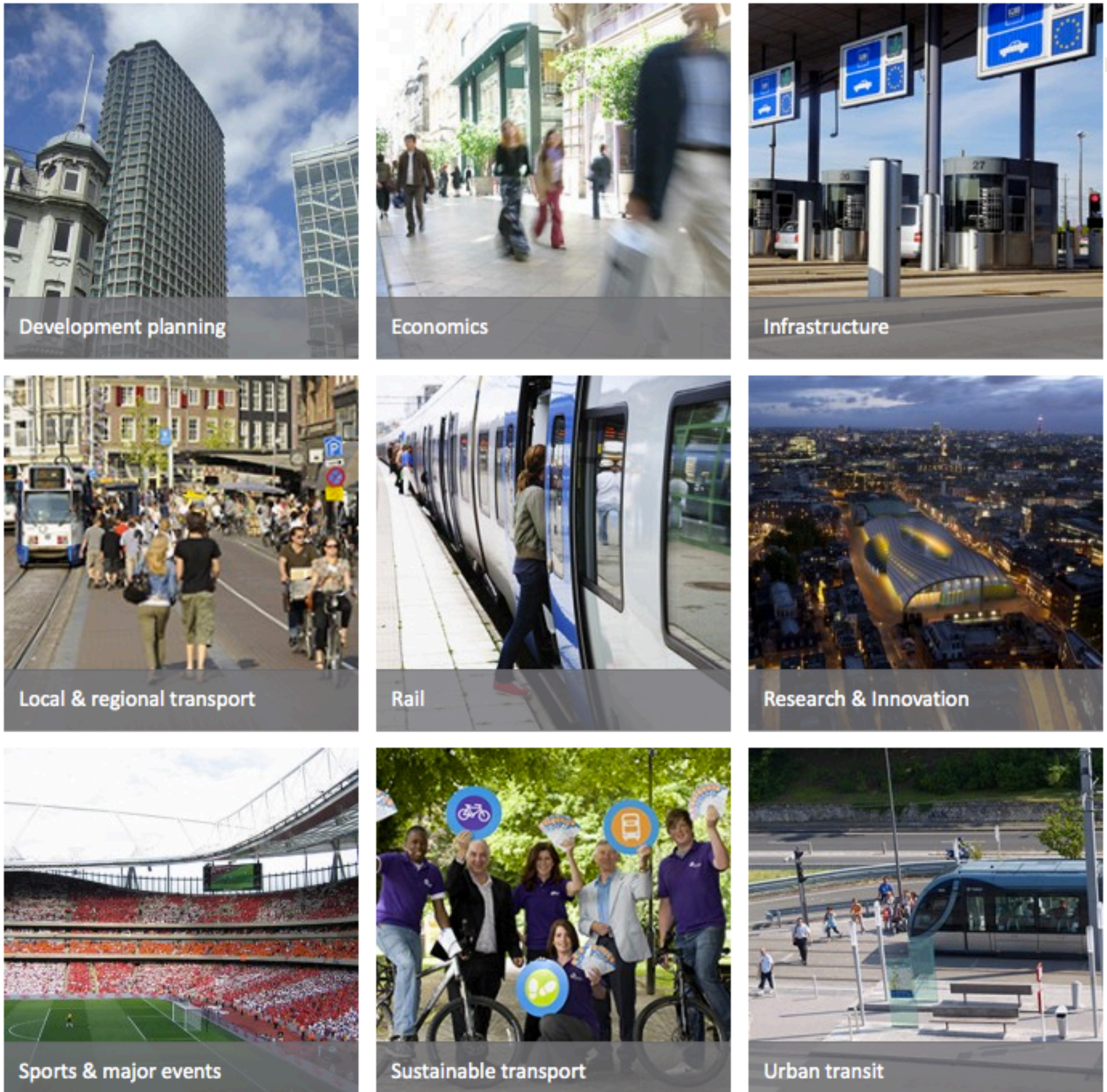
Big Data to Change Urban Demand Modeling Paradigms

Marta C. Gonzalez

Why do we need a Transportation Model?



steer davies gleave

Founded in 1978 has 16 Companies around the World

- Bogota
- Bologna
- Boston
- Leeds
- London
- Los Angeles
- Madrid
- Mexico City
- New Delhi
- New York
- Rome
- San Juan
- Santiago
- Sao Paulo
- Toronto
- Vancouver

Where do they apply a Transportation Model?

Projects [edit]

- Transport planning and design for London's new financial district, Canary Wharf
- Eurotunnel (largest Public Private Partnership project in the world)
- New Lines Programme – developing the case for high-speed rail in the UK
- Multi-modal transport masterplan for Abu Dhabi
- Design of Bogota's award-winning Transmilenio bus rapid transit network
- 30-year transit masterplan for Sacramento
- Review of air passengers' rights for European Commission
- Transport planning for the 2012 London Olympics
- Transport plans for World Cup 2014 in Brazil
- Best practice interchange guidelines for London
- Bikeability cycle training

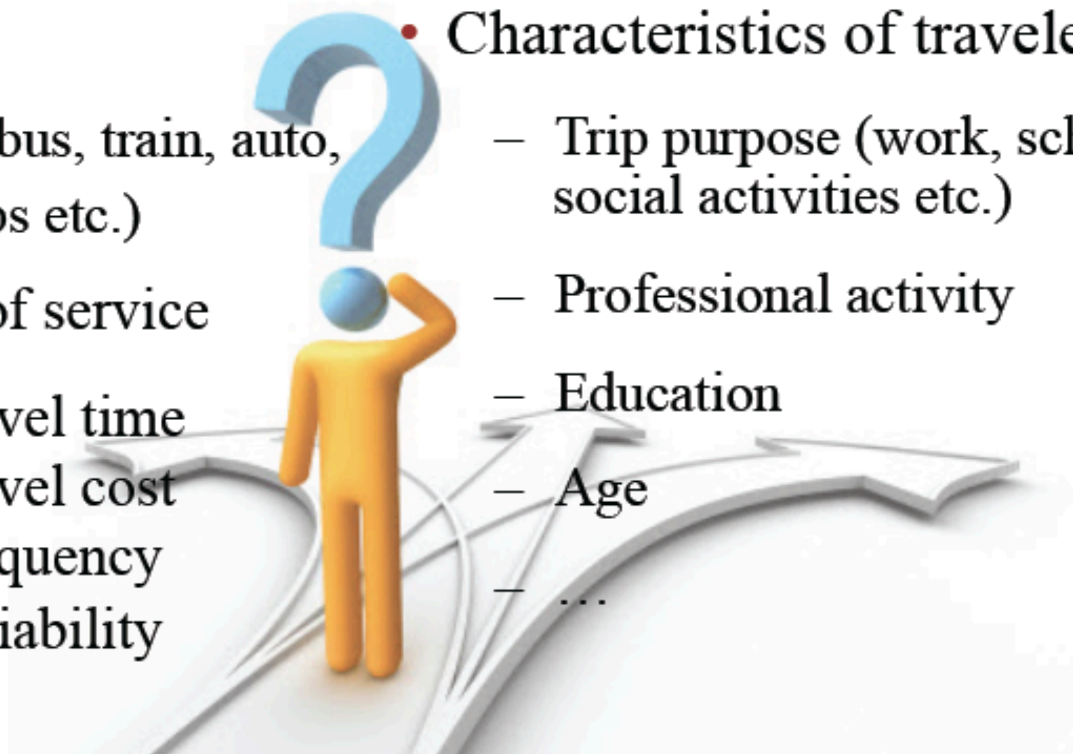
steer davies gleave

Transportation demand modeling

- Choice models based on the attributes of the transport alternatives and characteristics of travelers

- Attributes
  - Mode (bus, train, auto, air, ships etc.)
  - Level of service
    - Travel time
    - Travel cost
    - Frequency
    - Reliability
    - …

- Characteristics of travelers
  - Trip purpose (work, school, social activities etc.)
  - Professional activity
  - Education
  - Age
  - …



How much a Travel Survey Cost?

13. ESTIMATED COST

Cost estimates must be broken down by the basic project phases. There are two basic classifications, but a breakdown below this level to tasks will be required:

A. Instrument development and design, including determining the sampling technique and preparation for the survey.

B. Survey fielding, retrieval, checking and coding.

It is the intention at present, to choose the contractor as a group of MPOs. The contract will at that point be broken up into 5 contracts with the legal contracts for each MPO being entered into. The development costs will at that stage be pro-rated (based on sample size) to each of the participants, in essence as a per completed household overhead rate. In the event that any one of the participants withdraws at that stage, it is important that there is a mechanism for re-allocating the shared development costs.

14. PROJECT RESOURCES:

A. Metro: approximately $550,000
B. State of Oregon (Statewide) approximately $ 500,000
C. Southwest Washington RTC approximately $ 200,000

Source: DOT, 1996, prepared by Cambridge Systematics

- Opportunity:

  Call data records and ICT (information and communication technologies) from mobile devices is a proxy for individual movement.

- Challenge:

  Can we use Information and ICT to Simplify the Generation of Urban Transportation Models and change the current paradigm?

Greater Good: Optimize the whole, not an individual!

TRB TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

Home Contact Us Directory

About TRB | Annual Meeting | Calendar | Committees & Panels | Programs | Projects | Publications | Resources & Databases

NCHRP 08-95 [Active]

Cell Phone Location Data for Travel Behavior Analysis

Project Data	
Funds:	$250,000
Staff Responsibility:	Lawrence D. Goldstein
Research Agency:	Cambridge Systematics Inc,/MIT
Principal Investigator:	Kimon Proussaloglou
Effective Date:	5/30/2014
Completion Date:	3/1/2016
Comments:	Research in progress

The objectives of this research are to: (1) evaluate the extent to which cell phone location data and associated products accurately depict travel and (2) model travel patterns and behavior. The results will be used primarily by transportation planners, travel modelers, and travel survey practitioners.

This will a review work of all our methods, to be published in 2016.

Our Work

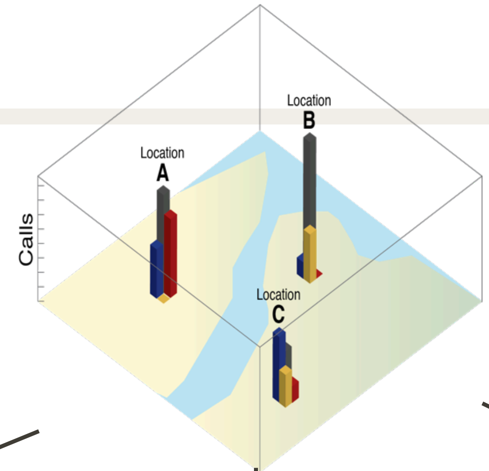
- How do people move around the city?
  - **Individual Trip Profiles**
  - **Contextual Mobility Patterns**
  - **Probabilistic Dynamic Model of City Traffic**

- How do we optimize solutions?
  - **Portable platform**
  - **"Routing" Test Case**

Mobile Data is the next frontier, in which people and spaces are connected in time!

Overview

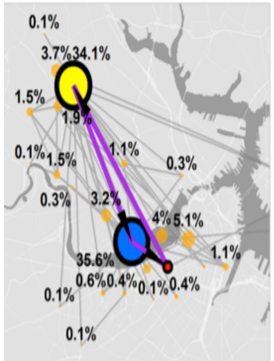
Information and Communication Technologies (ICT) Discrete Data

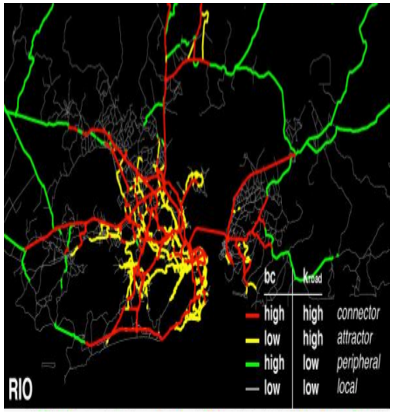


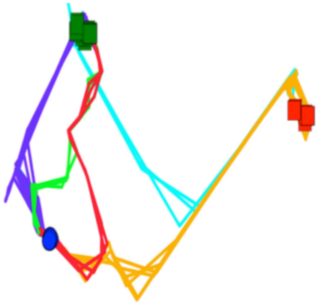
Inference of daily Journeys

Validated Models of Urban Traffic

Optimization: Routing Applications







From discrete data to social benefits

What are we using Travel Model for?

We have developed a portable pipeline to generate urban demand models from mobile phone data and we are interested now in its applications for better cities.

Examples of current projects:

1-WithinBlocks: Building Occupancy Models for Energy Modeling
2- Metro: Helping to plan the first metro in Riyadh
3- Planning for Electric Vehicles in the Bay Area
4- Understanding Travel time Reliability
5- Understanding Air quality in Bejing integrating AQ Sensors and a Traffic Model
6-Distributing Travel Demand during the Olympic Games

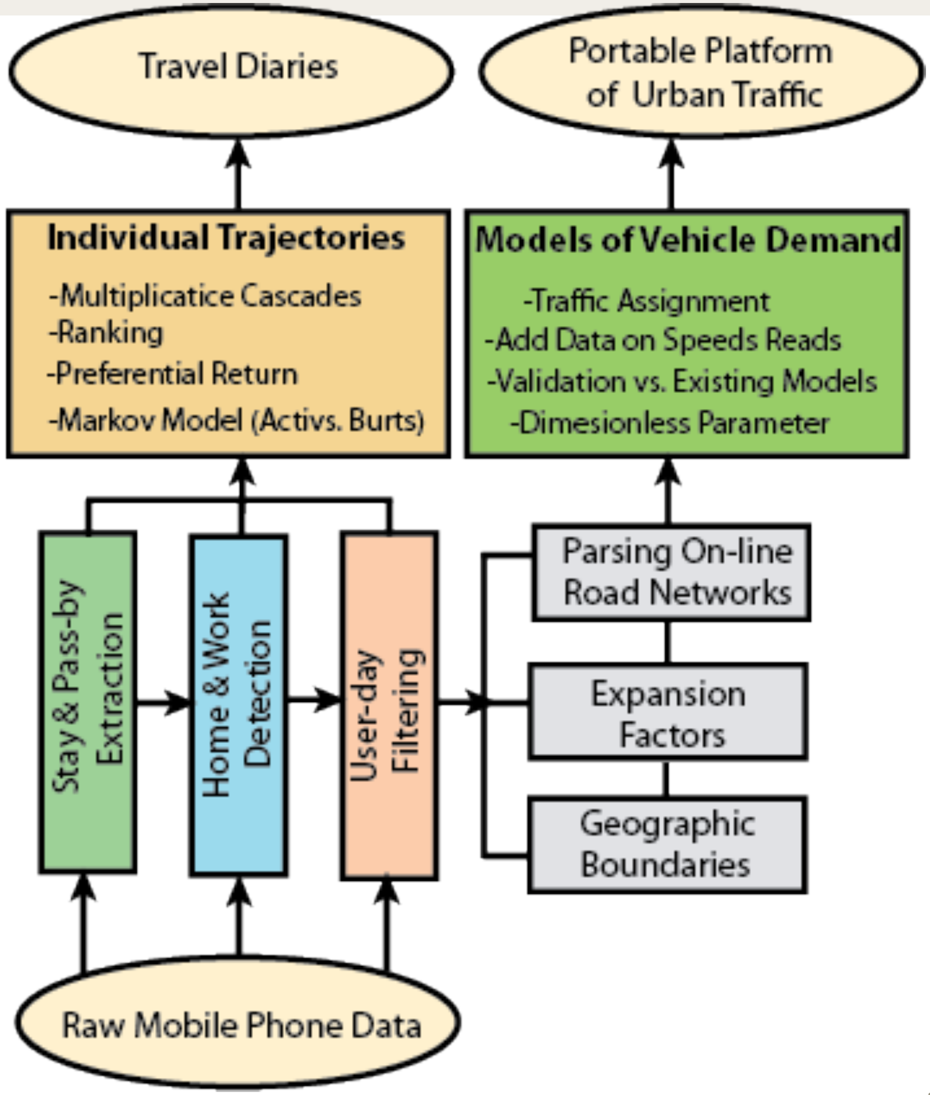

Jiang
Çolak
Yang
Sturt

Doctoral Students

Masters Students

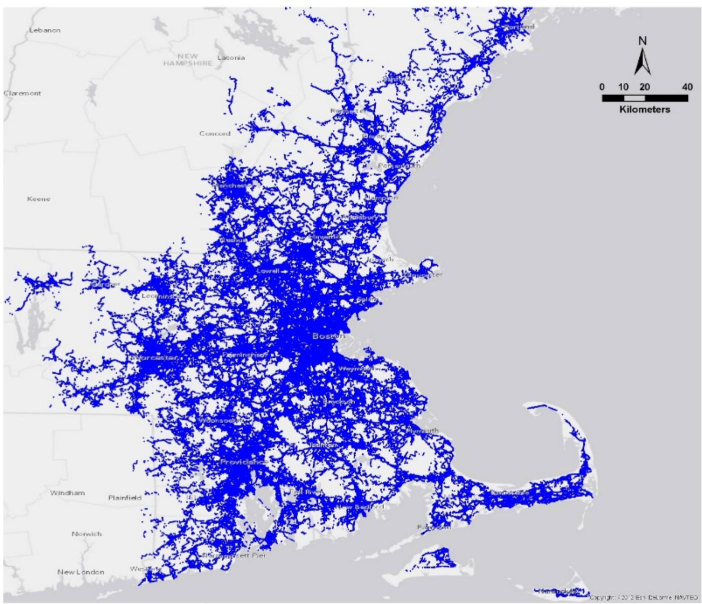
Former Postdocs

Spatiotemporal Mapping of Mobile Phone for Urban Demand Modeling

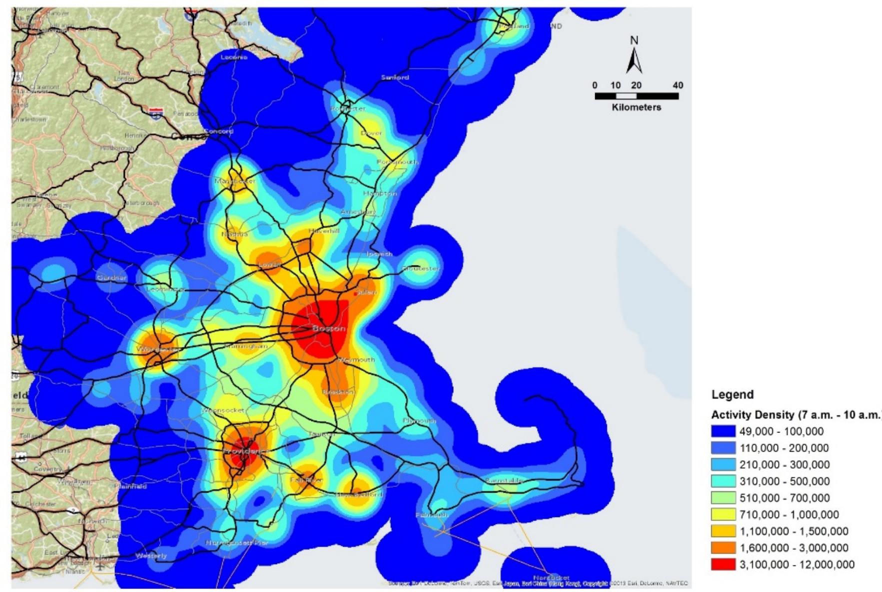


Overarching Goal

How to extract human daily activities (e.g., types, sequences, and chains) from these massive, passive and noisy Big Data that are comparable to travel demand models from travel surveys? and asses the role of Social Routing?



