Pacific Northwest
NATIONAL LABORATORY

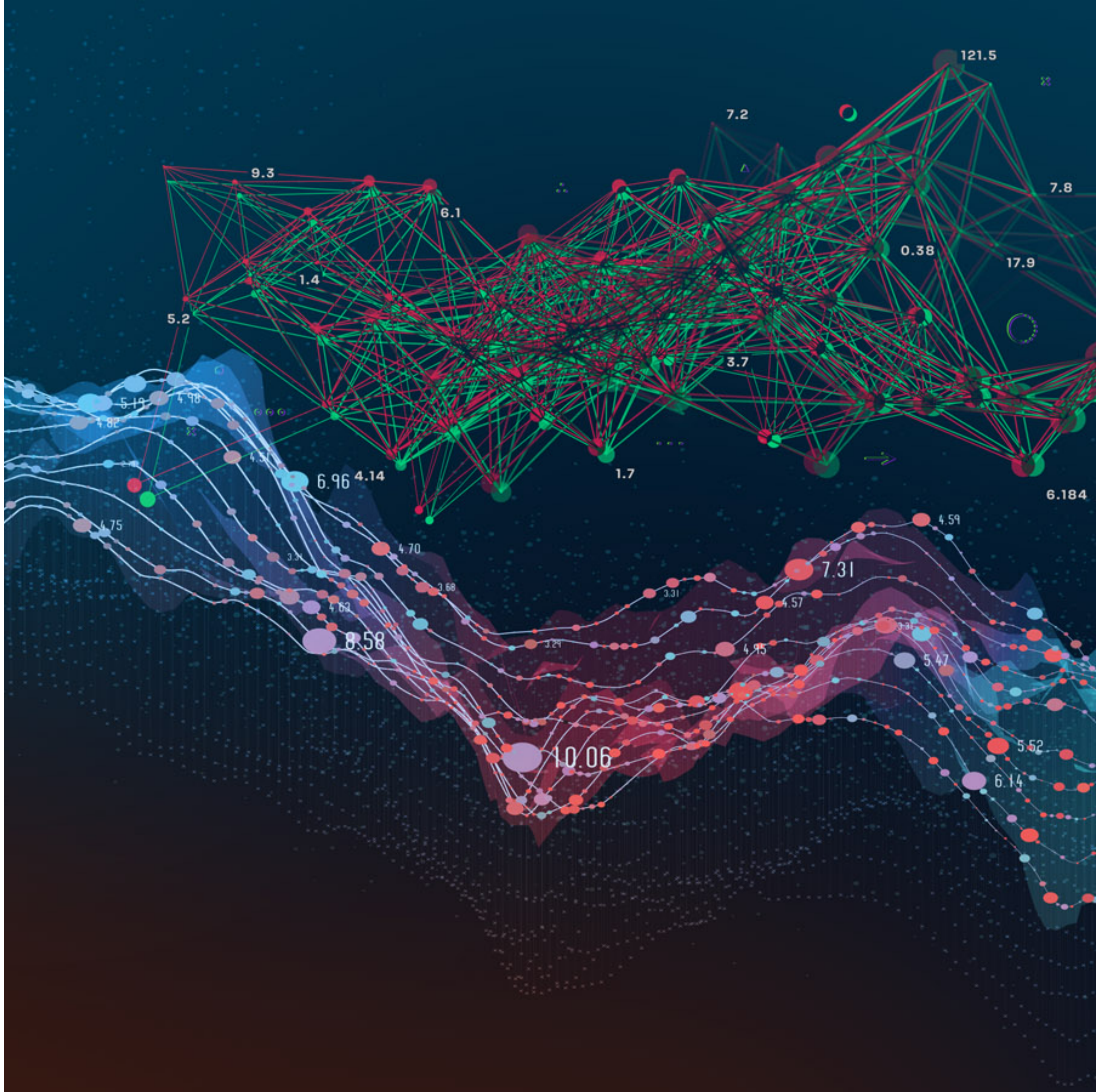
Street-level Travel-time Estimation using sparse and coarse data

Arif Khan, Kelsey Maass, Arun V. Sathanur, Robert Rallo and Katie Wolf

July 14, 2020

U.S. DEPARTMENT OF ENERGY BATTELLE

PNNL is operated by Battelle for the U.S. Department of Energy



Goal

- Two major areas of traffic forecasting:
  - Demand modeling (number of trips between O-D pairs)
  - **Travel-time estimation** (travel time between O-D pairs)

- Due to data availability, there is more research on freeways than on **arterials**.
  - Many sources but **sparse**.

- **Goal:** Develop a method to leverage open (sparse) data sets to estimate arterial **street-level** travel times on a metropolitan road network.

