# Privacy Issues on Social Networks

Serap Şahin*

* Department of Computer Engineering, Izmir Institute of Technology, Turkey

Tel: +090-2327507871, e-mail: serapsahin@iyte.edu.tr

**Abstract-** The privacy is a need for humanity since the creation of civilizations. In social networks the common point is that each user has to create a profile to define his or her own identity. The profile description includes many items with privacy settings to tune their visibility degrees for only owner, friends, friends of friends and sometimes for public. After enrolment stage, users extend their social connection graphs with accepted new friends and these graphs grow without the control of individual due to the new friends of friends. Hence, with high probability, the shared information of member is generally available to public and can be retrieved by users around the world. This article is prepared to give an overview on the reasons of privacy concerns and risks of SNs, and summarize the current and future possible solution directions for researchers and governments.

**Keywords** - Privacy; social networks; data privacy regulations; information spreading; privacy policies; static and dynamic big data analysis.

## 1. Introduction

Humanity has been experiencing a transformation since the emergence of the information and communication technologies (ICT). ICT gives a new virtual world with social networks (SNs) and each of us is volunteers to be a part of it with different privacy preferences. The privacy is a need for humanity since the creation of civilizations. According to the global study of Hsiao et. al. [1], the privacy is a human right with the global agreement of nearly 80% of people with some perceptions such as “right” to be alone, to prevent unauthorized publishing, to keep your secret, to control your information flow and not to be annoyed. Additional findings of their study also show that socio-cultural factors such as age, gender, religion and region influence people’s privacy perceptions. The region has the most influence on the people’s privacy beliefs. For instance; according to this survey due to the effect of living regions, nearly 70% do not accept that the privacy is a human right and accept that “privacy is a concern only for those having something to hide”. But, the SNs have no borders and, so they increase the importance of the following two points:

- The completeness and harmony of the international laws and regulations
- The privacy design principles of technologies.

Due to this reason the EU Data Protection Regulation includes a consensus that privacy is a human right and its created policies force to technologists to rethinking on the design of privacy-enhancing systems.

When we evaluate this topic from technical point of view that in SNs the common point is that each user has to create a profile to define his or her own identity. The profile description includes items such as name, surname, city, age, gender, email, phone and education with privacy settings to tune their visibility degrees for only owner, friends, friends of friends (FoF) and sometimes for public according to owner user’s privacy perceptions. At this enrolment stage, the social connection graph of a member includes only one node. After this enrolment stage, users extend their social connection graphs by accepting new friends. Naturally, members’ graphs continue to grow without the control of the enrolled member, because SN members continuously share new information on this dynamic graph. The profile description and policy setup of users on SNs also identifies the user’s interaction graph based on spreading their information to other users. As a result, the shared private information turns to available for public with high probability and it can be retrieved by users around the world.

According to study of Ho et al. [2]; there are three reasons of privacy problems in SNs; lack of user awareness, lack of flexibility in current privacy tools and lack of control on what other users see about their profiles and posts. The continuity of these outcomes are checked with an online survey on 200 SN users by Ho et al. Results of this survey confirmed that current users of SNs have already suffer from each of three problems.

Tuunainen et al. [3] worked on privacy concerns and risks of SNs, and they specifically focused on the privacy awareness of users. They conducted a survey for 210 Facebook users and they observed the reactions of users for two aspects as privacy protection and information disclosing. The results of this survey strongly show that; most of the users do disclose a significant amount of sensitive information of themselves. The most of them are not sufficiently aware of the visibility of their information to people. And unfortunately many users do not know or understand the privacy policies of the SNs. Furthermore, some of the SNs share private information with third parties and many users also do not know this reality.

Additionally, the users of SNs have very strong modus operandi, which motivates users to stay in SNs and share their information. According to a survey by Min et al. [4] on Facebook users, the most important motivation of members is the relationship management through SNs, and others are their perceived usefulness of SNs for self-presentations, and their subjective social norms of using SNs. These are collectively having stronger impact to be member of SNs than their perceived privacy concerns. The individuals accept this relationship as a cost-benefit tradeoff.

