Travelers' Behavior Modeling, Demand Estimation and Traffic Management: a Data Perspective

Wei Ma
Mobility Data Analytics Canter (MAC)
Carnegie Mellon University

Outline

- ❑ Background
- ❑ Behavior modeling: Statistical Traffic Assignment
- ❑ Demand estimation: Probabilistic O-D Estimation
- ❑ Dynamic Network Loading
- ❑ Real-time Traffic Management

Background

Ranking	City	Hours spent in congestion	Total Cost per Driver	Total Cost per City
1		104	$ 2,408.00	9.7B
2	New York	89	$ 2,533.00	16.9B
3	San Francisco	71	$ 1,861.00	3.1B
	Pittsburgh			

http://inrix.com/resources/inrix-2016-traffic-scorecard-us/

Background

Ranking	City	Hours spent in congestion	Total Cost per Driver	Total Cost per City
1	Los Angeles	104	$ 2,408.00	9.7B
2	New York	89	$ 2,533.00	16.9B
3	San Francisco	71	$ 1,861.00	3.1B
	Pittsburgh			

http://inrix.com/resources/inrix-2016-traffic-scorecard-us/

Background

Ranking	City	Hours spent in congestion	Total Cost per Driver	Total Cost per City
1	Los Angeles	104	$ 2,408.00	9.7B
2	New York	89	$ 2,533.00	16.9B
3	San Francisco	71	$ 1,861.00	3.1B
25	Pittsburgh	33	$ 1,062.00	944M

http://inrix.com/resources/inrix-2016-traffic-scorecard-us/

Background

What causes traffic congestions?

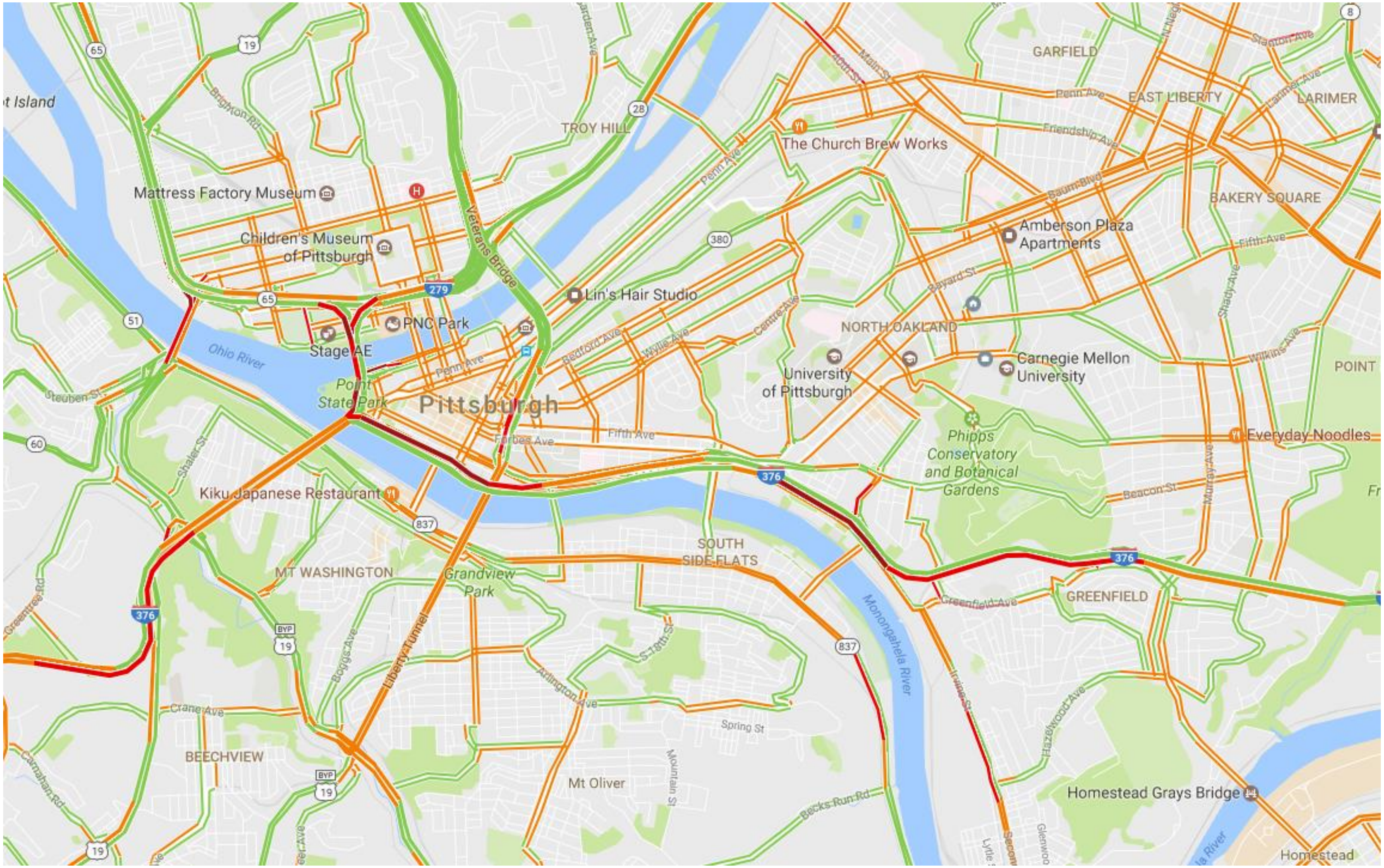
- Supply:
  - Network capacity --- more roads
  - Road capacity --- broader roads, forbid on-street parking
  - Speed density relationship --- autonomous vehicle
- Demand:
  - Heavy demand --- public transit, ride sharing
  - Imbalanced demand --- connected vehicle, information center

Background

A fundamental problem for all the traffic management models:

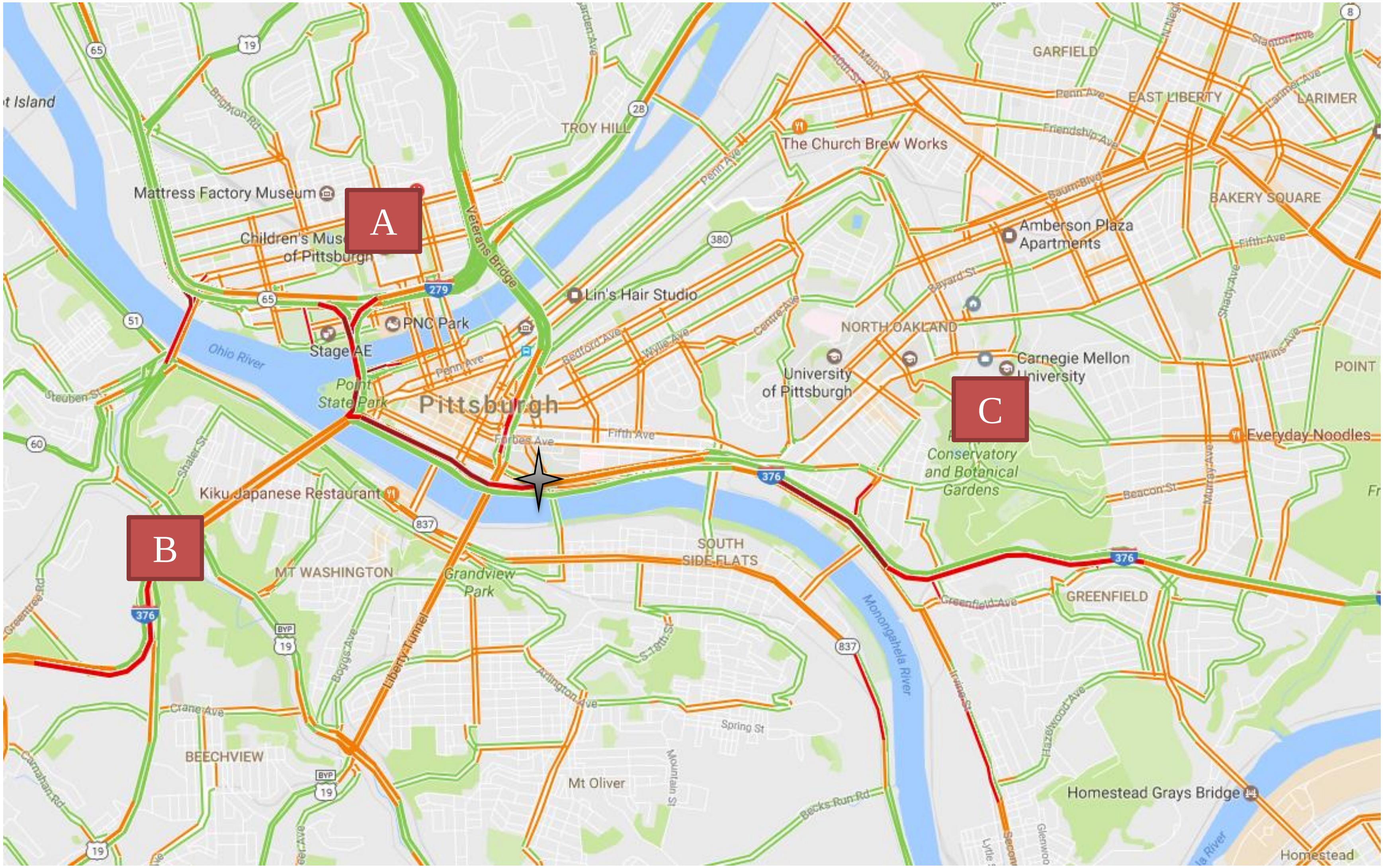
- **How to get:**
  - Network Conditions
  - Traffic Demand
  - Travelers’ Behavior