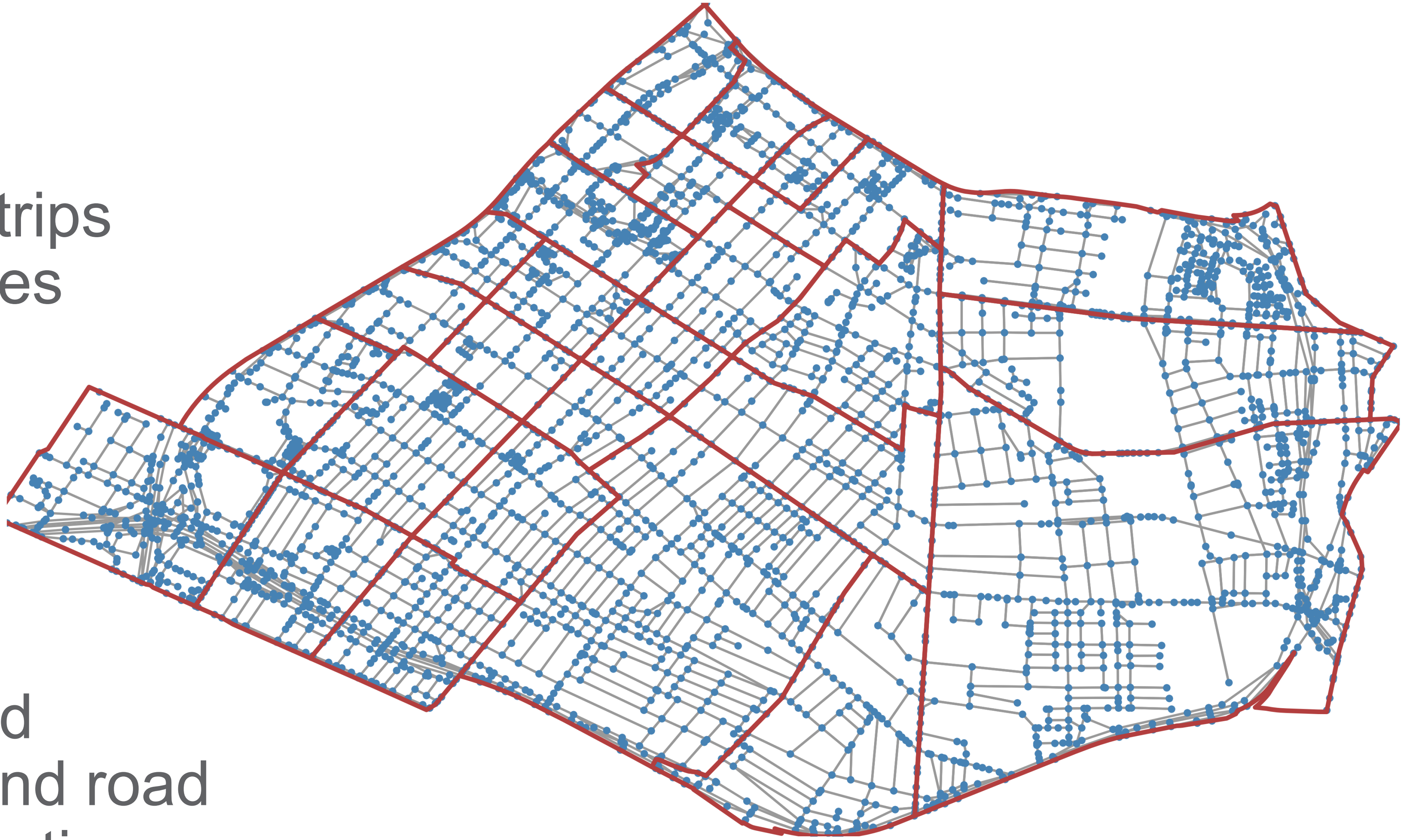
Data Sources

Uber Movement

- Summary statistics of Uber trips between traffic analysis zones (TAZ)
- movement.uber.com

OpenStreetMap

- Graph representation of road network with intersections and road segments represented as vertices and edges
- openstreetmap.org

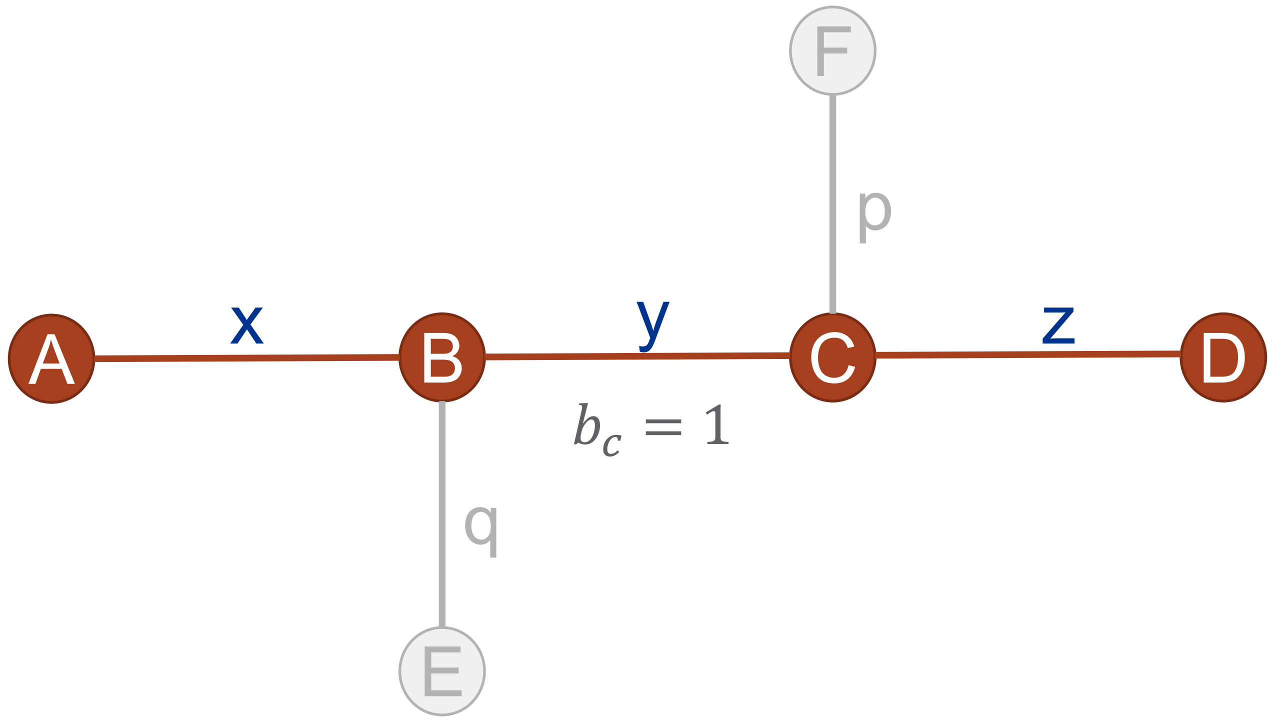


Road network for 27 TAZs in downtown Los Angeles

Pacific Northwest
NATIONAL LABORATORY

Edge Travel-time Estimation

	A	B	C	D	E	F
A				25		
B						
C						
D						6
E						13
F						



Solutions:

x = 12, y = 1, z = 12

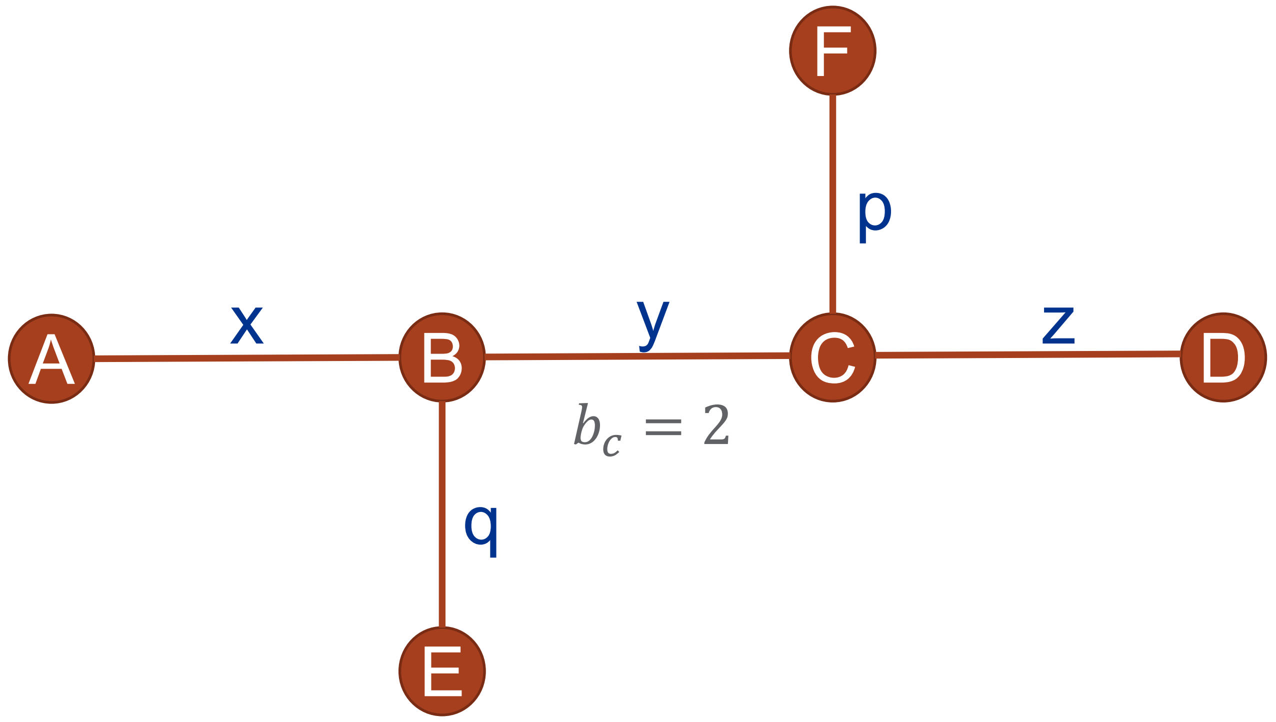
x = 1, y = 23, z = 1

y = [1, 23]

$$\begin{cases} x + y + z = 25 \\ p + z = 6 \\ q + y + p = 13 \end{cases}$$

Edge Travel-time Estimation

	A	B	C	D	E	F
A				25		
B						
C						
D						6
E						13
F						



Solutions:

$x = 12, y = 1, \quad z = 12$

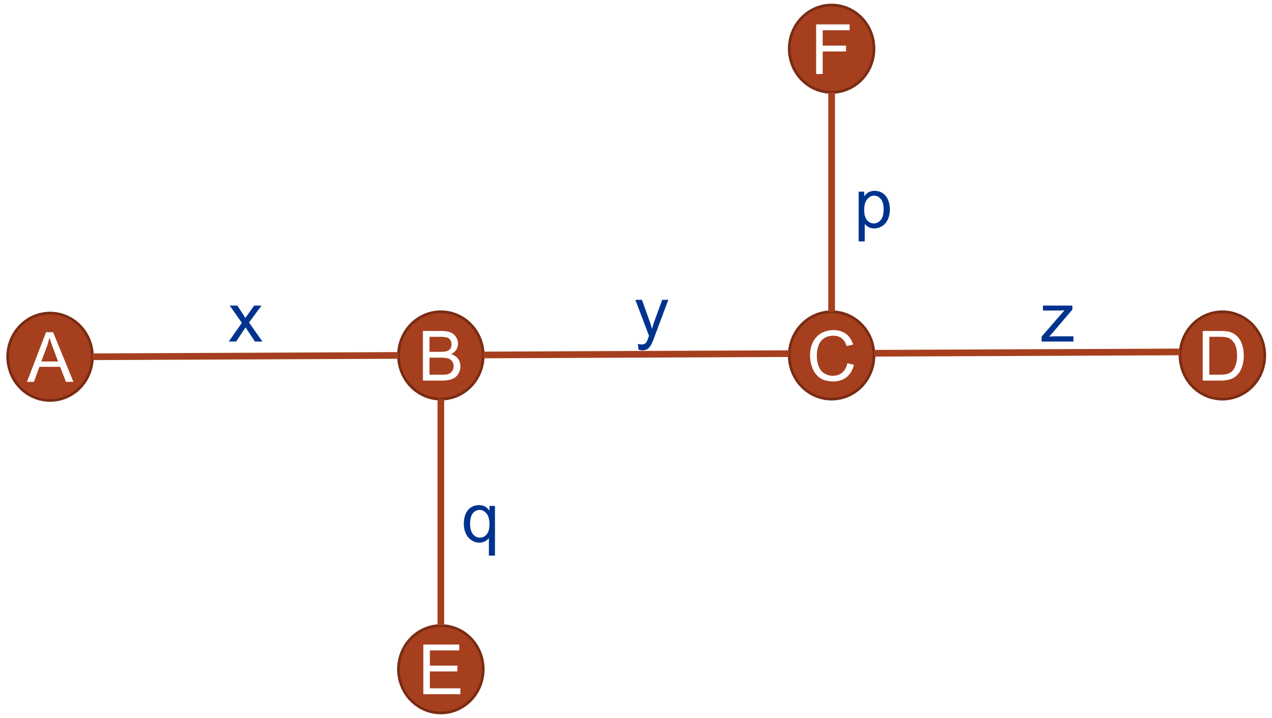
$q = 1, \quad y = 11, p = 1$

$y = [\,1, 11\,]$

$$\begin{cases} x + y + z = 25 \\ p + z = 6 \\ q + y + p = 13 \end{cases}$$

Edge Travel-time Estimation

	A	B	C	D	E	F
A		9	20	25	10	21
B			11	16	1	12
C				5	12	1
D					17	6
E						13
F						



Solution: x = 9, y = 11, z = 5, p = 1, q = 1

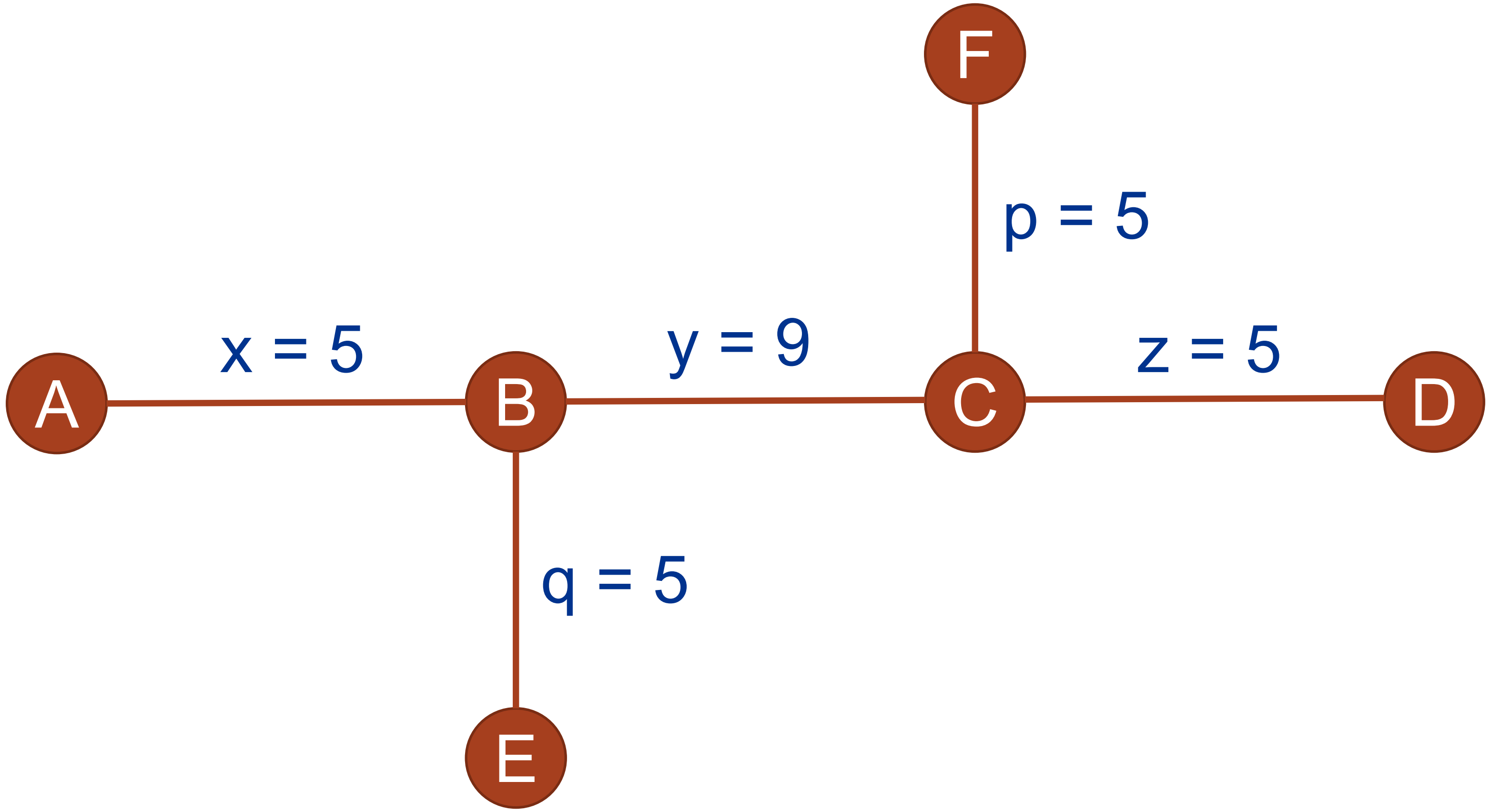
N ⟷ # of trips ⟷ # of equations

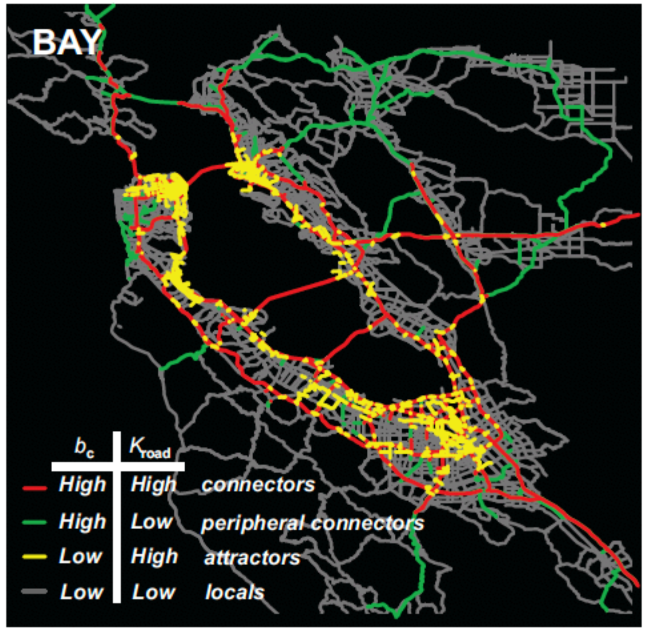
M ⟷ # of road segments ⟷ # of variables

$$\begin{cases} x + y + z = 25 \\ p + z = 6 \\ q + y + p = 13 \end{cases} \quad \longleftrightarrow \quad S^T t \approx y$$

Edge Betweenness (b_c) Centrality

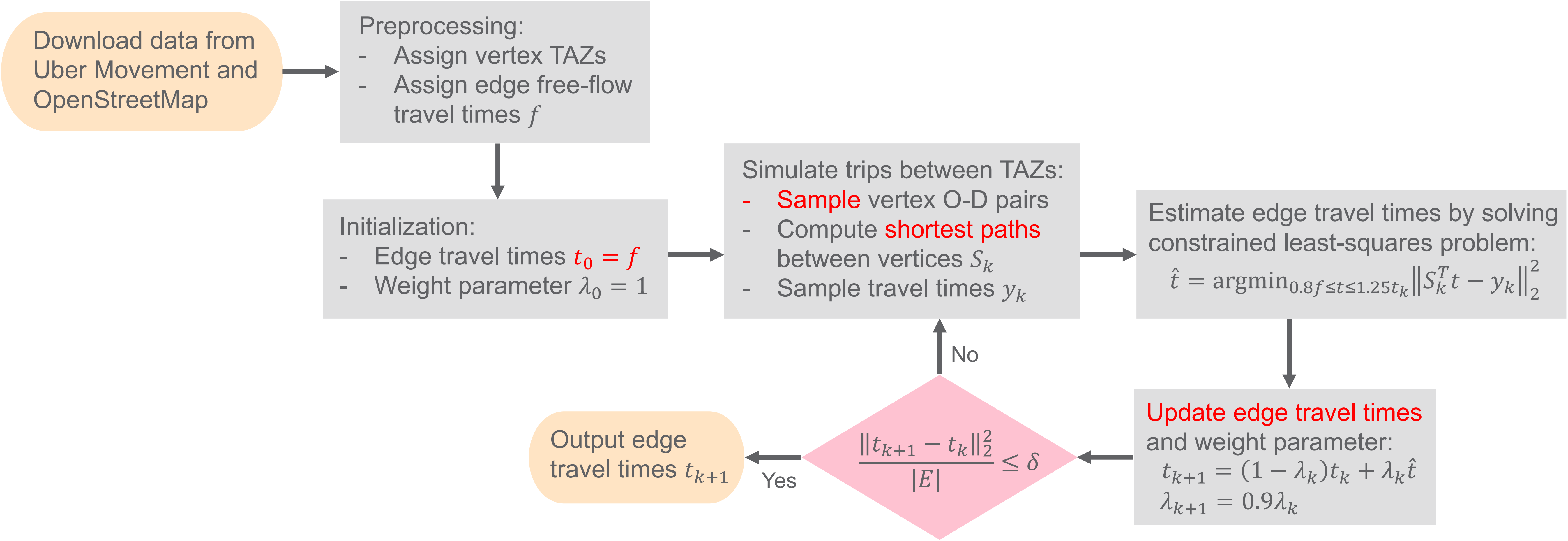
The betweenness of an edge is the number of shortest paths that go through it for all possible vertex O-D pairs.





