by other agents. The rows of this matrix represent the diffusion actions that a_i carries out over messages in the same stage, while columns represent the agents of the social network. We use $\gamma_{i_{t,a_j}}$ to refer to the entry in the tth row and a_jth column of $\gamma_{\mathbf{i}}$. This value represents the number of messages diffused by a_i in stage t that were seen by a_j. Note that the a_ith column of each row t ($\gamma_{i_{t,a_i}}$) represents the messages diffused by a_i in stage t that were seen by a_i (i.e., all of the messages published by a_i in t).

Given a stage t and a set of agents of level l, we define $p(a_i, t, l)$ as the average number of agents of this level that saw a message published by a_i in stage t:

$$p(a_i, t, l) = \frac{\sum_{a_j \in L_{a_i}(l)} \gamma_{i_{t,a_j}}}{\gamma_{i_{t,a_i}}} \tag{1}$$

Taking into account the above value, we estimate the PRS for an agent a_i at level l as the percentage of agents of that level that potentially see a message published by a_i at any stage. This can be calculated as:

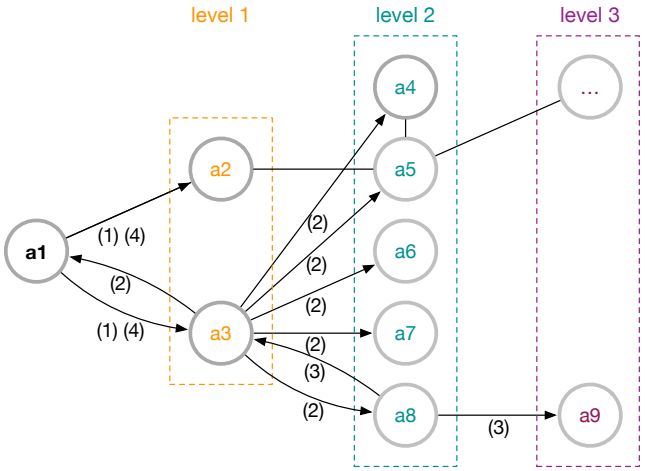
$$PRS(a_i, l) = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{p(a_i, t, l)}{|L_{a_i}(l)|} \right) \tag{2}$$

In a general view, by taking into account the whole population of the social network $\mathcal{G}$, we can estimate a general value of PRS for an agent a_i as the percentage of agents of the social network that potentially see a message published by a_i at any stage. This can be calculated by combining Equations 1 and 2:

$$PRS(a_i) = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\sum_{a_j \in N} \gamma_{i_{t,a_j}}}{\gamma_{i_{t,a_i}} \cdot |N|} \right) \tag{3}$$

Figure 3 shows a scenario where the privacy risk score is calculated for agent a_1 in a social network. This scenario represents an example of a social network with interactions between agents. We assume that all of the agents in $\mathcal{G}$ have the

privacy policy that only their direct friends can see their walls. As indicated in the definition above, the maximum value for parameters T and L cannot exceed the network diameter. Therefore, for this example of PRS calculation, we use the value 3 for parameters T and L.



– PRS metric of agent a_1:

$p(a_1, t = 1, l = 1) = 4/2 = 2$
$p(a_1, t = 1, l = 2) = 5/2$
$p(a_1, t = 1, l = 3) = 1/2$

$PRS(a_1, l = 1) = 1/3 * 2/2 = 1/3$
$PRS(a_1, l = 2) = 1/3 * (5/2)/5 = 1/6$
$PRS(a_1, l = 3) = 1/3 * (1/2)/1 = 1/6$

$PRS(a_1) = 1/3 * 10/(2 * 9) = 5/27$

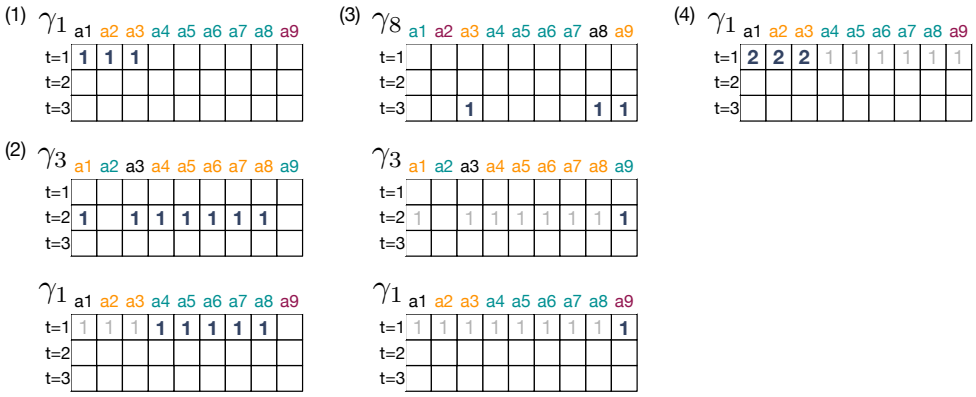


Figure 3: Example of social network activity and the PRS calculation process. The activities carried out on the social network are as follows (in this example, all agents share information with their friends): (1) agent a_1 publishes/shares a message m_1 on its wall; (2) agent a_3 shares the message m_1; (3) agent a_8 shares the message m_1; and (4) agent a_1 publishes/shares a new message m_2.

The message diffusion actions performed in this scenario are the following. (1) Agent a_1 publishes a message m_1 on its wall. Therefore, agents a_2 and a_3 can see the message. Since the interaction of agent a_1 with the message m_1 is in its initial stage, stage t is 1. The information about the agents that can see m_1 is stored in γ_1. (2) Agent a_3 then decides to share m_1 on its wall. Agents a_4, a_5, a_6, a_7, and a_8 can see message m_1. As in the previous case, the information about the agents that can see message m_1 is updated in γ_3. The interaction of agent a_3 with message m_1 occurs after agent a_1 shares it (i.e., the interaction is produced in the next stage $t = 2$). Note that the values of γ_1 are updated at $t = 1$ because agent a_1 interacts with the message in this stage, and in γ_1 we are measuring the reachability of the messages when agent a_1 interacts with it. (3) Agent a_8 then shares m_1 publishing it on its wall. Agents a_3 and a_9 can see it. Therefore, γ_8 is updated at $t = 3$, and γ_1 and γ_3 are updated in their corresponding t's (i.e., $t = 1$ and $t = 2$). (4) Agent a_1 then publishes a new message m_2 that agents a_2 and a_3 can see at stage $t = 1$. Then, γ_1 is updated at $t = 1$.

With the information stored in the γ matrix, the proposed PRS is calculated for each agent. In the scenario described in Figure 3, we show the values of PRS for agent a_1 at different levels (i.e., $PRS(a_1, l = 1)$, $PRS(a_1, l = 2)$, and $PRS(a_1, l = 3)$) and the general PRS value (i.e., $PRS(a_1)$).