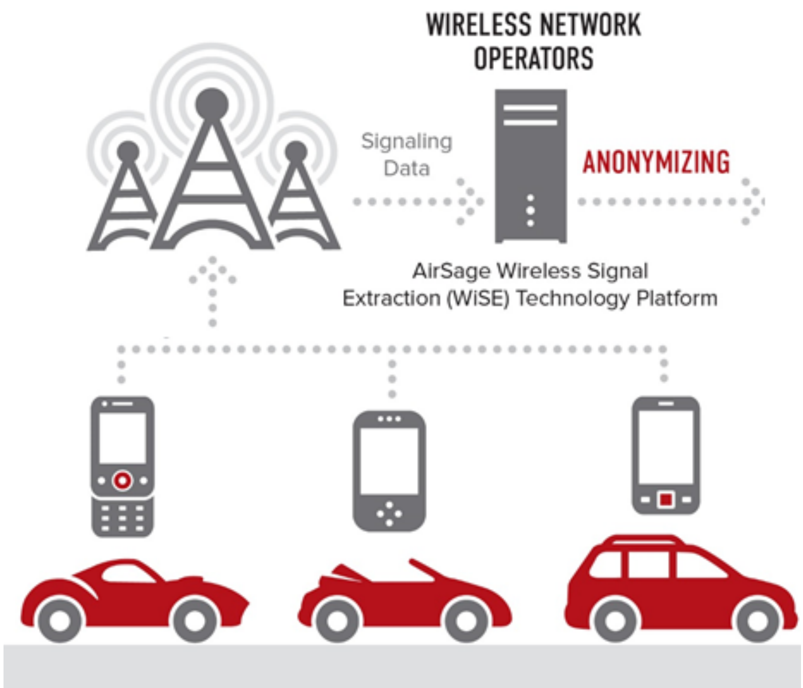
1.9 million total users observed in the 2 months, in Boston 2010.



Human Activity Density 4 P.M.-7 P.M.

Raw Data Description

Traces of People – Where and When



- **800 million of historical location records for 1 million anonymous individuals who use phones in the Boston metropolitan area**
- **Data for one anonymous user:**

Longitude	Latitude	Time
-71.059998	42.356132	1266513700
-71.059730	42.356391	1266513800
-71.063884	42.355315	1266513900
-71.063884	42.355315	1266514200
…	…	…

- **Estimation precision error**: ~ 300 meter

Reference: http://www.airsage.com/Technology/How-it-works/

Detailed Results Paper 1
Modeling Individual Trajectories

Data Sources

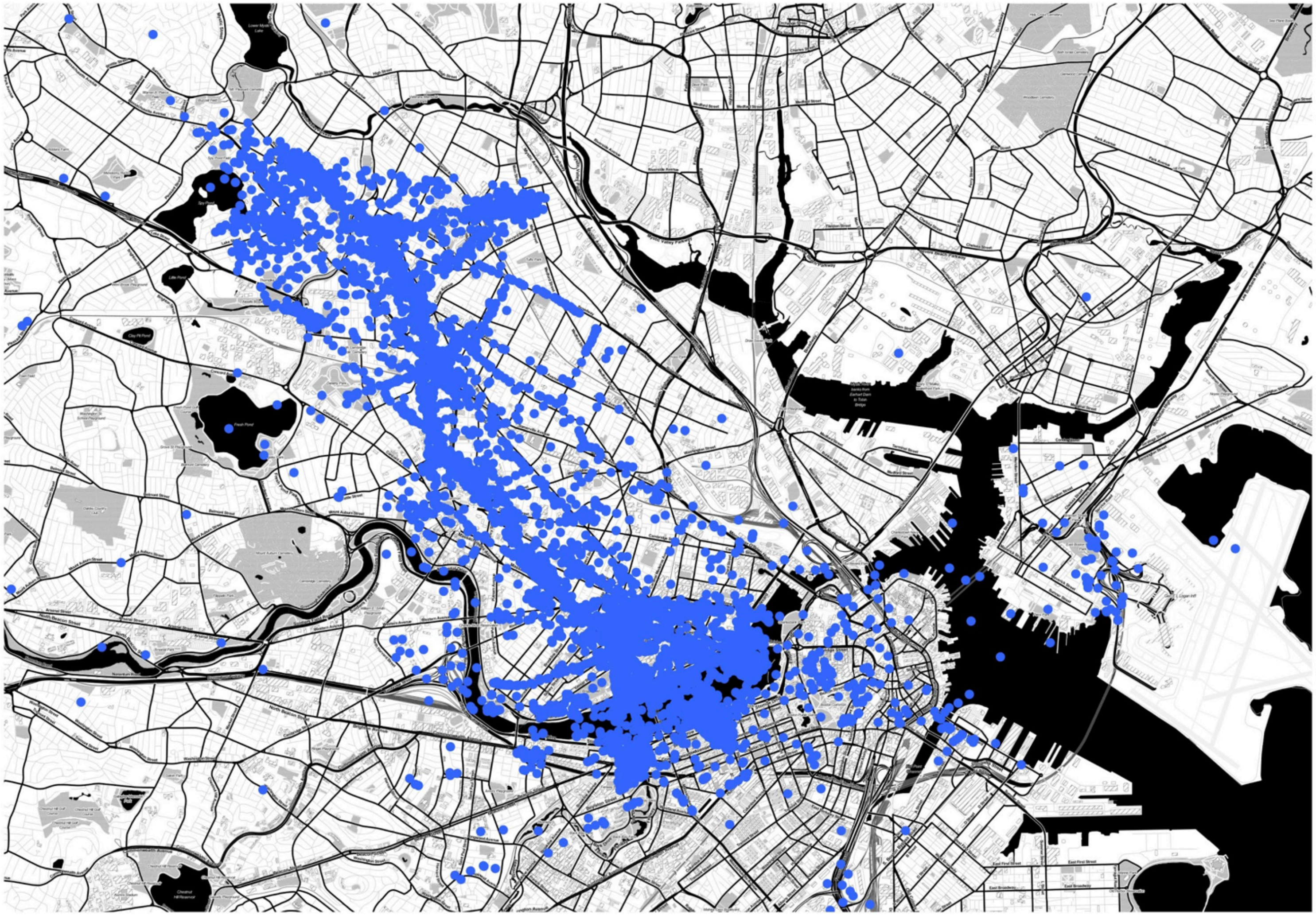
— 2 millions of individual phone users in Boston
(For purchase nationwide in AirSage.com)

— 14 Months of self-collected complete mobile phone data of 1 Student.

Goal

Model Individual Trajectories
(resolution 10min and 300mts radius)

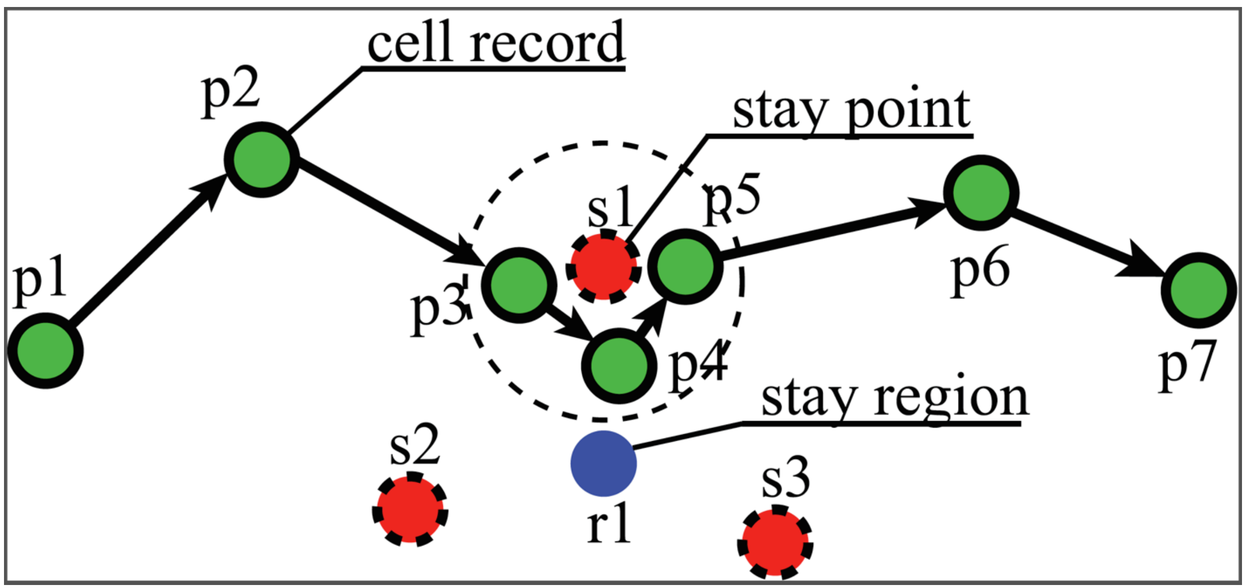
Stay Extraction



Stay extraction

How to detect a stay?

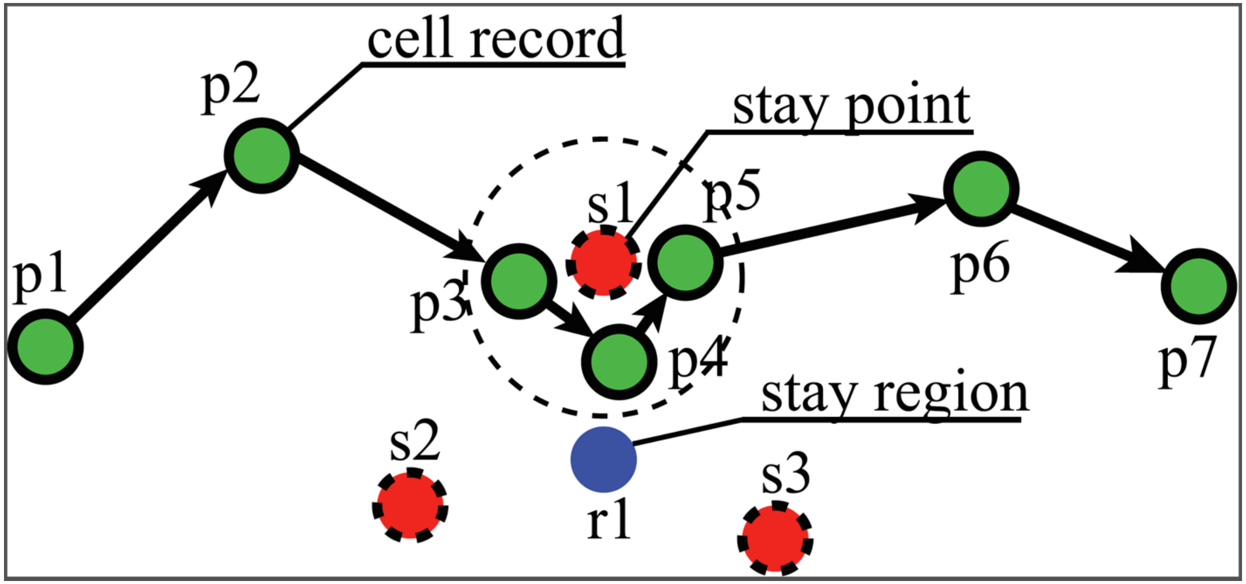
- Consecutive records within a circle of certain radius (300m)
- The time interval between the first/last record is larger than certain threshold (10min)



Stay region extraction

From stay to stay region

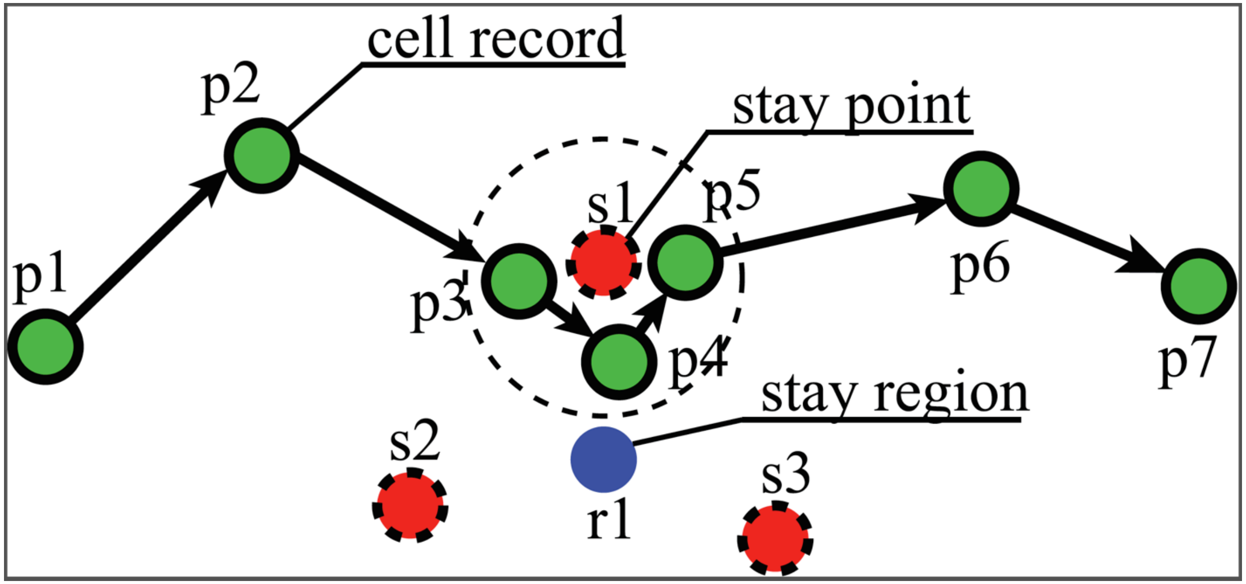
- Stay Region: stays from different trajectories might represent the same location
- Interchangeable with “location”



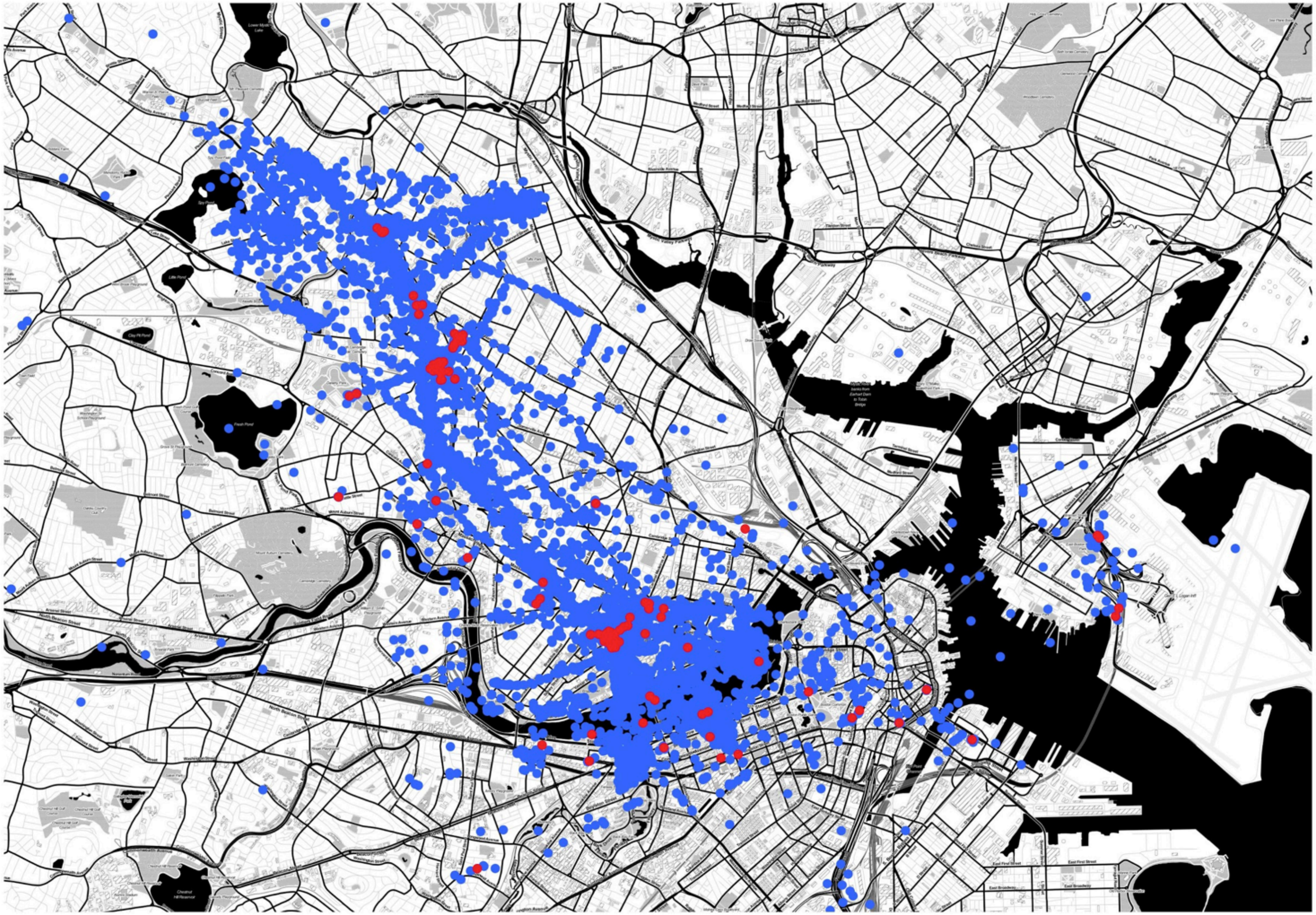
Stay region extraction

How to get stay regions?

