Big Data to Tackle Urban Mobility Challenges

Marta C. Gonzalez

Associate Professor of Civil and Environmental Engineering, Center for Advanced Urbanism

HUMNET

Civil and Environmental Engineering

Massachusetts Institute of Technology

The Urban Era

- Congestion
- Pollution, waste and water problems
- Unaffordable hosing

Occupy 3% of the earth surface

Over 50% of population

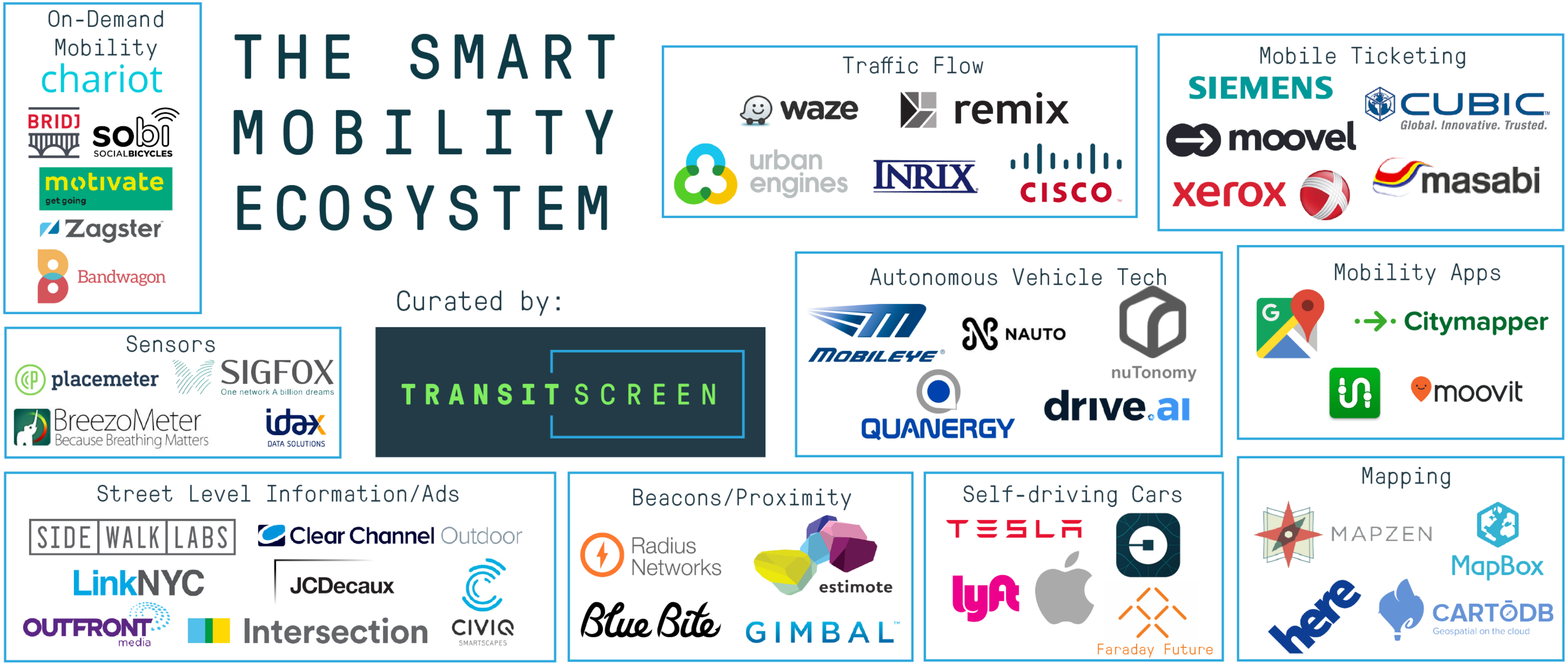
Generate 80% of global emissions and consume 75% of world resources

http://photomichaelwolf.com/#architecture-of-dens itiy/1

Internet on the move: Connected locations and transportation modes

source:fleet.be

New Business Opportunities



1- Big data to improve Urban Mobility?

2- Big data to improve Urban Policy?

3- Do this require new workforce and practices for planning cities?

http://photomichaelwolf.com/#architecture-of-densitiy/1

How do we govern cities related to Mobility?


Development planning

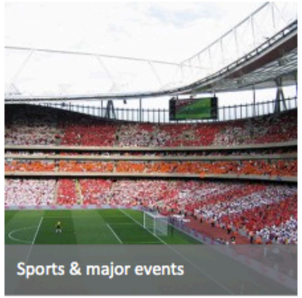

Economics


Infrastructure


Local & regional transport


Rail


Research & Innovation


Sports & major events


Sustainable transport


Urban transit

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

Urban Models today still use Travel Surveys

$200 per usable Survey

1 sample day,

$2.5x10^4$ households out of $2.6x10^6$

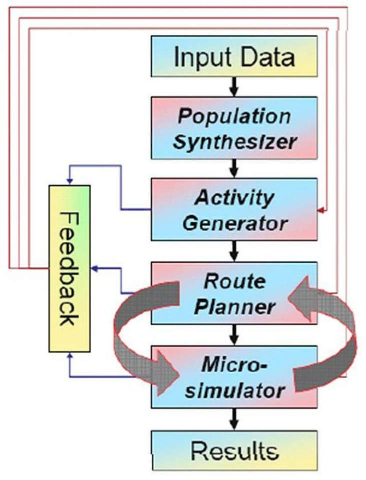
58% response rate.

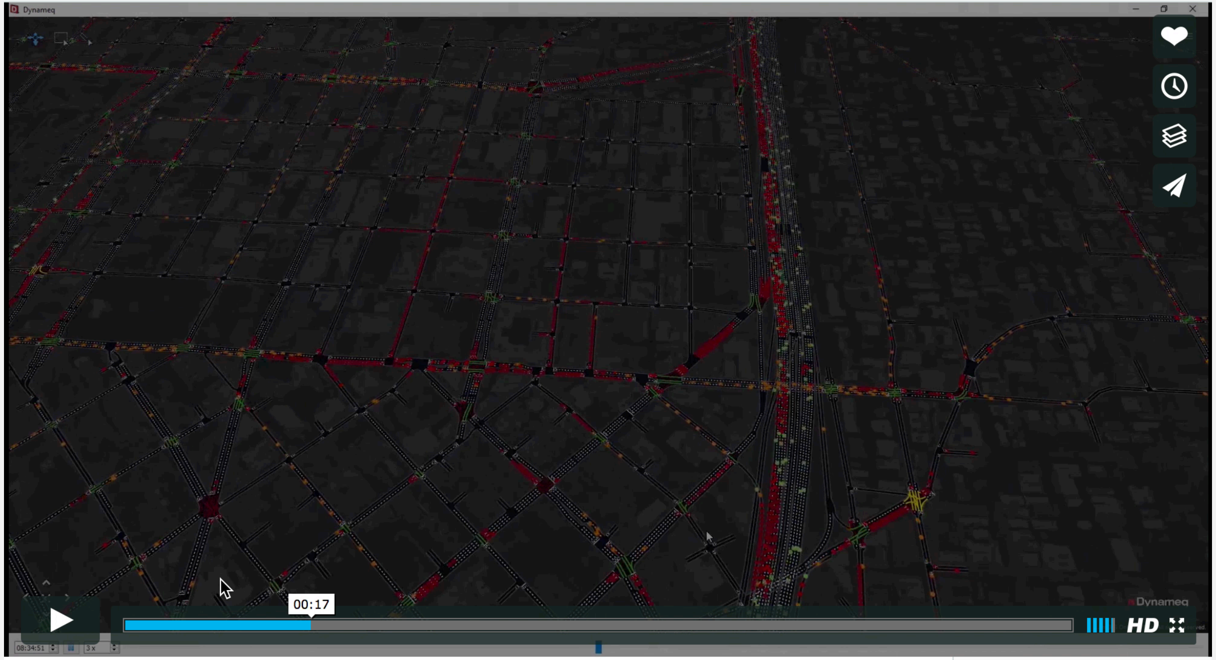
(3.7 calls and 17 minutes per survey)

MASSACHUSETTS TRAVEL SURVEY

2011-2012

TRANSIMS (TRansportation ANalysis SIMulation System) is an integrated set of tools developed to conduct regional transportation system analyses.





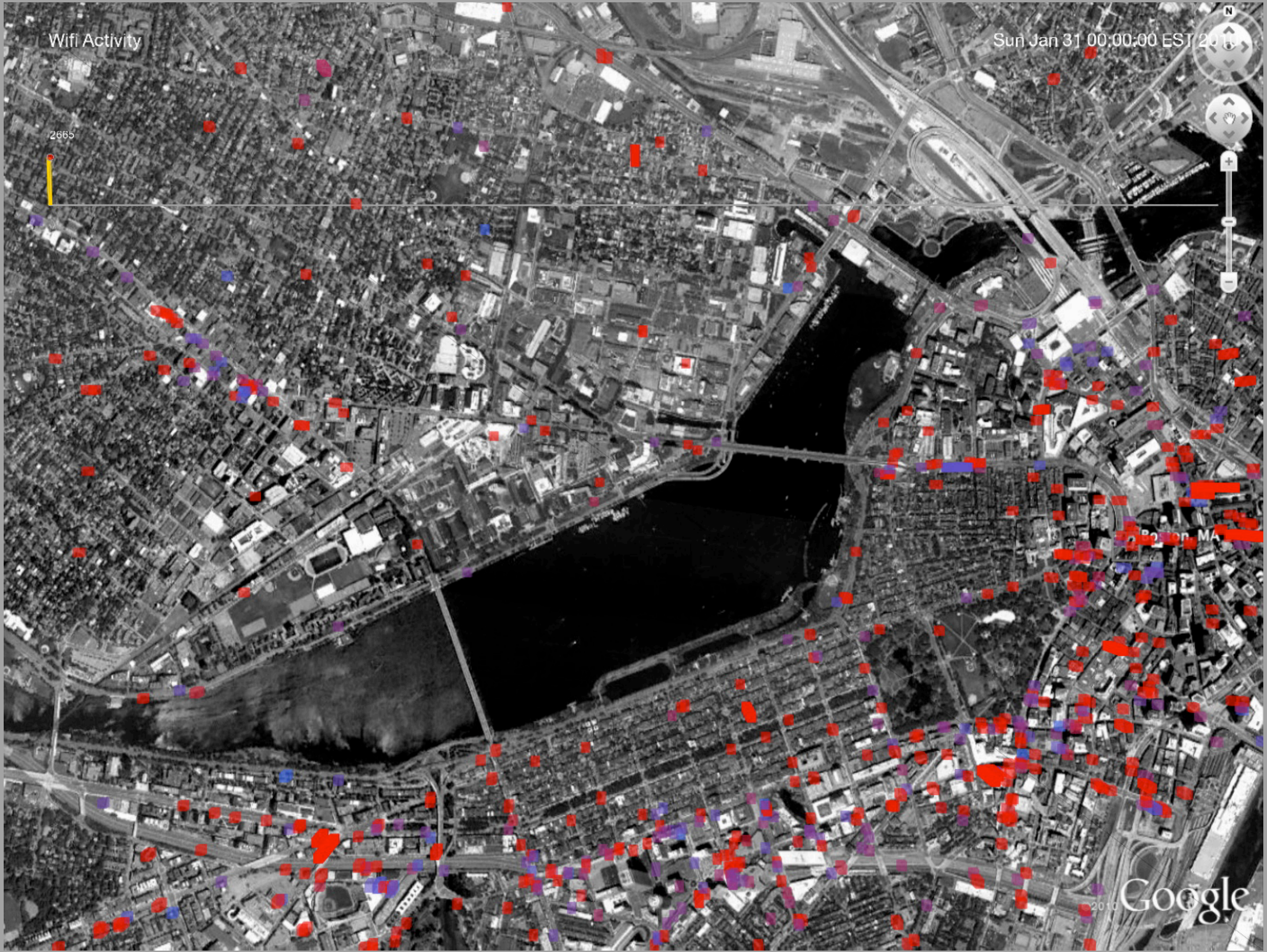
Sample Paper:
[Behavioral calibration and analysis of a large-scale travel microsimulation] G Flötteröd, Y Chen, K Nagel
Networks and Spatial Economics 12 (4), 481-502

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

Big Data From Cheap Phones

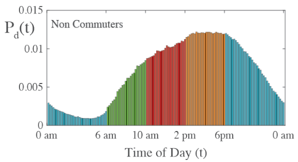
Source: Teralytics


Wifi Activity
Sun Jan 31 00:00:00 EST
2665
Google

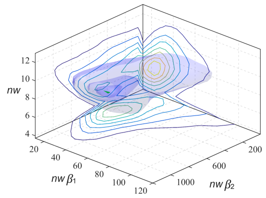
The Model

Features Extracted from data of Active Users

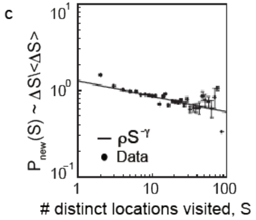
Global Trip prob.



