A Complex System approach to Study Cities and Human Mobility

Marta C. Gonzalez

Associate Professor of City and Regional Planning

Berkeley
UNIVERSITY OF CALIFORNIA

Lawrence Berkeley National Laboratory

HUMNET

Civil and Environmental Engineering

Massachusetts Institute of Technology

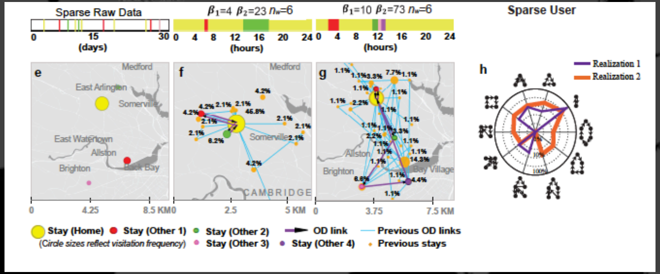
those who see similarities, patterns and universals…

those who see differences, variation and specifics.

P. Ball, Complexity: Decoding deep similarities. Nature, 545, 154–155 (11 May 2017)

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

Proceedings of the National Academy of Sciences (2016): 201524261.



BOSTON

Slower Faster

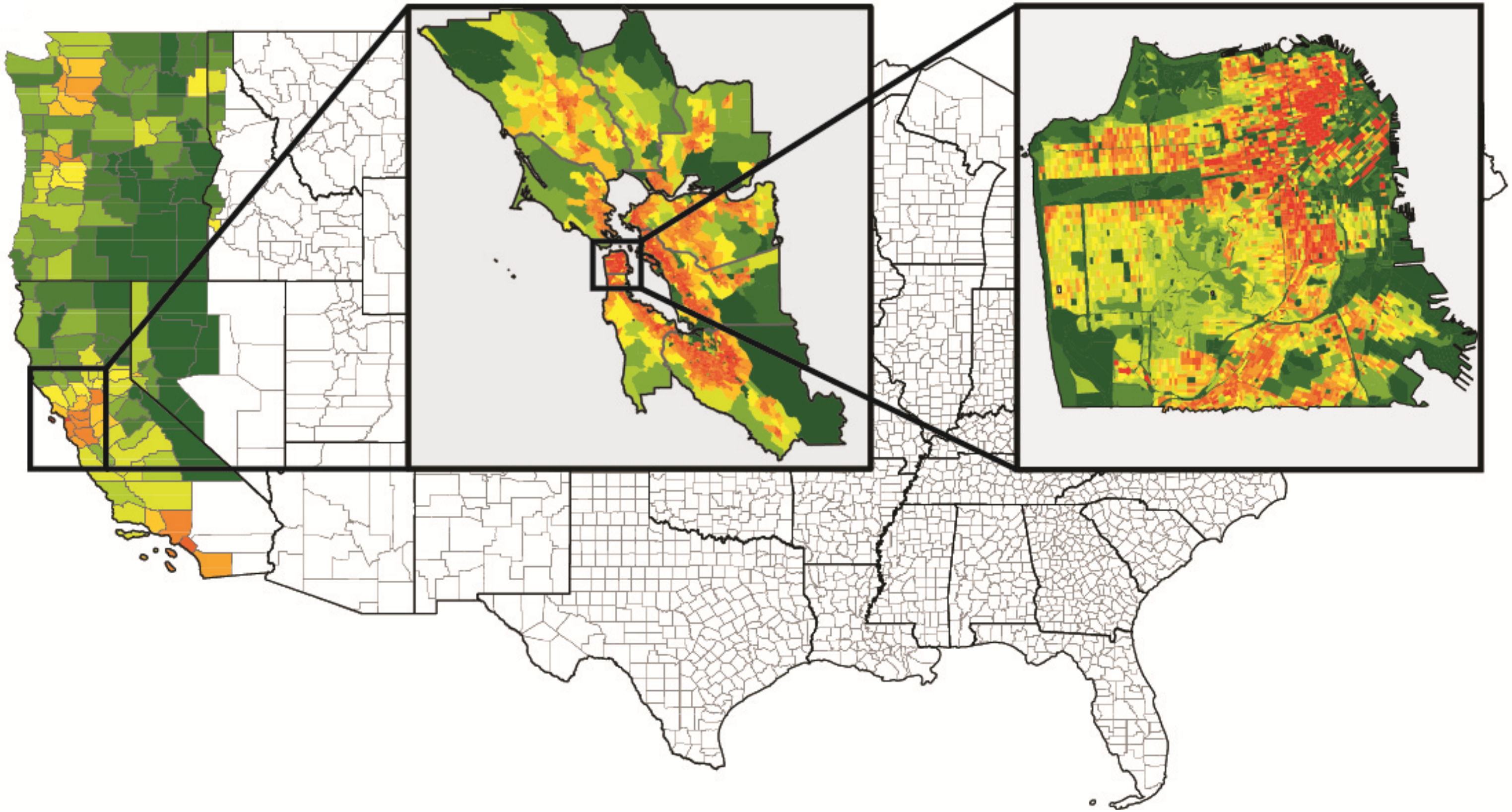
real minutes/animation second

Home

Work

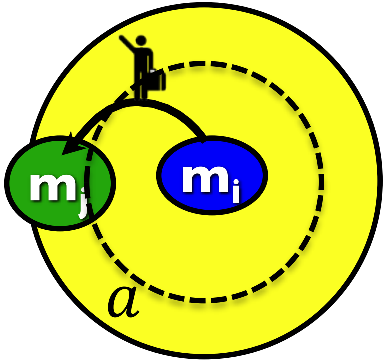
Other

Universality 1: Ranking of Opportunities



Gravity Laws and Opportunity Laws

$$p_{ij} = \mathbb{P}(i)\,\mathbb{P}(1|i,j)$$



$$p_{ij} \propto m_i\, m_j\, e^{-\beta\, d_{ij}}$$

$$\frac{\left[(m_i + m_j + s_{ij})^{\alpha} - (m_i + s_{ij})^{\alpha}\right](m_i{}^{\alpha} + 1)}{\left[(m_i + s_{ij})^{\alpha} + 1\right]\left[(m_i + m_j + s_{ij})^{\alpha} + 1\right]}$$

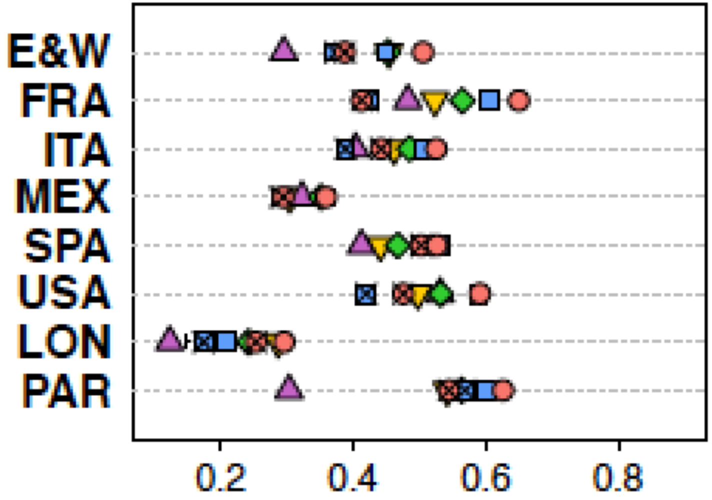
$$\mathrm{P}(1|i,j) = \frac{P_{>}(a) - P_{>}(a + m_j)}{P_{>}(m_i)}$$

$$a = s_{ij} + m_i$$

s_{ij} opportunities between *i* and *j*

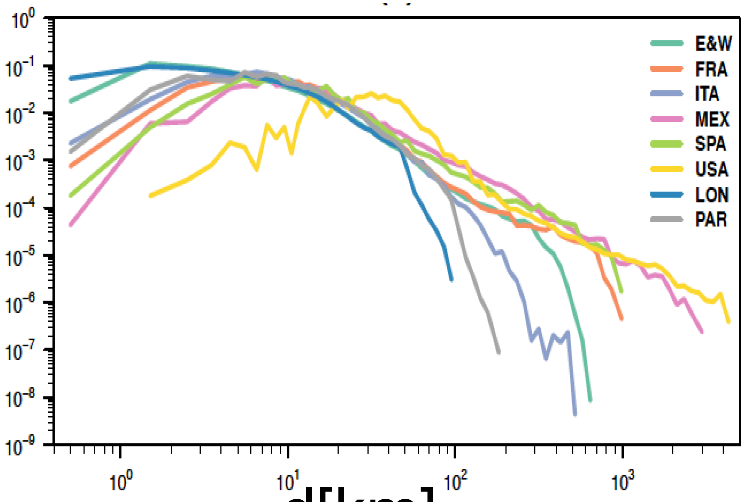
Simini, F., M. C. González, A. Maritan, and A-L Barabási. "A universal model for mobility and migration patterns." *NATURE* 484 (2012): 5.

Systematic comparison of trip distribution laws and models



Common parts per Commuter

P(d)



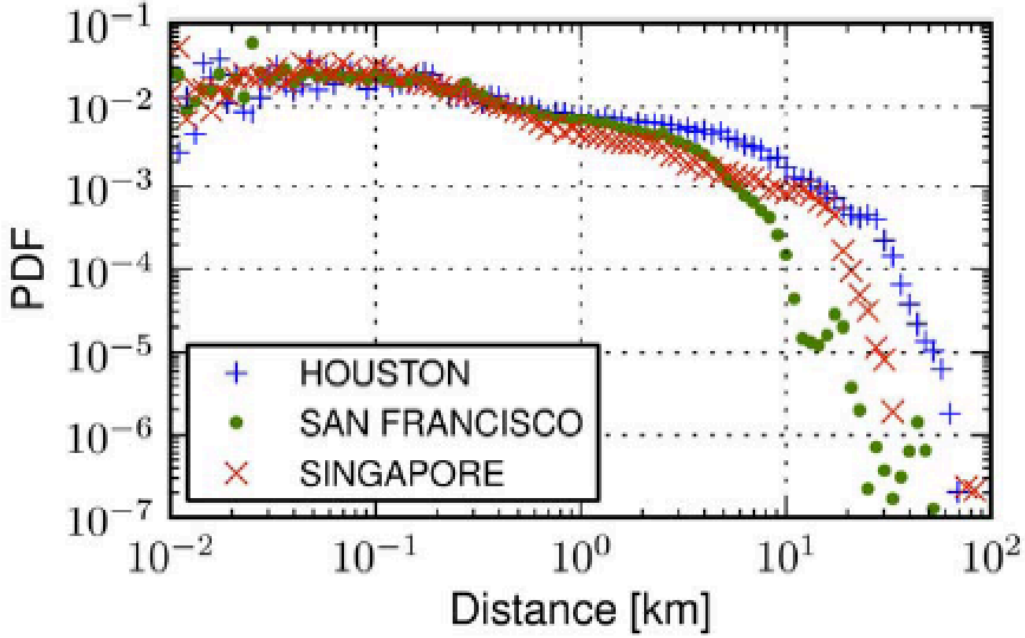
d[km]

Distribution of distance traveled

Random CPC = 0.01

Lenormand, Maxime, Aleix Bassolas, and José J. Ramasco. *Journal of Transport Geography* 51 (2016): 158-169.