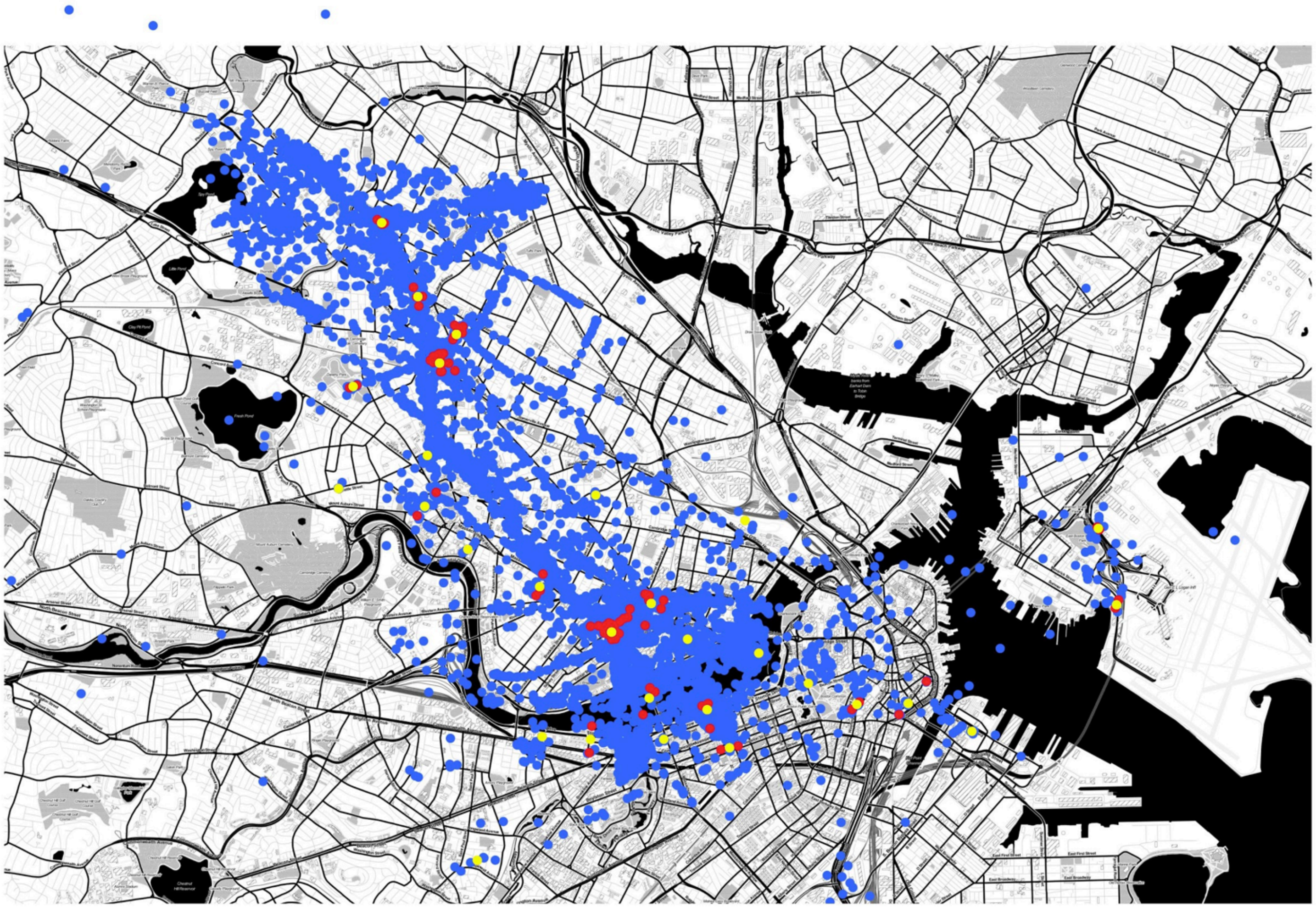
- Grid based clustering
- Divide the entire area into a grid
- Merge stays in neighboring grid cells



Stay Extraction

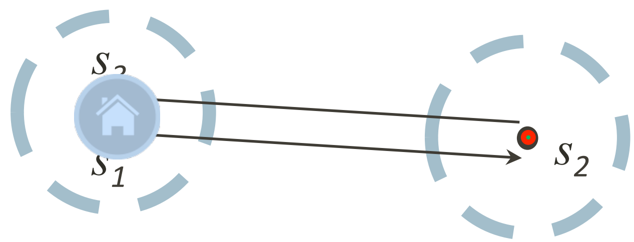


Stay Extraction



Measuring Individual Activities: Home, Work, and Other

- A phone user's "home" is defined as
  - the most frequently region <u>during nights of weekdays</u> & <u>days of weekends</u>
  - over the study period
  - Night time: a parameter (e.g., from 7 pm to 7 am)

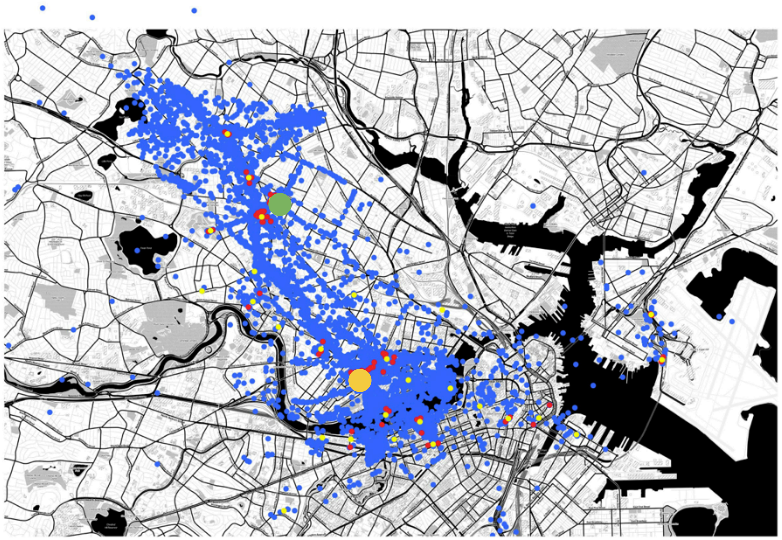


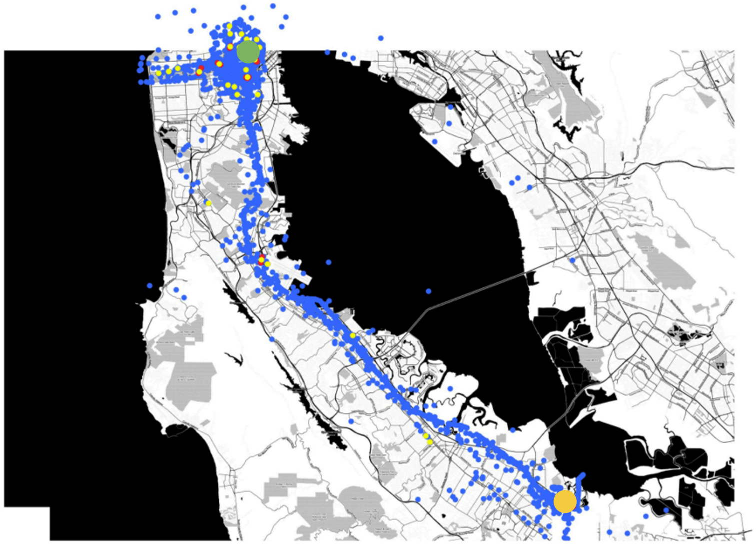
A phone user's "work" is defined as

- the most frequently stay region <u>working hours of the weekday</u>

A phone user's "other" is defined as the rest

Stay Extraction

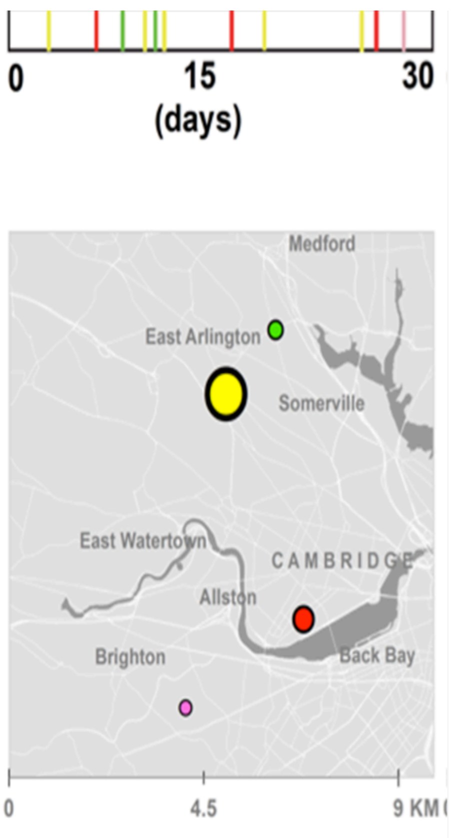


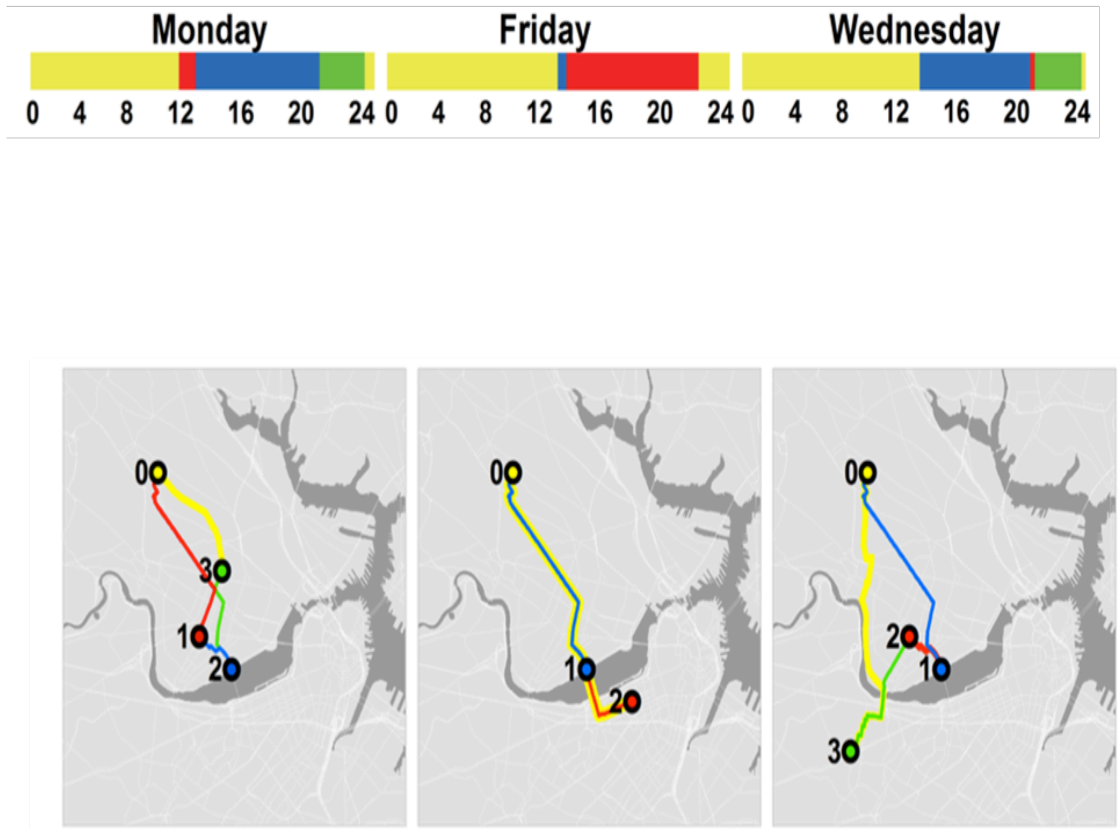


Same user can re-locate by rechecking the criteria periodically

Sparse vs. Active Users

Stays of Sparse User vs. Daily Journeys of Active User



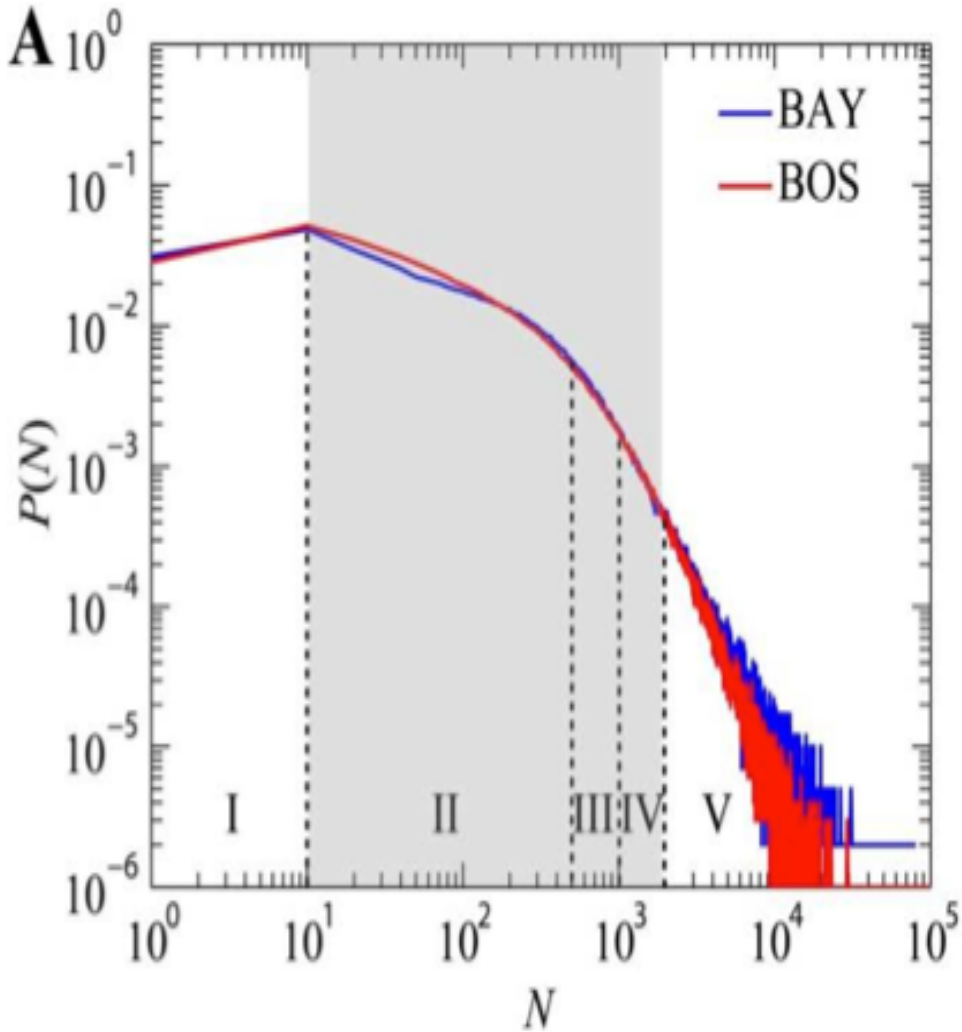


Reconstructing daily journeys is easier for active users.

Distribution of Phone Records per individual

Group II >> Sparse users

Group III and IV >> Representative Active Users

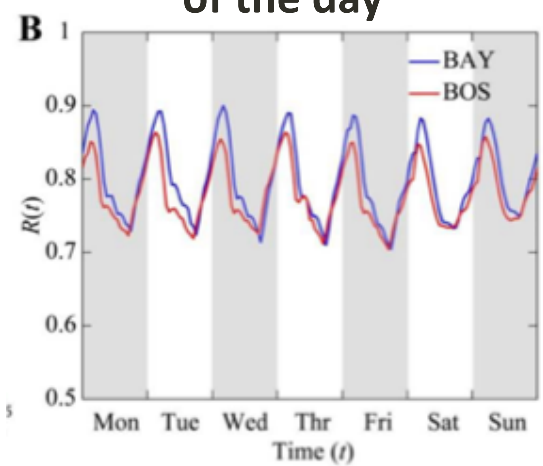


We can learn from the behavior of active users to effectively model sparse users?

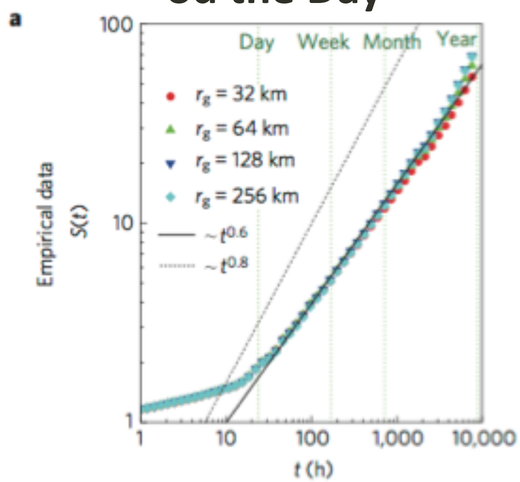
1. What is the likelihood of a trip to occur at different times of the days?
2. How long do they stay in flexible activities?
3. What is the like-hood to visit new Locations?
4. How do they select them?
5. How many trips I likely do per week?