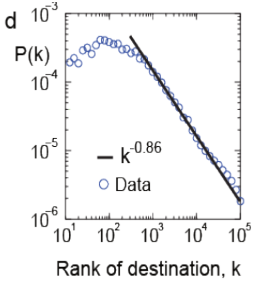
Individual Mobility Rates



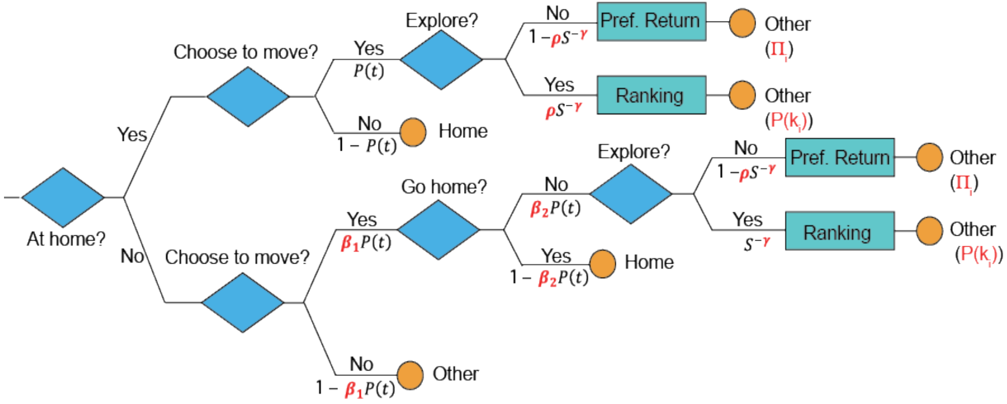
Preferential Return



Ranking of Explorations



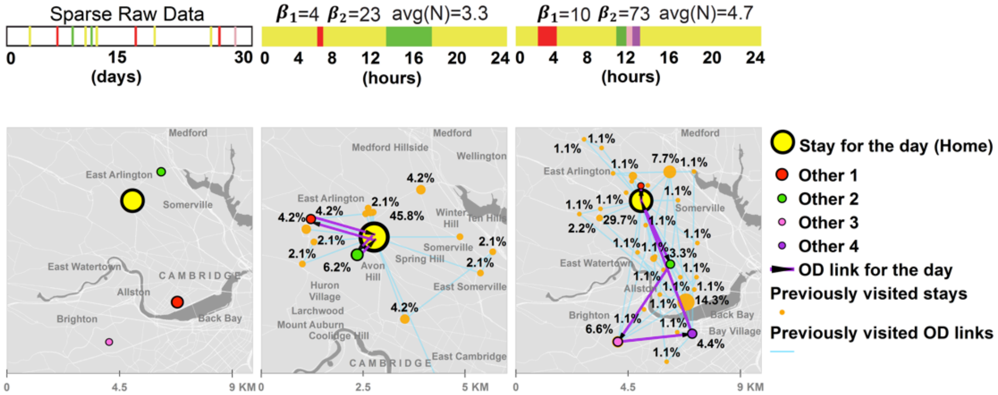
Flowchart of the Model

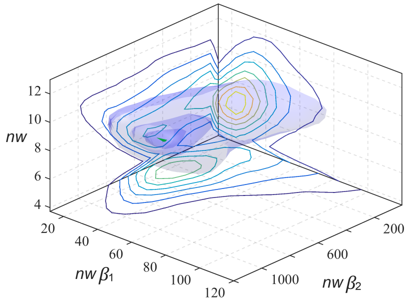


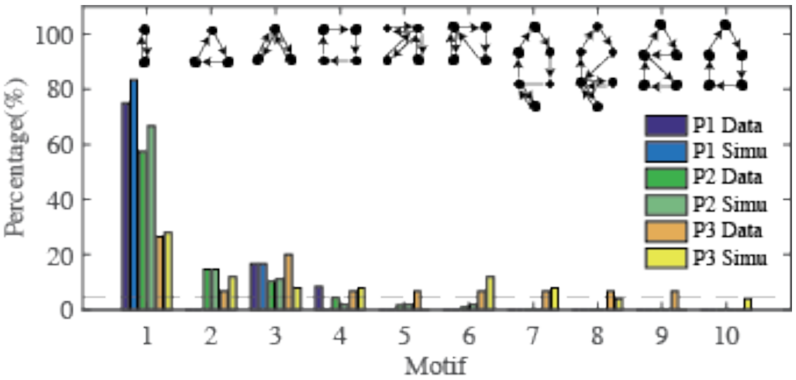
Marta C. Gonzalez

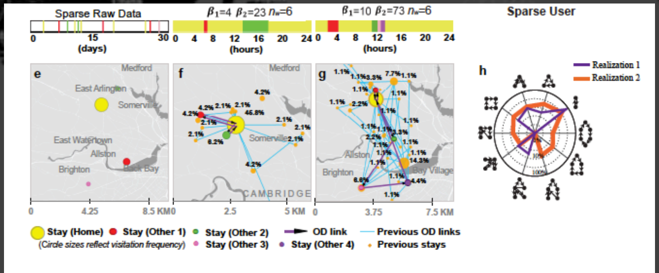
Models Results

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



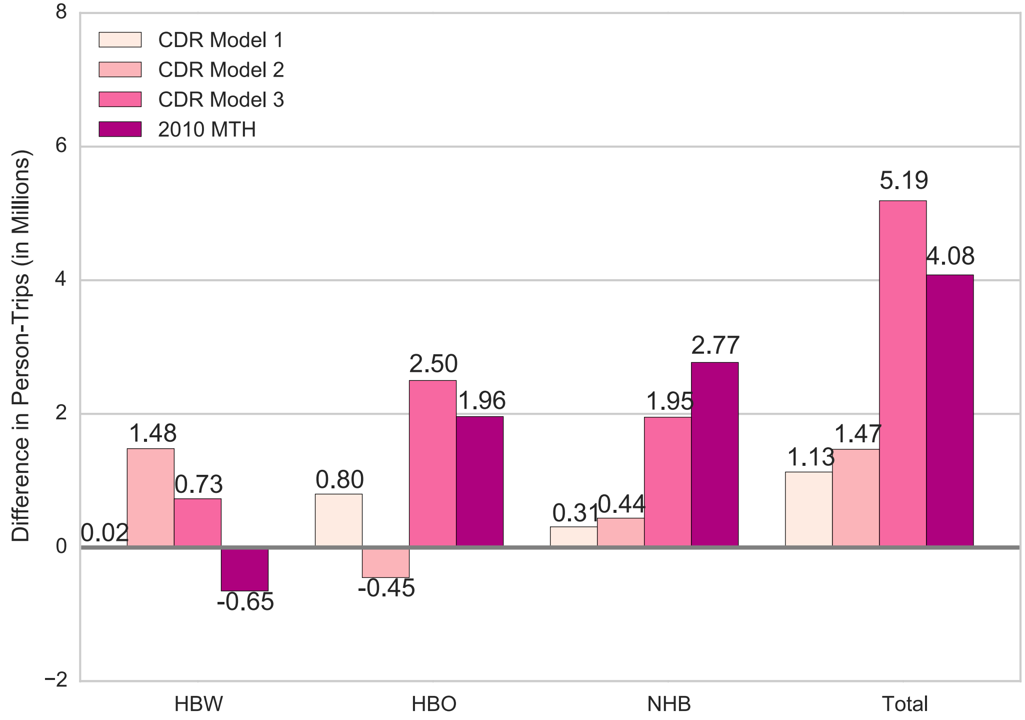






TimeGeo: a spatiotemporal framework for modeling urban mobility without surveys (Shan Jiang, Yingxiang Yang, Daniele Veneziano, Shounak Athavale, Marta C. Gonzalez), *PNAS* (2016)

Comparable results with Metropolitan Planning Office Models

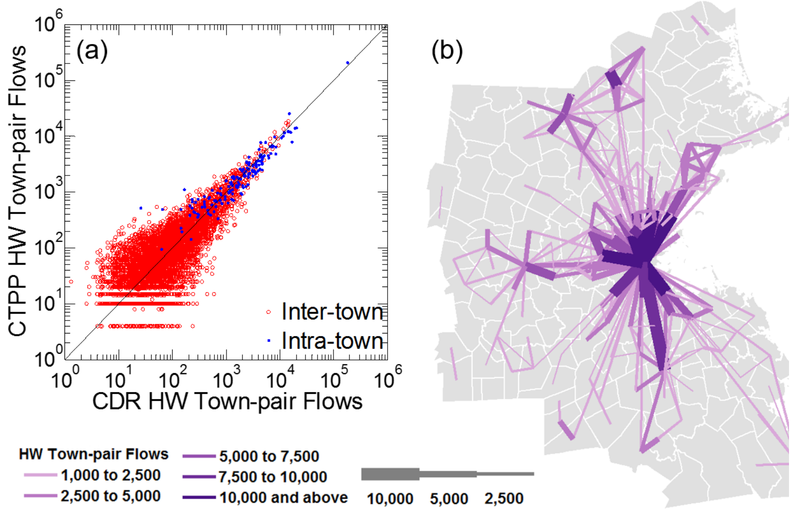


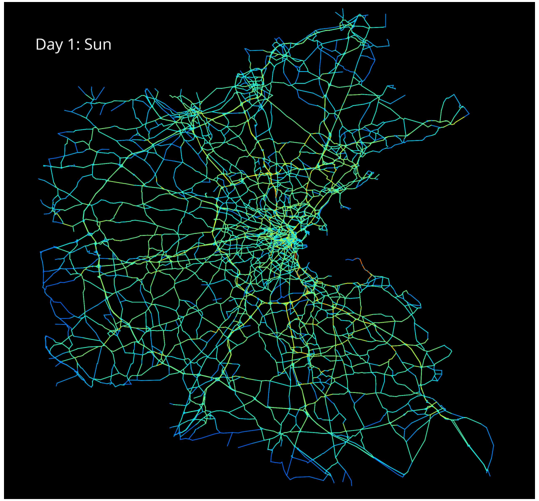
HBW=Home Based Work HBO=Home Based Other NHB=Non Home Based

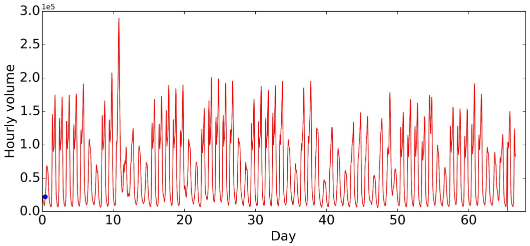
Trans. Research Records, "Practice Ready" (2014)

Trans. Res. C: Emergent Technologies (2015)

Validated Travel Demand







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

Demand Management for Large Events

Yanyan Xu PhD.
Postdoctoral Associate
Department of Civil and Environmental Engineering, MIT
Email: yanyanxu@mit.edu

Rio Olympic Games

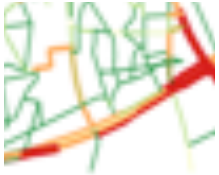
- Rio population: 6.4 millions
- International Olympic Committee (IOC) predicts 480,000 tourists in Rio for 2016 Olympics, that's about 7.5% of Rio population.
- How to evaluate the impact of Olympics to the travel of local population?
- How to manage the demand?

Data from Companies

- Mobile Phone Data 
- Waze (extend the seed OD to weekdays)

On-line

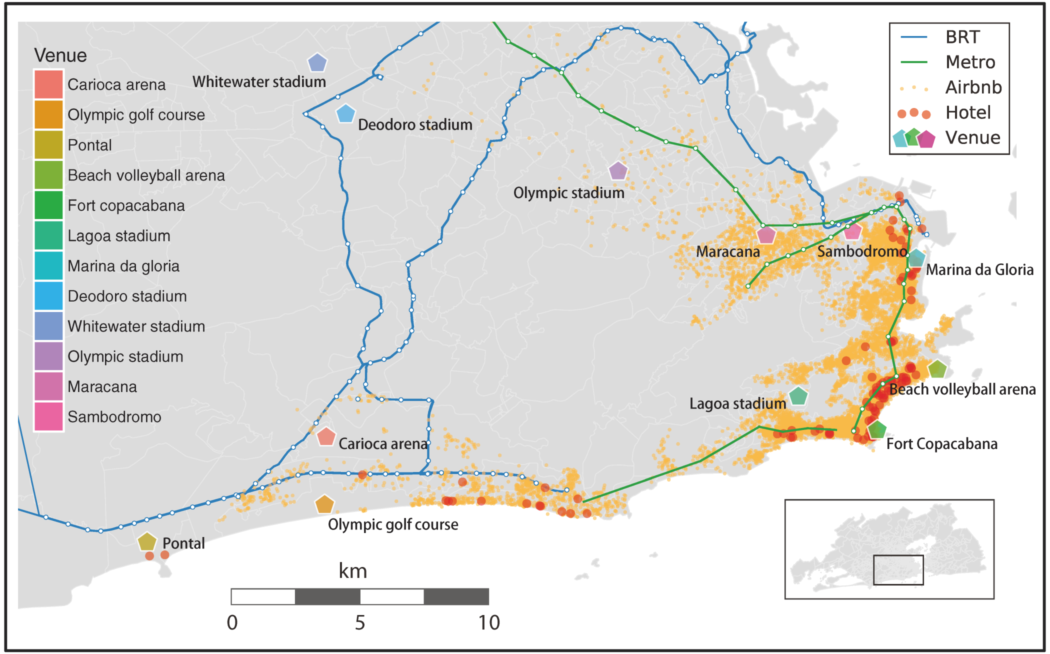
- Airbnb Supply
- GIS (OSM road network of Rio)