Background



Google Maps, Typical traffic, Wed, 4:00pm

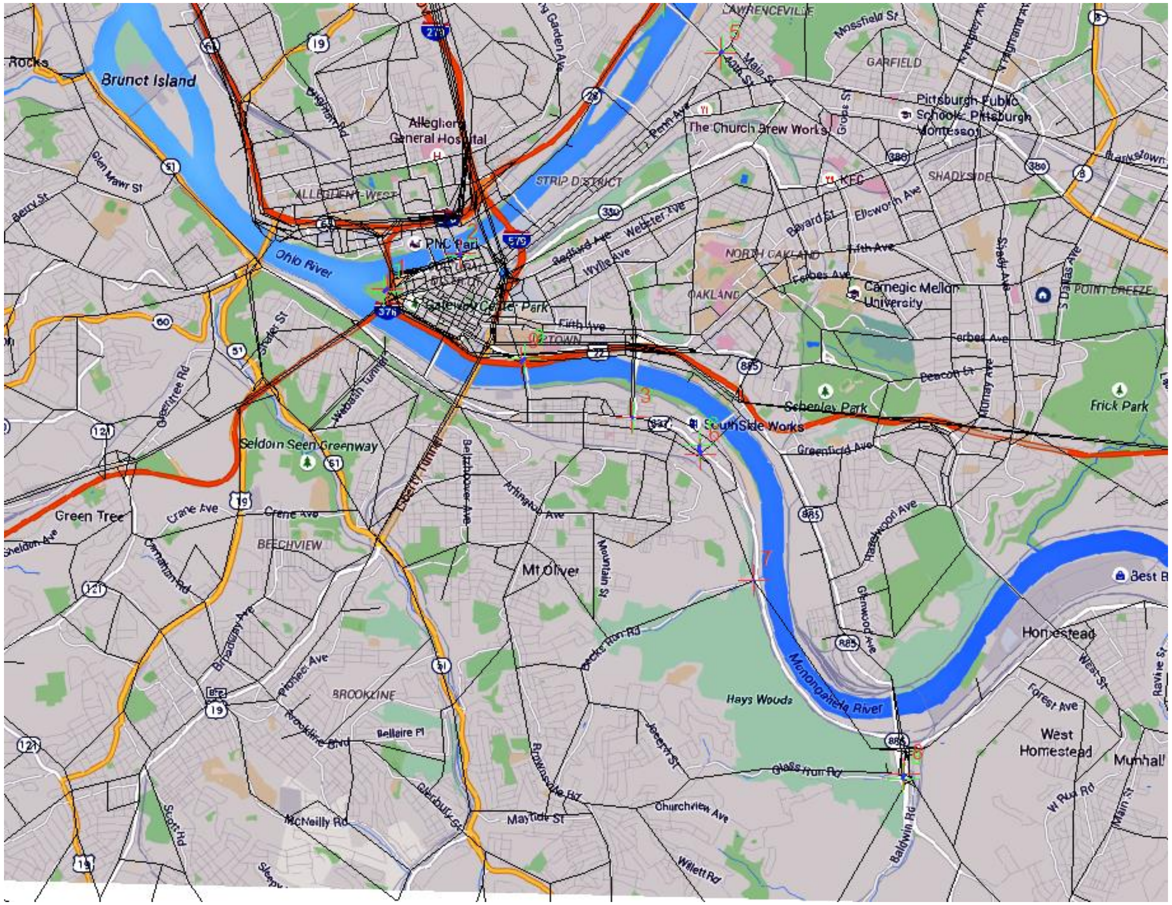
Background



Google Maps, Typical traffic, Wed, 4:00pm

Background

SUPPLY SIDE



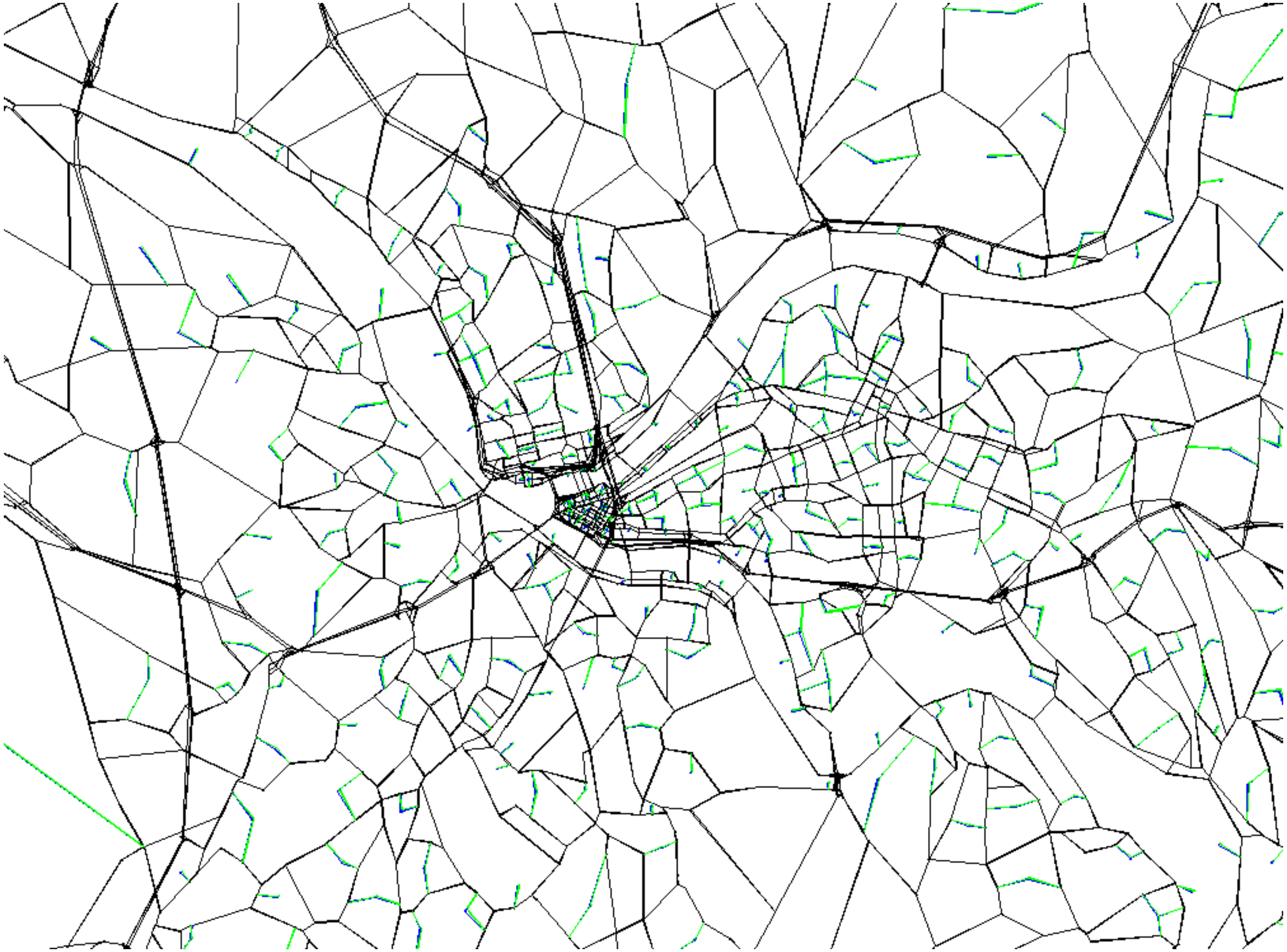
Background

What are the demand and supply for a road network?

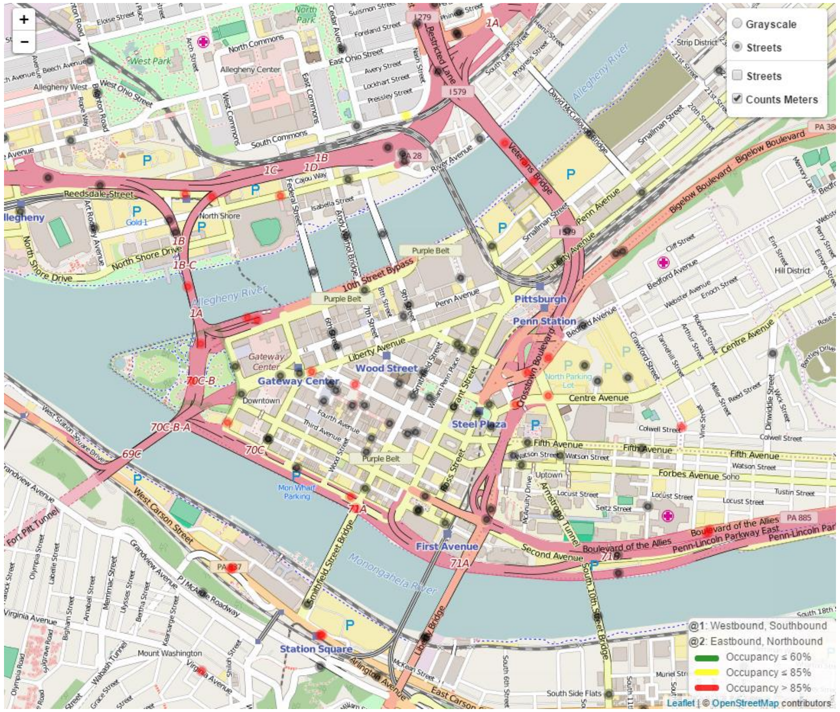
DEMAND SIDE

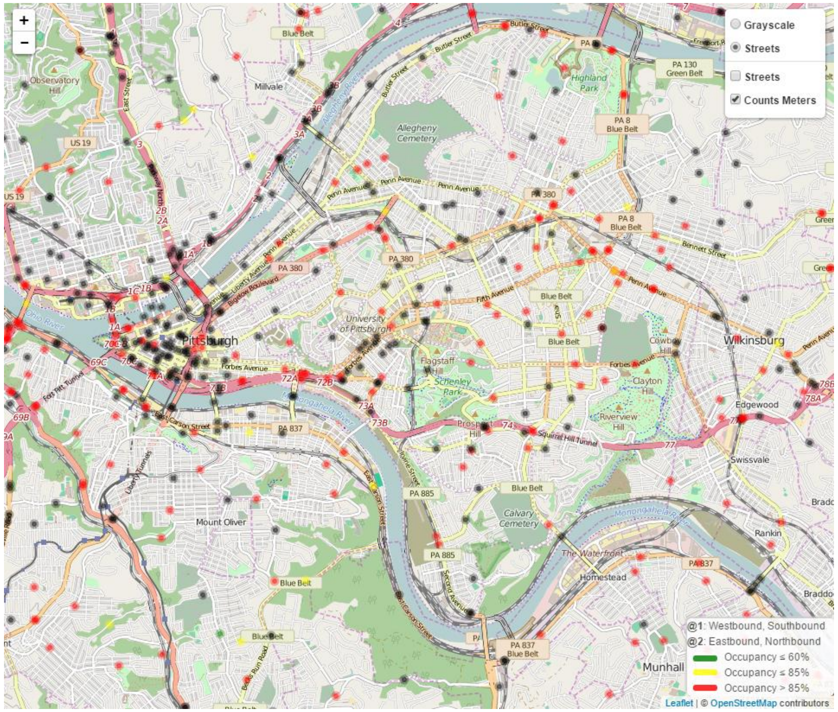
- Origin – Destination demand (O-D) demand
- Previously estimated by the density of residence, not accurate and not flexible
- We want to estimate the O-D demand from various traffic data.

86499 O-D pairs



Traffic flow





Traffic speed

www.ritis.org

OBJECTIVE

Infer the network conditions, travelers' behavior and traffic demand from partially observed traffic data.

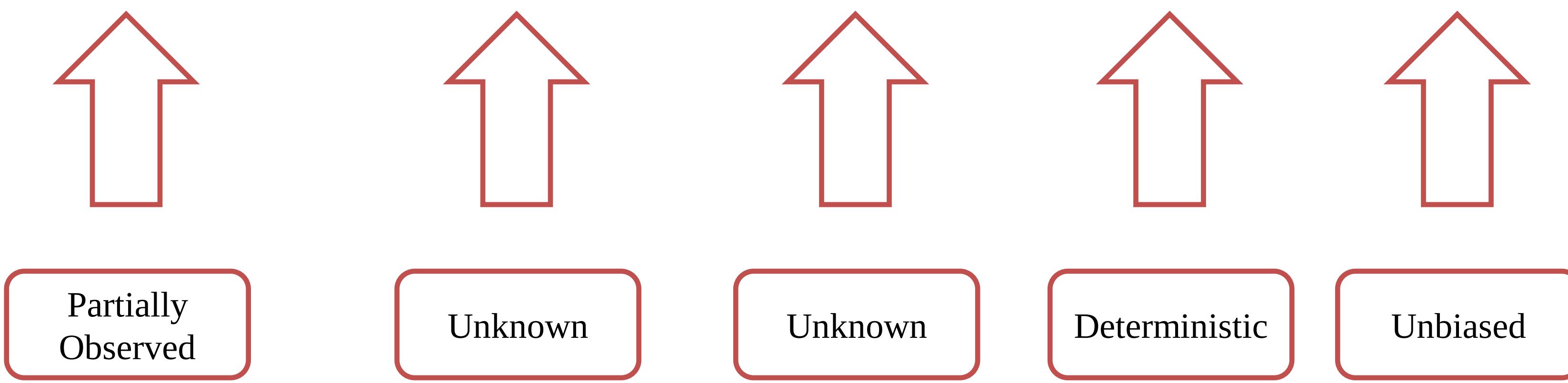
Input: partially observed traffic data
Output: whole network condition, underlying behavior

