A visual privacy tool to help users in preserving social network data

Stefano Cirillo[1], Domenico Desiato[2,*], Michele Scalera[2] and Giandomenico Solimando[1]

[1]Department of Computer Science, University of Salerno, via Giovanni Paolo II n.132, 84084 Fisciano (SA), Italy
[2]Department of Computer Science, University of Bari Aldo Moro, via Edoardo Orabona n.4, 70125 Bari (BA), Italy

Abstract

In the current era, social network platforms are increasingly important, especially for disseminating data that refers to virtual lives that, in most cases, are strictly coupled with real ones. For example, social networks permit us to share emotions, and ways of thinking, connect with people worldwide, find a job, etc. However, to have access to the virtual world, users need to register their data that, in most cases, univocally identify themselves. To this end, arise the necessity to make users aware of privacy issues that may occur when such an amount of data spread over social network platforms are mismanaged. In this work, we propose a visual privacy framework that improves the users' awareness concerning disseminating their data over social network platforms. Moreover, we define interactive visual metaphors that permit users to understand which kind of information they share and how to manage information disseminated over different social network platforms.

Keywords

Data wrapping, Data reconstruction, Privacy, Social Networks, Data Analysis

1. Introduction

Social networks interpret a crucial role in human interactions because they enable people to subscribe to multiple contents such as emotions, ways of thinking, points of view, and so on. Moreover, plenty of people have social profiles disseminated over several social network platforms, sharing a vast amount of information. Under this view, preserving users' privacy is challenging for social network platforms since they cannot permit to put at risk the privacy of their users [1].

Users exploit social networks to share information massively, and often, they do not privatize data and are unaware of the privacy threats they can be exposed to. Furthermore, the increasing number of users with social network profiles yields the necessity of monitoring how they manage their privacy, especially when they have multiple social network profiles.

Multiple studies have analyzed data privacy in social network domain [2, 3], but few of them provided tools exploited to improve users' awareness when they share data over social network platforms [4, 5, 6]. In our work, we perform cross-social network analysis over several social network platforms to understand which is the information that is most frequently shared over social networks and that can jeopardize users' privacy [7, 8, 9]. To this end, we define interactive

IS-EUD 2023: 9th International Symposium on End-User Development, 6-8 June 2023, Cagliari, Italy
*Corresponding author.
scirillo@unisa.it (S. Cirillo); domenico.desiato@uniba.it (D. Desiato); michele.scalera@uniba.it (M. Scalera); gsolimando@unisa.it (G. Solimando)
0000-0003-0201-2753 (S. Cirillo); 0000-0002-2455-2032 (M. Scalera); 0009-0000-6627-8820 (G. Solimando)
© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3408/short-s3-08.pdf

visual metaphors that permit users to understand which kind of information they share and how to manage information disseminated over different social network platforms.

In our proposal, we define a visual tool on top of the SOcial Data Analyzer (SODA) proposed in [10]. The latter can find and extract available information of users on different platforms considering only their photos. In particular, SODA allowed us to perform an accurate analysis for revealing privacy threats linked to incorrect usage of data sharing in social networks. Furthermore, (SODA) also allowed us to evaluate the sensitiveness of information shared by users and perform an exhaustive analysis to understand how social networks can reconstruct users' data even if some of them are privatized on other platforms.

The proposed visual tool is independent of the privacy settings offered by social networks since it simulates the search of a real user and retrieves data publicly available in social network profiles. In other words, if a user has privatized specific information over a specific social network, our visual tool is not able to retrieve that information. However, if the user has some information not privatized over different social networks, the proposed tool retrieves such information. Thus, our visual tool can help users in managing privacy settings offered by social network platforms.

In summary, the main contributions of our study are $i)$ a new visual tool capable of managing users' data from different social network platforms, and $ii)$ visual metaphors that permit to have a detailed analysis of users' data extracted from different social networks aiming to evaluate their privacy and improve their awareness concerning privacy threats in social network platforms.

The paper is organized as follows, Section 2 describes related works, whereas Section 3 presents the architecture of the proposed visual tool. Section 4 presents data reconstruction through multiple social networks, and Section 5 describes the experimental evaluation. Finally, conclusions and future research directions are discussed in Section 6.