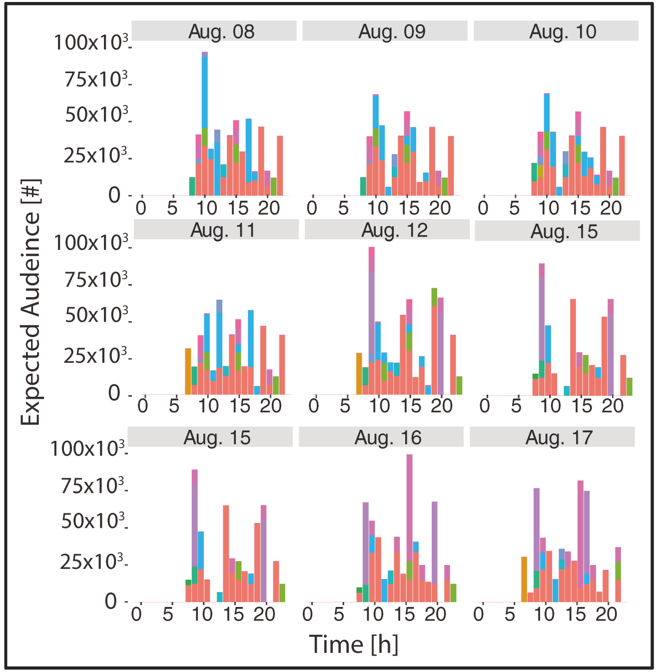


Data from Government

- Hotel, Venues and Schedules
- Camera Data 

Venues, Airbnb, hotels, BRT & Metro

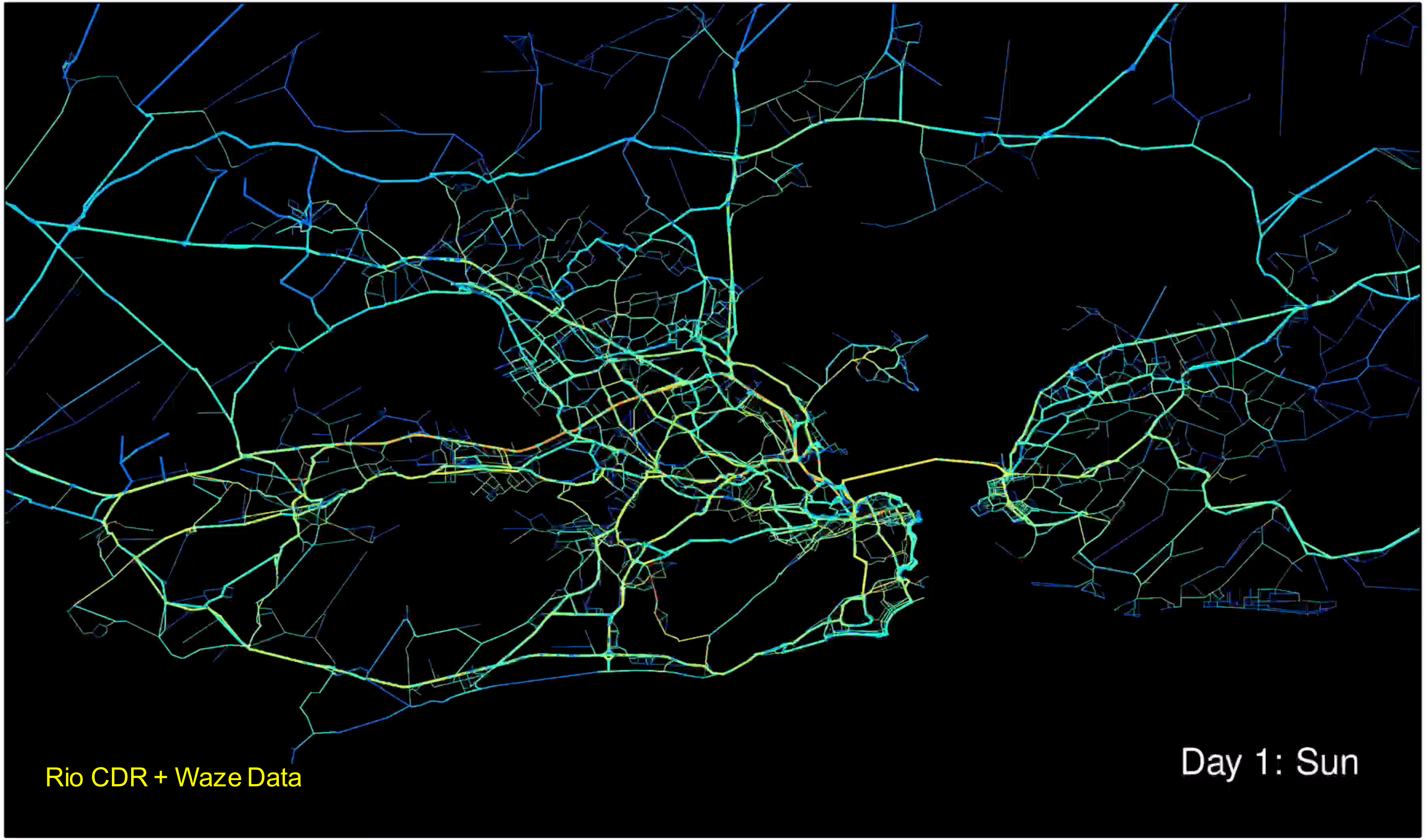




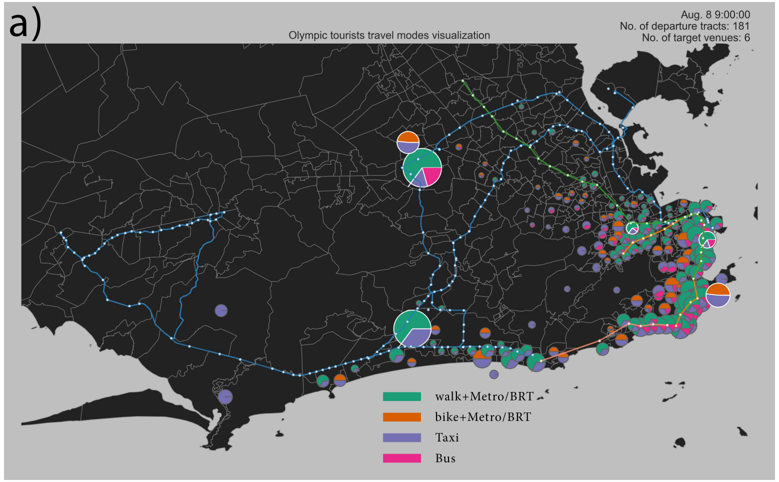
(a) Data Integrated

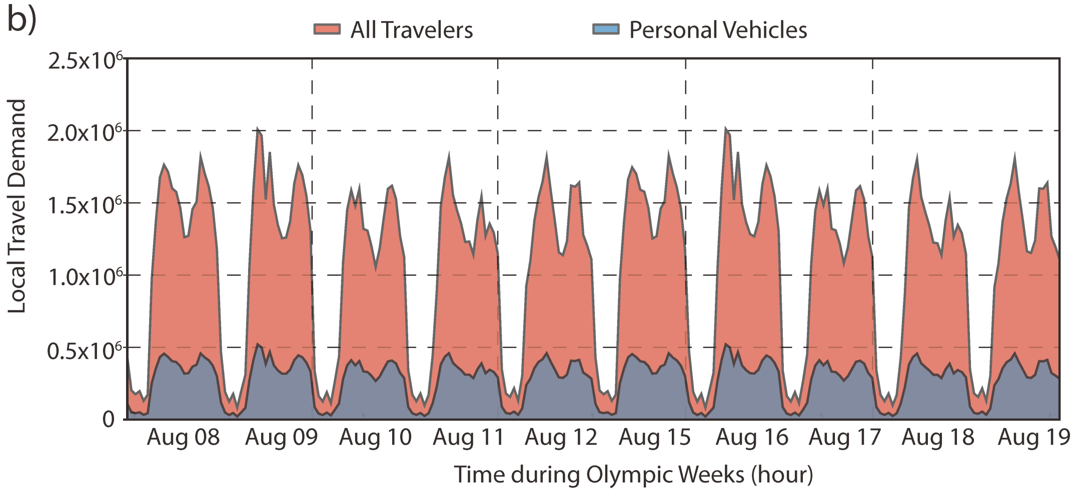
(b) Number of audiences arrive venues and when? (used data: Olympics schedule, capacity of venues)

Traffic Model



Travel demand prediction during Olympics





(a) Tourist travel mode split

(b) Total travel demand during Olympics: 1.8 million person / 0.5 million vehicles in peak hour.

Smart-app (routing)

Modifications on the level of altruism:

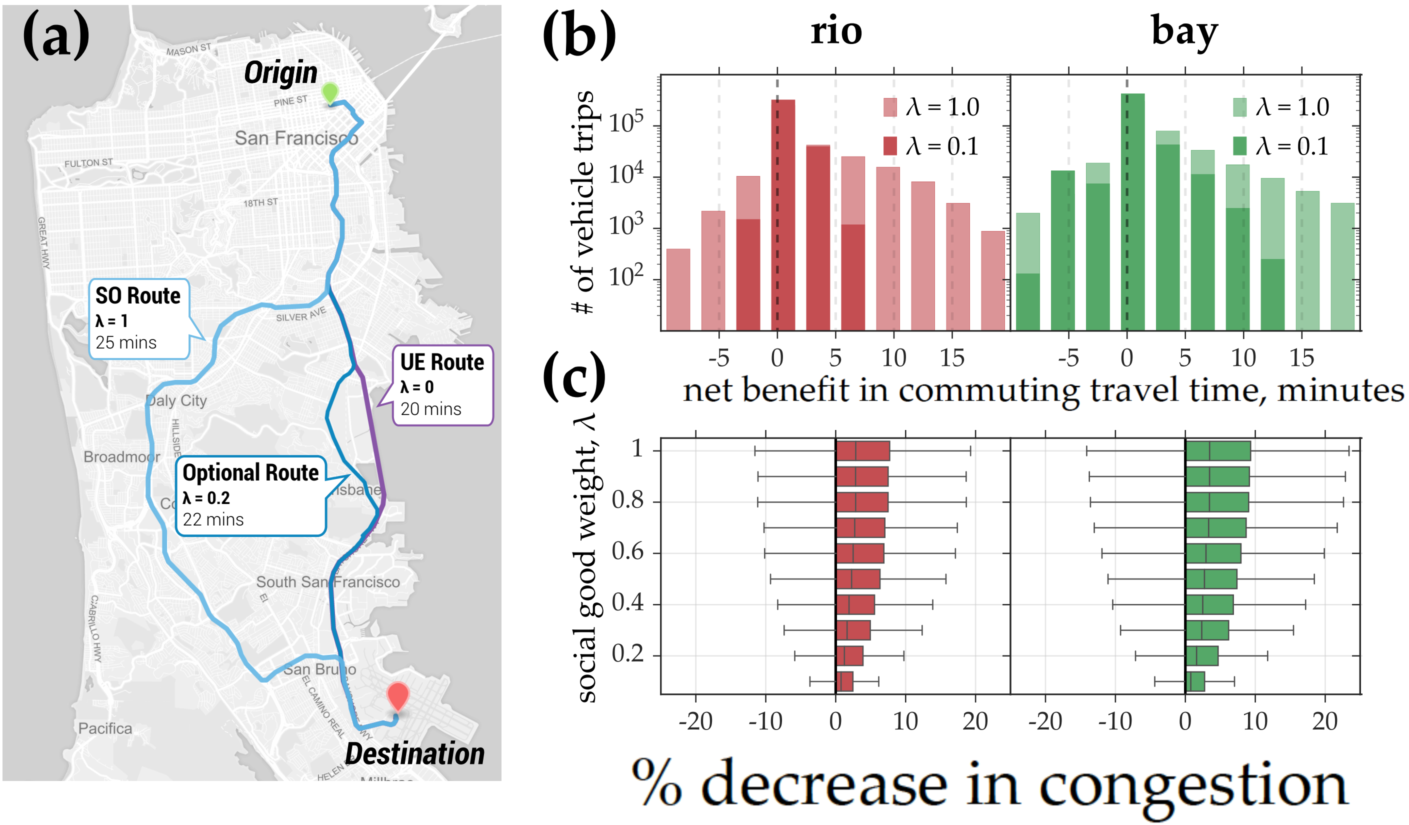
$$c_e^{\lambda}(x_e) = (1-\lambda)t_e(x_e) + \lambda \frac{d\left[x_e t_e(x_e)\right]}{dx_e}$$

$\lambda=[0..1]$

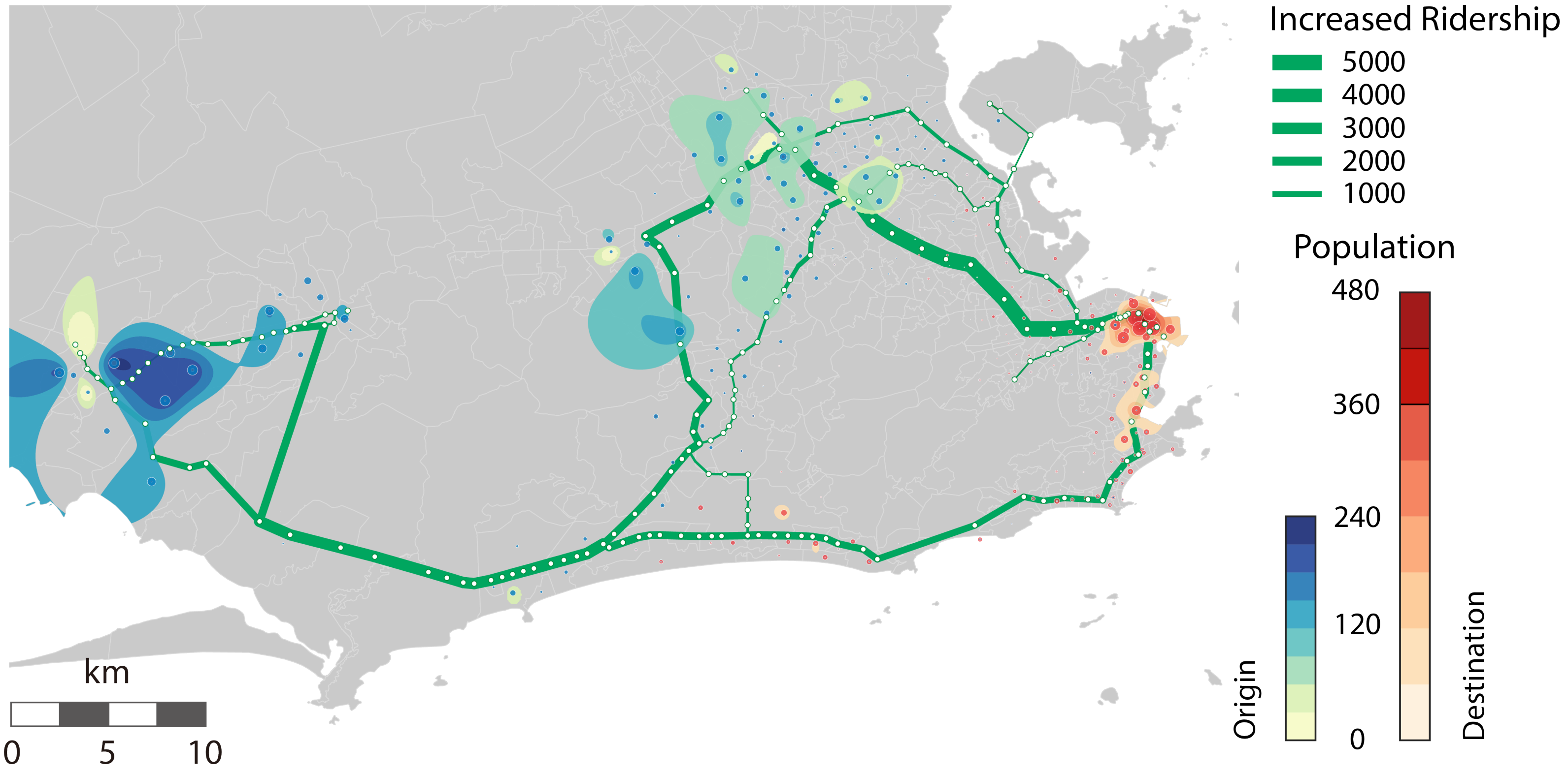
User Equilibrium component

Social Optimun component

Smart-app (routing)