Explore vs. Return

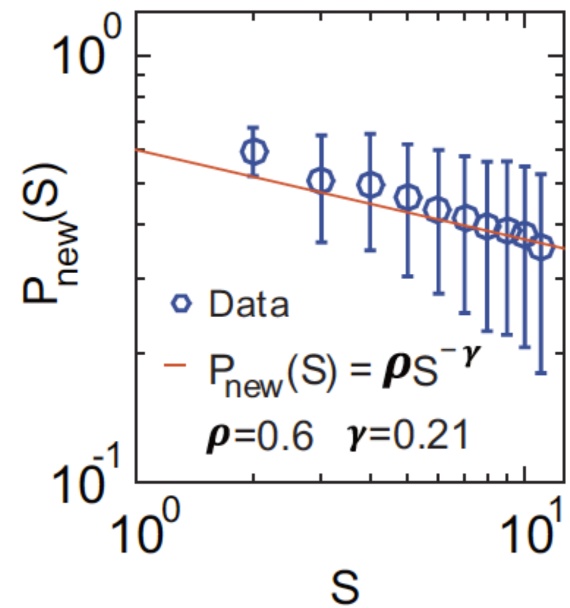
Like hood to be found in the most visited Location by time of the day



Number of Locations by Time od the Day



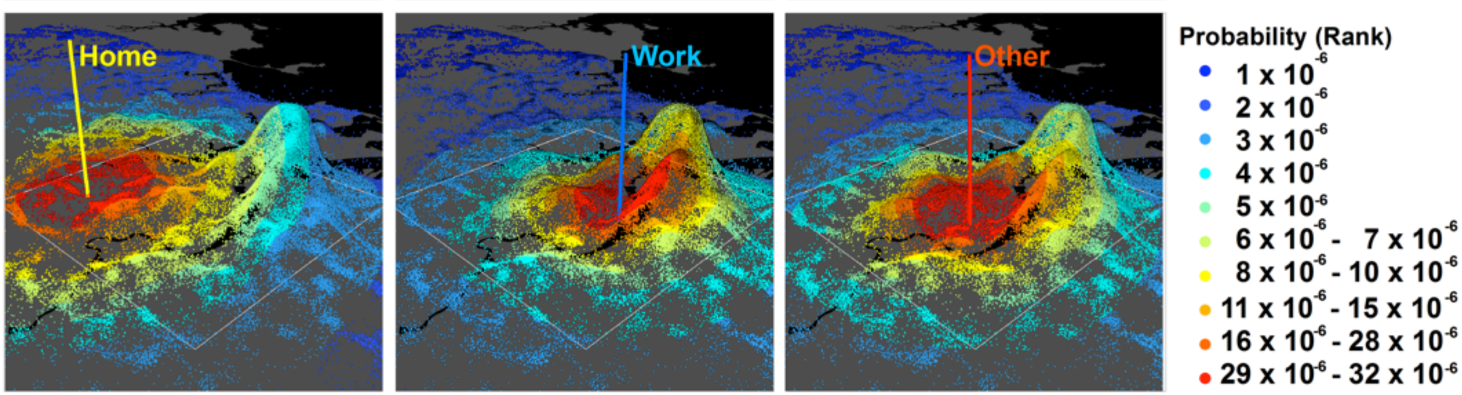
Probability of Exploring having S locations

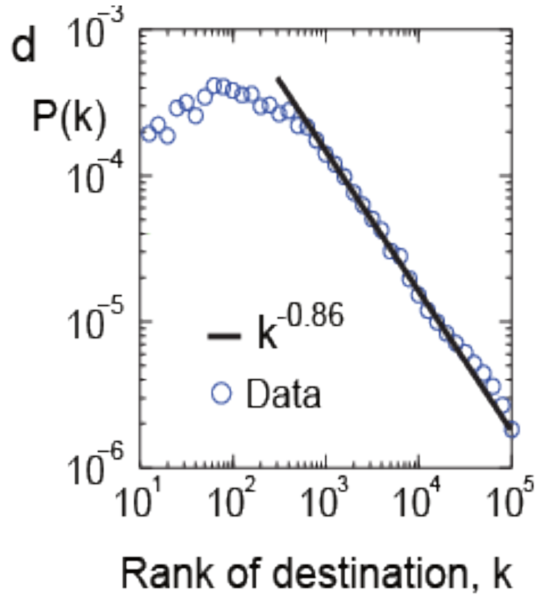


Wang, Pu, et al. "Understanding road usage patterns in urban areas." *Scientific reports* 2 (2012).

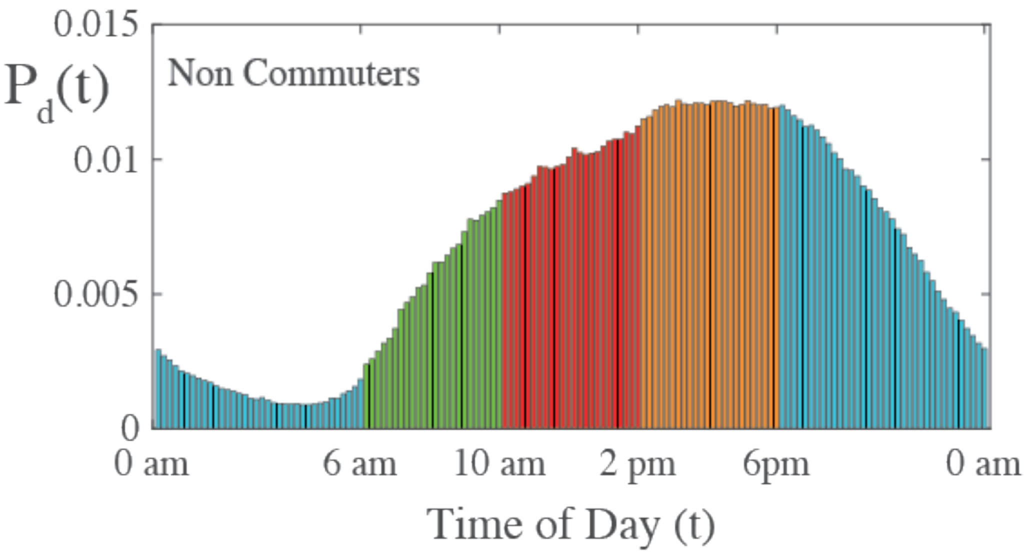
Song, Chaoming, et al. "Modelling the scaling properties of human mobility."*Nature Physics* 6.10 (2010): 818-823.

Ranking of POIs to select New Destination

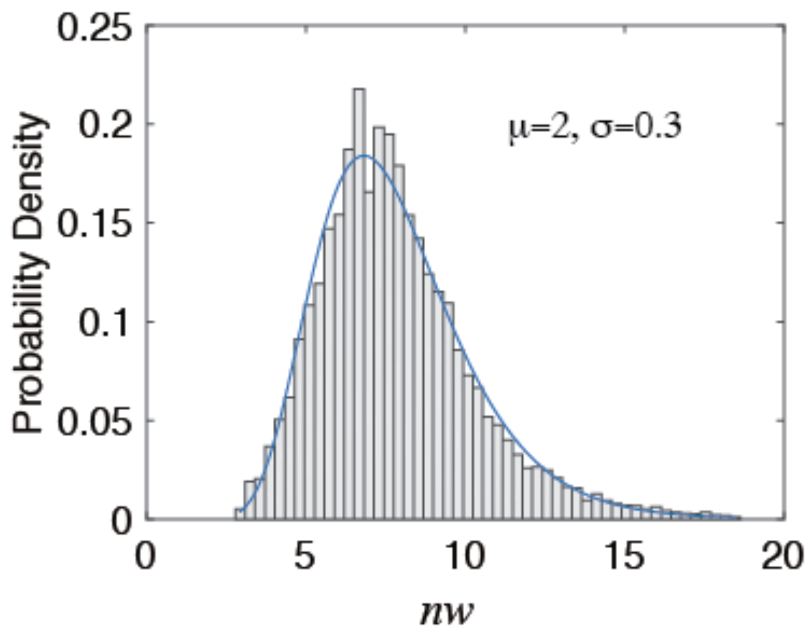




What is the probability of departing from home to do a flexible activity at time t?



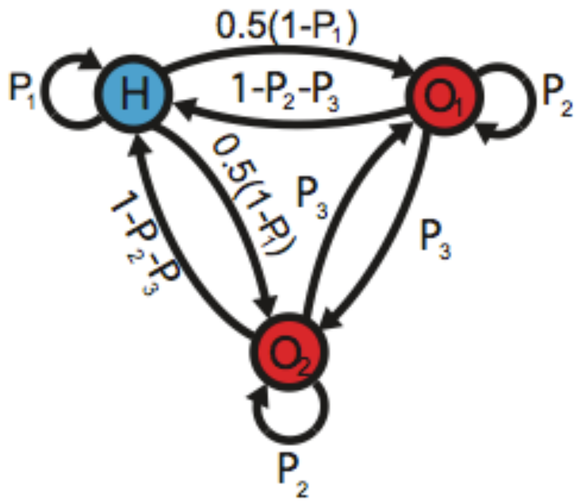
Fraction of Total Trips per time step



Weekly trips from home

$$P_i(t) = n^i_w P_d(t)$$

Time independent Markov Model



$$P(t) = n_w P_d(t)$$

$$P_1 = 1 - P(t)$$ Moving from "home"

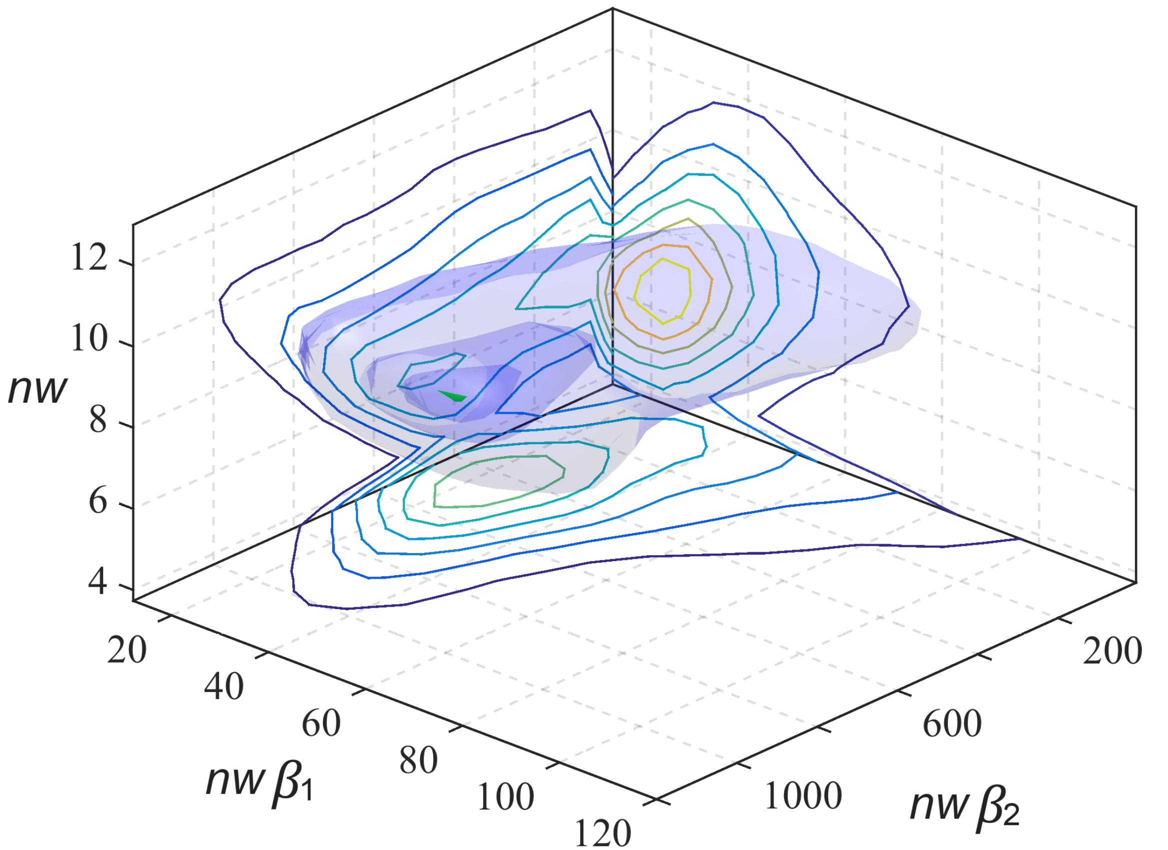
$$P_2 = 1 - \beta_1 P(t)$$ Moving from "other"

$$P_3 = \beta_1 P(t) \beta_2 P(t)$$ Moving from "other" to "other"

β_1 Generates shorter stays in the flexible location state.

β_2 Generate Different number of activities in a row per **active** cycle

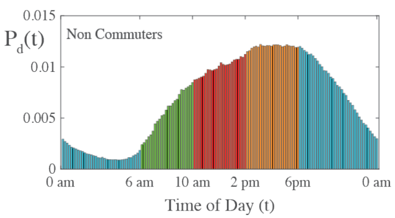
Measure the 3 key features from Active Users



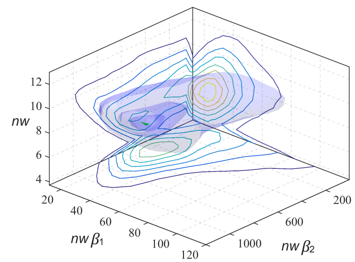
The Model

Features Extracted from data of Active Users

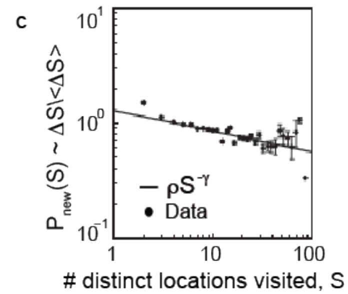
Global Trip prob.