SNs have also many important aspects with globalization, such as social and cultural integration, developing global economy and a resource for intelligence to prevent terrorist attacks and other criminal activities. Certainly, this cost-benefit relationship is represented in Fig.1. If we increase the benefit part, the number of members and conformity of SNs also increase. This is directly related with the reliability and functionality of having data on SNs, which is realized by the security and privacy services of SNs. These services while increasing benefit and decreases costs.

Therefore, in the last decade the researchers and governments have very intensively worked to produce solutions for privacy concerns.

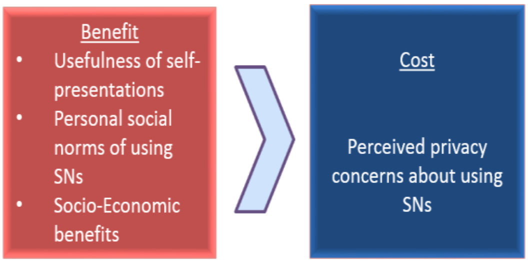
Benefit
Usefulness of self-presentations
Personal social norms of using SNs
Socio-Economic benefits
Cost
Perceived privacy concerns about using SNs

**Fig. 1.** The cost-benefit tradeoff of to be member of SNs.

The following parts address the reasons of privacy concerns, their impacts on users, and approaches of governments and researches to improve the trust to SNs.

## 2. Privacy setting oriented Problems, Privacy Measurement and Solutions

As mentioned above, defining a profile of individual on SN can be imagined as a creation of single node for a social connection graph of this member. Naturally this graph continuously grows with new connected member nodes and with the already exist graphs of those new friend nodes. So, the visibility and sensibility levels of items in a profile setting of member change with the growing social connection graph. Each node of graph has mutual impact on the privacy settings of other nodes and creates an interaction graph for information spreading area. In a timeline, the joining of new nodes or any changes to one of the node's privacy settings continuously affect the privacy risks of individuals in different levels of graph. None of the members can know which new nodes were added or removed from their own friendship graph. Therefore, it is certain that the members should monitor and manage their privacy risk levels on their own social friendship and interaction graph. Furthermore, they should tune their own privacy thresholds according to their privacy preferences. They may get support from specific applications, which have satisfied the following requirements:

1. The privacy settings should have the simplest structure to be easily tuned by members. Members should clearly understand, track and manage the visibility and sensibility degrees of their profile items and posts.

2. Each member can manage her friendship graph, and tune the threshold level between its cost of losing privacy and benefit of having growing friendship network.

3. Naturally some of the nodes on friendship graph can be named as main hubs. So; if one of the nodes increases privacy risk for other connected members; the system gives an alert and members can decide that to keep this node in her graph or not. Members' cost-benefit threshold values have valuable effect on the taken decision.

4. The members require privacy service from the owner of SNs. They could not trust to third-party solutions to privacy and security issues.

5. Due to the nature of SNs, we face with big and dynamic data. Hence, scientists have to do research on the complexity of graph algorithms, to build a solution.

In last decade unfortunately, there are not feasible and reliable applications yet matching all above requirement definitions. Therefore, most people prefer to use default privacy settings assigned by the SNs. There are some studies which calculate the privacy risk to help users on their privacy settings. The primary studies mostly use classification-based models and require training data to produce offline advices for users' profile settings such as the studies in [5, 6]. They try to find the meaningful correlations about the privacy settings of profile items but if the training data cannot support high-accuracy classification models then the data set cannot be suitable with new users' privacy preferences. The study of Talukder et. al. [6] from Purdue University produced Privometer as Facebook tool which measures the privacy risks on the users' privacy settings and interaction graph. Additionally, they use augmented inference model to calculate the information leakage probabilities and its privacy risk ratio where a potentially malicious applications are installed by directly connected friends in the social graph of individual. When the user logs in, the measure of information leakage and self-sanitization recommendations are ready to be viewed. The Privometer only measures the privacy risk from directly connected friends and ignores the risks from FoF in interaction graph of individual. However, in [7], Wang et. al. considers both visibility and sensibility of attributes and also introduces virtual group attributes to consider the effect of privacy setting of the connected FoF's social group members.

