Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 108C (2017) 28–37

Procedia Computer Science

International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland

Impact of Neighbors on the Privacy of Individuals in Online Social Networks

Livio Bioglio and Ruggero G. Pensa

University of Turin - Dept. of Computer Science, Turin, Italy
{livio.bioglio,ruggero.pensa}@unito.it

Abstract

The problem of user privacy enforcement in online social networks (OSN) cannot be ignored and, in recent years, Facebook and other providers have improved considerably their privacy protection tools. However, in OSN's the most powerful data protection "weapons" are the users themselves. The behavior of an individual acting in an OSN highly depends on her level of privacy attitude: an aware user tends not to share her private information, or the private information of her friends, while an unaware user could not recognize some information as private, and could share it without care to her contacts. In this paper, we experimentally study the role of the attitude on privacy of an individual and her friends on information propagation in social networks. We model information diffusion by means of an extension of the Susceptible-Infectious-Recovered (SIR) epidemic model that takes into account the privacy attitude of users. We employ this diffusion model in stochastic simulations on a synthetic social network, designed for miming the characteristics of the Facebook social graph.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Science

Keywords: complex networks, modeling, information diffusion, privacy

1 Introduction

The problem of user privacy in the so-called "Big Data Era" cannot be ignored and many companies are realizing the necessity to consider it at every stage of their business. In practice, they have been turning to the principle of *Privacy by Design* [6] by integrating privacy requirements into their business model. Online social network (OSN) providers are embracing this model as well. In recent years, Facebook has improved considerably the privacy protection tools provided within its Web and mobile products, and periodically suggests its users to review their privacy settings using a simplified but still flexible interface. However, differently from other Web, mobile and IoT services where data protection mostly involves access control rules, data anonymization techniques and other centralized or decentralized precautions that are invisible to the users, in OSN's the most powerful data protection "weapons" are the users themselves. In fact, social media (e.g., Facebook, Instagram, Twitter) are essentially human-generated logs

1877-0509
Peer-review under responsibility of the scientific committee of the International Conference on Computational Science
10.1016/j.procs.2017.05.030

that can be used to reconstruct life events and private facts of those users that carelessly disclose their personal information. As shown by the research project myPersonality [13] carried out at the University of Cambridge, by leveraging Facebook user's activity (such as "Likes" to posts or fan pages) it is possible to "guess" some very private traits of the user's personality. To alleviate this issue, social media usually provide advanced tools for controlling the privacy settings of the user's profile [22], but it has been shown that yet a large part of Facebook content is shared with the default privacy settings and exposed to more users than expected [15]. Moreover, even though OSN users can specify which of their contacts are allowed to see their notifications, they do not have any control on how these contacts will use their information: friends could spread the rumor through other social networks, blogs, websites, medias or simply with face-to-face communication.

The behavior of an individual in these situations highly depends on her level of privacy awareness: an aware user tends not to share her private information, or the private information of her friends, while an unaware user could not recognize some information as private, and could share it without care to her contacts, even to untrusted ones, putting her privacy or the privacy of her friends at risk. Users' privacy awareness then turns into the so-called "privacy attitude", i.e., the users' willingness to disclose their own personal data to other users, that can be measured by leveraging the way users customize their privacy settings in social networking platforms [14, 21].

A new question may arise now: how safe is the privacy of a social network user who is mostly surrounded by friends with a good privacy attitude? The question is not trivial, since the way most people set their privacy settings is based on the notion of closeness: close friends are usually allowed to see all user's updates, while acquaintances can only see "public" or less sensitive updates[1]. The common assumption is that closed friends are trusted ones and thus will not disclose friends' posts to other friends. In this paper, we model the effects of privacy attitude on information propagation in social networks with the aim of studying what happens to information diffused to friends with different levels of privacy awareness. We employ a model, proposed by us in [5] and inspired by the classic Susceptible-Infectious-Recovered (SIR) epidemic model [11], for representing privacy attitude of individuals by means of parametric values. By tuning the values of parameters, we model the attitude on privacy of single users, from more to less aware on privacy. Our objective is to investigate the role of privacy attitudes of the initial spreader, of her neighbors and of the whole population of the social network on the diffusion of information into a Facebook-like social network. We analyze, by means of stochastic simulations, the spreading of information in a Facebook-like network starting from a unique initial spreader. We assign to each user of the network a privacy class, representing its attitude on privacy: in order to study the role of privacy awareness of the whole population, we simulate information spreading on different assignment distributions of privacy classes of the nodes, from safer (where the majority of users has a high awareness on privacy) to unsafer ones (where the majority of users has a low level of attention on privacy issues). With the goal of studying the role of the neighborhood of a user in information diffusion, we set all the nodes directly linked with the initial spreader to the same privacy class, repeating the assignment for all privacy classes. Finally, we repeat the simulations choosing a node in every privacy class as initial diffuser to analyze the role of the privacy attitude of the initial spreader.