2. Related work

This section discusses relevant articles in which social network privacy-preservation is addressed to evaluate risks connected to personal user data.

In the context of privacy preservation for sharing data in social network platforms, several approaches define strategies to make users aware of the privacy issues linked to their posted data. In [11], the authors define a new approach for helping social media users to evaluate their privacy disclosure score (PDS). They assess PDS by taking into account user data shared across multiple social networking sites. Besides, they highlight sensitivity and visibility as the main points that significantly impact user privacy to derive the PDS for each user. The proposed approach exploits the statistical and fuzzy systems for specifying potential information loss derived from the PDS. The authors have analyzed data concerning 15 users registered over different social networks (Facebook, ResearchGate, LinkedIn, and Google+) to perform their analysis. The main differences concerning our work are the methodology used for collecting data and the analysis made over them, i.e. the number of examined users and the social networks considered.

Social network data represents a rich source of information, mainly when it characterizes users, and malicious users can jeopardize the user's privacy by performing targeted attacks to recover sensitive information. In [12], the authors define two modes of users' private information disclosure behavior: voluntary sharing and mandatory provision. They exploit the Communication Privacy Management theory to build a framework for explaining the impact of individual characteristics,

context, and benefit-risk ratio on the user's willingness to disclose voluntarily or mandatorily. Authors show that voluntary sharing is more likely to be driven by positive factors, such as perceived benefits, social network size, and customization. Simultaneously, mandatory provision is affected by individual characteristics such as age, privacy policy, and perceived risks. They highlight that perceived risk has less impact on voluntary sharing than previous studies suggested.

Concerning machine learning applications to preserve privacy in social network contexts, in [13], a comprehensive survey of multiple applications of social network analysis using robust machine learning algorithms is reported. In [14], the authors defined a privacy preservation algorithm that incorporates supervised and unsupervised machine learning anomaly detection techniques with access control models. They evaluated the algorithm over real datasets achieving over 95% accuracy using a Bayesian classifier and 95.53% using deep neural networks. In [15] perform a depression analysis using machine learning approaches over Facebook data collected from an online public source. They evaluated the efficiency of their method using a set of various psycholinguistic features. The authors put evidence that their method can significantly improve the accuracy and classification error rate by revealing that the Decision Tree obtains the highest accuracy than other machine learning approaches to discriminate the user's depression.

Finally, a recent study used data from people from social networks to find Multi-SIM subscribers within the same operator or between operators for improving campaigns and churn prediction models of Telecom customers [16].

3. Visual Social Network Privacy Tool

As previously introduced, we have designed a visual interactive tool on the top of the tool SODA [10]. In particular, the tool combines the effectiveness of SODA with a new tool named Profil3r, which is capable of finding the URLs of people's profiles on different social network sites, websites, and web applications[1]. Profil3r is an OSINT tool that can be executed through a command-line interface. However, these types of tools can be challenging to use, especially for non-expert users, since they cannot provide direct and clear feedback due to the lack of graphical interfaces. For example, a command-line program can be complex because it requires learning the correct syntax of the command, which often needs several parameters.

In this paper, we have chosen to integrate a lite version of SODA, limited to reconstructing information only from Facebook and Instagram, with the tool Profil3r that, on the other hand, is limited to finding the URLs related to a user on different social network sites and websites starting from a few basic information. It is important to notice that the original version of Profil3r cannot extract information from the URLs linked to a user. This functionality has been integrated into Profil3r through the use of SODA. However, the SODA requires as a mandatory input an image of a user and/or general information such as his/her name or surname to work correctly. Without these inputs, the SODA is not able to operate.

Figure 1 shows an overview of the architecture of the proposed tool. As we can see, the tool starts from a set of specified data according to the input parameters defined by Profil3r. Then, it performs a first-level search on the web to find URLs to the social profiles of the user who requested the analysis. After completing the analysis, we filtered the URLs extracted by Profil3r to obtain only those related to Instagram and Facebook. By starting from these, it is possible