By the study of James et al. [8], the "dual privacy decision" idea is proposed. In the study, user has a chance to be information manager to address which information will be spread and to be interaction manager who will be able to see it. They try to handle privacy concerns under two parts. According to these two schemes, James et al. did some tests and shows that socialization, self-expression and pleasing influence information spreading and interaction behaviours on Facebook.

Dong et al. [9, 10] use behavioural predictive model for SN's users and propose a decision-making tool. Tool gives advices to SN user according to her expected preferences to help her in decision making process. The predictive model based on a set of psychological and contextual factors [9, 10], such as trustworthiness of the receiver users, the information sharing tendencies of her, the degree of sensitivity of shared information, the appropriateness of disclosure and some traditional contextual factors. They use binary classification model to measure the influence of each factor. By using this proposed model, people can handle the tradeoff between benefit and risk of information disclosure decision on SNs.

By the study of Aghasian et al. [11], they proposed a framework to measure the privacy disclosure scores (PDS) on SNs. The user's privacy is defined with sensitivity and visibility main factors. They defined a scoring function, which also considers a set of common personal attributes, either with the form of structured (user name, age etc.) or unstructured (posts, comments, photos etc.) data. They first compute the sensitivity degree for each user, using the determined values by Srivastava et al. [12]. Then, they compute the visibility of each user, considering the three factors, respectively: (i) ease of accessibility, (ii) difficulty of data extraction, and (iii) data reliability. After calculating the effect of each three factor, they use a set of fuzzy rules to find the overall visibility score for the attributes of each user. Specifically, they use a fuzzy inference system, based on the Mamdani fuzzy inference [13]. Hence, they measure the sensitivity and visibility of each user, and then they combine these results to calculate users' PDS. Their results showed that users' PDS highly depend on the amount of information, disclosed by the users themselves.

A common point of all these studies is that they work with offline static analysis and give some advices to users for future decisions. However, online privacy risk monitoring and cost-benefit tradeoff management are important requirements for each user because their values change dynamically as mentioned before. If the users can have an application to measure the privacy risks and control the cost-benefit tradeoff value; users can tune their personal setting items and connections according to the online feedbacks of this application. So, the dynamic and real time tools and related models and algorithms have to be studied in details. Another important point is the reliability of these tools. SNs should give these services to users rather than the third party solutions. Also, international common regulations and rules have to organize the qualifications of these services. Users can clearly be sure about the objectivity and quality of these tools.

## 3. Risk of Privacy Disclosure on SNs and Ways to Handle

If we cannot control the spreading of our private information, it can be open to public access and turn into a valuable resource for open source intelligence (OSINT). Essentially, there are two main tools for OSINT; these are *data mining* and usage of *security vulnerabilities and cyber-attacks*;

i. ***Data mining***; users upload different kinds of fragmented data, such as text, photos and videos to SNs' sites. The public, private or third party organizations can integrate and verify them by statistical models and data mining techniques among them to find new relations and patterns which are not known, as well to identify data owner. For instance; the private sector wants to know the product perceptions of customers i.e. complaints and new requirements by tracking their sharing on SNs. The public sector, police and military forces also use open source environment to discover and track the criminal threats, organizations and events. Many organizations prefer to use this very cost effective and beneficial method to reach decisions.