[1]Facebook, for instance, allows its users to distinguish between friends, close friends and acquaintances during any posting action.

1.1 Related work

In epidemiology, the Susceptible-Infectious-Recovered (SIR) epidemic model [11] is employed for modeling infectious diseases that confer lifelong (or long-term) immunity, such as measles, rubella or chickenpox. In this model a susceptible (S) node can become infected, because of the presence of infectious (I) nodes, and an infectious node can naturally recover (R) after few time, gaining immunity to the disease. The SIR model has been applied to information spreading since early years, even if these applications slightly differ from the common model [7, 16, 10]: in fact, in this case, if an infectious node contacts another infectious one, it (or both of them, depending on the implementation) loses interest in the rumor and becomes recovered. An extension of this model also allows spontaneous recovery, justified as forgetting mechanism: in this case, the model behaves more similarly to the classical SIR model, as observed in [17].

A large part of research works on privacy issues in online social networks focuses on the anonymization of networked data [23], while our work can be positioned in another branch of research that focuses on modeling, measuring and preventing privacy leakage in online social networks. In this regard, one of the most prominent work is [14] where Liu and Terzi propose a framework to compute a privacy score measuring the users' potential risk caused by their participation in the network. An extension to this score taking into account the centrality of users in the social graph has been presented in [18]. However, it has been shown that these approaches are based on the wrong assumption that all user's friends are equally close [19]. Instead, Fang and LeFevre [9] describe a social networking privacy wizard based on active learning to help users customize their privacy settings. [4] presents a tool to detect unintended information loss in online social networks by quantifying the privacy risk attributed to friend relationships in Facebook. In [20] the authors measure the inference probability of sensitive attributes from friendship links, while the authors of [2] define a measure of how much it might be risky to have interactions with them, in terms of disclosure of private information.

Differently from those studies, in our work we focus on rumor spreading in presence of a sort of "immunization parameter" that models the privacy attitude of users, i.e., their willingness to disclose their own personal data to other users directly or indirectly.

1.2 Contribution

We adopt the SIR model for modeling the spread of information in a social network: susceptible (S) individuals do not know the information item, then are susceptible to be informed; infectious (I) individuals know and spread the information item, while recovered (R) individuals already know the information item but do not spread it anymore. In the considered extension, each individual belongs to a privacy attitude class that tunes the parameters of the model. By means of stochastic simulations, we study the role of privacy on information diffusion in a synthetic network reproducing the characteristics of the real Facebook social graph, with different distributions of privacy attitude of their nodes. We then try to answer the following research question: is the safe privacy attitude of friends a sufficient condition to avoid the diffusion of private information in the network?

The practical outcomes of our study involves the way user privacy protection is handled by social media providers. If the common assumed model turned out to be false, at least under certain conditions, it should be reconsidered.

The remainder of the paper is organized as follows: the privacy-aware propagation model is presented in Section 2; Section 3 provides the report of our experimental research; finally, we draw some conclusions in Section 4.