Types of roads based on betweenness centrality (b_c) and degree (k_{road}). (Source: Wang et al. *Sci. Reports*, 2012, 2, 1001)

Project Workflow



Experimental Results

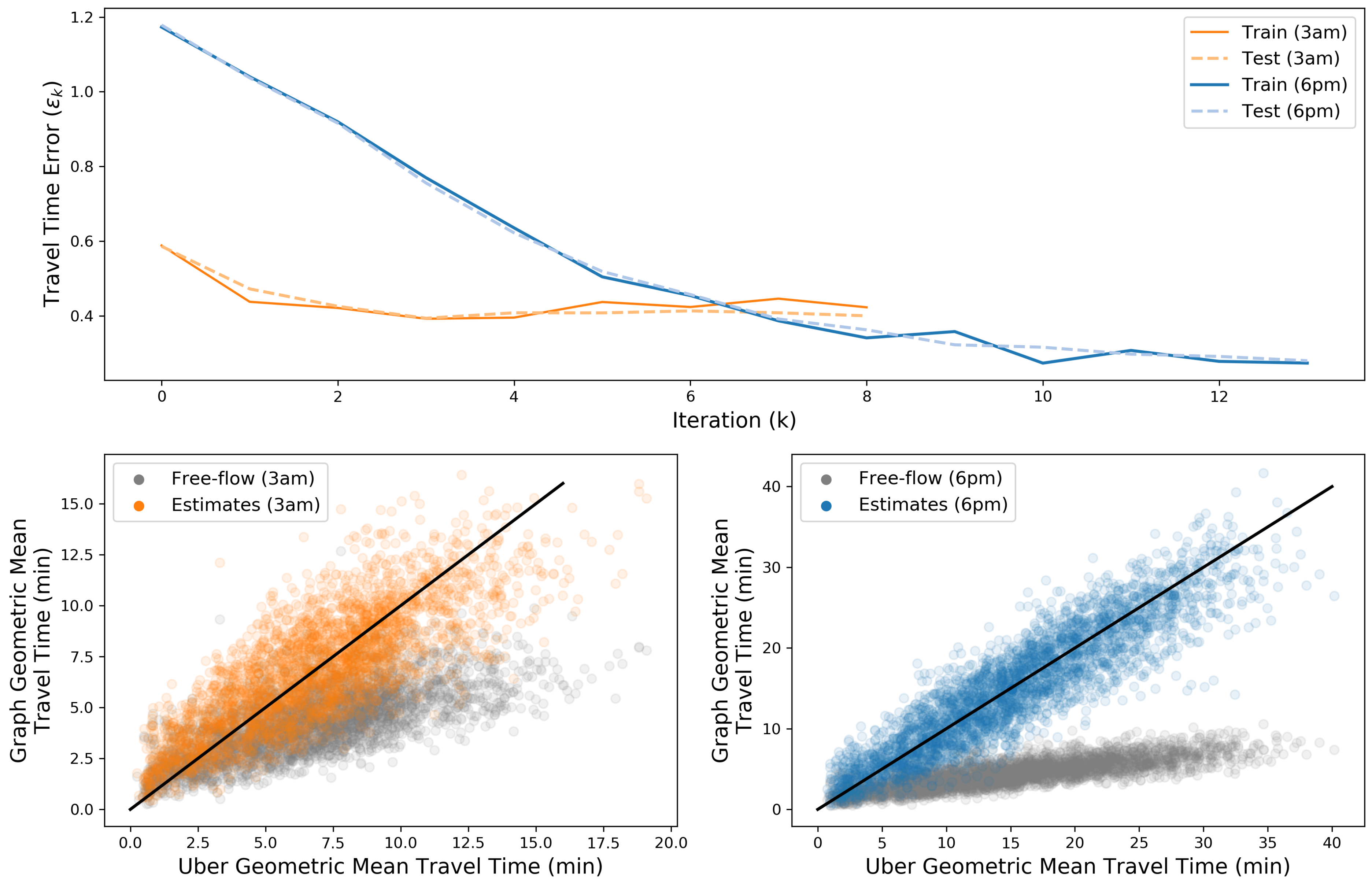
We estimated edge travel times for downtown LA graphs using

- Different times of day:
  - 3am = 507,449 TAZ O-D pairs
  - 6pm = 1,115,432 TAZ O-D pairs
- Different graph sizes:
  - 1-mile ($M = 7{,}138$), 2-mile ($M = 22{,}756$) and 3-mile radii ($M = 40{,}929$)
- Different number of sampled trips:
  - N = 1,000, 3,000, and 5,000

Error measured by $\epsilon_k = \sqrt{{}^1/_N \sum_{(i,j)\in\mathcal{U}_k} n_{ij}\left(\log g_{ij}(t_k) - \log G_{ij}\right)^2}$

Pacific Northwest
NATIONAL LABORATORY

Experimental Results: Convergence



Initial Results on metro-scale estimation

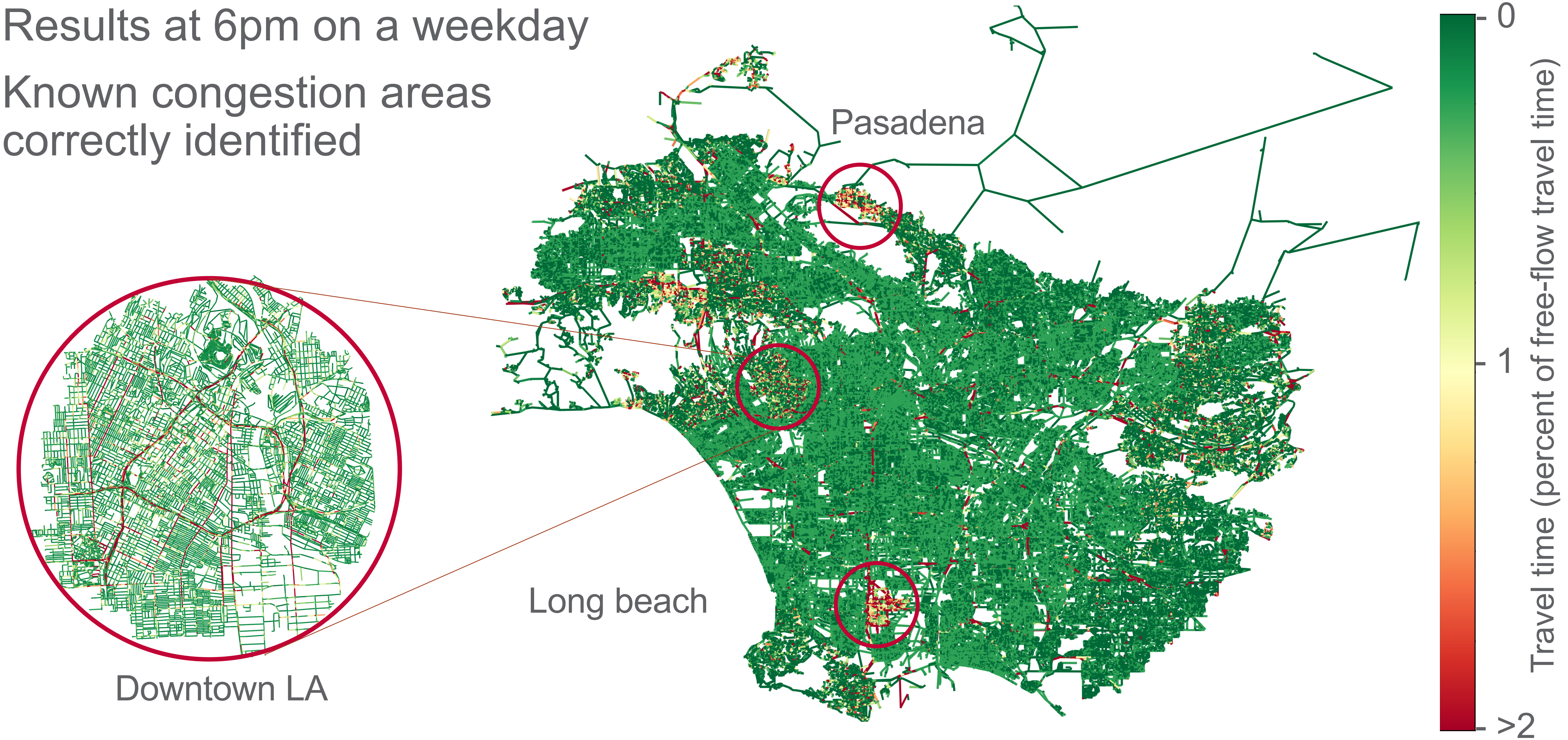
- LA and associated Metro Areas (160k intersections / 470k links – no sparsification):
  - Total coverage - Approx. 1200 sq. mi
  - 64 chunks @ approx. 20 sq. mi. per chunk
  - Total running time on a single (8-core) server node : 18.5 hours per hourly window
  - With 64-node cluster ~ Approx. 25 mins (including serial overhead)