Universality 2: Ranking of destinations

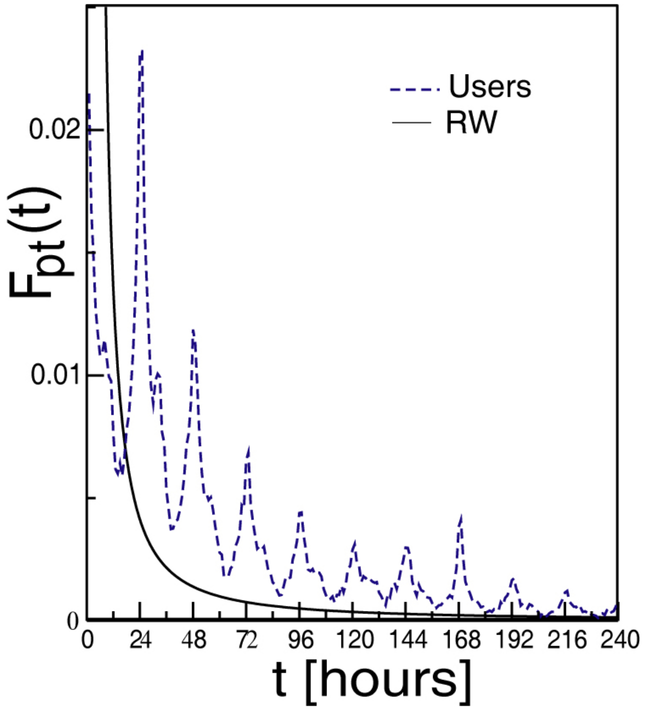


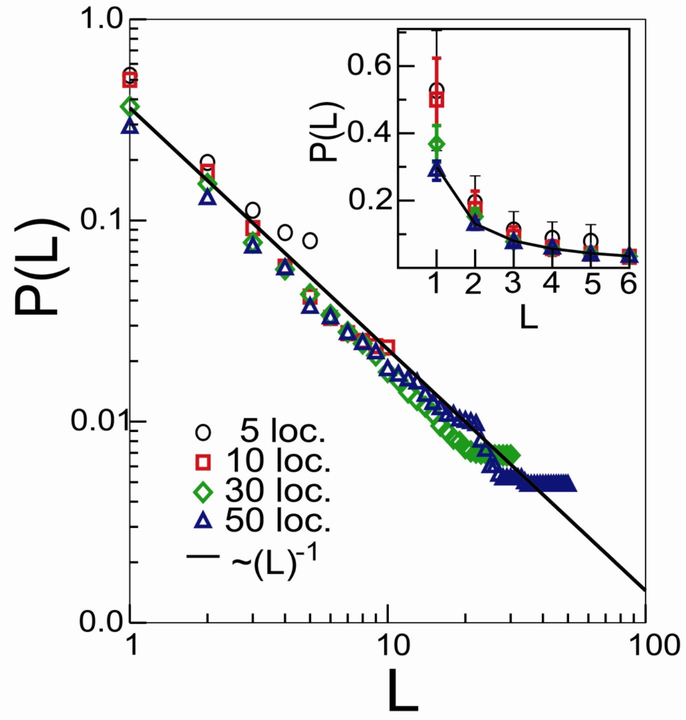
$$P_r[u \rightarrow v] \propto \frac{1}{rank_u(v)^a}$$

$$rank_u(v) = |\{w : d(u,w) < d(u,v)\}|.$$

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). **A tale of many cities: universal patterns in human urban mobility**. *PloS one*, *7*(5), e37027.

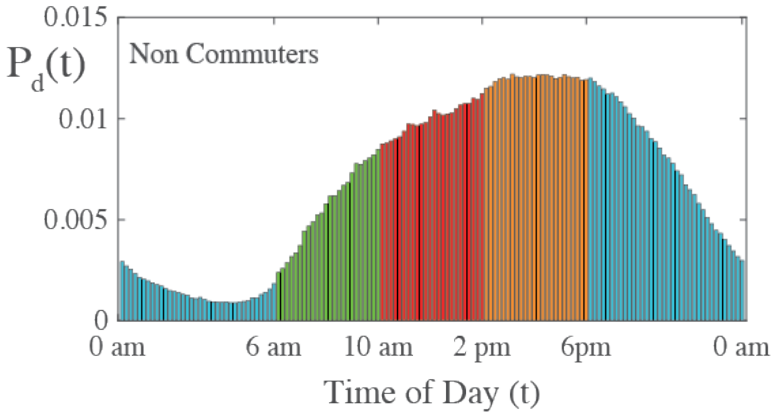
Universality 3: Preferential returns



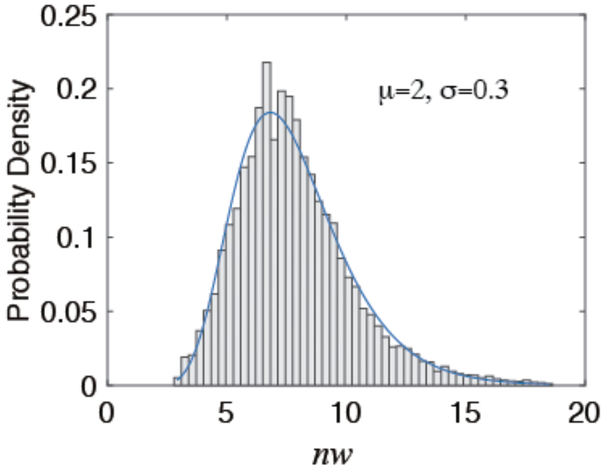


González, Hidalgo, Barabasi, *Nature* 2008

Circadian Rhythm + heterogeneity 1:



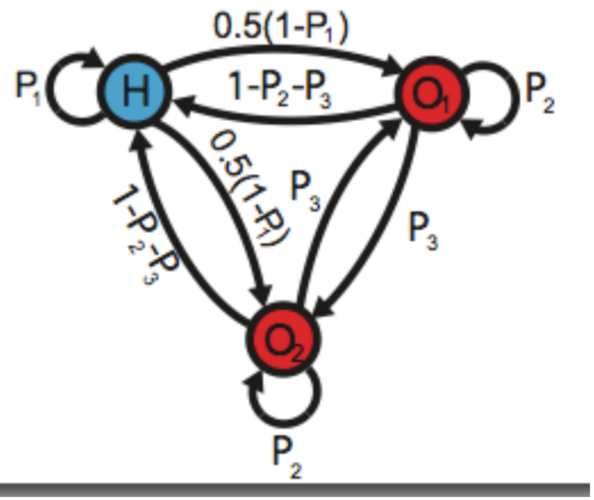
Fraction of Total Trips per time step



Weekly trips from home

$$P_i(t) = n_w^i P_d(t)$$

Markov Mobility Model



rhythm(t)

$$P_1 = 1 - P(t)$$ Staying at "home"

$$P_2 = 1 - \beta_1 P(t)$$ Staying in current errand

$$P_3 = \beta_1 P(t) \beta_2 P(t)$$ Moving from current errand to a new current errand

"bursts" rate

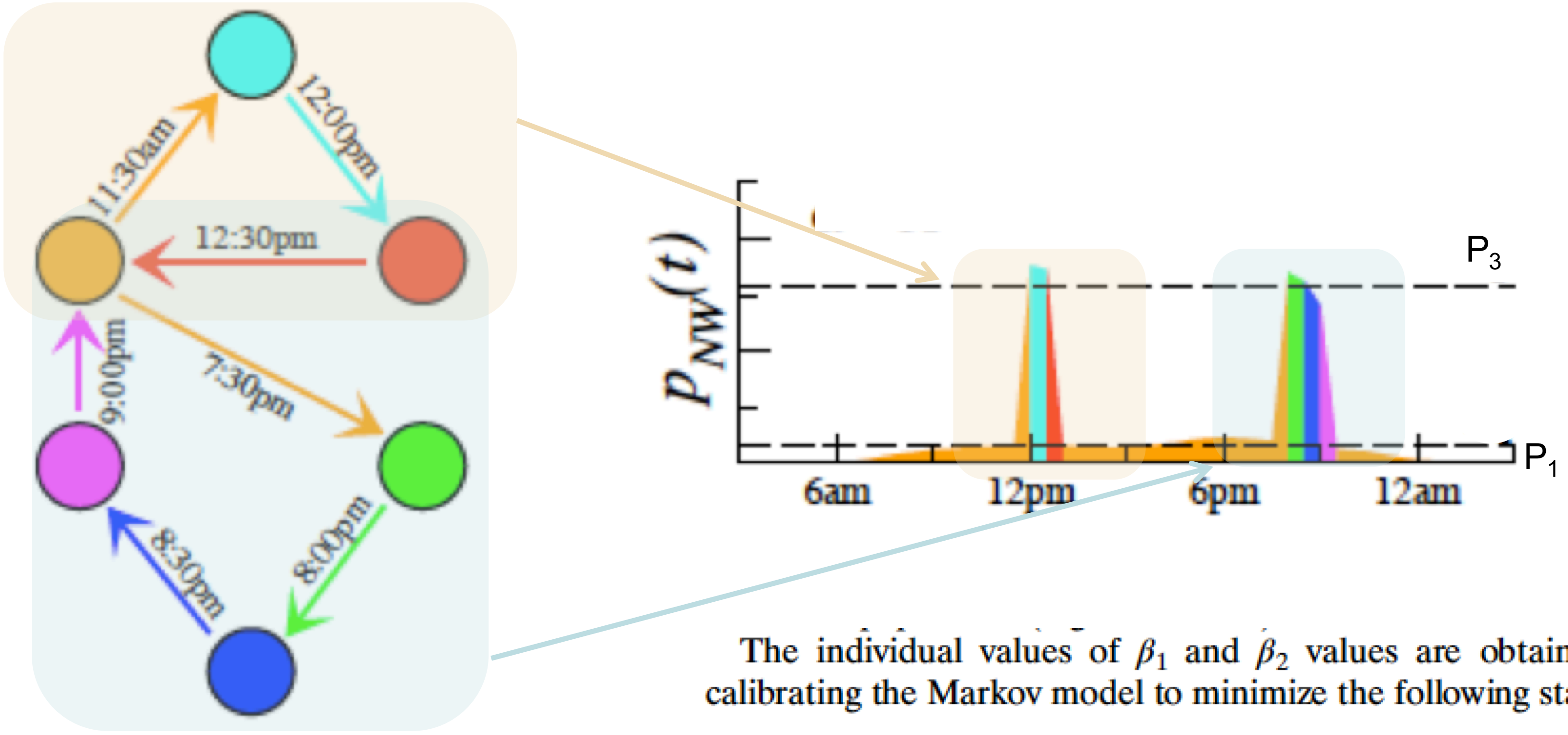
exploration rate

Generates shorter stays in the errand states.