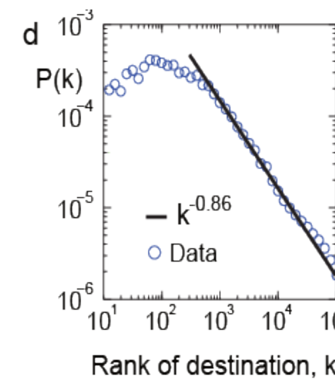
Individual Mobility Rates



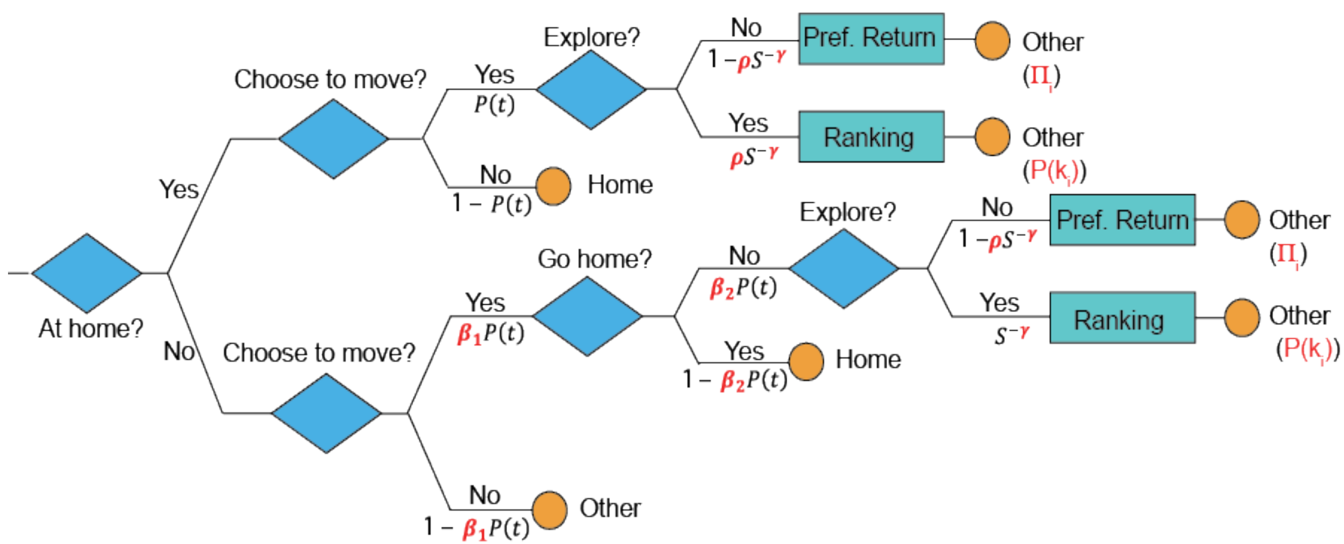
Preferential Return



Ranking of Explorations



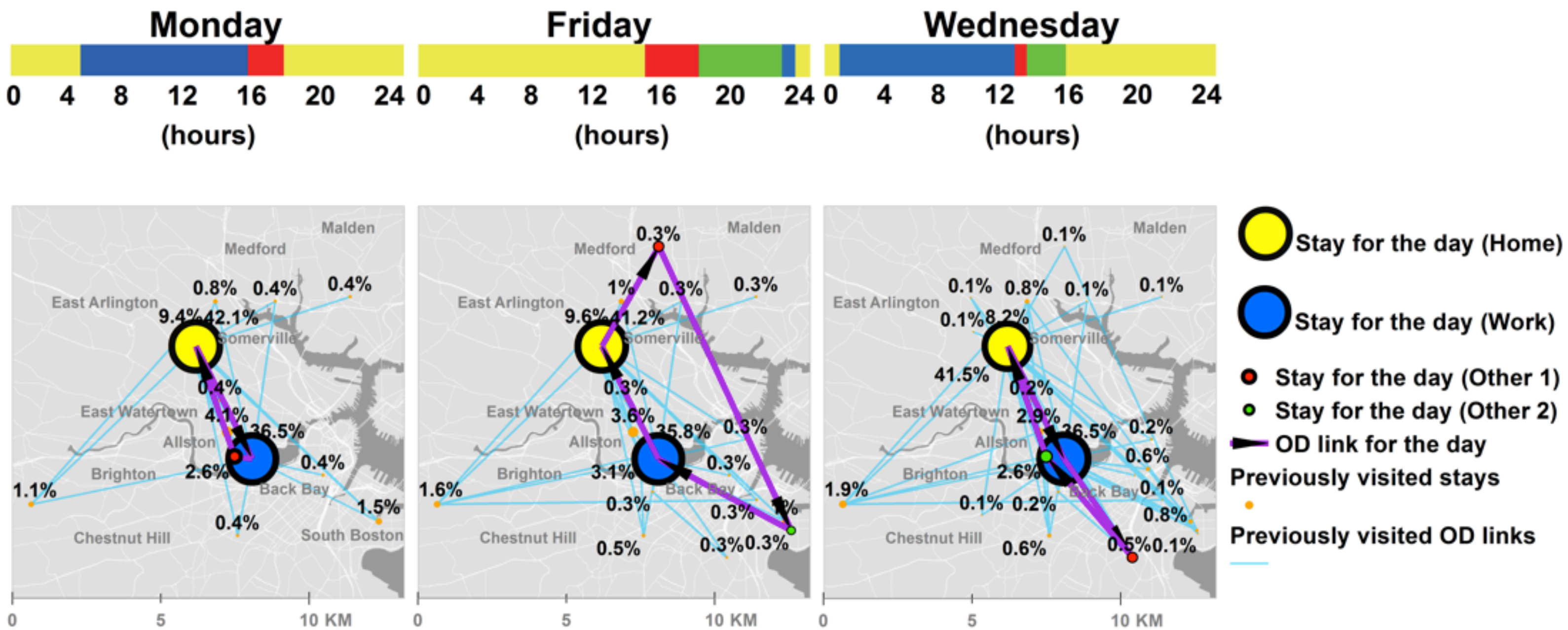
Flowchart of the Model



 School of Engineering

Marta C. Gonzalez

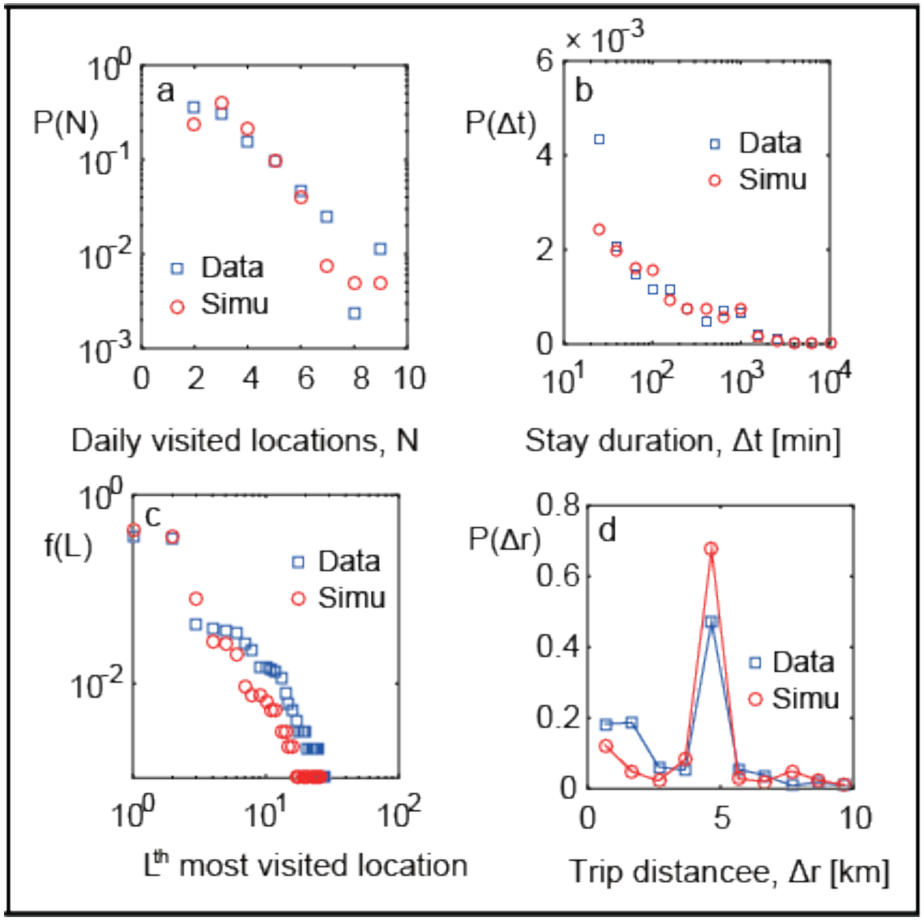
Models Results



Modeled Trajectories of our Student
and see the results (only home and work data used)

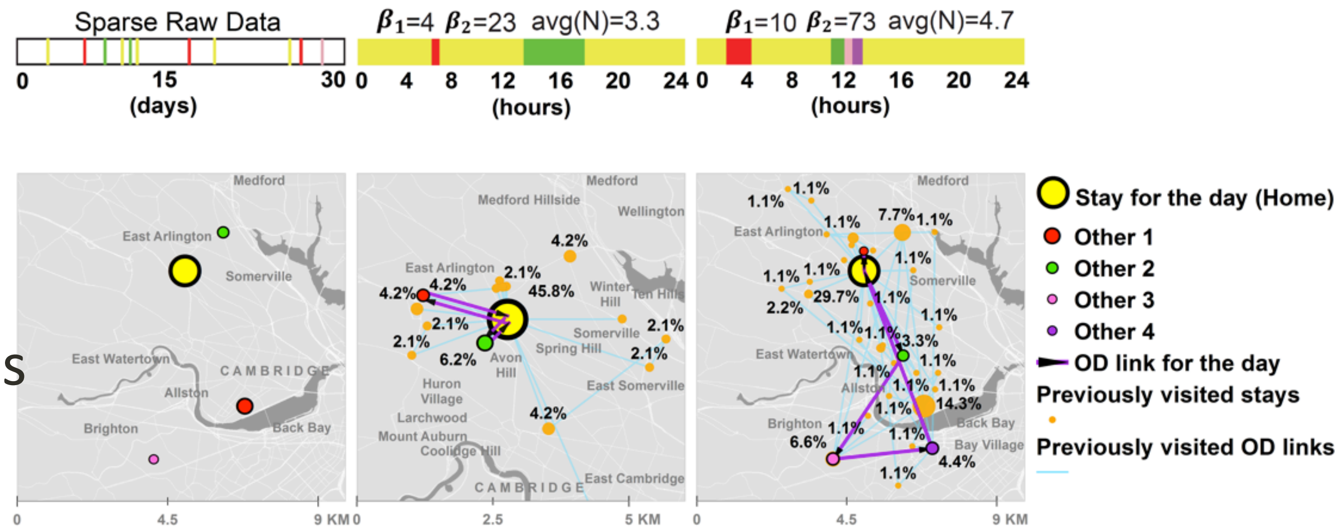
Individual Mobility Patterns

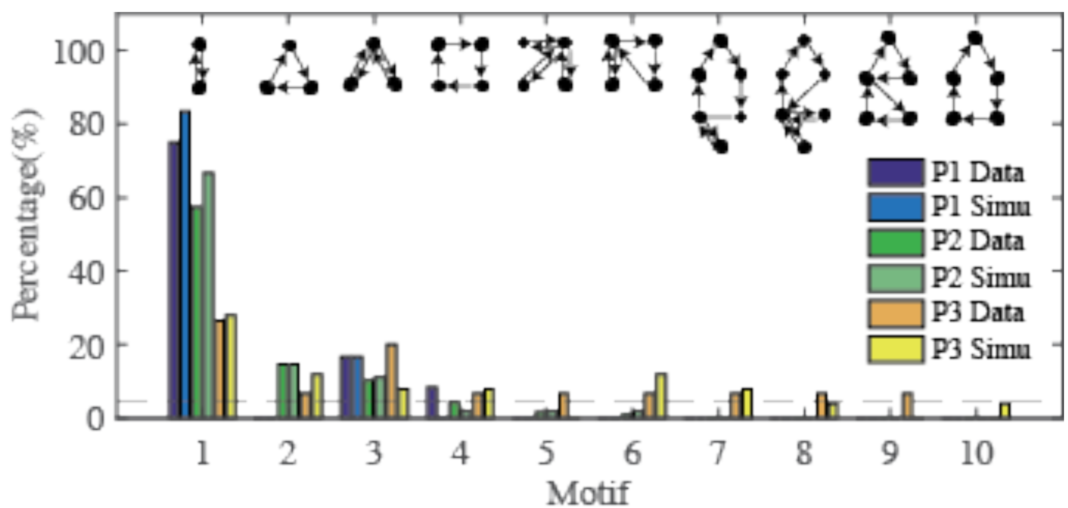
Student vs. his synthetic version



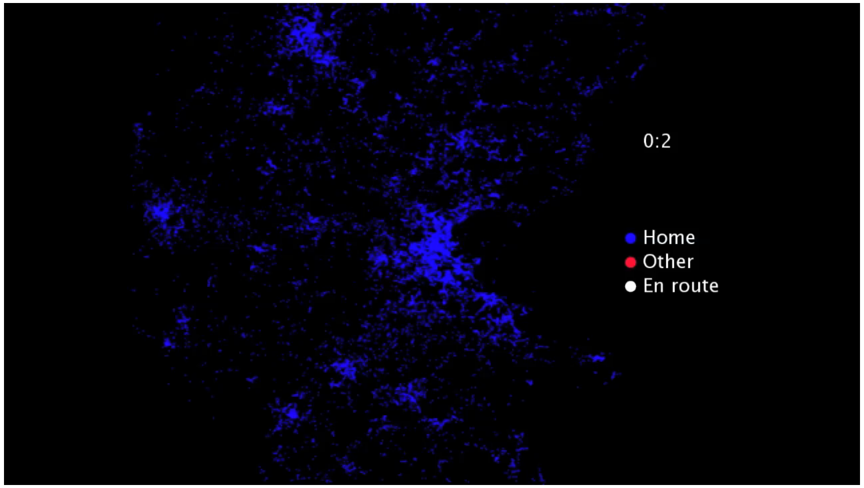
Models Results

Modeled Trajectories From Sparse Data of sample User (previous locations used)





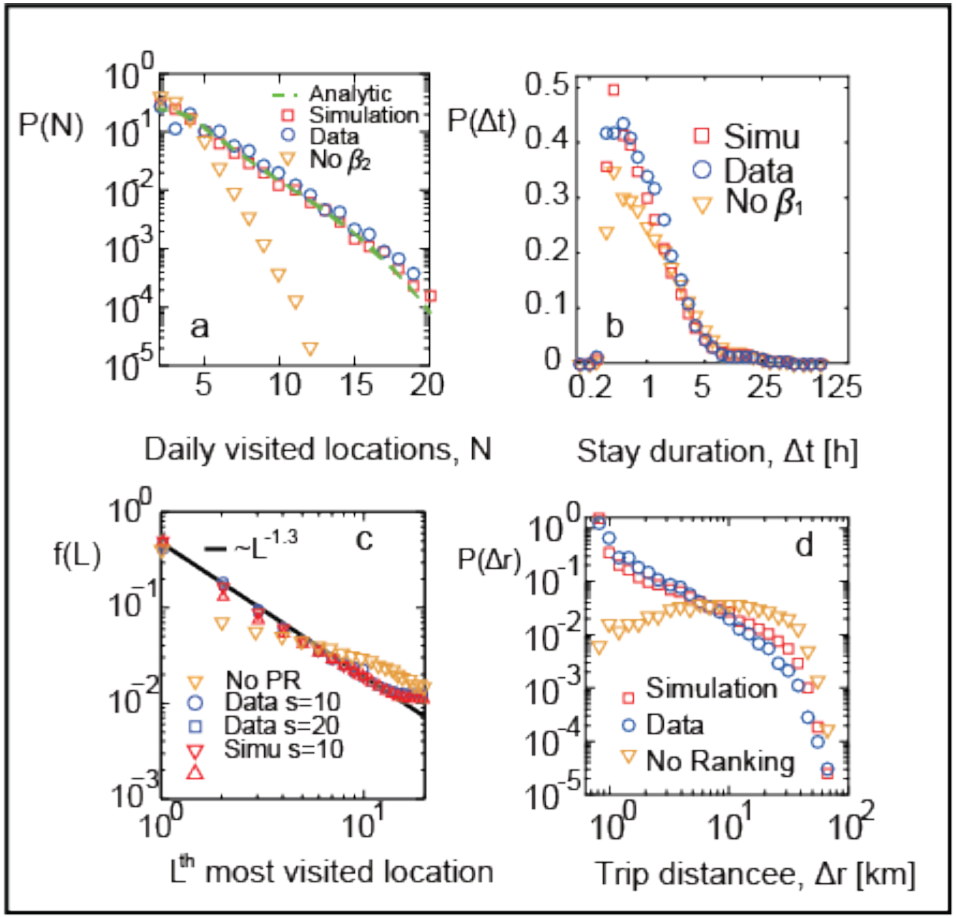
Models Results



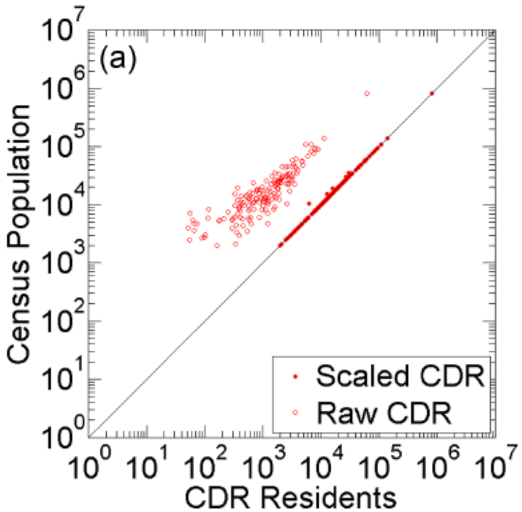
The combined algorithm

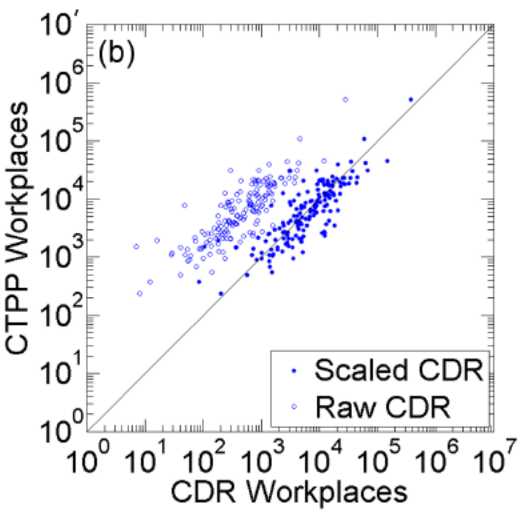
Input:
$p(t)$, ρ, γ, α, β_1, β_2, the home location and the set of other locations
Output:
Location at each time step;
Set $t = 0$; $l = home$; $S = 0$; //S is the number of visited location
while $t < t_{max}$ **do**
  if $l == Home$
    if $rand < p(t)$ //decide to move
      if $rand < \frac{\rho S^{-\gamma}}{\rho S^{-\gamma}+0.6\times(1-\rho S^{-\gamma})}$ //Choose a previously unvisited location;
        Choose rank k location with probability $p(k) = \frac{k^{-\alpha}}{\sum_{l=1}^{M} l^{-\alpha}}$;
        $S++$;
        $l = other$;
      else
        Choose a previously visited location k with $p(k) = f(k)$;
        $f(k)++$;
      end if
    end if
  else
    if $rand < \beta_1 p(t)$ //decide to move
      if $rand < \beta_2 p(t)$ //go to a place other than home
        if $rand < \frac{\rho S^{-\gamma}}{\rho S^{-\gamma}+0.6\times(1-\rho S^{-\gamma})}$ //Choose a previously unvisited location.
          Choose rank k location with probability $p(k) = \frac{k^{-\alpha}}{\sum_{l=1}^{M} l^{-\alpha}}$;
          $S++$;
        else
          Choose a previously visited location k with $p(k) \sim f(k)$;
          $f(k)++$;
        end if
      else //go home
        $l = home$;
      end if
    end if
  end if
  $t++$;
end while

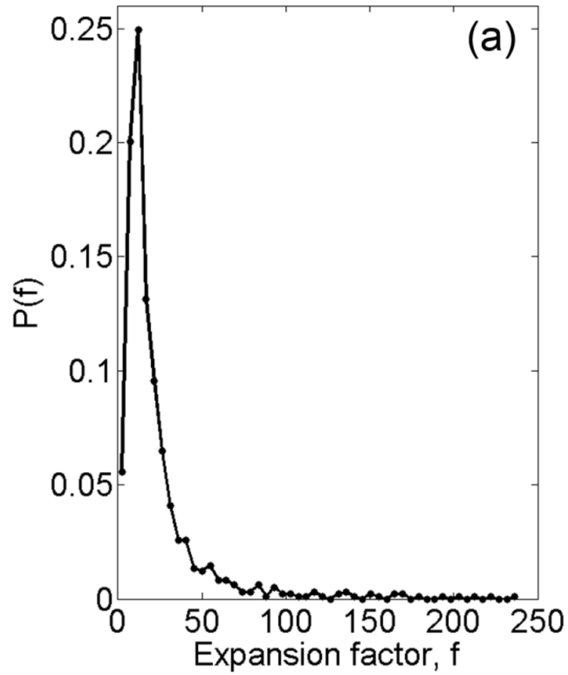
Aggregated Results Boston Trajectories

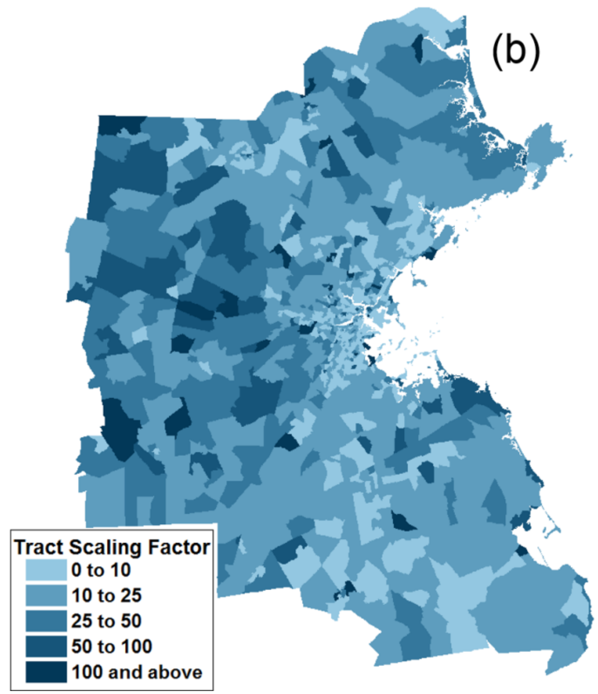


Are Active users Representative?