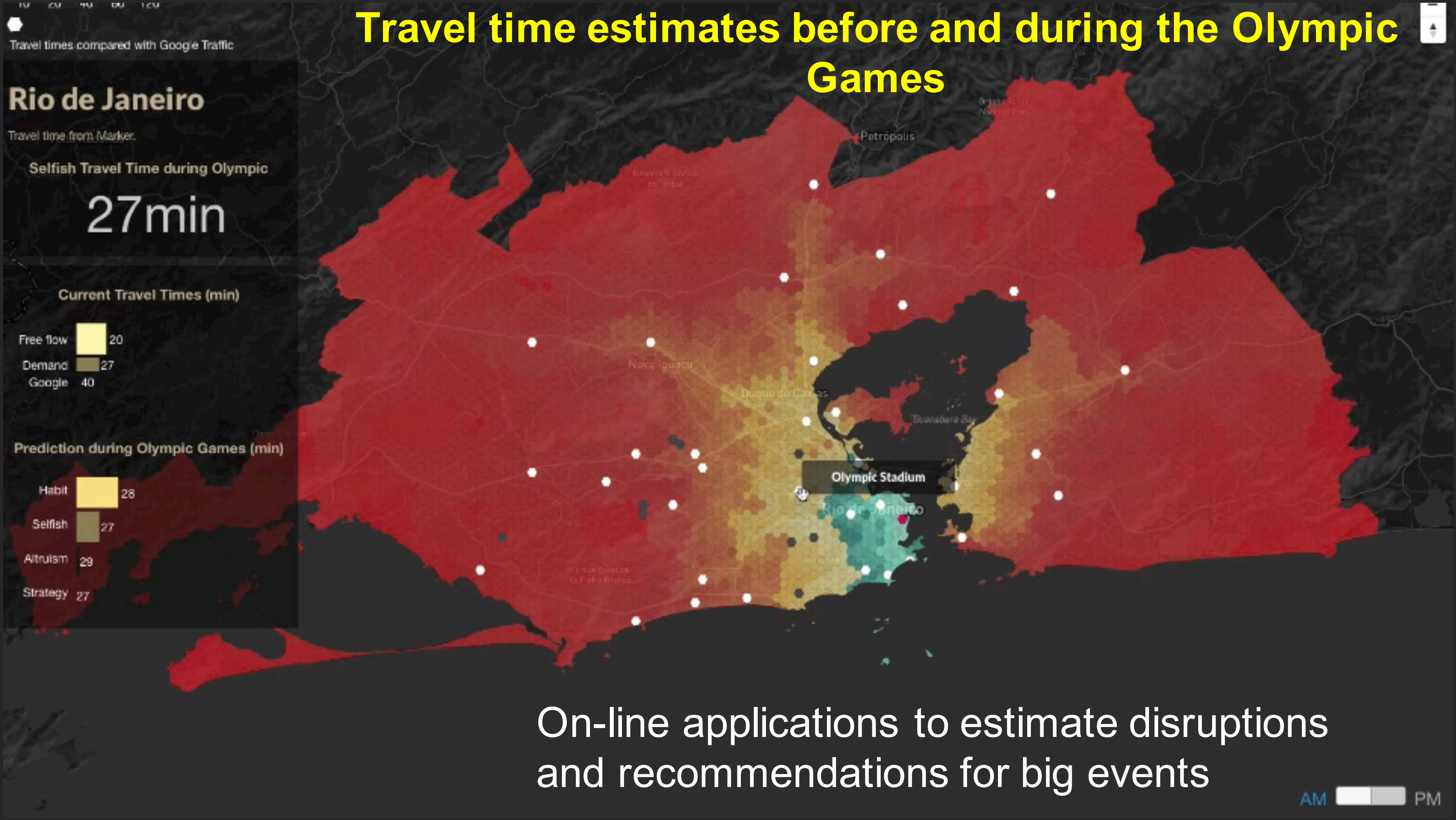


Recommendations of Car reduction per Origin and Destination



Vehicle demand decrease: ~1.3%
Total travel time decrease: ~10.5%

Travel time estimates before and during the Olympic Games



On-line applications to estimate disruptions and recommendations for big events

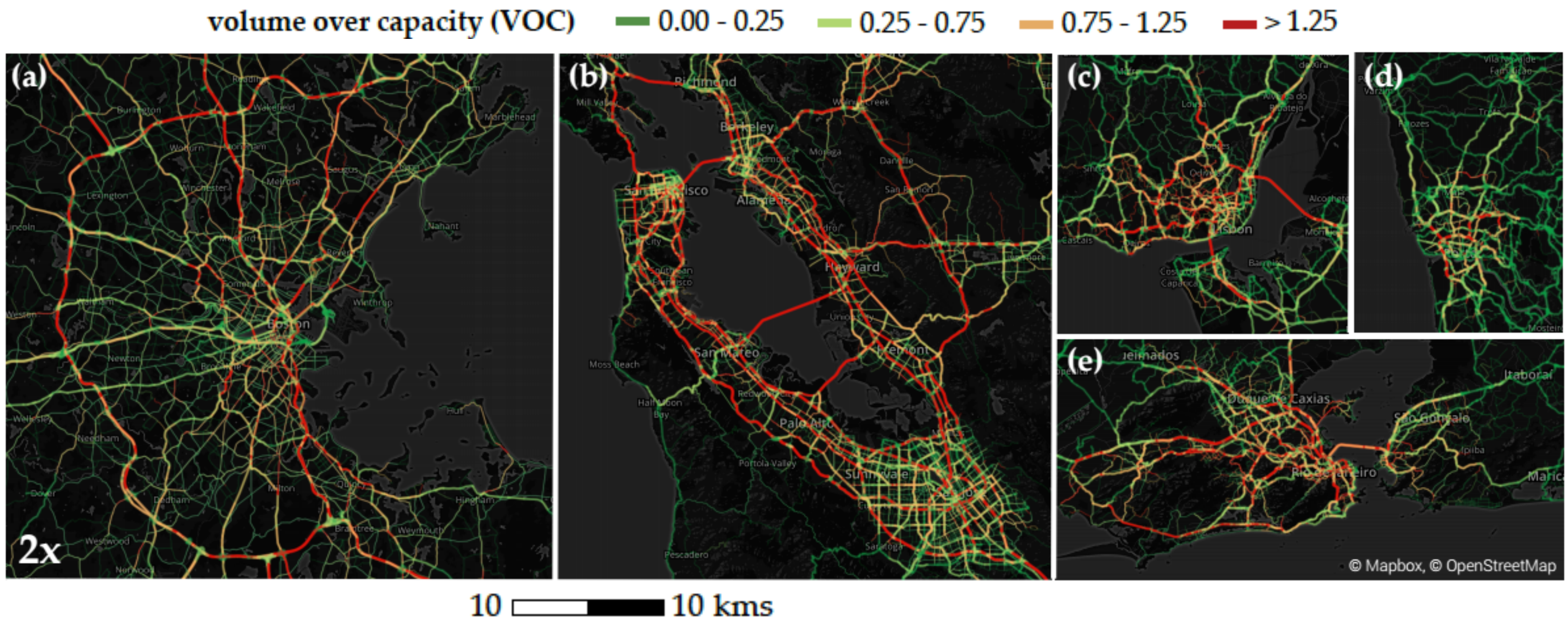
Understanding and Mitigating Congestion in Urban Areas

Luis Eduardo Olmos Sánchez
PhD Candidate
Physics Department
National University of Colombia

movistar telecom
Telefónica

alvaro.ramirez@telefonica.com

gonzalo.durban@telefonica.com



(a) Boston
(b) San Francisco Bay Area

(c) Lisbon
(d) Porto

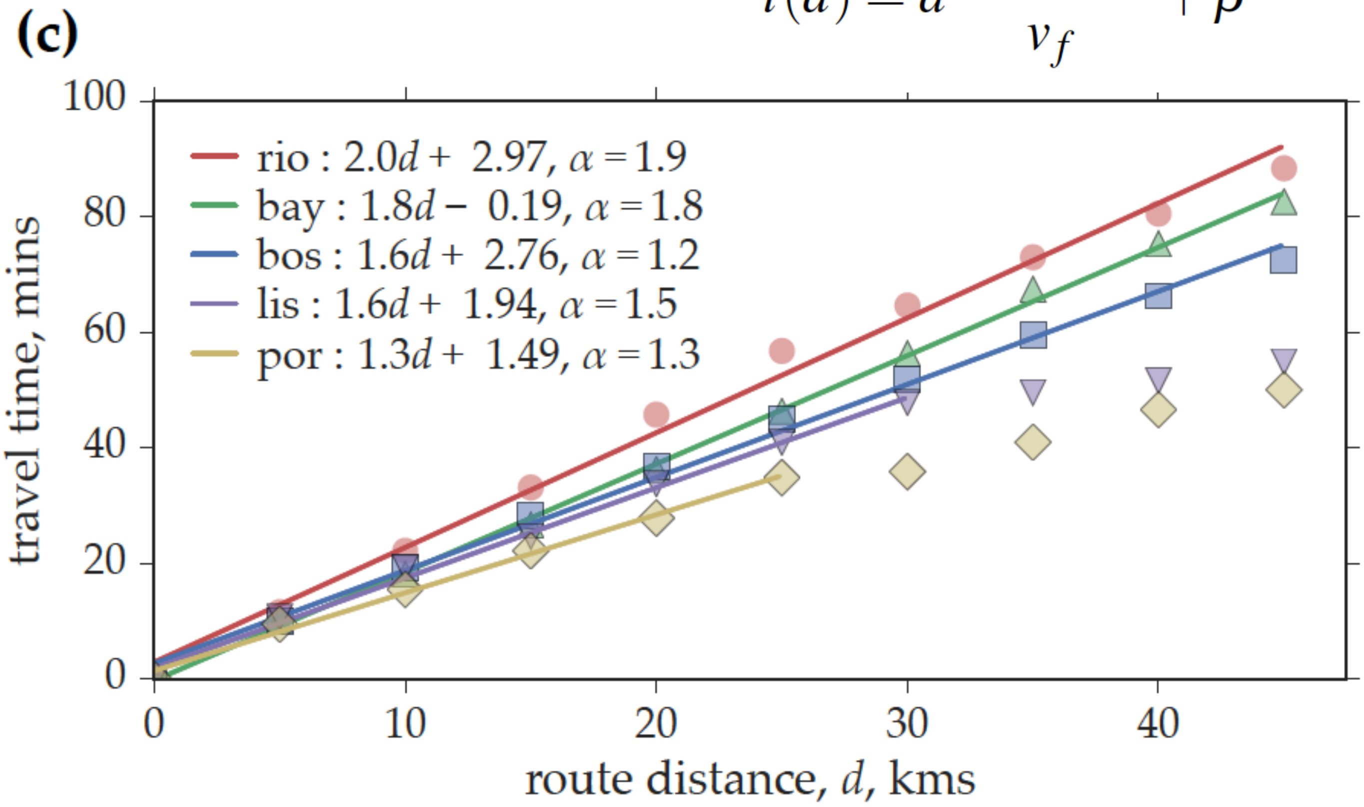
(e) Rio de Janeiro

Using 3 months of phone data
And Census Information on Population
and numbers of cars and their usage

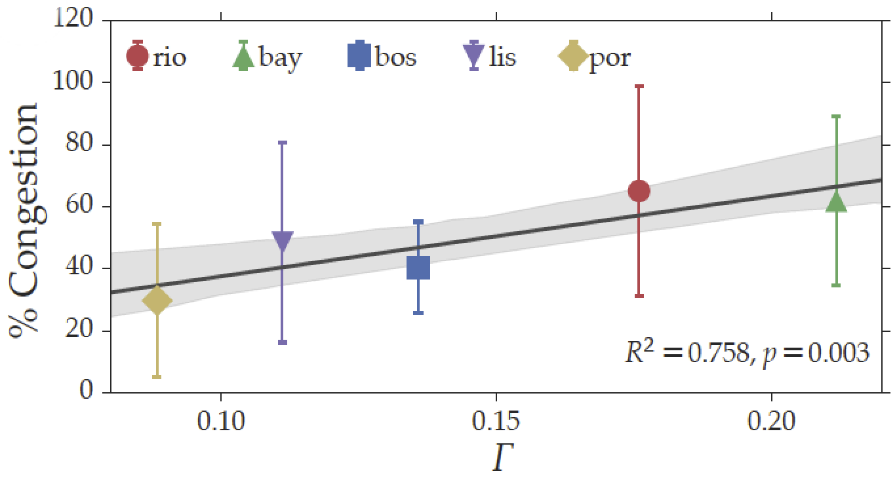
Open Street Map data for the Streets when
not better data is available

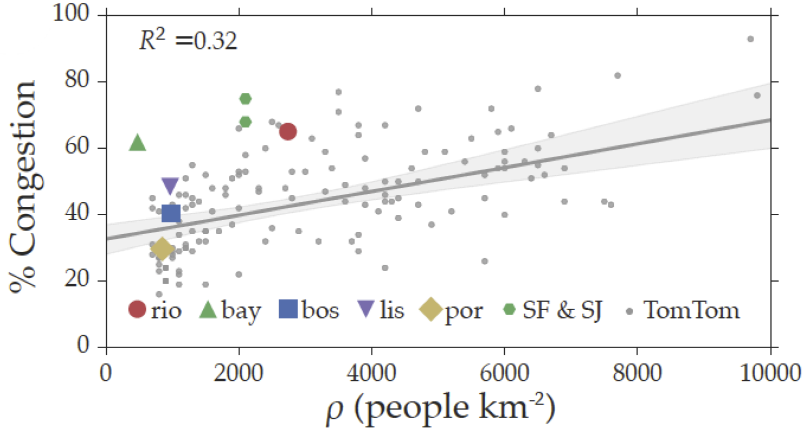
Commuting Time

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



Better information based on our pipeline





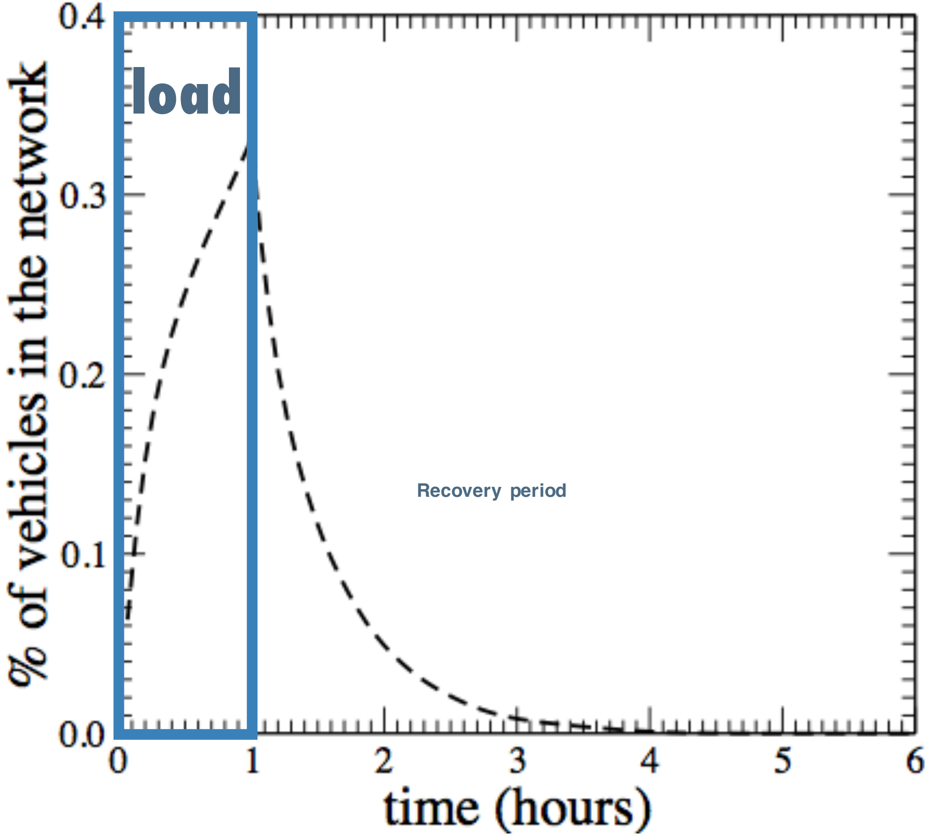
$$\Gamma = \frac{\sum_{e \in E} l_e x_e}{\sum_{x_e > 0, e \in E} l_e C_e}.$$

x_e # Cars in the road link e

l_e Road link length in miles

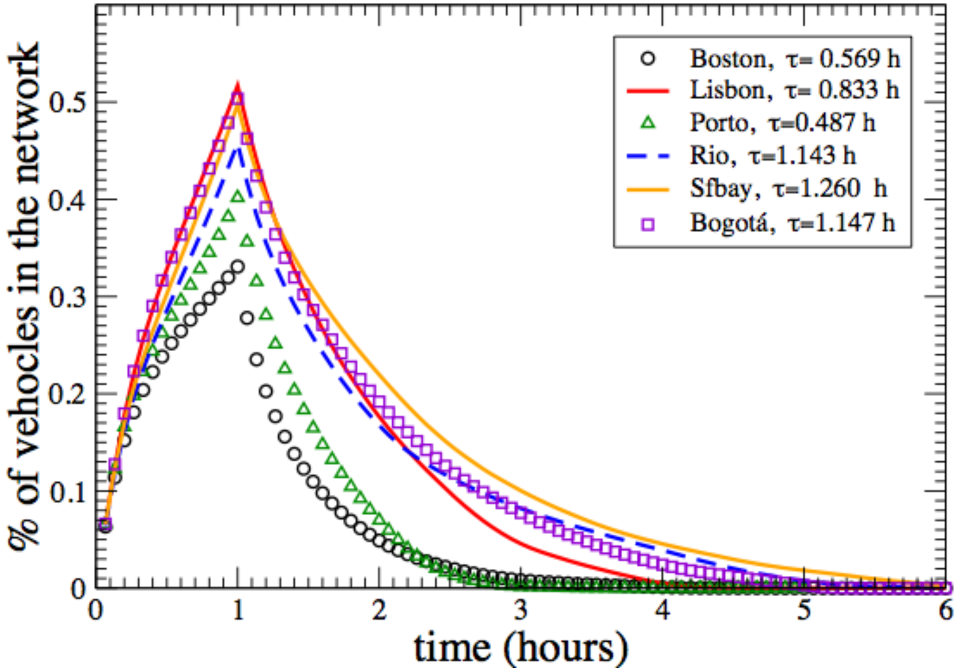
C_e Capacity in the road link [cars/miles^2]

Morning peak as a cycle of loading and unloading.