Generates different number of activities in a row per **active** cycle

$$\frac{P(O_1 \to O_2)}{P(O_1 \to H)} = \frac{\beta_2 n_w P(t)}{1 - \beta_2 n_w P(t)}.$$

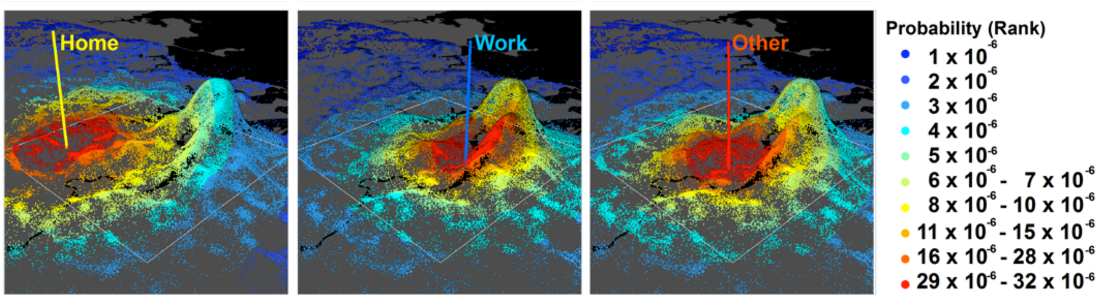
We fit the model to the observed heterogeneity



The individual values of β_1 and β_2 values are obtained by calibrating the Markov model to minimize the following statistic:

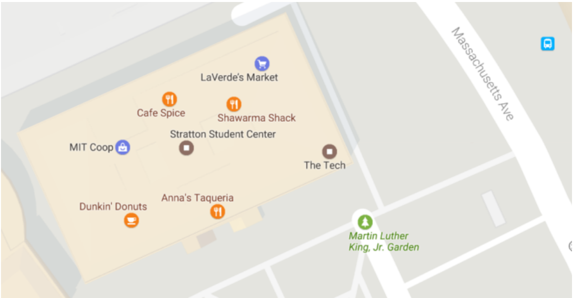
$$A(\beta_1,\beta_2)=\int |P_D(\Delta t)-P_M(\Delta t|\beta_1,\beta_2)|d\Delta t+\eta|\overline{N}_D-\overline{N}_M(\beta_1,\beta_2)|,$$

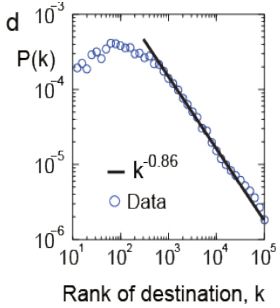
Explorations are selected via the opportunities law



Colors represent the P(rank), height is POIs (point of interests) numbers.

POIs = Point of Interests

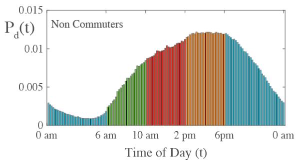




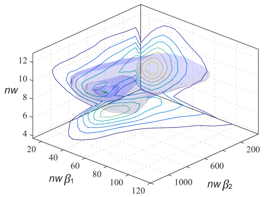
The TimeGeo Modeling framework

Features Extracted from data of Active Users

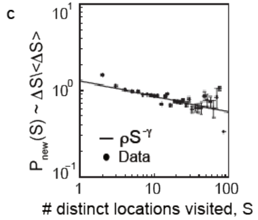
Global Trip prob.



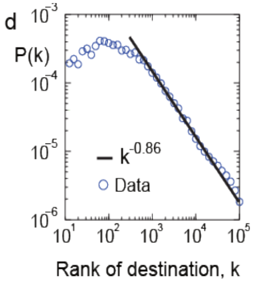
Individual Mobility Rates



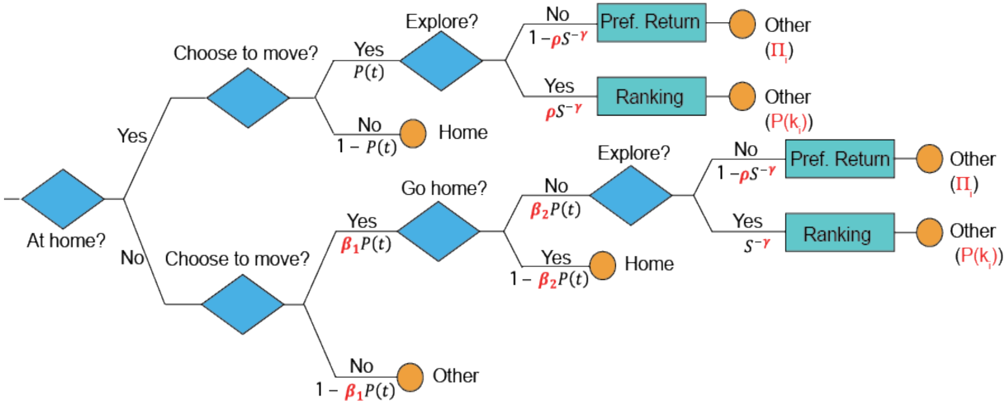
Preferential Return



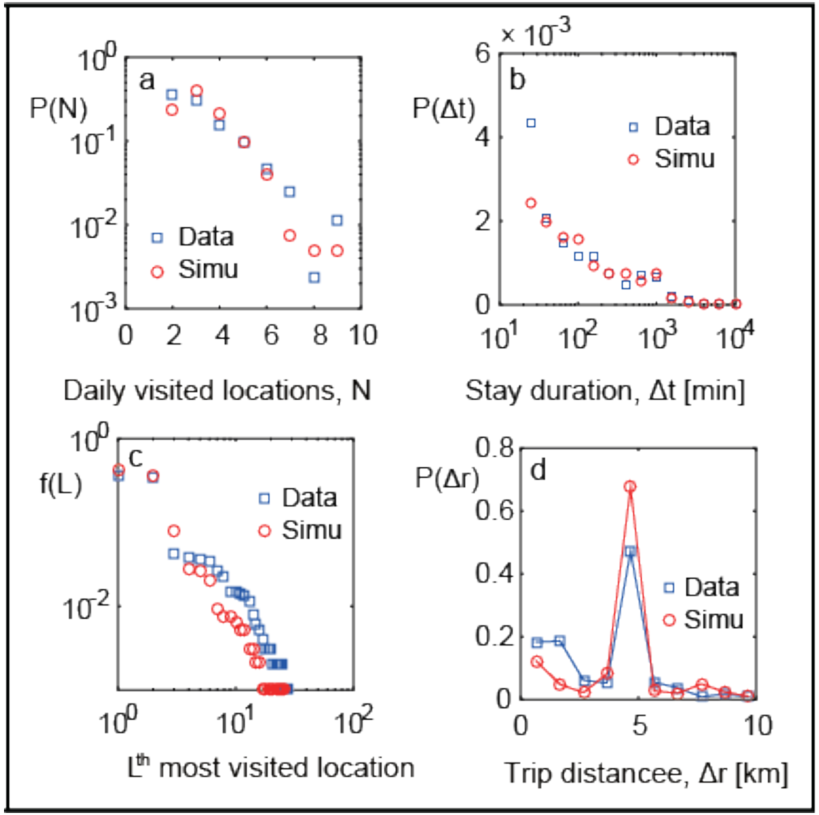
Ranking of Explorations

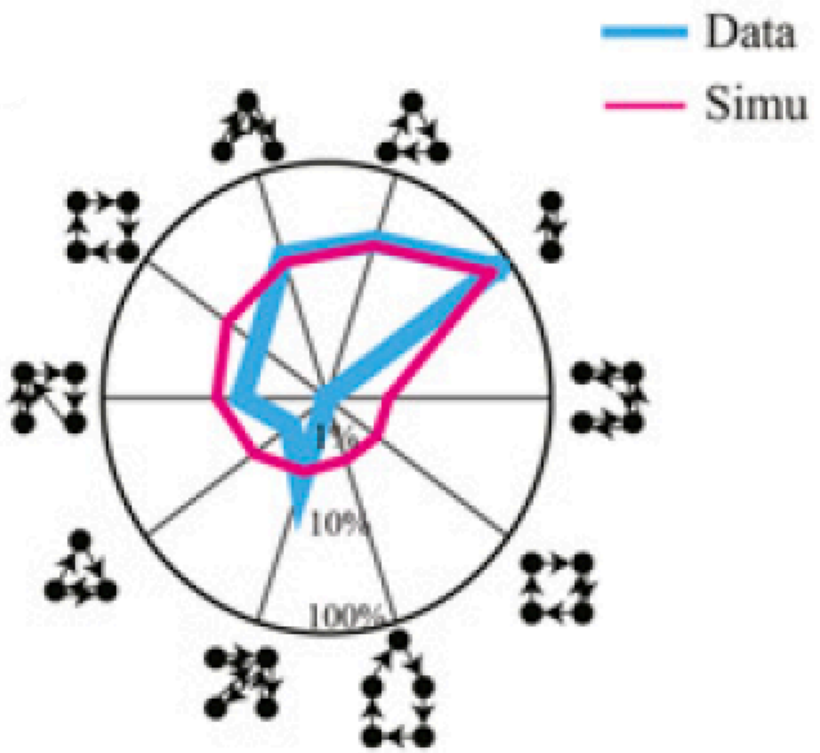


Flowchart of the Model



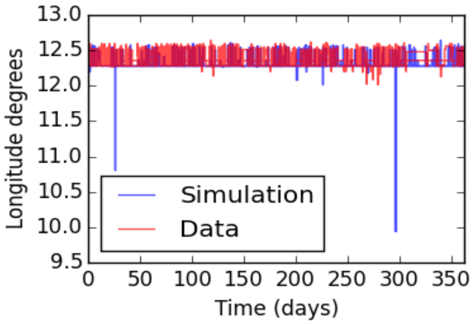
It generates good synthetic versions of each individual



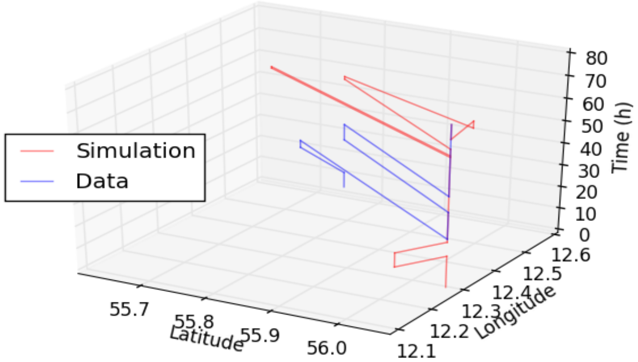


Model does not predict next location

364-day simulation

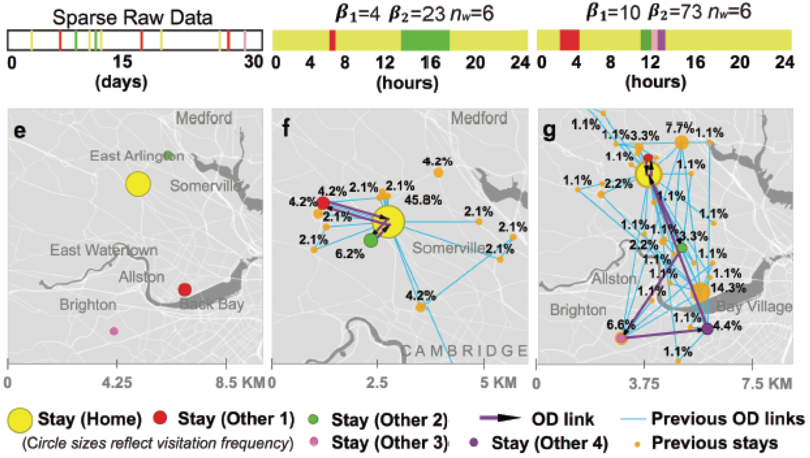


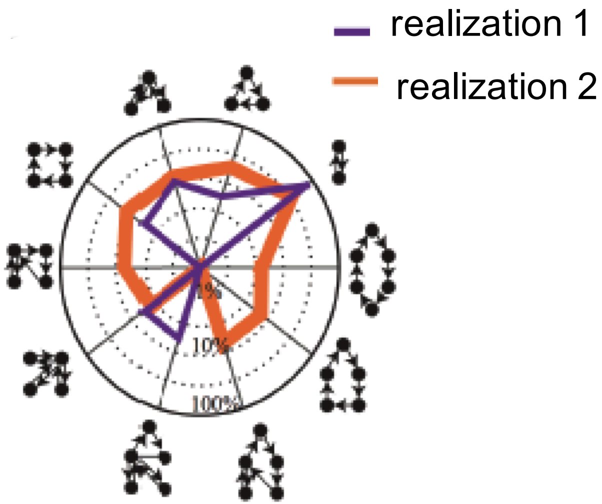
3-day simulation

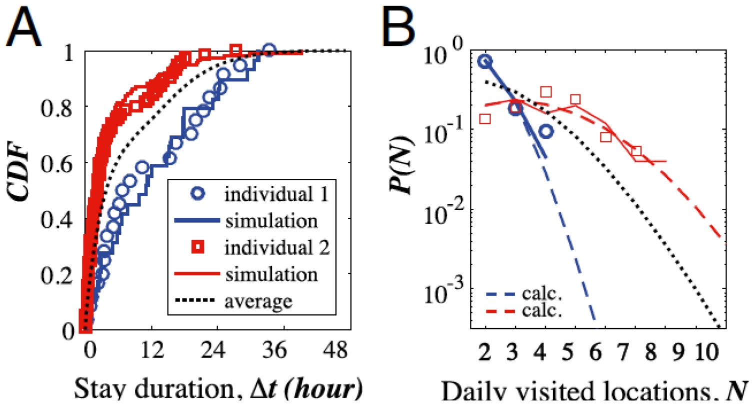


Period	Predicted locations
Day 1	58%
Day 2	70.8%
Day 3	79.8%
Overall (363 days)	46.9%

TimeGeo: From sparse user to synthetic trajectories







Adjusting parameters we generate various types of daily Profiles, keeping their home, work and frequent returns.