[1]Official Repository: https://github.com/Greyjedix/Profil3r

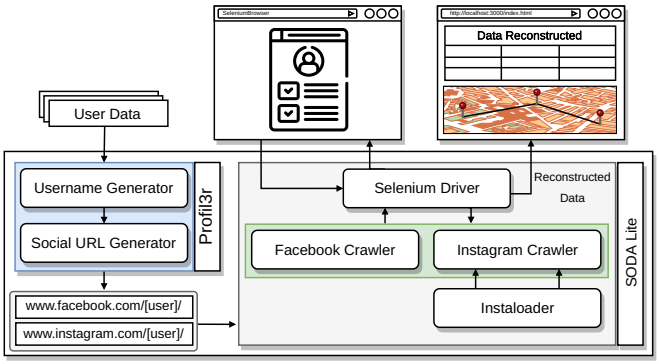

Figure 1: Overview of the architecture underlying the proposed visual social network privacy tool.

to execute the lite version of SODA. It is important to notice that we have re-designed all the input modules of SODA to work only using a link to a user's profile. More specifically, SODA receives both Instagram and Facebook URLs and is able to visit these web pages and extract publicly available user information from web pages using two focused crawlers, i.e., *Facebook* and *Instagram Crawler*, respectively. Furthermore, the proposed tool exploits the Instaloader[2]. framework to extend the set of information that can be reconstructed. In fact, by exploiting Instaloader, the proposed tool can retrieve new information from a public profile, such as hashtags, user stories, geotags, and captions of the posts. Finally, the extracted information is displayed through an interactive interface to help users properly manage their social network data.

4. Data reconstruction through multiple social networks

This section presents a cross-social evaluation to show the tool's effectiveness in analyzing sensitive data shared on various social networks. The collected data and experimental evaluation of the analyzed user data and the performance of the proposed tool in terms of extrapolated attributes are presented below. The experimental evaluation involved a set of real users who were unaware of privacy threats liked to the sharing of information over social networks. All users involved in our experimentation have used the toll only for a personal purpose with full awareness of its potential functions. Through the use of the proposed tool, a user is able to understand the information that can be reconstructed from social, despite any privacy requirement.

Figure 2 shows the interface defined for the proposed visual privacy tool, which is provided to the users for evaluating their privacy. In the upper part of Figure 2, the user can decide which social s/he wants to analyze by selecting Facebook, Instagram, or both. Based on the selected choice, s/he provides his/her data, such as first name, last name and username, in order to access the web platform. Submission of data will lead to the execution of forms that are based on *Profil3r*, which are capable of finding the URLs of users' profiles on Facebook and Instagram, respectively. Following its execution, the user selects the link to his/her account.

User information is identified and collected by executing a light version of *SODA*. The latter is able to visit the web pages and extract publicly available user information from web platforms using two crawlers. The extracted information belong to the various informative section of the platforms. For example, Facebook could provide data concerning work, education, the place lived,