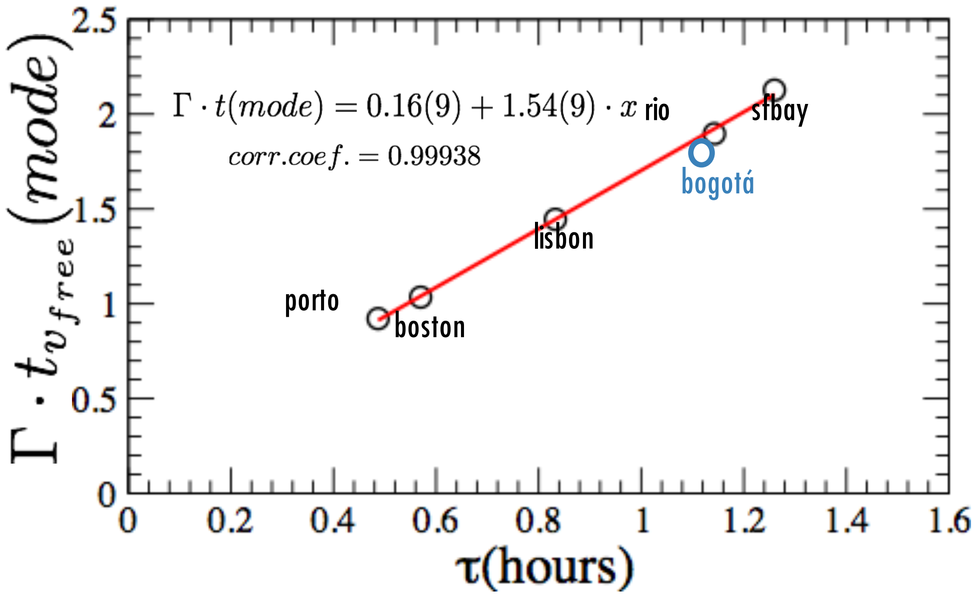
How long does it take to unload?

$$n(t) = n(1h)e^{-t/\tau}$$

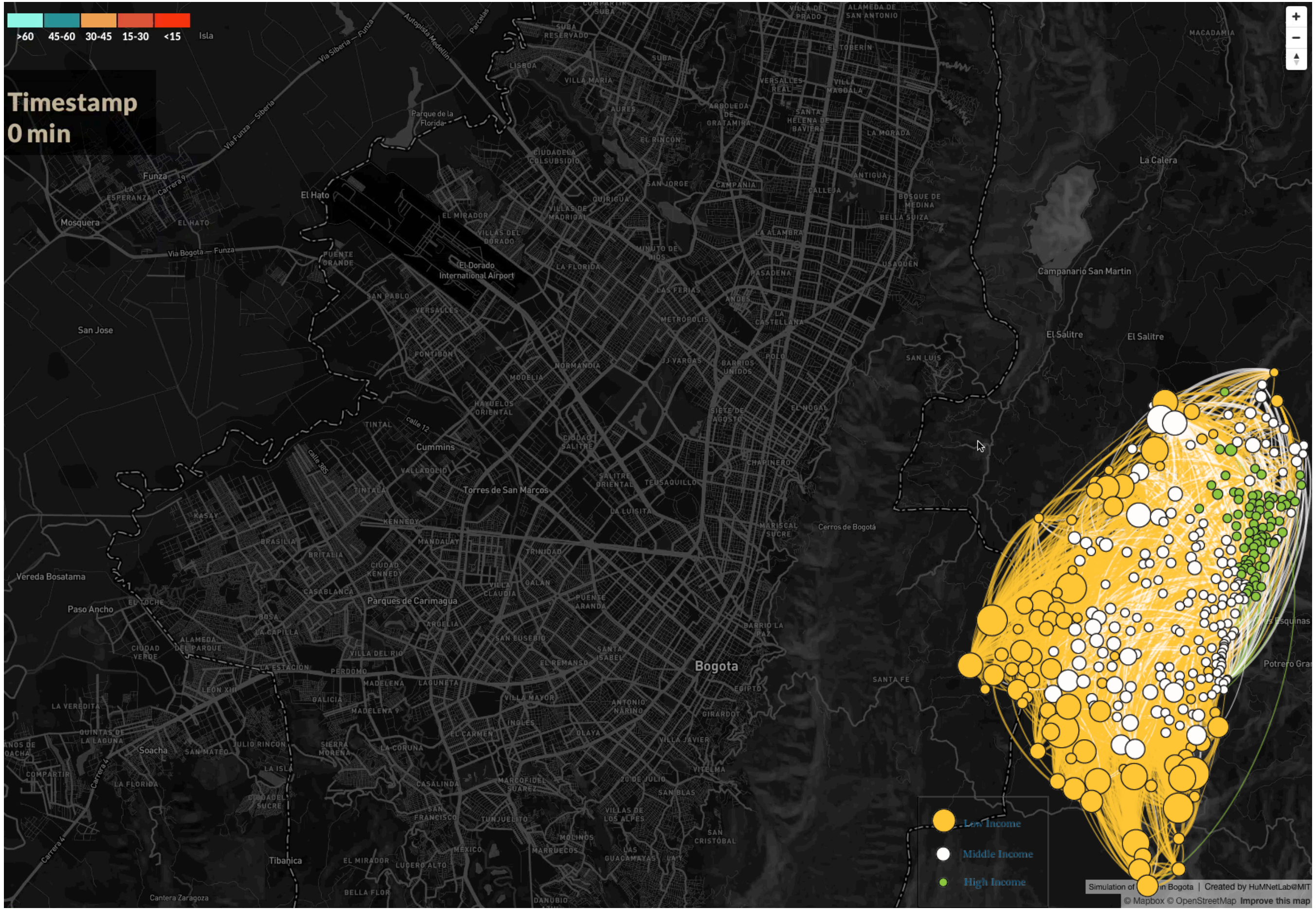


average over 20 configurations

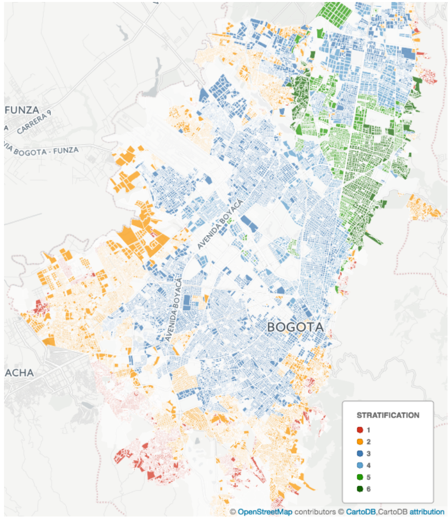
τ is the results of Travel distances in free flow, #cars and available space



$$\tau \propto t(mode) \cdot \Gamma$$



Current work: Decision platform for urban transportation



SIDE WALK LABS

Alphabet Inc's (NASDAQ:GOOGL)

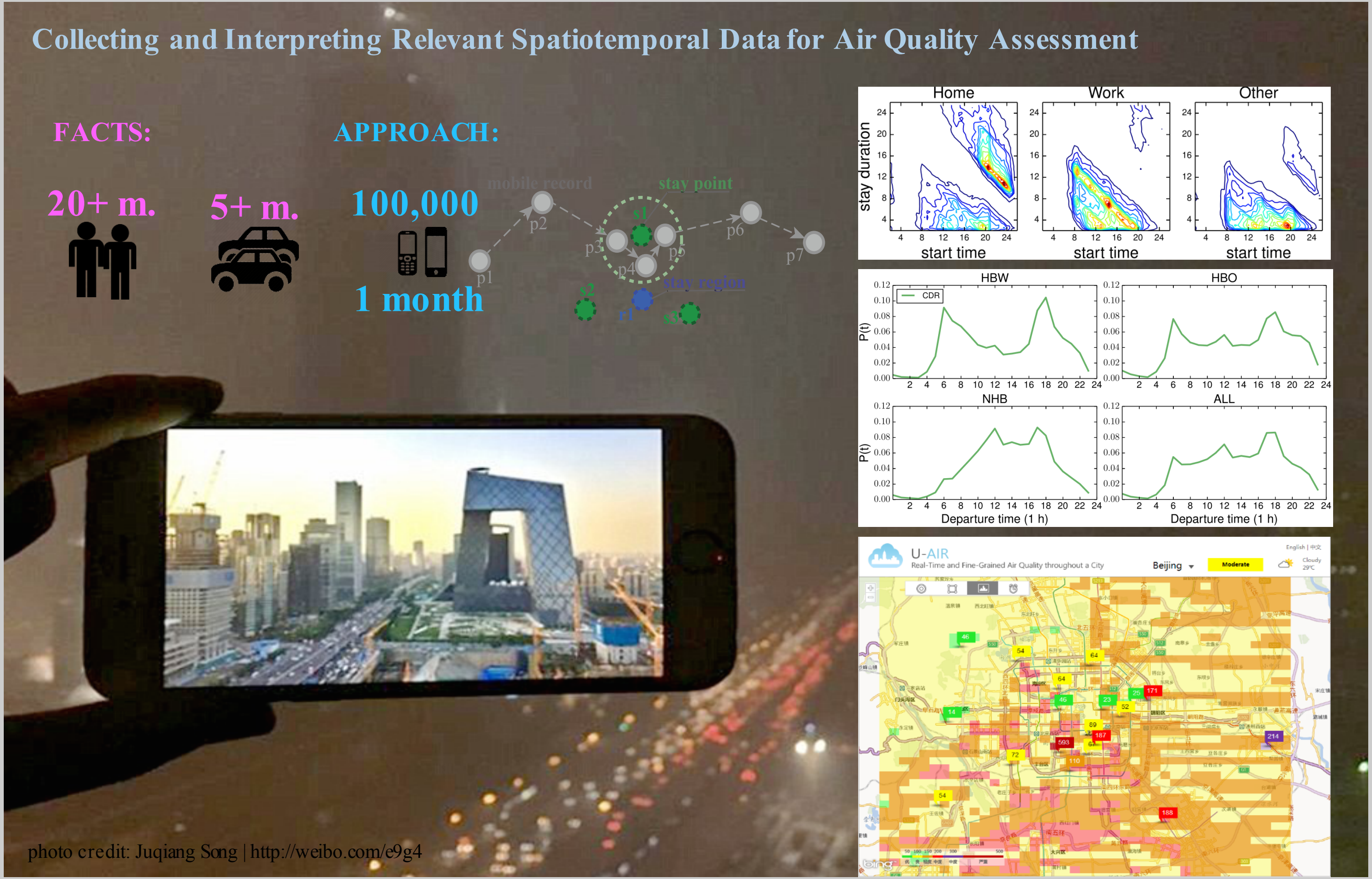
FLOW USDOT

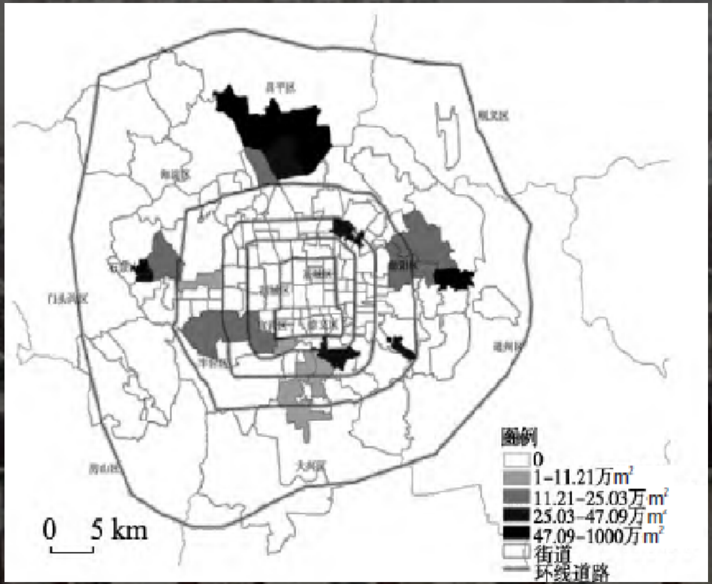
A transportation coordination platform that uses analytics and messaging to help cities work with citizens to increase the efficiency of road, parking, and transit use, improving access to mobility for all.

Urbanization in China New Opportunities: Big Data Study Area: Beijing

- Housing Price to estimate housing affordability
- Linking Affordable Housing Policy with
- Traffic Congestion Mitigation

Method and Results

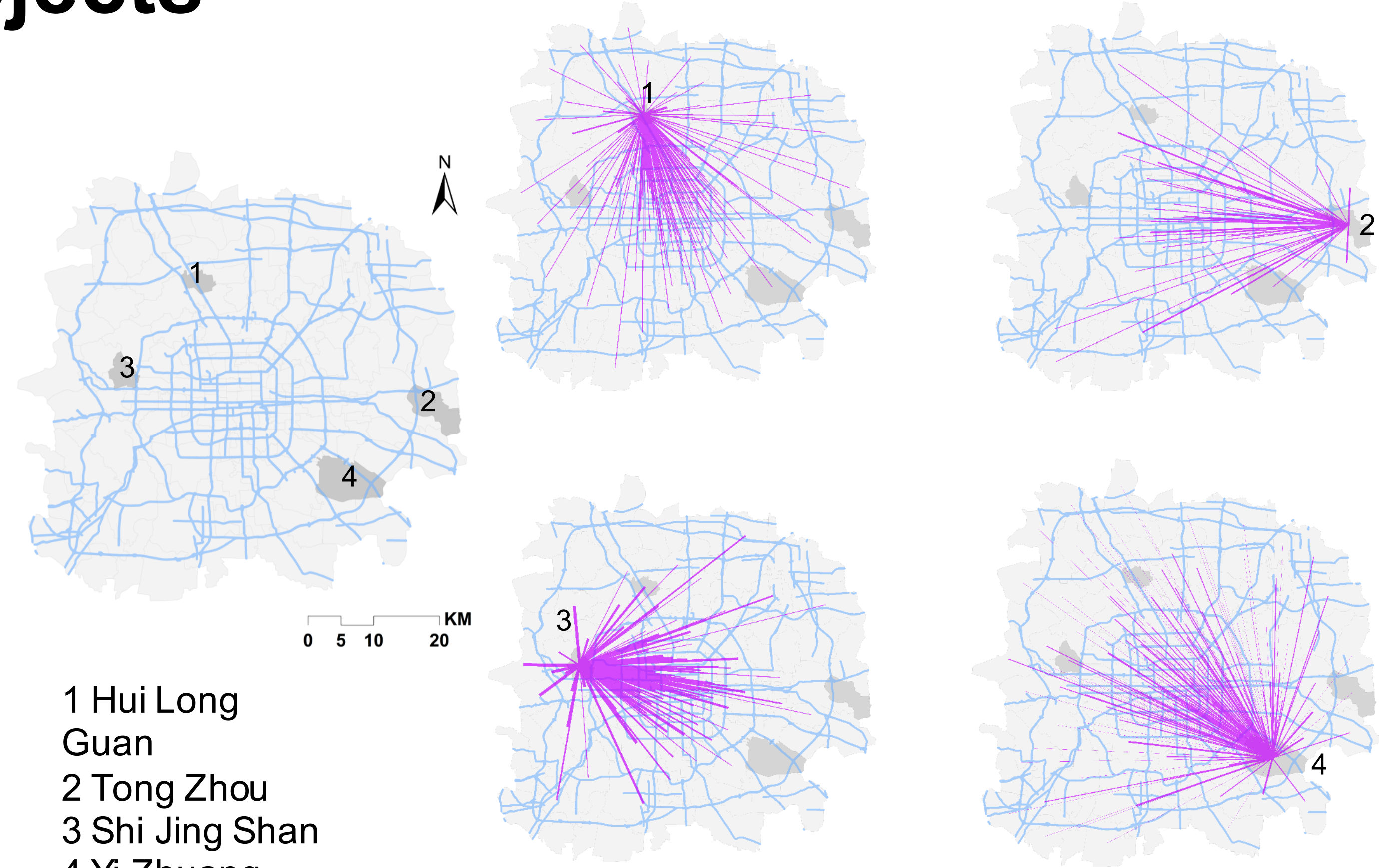




Distribution of Affordable Housing projects in Beijing 2009 (Source: Dang, et al 2014)

Source: rediff

AM OD flows from Affordable Housing Projects



1 Hui Long Guan
2 Tong Zhou
3 Shi Jing Shan
4 Yi Zhuang

Adoption of Financial Services for Low Income Groups

Credit Card Users	251,000
Debit Card Users	855,000
#Transactions	23 millions

Population	8.9 millions
Area	1,485 km^2

Dr. Riccardo Di Clemente
Postdoctoral Associate
MiT HuMNet Lab
rdicle@mit.edu

GRANDATA
data2x
BILL & MELINDA GATES foundation
CIUDAD DE MEXICO

Spending Pattern

Proxy for urban mobility

S. Sobolevsky et al., 2014 IEEE international congress on big data, 136–143 (2014).

Predictability of the shoppers' visitation pattens

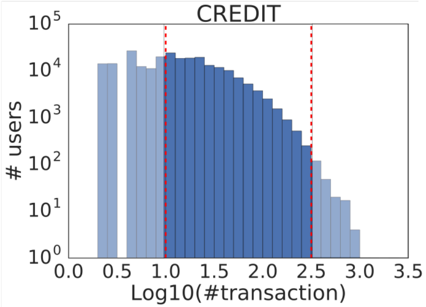
C. Krumme, *et al.*, *Sci. Rep.* **3,** 10.1038/srep01645 (2013).