Using Sparse Digital Traces to Fill in Individual Level Mobility Timelines

Nabeel Abdur Rehman, Kunal Relia, Rumi Chunara

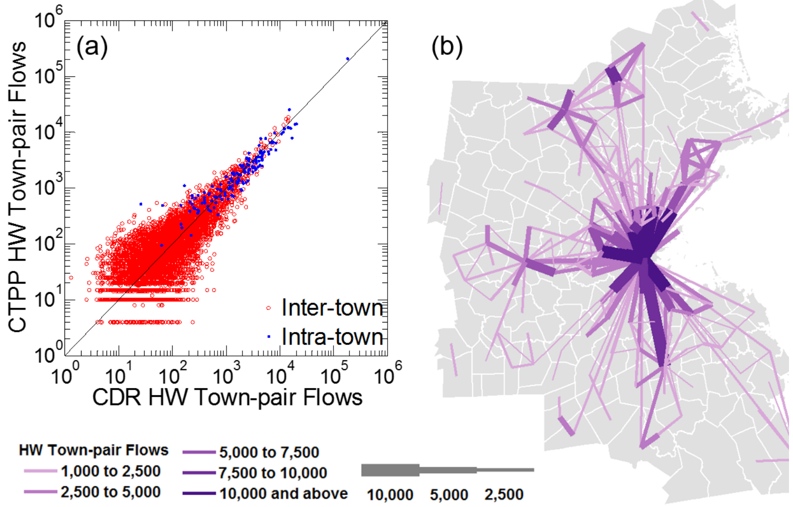
(Submitted on 6 Oct 2017)

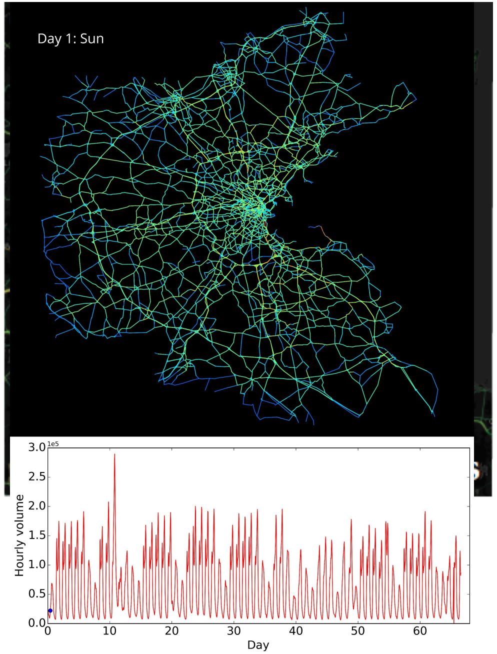
UBICOMP 2017

NetMob

UrbComp 2017 : UrbComp 2017 : The 6th International Workshop on Urban Computing (in conjunction with KDD'17)

Validated Travel Demand





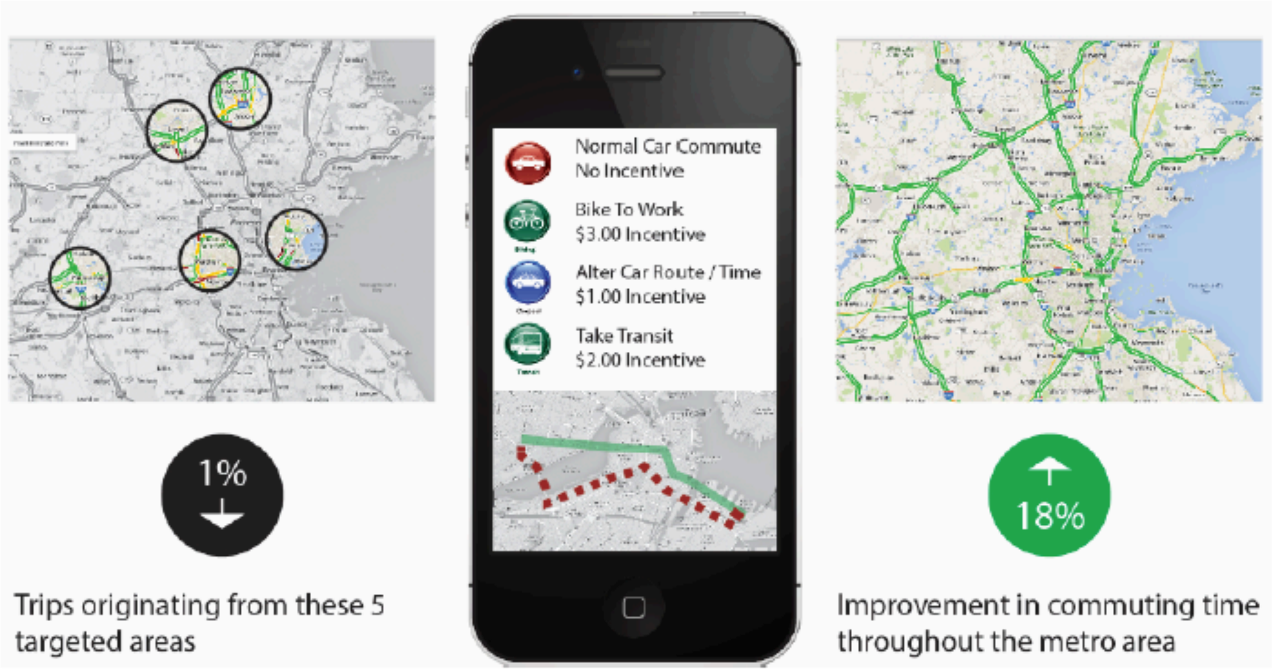
Contents lists available at ScienceDirect
Transportation Research Part C
journal homepage: www.elsevier.com/locate/trc

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

Smart Commute

A program that incentivizes drivers to avoid bottlenecks by choosing a different route or mode of transportation for their commute.



Trips originating from these 5 targeted areas

Improvement in commuting time throughout the metro area

Demand Management for Large Events

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

Rio Olympic Games

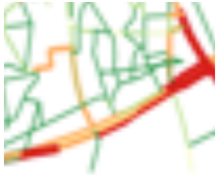
- Rio population: 6.4 millions
- International Olympic Committee (IOC) predicts 480,000 tourists in Rio for 2016 Olympics, that's about 7.5% of Rio population.
- How to evaluate the impact of Olympics to the travel of local population?
- How to manage the demand?

Data from Companies

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

On-line

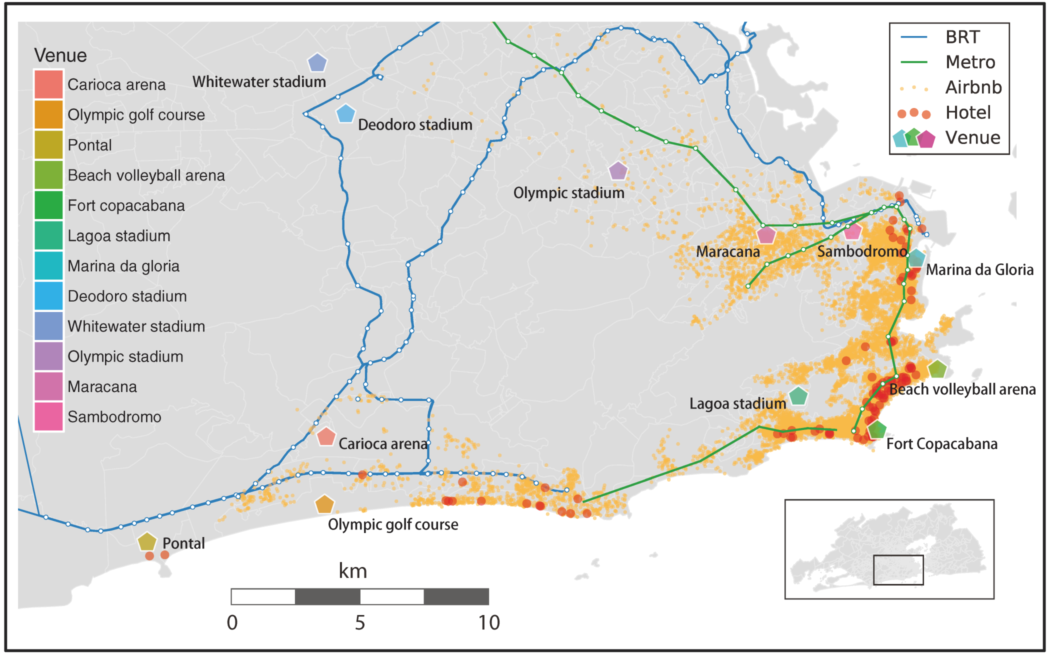
- Airbnb Supply
- GIS (OSM road network of Rio)