[2]Official Repository: https://github.com/instaloader

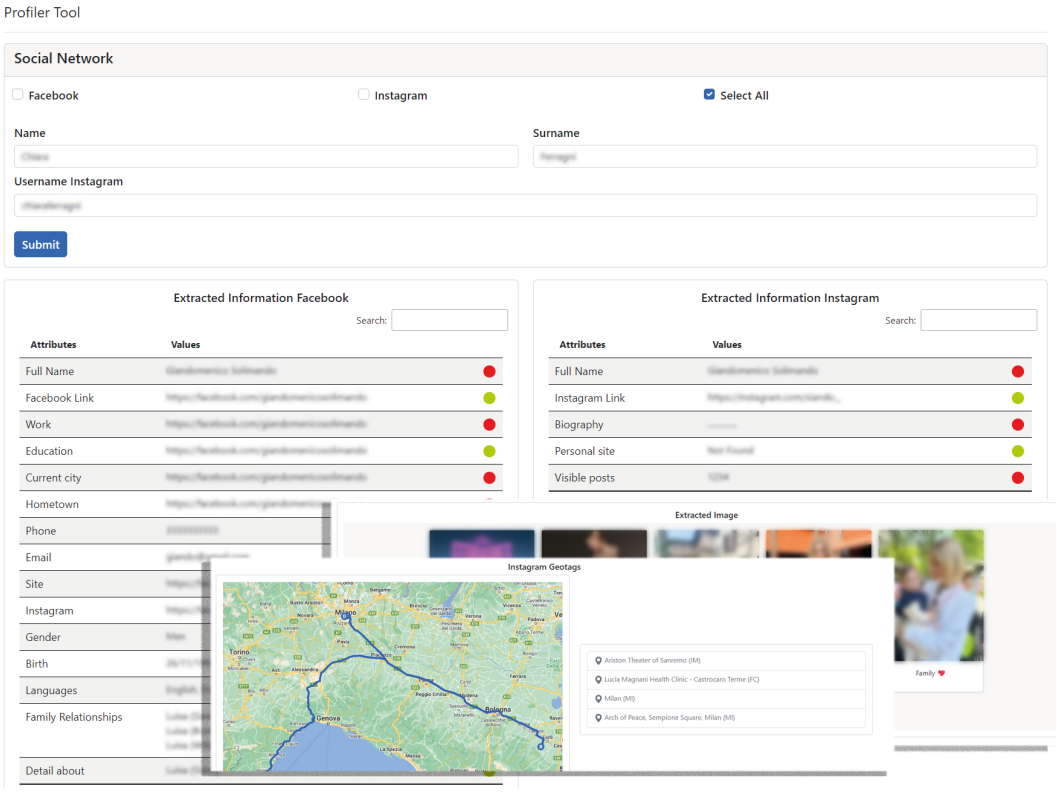

Figure 2: Overview of the interface of the proposed visual social network privacy tool.

family, relationships and personal contact, whereas Instagram could provide data concerning biography, personal site and publicly visible posts. The data are shown in tables to help users easily view the publicly available information extracted. Moreover, to help users to identify sensitive information, visual labels are employed to determine if the extracted data could violate users' privacy, either on an individual or aggregated level. At the bottom of Figure 2 are provided to the user two additional sections containing the posts and publicly available comments extracted via *InstaLoader* together with locations where the posts were defined. The latter exploits an interactive geographical map to show the user a history of the places visited by him/her.

5. Experimental Evaluation

This section reports an experimental evaluation for verifying the effectiveness of the proposed data reconstruction tool. In particular, we conducted a user study involving several participants. The user evaluation was performed in a research laboratory, where users accessed a pre-configured computer having the proposed tool installed. The study consisted of three phases: an initial survey, a task to be addressed, and a final survey. The initial survey aimed to assess the following aspects: (i)how much users concern about security and privacy, (ii) what behaviour the users adopt for sharing information on social networks, and (iii) what of the shared data the users consider to be sensitive. Instead, the objective of the final survey was to evaluate the participants' experiences and opinions about the proposed tool. In particular, we involved 10 participants for the user study, comprising individuals with different ages, educational backgrounds, and levels of social media

usage. Moreover, the study involved students and employees of the University of Salerno. These participants were informed about the study's objectives and methods, and before participating, they gave their informed consent.

Participants were given access to the tool and guided to provide limited personal information, such as their name and surname. The tool then explored publicly available data from different social networks to reconstruct the public information of the participants.The task submitted to users lasted 5 minutes. The experiment started with explaining the task and the tool to the participants. Then, they were introduced to the purpose of the experiment, i.e. understand users awareness through the tool. In addition, they were given a release attesting that they were aware of the purpose of the experiment and the possible reconstruction of sensitive information. Once the preliminary phase was completed, users were asked to log in using their social login information to extract data shared on the platforms.

After using the tool, participants were asked to complete a final survey consisting of Likert scale questions and open-ended prompts. The Likert scale questions assessed participants' perceptions of the tool's usefulness. The open-ended prompts allowed participants to provide qualitative feedback on their experience and suggest potential improvements. The purpose of this survey was to collect feedback from participants about the tool's effectiveness, their satisfaction with the reconstructed data, and their willingness to use the tool in the future. Moreover, participants were encouraged to provide additional feedback or suggestions for enhancing the tool's performance. It is important to notice that the initial and the post-task questionnaires share several questions, aiming to monitor whether the users' privacy perception changed after using the proposed tool.

The collected survey was analyzed using both quantitative and qualitative methods. The Likert scale responses were subjected to statistical analysis to determine the average ratings for each aspect of the tool. Open-ended responses from the initial and final surveys were analyzed using thematic analysis to identify common patterns within the participants' feedback.

