Enhancing Privacy Awareness in Online Social Networks: a Knowledge-Driven Approach

Ruggero G. Pensa
University of Turin, Dept. of Computer Science
Turin, Italy I-10149
ruggero.pensa@unito.it

Extended abstract

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

If, from one hand, many users are informed about the risks linked to the disclosure of sensitive information (private life events, sexual preferences, diseases, political ideas, among others), on the other hand the awareness of being exposed to privacy breaches each time we disclose information that apparently is not sensitive is still insufficiently widespread. In this regard, daily activities may reveal information that can be used by others in a negative manner. For example, a GPS tag far from home or pictures taken during a journey may alert potential burglars, or the disclosure of family relationships may expose our own or other family members’ privacy to criminal offence risks, as well as source of tort liability. Most troubling of all, it has been shown that by leveraging Facebook user’s activity it is possible to infer some very private traits of the user’s personality [KSG13]. This inference capability has been recently exploited to help propel Donald Trump to victory in the last U.S. presidential elections and was at the very center of the Facebook–Cambridge Analytica scandal in early 2018. This privacy breach event has multiplied the interests in the protection of human dignity and personal data, and privacy has become a primary concern among social network providers and web/data scientists.

Although social platforms often provide some kind of notification intended to inform their users about the risks of private information disclosure, many people simply overlook the dangers due to the uncontrolled disclosure of their (and others’) personal data. Therefore, following the recent scandals, most social media have considerably improved their tools for controlling the privacy settings of the user profile (e.g., Instagram can now limits the visibility of stories to “close-friend”), but such tools are often hidden and not that user-friendly. Consequently, they are barely utilized by most users. Recent machine learning and data mining studies try to go beyond these limitations by proposing some measures of users’ profile privacy based on the way they customize their privacy settings, or lightening the customization process of the privacy settings by means of guided tools and wizards [FL10, SWN$^+$18]. Privacy measures, in particular, when associated to popup alerts or other visual components, may enhance user’s perception of privacy, according to the principles of *Privacy by Design* specifications [Cav12]. These metrics usually require a *separation-based* policy configuration: in other terms, the users decide “how distant” a published item may spread in the network. Typical separation-based privacy policies for profile item/post visibility include: visible to no one, visible to friends, visible to friends of friends, public. However, this policy fails when the number of user friends becomes large. According to a well-known anthropological theory, in fact, the maximum number of people with whom one can maintain stable social (and cybersocial) relationships (known as Dunbar’s number) is around 150, but the average number of user friends in Facebook is more than double. This means that many social links are weak (offline and online interactions with them are sporadic), and a user who sets the privacy level of an item to "visible to friends" probably is not willing to make that item visible to *all* her friends. Other studies try to make the customization process of the privacy settings less frustrating. However, a consensus on how to identify a trade-off between privacy protection and exploitation of social network potentials is still far from being achieved.

Hence, in this talk, we show our theoretical framework (first presented in [PB17]) to i) measure the privacy risk of the users and alert them whenever their privacy is compromised and ii) help the exposed users customize semi-automatically their privacy level by limiting the number of manual operations thanks to an active learning approach. Moreover, instead of using a *separation-based* policy for computing the privacy risk, we adopt a *circle-based* formulation of the privacy score proposed in [LT10]. We show experimentally that our circle-based definition of privacy score better capture the real privacy leakage risk. Moreover, by investigating the relationship between the privacy measure and the privacy preferences of real Facebook users, we show that our framework may effectively support a safer and more fruitful experience in social networking sites.

Additionally, we argue that the privacy risk is not just a matter of users' preferences (i.e. to which friends a user is wishing to disclose each particular action/post); it is also heavily affected by the characteristics of the social network they belong to., i.e., their centrality within the network and the attitude of their friends towards privacy. According to a recent computational science study [BP17], even restraining privacy settings are ineffective when the user is located within an unsafe network, i.e., a network where the majority of nodes have little or no awareness about their own and others' privacy. This leads to the intuition that privacy risk in a social network may be modeled similarly as page authority in a hyperlink graph of web pages. According to a well-known theory, more authoritative web sites are likely to receive more links from other web sites that are authoritative in their turn. In this talk, we make the hypothesis that the concept of "importance" of a web-page can be transposed into the concept of "privacy risk" of users in a social network as follows: the more an individual is surrounded by friends that are careless about their privacy, the more the privacy of that individual is likely to be exposed to concrete privacy leakage risks. Then, we present a new network-aware computational method for measuring the privacy risk (first published in [PBB19]), and report on a social experiment we performed, which involves more than one hundred Facebook users. Thanks to this experiment, we show the effectiveness of our privacy measure not only on two simulated networks but also on a large network of real Facebook users.

Acknowledgements

The work presented in the talk is supported by Fondazione CRT (grant numbers 2015-1638 and 2017-2323).

References

[BP17] Livio Bioglio and Ruggero G. Pensa. Impact of neighbors on the privacy of individuals in online social networks. In *Proceedings of the International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland*, volume 108 of *Procedia Computer Science*, pages 28–37. Elsevier, 2017.

[Cav12] Ann Cavoukian. Privacy by design [leading edge]. *IEEE Technol. Soc. Mag.*, 31(4):18–19, 2012.

[FL10] Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. In *Proceedings of WWW 2010*, pages 351–360. ACM, 2010.

[KSG13] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *PNAS*, 110(15):5802–5805, 2013.

[LT10] Kun Liu and Evimaria Terzi. A framework for computing the privacy scores of users in online social networks. *TKDD*, 5(1):6:1–6:30, 2010.

[PB17] Ruggero G. Pensa and Gianpiero Di Blasi. A privacy self-assessment framework for online social networks. *Expert Syst. Appl.*, 86:18–31, 2017.

[PBB19] Ruggero G. Pensa, Gianpiero Di Blai, and Livio Bioglio. Network-aware privacy risk estimation in online social networks. *Social Netw. Analys. Mining*, 9(1):15:1–15:15, 2019.

[SWN$^+$18] Xuemeng Song, Xiang Wang, Liqiang Nie, Xiangnan He, Zhumin Chen, and Wei Liu. A personal privacy preserving framework: I let you know who can see what. In *Proceedings of ACM SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 295–304. ACM, 2018.