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 naive anonymization (of medical records, AOL internet search queries, Netflix movie ratings and NYC-taxis 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 Center 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 Politcnica 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