

Comparing Twitter and LODES Data for Detecting Commuter Mobility Patterns

Project 2037
May 2021

Jochen Albrecht, PhD, GISP
Andreas Petutschnig
Laxmi Ramasubramanian, PhD, AICP
Bernd Resch, PhD



Local and regional planners struggle to keep up with rapid changes in mobility patterns. This exploratory research is framed with the overarching goal of asking if and how geo-social network data (GSND), in this case, Twitter data, can be used to understand and explain commuting and non-commuting travel patterns. Statistics capturing human mobility are expensive to obtain and deteriorate quickly as existing mobility patterns change and new ones emerge. Planners have been relying on US Census LODES data, which explicitly captures only commuting trips, and seems unsatisfying because only some 16.6% of all vehicle trips are work-related (FHWA 2017). GSND potentially offers a solution, as data derived from repeat origin-destination pairs of the same Twitter ID indicate trips regardless of purpose.

We set out to answer the following research questions:

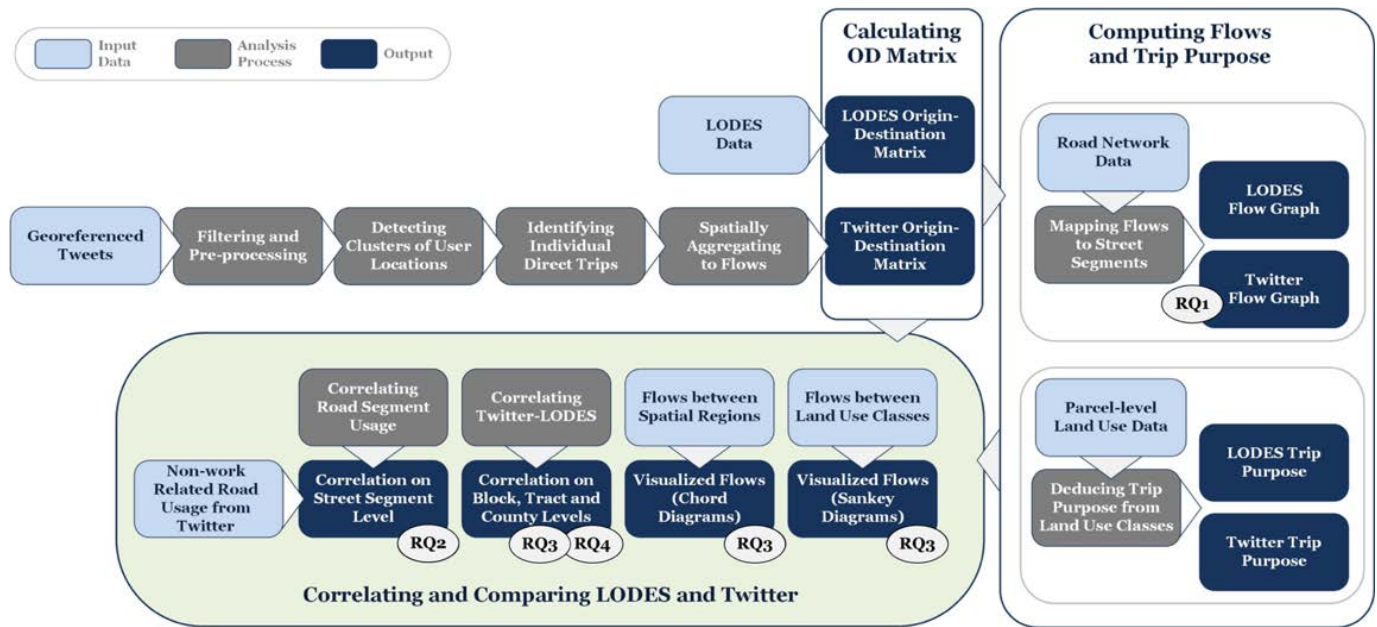
1. Is it possible to extract travel flow patterns in the Bay Area from GSND and if so, how can this be done efficiently?

2. To what degree do commuter flow patterns identified in GSND correlate with official LODES commuting data?

3. Can GSND be used to explain non-commuting trips?

Study Methods

Approximately 33 million geo-referenced Bay Area tweets were harvested for the study period from 2010 until early 2020. They were filtered by repeat occurrences of origin/destination (O/D) pairs and categorized by time of day and day of week. Each of these pairs, as well as all LODES O/D's were then routed as shortest paths on the Open Street Maps network of roads. This study is limited to road trips only; further research should apply routing procedures for transit trips as well. For the GSND, we attributed trip purpose by the dominant land use in the O/D census blocks. We then compared the road segment loads of the two input data sets and found not only incredible high rates of correlation but also nearly complete spatial randomness among their differences, which suggests that the findings



below are scale-independent and applicable in all parts of the study region. Twitter's 2015 geolocation policy change resulted in a dramatic reduction of available GSND. Since then, smaller temporal samples have shown to be a poor predictor of local traffic loads.

US Census LODS data is highly representative for all road segment loads regardless of trip purpose.

Findings

GSND are suitable to capture the over 80% of non-commuting trips that keep our roads busy. GSND is not suitable for characterizing real-time or short-term commuting patterns but is complementary to existing commuting data. Translated into road segment loads, GSND and LODS data are virtually indistinguishable, which means that LODS data are an excellent substitute for overall transportation demand.

Policy/Practice Recommendations

The research project set out to determine whether GSND may be used to augment LODS data beyond commuting trips and whether it may serve as a short-term substitute for commuting trips. It turns out that the reverse is true and the common practice of employing LODS data to extrapolate to overall traffic demand is indeed justified.

Figure 1. Schematic project work flow illustrating input data, analytical steps, and output.

This means that expensive and rarely comprehensive surveys are now only needed to capture trip purposes. Regardless of trip purpose (e.g., shopping, regular recreational activities, dropping kids at school), the LODS data is an excellent predictor of overall road segment loads.

About the Authors

Drs. Laxmi Ramasubramanian and Jochen Albrecht are professors at the SJSU and Hunter College, CUNY, respectively. Dr. Bernd Resch is a professor at the University of Salzburg, Austria, as is his PhD student Andreas Petutschnig. Aleisha Wright is a MA student in the SJSU Urban and Regional Planning program.

To Learn More

For more details about the study, download the full report at transweb.sjsu.edu/research/2037



MTI is a University Transportation Center sponsored by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology and by Caltrans. The Institute is located within San José State University's Lucas Graduate School of Business.