

Quantifying Privacy Loss of Human Mobility Graph Topology

Dionysis Manousakas¹, Cecilia Mascolo^{1,2}, Alastair R. Beresford¹, Dennis Chan¹,
and Nikhil Sharma³

¹University of Cambridge

²The Alan Turing Institute

³UCL

Abstract

Human mobility is often represented as a mobility network, or graph, with nodes representing places of significance which an individual visits, such as their home, work, places of social amenity, etc., and edge weights corresponding to probability estimates of movements between these places. Previous research has shown that individuals can be identified by a small number of geolocated nodes in their mobility network, rendering mobility trace anonymization a hard task. In this presentation we build on prior work and demonstrate that even when all location and timestamp information is removed from nodes, the graph topology of an individual mobility network itself is often uniquely identifying. Further, we observe that a mobility network is often unique, even when only a small number of the most popular nodes and edges are considered. We evaluate our approach using a large dataset of cell-tower location traces from 1500 smartphone handsets with a mean duration of 430 days. We process the data to derive the top- N places visited by the device in the trace, and find that 93% of traces have a unique top-10 mobility network, and all traces are unique when considering top-15 mobility networks. Since mobility patterns, and therefore mobility networks for an individual, vary over time, we use graph kernel distance functions, to determine whether two mobility networks, taken at different points in time, represent the same individual. We then show that our distance metrics, while imperfect predictors, perform significantly better than a random strategy and therefore our approach represents a significant loss in privacy

Keywords

Mobility privacy; De-anonymization; k -anonymity; Graph kernels.