- Seattle and associated Metro areas (/ 160k links – no sparsification)
  - Total coverage – Approx. 500 sq. mi
  - 16 chunks @ approx. 30 sq. mi per chunk
  - Total running time on a single (8-core) server node : 5.00 hours per hourly window
  - With 16-node cluster ~ Approx. 20 mins (including serial overhead)

- Current approach cannot be used to drive-down chunk sizes as it affects accuracy
  - Imposes a limit on scaling gains
  - Projections for a 64-node cluster (12K Processors) are of about 30 mins (including overhead)
  - Potential areas of performance improvement: Python code, constrained least squares solver

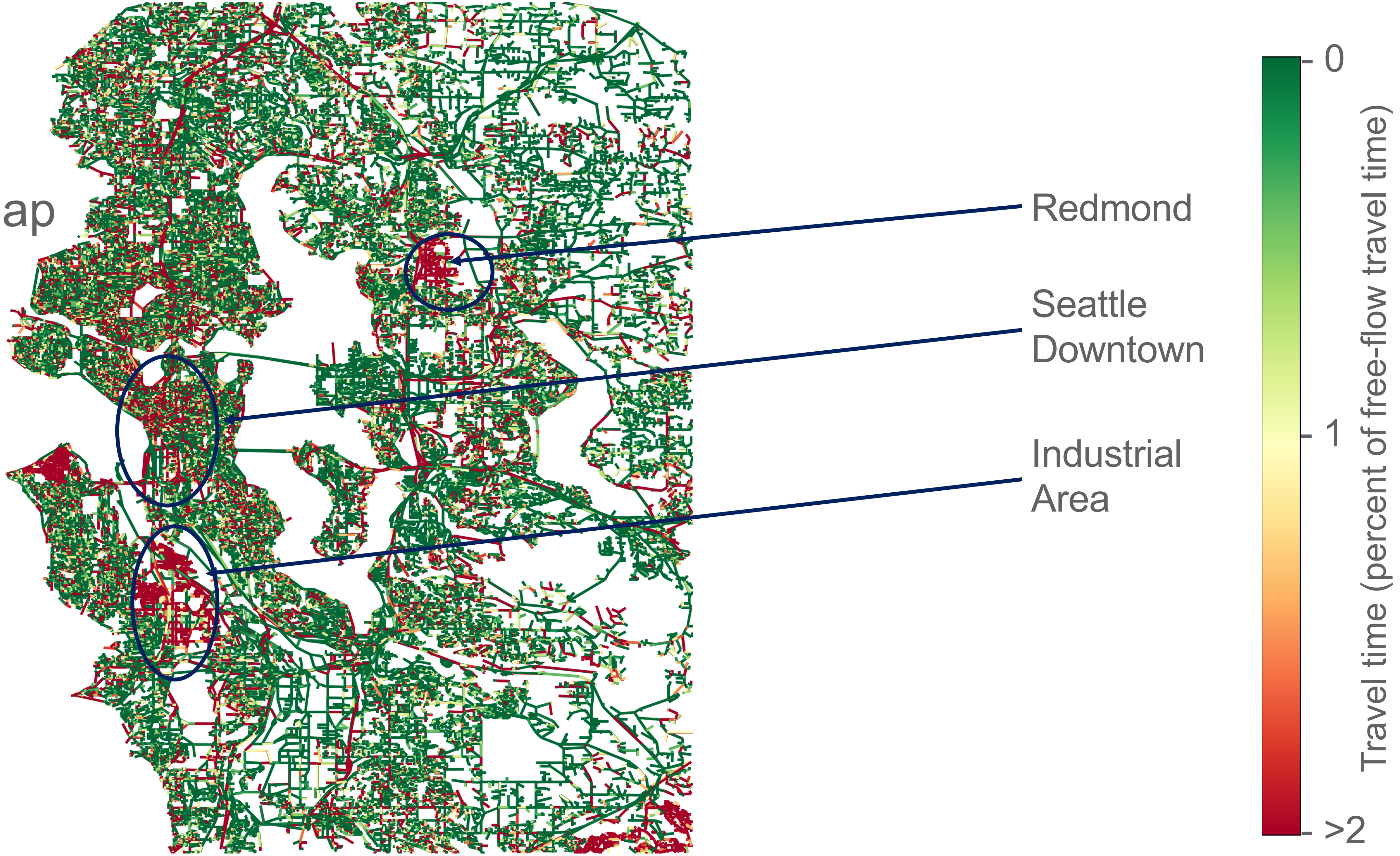
Visualization of congestion on LA metro-area map