4.2. PRS metric in OSN

The integration of the PRS metric in OSN must be done as a service for users. This privacy service will help users to manage their sensitive and non-sensitive information and aware its scope, improving their experiences in OSN. In Figure 4, we show a block diagram of OSN where the PRS metric was included as a service in the OSN platform layer. The diagram is composed of a User layer, OSN Platform layer, and Privacy Risk Module. The User layer manages users contacts, information related to the user (e.g., profile info, posts, comments, etc.), and setting parameters to control who has access to the information when a sharing action is carried out. The OSN Platform layer provides the whole

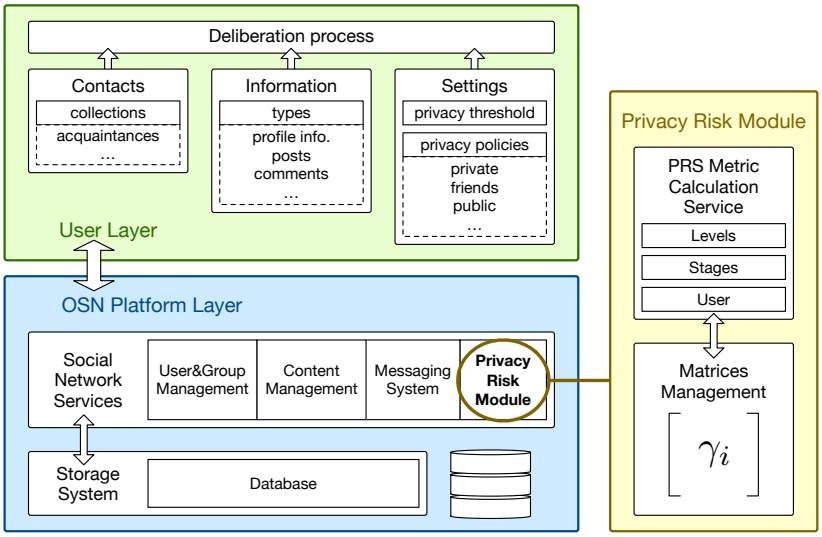


Figure 4: Block diagram of the integration of PRS metric as a service in OSN.

functionality of a OSN (e.g., management of users, messaging system, etc.). The Privacy Risk module is included as a service of the OSN Platform layer. This service is responsible for the PRS metric calculation.

Figure 5 shows the workflow to estimate the PRS value of an individual agent when he performs a message sharing action in the OSN. The process starts when an agent a_i sees a publication or when he creates content for a new publication (m_j). Then, this agent evaluates the risk of sharing/publishing m_j considering its PRS value ($PRS(a_i)$). If the value is greater than his individual risk threshold (θ_{a_i}), a_i does not perform the action. Otherwise, a_i shares m_j, which in turn, could be seen by other agents. In this case, the matrix γ_i of a_i is updated as well as the matrices of other agents that previously participated in the sharing process of m_j.

5. PRS and centrality metrics

Even though the PRS estimation provides accurate measurements of the privacy risk associated to a diffusion action, this estimation requires a detailed record of sharing activity in a social network. However, the management of this

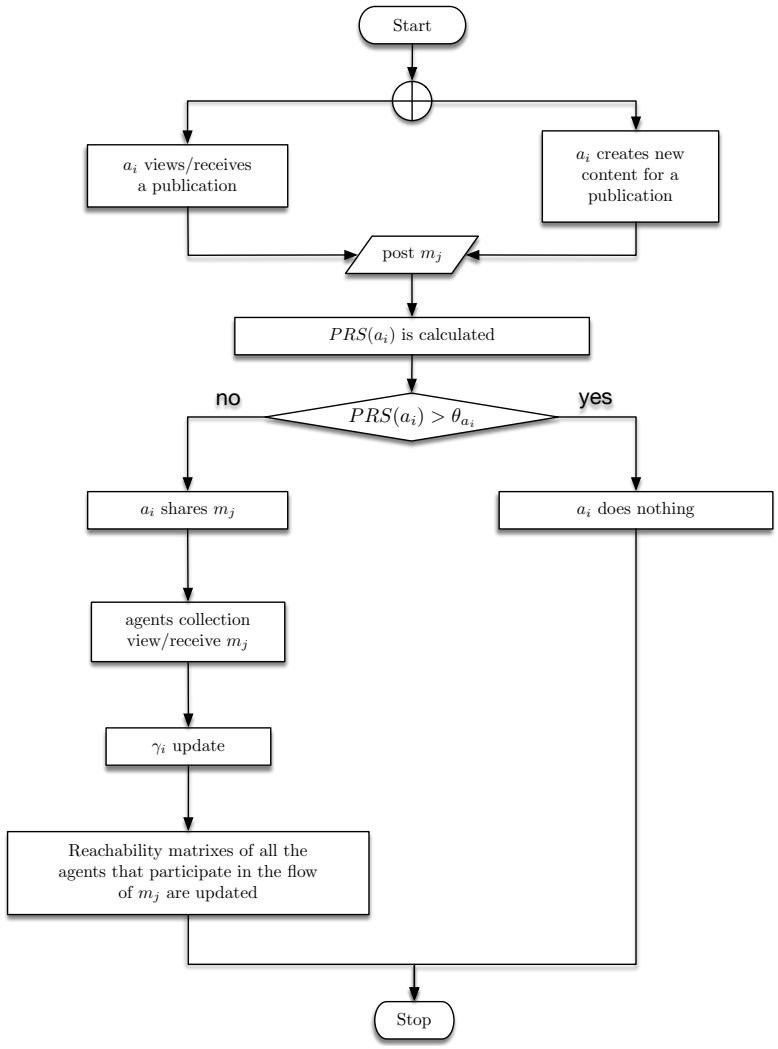


Figure 5: Flowchart of the PRS calculation process.

information is not always feasible in large networks with high activity, and, in some scenarios, this knowledge is not even accessible. As a result, in certain circumstances, we would require metrics that approximate PRS values in a feasible way.

Influential users may play a critical role in paths that information follows. If an influential user sees a publication and performs a sharing action, it is

more likely for the publication to reach more people. It is important to have a reliable and efficient predictor of these nodes based on topological properties. From the area of Complex Networks, there is no consensus on the best metric for predicting this influence. Researchers have proposed several structural metrics for identifying influential users [47]. According to the information they used, these metrics can be classified into three classes: global, local, and social [48].

Global metrics are based on structural properties that require a complete view of the network structure to be computed. Among the global metrics, we considered the following commonly used metrics: betweenness, closeness, and pagerank. Betweenness metrics are based on assumptions about the paths that information follows. Shortest-path betweenness assumes that information is transmitted along the shortest paths. It is defined as the fraction of the shortest paths between pairs of agents in a network that pass through the agent of interest [27],

$$\text{bet-sp}_i = \sum_{a_j, a_k \in N} \frac{\sigma(a_j, a_k|a_i)}{\sigma(a_j, a_k)}, \tag{4}$$

where $\sigma(a_j, a_k)$ is the number of shortest $(a_j,\ a_k)$-paths, and $\sigma(a_j, a_k|a_i)$ is the number of those paths passing through some node a_i other than a_j, a_k.

Random-walk betweenness was proposed by Newman [49], and, instead of considering the shortest paths, it considers the number of times a random walk between each pair of agents passes through the agent of interest. Thus, random-walk betweenness can be defined as follows,

$$\text{bet-rw}_i = \sum_{a_j, a_k \in N} \frac{\sigma_r(a_j, a_k|a_i)}{\sigma_r(a_j, a_k)}, \tag{5}$$

where $\sigma_r(a_j, a_k)$ is the number of random $(a_j,\ a_k)$-paths, and $\sigma_r(a_j, a_k|a_i)$ is the number of those random paths passing through some node a_i other than a_j, a_k.