Background

Look the problem as a regression model

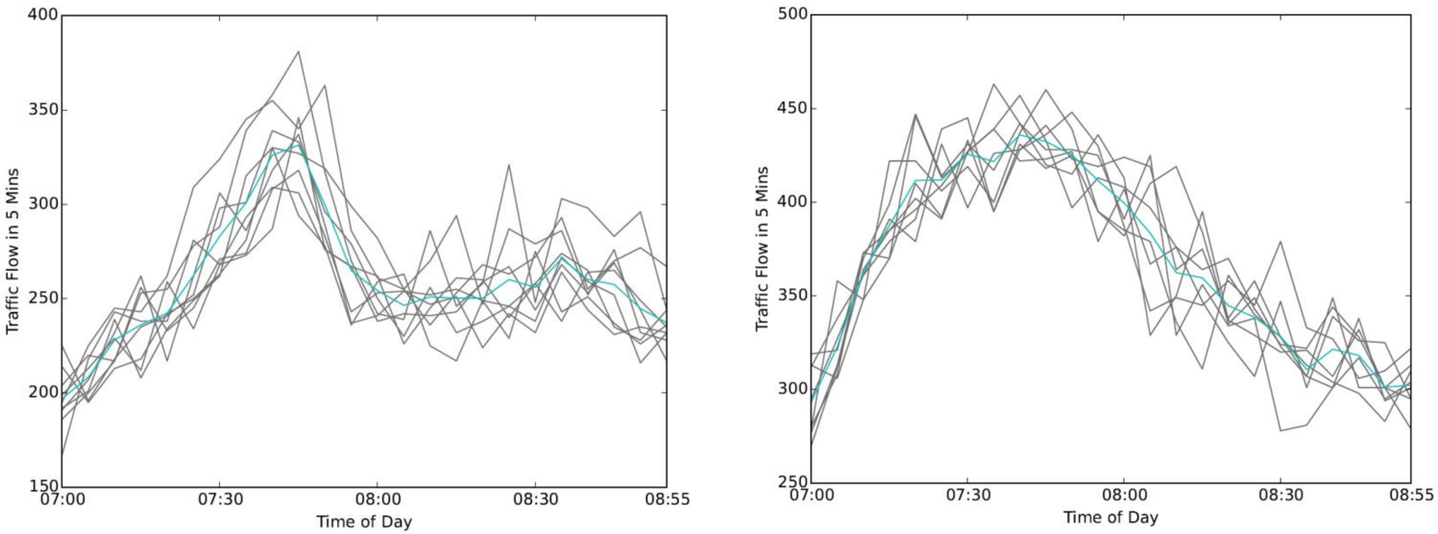
$$Y = f(q, s) + e$$

(speed, flow) = Travelers' behavior(O-D demand, network supply) + error



Statistical Traffic Assignment

(speed, flow) = Travelers' behavior(O-D demand, network supply) + error



Daily time-varying traffic counts on SR41 SB and NB

- Statistical traffic assignment explores the statistical features of recurrent flow patterns

$$f: Q \longmapsto X, F$$

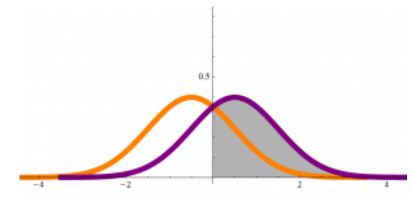
Statistical Traffic Assignment

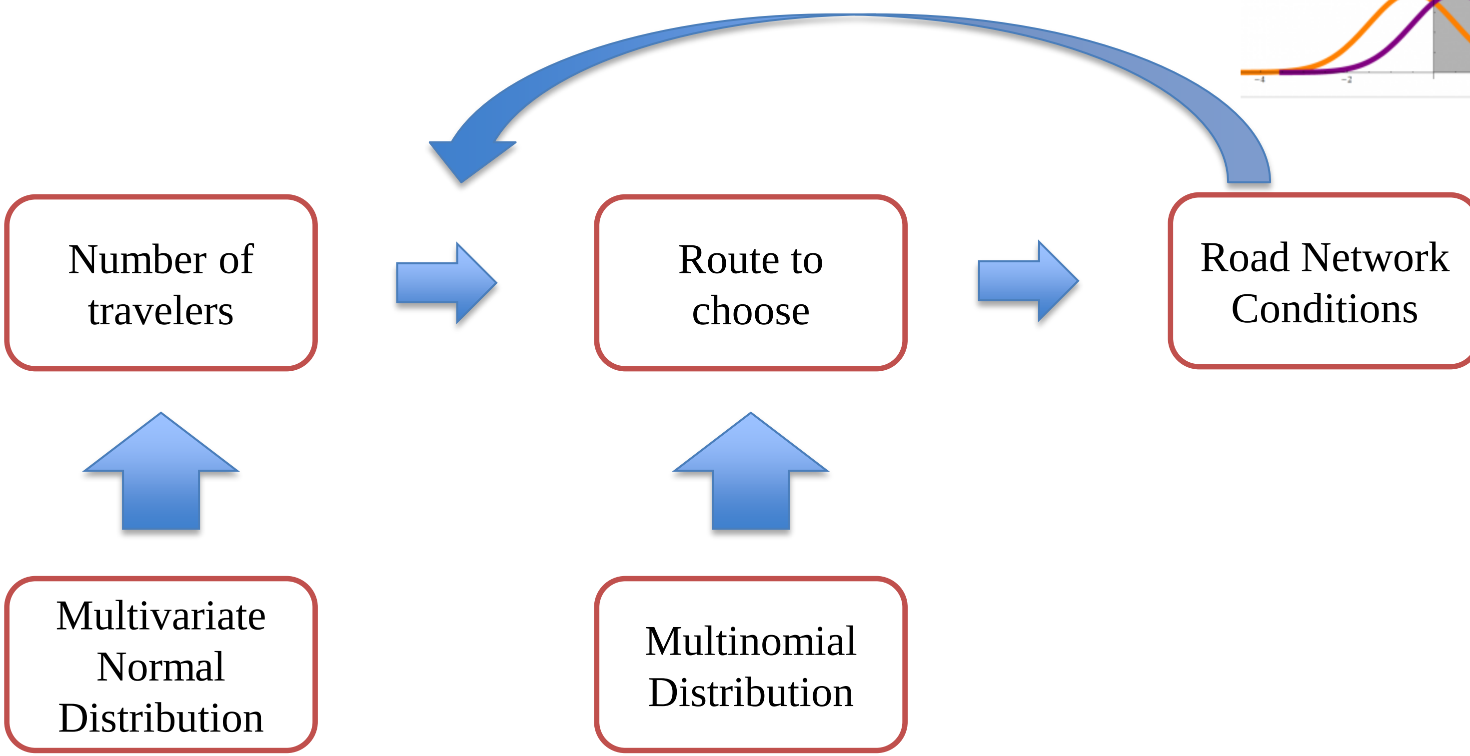
- ❑ O-D demand variation, route choice variation, unknown error
- ❑ Variance and covariance in O-D demand
- ❑ Random selection of the route
- ❑ Unknown error caused by measurement

Statistical Traffic Assignment

How to generate traffic?

- Travelers' Information Structure: Know the distribution of path costs
- Travelers' Route Choice: Probability based





http://www.nationalpower.info/winning-probability/

Statistical Traffic Assignment

- OD
  - MVN: $Q \sim N(q, \Sigma_q)$
  - Scalable, stable and consistent
- Route choice
  - p is deterministic for a recurrent traffic network
  - A generalized probability function: $p = \psi(C; \Theta)$
- Path flow
  - Path flow follows multinomial distribution
  - $F_{rs}|Q_{rs} \sim MN(Q_{rs}, p_{rs})$
  - Normal approximation
- Link flow
  - $X = \Delta F$, where Δ is the path/link incidence matrix
- Path cost
  - A generalized path cost function: $C \sim t(X; \Theta)$
  - Normal approximation

Model: Hierarchy

Level 1:

$$X_m \sim N(X + e, \Sigma_x + \Sigma_e)$$ Unknown error

Level 2:

$$X \sim N(\Delta pQ, \Sigma_x)$$
$$F \sim N(pQ, \Sigma_f)$$ Route choice variation

Level 3:

$$Q \sim N(q, \Sigma_q)$$ OD variation

- ❑ Level i is conditional on level i+1
- ❑ Each level reflects one single source of the link flow variation

Model: Property

- ❑ Existence and consistence of the solution
  - Fixed point problem
  - If the path cost function is continuous, the solution $(x, \Sigma_x, f, \Sigma_f, p)$ exists.
  - Once the solution $(x, \Sigma_x, f, \Sigma_f, p)$ is determined, then it's consistent.
- ❑ Reducible
  - Model can be reduced to existing statistical assignment models
- ❑ Data driven
  - Friendly to model learning techniques for large scale networks
  - Dimension reduction, sparse regularization, model selection

Model: Application

- ❑ Marginal distribution

  Marginal distribution of X_m can be written as:

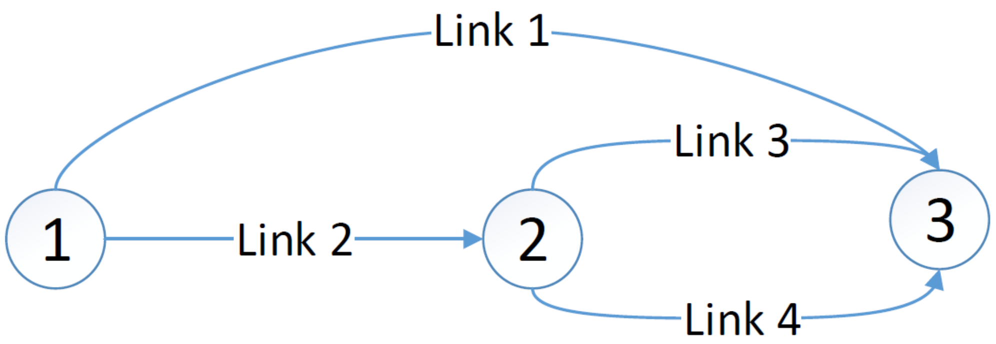
$$X_m \sim N(x_m, \boxed{\Delta p \Sigma_q p^T \Delta^T} + \boxed{\Delta \Sigma_{f|q} \Delta^T} + \boxed{\Sigma_e})$$

  Three matrices represent the variance from OD, route choice and measurement error separately.