In this environment we expect that the users have to show some reactions to protect anonymity. Bayerl et al. [14] analyse the behavioural tendency of SNs' users. According to their survey; the most of the participants accepts the online surveillance but participants especially concern about threats on their freedom of expression from their governments. The authors confirm a correlation between surveillance awareness and falsification of SNs' users' private attributes. This reality creates a negative effect on the validity of OSINT based outcomes and increases the cost to access more reliable data and results, and requires cross-validation of information by many sources.

ii. ***Security vulnerabilities and cyber-attacks***; is another important reason of privacy disclosure. The third party organizations use illegal methods such as cookies, malicious applications and intrusions to networks or information systems etc. to access or manipulate the private information. SN's users have to be aware of these threats and should take their preventions. However, security assurance is a very complicated problem and cannot be solved with standard user's awareness and prevention methods. Therefore, privacy preservation is a very hot study area for database organization and cryptology. The study of Schwittmann et. al. [15] proposes a social network structure which focuses on user privacy and data availability. This study proposes that all user content should be encrypted and decrypted on end-user devices, and all content are hidden from the SN provider.

The social graph of user should also be hidden from the SN provider. Users authenticate to each other without revealing their identities to a potential attacker. This proposed solution looks like strong enough to decrease privacy concerns of the users.

However, both falsification and crypto solutions increase the costs of data access and analysis. These negatively affect the integration of social and economic worlds in the global scale. The governmental regulations and rules are the missing piece of the solution, which are important to increase the users' trust on the protection of their privacy in SNs.

## 4. Regulations and Trustable Virtual World

The legal certainty, regulations and their international completeness for SNs have prominence to assure trust and have the key role in economic and social development in the global world. As mentioned before; the regulations should be in harmony with the ICT architecture of SNs. For instance, SNs have no borders with the cloud architecture; users cannot be sure about the location of the processing or storage of data. Naturally users expect to control, manage and trace their information with the support of regulations and ICT solutions. Another question in the mind of investors in market is who will tolerate the financial costs to implement rules of privacy regulations. Also the unmatched implementations of these regulations create additional costs for the companies. In 2012, an important solution idea was explained by Deirdre et al. [16] about the privacy regulations of EU and US, both of them update their regulations with the principle of "privacy by design". It is known that the most likely the reason of the security problems of ICT is that their designs are created without security perspectives. This valuable initiative will trigger the new business models to include privacy and other security expectations in the design stage. In 2015, EU[†] also updated its cross-border rules and regulations for the consumers and companies that have trust among them to create single digital market. With this development, the rules that apply to business transactions can be clear and the same rules should be applies to all EU Member States. The different national protection and contract laws discourage companies and users from cross-border activities and prevent them from benefitting of online services, which creates a concentrated expectation to have objective and common rules and regulations in the global scale with reliable implementations.

In 2018 General Data Protection Regulation (GDPR) [17] which came into play on 25th May. According to this regulation SNs, (i) do not keep personal data for longer than they need, (ii) they have to define a policy setting standard with documentation requirements, (iii) they should also periodically review the data they hold, and erase or anonymise it when they no longer need it, (iv) individuals have a right to erasure if they no longer need the data, (v) the "integrity and confidentiality" is defined under the security principle of GDPR. These are only small part of GDPR, the policy definitions and tools for privacy protection is an open and important research area with many critical requirements as mentioned under section 2.

## 5. Conclusion and Future Work

At this current state we see that social networks are eliminating all borders in globe and add new values to our lives. However, users should be aware of the risks of the privacy setting parameters about social networks' profiles. We should know how much risk arises when our friends add new friends to their friend lists. So, the social networks should have some privacy measurement services to support members. For this purpose, the researchers are continuing their studies but they have not proposed an efficient and complete solution yet. From another perspective, if the governments can be more voluntary to improve and sign the global regulations and laws for social networks, they can increase the chance to have more trustable social networks and help us to solve our privacy concerns. On top of that, governments can have powerful economic, social and cultural integrations and more reliable data to take

[†] http://ec.europa.eu/priorities/digital-single-market/docs/dsm-communication_en.pdf

precautions for terrorist attacks and for many other open source intelligence analyses in this era. At last, the humanity never escapes from the challenges or risks to explore new worlds, because the nature of the human strongly forces us to take these risks. Therefore, the humanity will always evolve with the inventive science and its followers as laws, regulations, new technologies and standards. The coming regulations with General Data Protection Regulation in European Union will force all these requirements to increase privacy on social networks.