The experiment revealed that users shared their concerns after utilizing our tool. In particular, they noted that certain information they initially deemed non-sensitive in the first survey was able to jeopardize their privacy. Moreover, users highlighted that the tool could be used to support users in understanding how personal information is spread and how it can be reconstructed from different social platforms. Finally, based on the latest survey results, many users expressed concern about the tool's capability to track their visited locations through Instagram posts.

6. Conclusion

In our work, we defined a visual social network privacy tool that helps users to manage their data over social network platforms. In particular, we performed a cross-social evaluation concerning users' data to help them figure out the sensitivity of their data. In the future, we would like to collect more data concerning users by integrating information over other social networks.

Acknowledgments

This Publication was produced with the co-funding of the European union - Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3 – Partnerships extended to universities, research centers, companies and research D.D. MUR n. 341 del 5.03.2022 – Next Generation EU (PE0000014 - "Security and Rights In the CyberSpace - SERICS" - CUP: H93C22000620001).

References

[1] M. T. Baldassarre, V. S. Barletta, D. Caivano, A. Piccinno, M. Scalera, Privacy knowledge base for supporting decision-making in software development, in: Sense, Feel, Design: INTERACT 2021 IFIP TC 13 Workshops, Bari, Italy, August 30–September 3, 2021, Springer, 2022, pp. 147–157.

[2] M. Teresa Baldassarre, V. Santa Barletta, D. Caivano, A. Piccinno, Integrating security and privacy in hcd-scrum, in: CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter, 2021, pp. 1–5.

[3] L. Caruccio, D. Desiato, G. Polese, Fake account identification in social networks, in: 2018 IEEE international conference on big data (big data), IEEE, 2018, pp. 5078–5085.

[4] S. Cirillo, D. Desiato, B. Breve, Chravat-chronology awareness visual analytic tool, in: 2019 23rd International Conference Information Visualisation (IV), IEEE, 2019, pp. 255–260.

[5] B. Breve, L. Caruccio, S. Cirillo, D. Desiato, V. Deufemia, G. Polese, Enhancing user awareness during internet browsing., in: ITASEC, 2020, pp. 71–81.

[6] V. S. Barletta, G. Desolda, D. Gigante, R. Lanzilotti, M. Saltarella, From GDPR to privacy design patterns: The MATERIALIST framework, in: S. D. C. di Vimercati, P. Samarati (Eds.), Proceedings of the 19th International Conference on Security and Cryptography, SECRYPT 2022, Lisbon, Portugal, July 11-13, 2022, SCITEPRESS, 2022, pp. 642–648.

[7] D. Desiato, G. Tortora, A methodology for gdpr compliant data processing., in: SEBD, volume 2161, 2018, pp. 1–4.

[8] L. Caruccio, D. Desiato, G. Polese, G. Tortora, Gdpr compliant information confidentiality preservation in big data processing, IEEE Access 8 (2020) 205034–205050.

[9] L. Caruccio, D. Desiato, G. Polese, G. Tortora, N. Zannone, A decision-support framework for data anonymization with application to machine learning processes, Information Sciences 613 (2022) 1–32.

[10] F. Cerruto, S. Cirillo, D. Desiato, S. M. Gambardella, G. Polese, Social network data analysis to highlight privacy threats in sharing data, Journal of Big Data 9 (2022) 1–26.

[11] E. Aghasian, S. Garg, L. Gao, S. Yu, J. Montgomery, Scoring users' privacy disclosure across multiple online social networks, IEEE access 5 (2017) 13118–13130.

[12] K. Li, L. Cheng, C.-I. Teng, Voluntary sharing and mandatory provision: Private information disclosure on social networking sites, Information Processing & Management 57 (2020) 102128.

[13] T. Balaji, C. S. R. Annavarapu, A. Bablani, Machine learning algorithms for social media analysis: A survey, Computer Science Review 40 (2021) 100395.

[14] R. Aljably, Y. Tian, M. Al-Rodhaan, Preserving privacy in multimedia social networks using machine learning anomaly detection, Security and Communication Networks 2020 (2020).

[15] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, A. Ulhaq, Depression detection from social network data using machine learning techniques, Health information science and systems 6 (2018) 1–12.

[16] N. R. Al-Molhem, Y. Rahal, M. Dakkak, Social network analysis in telecom data, Journal of Big Data 6 (2019) 1–17.