While betweenness centrality measures represent the degree to which an agent is between pairs of other agents, closeness is just the inverse of the average distance to other agents. Closeness is defined as the mean geodesic distance from

the agent of interest to the rest of the reachable agents in the network,

$$\text{closeness}_i = \frac{|N| - 1}{\sum_{j=1}^{j=|N|-1} d(a_j, a_i)}, \tag{6}$$

where $d(a_j, a_i)$ is the shortest-path distance between a_j and a_i, and $|N|$ is the number of nodes in the network. This metric reflects the efficiency of an agent distributing information to any agent in the network [25].

PageRank is based on the idea that an agent has a high rank if the sum of the ranks of its neighbors are high. The ranks are calculated based on the structure of the links of the agent of interest. Then, pagerank centrality can be defined as follows,

$$\text{pagerank}_i = \alpha \sum_{j=1}^{j=|N|} \mathbf{A}_{a_j, a_i} \frac{\text{pagerank}_j}{k_j} + \beta, \tag{7}$$

where α and β are constants and k_j is the degree of node j. This metric implies a relatively low computational complexity and has been used to identify pivotal individuals in social networks who lead to quick and wide spreading of useful items [20].

Global metrics can be suitable to estimate the risk of a sharing action in the network since they capture the user's relevance in the transmission of information and do not require data about information flows. The computation of a global metric requires the analysis of structural properties that involve the consideration of the whole social network. However, in real-world scenarios, these metrics are not always computationally affordable and information about friendship relationships is not always accessible. Moreover, some social applications do not facilitate access to users' information to third party applications; therefore, it is not possible to infer the social network structure beyond the first level.

As an alternative, local and social metrics efficiently identify influential agents when there is no global information about network structure and information diffusion [50]. These metrics are focused on the user's ego networks.

Ego networks consist of a focal agent (*ego*) and the agents to whom the ego is directly connected to (these are called *alters*) plus the links [51]. Local metrics such as degree and ego-betweenness only use information from the agent itself to be computed. Degree is the simplest centrality measure and considers the number of direct neighbors (*alters*) that the ego is directly connected to,

$$\text{degree}_i = \sum \mathbf{A}_i(a_i, a_j). \tag{8}$$

Ego-betweenness is an ego-centric method for approximating the betweenness centrality [52]. This metric calculates the sum of the ego's proportion of times that the ego lies on the shortest path between each part of the alters. Ego-betweenness is the sum of the reciprocal values $\mathbf{A}_i^2(a_j, a_k)$ such that $\mathbf{A}_i(a_j, a_k) = 0$. Thus, ego-betweenness can be defined as follows,

$$\text{bet-ego}_i = \sum_{\mathbf{A}_i(a_i,a_j)=0, j>i} \frac{1}{\mathbf{A}_i^2(a_i, a_j)} \tag{9}$$

Social metrics use strictly local information and topological information from an agent's first and second level neighbors. Social degree and Social ego-betweenness metrics consider the sum of the local centrality metrics of neighbors in the first two levels. We have considered the following four social centrality metrics:

$$\text{bet-ego}_{\text{sum}_i} = \sum_{a_j \in L_{a_i}(1)} \text{bet-ego}_j \tag{10}$$

$$\text{degree}_{\text{sum}_i} = \sum_{a_j \in L_{a_i}(1)} \text{degree}_j \tag{11}$$

$$\text{bet-ego}_{2\text{sum}_i} = \sum_{a_j \in L_{a_i}(2)} \text{bet-ego}_j \tag{12}$$

$$\text{degree}_{2\text{sum}_i} = \sum_{a_j \in L_{a_i}(2)} \text{degree}_j \tag{13}$$

Centrality metrics provide mechanisms to estimate the relevance of users in information transmission processes. Influential users play a key role in information diffusion and therefore in the increase of the privacy risk if they perform a re-sharing action. For this reason, considering global, local, and social centrality metrics might be appropriate to estimate the proposed PRS when there is no data available about information flows. Global centrality metrics can be used if the network structure is known. If there is no access to this information, local and social centrality metrics based on ego-networks provide metrics to estimate the relevance of users in information transmission processes.

6. Experiments

In this section, we evaluate the relationship between PRS values of an agent and its centrality in the social network. The social networks considered for the experiments can be viewed in terms of the friendship relationships and the activities carried out by agents. We analyze the relationship between the structural features of the friendship layer and the privacy risk resulting from the diffusion actions. We perform a set of experiments in different synthetic and real networks. For the experiments in synthetic networks we use a simulation tool to reproduce information flows in the network, and the proposed PRS metric to measure the individual risk of users. While in real networks, how there are already real information flows, we only measure the PRS values of users.

6.1. Simulation environment

We based our simulation environment on the Elgg engine[1] (Figure 6). Elgg is a popular open source engine to build a wide range of social environments. For our purpose, we required to collect message tracing information and manage them in matricial structures in order to calculate the PRS metric. Therefore, we needed to extend the functionalities of Elgg in order to fulfill our requirements. Following the Elgg policy, we extended Social Network Services by means of

[1] `https://elgg.org/`

plug-ins. First, we developed the Privacy Risk Module following the structure shown in Figure 4, which is a plug-in for PRS calculation according to our requirements. This module was focused on two different purposes: for being used in simulations and with real users.

Second, we developed the Simulation Tool, which is a plug-in for modelling social networks and generating activity. The Simulation Tool was designed to use the services of the OSN (properly supported by Elgg) such as the creation of users and relationships, message sending, and social interactions. Users are represented as software agents that interact among them in the OSN. Agent-based simulation is widely used in different areas [53]. The Simulation Tool is composed of three main components: Input Parameters, Simulator Core and Outputs. As Input Parameters, the simulation tool allows the definition of the number of simulations, the network model, and the customization of agent behaviours (i.e., message diffusion actions, probabilities, deliberation process,

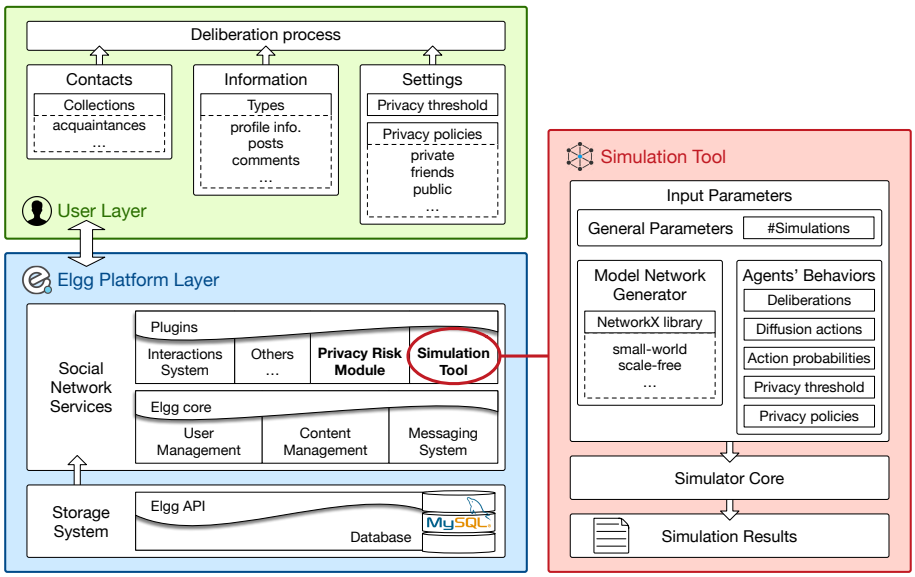


Figure 6: Block diagram of the integration of the Simulation Tool developed as a service in the OSN.

etc.). For modelling social network structures, we used the NetworkX[2], which is a widely tested and recommended library for research purposes in complex networks [54, 55, 56]. The Simulator Core carries out the simulation according to the input parameters. Finally, Simulation Results (i.e., Privacy Risk Score values of each agent) are stored for further analysis. These both plug-ins were integrated into the existing Elgg engine. Since this engine is open source, these plug-ins will be public available.