- ❑ Variance ratio
  - To measure the portion of each source of variance
  - To help to reduce the system variance
  - Trace norm, nuclear norm

Experiment

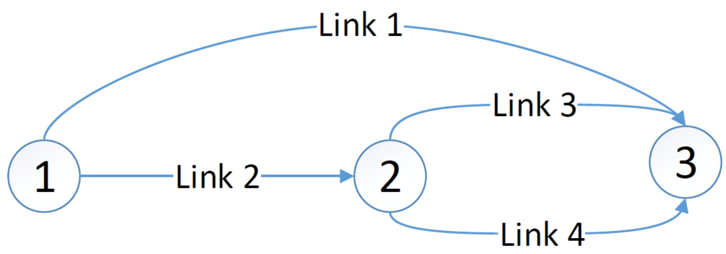
❑ Settings



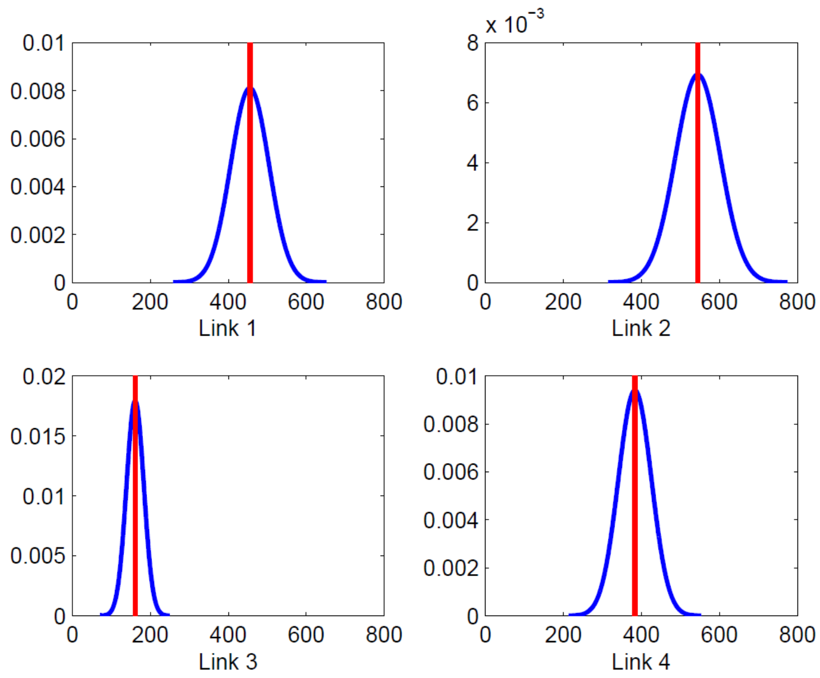
- OD: 1 -> 3, $q = 1000$, $\sigma^2 = 10000$
- Route choice model: Probit + Multinomial
- Link cost function: BPR function
- Measurement error: $e = 0$, $\sigma_e^2 = 100$

Experiment

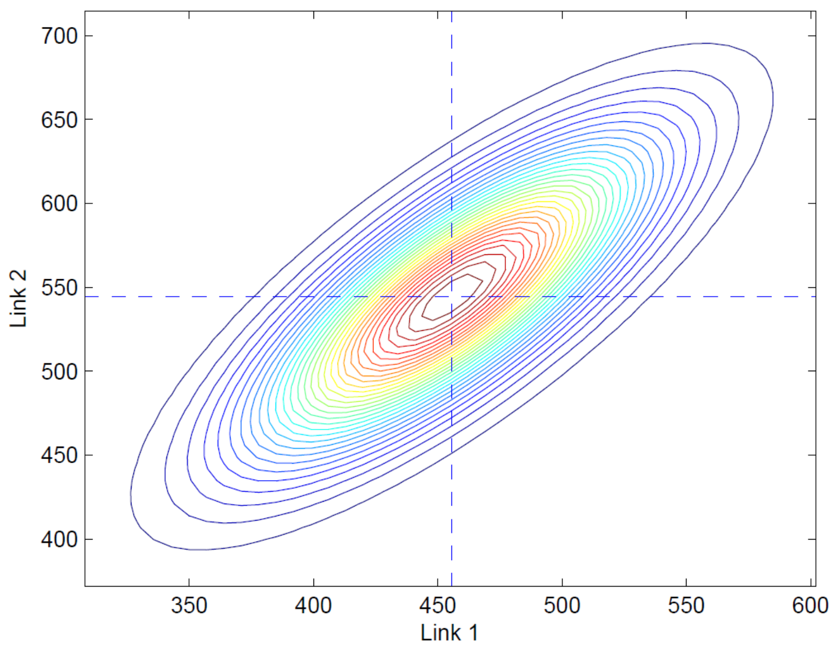
- ❑ Basic results



Network condition

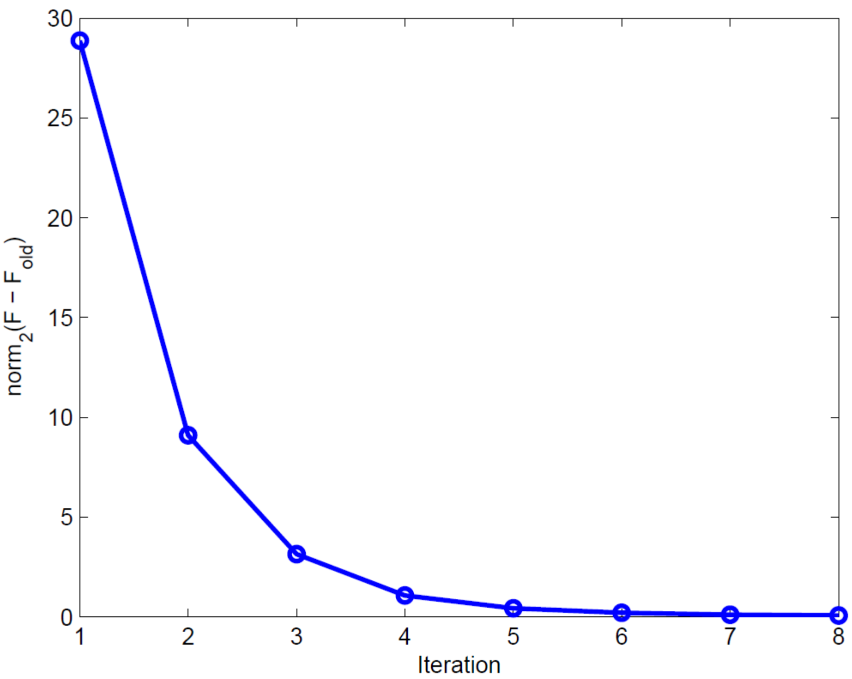


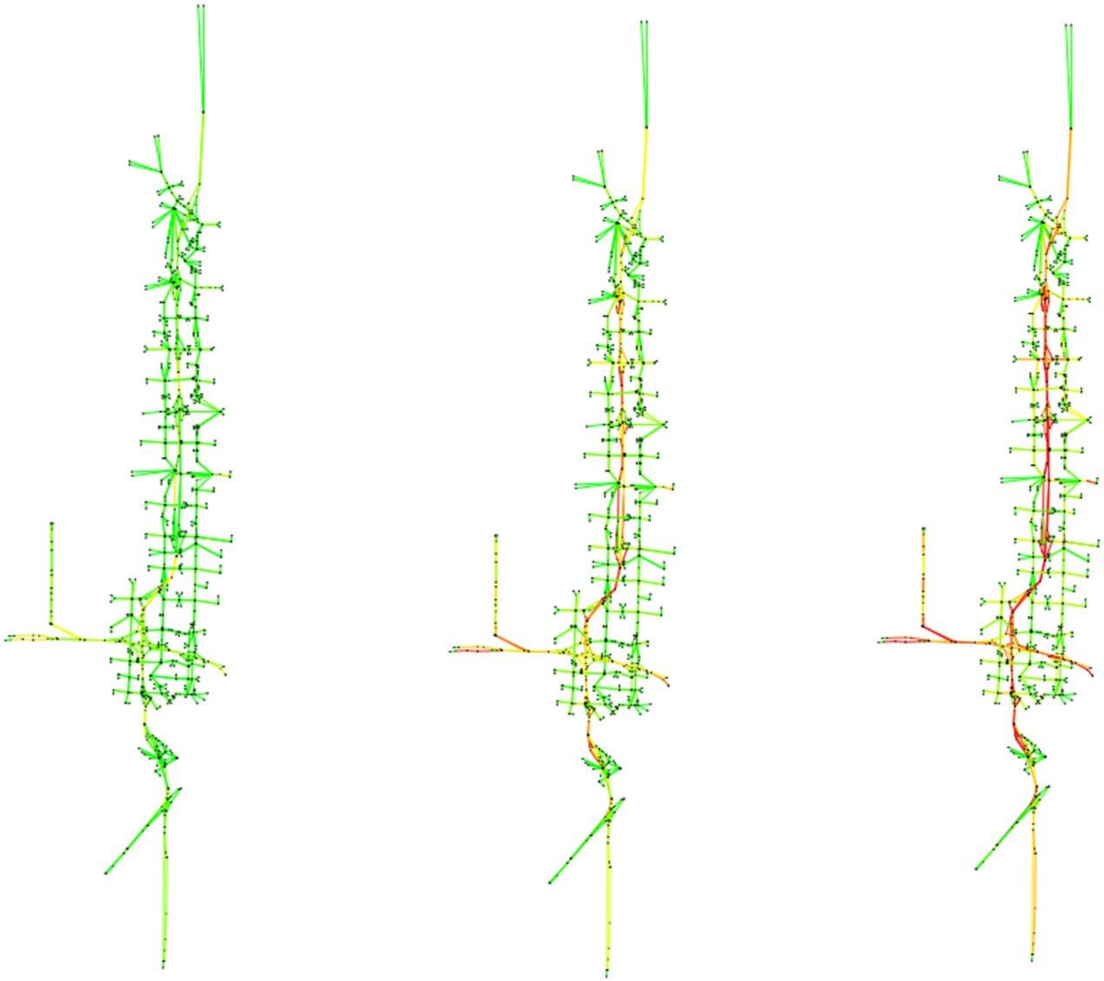
Confidence interval



Large Scale Network

- Settings
  - SR41 freeway network: 2413 links and 7110 OD pairs
  - OD demands were carefully calibrated
  - Assume the O-D demand variance is 20% of its mean
- Results
  - Terminated in 9 iterations and 467 seconds





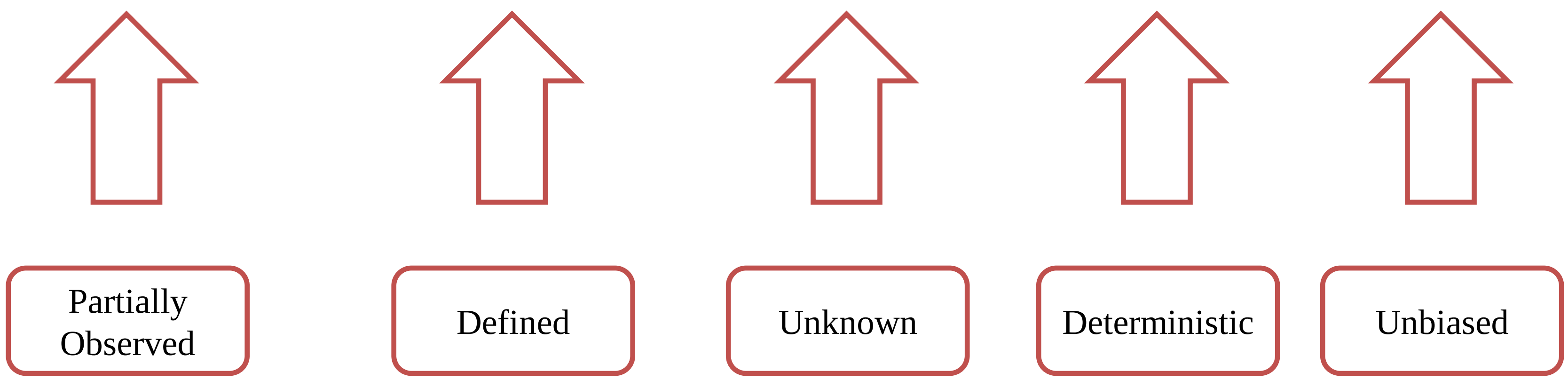
- Using Rectangle CI approximation.
- Red represents volume/capacity>1, and green represents volume/capacity=0, other colors are smoothly transitioned from green to red as volume/capacity increases from 0 to 1)

Background

Look the problem as a regression model

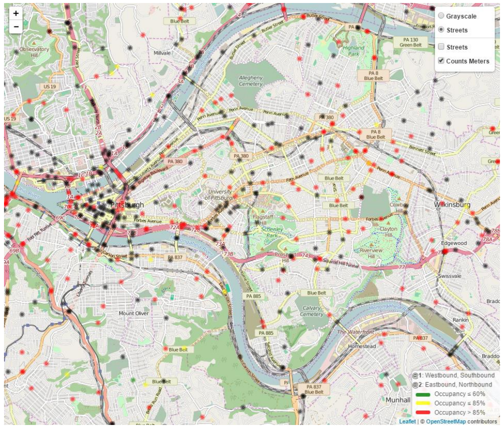
Y= f (q, s) + e

(speed, flow) = Travelers' behavior(O-D demand, network supply) + error



Probabilistic O-D Estimation

- ❑ Objective: estimate the mean and variance/covariance matrix of O-D demand from day-to-day traffic data
- ❑ Data: partial observations X_m^o, 10% of the total road segments, speed on major roads



Probabilistic O-D Estimation

Why this is difficult?