We can summarize future directions in three layers;

(i) At bottom layer is built by scientific researches. As mentioned above, social networks include big and dynamic data to process. At this area, we need dynamic new data structures and efficient dynamic algorithms to process data and measure privacy leakage possibilities or detect possible risks and malicious real time attacks. Another requirement is encryption of private data on social networks. But, current schemes cannot be feasible for big data due to their computational costs. More practical and light crypto solutions should be studied.

The rise of social networks has led to the rise of unwelcome social bots as automated social actors. Those actors can play many malicious roles including infiltrators of human conversations, scammers, impersonators, misinformation disseminators, stock market manipulators, astroturfers, and any content polluter (spammers, malware spreaders) and so on. It is undeniable that social bots have major importance on social networks. To prevent and resolve these attacks, community detection algorithms and resource detection of these malicious activities on social networks have gained importance. This problem was analysed by Karatas et al. [18], and as future research directions they proposed distributed two new methods as (i) usage of autonomous intelligent based approaches to detect local malicious activities and, (ii) identification based approaches to detect un-legitimate entities.

(ii) At middle layer; we have to build correct and sufficient models to fully represent social networks. Some of the models use SIR information spreading theory [19, 20, 21], and some of them uses information cascading models [22]. However, these models are not sufficient to represent information spreading on social networks. Therefore, the study of Sayin et al. [23] proposed a hybrid model, which focuses on behavioural model and privacy policy of social networks.

(iii) At top of the layer; we have to develop audit and control tools to guarantee the execution of regulations and protect the privacy of users' information on social networks. Some tools should be developed to audit the execution of regulation on social networks. Some others should be developed to give the user management rights of personal information and dynamic message spreading over social networks.

## Acknowledgement

Many thanks for valuable discussions of Asst. Prof. Dr. Selma Tekir and MSc Burcu Sayin.

## References

[1]. H. Hsiao-Ying; M. Bashir, "Is privacy a human right? An empirical examination in a global context," in Privacy, Security and Trust (PST), 13th Annual Conference on, vol., no., pp.77-84, 21-23 doi: 10.1109/PST.2015.7232957, July 2015.

[2]. A. Ho, A. Maiga, and E. Aimeur. "Privacy protection issues in social networking sites", In 2009 IEEE/ACS International Conference on Computer Systems and Applications, pages 271–278, May 2009.

[3]. V. Tuunainen, O. Pitkänen, and M. Hovi, "Users' Awareness of Privacy on Online Social Networking Sites – Case Facebook", BLED 2009 Proceedings. 42, http://aisel.aisnet.org/bled2009/42, 2009.

[4]. J. Min, and B. Kim, "How are people enticed to disclose personal information despite privacy concerns in social network sites? The calculus between benefit and cost", JASIST, 66: 839–857. doi: 10.1002/asi.23206, 2015.

[5]. L. Fang and K. LeFevre. "Privacy wizards for social networking sites", In *Proceedings of the 19th international conference on World wide web* (WWW

'10). ACM, New York, NY, USA, 351-360. doi=10.1145/1772690.1772727 http://doi.acm.org/10.1145/1772690.1772727, 2010.

[6]. N. Talukder, M. Ouzzani, A. Elmagarmid, H. Elmeleegy, “Privometer: Privacy protection in social networks”, vol. 1, no. 2. VLDB Endowment, pp. 141–150, 2010

[7]. Y. Wang, R.K. Nepali, and J. Nikolai. “Social network privacy measurement and simulation”, In Computing, Networking and Communications (ICNC), 2014 International Conference on, vol., no., pp.802-806, 3-6 doi: 10.1109/ICCNC.2014.6785440, Feb. 2014.

[8]. T. L. James, M. Warkentin, and S. E. Collignon. “A dual privacy decision model for online social networks”, Inf. Manage., 52(8):893–908, December 2015.