6.2. Settings

The experiments carried out using the simulation tool use synthetic networks generated follow three classic models: Erdös-Rényi [37] (ER, random), Barabási-Albert [40] (BA, scale-free), and Watts-Strogatz [42] (WS, small-world). The networks are undirected, have 1000 agents with a diameter of 5, and an average degree of about 12 (see Table 2). The number of simulations is 400 per each agent. In each simulation an agent is randomly selected and the simulation starts if the agent decides to post a message. Figure 7 shows the deliberation process of an agent during the simulation. Each agent decides whether or not a message diffusion action is carried out (i.e., commenting on an existing post, sharing a post, etc.) according to his probabilities of performing each action. If the agent decides to perform a diffusion action, then he selects the privacy policy for this message. In case that the message was previously received by this agent or if the agent decides not to carry out a message diffusion action, then, the message is not diffused by this agent. Each simulation finishes when there is not any message diffusion action in the OSN.

Simulation parameters are shown in Table 3. The #Simulation parameter allows to define the simulation rounds. Network topology parameter establishes the underlying social network structure (i.e., scale-free, random, small-world). Diffusion action parameter allows to define the permitted actions in the simulation (i.e., posting a message, sharing a message, commenting a post and liking

[2] `https://networkx.github.io`

	Random network	Scale-free network	Small-world network
Nodes	1000	1000	1000
Edges	6464	5875	6000
Density	0.01292	0.01175	0.012
Maximum degree	32	117	21
Minimum degree	3	5	7
Average degree	12.93	11.75	12.00
Assortativity	0.00077	-0.07481	-0.02096
Triangles	631	1022	1963
Diameter	5	5	5

Table 2: Structural properties of synthetic networks.

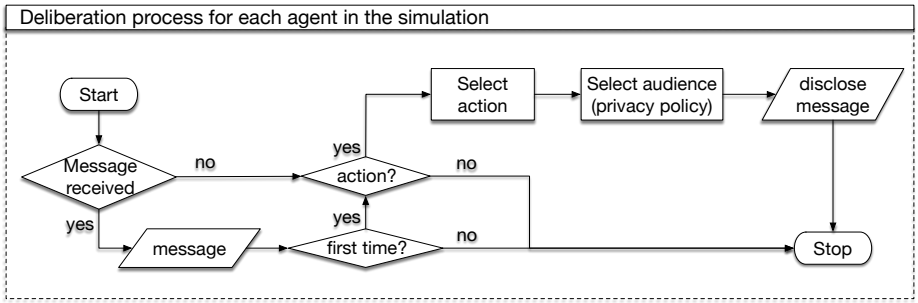


Figure 7: Flowchart of the agent deliberation process.

#Simulations	400×1000 (agents)
Network topology	{scale-free, random, small-world}
Diffusion action	{publish, share, comment, like}
Action probability	uniform
Privacy threshold	uniform
Privacy policy	friends

Table 3: Simulation parameters.

a post). Action probability parameter establishes the probability of an agent to perform an action. Privacy threshold parameter specifies the value from which an agent considers that an action is risky for him. Privacy policy parameter describes the audience of an agent action (i.e., friends).

Regarding real networks, we used PHEME dataset[3] that is based on the dynamics of the life cycle of social media rumours on Twitter [57]. The dataset contains 330 conversational threads. We analyzed the PRS and centrality values of the 330 users that initiated a thread through the publication of a message.

To evaluate the relationship between the PRS and structural centrality metrics in synthetic and real networks, we consider message stages from 1 to 4 and also relationship levels from 1 to 4. The reason for the number of relationship levels is based on the analysis presented in [13] where it is reported that most of the cascades in reality are small.

In the next subsections, calculations about agents' PRS based on information sharing activities are used to find a relation with centrality metrics. In this way, approximations using centrality metrics would allow us to calculate agent privacy risks in scenarios where there is no access to data about social interaction, when there is no previous activity, or when new users join the network.

6.3. PRS and global centrality metrics

In this section, we analyze whether or not there is a correlation between agents' PRS (i.e., dependent variable) and their global centrality (i.e., independent variable) in synthetic networks. Real networks were not considered in these experiments since the global structures of the rumor networks are not available. We considered the global centrality metrics described in Section 4: random-walk betweenness (bet-rw), shortest-path betweenness (bet-sp), closeness, and pagerank. The values of centrality properties are normalized in the range $[0, 1]$. We used analytical regressors to estimate the dependence relationship between centrality metrics and PRS. We considered the R^2 coefficient to determine how close the data are to the fitted regression line. In this case, values close to 1 indicate that there is a high correlation between centrality and PRS values.

Figure 8 displays the comparison between PRS and global centrality values.

[3] `https://figshare.com/articles/PHEME_rumour_scheme_dataset_journalism_use_case/2068650`

In 8, a centrality metric is analyzed in each row, and a network topology is considered in each column. The x axis shows the values of the agents' PRS and the y axis shows the values of the agents' centrality metrics. Colors represent the number of agents with certain values of PRS and centrality. The relationship between PRS and centrality metrics is also shown by the coefficient of determination (R^2). Due to the logarithmic behavior of the centrality metrics (especially in the case of the scale-free network), a *linear-log* filter was applied to all of the data.

First, the results reflect the variability of agents' PRS depending on the type of network. The scale-free BA networks (see Figure 8 – first column) favor higher values of PRS (close to 0.5). In contrast, in the small-world WS networks (Figure 8 – third column), PRS does not reach 0.3. It can also be observed that the type of network reflects the existence of different groups of agents based on their privacy risk. As an example, in the scale-free BA networks there is a small group of agents with high values of PRS (i.e., values that range in the interval $[0.3, 0.5]$), while the rest are distributed between 0.1 and 0.3. In the random ER networks, there is a majority group with relatively high values (i.e., values between 0.25 and 0.4) and a minority with very low values of PRS. In the small-world WS network, it can be observed that most of the agents have low PRS values (between 0.125 and 0.2) compared to other network topologies, and there are two minorities: one with slightly lower PRS values and another with slightly higher PRS values.

Second, there is a high correlation between global centrality metrics and the PRS values (see Figure 8). The R^2 value is around 0.9 in scale-free networks (Figure 8 – first column [`bet-sp`, `pagerank`]); 0.93 in random networks (Figure 8 – second column [`closeness`]); and 0.92 in small-world networks (Figure 8 – third column[`closeness`]). Thus, we can conclude that PRS values can be approximated through global centrality metrics in scenarios without data about information flows in the social network.