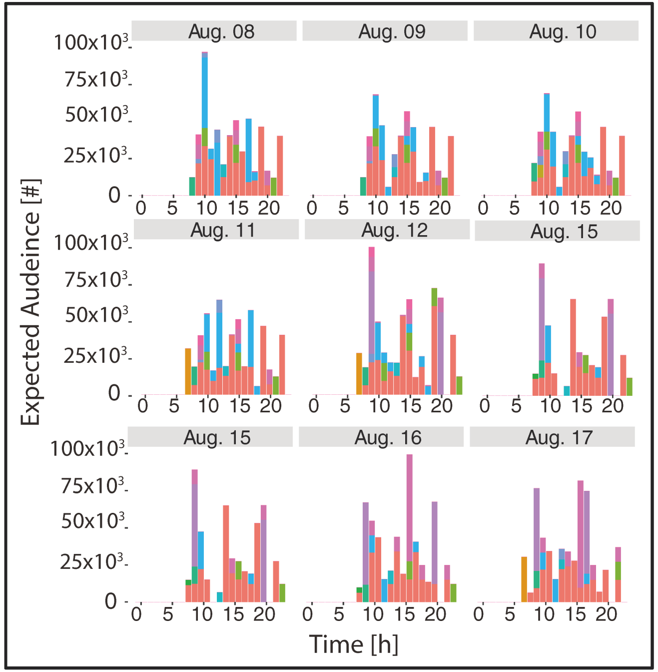


Data from Government

- Hotel, Venues and Schedules
- Camera Data 

Venues, Airbnb, hotels, BRT & Metro

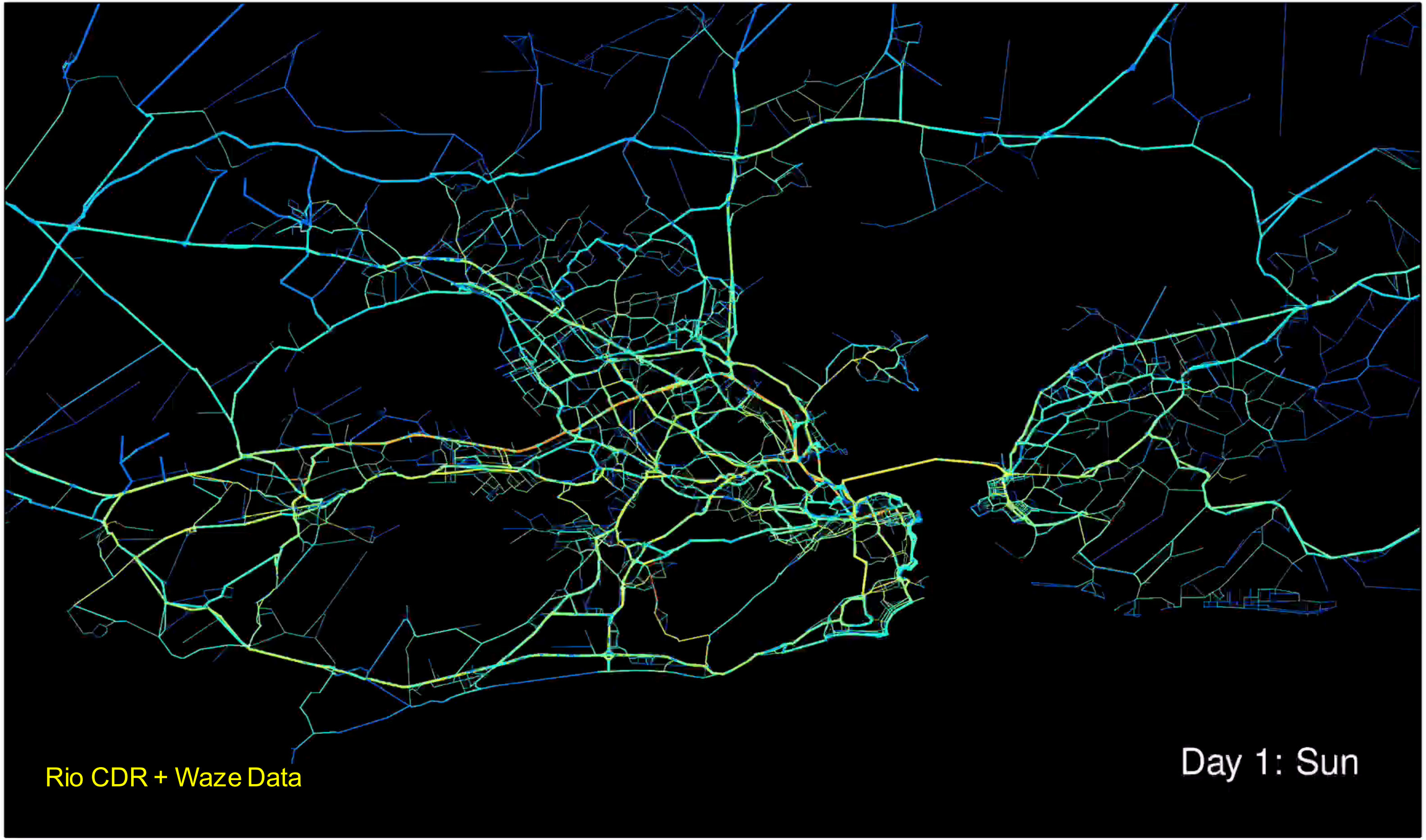




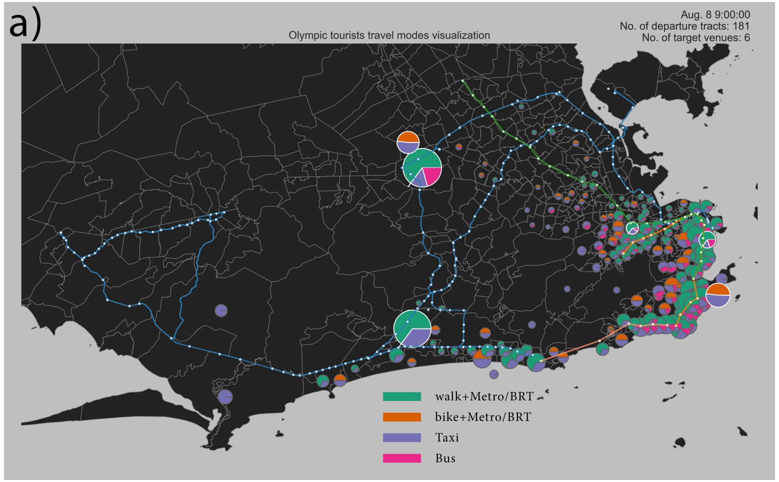
(a) Data Integrated

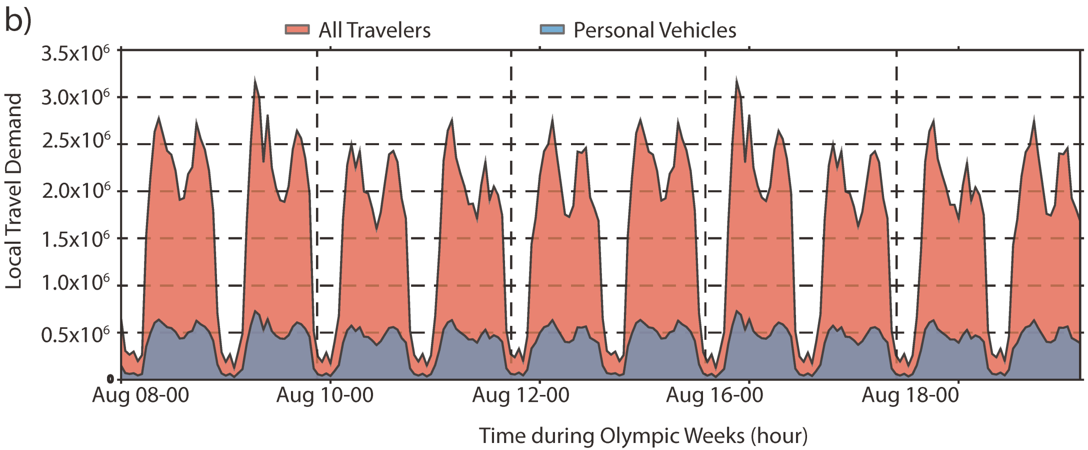
(b) Number of audiences arrive venues and when? (used data: Olympics schedule, capacity of venues)

Traffic Model



Travel demand prediction during Olympics





(a) Tourist travel mode split

(b) Total travel demand during Olympics: 2.8 million person / 0.7 million vehicles in peak hour.

Smart-app (routing)

Modifications on the level of altruism:

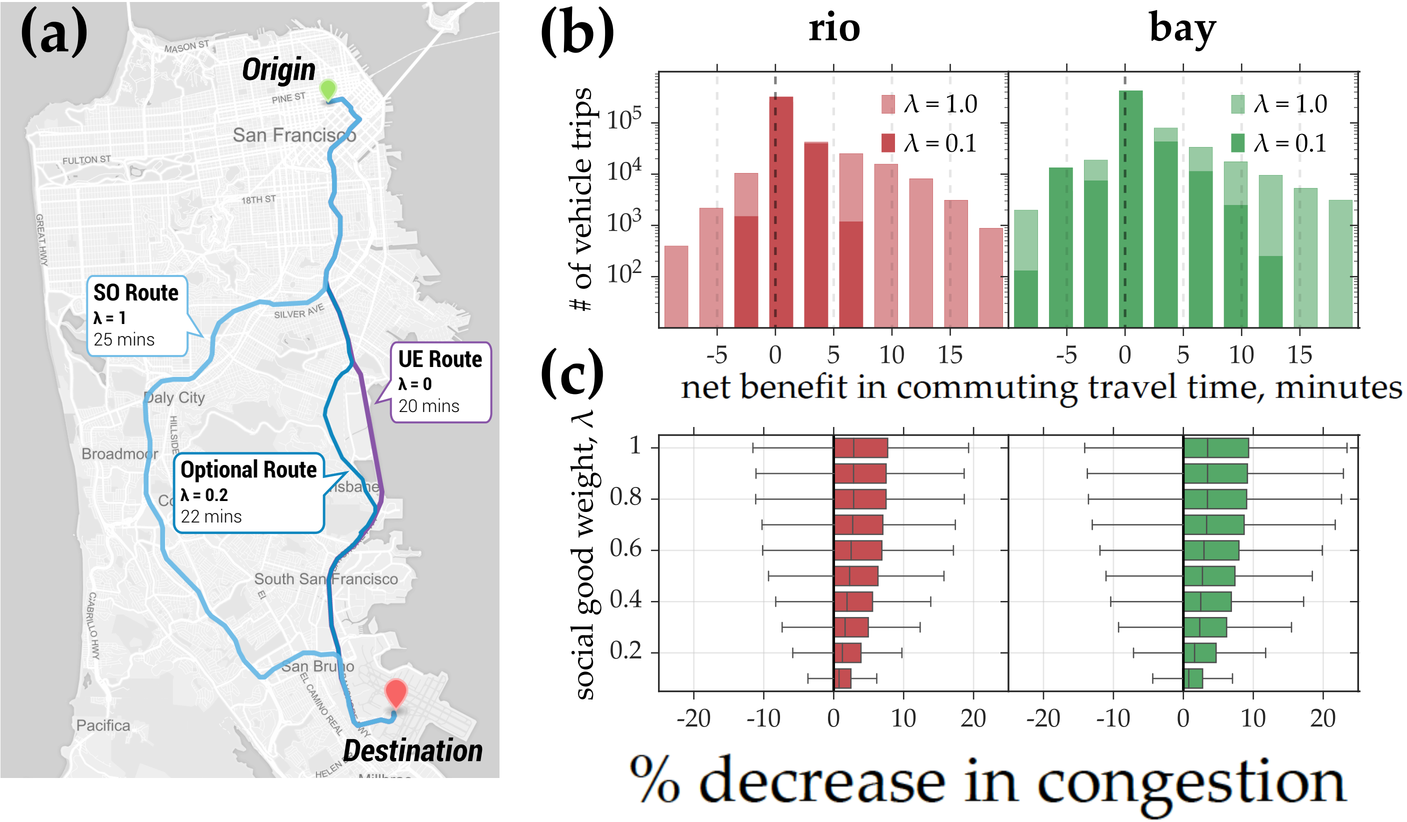
$$c_e^{\lambda}(x_e) = (1-\lambda)t_e(x_e) + \lambda \frac{d\left[x_e t_e(x_e)\right]}{dx_e}$$