[9]. C. Dong, H. Jin, and B. Knijnenburg. “Predicting privacy behavior on online social networks” Ninth International AAAI Conference Web and Social Media, April 2015.

[10]. C. Dong, H. Jin, and B. P. Knijnenburg, “PPM: A Privacy Prediction Model for Online Social Networks”, In: Spiro E., Ahn YY. (eds) Social Informatics. SocInfo 2016. Lecture Notes in Computer Science, vol 10047. Springer, Cham, https://doi.org/10.1007/978-3-319-47874-6_28, 2016.

[11]. E. Aghasian, S. Garg, L. Gao, S. Yu, and J. Montgomery, “Scoring users’ privacy disclosure across multiple online social networks”, IEEE Access, 5:13118–13130, 2017.

[12]. A. Srivastava and G. Geethakumari. “Measuring privacy leaks in online social networks”, In 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp 2095–2100, Aug 2013.

[13]. D. Wang, X. Zeng, and J. A. Keane. “A simplified structure evolving method for Mamdani fuzzy system identification and its application to high-dimensional problems”, Information Sciences, 220(Supplement C):110 – 123. Online Fuzzy Machine Learning and Data Mining, 2013.

[14]. P.S. Bayerl and B. Akhgar. “Surveillance and falsification implications for open source intelligence investigations”, Commun. ACM 58, 8, 62-69. DOI=10.1145/2699410 http://doi.acm.org/10.1145/2699410, July 2015.

[15]. L. Schwittmann, C. Boelmann, M. Wander, and T. Weis, “SoNet -- Privacy and Replication in Federated Online Social Networks”, In Proceedings of the 2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops (ICDCSW '13). IEEE Computer Society, Washington, DC, USA, 51-57. DOI=10.1109/ICDCSW.2013.20 http://dx.doi.org/10.1109/ICDCSW.2013.20, 2013.

[16]. K.M. Deirdre and A.B. Kenneth. “What regulators can do to advance privacy through design”, *Commun. ACM* 56, 11, 20-22. DOI=10.1145/2527185 http://doi.acm.org/10.1145/2527185, November 2013.

[17]. Guide to the General Data Protection Regulation (GDPR), https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/ Latest Access Time for the website is 06 June 2018.

[18]. A. Karatas, S. Sahin, “A Review on Social Bot Detection Techniques and Research Directions”, in Proceedings of the 10th International Conference on Information Security and Cryptology (ISCTurkey 2017), Article No. 29, Ankara, Turkey, October 20-21,2017, pp. 156- 161. Accessed on Oct. 30,2017[Online]

[19]. Y. Bao, C. Yi, Y. Xue, and Y. Dong. “A new rumor propagation model and control strategy on social networks”, In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13). ACM, New York, NY, USA, 1472-1473, DOI=http://dx.doi.org/10.1145/2492517.2492599, 2013.

[20]. E. Serrano, C. Á. Iglesias, and M. Garijo, “A Novel Agent-Based Rumor Spreading Model in Twitter”, In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, New York, NY, USA, 811-814. DOI: http://dx.doi.org/10.1145/2740908.2742466, 2015.

[21]. G. Cordasco, L. Gargano, A. A. Rescigno, and U. Vaccaro, “Brief Announcement: Active Information Spread in Networks”, In Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing (PODC '16). ACM, New York, NY, USA, 435-437. DOI: http://dx.doi.org/10.1145/2933057.2933069, 2016.

[22]. Chao Tong, Wenbo He, Jianwei Niu and Zhongyu Xie, “A novel information cascade model in online social networks”, Physica A: Statistical Mechanics and its Applications, Volume 444, p. 297-310. DOI:10.1016/j.physa.2015.10.026, 2015.

[23]. B. Sayin, S. Sahin, “A Novel Approach to Information Spreading Models for Social Networks”, IARIA, The Sixth International Conference on Data Analytics Conference 2017, ISBN: 978-1-61208-603-3 Page 23-27, November 2017.