

Basics of Privacy on Social Networks

Julián Salas

Internet Interdisciplinary Institute (IN3)

Universitat Oberta de Catalunya (UOC)

Center for Cybersecurity Research of Catalonia (CYBERCAT)

Barcelona, Spain

jsalaspi@uoc.edu

Abstract—Social network analysis brings many benefits for understanding our society. At the same time, it reveals a lot about ourselves, from our habits to our personality traits. These characteristics may also be used to re-identify our data and possibly disclose private attributes.

In this tutorial, we review some of the past attacks to privacy resulting from nave anonymization (of medical records, AOL internet search queries, Netflix movie ratings and NYC-taxi geo-located data) to motivate privacy enhancing techniques and present some measures to prevent attribute and identity disclosure (namely k-anonymity and differential privacy).

We focus on social network anonymization and discuss the trade-off between the risk of disclosure and the utility loss. We explain generic information loss measures (such as centrality, betweenness, average path length, etc.), together with task-specific information loss measures such as community detection, and how to relate them with the risk of disclosure.

We discuss different networks that may arise from recommender systems or mobility data, and their specifics regarding privacy protection.

We comment privacy by design strategies, based on the legal framework, to cover possible industrial needs. Finally, we present the emerging problems that arise in the privacy field when considering dynamic data.

I. AUTHORS' SHORT BIO

Julián Salas is currently a postdoctoral fellow at the Internet Interdisciplinary Institute (IN3) from Universitat Oberta de Catalunya (UOC) and member of the Cybersecurity Research Centre of Catalonia (CYBERCAT).

He has previously been a postdoctoral researcher at the CRISES group of the University Rovira i Virgili (URV) and at the Artificial Intelligence Research Institute-National Research Council (IIIA-CSIC). He received his Ph.D degree in Applied Mathematics at the Universitat Politècnica de Catalunya - BarcelonaTech (UPC) with a European mention in 2012.

His research interests are on privacy preserving data mining, social network anonymization and big data privacy, with special focus on dynamic data.

II. TARGET AUDIENCE AND PREREQUISITES

The target audience for this tutorial is all those researchers interested in understanding the basic concepts and the motivations for privacy protection on social networks. No particular background is expected from the audience.

III. OUTLINE OF THE TUTORIAL

- Medical record de-anonymization (20 min)
k-anonymity and related privacy definitions
- AOL search data de-anonymization (20 min)
Privacy by design concepts
- Netflix de-anonymization (20 min)
Collaborative filtering anonymization
Network anonymization
Privacy vs Utility tradeoff
- NYC-taxi de-anonymization (20 min)
Mobility data anonymization
Differential privacy
- Open problems for dynamic data privacy (20 min)

REFERENCES

- [1] J. Salas and J. Domingo-Ferrer, Some basics on privacy techniques, anonymization and their big data challenges, *Math. Comput. Sci.*, 2018.
- [2] A. Narayanan and V. Shmatikov, De-anonymizing social networks, in *Proceedings - IEEE Symposium on Security and Privacy*, 2009.
- [3] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, Unique in the Crowd: The privacy bounds of human mobility, *Sci. Rep.*, vol. 3, 2013.
- [4] G. Danezis et al., *Privacy and Data Protection by Design - from policy to engineering*, 2015.
- [5] P. Samarati and L. Sweeney, Protecting Privacy when Disclosing Information: k-Anonymity and its Enforcement Through Generalization and Suppression., *Proc IEEE Symp. Res. Secur. Priv.*, 1998.
- [6] N. Li, T. Li, and S. Venkatasubramanian, t-Closeness: Privacy Beyond k-Anonymity and l-Diversity, in *2007 IEEE 23rd International Conference on Data Engineering*, 2007, pp. 106115.
- [7] C. Dwork and A. Roth, The Algorithmic Foundations of Differential Privacy, *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 34, pp. 211407, 2014.
- [8] B. Zhou, J. Pei, and W. Luk, A brief survey on anonymization techniques for privacy preserving publishing of social network data, *ACM SIGKDD Explor. Newsl.*, 2008.
- [9] E. Zheleva, E. Terzi, and L. Getoor, Privacy in Social Networks, *Synth. Lect. Data Min. Knowl. Discov.*, 2012.
- [10] B. Zhou and J. Pei, Preserving Privacy in Social Networks Against Neighborhood Attacks, in *2008 IEEE 24th International Conference on Data Engineering*, 2008.
- [11] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis, Resisting Structural Re-identification in Anonymized Social Networks, *Proc. VLDB Endow.*, vol. 1, no. 1, pp. 102114, 2008.
- [12] K. Stokes and V. Torra, Reidentification and k-anonymity: A model for disclosure risk in graphs, *Soft Comput.*, 2012.
- [13] J. Soria-Comas and J. Domingo-Ferrer, Big Data Privacy: Challenges to Privacy Principles and Models, *Data Sci. Eng.*, 2016.
- [14] Clifton, C., Tassa, T.: On syntactic anonymity and differential privacy. *Trans. Data Priv.* 6(2), 161183 (2013)