$\lambda=[0..1]$

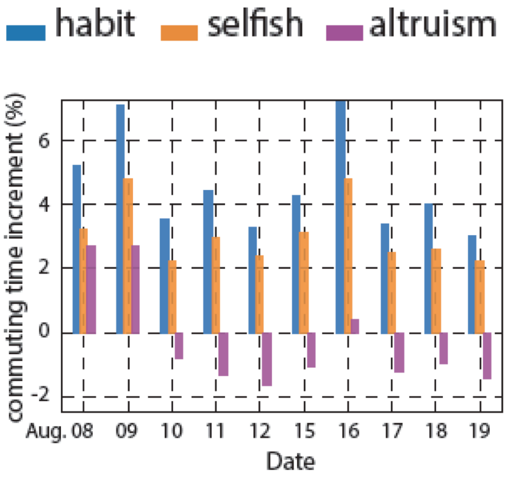
User Equilibrium component

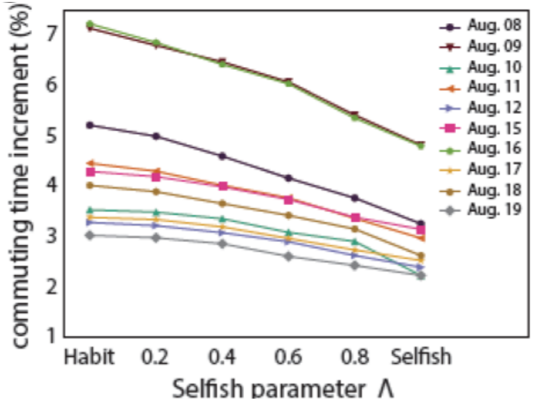
Social Optimun component

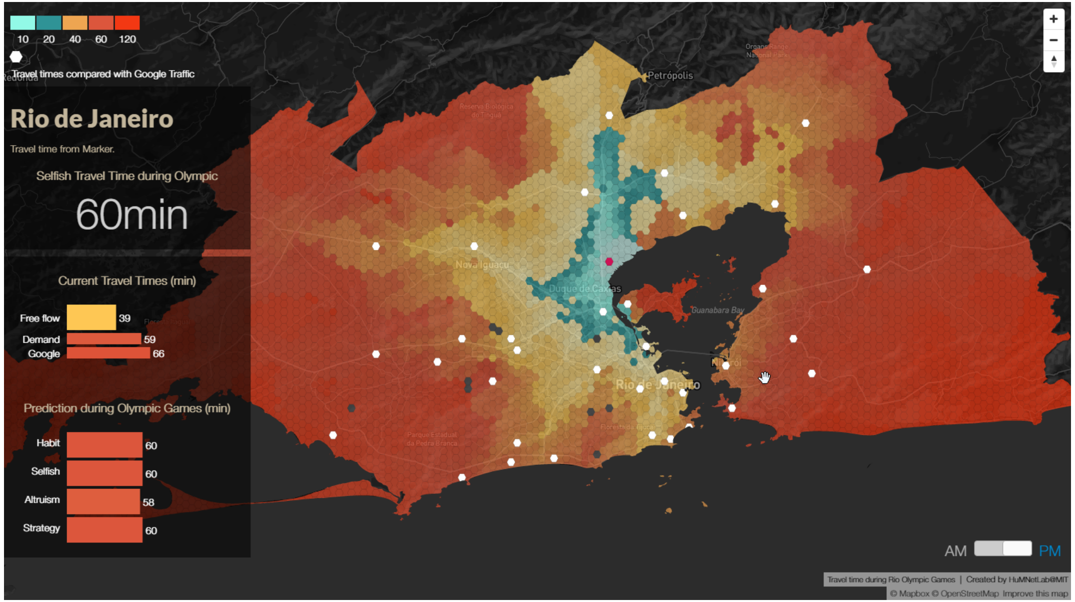
Smart-app (routing)



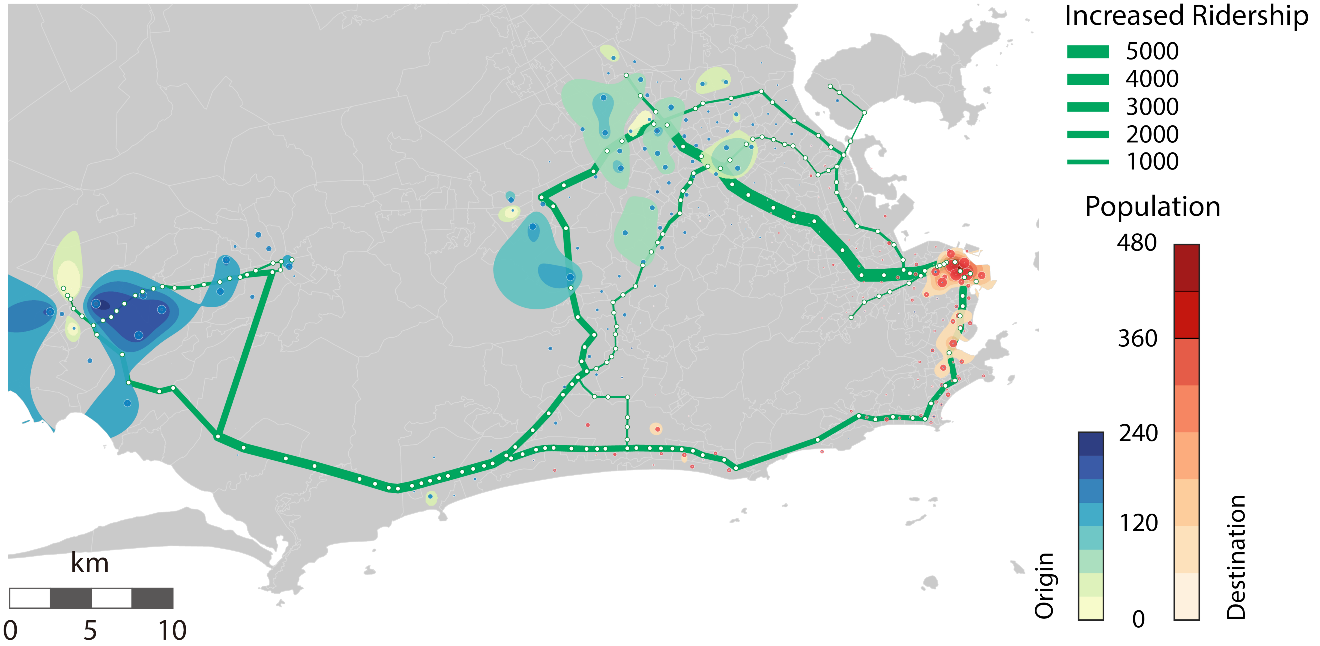
Travel time estimates before and during the Olympic Games

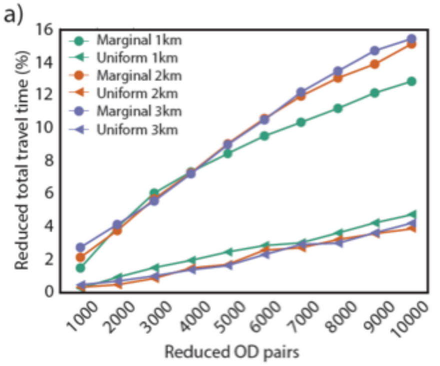


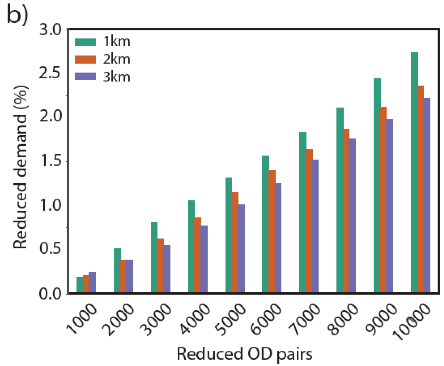




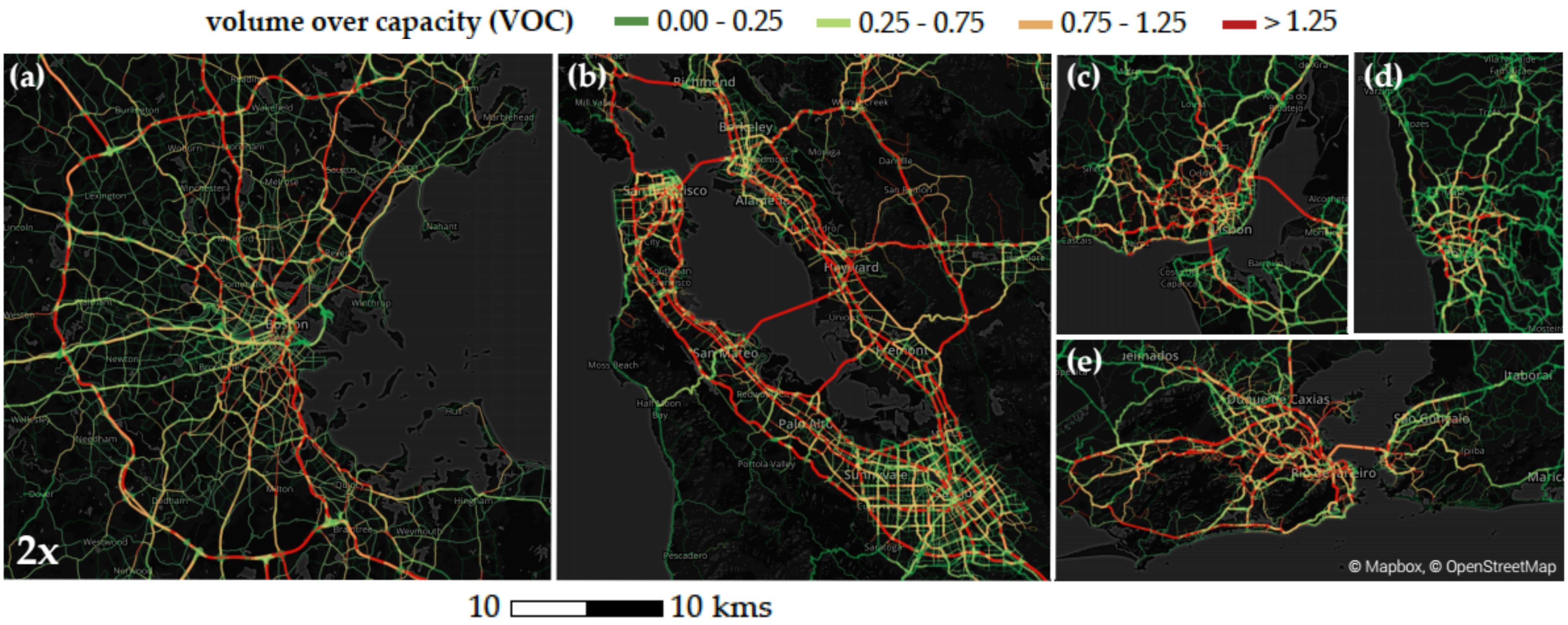
Recommendations of Car reduction per Origin and Destination