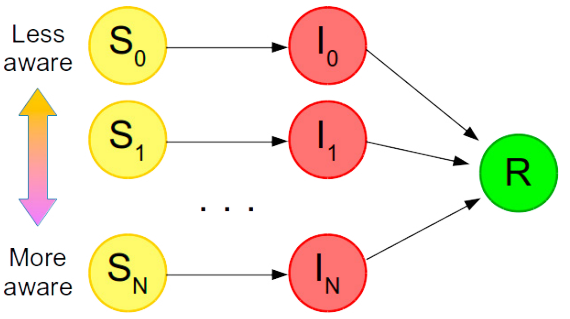

Figure 1: Transmission model. Each index of compartments S and I represents a privacy class

2 A privacy-aware model for information spreading

We consider a social graph G composed by a set of n nodes $\{v_1, \ldots, v_n\}$ representing the users of G. We represent the social network as a directed graph $G(V, E)$, where V is a set of n nodes and E is a set of directed edges $E = \{(v_i, v_j)\}$. Given a pair of nodes $v_i, v_j \in U$, $(v_i, v_j) \in E$ iif there exists a link from v_i to v_j (e.g., users v_i is in the friend list/circle of v_j or v_j follows v_i). We define the neighborhood of a node $v_i \in V$ as the set of nodes v_k directly connected to the node v_i, i.e., $\mathcal{N}(v_i) = \{v_k \in V \mid (v_i, v_k) \in E\}$. We can see the neighborhood as the list of friends or followers of user v_i. Finally, we consider a set P of privacy classes, representing the propensity of a user of the class to disclose her own or other's information, directly or indirectly. In practical terms, in online social networks (such as Facebook, Twitter, Instagram or Google+) the privacy class may be unveiled by the way users configure their privacy settings, or the way they post or share/comment other users' posts. Each user of G belongs to a privacy class $p \in P$.

2.1 Information spreading model

In the SIR model, at any time step an individual v_i may belong to susceptible (S), infectious (I) or recovered (R) compartment. A susceptible (S) individual may be infected by an infectious (I) individual in contact with her with a probability λ, called infection probability, becoming infectious (I) herself. An infectious (I) individual v_i may spontaneously recover from infection with probability μ, called recovery probability, entering the recovered (R) compartment. An infectious individual (I) may infect each user in her neighborhood belonging to the susceptible (S) compartment. We denote with $c(v_i, t) \in \{S, I, R\}$ the compartment of user v_i at time t.

The SIR model is usually applied for modeling infectious diseases that confer lifelong (or long-term) immunity, but it can be also used for modeling the diffusion of information in a population: susceptible individuals are those who not already know the information item, and then they are susceptible to be informed; infectious individuals are the ones who know it and actively spread it; finally, recovered individuals already know the information item but do not spread it anymore. We can see the recovery process as an aging mechanism that lets the individual loose interest on information after few time steps, stopping its diffusion.

Here we employ an extension, proposed by us in [5], of the SIR model for considering the explicit or implicit privacy polices of an individual during the spread of information on a social network. A set of privacy classes $P = \{p_0, p_1, \ldots, p_N\}$ is assigned to the susceptible (S) and infectious (I) compartments, representing the privacy class of an individual belonging to the

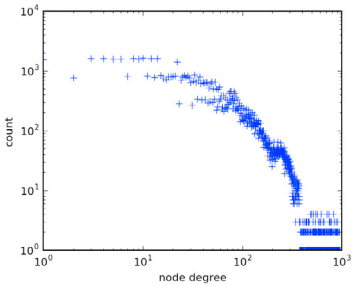

Figure 2: Degree distribution of the Facebook-like synthetic network

compartment, and consequently her behavior on information spreading, from less aware (p_0) to more aware (p_N). This behavior is reached by assigning different values to the parameters λ and μ of each privacy class: every privacy class $p \in P$ is linked to a different pair of values λ_p and μ_p. A graphic representation of our extension of classic SIR model is given in Figure 1. We also introduce a novel parameter $\beta_p \in [0, 1]$ in the SIR transmission model, symbolizing the interest in information of the users in privacy class p.

The evolution of information diffusion follows the Reed-Frost chain-binomial model [1]: it consists in a stochastic approach, where time is measured in discrete units and infection occurs because of direct contacts. The evolution probabilities are obtained as follows. Let $p(v_i) = p \in P$ be the privacy class of an individual v_i. If it belongs to the susceptible compartment, it may be infected at time $t + 1$ with probability:

$$P_{inf}(v_i, t+1) = \beta_p \cdot (1 - \prod_{p' \in P} (1 - \lambda_{p'})^{n_I(v_j,t)}) \tag{1}$$

where $n_I(v_j, t) = |\{v_j \in \mathcal{N}(v_i) \mid c(v_j, t) = I \ \wedge \ p(v_j) = p'\}|$ is the number of individuals in infectious (I) compartment and privacy class p' at time t among the neighbors of individual v_i. Otherwise, if the individual v_i of privacy class p belongs to the infectious (I) compartment at time t, it may recover with probability μ_p at time $t + 1$.