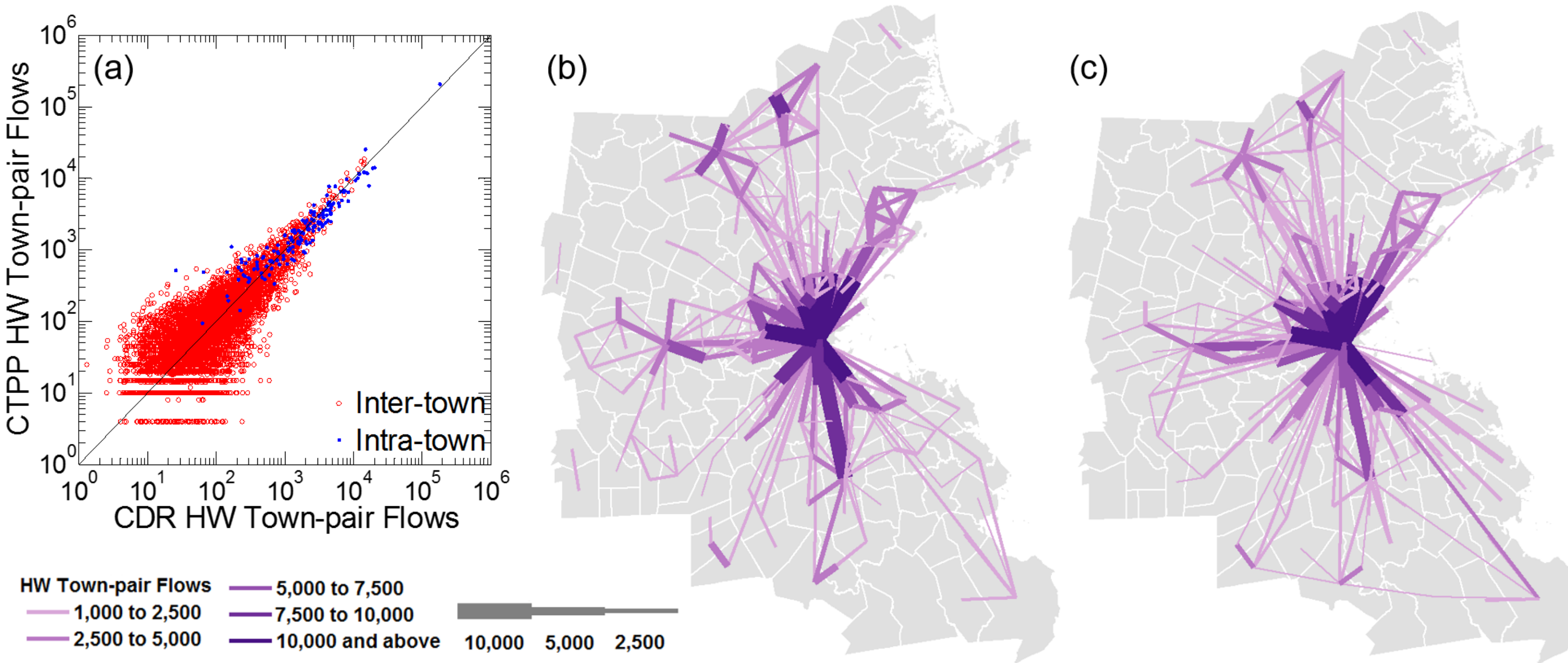








Comparison with Traditional Models



Contents lists available at ScienceDirect

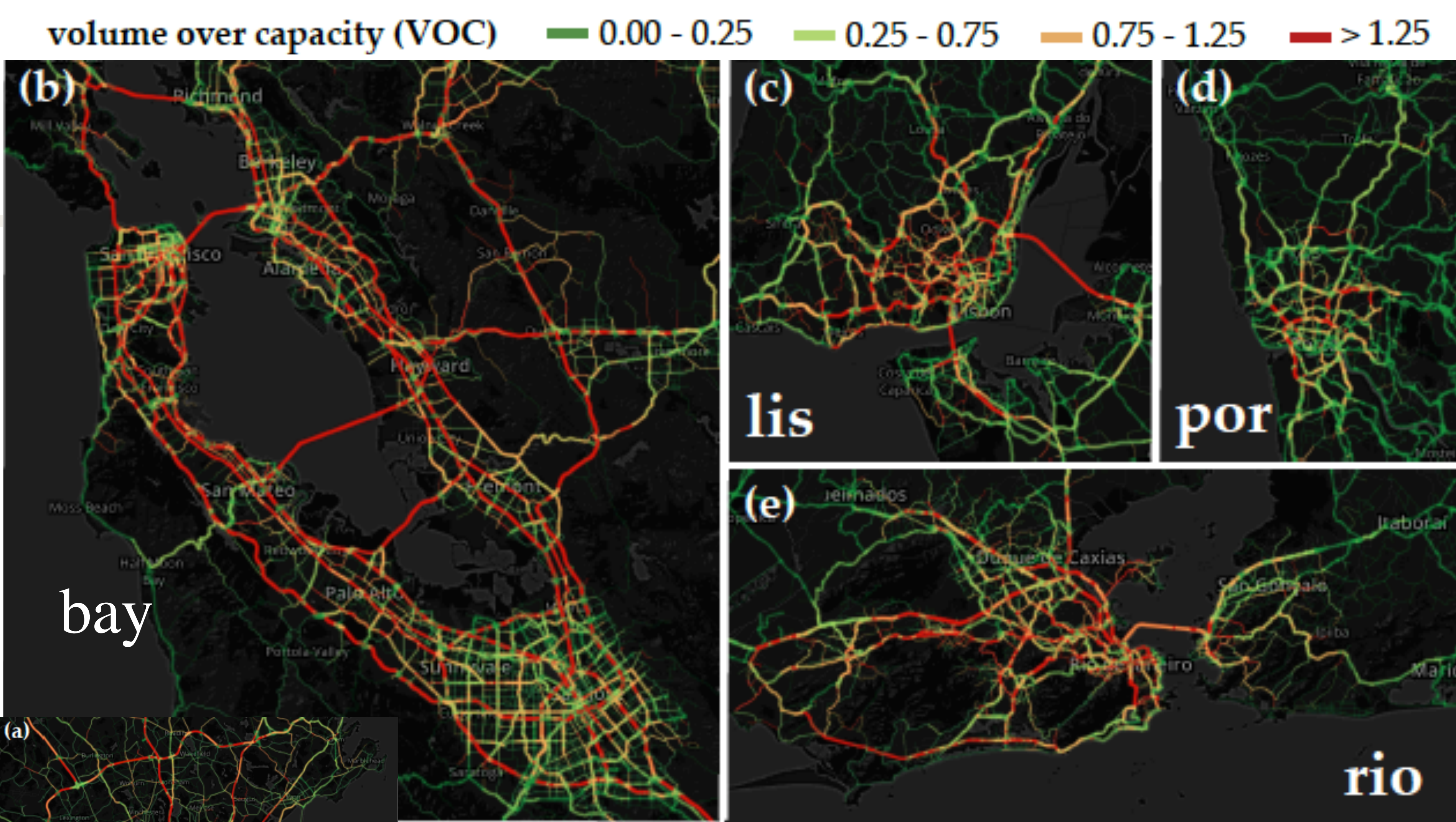
Transportation Research Part C

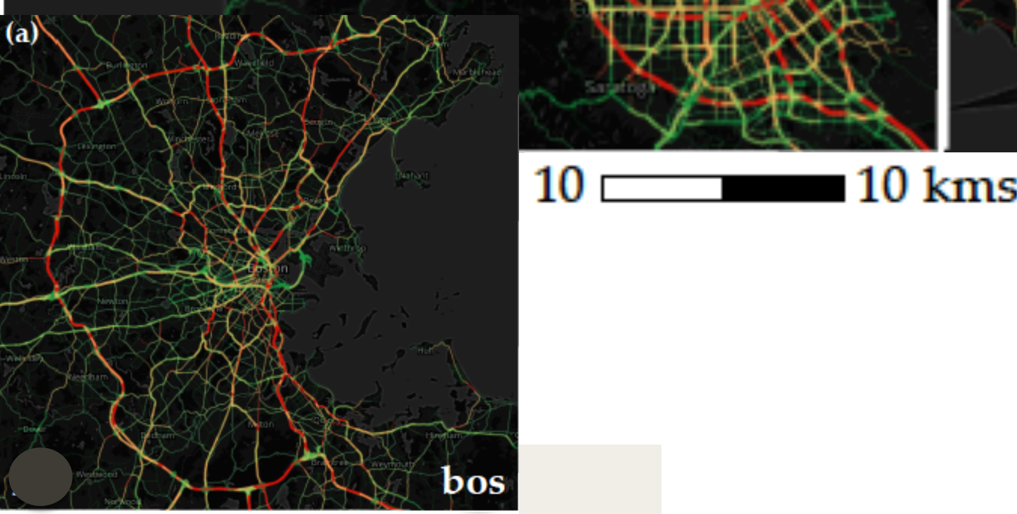
journal homepage: www.elsevier.com/locate/trc

Origin–destination trips by purpose and time of day inferred from mobile phone data

Lauren Alexander [a,*], Shan Jiang [b], Mikel Murga [a], Marta C. González [a]

[a] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States
[b] Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA, United States





Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

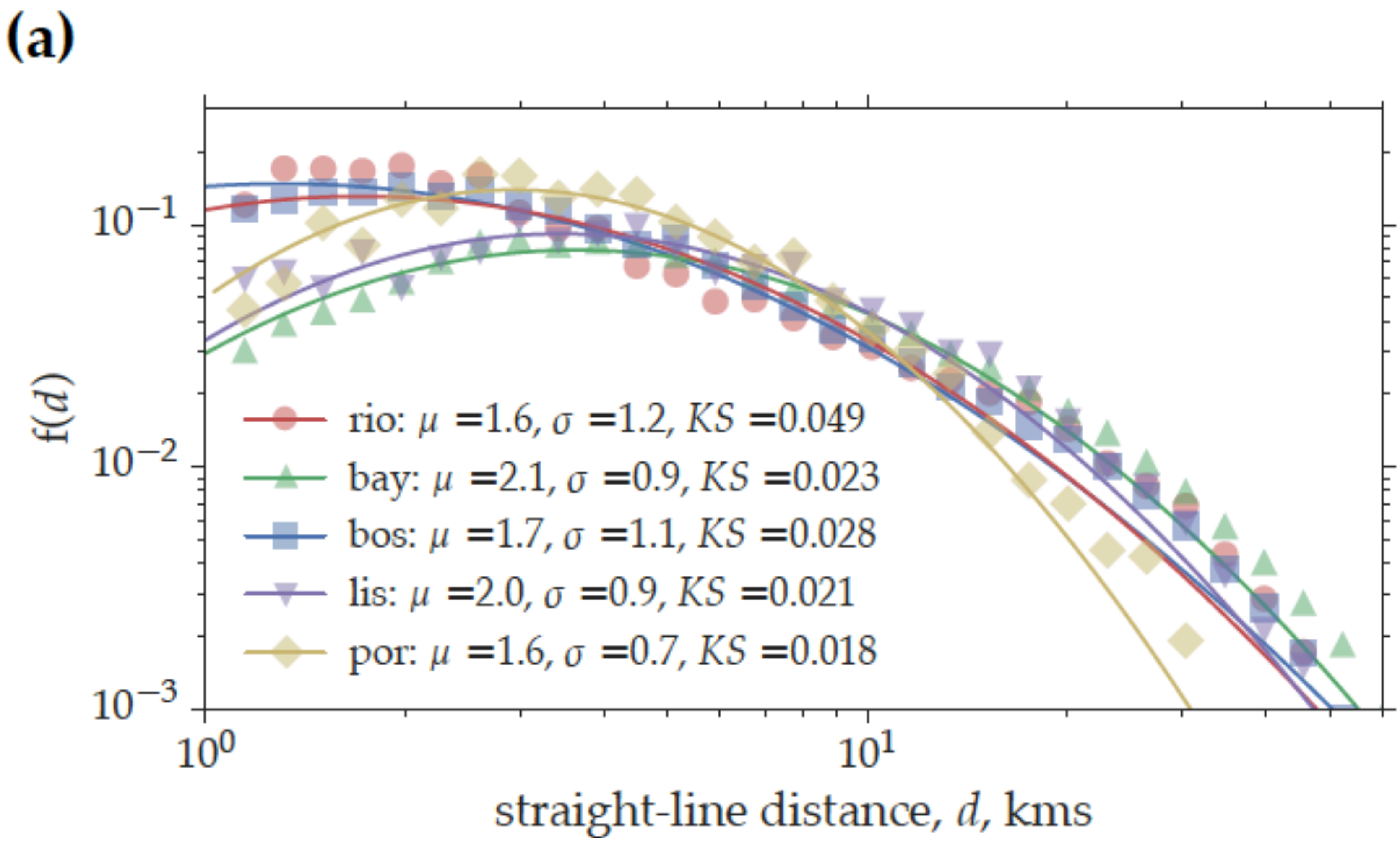
ELSEVIER

The path most traveled: Travel demand estimation using big data resources

Jameson L. Toole [a,1], Serdar Colak [b,*,1], Bradley Sturt [a], Lauren P. Alexander [b], Alexandre Evsukoff [c], Marta C. González [a,b]

[a] Engineering Systems Division, MIT, Cambridge, MA 02139, United States
[b] Department of Civil and Environmental Engineering, MIT, Cambridge, MA 02139, United States
[c] COPPE/Federal University of Rio de Janeiro, Brazil

Commuting Distance



Route Assignment

1. Road networks from OpenStreetMap data.

2. Algorithm B, implements equilibration on a directed acyclic graph (DAG).

3. Keep track of where flow is sent two and from.

$$r_g = 1 - \frac{\sum_{o,d} t_{od} d_{od}}{\sum_{e \in E} t_e v_e},$$

where t_{od} and d_{od} represent the demand and the travel time between an origin and a destination, and t_e and v_e represent the travel time and the volume on a road segment e.

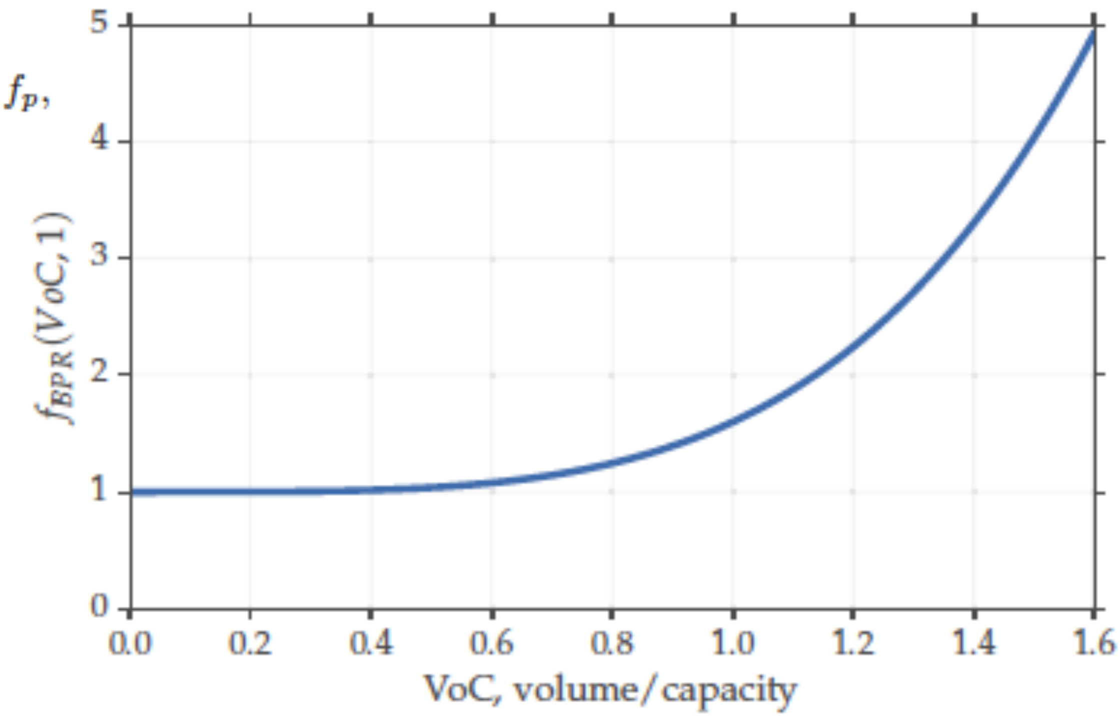
This ensures that all drivers in the system are in fact taking the shortest possible routes,

Assigned volumes are converted to link travel times using a standard BPR function

$$f_p^{rio} = f_p^{boston} = f_p^{sfbay} = 1.3 \text{ and } f_p^{lisbon} = f_p^{porto} = 1.1.$$