Table 4 shows the relationship between PRS and global centrality metrics for each level expressed as the R^2 coefficient. Level 1 (i.e., direct neighbors) is

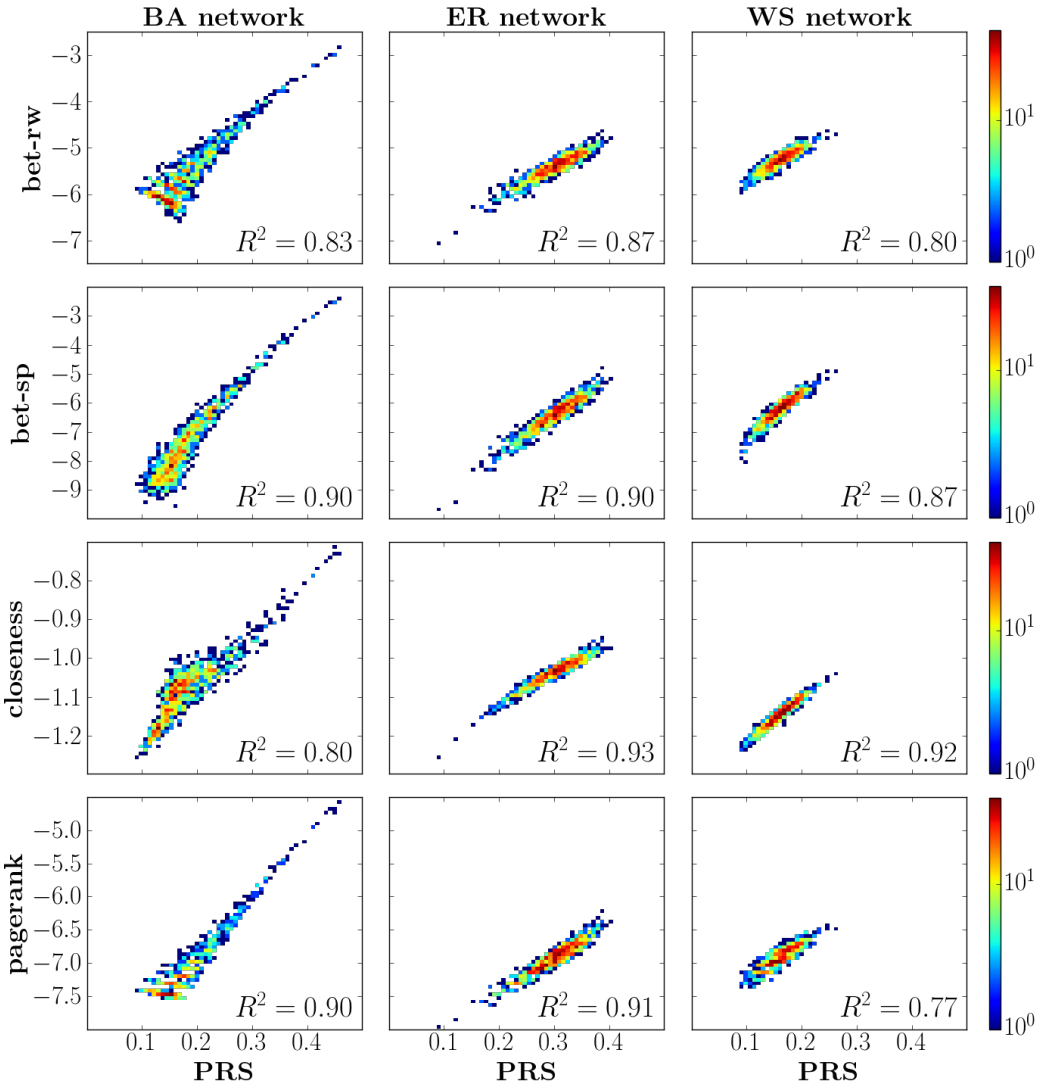


Figure 8: Correlation between global centrality metrics and PRS for different social network topologies.

not shown due to its irrelevance, since it corresponds to the agent that initiates the activity (i.e., publishes a message). As can be seen from the results, the R^2 coefficient generally decreases according to the depth of the target level, except for random ER network topology.

Network type	Level	R^2 score			
		closeness	pagerank	bet-sp	bet-rw
scale-free (BA)	2	0.82	0.79	0.75	0.66
	3	0.38	0.29	0.43	0.32
	4	0.82	0.25	0.40	0.18
random (ER)	2	0.93	0.90	0.88	0.83
	3	0.80	0.74	0.77	0.74
	4	0.95	0.82	0.84	0.78
small-world (WS)	2	0.93	0.73	0.87	0.78
	3	0.96	0.74	0.86	0.77
	4	0.78	0.49	0.59	0.50

Table 4: Evaluation of the relation between global centrality metrics and PRS by levels for different social network topologies.

As the results show, the estimation of PRS using global centrality metrics yields promising results. However, as we stated in the previous section, global centrality metrics present several limitations: their calculation requires a knowledge of the whole network structure, and they suffer from performance issues in large networks. Moreover, a recalculation is needed when the network structure changes (i.e., when a new agent joins/leaves the network or a relationship is created/removed). Taking into account these challenges in calculating global centrality metrics, we examine local and social centrality metrics in the following subsection.

6.4. PRS, local, and social centrality metrics

In this section, we evaluate the relationship between local and social centrality and PRS values in synthetic and real social networks. First, we analyze degree centrality and the ego-betweenness centrality [52] (i.e., a local approximation of the betweenness centrality metric). Second, we analyze social degree and social ego-betweenness centrality. These experiments have the same settings considered in previous experiments (subsection 6.2).

Figure 9 shows the results of the linear-log regression analysis to determine if there is a relationship between local centrality and PRS values. Although

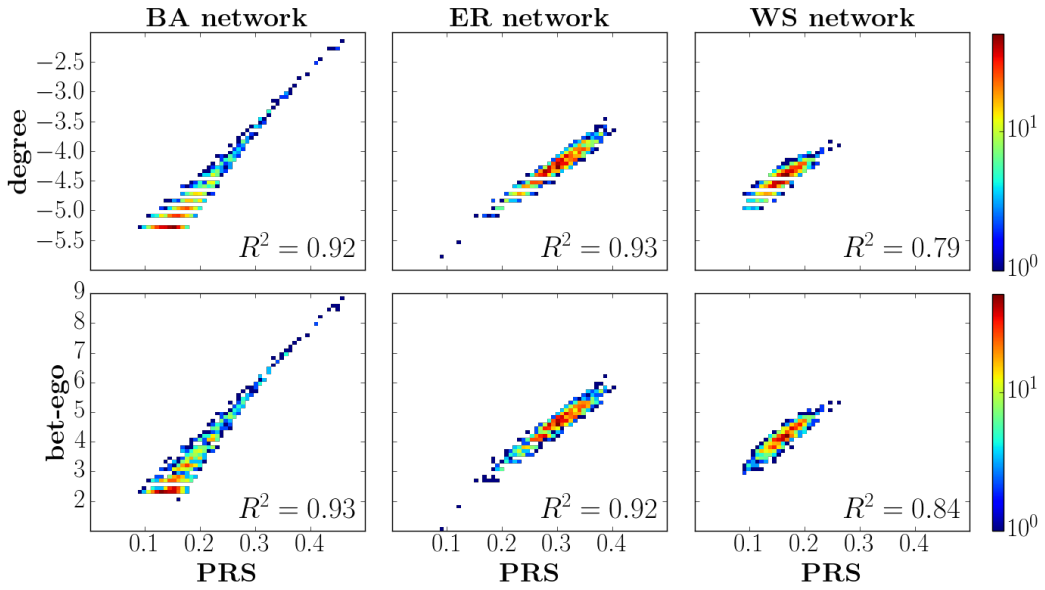


Figure 9: Correlation between local centrality metrics and PRS for different social network topologies.

ego-betweenness and degree centrality metrics rely on local data, they provide values in scale-free and random network topologies that can be used to provide a fitted approximation of the PRS. Based on agents' privacy risk, both local metrics detect the same groups of agents that were detected with global metrics. The R^2 values obtained with local centrality metrics in some cases improve the results provided by global centrality metrics, or these results are at least as good as those provided by global metrics.