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



(a) Boston
(b) San Francisco Bay Area

(c) Lisbon
(d) Porto

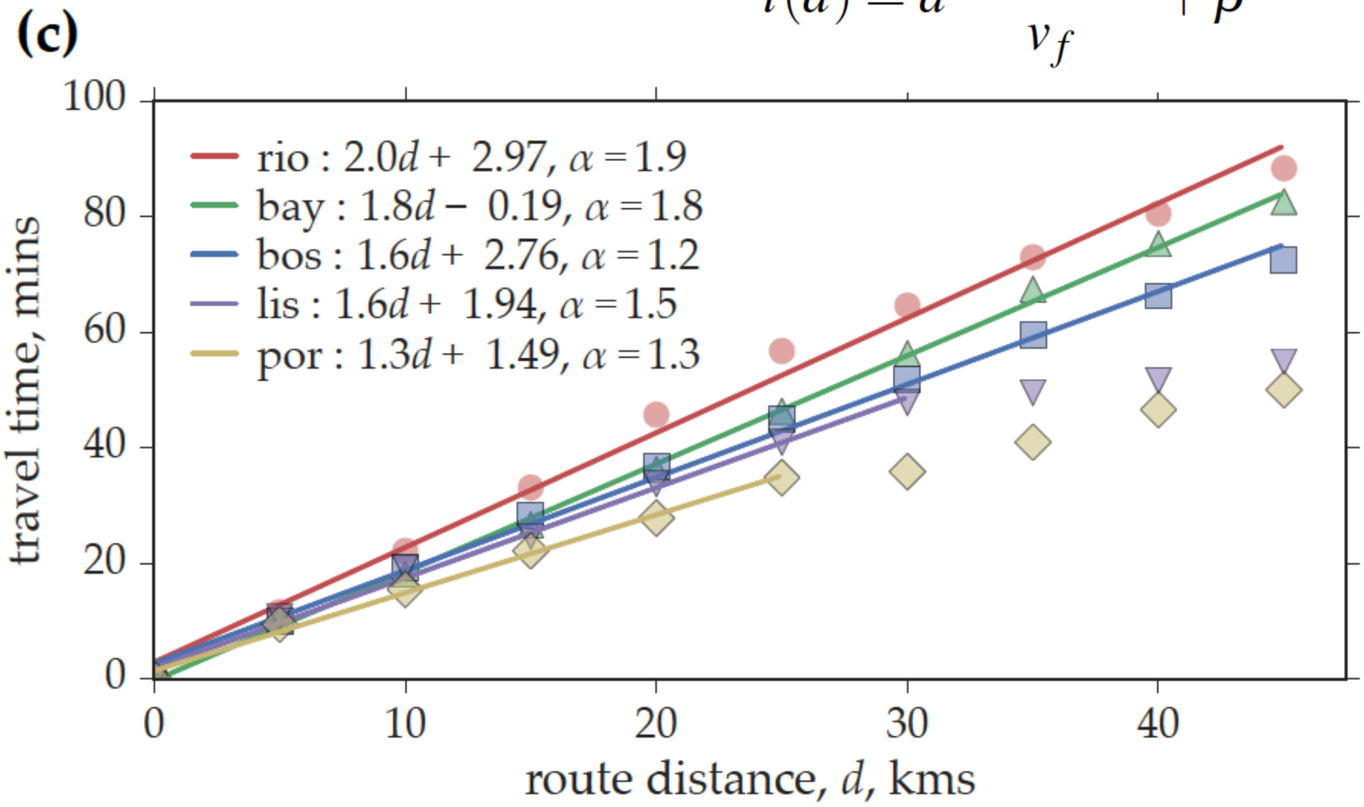
(e) Rio de Janeiro

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

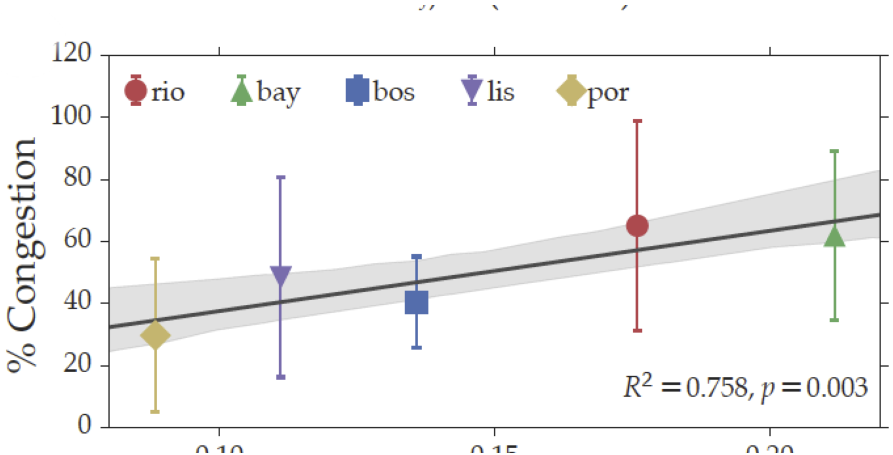
Open Street Map data for the Streets when
not better data is available

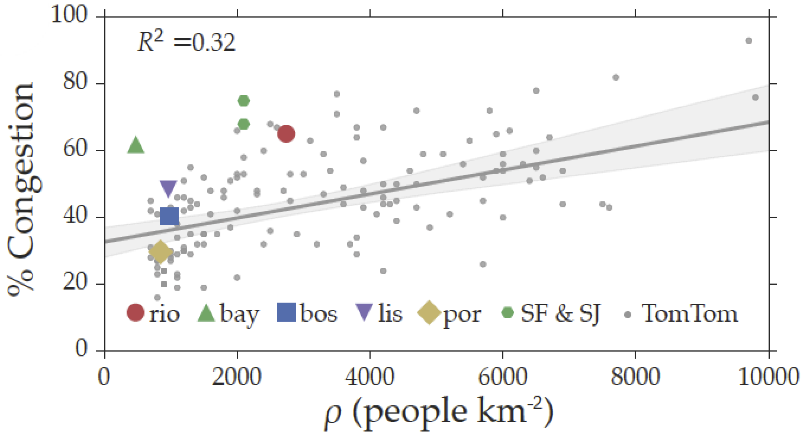
Commuting Time

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



Understanding Observations





Demand to Supply Ratio

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

$x_e =$ flow on the road link e (veh/h)

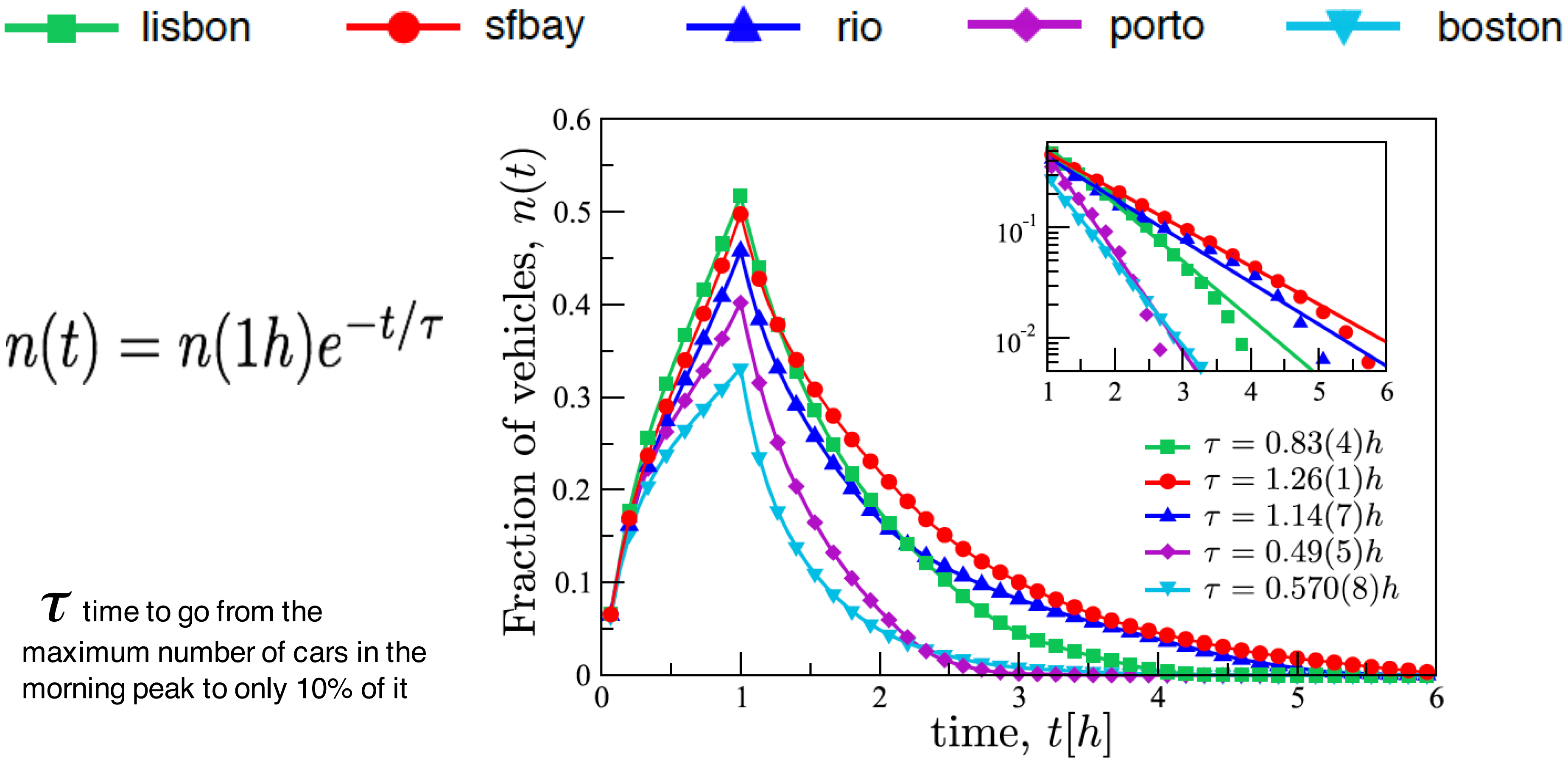
$\ell_e =$ road link length e (km)

$C_e =$ capacity of the road link e (veh/h)

	City				
	Boston	Porto	Lisbon	Rio	SF Bay
Γ	0.129	0.101	0.121	0.180	0.213
$t_{ff}[h]$	0.184	0.188	0.267	0.218	0.234

S.Colak et al. Nat. Comm. 10793, 2016

Analysis



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

τ time to go from the maximum number of cars in the morning peak to only 10% of it

Informing Planning Decisions

Median free-flow travel time

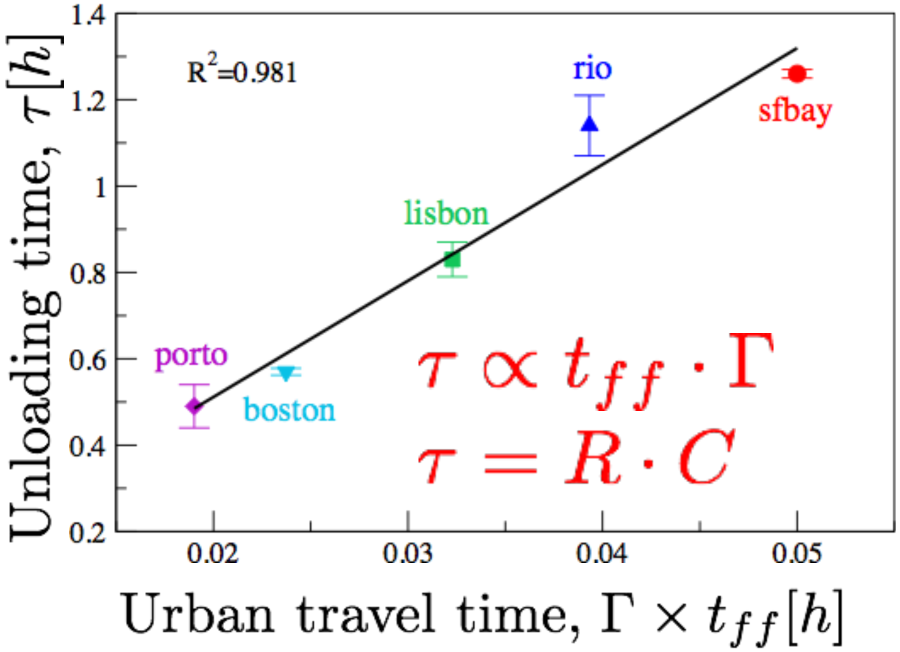
$$t_{ff}[h]$$

"Resistance"

Demand to Supply Ratio

"Capacitance"

