UNIVERSITY TRANSPORTATION CENTER
RESEARCH BRIEF

Privacy Risk Evaluation of Human Mobility Data for Urban Transportation Planning

Jan Whittington, PhD

Background

Human mobility analysis has many applications, ranging from urban planning to transportation to public health. These analyses generally rely on large datasets storing spatiotemporal points visited by individuals in urban areas (i.e., GPS traces). The emergence of new mobility options and location-based services has led to an unprecedented surge in mobility data sources. While the new sources offer opportunities to discover subject-level knowledge and expand fields of inquiry, they also allow the re-identification of individuals, thus raising privacy risk by revealing intimate information about persons. For example, de Montjoye et al. has shown that only four spatiotemporal points in a day are needed to uniquely identify 95% of individuals in mobility datasets, and reporters from the New York Times illustrated this ease and associated privacy impacts by identifying persons, including US senators, from just such a dataset. To readers of this widely publicized exposé, it is obvious that mobility data may be analyzed with malicious intent, with serious consequences for the persons identified through the data.

Research Project

Practitioners in the public and private sector have yet to resolve the pressing issue of how to condition (e.g., aggregate, suppress, modify) data to be shared or published to ensure its usefulness in analysis while preventing the re-identification of persons represented in the data. Built on literature of the re-identifiability of persons from mobility data (e.g., De Montjoye et al. 2013), the purpose of this project is to contextualize privacy risk in the built environment and to place the resulting probabilities into a tool for public agencies. This project builds upon existing studies and addresses the problems by testing three hypotheses.

1. Linking human mobility data with built environment data increases the uniqueness of human mobility traces, thus increasing the privacy vulnerability of mobility datasets (i.e., probability of re-identification of individuals);
2. Due to the differences in travel patterns, privacy vulnerability is heterogeneous across built environments, categorized by urban density, land use mixture, and property value, and user groups, which may be categorized by age, gender, and amount of travel;
3. Built environment types require different approaches to preserve privacy (i.e. different levels of generalization and suppression) to optimize the tradeoffs between privacy protection and loss of information utility in transportation planning and operations.

ABOUT THE AUTHORS

The research team consisted of Jan Whittington of the University of Washington.

ABOUT THE FUNDERS

This research was funded by the Pacific Northwest Transportation Consortium, with additional support from the University of Washington.

EXPECTED DATE OF COMPLETION

August 2022

FOR MORE INFORMATION

http://depts.washington.edu/pactrans/research/projects/privacy-risk-evaluation-of-human-mobility-data-for-urban-transportation-planning/