Nevertheless, there are some situations where the degree or ego-betweenness centrality of an agent can be misleading for detecting privacy risk. For instance, an agent a_i can be highly connected to other agents with a low degree of connection and a_i has a high PRS value. However, the message diffusion actions that its neighbors may perform will not have a real risk impact on its privacy. Therefore, it would be interesting to consider not only the local centrality metrics of an agent, but also the centrality values of its neighbors. Hence, in the following experiments, we evaluate the relation between social degree and social

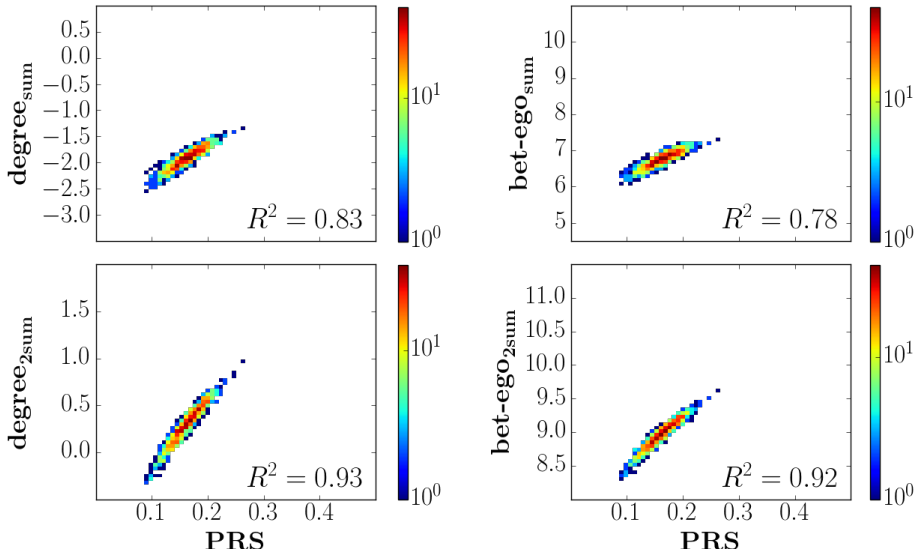


Figure 10: Correlation between social centrality metrics (i.e., $\text{degree}_{\text{sum}}$,$\text{degree}_{\text{2sum}}$, $\text{bet-ego}_{\text{sum}}$, and $\text{bet-ego}_{\text{2sum}}$) and PRS for small-world WS network.

ego-betweenness metrics and PRS. Specifically, we examine four measures in the first and second level: $\text{bet-ego}_{\text{sum}}$, $\text{degree}_{\text{sum}}$, $\text{bet-ego}_{\text{2sum}}$, and $\text{degree}_{\text{2sum}}$ (see Equation 10, 11, 12, and 13). We do not consider further distance since the majority of diffusion cascades in reality are small [13].

Figure 10 shows the results achieved with social degree and social ego-betweenness centrality metrics for the small-world WS network. The relationship between social centrality and PRS values in scale-free and random structures is not shown since the values obtained were similar to those obtained by using previous centrality metrics. The correlation between centrality and PRS values in the small-world WS network improves considerably for $\text{bet-ego}_{\text{2sum}}$ and $\text{degree}_{\text{2sum}}$, while there is not any improvement for $\text{bet-ego}_{\text{sum}}$ and $\text{degree}_{\text{sum}}$. The reason for this could be that the ability to disseminate information in level 2 (i.e., direct neighbors of neighbors) has a great impact on the final PRS. $\text{bet-ego}_{\text{2sum}}$ and $\text{degree}_{\text{2sum}}$ capture this effect better than $\text{bet-ego}_{\text{sum}}$ and $\text{degree}_{\text{sum}}$.

Network type	Level	R^2 score					
		local centralities		social centralities			
		degree	bet-ego	$\text{degree}_{\text{sum}}$	$\text{bet-ego}_{\text{sum}}$	$\text{degree}_{\text{2sum}}$	$\text{bet-ego}_{\text{2sum}}$
scale-free (BA)	2	0.79	0.79	0.78	0.47	0.63	0.35
	3	0.33	0.36	0.42	0.36	0.53	0.63
	4	0.28	0.30	0.83	0.90	0.94	0.88
random (ER)	2	0.91	0.90	0.94	0.88	0.93	0.92
	3	0.78	0.78	0.78	0.72	0.79	0.79
	4	0.85	0.85	0.91	0.89	0.93	0.93
small-world (WS)	2	0.75	0.81	0.81	0.78	0.95	0.93
	3	0.77	0.82	0.82	0.79	0.94	0.94
	4	0.51	0.53	0.59	0.61	0.71	0.70

Table 5: Evaluation of the local and social centrality metrics correlation with PRS by levels for different network topologies.

When analyzing the relationship between the local and social version of degree and ego-betweenness and the PRS values by levels (see Table 5), we detect that local centrality metrics have a behavior similar to social centrality metrics. In general, if we compare local centrality with social centrality metrics, we find that the estimation of the PRS by levels improves for the three topologies, especially for deep levels such as level 4. Finally, comparing both social and local centrality metrics, $\text{degree}_{\text{2sum}}$ obtains a slightly higher degree of correlation with PRS by levels than the other centrality metrics.

Figure 11 shows the results obtained in real networks. Most users have low PRS values (i.e., values in the range [0,0.2]). Social and local ego-betweenness are not suitable to distinguish between users with high or low PRS. The degree of correlation is lower than 0.5 (see Figure 11 – second column). However, social and local degree centrality metrics provide better results. Degree and $\text{degree}_{\text{sum}}$ show a high degree of correlation (i.e., 0.66 with degree and 0.82 with $\text{degree}_{\text{sum}}$). The results are close to those obtained in synthetic networks, where $\text{degree}_{\text{2sum}}$ obtained a high degree of correlation with PRS.