- Results at 6pm on a weekday
- Known congestion areas correctly identified



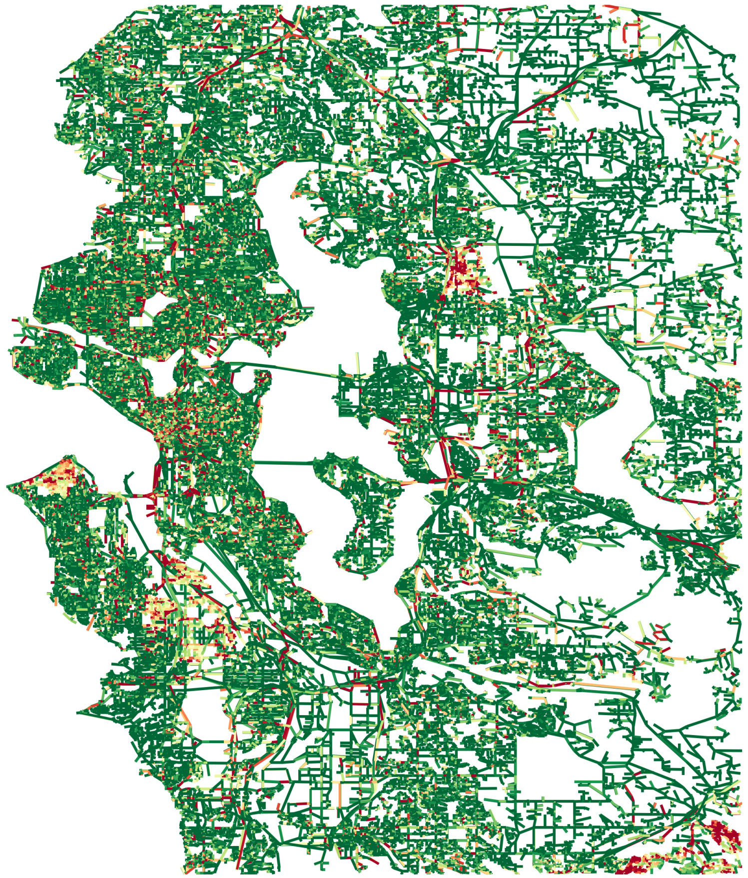
Visualization of congestion on Seattle metro-area

- 6PM
- Weekday
- OpenStreet Map

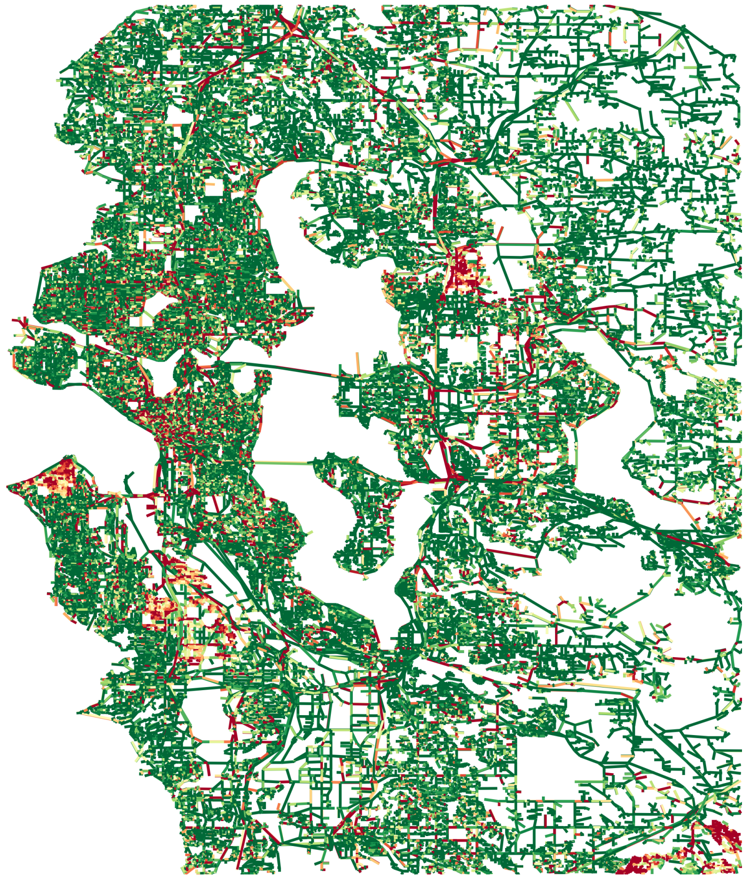


Time evolution of the Seattle network 4pm-6pm

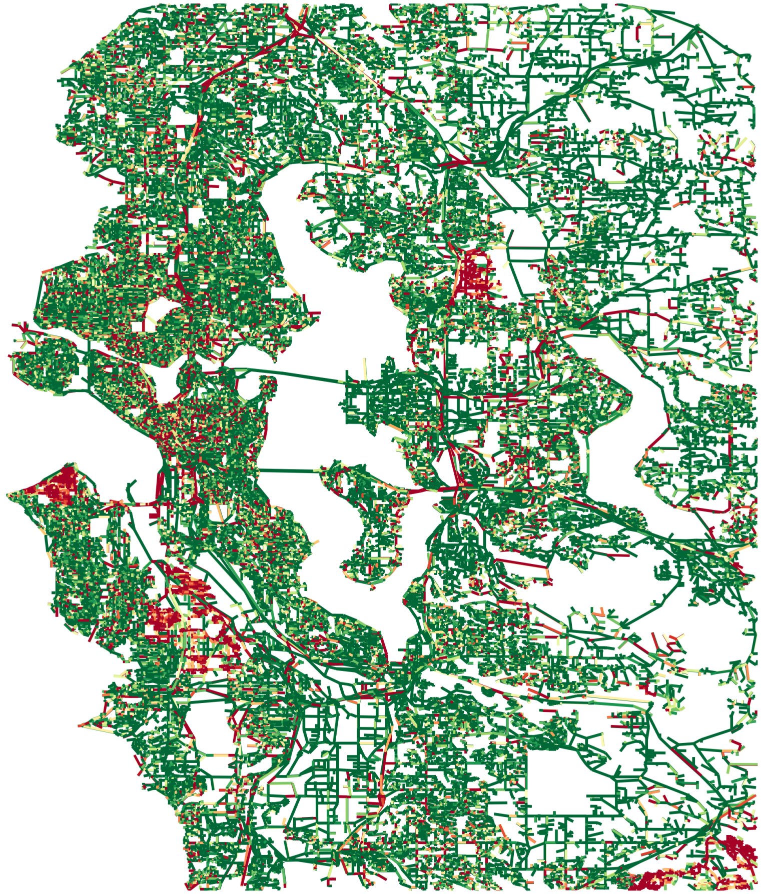
Increase in traffic congestion → larger travel times in congested arterials