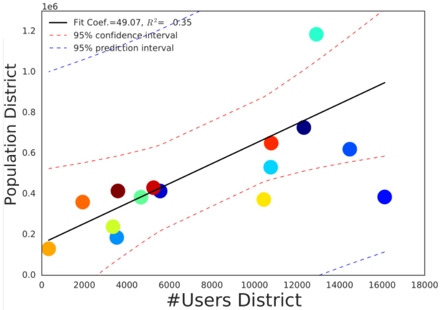
Users' Reidentificability

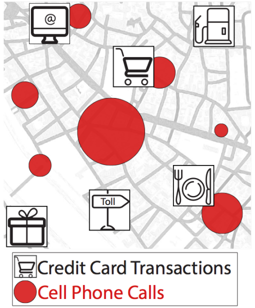
Y.-A. De Montjoye, *et al.*, *Science,* **347,** 536–539 (2015).

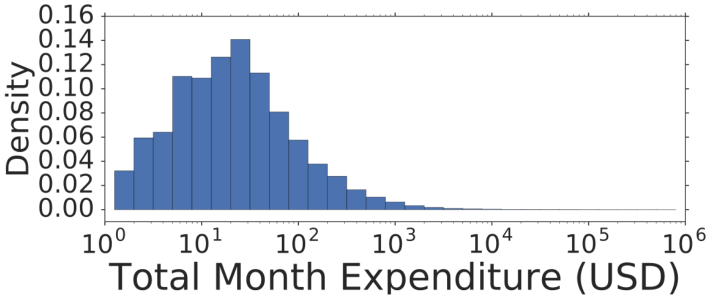
How much of our life-style can be described by the sequence of credit card swipes?

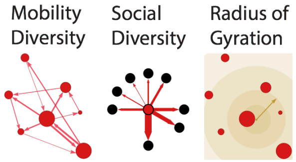
Coupling Credit Card Data with Mobile Phone Metrics



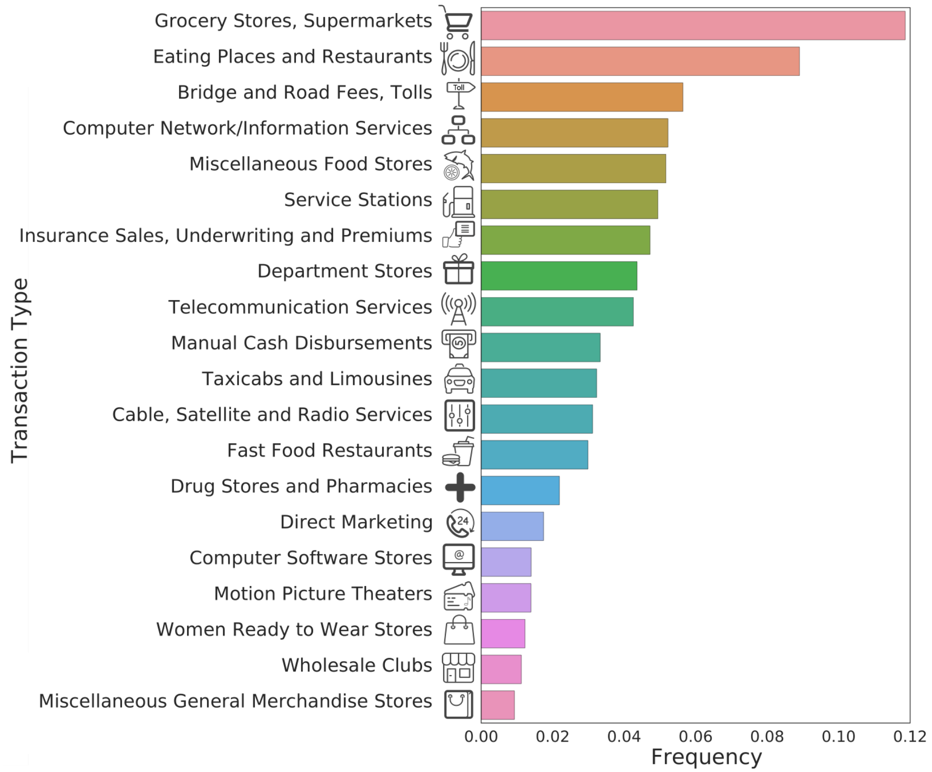


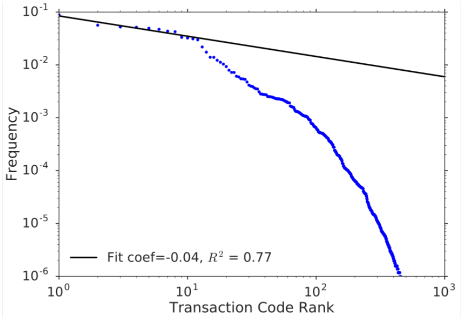




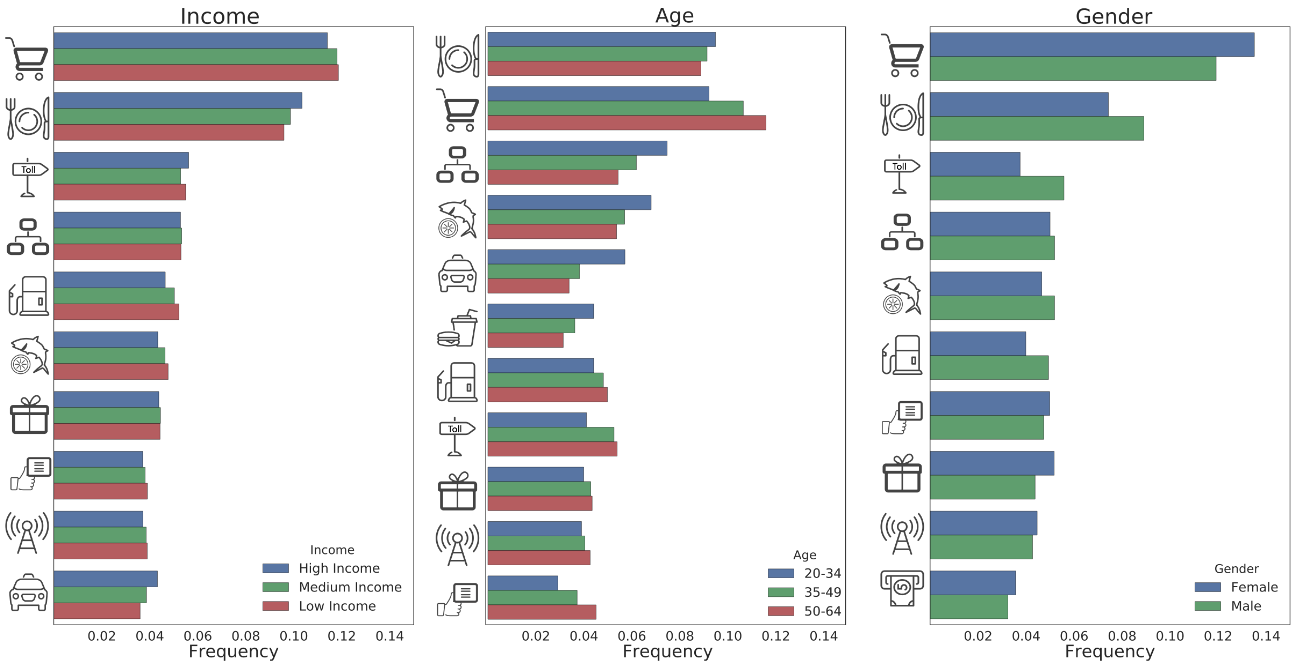


Emerging Trends in Transactions' Type





Comparative Analysis of transactions' type

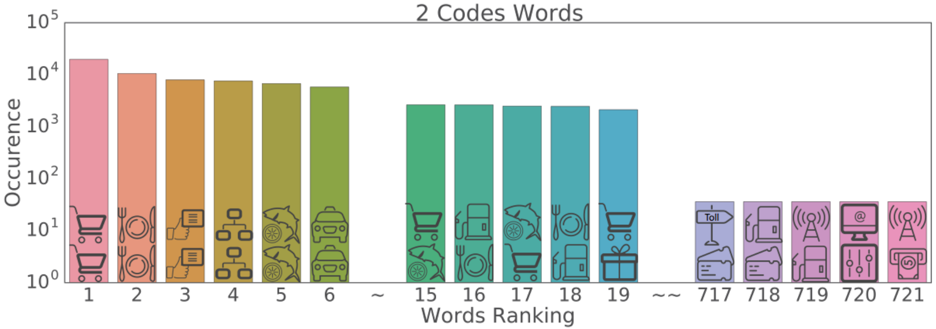


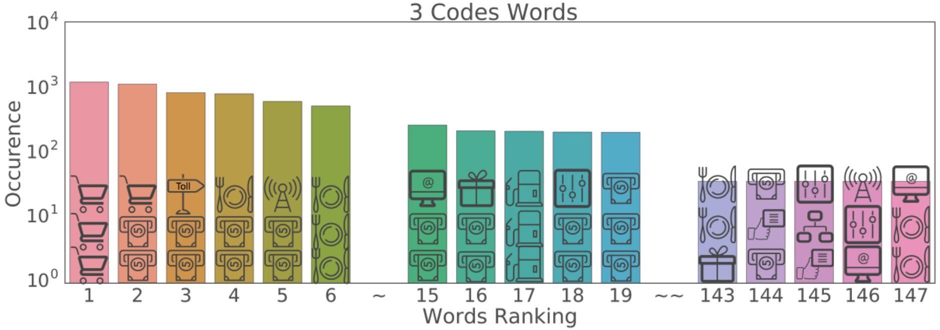
Sequitur Algorithm

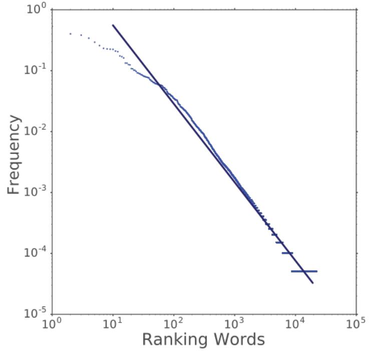
equence of discrete symbols.