The experiments validate the use of centrality metrics to approximate PRS values in scenarios where there is no information about the activity generated

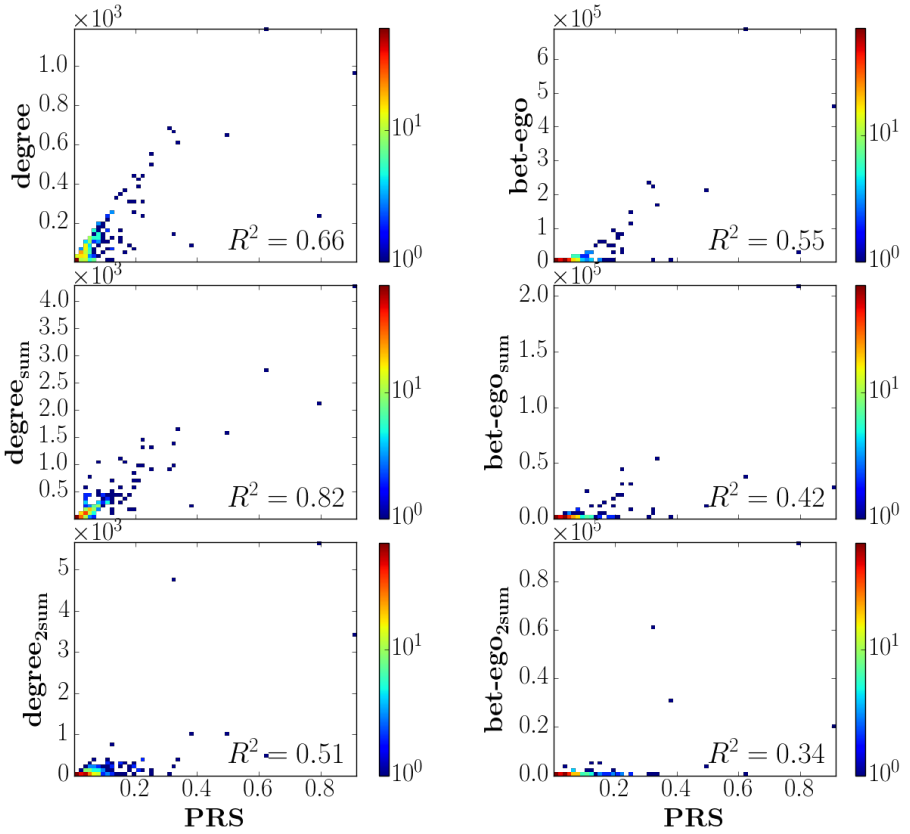


Figure 11: Correlation between local (bet-ego and degree) and social ($\text{degree}_{\text{sum}}$ and $\text{degree}_{\text{2sum}}$) centrality metrics in rumour social networks.

in the social network. In scenarios where there is information about network structure, global metrics such as closeness show a high degree of correlation with PRS and PRS in levels. In scenarios where there is only local knowledge, local and social centrality metrics based on ego-networks also provide good results. Specifically, local centrality metrics provide results estimating PRS values that are just as good as those obtained with global metrics or even better in some topologies such as scale-free networks. Social centrality metrics have also been evaluated and the metrics that consider centrality properties based on neighbors of neighbors ($\text{degree}_{\text{2sum}}$ and $\text{bet-ego}_{\text{2sum}}$) obtain the best degree of correlation

with PRS and PRS in levels. Finally, we have tested local and social centrality metrics to estimate PRS values with real data from rumor networks. The results show that degree and $\text{degree}_{\text{sum}}$ provide the best approximation to estimate the privacy risk of an action.

7. Conclusions

Most privacy approaches focus on mechanisms that semi-automatically facilitate the definition of privacy policies to define the audience that a user expects is going to receive the information published. However, there is still an open problem of making users aware of the extent of sharing information on the social network, even if such information reaches the audience previously defined. In this paper, we have focused on solving this problem. A measure of the privacy risk of a user-sharing action, PRS, has been proposed based on the scope of its dissemination in the network with the following main contributions:

- The PRS is oriented to estimating the reachability of users' sharing actions instead of being focused on the misalignment of their users' expected audience with the actual audience.
- This measure is provided globally and in levels in order to be able to adjust to the user's perception of risk.
- The PRS takes into account the paths that the publications follow in the social network without the need for the user to have to provide information explicitly.
- Centrality metrics have proven to be good estimators in establishing an approximation of the PRS in those social networking environments whose detailed record of the information sharing activity in the social network is not available.

As shown in Section 6, despite the topological properties of the network, centrality metrics can evaluate the user's relevance in information transmission

processes. We have considered global metrics (i.e., betweenness, closeness, and pagerank) for scenarios where a complete view of the network it is available, and local and social measures (i.e., degree, ego-betweenness) for scenarios where you only have a local view of the structure of the network. To evaluate the relationship between these measures of centrality and the proposed measure of PRS, we have performed a set of experiments in different topologies of synthetic networks and in real networks of rumors. The results showed that in scenarios where there is information about network structure, global metrics such as closeness show a high degree of correlation with PRS and PRS in levels. In scenarios where there is only local knowledge, local and social centrality metrics based on ego-networks provide a suitable approximation to PRS and PRS in levels. The results in real social networks confirm that local and social centrality metrics based on degree perform well in estimating a user's privacy risk and could be integrated in social network applications that offer limited information access.

As future work, we plan to validate the proposed privacy risk score through experiments in real environments. These experiments will provide feedback about the effect of the use of PRS on user behavior in social networks. We also plan to evaluate different methods (i.e., numeric values, text messages, color gradient, etc.) to show PRS values in order to inform the user about the risk of certain actions in the network. We will also evaluate the inclusion of new parameters (i.e., tie-strength between users, user personality, type of content posted, etc.) that may influence the privacy risk in order to obtain more accurate values.

8. Acknowledgements

This work is partially supported by the Spanish Government project TIN2014-55206-R and FPI grant BES-2015-074498.

References

[1] E. Del Val, M. Rebollo, V. Botti, Does the type of event influence how user interactions evolve on Twitter?, PloS one 10(5) (2015) 1–32.

[2] D. Christin, Privacy in mobile participatory sensing: Current trends and future challenges, Journal of Systems and Software 116 (2016) 57 – 68.

[3] K. Liu, E. Terzi, A framework for computing the privacy scores of users in online social networks, ACM Transactions on Knowledge Discovery from Data (TKDD) 5 (1) (2010) 1–6.

[4] R. K. Nepali, Y. Wang, Sonet: A social network model for privacy monitoring and ranking, in: Proc. of 33rd International Conference on Distributed Computing Systems Workshops (ICDCSW), 2013, pp. 162–166.

[5] M. Shehab, H. Touati, Semi-supervised policy recommendation for online social networks, in: Proc. of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2012, pp. 360–367.

[6] L. Fang, K. LeFevre, Privacy wizards for social networking sites, in: Proc. of the WWW, ACM, 2010, pp. 351–360.

[7] B. Vidyalakshmi, R. K. Wong, C.-H. Chi, Privacy scoring of social network users as a service, in: SCC, IEEE, 2015, pp. 218–225.

[8] I. Bilogrevic, K. Huguenin, B. Agir, M. Jadliwala, J.-P. Hubaux, Adaptive information-sharing for privacy-aware mobile social networks, in: Proc. of the Ubicomp, 2013, pp. 657–666.

[9] Z. Sun, L. Han, W. Huang, X. Wang, X. Zeng, M. Wang, H. Yan, Recommender systems based on social networks, Journal of Systems and Software 99 (2015) 109 – 119.

[10] G. Calikli, M. Law, A. K. Bandara, A. Russo, L. Dickens, B. A. Price, A. Stuart, M. Levine, B. Nuseibeh, Privacy dynamics: Learning privacy

norms for social software, in: Proc. of the 11th SEAMS, ACM, 2016, pp. 47–56.

