MTI Research Snaps presents:

# **Twitter to Model Traffic Flows**







#MTIResearchSnaps

Presented by: Dr. Laxmi Ramasubramanian and Dr. Jochen Albrecht



17 June 2021

### Acknowledgments

- Dr. Bernd Resch, Associate Professor, University of Salzburg
- Andreas Petutschnig, PhD candidate, University of Salzburg
- Aleisha Wright, MUP student, SJSU
- Mineta Transportation Institute

### **Research Motivation**

• Can geo-social network data be used to model the flow of people?

• Can we identify commuter flows using geosocial network data?

• Can we identify other types of regular flows?

• At what spatial scale can we match and augment existing flow data?

### **Geo-Social Network Data**

• Twitter, Facebook, Instagram, etc.

- Location derived from tweets
  - Explicit: user allows sharing of tweet lat/lon
    - Pre April 2015
    - 90% drop since 2015
  - Implicit: semantic analysis of tweet content augmented by gazetteers



### **Understanding Geospatial Tweets**

- Usefulness of tweet locations is a function of
  - Spatial resolution (address / census tract / county)
  - Temporal extent (day / week / month/ year / multi-year)
- As with Census data, researchers must compromise one or the other
- Scope of analysis influenced by tradeoffs between spatial specificity and temporal extent
- Presently, geospatial Twitter data is *not* abundant enough to conduct detailed short-term analyses such as capturing the effect of a temporary road or bridge closure

### **Volume of Geo-Spatial Tweets by County**



### **Authoritative Data Sources To Measure Flows**

- US Census-based
  - 1. <u>Census Transportation Planning Package</u>
  - 2. <u>LEHD Origin-Destination Employment Statistics</u>
- Both provide Origin-Destination (OD) pairs
  - CTTP: Tract- and TAZ-level; many attributes; latest version 2012-16
  - LEHD: block-level; no further attributes; latest version 2018
- 3,252,286 block-to-block LODES flows ≡ commutes

### All 3 Million + LODES Routes

Sonoma

Napa

Solano

Alameda

Marin

Contra Costa

100 km

#### San Francisco

50

There are 9 levels of line thickness. The lowest level are roads that have up to 420 riders or fewer on them. The fattest lines represent roads with 45 to 94K commuters on them. In [parentheses], we have the

number of observations in each of the nine categories.

### San Mateo

75

Santa Clara

All LODES trips [962386] 1 - 420 [894876] 420 - 1560 [45071] 1560 - 3567 [12374] 3567 - 6828 [4992] 6828 - 11540 [2442] 11540 - 18083 [1386] 18083 - 29096 [773] 29096 - 45242 [328] 45242 - 94548 [144] NineCounty

Google Satellite

### **Project Workflow – calculating OD matrix**



### **OD** Pairs in Twitter and LODES

- Built a O-D dataset using our Twitter dataset to make it as comparable as possible to the LODES dataset
  - 1. 44,812,476 georeferenced tweets (Oct 2010 Apr 2020)
  - 2. Filter and pre-processing (removing automated tweets)
  - 3. Clusters of user locations by Twitter-ID
  - 4. Identifying individual direct trips
  - 5. Spatially aggregating these to block-to-block flows (to mirror LODES)
  - 6. Creating a Twitter O-D matrix

### County-to-County Movements

Twitter rush hour Mo-Fr 6-8 AM and 3-5 PM Twitter outside weekday rush hours

LODES (commutes)



## **Commuting Flows and Trip Purpose**

- We distinguish trips by
  - Rush hour (Mo-Fr 6-8 AM, 3-5 PM)
  - Weekday non-rush hour
  - Weekends
- Twitter trips are a lot shorter than LODES trips (3 minutes versus 14)
- Trip purpose derived from land use to be discussed a little later
  - Twitter trips are more commonly residential to residential

### **Non-commute Trips**





Weekends

.. But not during weekday rush hour

14

Non-rush hour weekdays



### **Deriving Trip Purpose from Land Use**



### Conclusions

- 542K OSM road segments
- 136K used by LODES
- 116K used by Twitter
- Map shows tiny portions of road segments, where Twitter and LODES don't correlate well



Thank you for joining us for: Twitter to Model Traffic Flows

View the full report at: <u>https://transweb.sjsu.edu/research/2037-Detecting-Commuter-Mobility-Patterns</u>



@MinetaTransportation

#MTIResearchSnaps

Tune in for our next MTI Research Snap "Getting There with Google's Via2G and the Future of Commuting" on July 22, 2021 at 10a.m. (PT)! Visit <u>https://transweb.sjsu.edu/events</u> for details and registration.

Have a suggestion for a webinar topic you'd like to see featured? Email irma.garcia@sjsu.edu