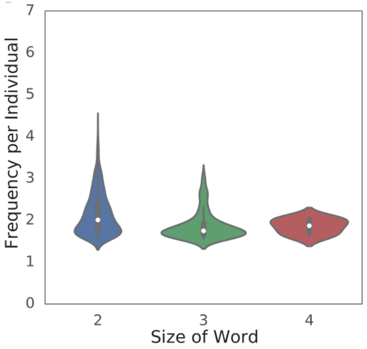
Compressed Sequence

Words as Ordered Sequence of transactions

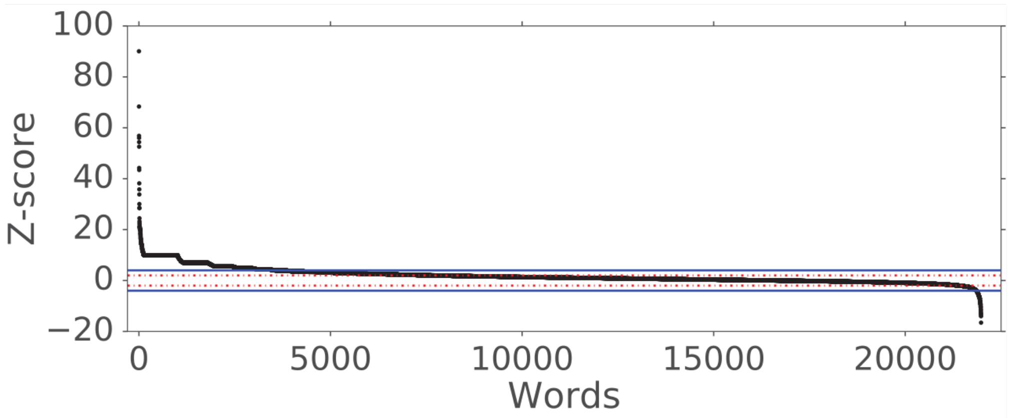


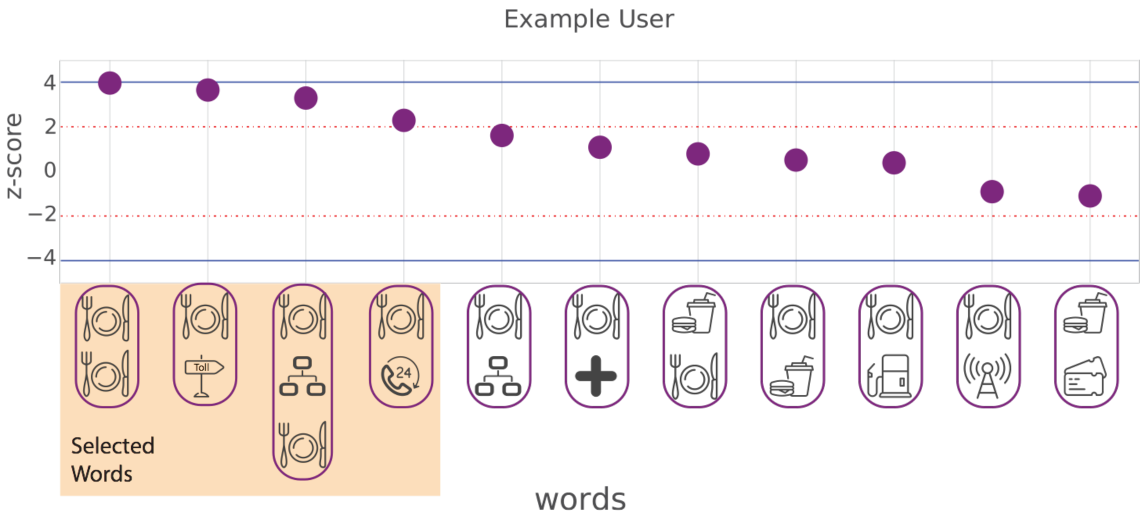


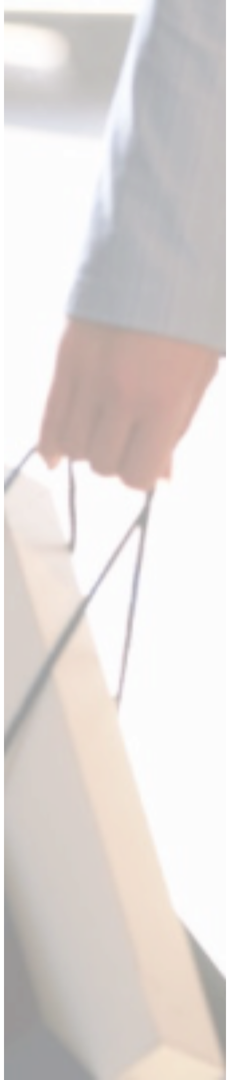




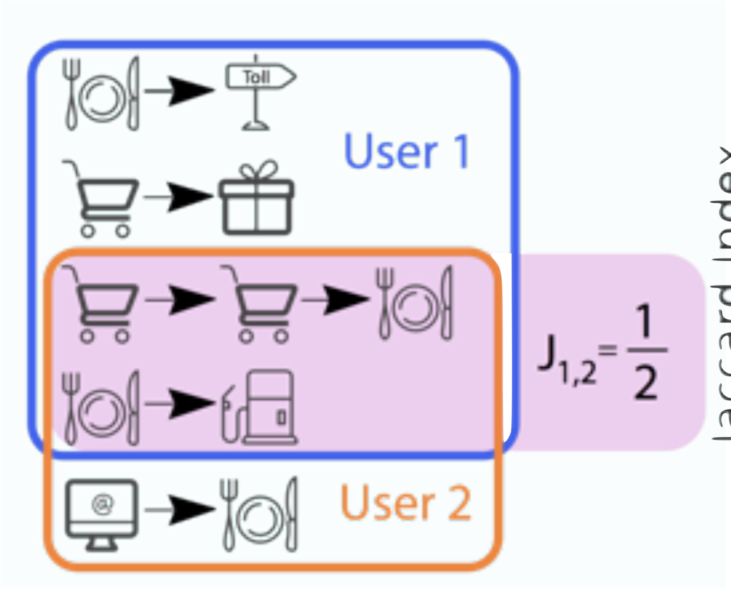
Quantifying Words significance

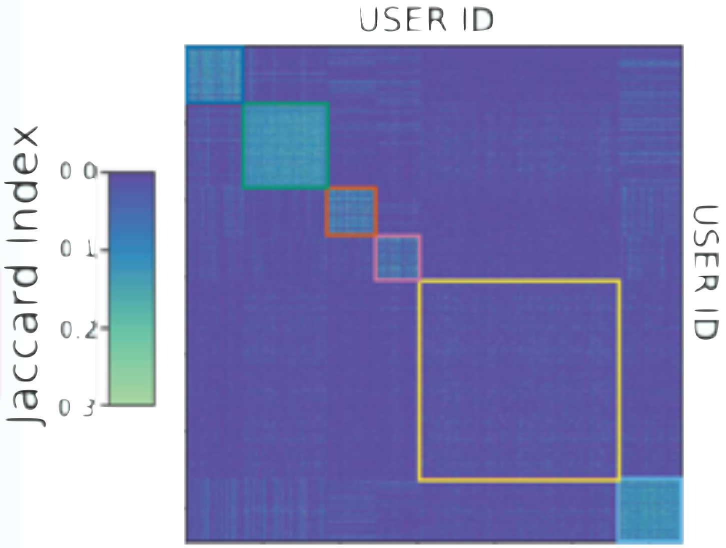




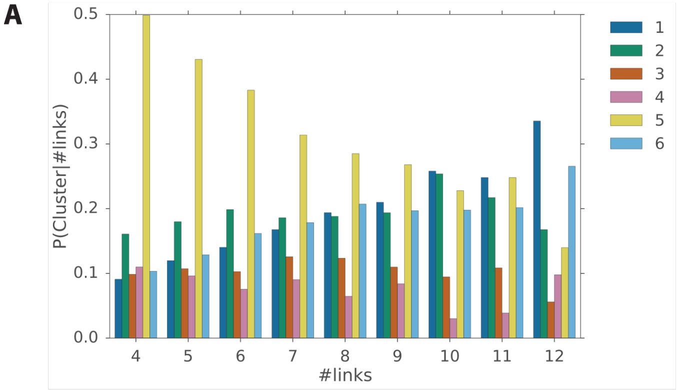


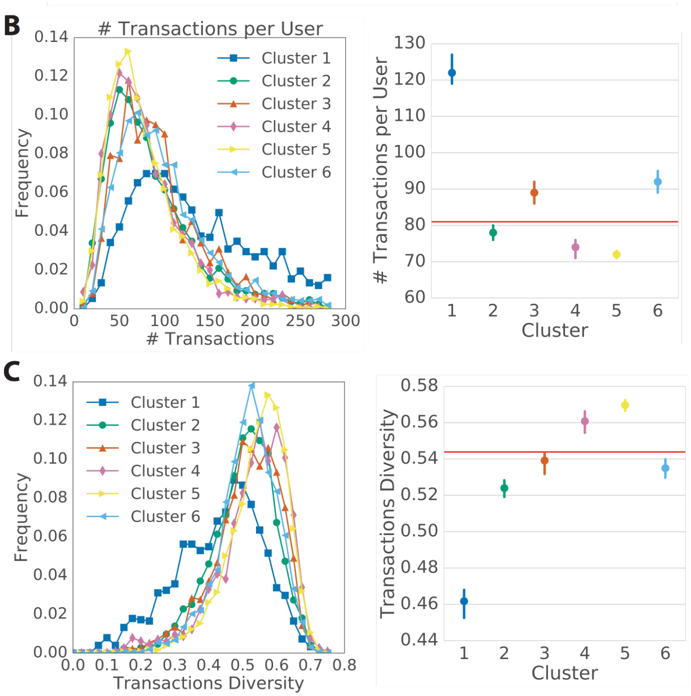
How to Cluster the Users?



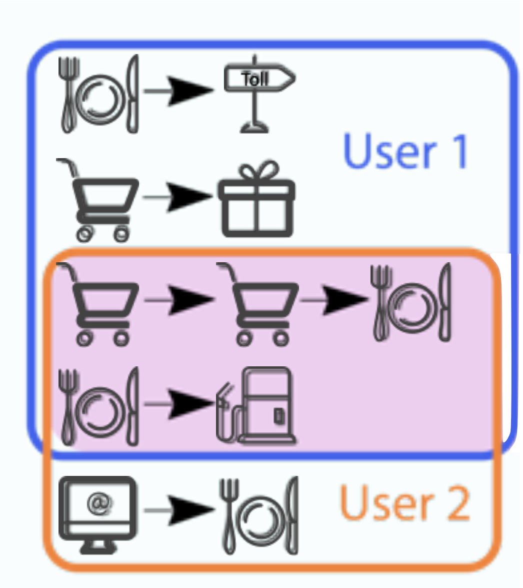


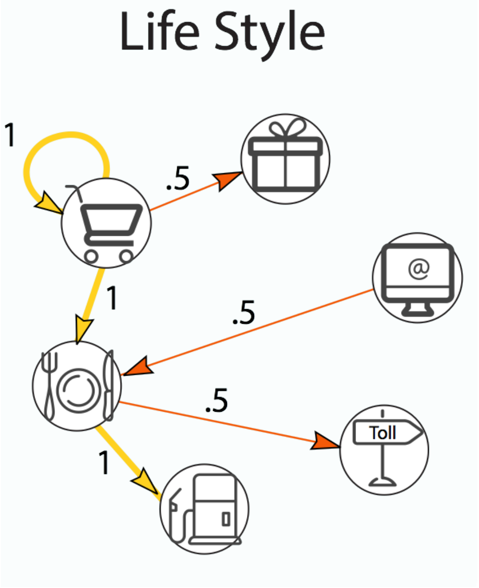
Characterizing the clusters



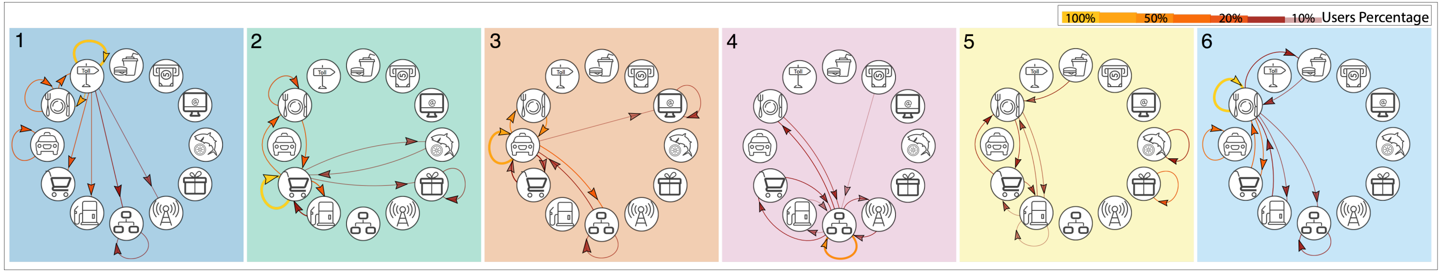


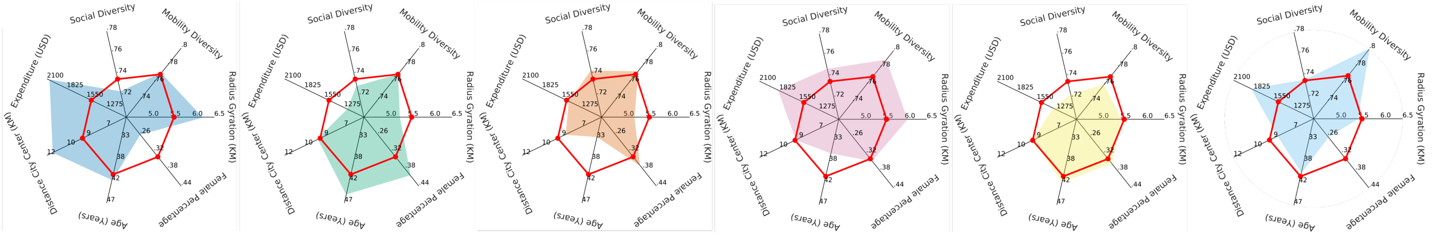
Defining Life-Styles





Life - Style





Commuters

Household-chief

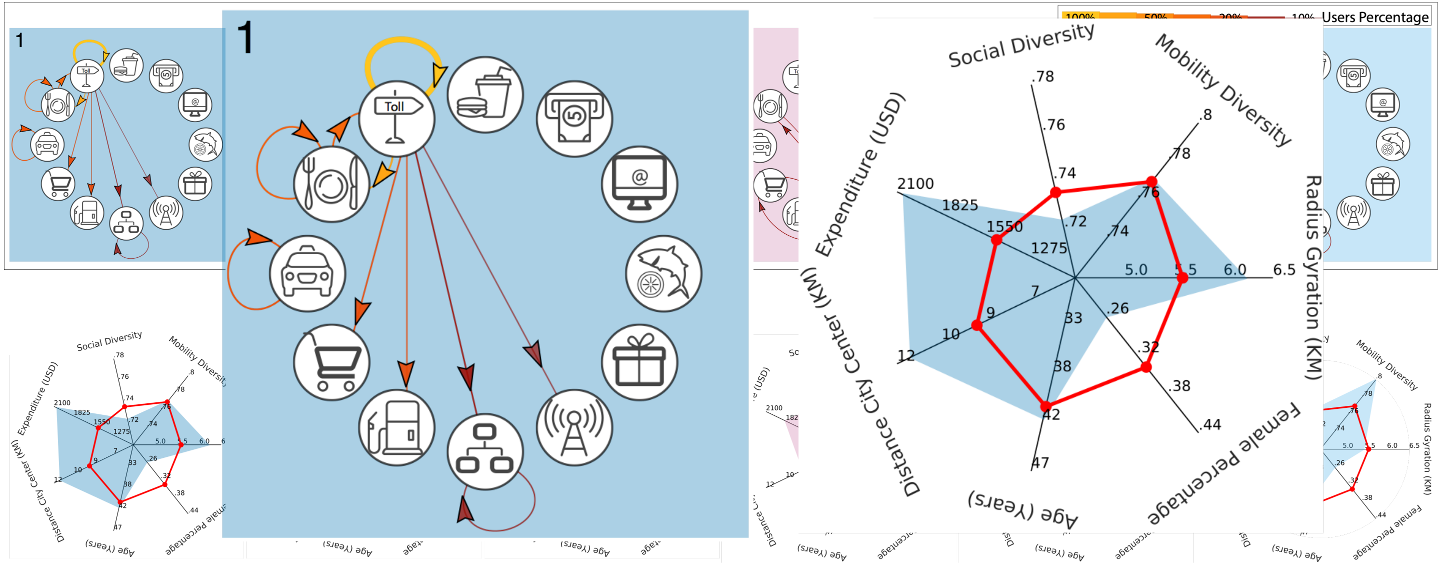
Youth

High-tech

Average

Diners

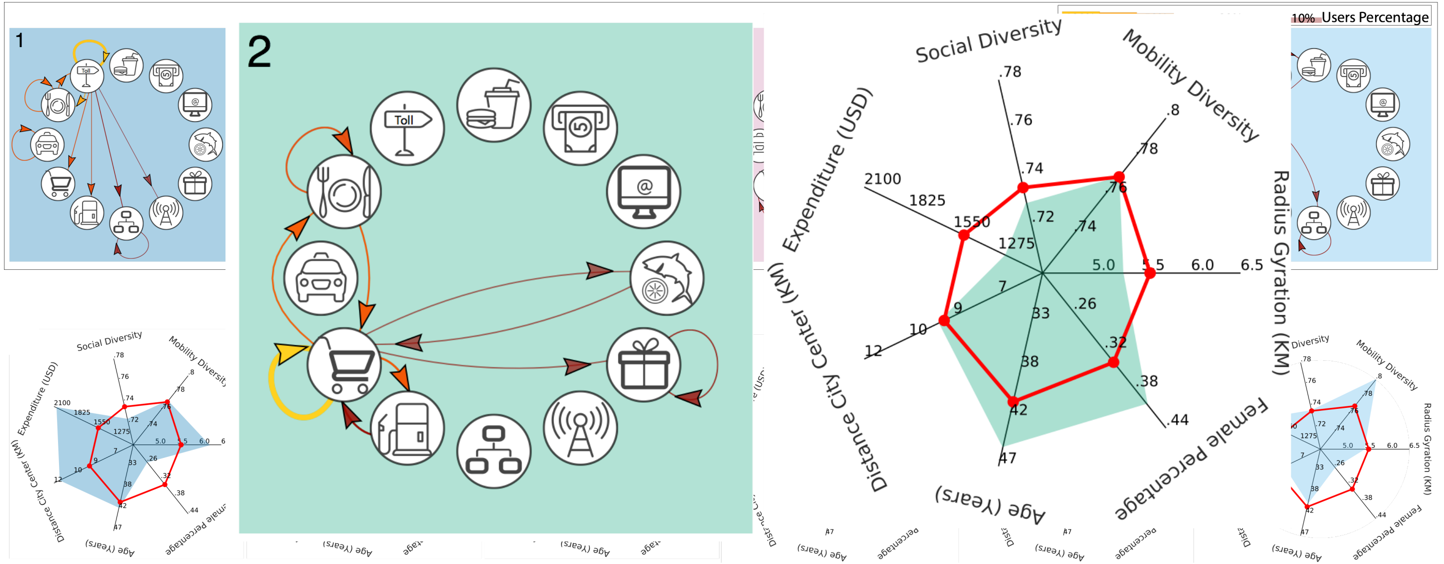
Life - Style



Commuters:

High expenditure, living far from the city center, higher radius of gyration, low social diversity and female %

Life - Style

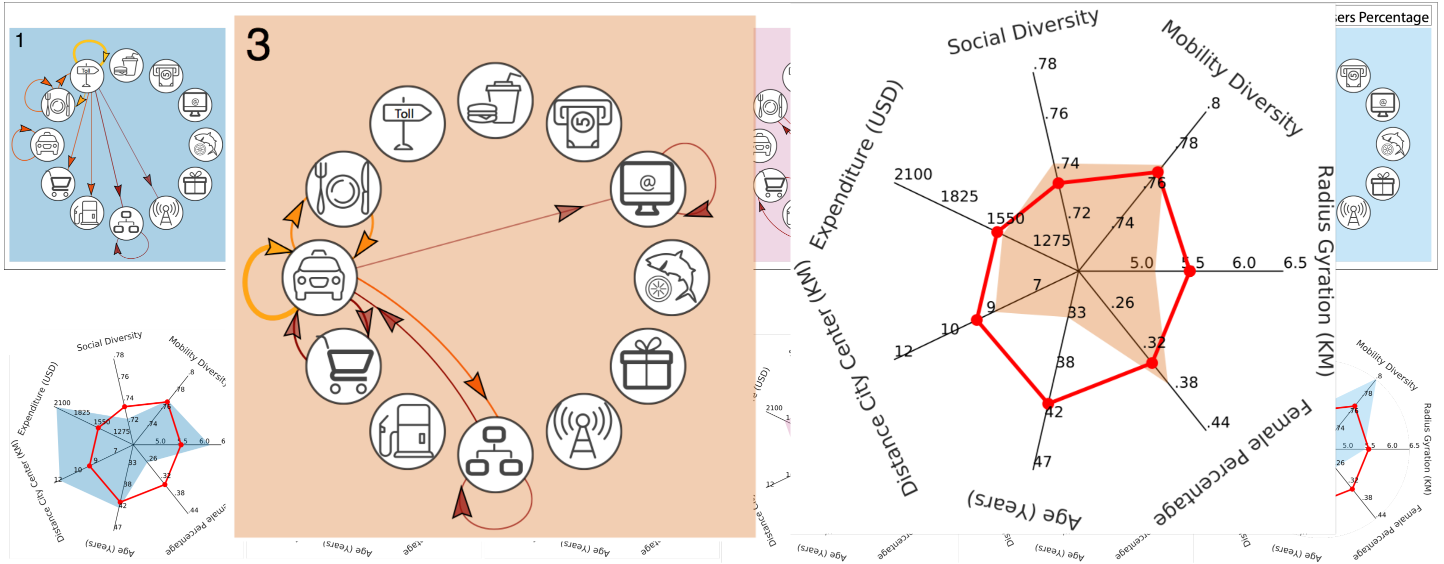


Household-chiefs:

Lower expenditure, older Age, more females

Commuting

Life - Style

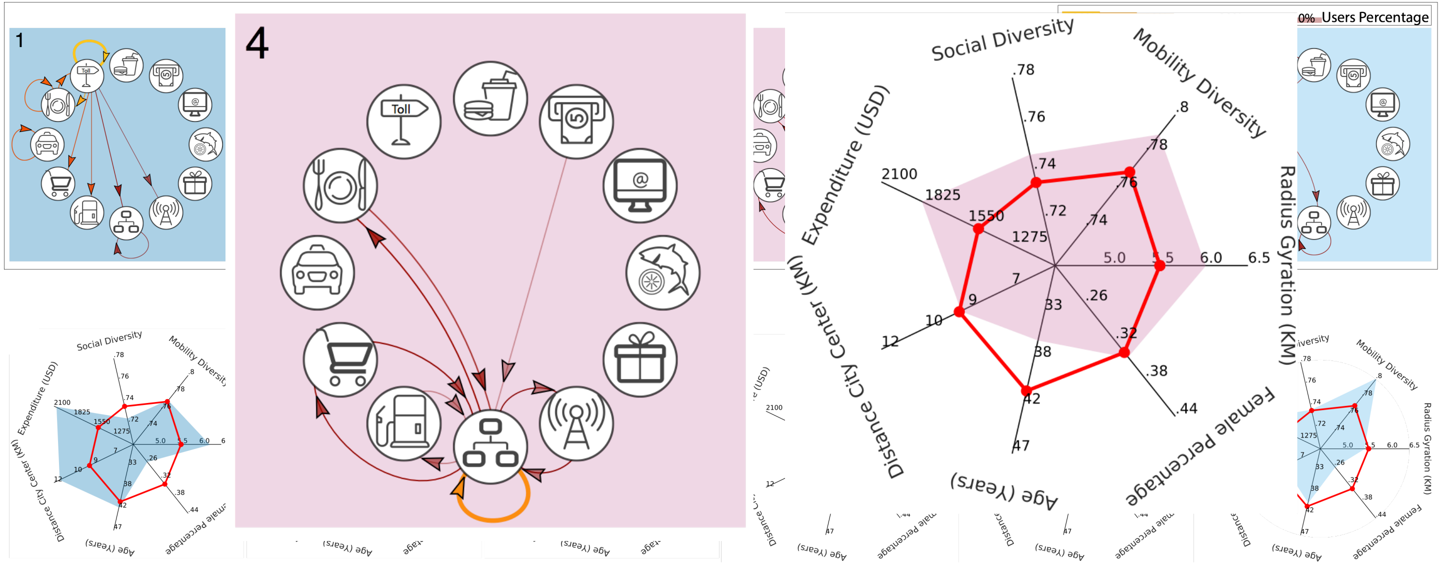


Youth:

Younger age, taxi and communication technologies

Commuting

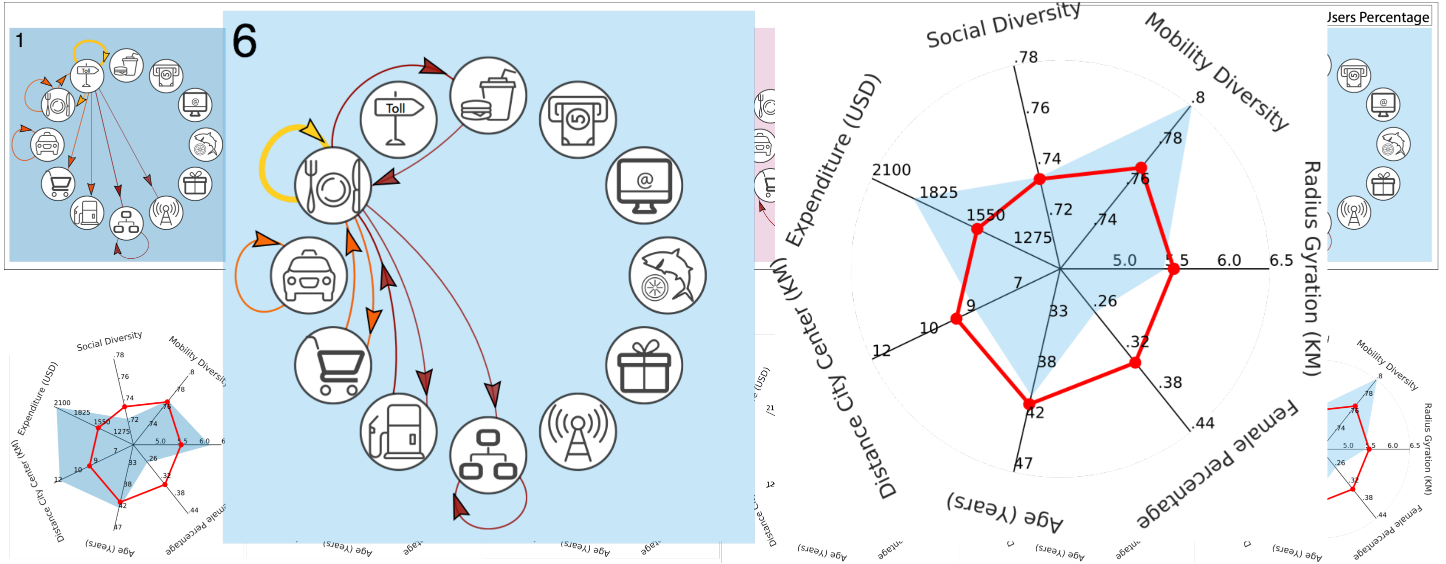
Life - Style



High-techs:

Higer expenditure, social and mobility diversity, core transaction is information technologies

Commuting

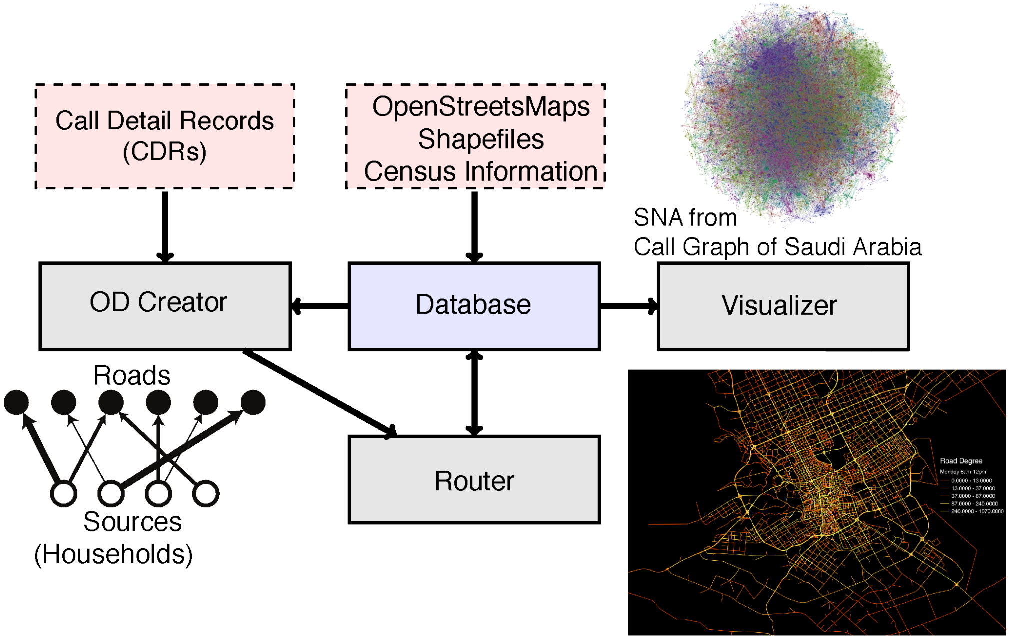


Diners:

High expenditure and mobility diversity, less females.
Restaurants is core transaction

Commuting

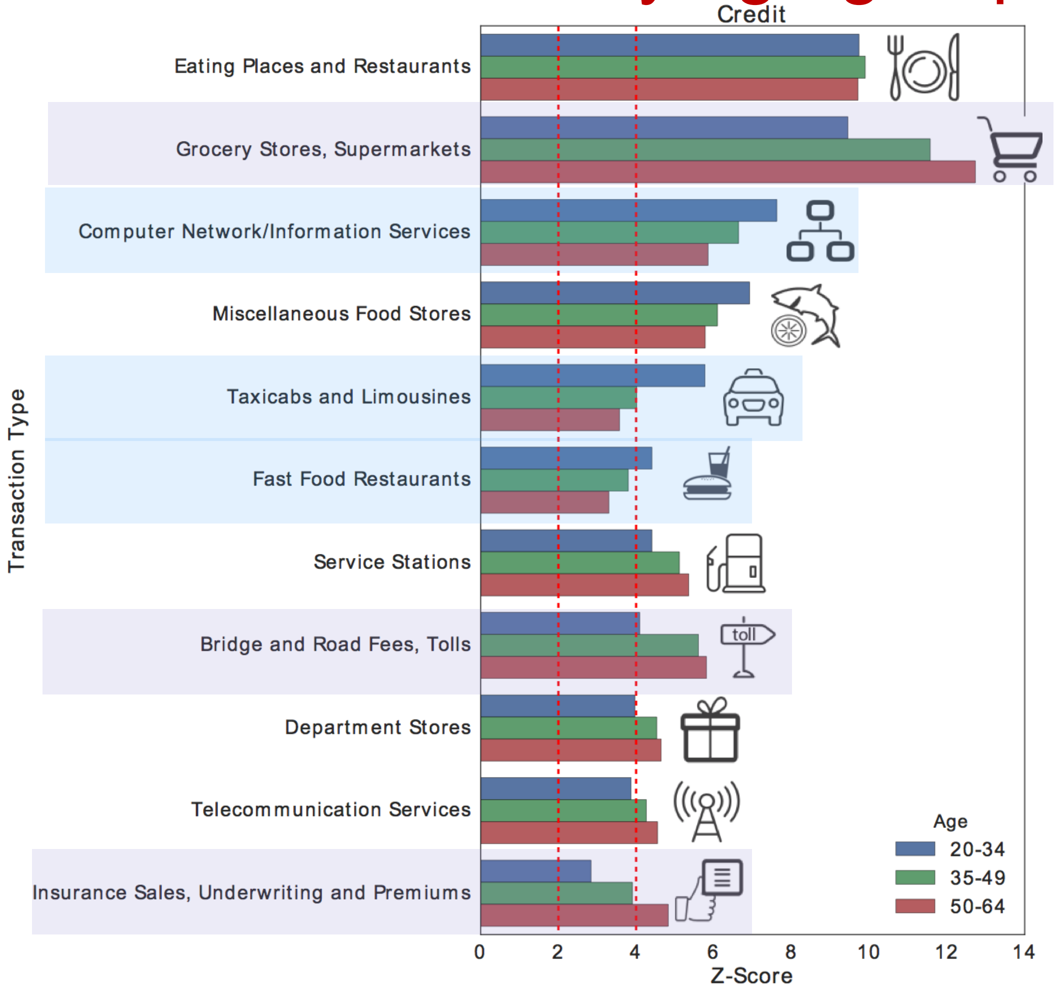
Conclusions



Riyadh road attractors

- Today we can measure behavior from communication technologies and can plan cities with them.
- Real time incentives is the next frontier.
- Integration of data from companies, governments and online platforms is needed.
- Platforms for urban data, and flow of information there may be the next Internet.

Expenditures behavior by age group



Age 20-34:

Computer store,
TAXI,
Fast Food

Age 50-64

Grocery store,
Insurance Sales,
Highway Tolls