[11] Ö. Kafali, A. Günay, P. Yolum, Protoss: A run time tool for detecting privacy violations in online social networks, in: Proc. of ASONAM, 2012, pp. 429–433.

[12] Y. Mester, N. Kökciyan, P. Yolum, Negotiating privacy constraints in online social networks, in: Proc. of CARE, Springer International Publishing, 2015, pp. 112–129.

[13] S. Goel, D. J. Watts, D. G. Goldstein, The structure of online diffusion networks, in: Proc. of the ACM SIGecom, 2012, pp. 623–638.

[14] G. Lawyer, Understanding the influence of all nodes in a network, Scientific reports 5.

[15] J.-G. Liu, J.-H. Lin, Q. Guo, T. Zhou, Locating influential nodes via dynamics-sensitive centrality, Scientific reports 6.

[16] G. F. de Arruda, A. L. Barbieri, P. M. Rodríguez, F. A. Rodrigues, Y. Moreno, L. da Fontoura Costa, Role of centrality for the identification of influential spreaders in complex networks, Physical Review E 90 (3) (2014) 032812.

[17] E. Y. Daraghmi, S.-M. Yuan, A small world based overlay network for improving dynamic load-balancing, Journal of Systems and Software 107 (2015) 187 – 203.

[18] S. Pei, L. Muchnik, J. S. Andrade Jr, Z. Zheng, H. A. Makse, Searching for superspreaders of information in real-world social media, Scientific reports 4.

[19] S. Wen, J. Jiang, B. Liu, Y. Xiang, W. Zhou, Using epidemic betweenness to measure the influence of users in complex networks, Network and Computer Applications 78 (2017) 288–299.

[20] L. Lü, Y.-C. Zhang, C. H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PloS one 6 (6) (2011) e21202.

[21] M. Šikić, A. Lančić, N. Antulov-Fantulin, H. Štefančić, et al., Epidemic centrality – is there an underestimated epidemic impact of network peripheral nodes?, The European Physical Journal B-Condensed Matter and Complex Systems 86 (10) (2013) 1–13.

[22] F. Bauer, J. T. Lizier, Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: A walk counting approach, EPL (Europhysics Letters) 99 (6) (2012) 68007.

[23] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Physical review letters 86 (14) (2001) 3200.

[24] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, H. A. Makse, Identification of influential spreaders in complex networks, Nature physics 6 (11) (2010) 888–893.

[25] U. Brandes, D. Fleischer, Centrality measures based on current flow, in: Annual Symposium on Theoretical Aspects of Computer Science, 2005, pp. 533–544.

[26] L. C. Freeman, et al., Centrality in social networks: Conceptual clarification, Social networks 1 (3) (1979) 215–239.

[27] L. C. Freeman, A set of measures of centrality based on betweenness, Sociometry (1977) 35–41.

[28] Q. Li, T. Zhou, L. Lü, D. Chen, Identifying influential spreaders by weighted leaderrank, Phys. A: Statistical Mechanics and its Applications 404 (2014) 47–55.

[29] L. C. Freeman, S. P. Borgatti, D. R. White, Centrality in valued graphs: A measure of betweenness based on network flow, Social networks 13 (2) (1991) 141–154.

[30] P. Bonacich, Power and centrality: A family of measures, American journal of sociology (1987) 1170–1182.

[31] K. Lerman, P. Jain, R. Ghosh, J.-H. Kang, P. Kumaraguru, Limited attention and centrality in social networks, in: Proc. of International Conference on Social Intelligence and Technology (SOCIETY), IEEE, 2013, pp. 80–89.

[32] N. L. Muscanell, R. E. Guadagno, Make new friends or keep the old: Gender and personality differences in social networking use, Computers in Human Behavior 28 (1) (2012) 107–112.

[33] E. Christofides, A. Muise, S. Desmarais, Hey mom, what's on your facebook? comparing facebook disclosure and privacy in adolescents and adults, Social Psychological and Personality Science 3 (1) (2012) 48–54.

[34] F. Stutzman, R. Capra, J. Thompson, Factors mediating disclosure in social network sites, Computers in Human Behavior 27 (1) (2011) 590–598.

[35] J. Fogel, E. Nehmad, Internet social network communities: Risk taking, trust, and privacy concerns, Computers in human behavior 25 (1) (2009) 153–160.

[36] M. Yang, Y. Yu, A. K. Bandara, B. Nuseibeh, Adaptive sharing for online social networks: a trade-off between privacy risk and social benefit, in: 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, 2014, pp. 45–52.

[37] P. Erdos, Graph theory and probability, canad. J. Math 11 (11) (1959) 34–38.

[38] R. Van Der Hofstad, Random graphs and complex networks (2016).

[39] B. Bollobás, Modern graph theory, Vol. 184, Springer Science & Business Media, 2013.

[40] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.

[41] G. Caldarelli, Scale-Free Networks: Complex Webs in Nature and Technology, Oxford University Press, 2007.

[42] D. J. Watts, S. H. Strogatz, Collective dynamics of small-world networks, Nature 393 (6684) (1998) 440–442.

[43] D. Centola, The spread of behavior in an online social network experiment, science 329 (5996) (2010) 1194–1197.

[44] M. Newman, Networks: an introduction, Oxford university press, 2010.

[45] S. P. Borgatti, Centrality and network flow, Social networks 27 (1) (2005) 55–71.

[46] C. Haythornthwaite, Social network analysis: An approach and technique for the study of information exchange, Library & information science research 18 (4) (1996) 323–342.

[47] A. Landherr, B. Friedl, J. Heidemann, A critical review of centrality measures in social networks, Business & Information Systems Engineering 2 (6) (2010) 371–385.

[48] T. C. Silva, L. Zhao, Machine learning in complex networks, Vol. 2016, Springer, 2016.

[49] M. E. Newman, A measure of betweenness centrality based on random walks, Social networks 27 (1) (2005) 39–54.

[50] M. Everett, S. P. Borgatti, Ego network betweenness, Social networks 27 (1) (2005) 31–38.

[51] P. V. Marsden, Egocentric and sociocentric measures of network centrality, Social networks 24 (4) (2002) 407–422.

[52] B. Guidi, M. Conti, A. Passarella, L. Ricci, Distributed protocols for ego betweenness centrality computation in dosns, in: Proc. of PERCOM, IEEE, 2014, pp. 539–544.

[53] S. Abar, G. K. Theodoropoulos, P. Lemarinier, G. M. OHare, Agent based modelling and simulation tools: A review of the state-of-art software, Computer Science Review 24 (Supplement C) (2017) 13 – 33.

[54] M. Kaur, H. Kaur, Implementation of enhanced graph layout algorithm for visualizing social network data using networkx library, International Journal of Advanced Research in Computer Science 8 (3).

[55] A. Nino, C. Munoz-Caro, S. Reyes, M. Castillo, A java api for the description of large complex networks under the object-oriented paradigm, Int. J. Complex Systems in Science 5 (1) (2015) 9–11.

[56] N. Akhtar, Social network analysis tools, in: Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on, IEEE, 2014, pp. 388–392.

[57] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, P. Tolmie, Analysing how people orient to and spread rumours in social media by looking at conversational threads, PloS one 11 (3) (2016) e0150989.