3 Experiments and results

In this section we provide the results of our experiments performed over a Facebook-like synthetic network. We firstly present the contact network, then the details of our experimentation and the results. Finally, we provide an explorative analysis of the parameter space.

3.1 Contact network

Information spreads on a contact network, in which nodes represent individuals, and edges between nodes represent contacts between two individuals. In our experiments we employ a Facebook-like network generated using LDBC–SNB Data Generator[2] which produces graphs that mimic the characteristics of real Facebook networks [8]: in particular, we generate a network with 80,000 nodes, but here we consider only the greatest connected component of such network, composed by approximately 75,000 nodes and 2,700,000 edges. The degree distributions of this network is given in Figure 2. The graph is undirected.

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

Table 1: Values of the parameters for the three privacy classes

	Classes		
Parameter	0	1	2
β	0.85	0.5	0.15
μ	0.5	0.5	0.5
λ	0.85	0.5	0.15

Figure 3: Class distribution in the three kinds of class assignments

3.2 Experimental settings

Our experiments are conducted as follows. We perform 100 stochastic simulations of information spreading on a completely susceptible population, except for one infectious node. We select three privacy classes, numbered from 0 to 2, representing users from unaware (class 0) to more aware on privacy (class 2), in order to provide a few grades of awareness. Table 1 reports the values assigned to parameters of the spreading model for each privacy class. Users in class 0 have a high probability of being interested in information and spreading it, and the opposite for users in class 2; class 1 represents average users, then its parameters have been tuned accordingly. We randomly assign to each node of network in Section 3.1 a privacy class. In order to study the role of the global privacy attitude of the whole network, we repeat the set of simulations for 3 different class distributions of privacy classes of the nodes: a safer assignment, where the majority of nodes are in the most aware class (the number 2); a medium assignment, where the majority of nodes are in the middle privacy class (1); an unsafer assignment, where the majority of nodes are in the less aware class 0. The number of nodes in each privacy class of these three class distributions are graphically summarized in Figure 3. For studying the role of the neighborhood, we assign the same privacy class $p \in P$ to all the nodes in the neighborhood of the initial infectious node, repeating the set of simulations for each privacy class in $p \in P$. Finally, in order to study the role of the privacy class of the initial spreader, we choose the initial infectious individual among all the nodes of privacy class $p' \in P$. More precisely, with the aim of reducing biased results we collect and aggregate the data from 10 different nodes for each privacy class, having degree close to the average value of the network. We repeat this set of simulations for each privacy class in $p' \in P$. Summarizing, for each class distribution in Section 3.1, for each pair of privacy classes in $p, p' \in P$, for 10 times we randomly choose a node of the network in privacy class p', having degree close to the average value of the network, we change the privacy class of all the nodes in its neighbors list to p, we perform 100 simulations and then we aggregate the results.