4pm



5pm

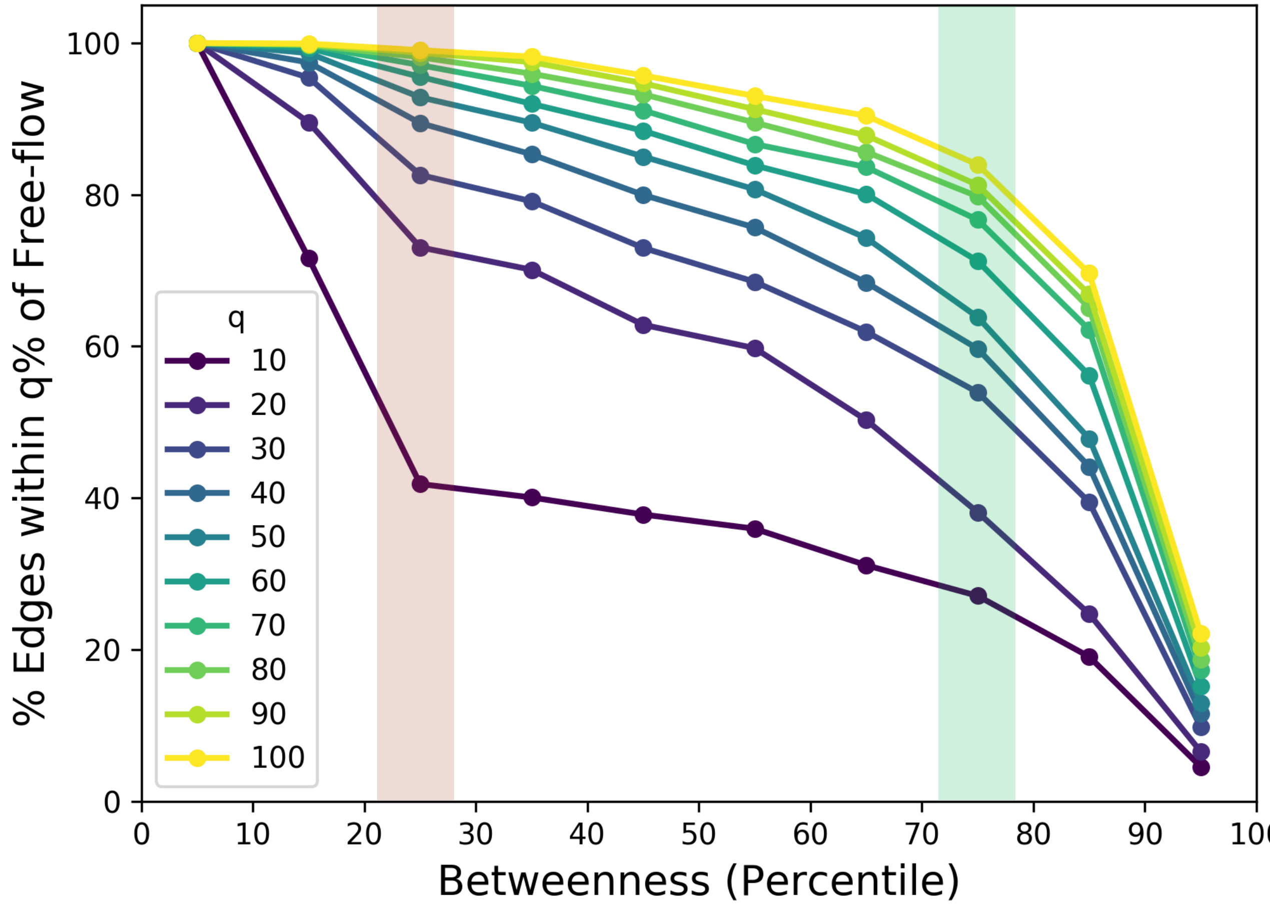


6pm

Graph Pseudo-Sparsification

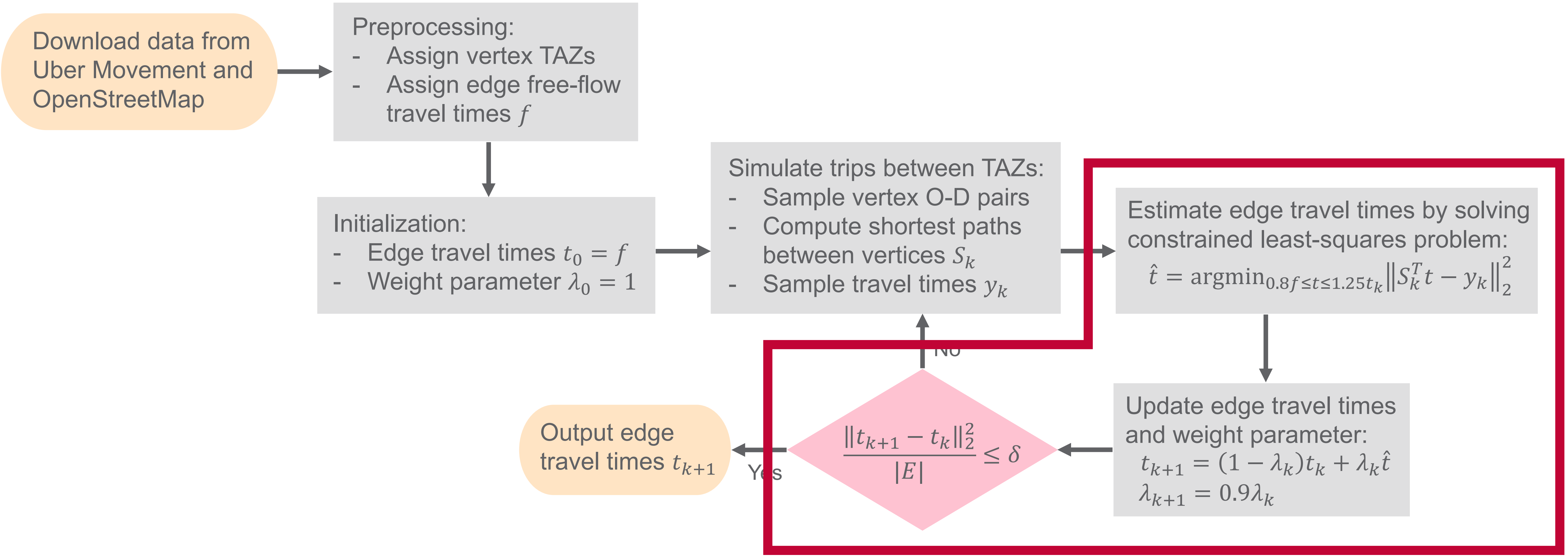
Edges with lower betweenness are more likely to have estimated travel times closer to their free-flow travel times.

Impact of using the 25% betweenness percentile as sparsification threshold



$$q = \frac{|t - f|}{f}$$

Graph Pseudo-Sparsification: Method



Summary and Future Work

- Leveraged coarse-grained Uber Movement data in the form of TAZ O-D pair summary statistics to provide estimates of fine-grained, street-level travel times.
- Solved Large metropolitan areas like Los Angeles and Seattle.
- Implement HPC based software for (near)real time estimation.
- Include Multi-Modal information: bi-cycles, bus, peds, etc.
- Open source the software with Laptop/Cloud/Supercomputer support.

Pacific Northwest
NATIONAL LABORATORY

Thank you