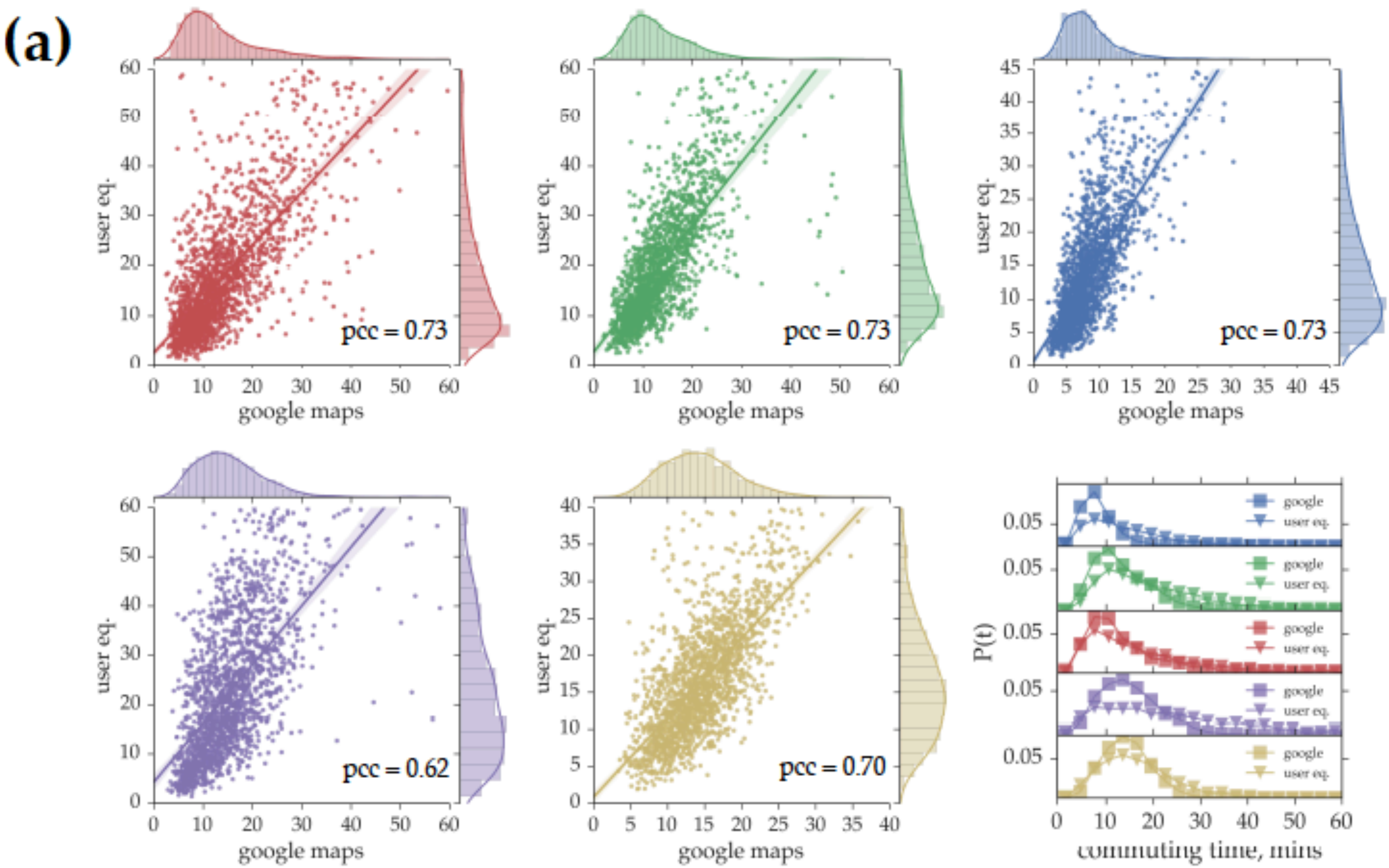
$$f_{BPR}(VoC, f_p) = t_f * \left(1 + \alpha\,(VoC)^{\beta}\right) * f_p,$$

$$\alpha = 0.6 \text{ and } \beta = 4$$

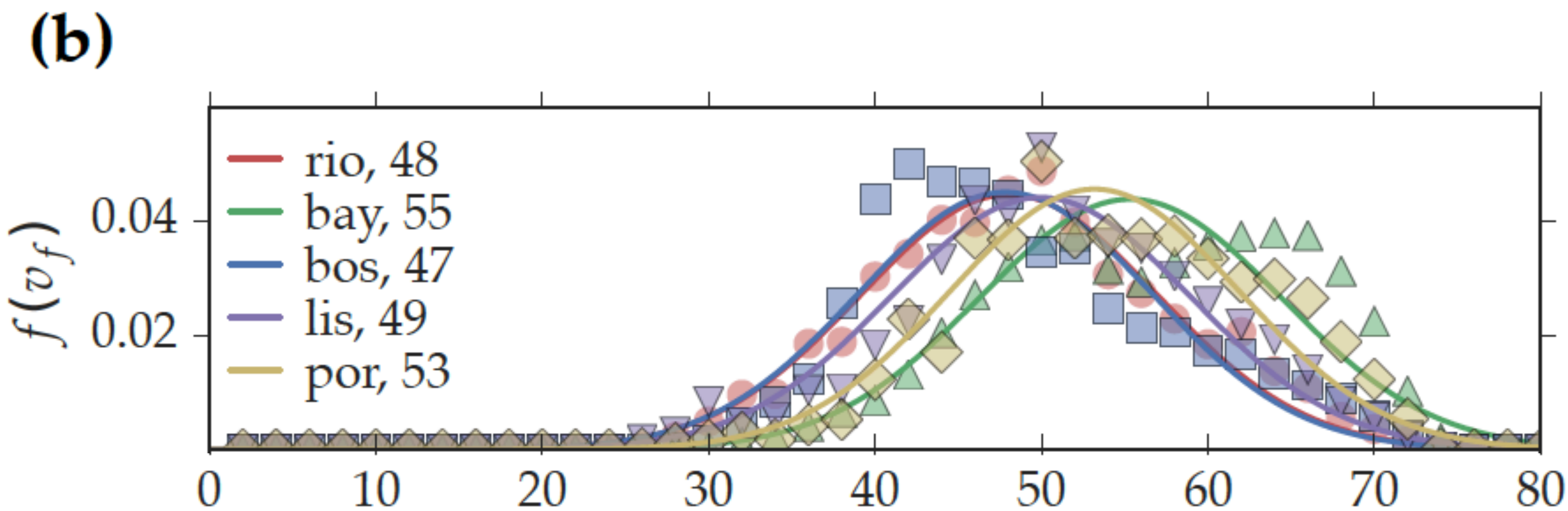


Note: The results of validated travel time at the level of routes act as a validation of the OD flows and show an application of the urban mobility platform to compare cities and the cause of their congestion.

Validation of CDR OD assignment

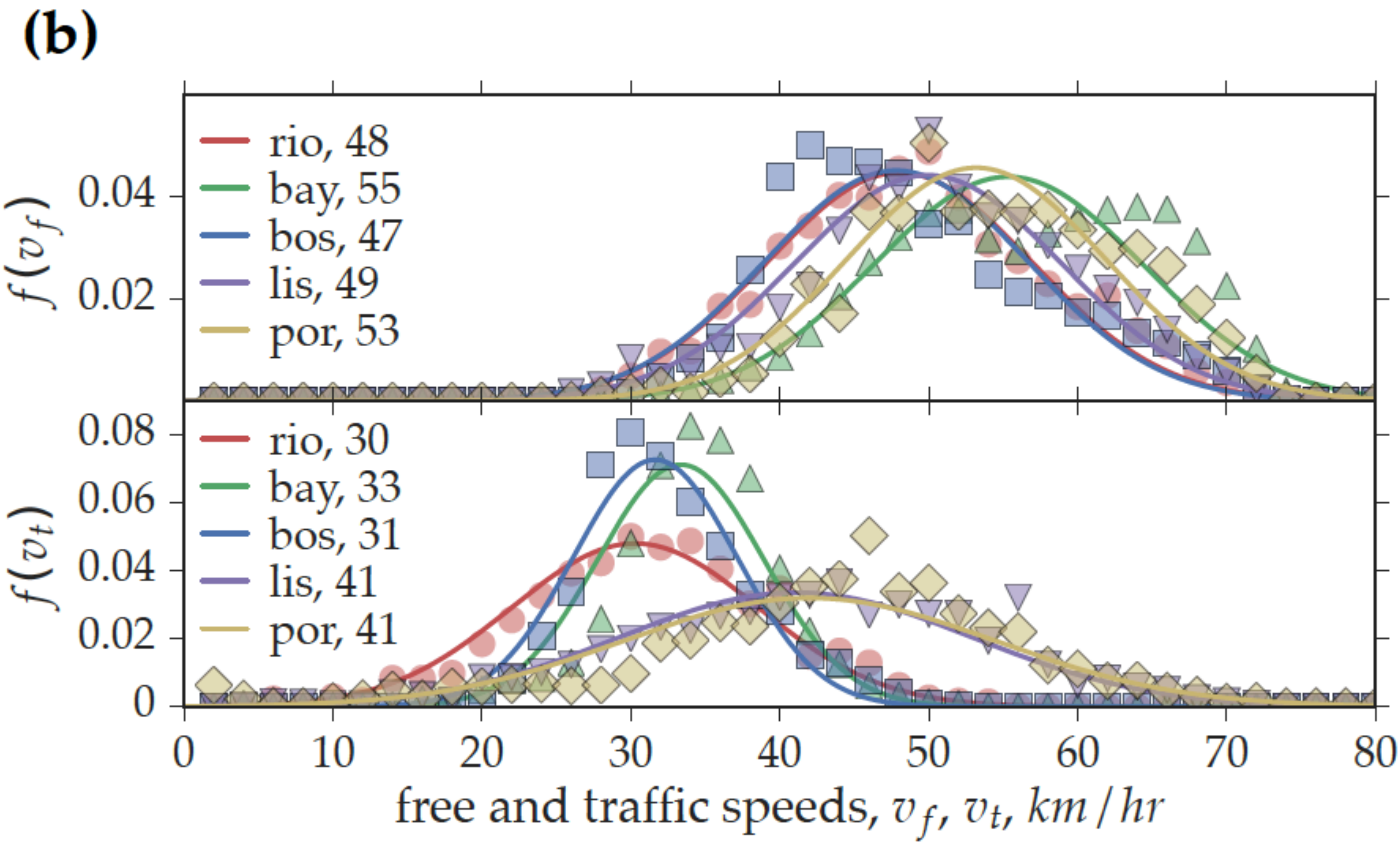


Free Traffic Speed Comparison



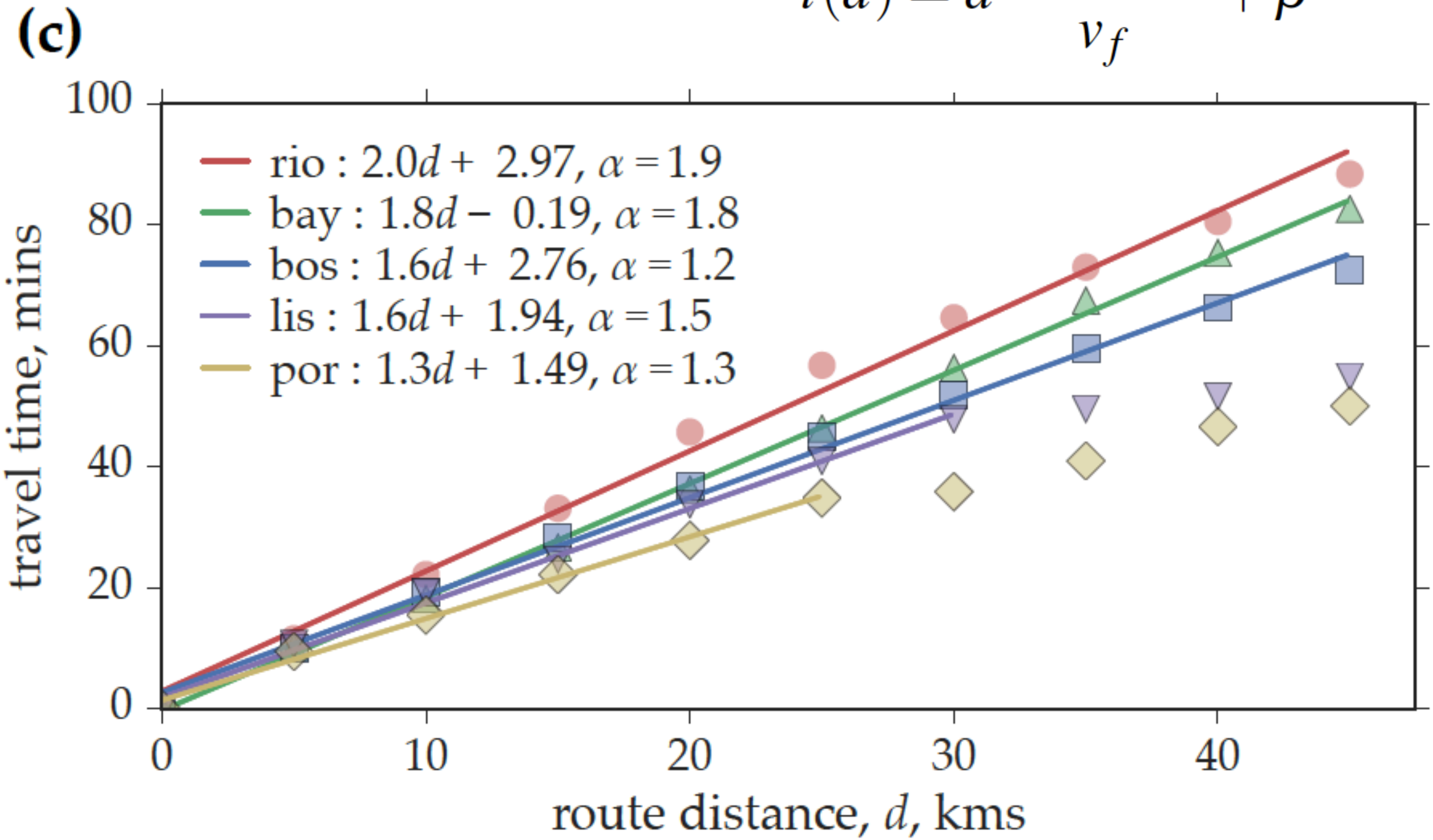
Free traffic speed, v_f [km/h]

Traffic Speed Comparison

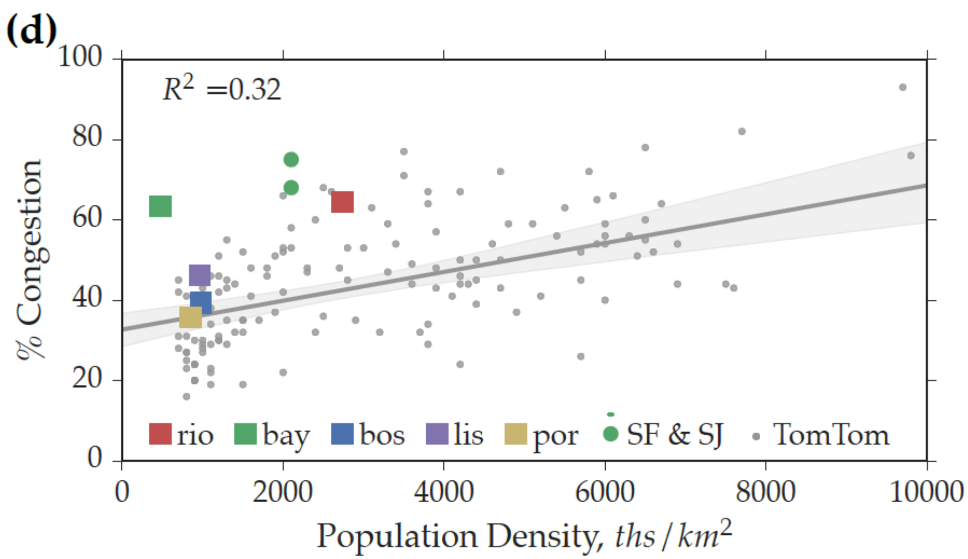


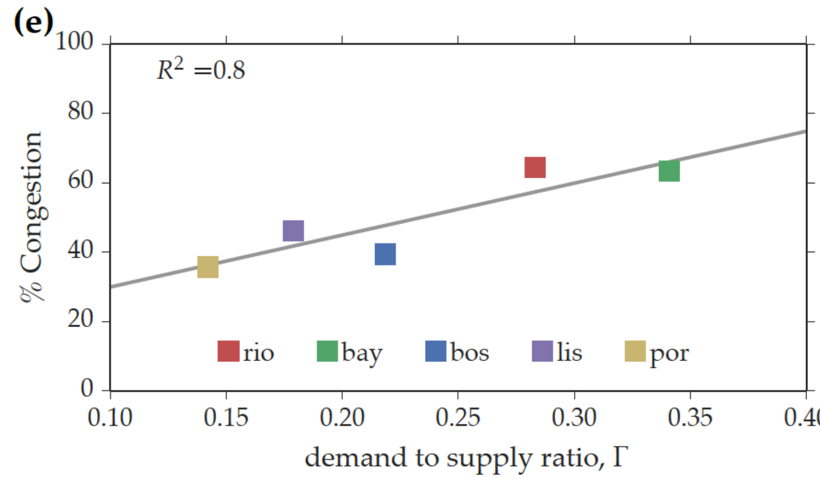
Commuting Time

$$t(d) = d\,\frac{(1+\Gamma)^{\alpha}}{v_f} + \beta$$



% Time Lost vs. Pop. Density





$$\Gamma = \frac{\sum_{e \in E} l_e x_e}{\sum_{x_e > 0, e \in E} l_e C_e}.$$

x_e # Cars in the road link *e*

l_e Road link length in miles

C_e Capacity in the road link [cars/miles^2]

Smart-app (routing)

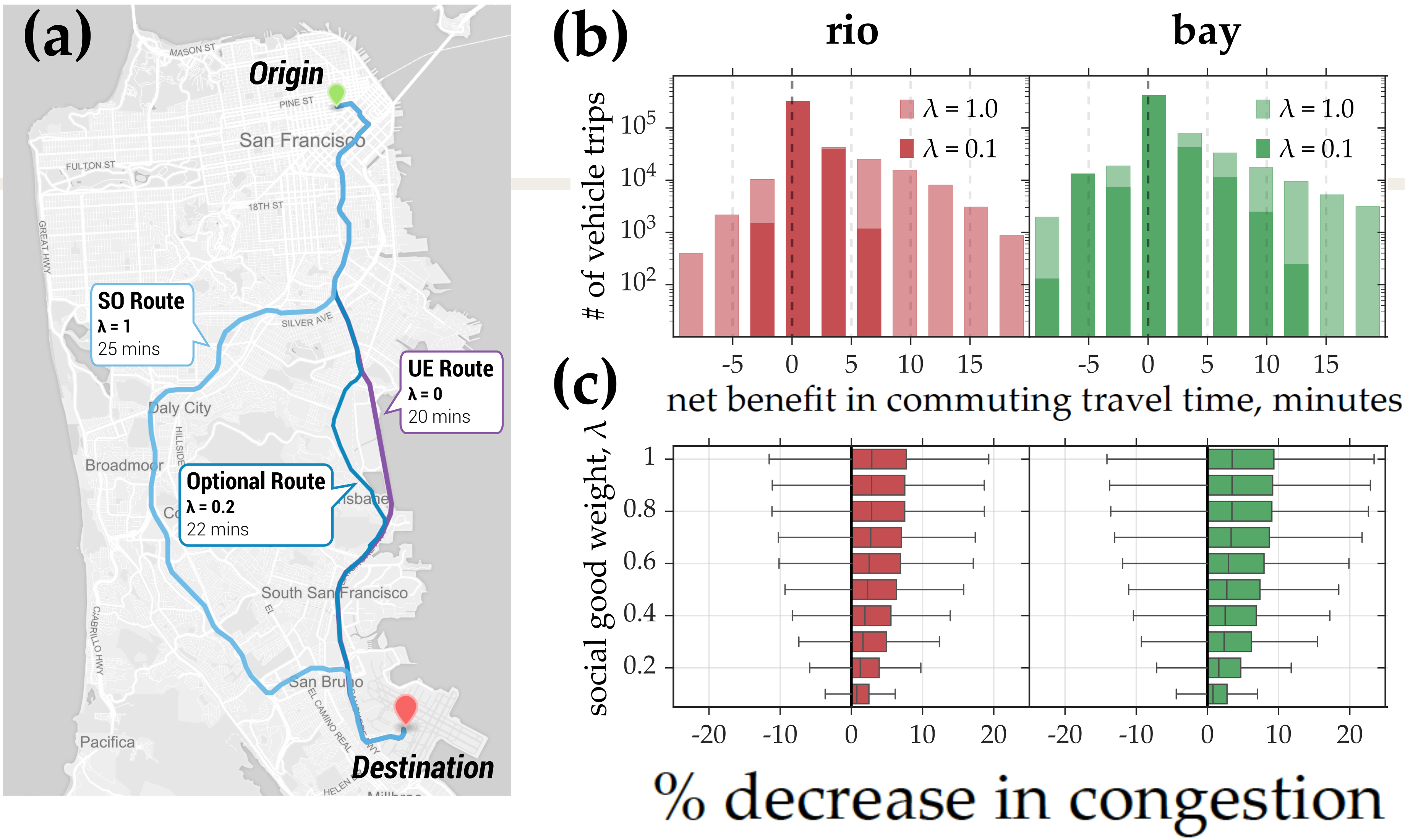
Modifications on the level of altruism:

$$c_e^{\lambda}(x_e) = (1-\lambda)t_e(x_e) + \lambda \frac{d\left[x_e t_e(x_e)\right]}{dx_e}$$