- Hierarchical statistical model
- For Pittsburgh area: number of traffic flow data: ~2300

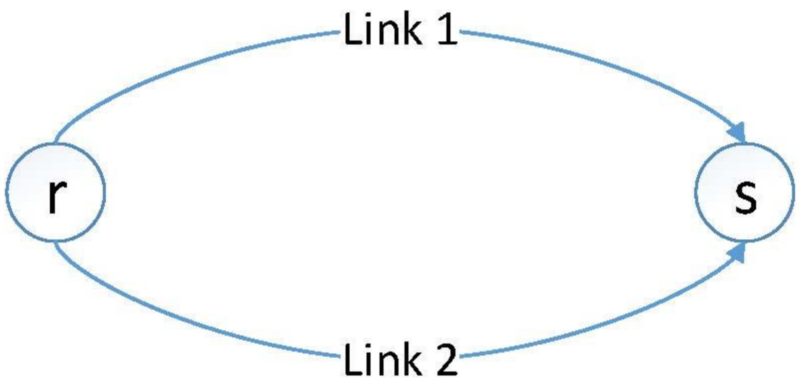
  number of traffic speed data: ~1000

  number of O-D pair: ~86000
- Under-determined problem, multiple solutions for O-D demand
- Number of parameters in variance/covariance matrix for O-D demand: 86000 * 43000 !

Probabilistic O-D Estimation

Equilibrium:

- Rational travelers in recurrent traffic conditions
- If link 1 and link 2 are exactly the same, then the traffic flow on both links should be the same
- Multiple player equilibrium: Nash Equilibrium



Probabilistic O-D Estimation

General Framework:

- Iterative estimation framework (IGLS).
- EM algorithm in ML

Probabilistic O-D Estimation

Estimate OD mean:

GLS, equilibrium constraint, single level relaxation

$$\begin{aligned} \min_{f} \quad & n\left(\Delta^{o} f - \hat{x^{o}}\right)^{T} \Sigma_{x}^{o\,-1}\left(\Delta^{o} f - \hat{x^{o}}\right) + (q^{H} - Mf)^{T} \Sigma_{q}^{H\,-1} (q^{H} - Mf) \\ \text{s.t.} \quad & f \in \Phi^{+} \end{aligned}$$

Estimate OD variance:

MLE, convex relaxation, LASSO model selection.

$$\begin{aligned} \min_{\Sigma_q} \quad & \left\| S_x^o - \Sigma_x^o \right\|_F^2 + \lambda \left\| \Sigma_q \right\|_1 \\ \text{s.t.} \quad \Sigma_x^o \; & = \; \Delta^o \Sigma_{f|q} \Delta^{oT} + \Delta^o \tilde{p} \Sigma_q \tilde{p}^T \Delta^{oT} \\ \Sigma_q \; & \in \; \text{semidefinite}(\Re^{|K_q| \times |K_q|}) \end{aligned}$$

Probabilistic O-D Estimation

Model Observability:

- Non-unique solution
- No worse than deterministic O-D estimation

Goodness of fit:

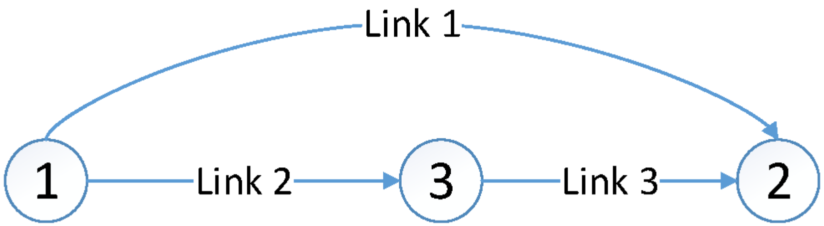
- Distribution based
- Hellinger distance or Kullback-Leibler distance

$$D_H((\mu_1,\Sigma_1)^T,(\mu_2,\Sigma_2)^T) = 1-\frac{|\Sigma_1|^{\frac{1}{4}}|\Sigma_1|^{\frac{1}{4}}}{\left|\frac{1}{2}\Sigma_1+\frac{1}{2}\Sigma_2\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{8}(\mu_2-\mu_1)^T\left(\frac{1}{2}\Sigma_1+\frac{1}{2}\Sigma_2\right)^{-1}(\mu_2-\mu_1)\right) \tag{27}$$

$$D_{KL}((\mu_1,\Sigma_1)^T,(\mu_2,\Sigma_2)^T) = \frac{1}{2}\left(\log\frac{|\Sigma_2|}{\Sigma_1}-d+\mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right)+(\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1)\right) \tag{28}$$

Probabilistic O-D Estimation

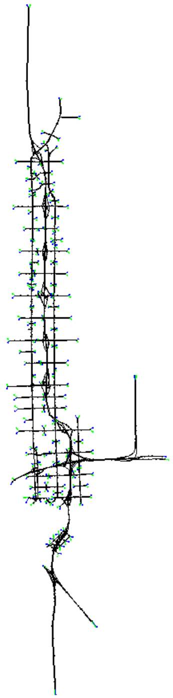
Link 1 and Link 3 Observed,
500 samples

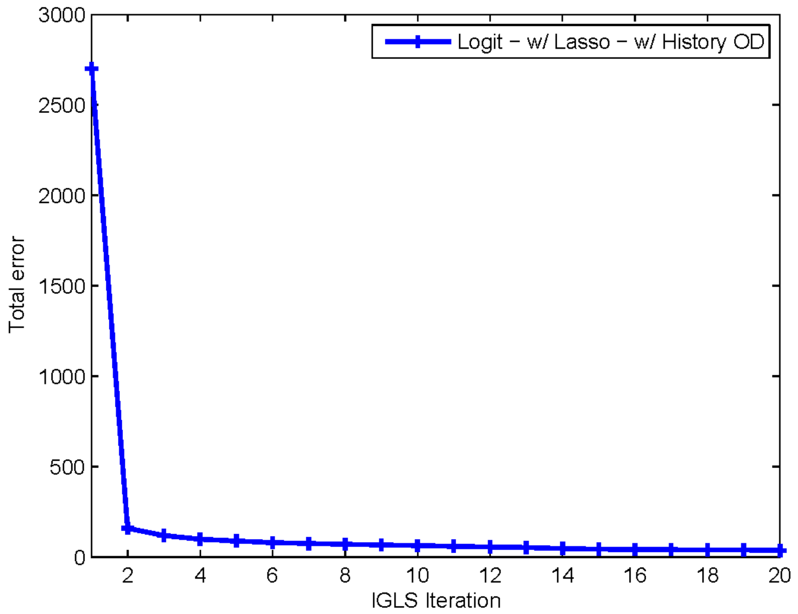


True ρ	Method	$\hat{q}_{1\to2}$	$\hat{q}_{3\to2}$	$\hat{\sigma}_{1\to2}$	$\hat{\sigma}_{3\to2}$	$\hat{\rho}$	RMPSE	KL-distance
	True	700	500	13.23	11.18	NA	NA	NA
0.5	w/o EC - w/o Lasso	611.02	588.01	14.18	11.46	0.37	14.64%	106.80
	Logit - w/o Lasso	728.55	588.85	12.81	11.07	0.55	11.00%	33.78
	Probit - w/o Lasso	618.63	590.17	15.34	11.36	0.45	14.31%	101.69
0	w/o EC - w/o Lasso	765, 43	588.78	11.95	10.25	0.03	12.90%	43.08
	Logit - w/o Lasso	727.94	588.05	13.05	11.00	0.05	10.89%	33.24
	Probit - w/o Lasso	618.58	587.79	13.79	11.16	0.07	14.02%	49.07
	Logit - w/ Lasso	728.27	588.48	12.87	11.53	0.00	10.95%	33.60
	Probit - w/ Lasso	621.39	588.56	13.71	11.67	0.00	13.94%	49.04
-0.5	w/o EC - w/o Lasso	780.12	586.04	7.80	11.18	-0.58	13.85%	95.26
	Logit - w/o Lasso	726.41	586.08	12.99	11.12	-0.58	10.61%	52.44
	Probit - w/o Lasso	621.40	588.66	13.29	10.68	-0.37	13.79%	32.90

w/o EC: without equilibrium constraint; Logit/Probit: using Logit/Probit based SUE constraint;
w/o Lasso: without Lasso regularization; w/ Lasso: using Lasso regularization

Probabilistic O-D Estimation





- Logit - w/ Lasso - w/ History O-D method on a desktop computer (Inter(R) Core 529 i5-4460 3.20 GHz 2, RAM 8 GB)
- Average computation time for each IGLS iteration is 233.05s.
- Converge in 20 iterations.

Big Map

	Model reliability	
Model complexity	Deterministic Dynamic	Stochastic Dynamic
	Deterministic Static	Stochastic Static

Moving to Dynamic Network

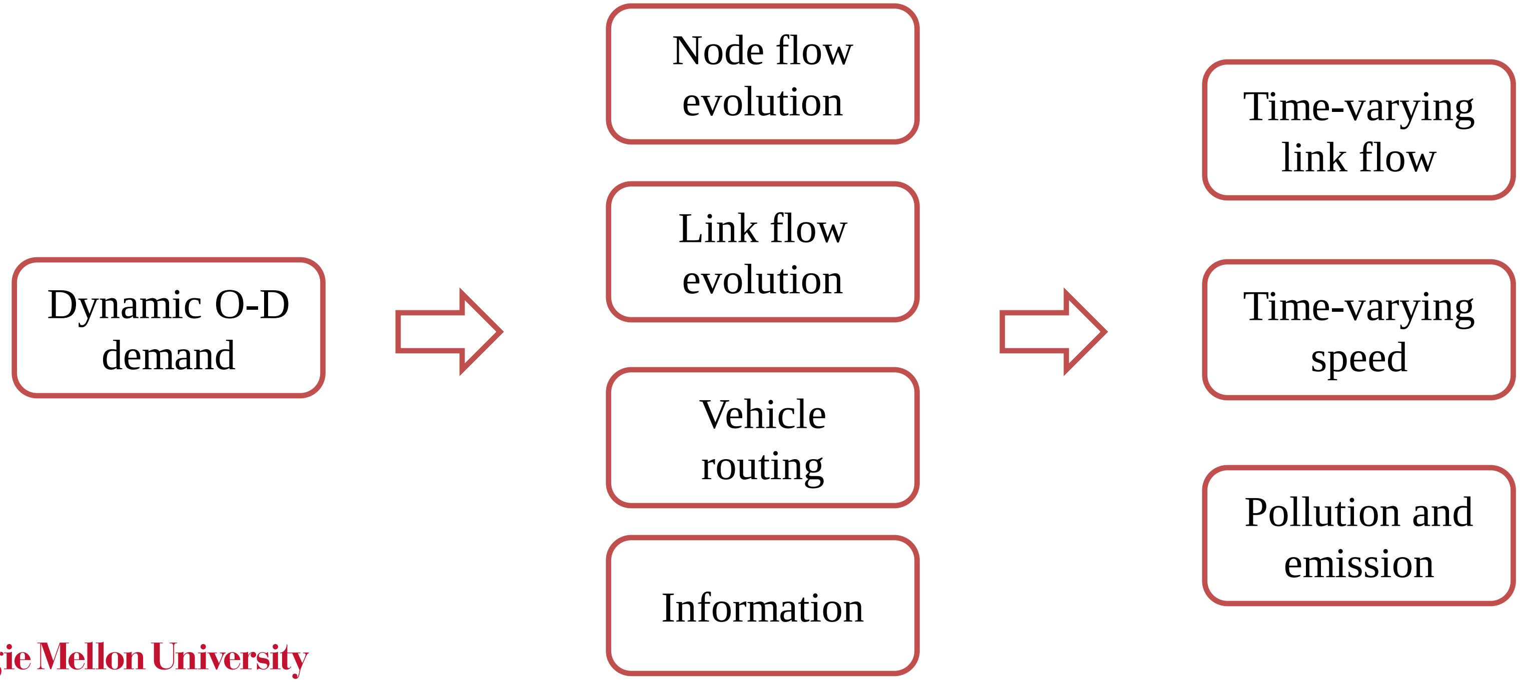
- Static O-D demand --- Dynamic O-D demand
- Travelers' behavior --- **Dynamic network loading**
- Static traffic condition --- Network flow evolution