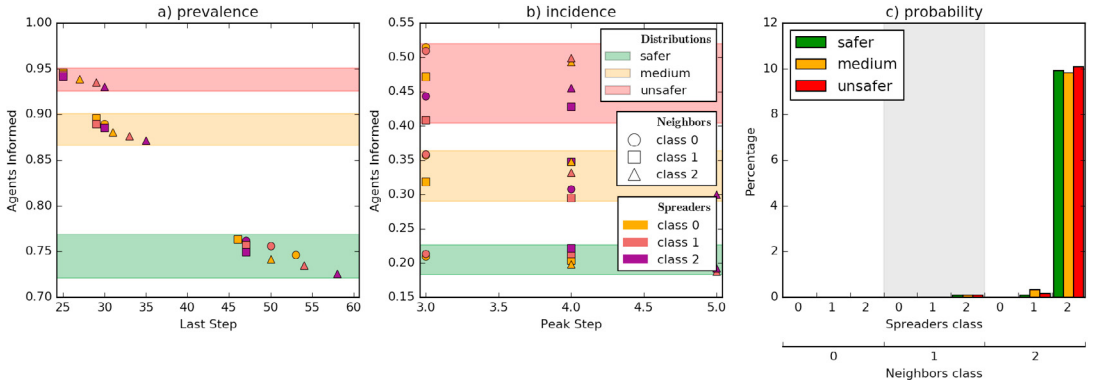

Figure 4: a) Median of prevalence of informed individuals (ratio) in each set of simulations. b) Median of incidence of informed individuals (ratio) in each set of simulations. c) Probability that information does not reach more than 1% of the population.

3.3 Results

For each set of simulations we observe the proportion of informed individuals at each time step (prevalence), that is the number of nodes in infectious (I) or recovered (R) compartments, and the proportion of new infection cases at each time step (incidence), that is the number of nodes changing their compartment from susceptible (S) to infectious (I).

In order to compare the results of several simulations, we collect some key features of the prevalence and the incidence curves. For prevalence curves, we gather the proportion of informed individuals at the end of simulations, when the number of infectious individuals is equal to zero (then nobody can spread information anymore), and the step when simulation ends: such data convey the duration of the spread and its diffusion among the population. For incidence curves, we collect two types of information related to the peak of new cases of informed individuals: i) the step where this peak is reached and ii) the proportion of population involved. Such result offers a snapshot on the speed of information diffusion.

The features extracted from our simulations are graphically summarized in Figure 4(a) for prevalence curves and in Figure 4(b) for incidence curves. For each point of these graphs, the marker color identifies the privacy class of the initial spreader, its shape identifies the privacy class of the neighbors of initial spreader node, while its background color identifies the distribution of privacy classes of the nodes in the whole network. Each point shows the median value resulted from 100 simulations performed on 10 initial spreaders.

As we have already noticed in [5], the attitude on privacy in the whole network is responsible of the proportion of population informed, both at the end of diffusion and at the peak of new cases. Where the majority of nodes is unaware, information immediately spreads over almost the entire population, while where the network is full of aware individuals, information reaches a smaller part of the population. It also greatly affects the duration of the information spreading process: in a network where users are more attentive on privacy, the diffusion of information is few times longer than in other cases. Interestingly, even in this case information reaches a huge portion of the population, around 75%. The speed of diffusion is more evident in the curves of incidences, depicting the proportion of new cases of informed individuals in each time step (Figure 4(b)): under the least safe distribution, information immediately reaches around half of population, while for safer distributions this peak is lower. However, the step where

this peak is reached does not depend on the attitude on privacy of the population: its value is approximately the same in every case.

The duration of spread is a key feature for evaluating the impact of an information item on a population. If the diffusion is slow, and takes more time for reaching a huge portion of population, there is more probability to block it before its natural death. For example in case of diffusion of hoaxes or false news, authorities of the social network have more time for applying containment policies for contrasting its proliferation, like the isolation of spreaders or the diffusion of official news revealing the hoax. Moreover, in our experiment we propagate only one information item at a time, but in real scenarios there are several news diffusing at the same time, competing for the attention of users: an information item loses of interest quite quickly. If it is not able to reach rapidly a huge portion of population, it is possible that information naturally stops being spread in favor of other news. We recall that the privacy attitude of a user influences the way she sets her privacy settings that, in turn, has an impact on the visibility of any (both public and private) of her sharing actions.

The most noticeable result for information diffusion is the role of the attitude towards privacy of the initial spreader and its neighbors: a safer attitude of the node and its neighbors decreases the portion of informed population, and extends the duration of information diffusion, but its impact is not as influential as the behavior of the whole network. This is likely due to the high average degree of the nodes, equal to 71: for an aware user, even if the probability of diffusing information to a friend is low, the number of friends is so high that a small number of friends can become spreaders themselves. This consideration is applicable to the informed friends too, and as soon as information reaches a node out of the neighborhood, its diffusion depends only on the attitude on privacy of the whole network. For this reason we decide to analyze the portion of simulations where information has reached only a small portion of the population, lower than 1%, before dying: our results are reported in Figure 4(c). In this case we notice that the attitude of the network is irrelevant, and only the privacy classes of the spreader and its neighborhood is crucial. Interestingly, a safe attitude of the spreader or of the neighbors is not sufficient on its own to block information diffusion. An information item on a safer user with unsafer friends, and vice versa, can easily overtake the circle of friends.

