

Anonymized Data Linkability via Approximate Graph Matching

Proposed by :

Dionysis Manousakas
University of Cambridge

dm754@cam.ac.uk

Abstract

Privacy concerns of users have led data publishers to anonymize personal data using independent pseudonyms. For example, mobility recorded on devices is often released as a sequence of states that have been pseudonymised independently per device. This excludes the possibility of crossreferencing the states among different devices. However, it leaves untouched the structural patterns in individual users mobility routines.

This study will aim to study the linkability of such datasets over large populations, in two respects:

- Given multiple anonymized representations of the mobility of one user across different observation windows, can we match states that correspond to the same location?
- Given a population of such anonymized representation, can we match profiles that correspond to the same user?

To answer these questions we will rely on graph representations of the profile and approximate graph matching/network alignment techniques.

Keywords

network alignment, graph isomorphism, generative models for graphs, graph embeddings, metric learning, Wasserstein distance, optimal transport, data deanonymization