Time varying speed, flow = DNL(dynamic O-D demand, network supply)

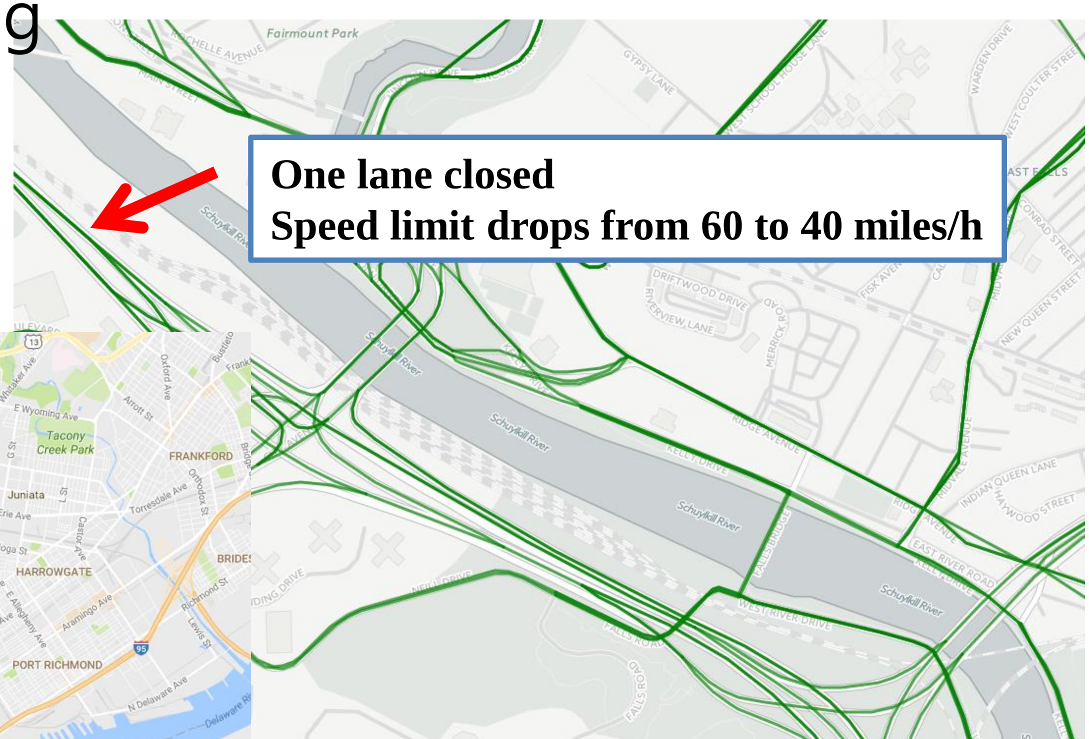
Dynamic Network Loading

MAC-POSTS:

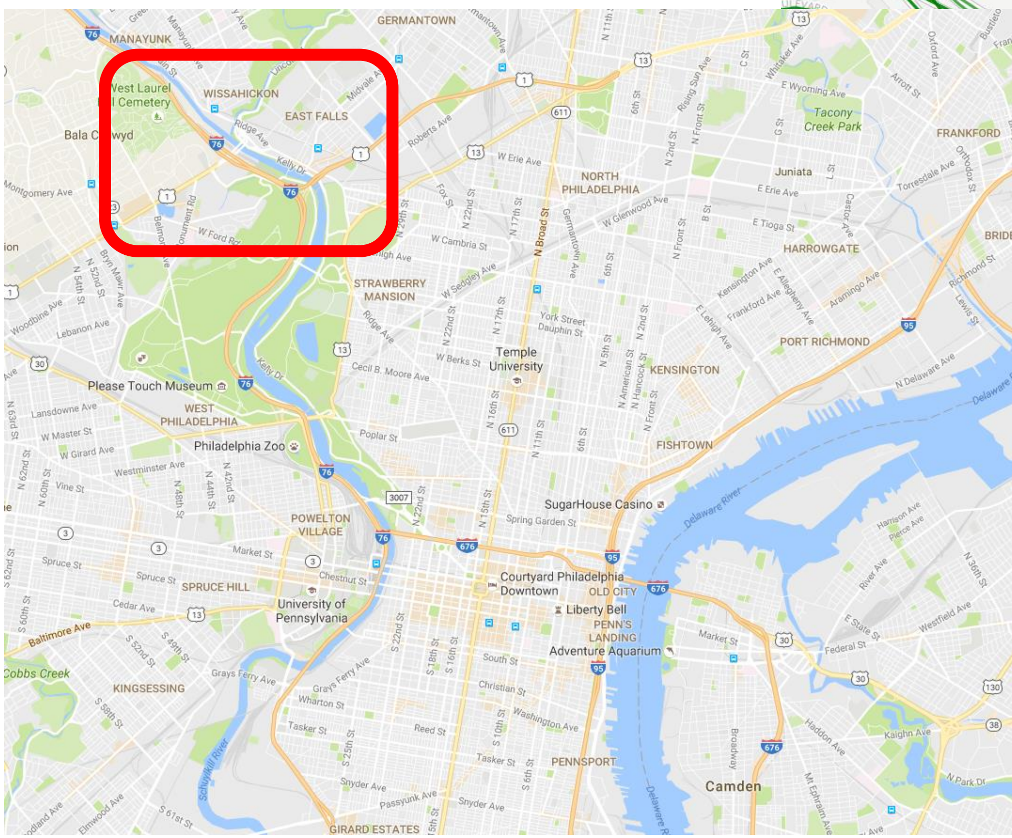
- Mobility Data Analytics Center - Prediction, optimization, and simulation toolkit for transportation systems
- https://github.com/Lemma1/MAC-POSTS



Dynamic network loading

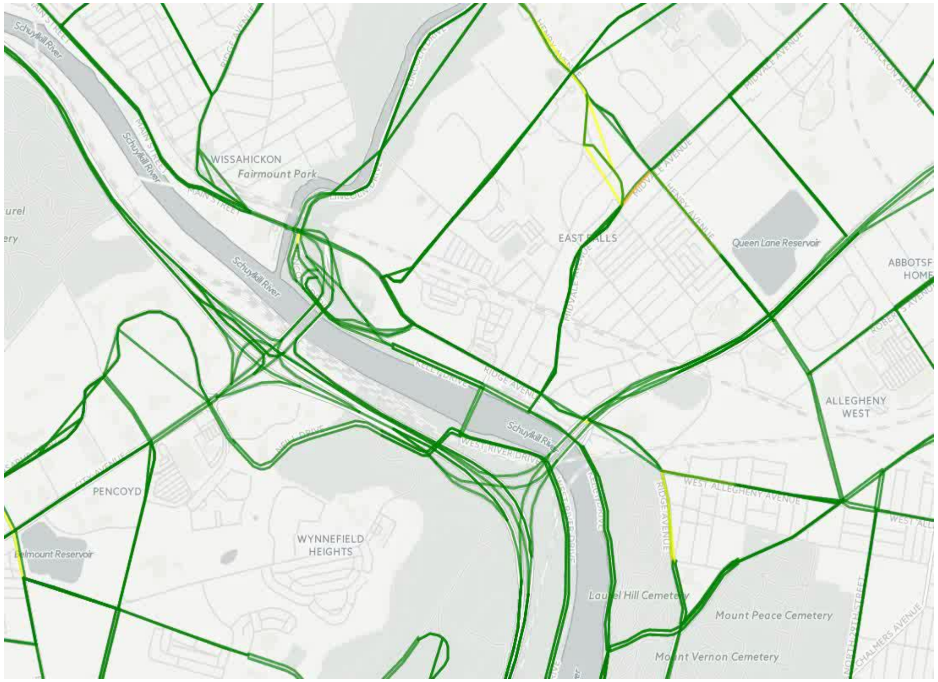


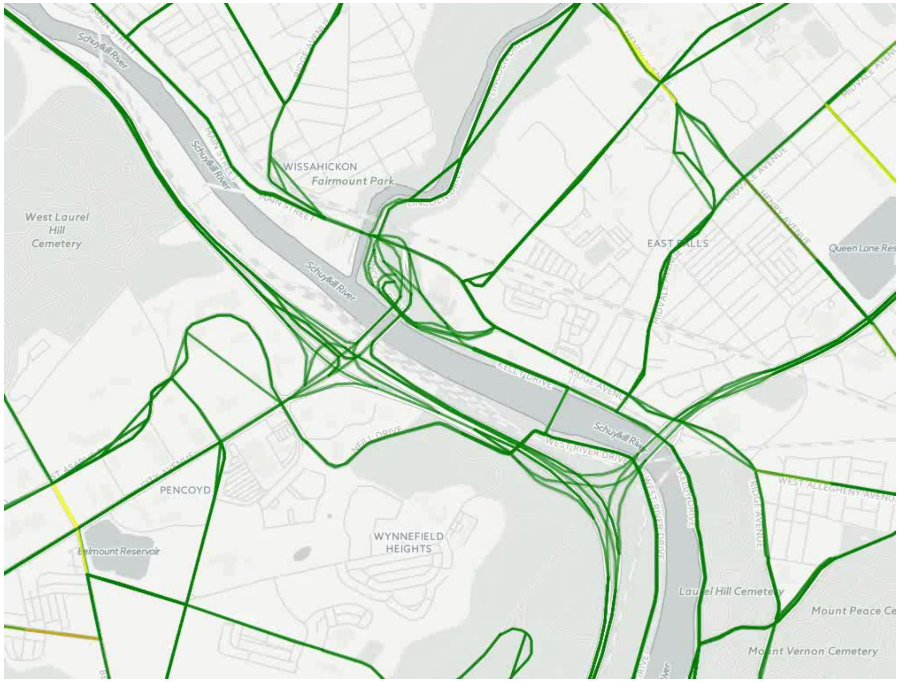
One lane closed
Speed limit drops from 60 to 40 miles/h



Philadelphia, DVRPC area

Dynamic network loading





PLAY

2:00 PM – 3:00 PM

Real-time traffic management

Objective:

- real-time control of dynamic message sign on I-95 corridor to reduce network congestion

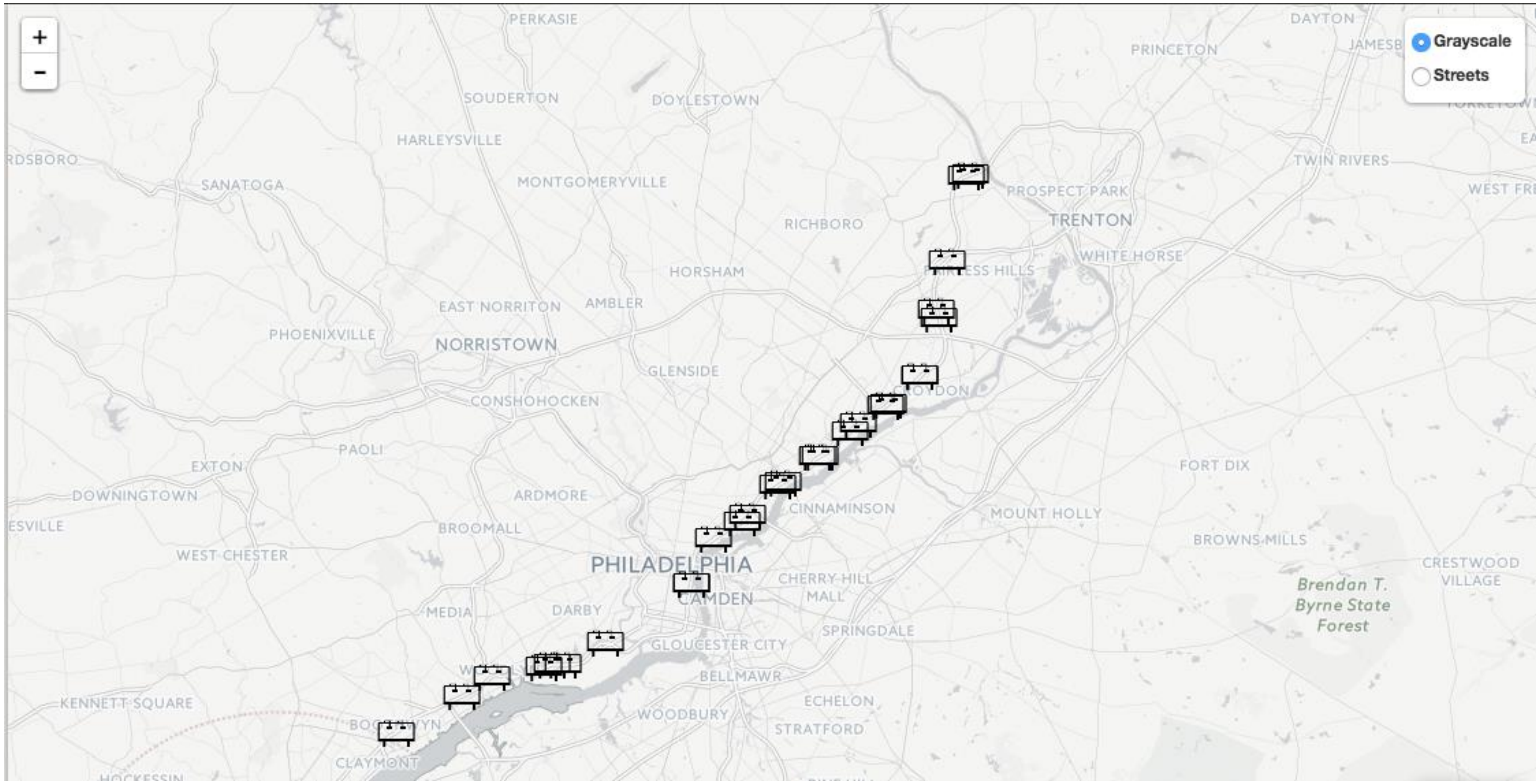
Data Input:

- Estimated dynamic O-D demand, calibrated behavior parameters
- Real-time traffic speed feeds
- Real-time accident report

Prediction Method:

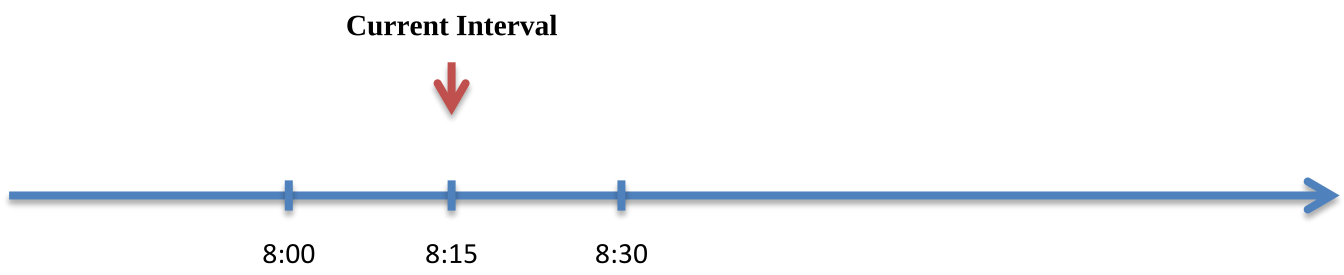
- Dynamic network loading

Real-time traffic management

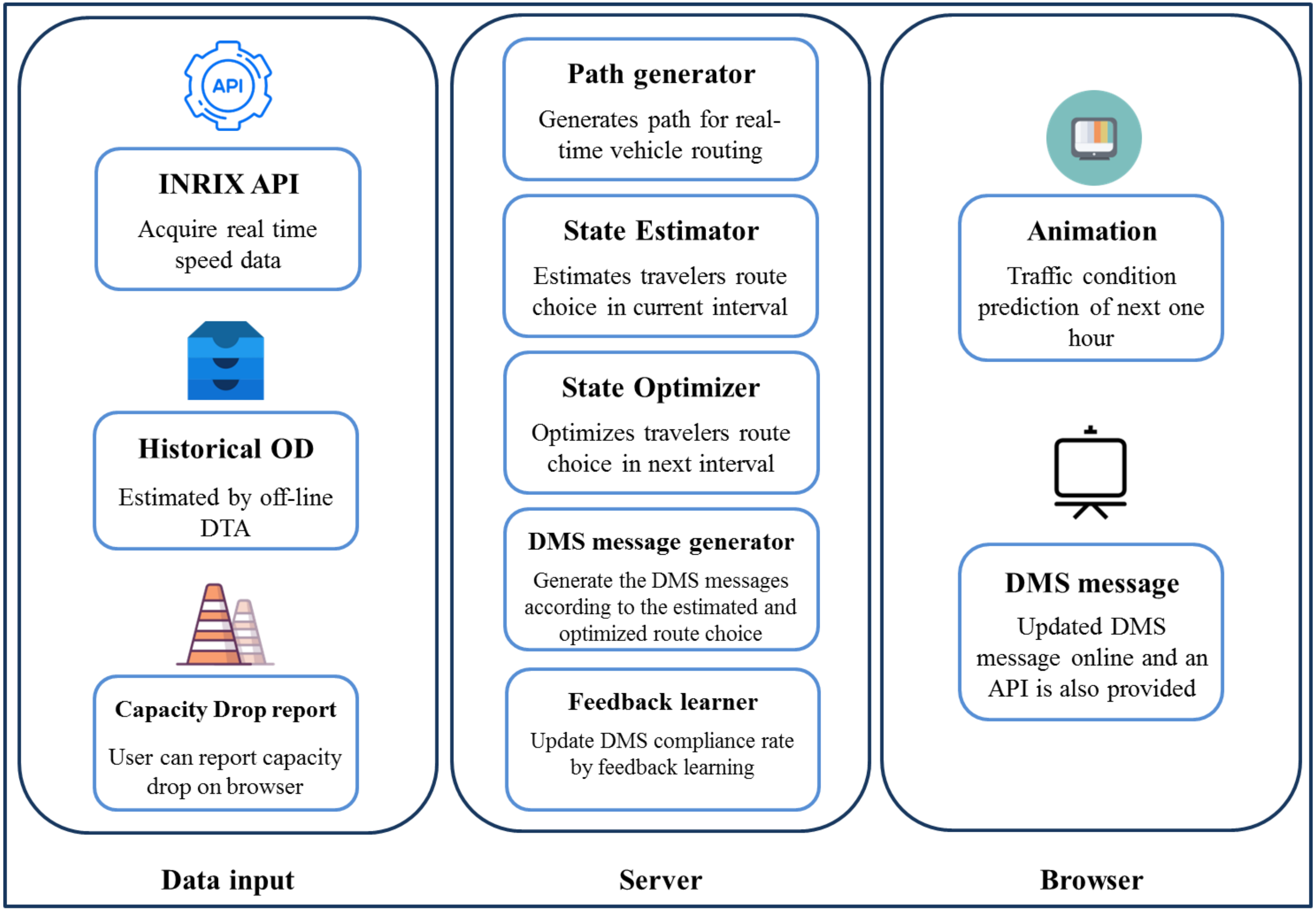


Dynamic message signs along the I-95 corridor in Philadelphia

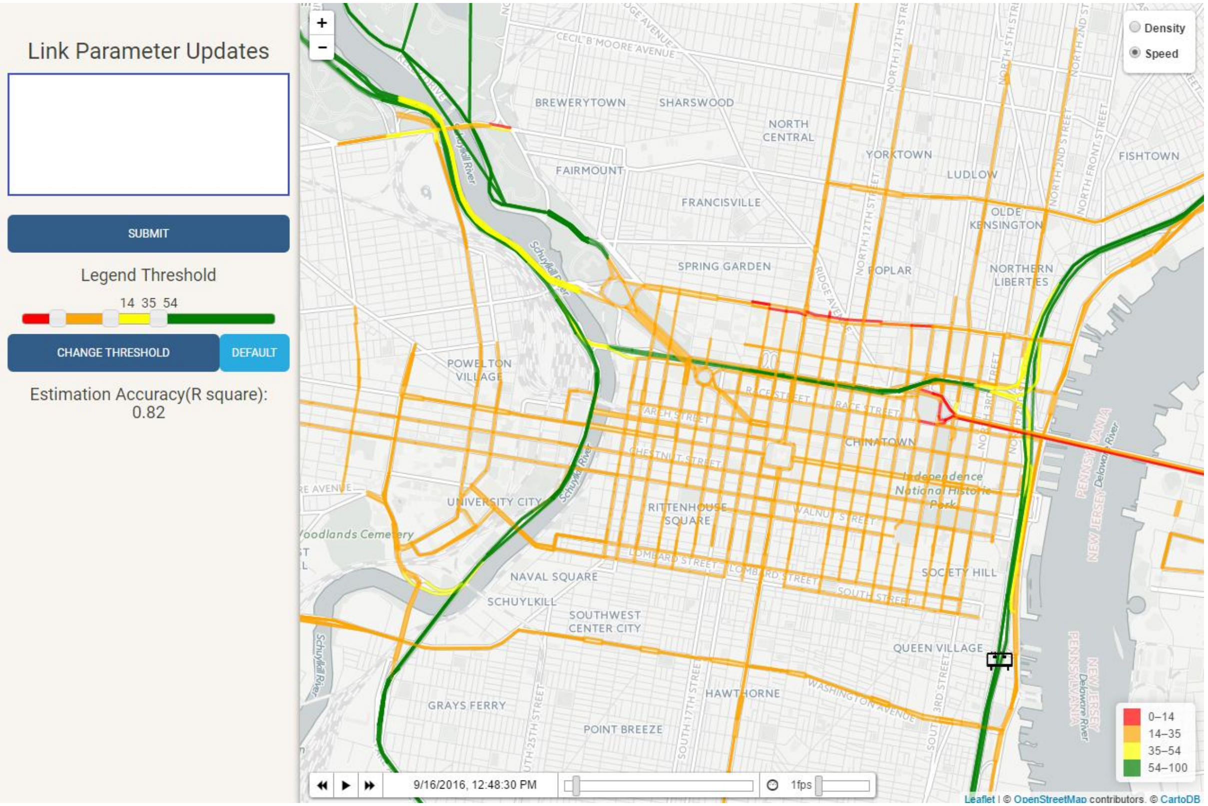
Rolling Horizon Framework



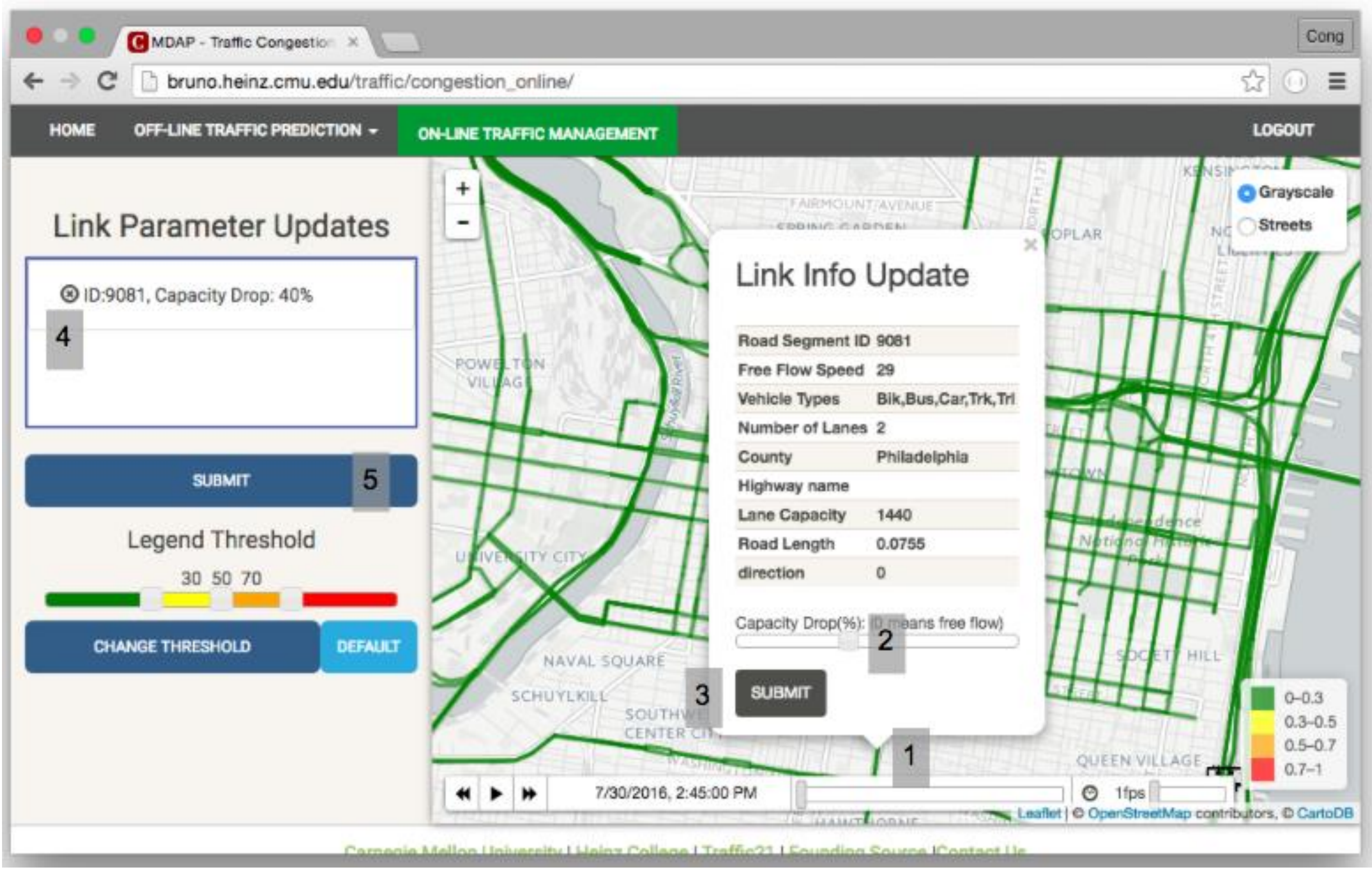
1. **Acquire Speed**: acquire the real-time traffic speed in [8:00, 8:15]
2. **Estimate Previous**: estimate the route choice probability in [8:00, 8:15]
3. **Compliance ratio update**: update the DMS compliance ratio calculated in [7:45, 8:00]
4. **Optimize Next interval**: optimize the route choice probability in [8:15, 8:30]
5. **Generate DMS Message**: based on the optimized and estimated route choice, generate DMS message for [8:15, 8:30]
6. **Predict Future**: Predict traffic state in next one hour using updated compliance ratio
7. Move current time to 8:30, go back to step 1

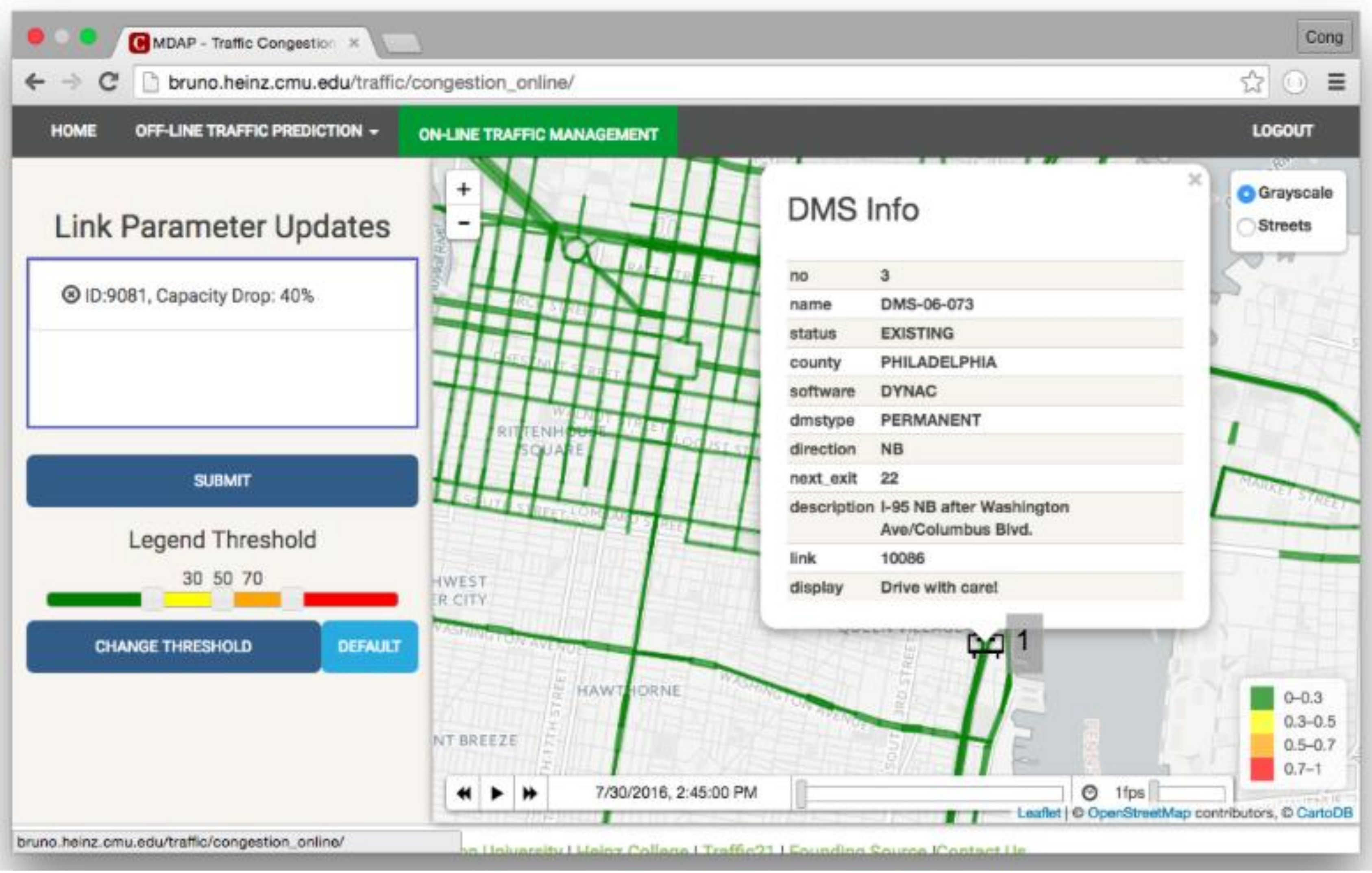


Overview of real-time traffic management framework


Link Parameter Updates
SUBMIT
Legend Threshold
14 35 54
CHANGE THRESHOLD
DEFAULT
Estimation Accuracy(R square):
0.82
Density
Speed
0–14
14–35
35–54
54–100
9/16/2016, 12:48:30 PM
1fps
Leaflet | © OpenStreetMap contributors, © CartoDB


Data for the next 60 minutes
PostgreSQL
Query
Data
django
Query
Data
www
Query
Data


MDAP - Traffic Congestion
bruno.heinz.cmu.edu/traffic/congestion_online/
HOME
OFF-LINE TRAFFIC PREDICTION
ON-LINE TRAFFIC MANAGEMENT
LOGOUT
Link Parameter Updates
ID:9081, Capacity Drop: 40%
4
SUBMIT
5
Legend Threshold
30 50 70
CHANGE THRESHOLD
DEFAULT
Link Info Update
Road Segment ID 9081
Free Flow Speed 29
Vehicle Types Bik,Bus,Car,Trk,Trl
Number of Lanes 2
County Philadelphia
Highway name
Lane Capacity 1440
Road Length 0.0755
direction 0
Capacity Drop(%): (0 means free flow)
2
3
SUBMIT
1
Grayscale
Streets
0-0.3
0.3-0.5
0.5-0.7
0.7-1
7/30/2016, 2:45:00 PM
1fps
Leaflet | © OpenStreetMap contributors, © CartoDB


MDAP - Traffic Congestion
Cong
bruno.heinz.cmu.edu/traffic/congestion_online/
HOME
OFF-LINE TRAFFIC PREDICTION
ON-LINE TRAFFIC MANAGEMENT
LOGOUT
Link Parameter Updates
ID:9081, Capacity Drop: 40%
SUBMIT
Legend Threshold
30 50 70
CHANGE THRESHOLD
DEFAULT
DMS Info
no 3
name DMS-06-073
status EXISTING
county PHILADELPHIA
software DYNAC
dmstype PERMANENT
direction NB
next_exit 22
description I-95 NB after Washington Ave/Columbus Blvd.
link 10086
display Drive with care!
Grayscale
Streets
0–0.3
0.3–0.5
0.5–0.7
0.7–1
RITTENHOUSE SQUARE
HAWTHORNE
MARKET STREET
7/30/2016, 2:45:00 PM
1fps
Leaflet | © OpenStreetMap contributors, © CartoDB
bruno.heinz.cmu.edu/traffic/congestion_online/

Thanks!

Contact Info:
Carnegie Mellon University
Mobility Data Analytics Center
Wei Ma: weima@cmu.edu