3.4 Exploration of parameters space

In our experiments we have fixed the parameters of the diffusion model described in Section 2 to arbitrary albeit reasonable values. In order to generalize our findings, we explore the space of parameters for two extreme privacy classes (the least and the most aware): for reducing the depth of parameter search space, we fix μ, and we use the same value for β and λ. In details, we let β_0 and λ_0 vary between 0.05 and 0.3 (with 0.05 step), while β_2 and λ_2 range between 0.7 and 0.95 (with 0.05 step). In Figure 5 we show, for each class distribution, the portion of informed users at the end of simulations (the value on y axis in Figure4(a)) for each couple of values assigned to the two privacy classes. The color of each tile denotes the median value of several simulations: we aggregate the results of all the assignments of privacy classes for neighborhoods and initial spreaders in the same class distribution on population. We notice that the value chosen for the more aware class (class 0) is the most important for determining the final portion of population informed, especially when the attitude on privacy of the network is safer. In this case less aware users are probably surrounded by more aware friends, who are not interested in the information item, and do not let it diffuse over the network, even if less aware individuals are spreading it to their friends with high probability. In fact this effect is less noticeable in the unsafer distribution, where less aware users can diffuse information, encountering a weaker

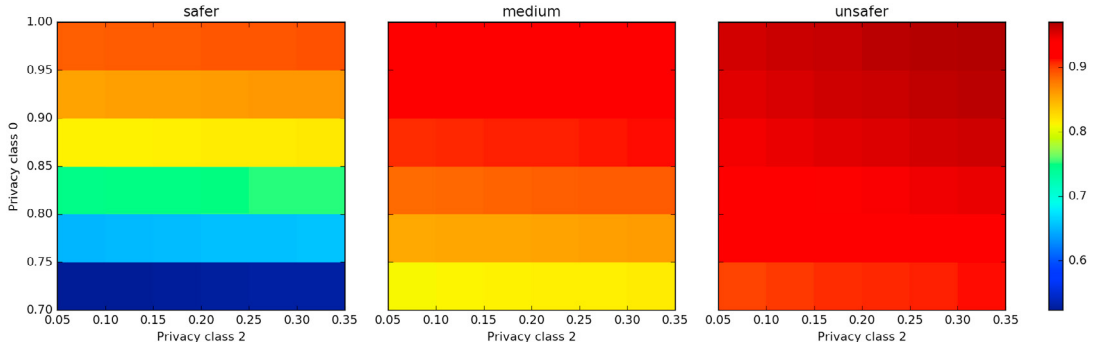

Figure 5: Portion of population informed at the end of information diffusion for each couple of values assigned to β and λ of privacy classes 0 and 2 in the three privacy class distributions.

resistance. This is due to the fact that their friends belong to their same less aware class with high probability. However, the behavior observed in Section 3.3 is confirmed: the portion of informed population, if we fix the parameters β and λ of the privacy classes, greatly depends on the attitude towards privacy of the whole population.

4 Conclusions

In this paper we have presented an information propagation model, inspired by the classical SIR epidemic model, that takes into consideration the privacy awareness of the user. Through stochastic simulations we have employed this model for studying the role of privacy in the diffusion of information in a Facebook-like network, and in particular its influence at level of the single initial spreader, her circle of friends, and the entire population.

Our results show how privacy attitude can affect the diffusion of information on social networks. On one hand, the behavior of the entire population can reduce or increase, in case of more or less aware users respectively, the portion of population which receives the information item, and its speed of diffusion. On the other hand, the attitude on privacy of the initial spreader and her friends has a marginal impact on this behavior, but they play a more important role for preventing information spreading. An interesting result is that a safer attitude of friends of an unaware user is not sufficient to block the diffusion of information. In future, we plan to extend our work by studying the role of privacy awareness of the K-th order neighbors of an individual, that is all individuals which can be reached in exactly K hops, and not only to friends. We also plan to extend such analysis to a set of node's features, such as its degree, its closeness centrality, and its clustering coefficient. The final objective is to employ all these characterizing features to predict the risk, for a particular individual, that her private information spreads on the network, looking only at her local features.

