Chapter 33

Protecting Privacy by Secure Computation:

Privacy in Social Network Analysis

Florian Kerschbaum
SAP Research Karlsruhe, Germany

Daniel Funke
SAP Research Karlsruhe, Germany

ABSTRACT

We consider collaborative social network analysis without revealing private inputs of the participants. This problem arises in criminal investigations of federal police organization where single organizations may not reveal their data without probable cause, but the aggregation of all data entails new information, such as the entire social network structure. We present algorithms for securely computing either the entire, anonymized graph or only specific metrics for individuals. We use secure computation protocols to disclose nothing, but the output of the analysis, i.e. anything that cannot be derived from one's input and output – including other parties' input – remains private. We have implemented a prototype for SAP's investigative case management system – a derivate of its customer relationship management.

INTRODUCTION

In federated states or organization of states, such as the European Union or the United States, a mutual approach to organized crime is necessary. For this purpose, federal law enforcement agencies, such as Europol or the FBI, have been established. Nevertheless, data privacy laws or simply data governance concerns restrict institutions from sharing their data, unless there is hard corroborating evidence on a case and subject under investigation. In particular, in the European Union (EU, 1995) data privacy is regarded as a high social and political value and the dilemma on how to generate evidence without violating privacy laws is evident.

A common tool for the criminal investigator is social network analysis. It graphically depicts the suspects and their connections to other people or artifacts, such as telephone numbers or bank

DOI: 10.4018/978-1-4666-2625-6.ch033

accounts, and allows the computation of certain metrics. Not all the facts composing the entire picture of a case may be known to one investigator. In particular, in pan-European organized crime, local police forces may only be aware of a partial view of the picture.

This necessitates data exchange between the institutions, but European data privacy laws prohibit data exchange without probable cause and in excessive amounts. Therefore we propose a solution where the local investigator or an investigator at the superordinate institution has access to all information, but without revealing sensitive or private details. This allows the investigator to still use SNA and profit from its achievements without breaking individual privacy rights or guidelines of other institutions.

Privacy-Preserving SNA has been suggested in the literature before, but we have found the solutions to be insufficient for the requirements of our scenario. In (Frikken and Golle, 2006) a fully anonymized version of the social network is computed. This does not allow the investigator to track his suspect anymore and he cannot gain additional information or collect evidence about him. In (Canny, 2002) a recommendation value metric is proposed, suitable for privacy-preserving calculation. However, investigators are used to centrality metrics they are trained on, such as betweeness and closeness (Xu and Chen, 2005).

In this chapter we will present

- an algorithm that computes the entire social network from distributed sources without revealing personally identifying information while keeping track of the local view of each party.
- an algorithm to compute the important centrality metrics of betweeness and closeness without revealing personally identifying information and without revealing the entire social network.

The first algorithm, called "Compute Entire Network", allows the investigator to gain an overview of the entire social network. He can compute metrics on his subject as possible with the second algorithm, but he also gains additional information about the entire structure of the criminal organization. The second algorithm, called "Compute Metrics", provides higher privacy guarantees, as it does not even reveal the entire social network (except its size), but still allows important metrics to be computed about the subject. These metrics allow the identification of the role the subject is playing within the criminal organization (Xu and Chen, 2005).

RELATED WORK

SNA has been used for criminal investigations for a long time (Harper and Harris, 1975; Sparrow, 1991; Xu and Chen, 2005). Recent research suggests using graphical tools and investigates the impact of SNA (Xu and Chen, 2005). We can conclude that SNA is a widely accepted tool in criminal investigations.

Privacy-Preserving SNA has been first proposed by Frikken and Golle (2006). They compute an anonymized graph of the social network, such that no one should be able to track their position in the graph. They allow for certain modifications of the correctness of the anonymized graph in order to prevent tracking of one's position, e.g. they may bound the number of incoming connections or apply similar restrictions.

While this provides strong privacy guarantees it does not match the requirements of our scenario. An investigator intends to gather additional information to his present view of the social network. It is therefore unacceptable to anonymize his view, but the goal is to augment it with additional information about the entire network.