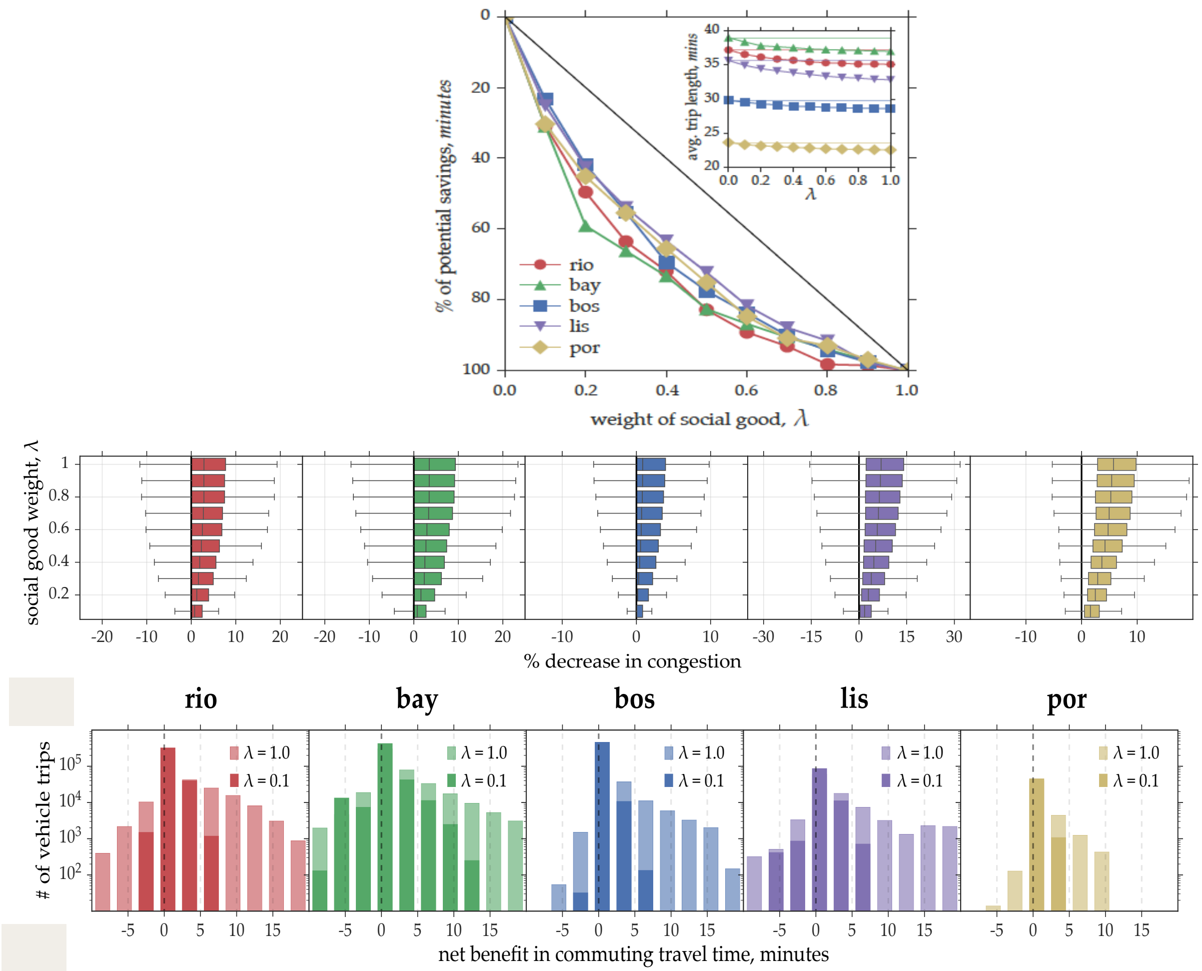
λ=[0..1]

User Equilibrium component

Social Optimun component

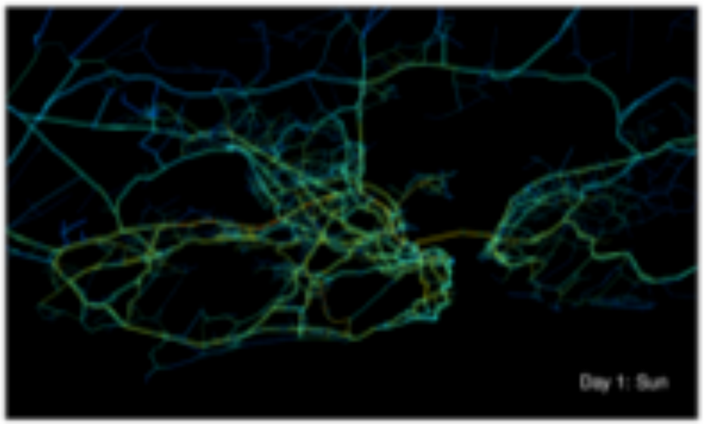

(a)
Origin
San Francisco
SO Route
λ = 1
25 mins
UE Route
λ = 0
20 mins
Optional Route
λ = 0.2
22 mins
Daly City
Broadmoor
South San Francisco
San Bruno
Pacifica
Destination
(b)
rio
bay
# of vehicle trips
λ = 1.0
λ = 0.1
net benefit in commuting travel time, minutes
(c)
social good weight, λ
% decrease in congestion

λ=0.3 is a good compromise

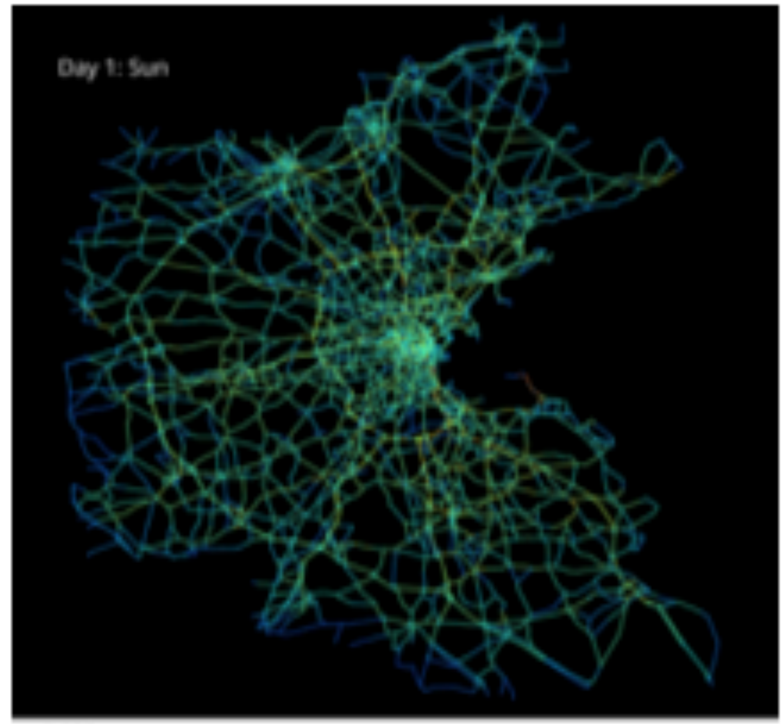


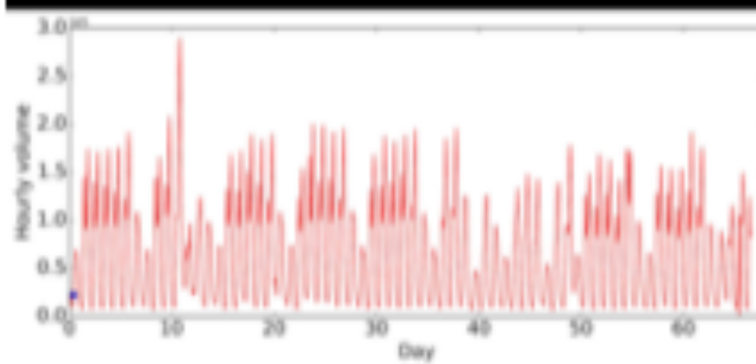
Next steps: Travel Time Reliability

CDR Rio de Janeiro + Waze Data



CDR Boston + TomTom Speed reads

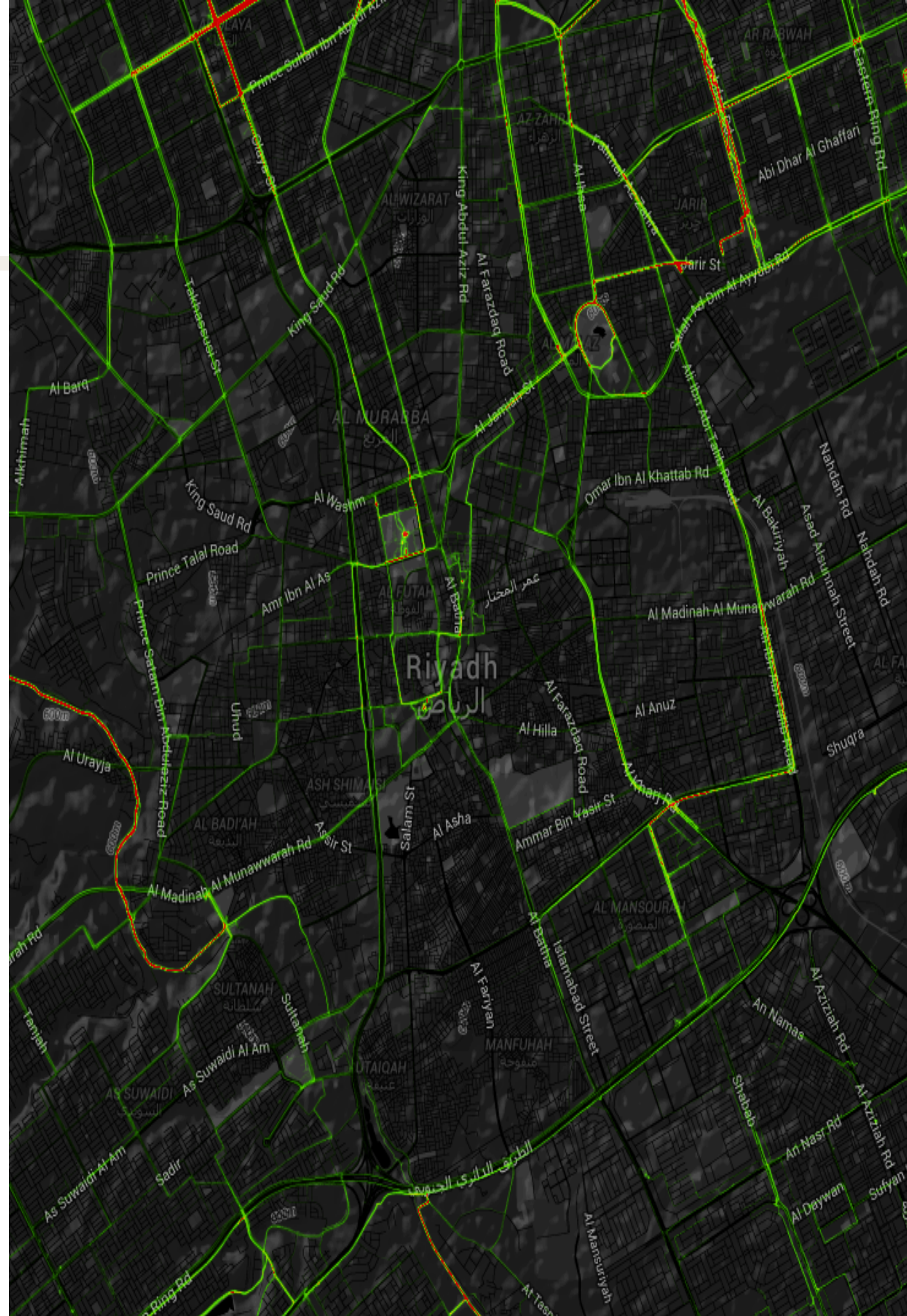




Recommendations for the new Riyadh Metro System (METRO)

- Methodology for optimizing catchment zones based on potential demand of the new metro system and to support the design of the bus system.

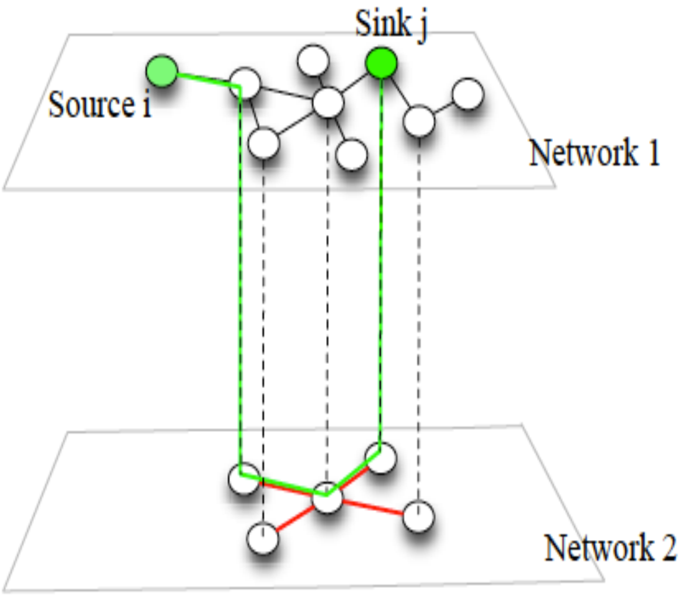
- Scenario analysis of various travel time outputs from diverse coupling mechanisms of the road vehicle trips (cars and buses) with the metro system.



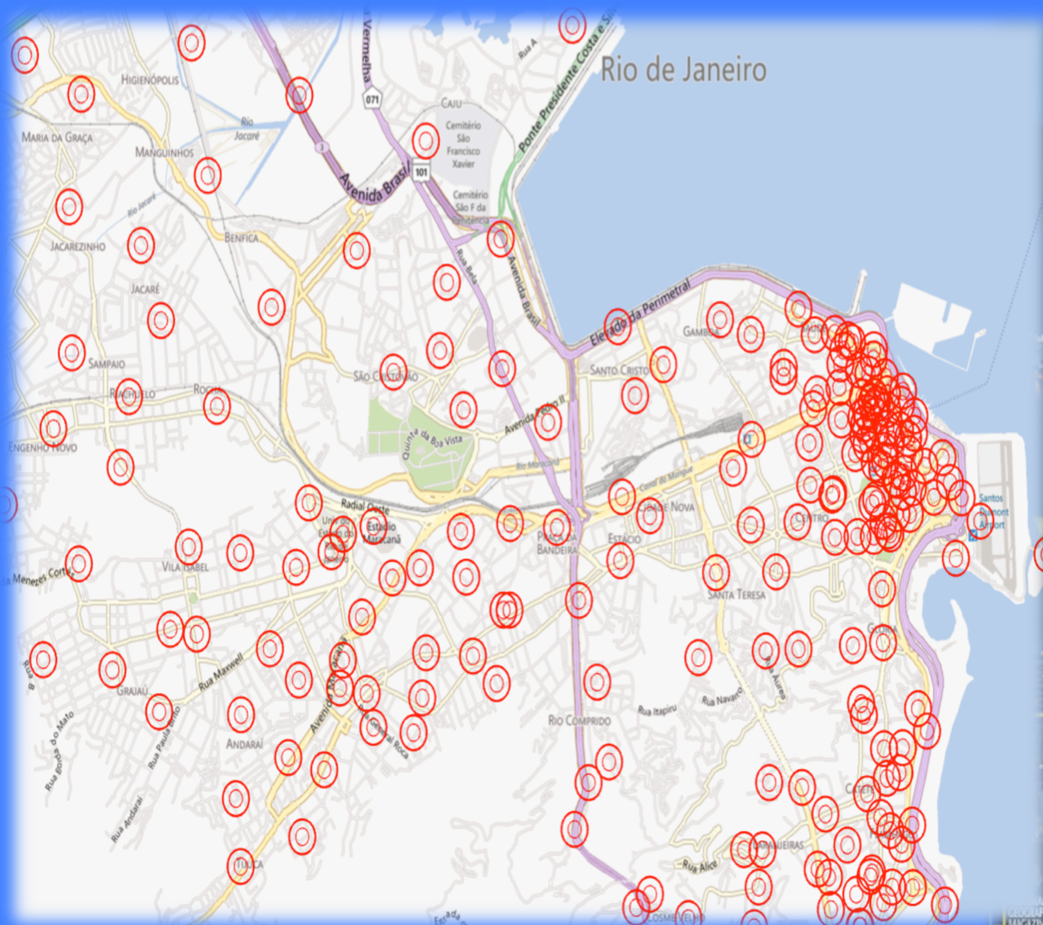
 MIT School of Engineering

APPROACH

- **Coupled Networks approach:** A small network 2 (metro) is coupled with the larger network 1 (road network). We will explore various coupling mechanisms that result in the best performance in terms of travel times for existing conditions of OD demand and travel times.



Current Application: Help Distribute Demand in the Rio Olympic Games



Overview of research

- Working at the intersection statistical physics and machine learning methods to generate urban transportation models. That intersection enables the generation of knowledge from data that cannot be extracted from one discipline alone.
- Current models of travel demand are based in Census and Activity Surveys which get **outdated** and are **expensive** and **limited in size**. While massive digital traces (phone usage, on-line activities, GPS records) are passively generated geo-located signals of human activities. But these signals are **sparse** and incomplete.
- We can convert raw data into **surveyless models** of **trip diaries, urban traffic** and **social behavior** at urban scale. This is key for urban and transportation planning.

Implementation Approach

Stage 1: Strategy

- Publish research approach
  - Incentives
  - Success
- Form partnerships
- Secure funding

Stage 2: Pilot

- Solicit proposals
- Select host city
- Conduct Pilot

Stage 3: Expand

- Advertise findings
- Implement Smart Commute in cities nationwide