Our study shows the importance of considering the privacy attitude of users in modeling the spreading of rumors, with direct and indirect implications on all applications that involve the dynamics of information spreading, such as influence maximization [12] and community detection [3], as well as on privacy enforcement models and techniques for online social networks.

Acknowledgments

This work was supported by Fondazione CRT (grant number 2015-1638).

References

[1] H. Abbey. An Examination of the Reed-Frost Theory of Epidemics. *Human Biology*, 24(3):201, 1952.

[2] C. G. Akcora, B. Carminati, and E. Ferrari. Risks of friendships on social networks. In *Proceedings of IEEE ICDM 2012*, pages 810–815. IEEE Computer Society, 2012.

[3] N. Barbieri, F. Bonchi, and G. Manco. Influence-based network-oblivious community detection. In *Proceedings of IEEE ICDM 2013*, pages 955–960. IEEE Computer Society, 2013.

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

[5] L. Bioglio and R. G. Pensa. Modeling the impact of privacy on information diffusion in social networks. In *Complex Networks VIII: Proceedings of the 8th Conference on Complex Networks CompleNet 2017*, pages 95–107, 2017.

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

[7] D. J. Daley and D. G. Kendall. Epidemics and rumours. *Nature*, 208:1118, 1964.

[8] O. Erling, A. Averbuch, J. Larriba-Pey, H. Chafi, A. Gubichev, A. Prat-Pérez, M. Pham, and P. A. Boncz. The LDBC social network benchmark: Interactive workload. In *Proceedings of ACM SIGMOD 2015*, pages 619–630. ACM, 2015.

[9] L. Fang and K. LeFevre. Privacy wizards for social networking sites. In *Proceedings of WWW 2010*, pages 351–360. ACM, 2010.

[10] D. Gruhl, D. Liben-Nowell, R. V. Guha, and A. Tomkins. Information diffusion through blogspace. *SIGKDD Explorations*, 6(2):43–52, 2004.

[11] M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[12] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of ACM SIGKDD 2003*, pages 137–146. ACM, 2003.

[13] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *PNAS*, 110(15):5802–5805, 2013.

[14] K. Liu and E. Terzi. A framework for computing the privacy scores of users in online social networks. *TKDD*, 5(1):6, 2010.

[15] Y. Liu, P. K. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of ACM SIGCOMM IMC '11*, pages 61–70. ACM, 2011.

[16] D. P. Maki and M. Thompson. *Mathematical models and applications: with emphasis on the social, life, and management sciences*. Prentice-Hall, 1973.

[17] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili. Theory of rumour spreading in complex social networks. *CoRR*, abs/0807.1458, 2008.

[18] R. G. Pensa and G. di Blasi. A centrality-based measure of user privacy in online social networks. In *Proceedings of IEEE/ACM ASONAM 2016*, pages 1438–1439. IEEE Computer Society, 2016.

[19] R. G. Pensa and G. di Blasi. A semi-supervised approach to measuring user privacy in online social networks. In *Proceedings of DS 2016*, pages 392–407, 2016.

[20] N. Talukder, M. Ouzzani, A. K. Elmagarmid, H. Elmeleegy, and M. Yakout. Privometer: Privacy protection in social networks. In *Proceedings of M3SN'10*, pages 266–269. IEEE, 2010.

[21] Y. Wang, R. K. Nepali, and J. Nikolai. Social network privacy measurement and simulation. In *Proceedings of ICNC 2014*, pages 802–806. IEEE, 2014.

[22] L. Wu, M. Majedi, K. Ghazinour, and K. Barker. Analysis of social networking privacy policies. In *Proceedings of 2010 EDBT/ICDT Workshops*. ACM, 2010.

[23] E. Zheleva and L. Getoor. Privacy in social networks: A survey. In *Social Network Data Analytics*, pages 277–306. Springer US, 2011.