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



L. Olmons, S.Colak, M.C Gonzalez, Nat. Comm. (under review)

Identifying Lifestyles with card transactions

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

Ciudad de México

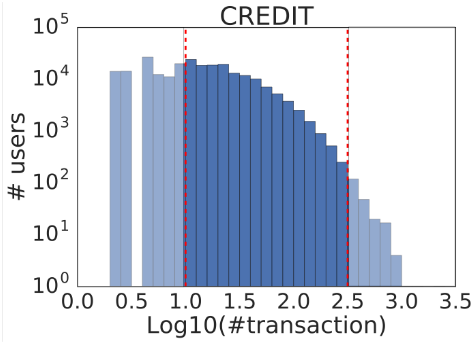
Coupling Credit Card Data with Mobile phone Data

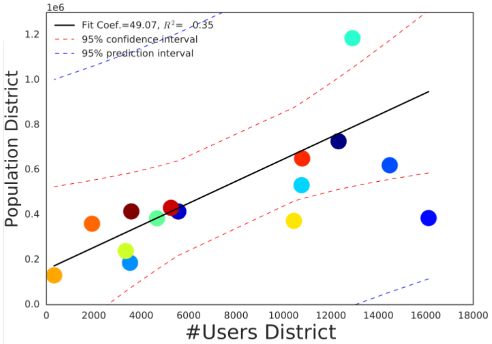
Mexico City

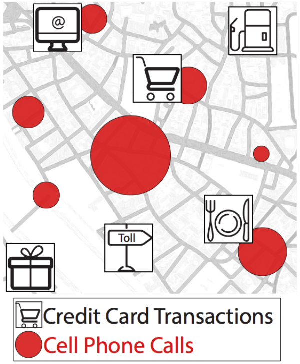
150,000 users

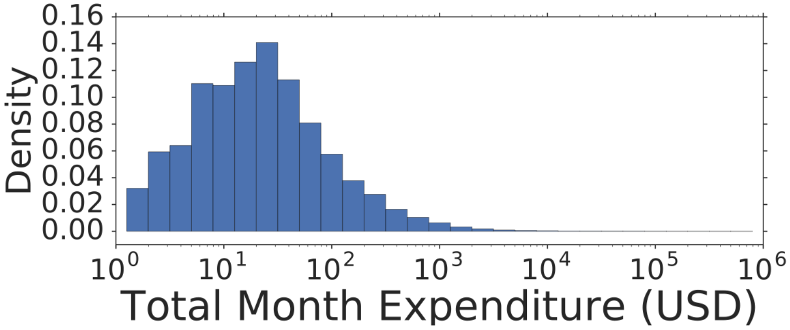
- Age
- Gender
- Zipcode
- Mobile Phone data

10 weeks period



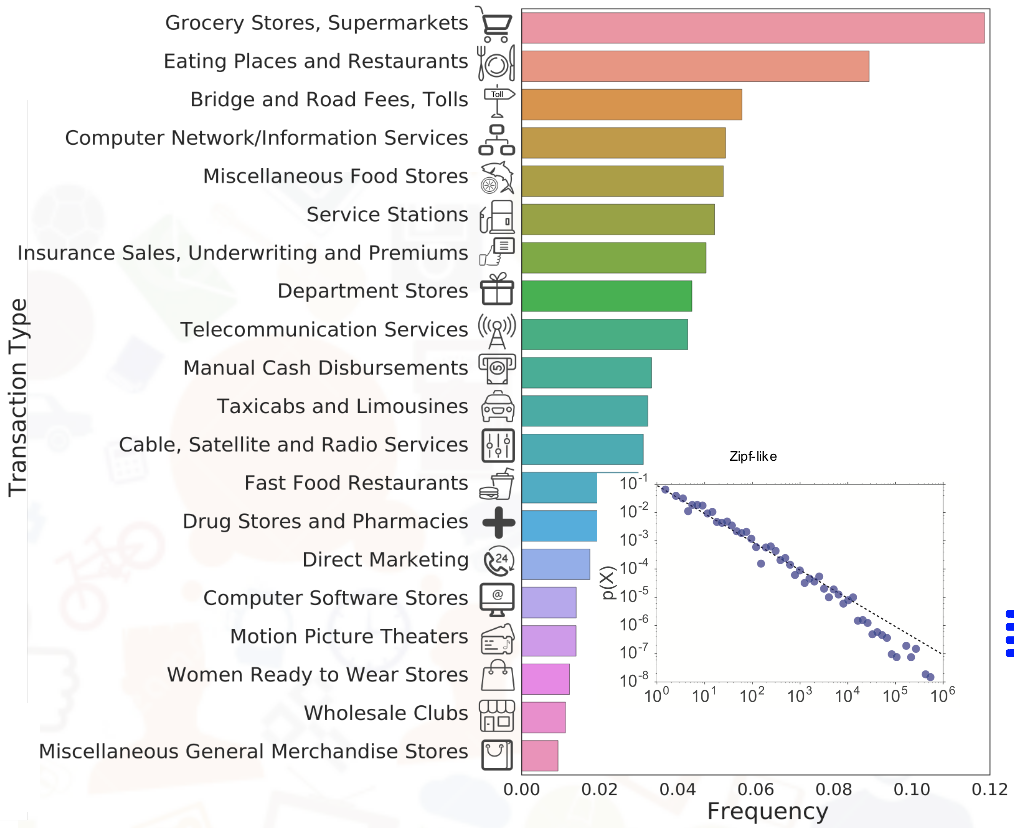




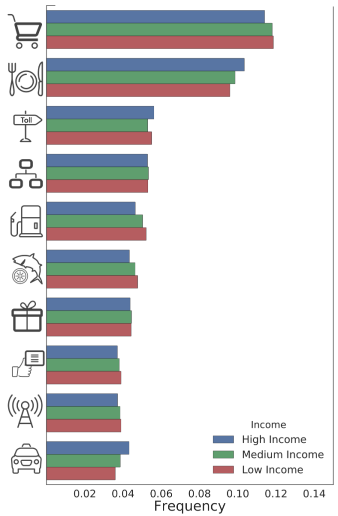


Mobility Diversity | Social Diversity | Radius of Gyration

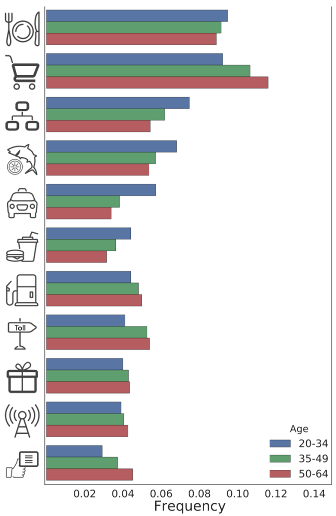
Distribution of Transactions



Income



Age



Affinity Algorithm

Latent Dirichlet allocation

TD-IDF

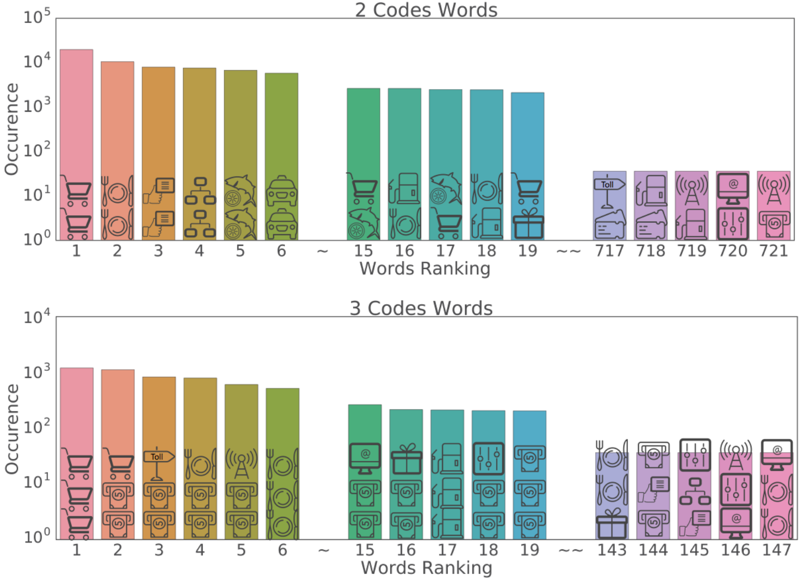
(term frequency-inverse document frequency)

Text analysis via the Sequitur algorithm

Sequitur is a recursive algorithm developed by Craig Nevill-Manning and Ian H. Witten in 1997 that infers a hierarchical structure from a sequence of discrete symbols.

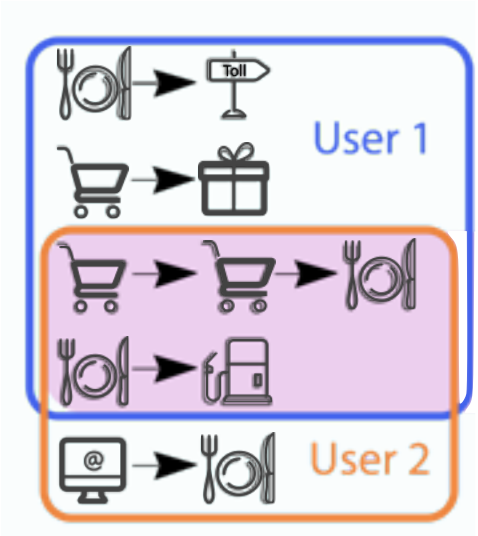
Compressed Sequence

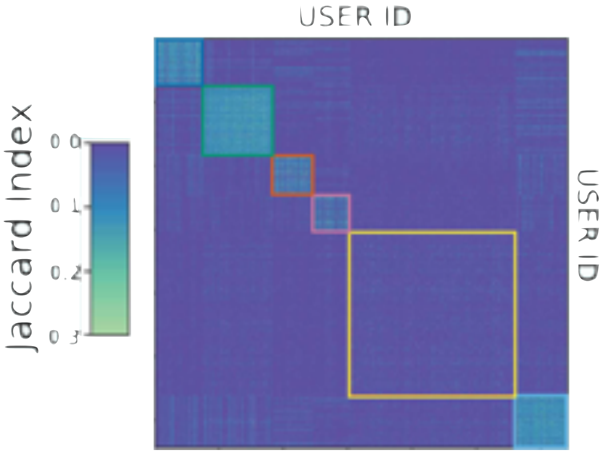
Words as Ordered Sequence of transactions

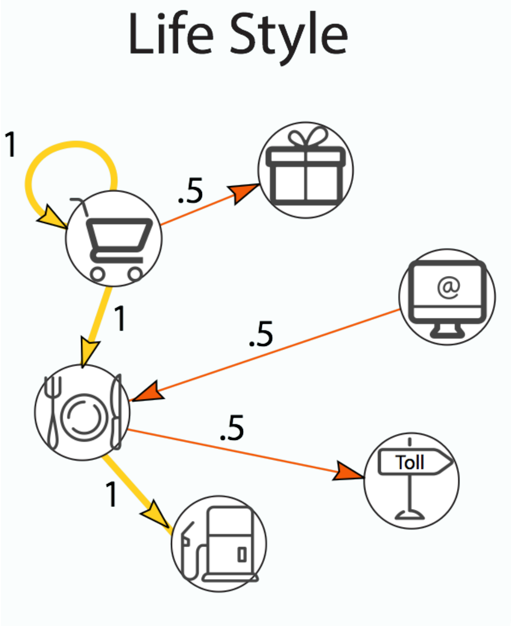


1. Algorithm from text analysis to infer the most common sequence of transaction → 2. Null model to select the most significative transaction → 3. Network Theory to cluster the users

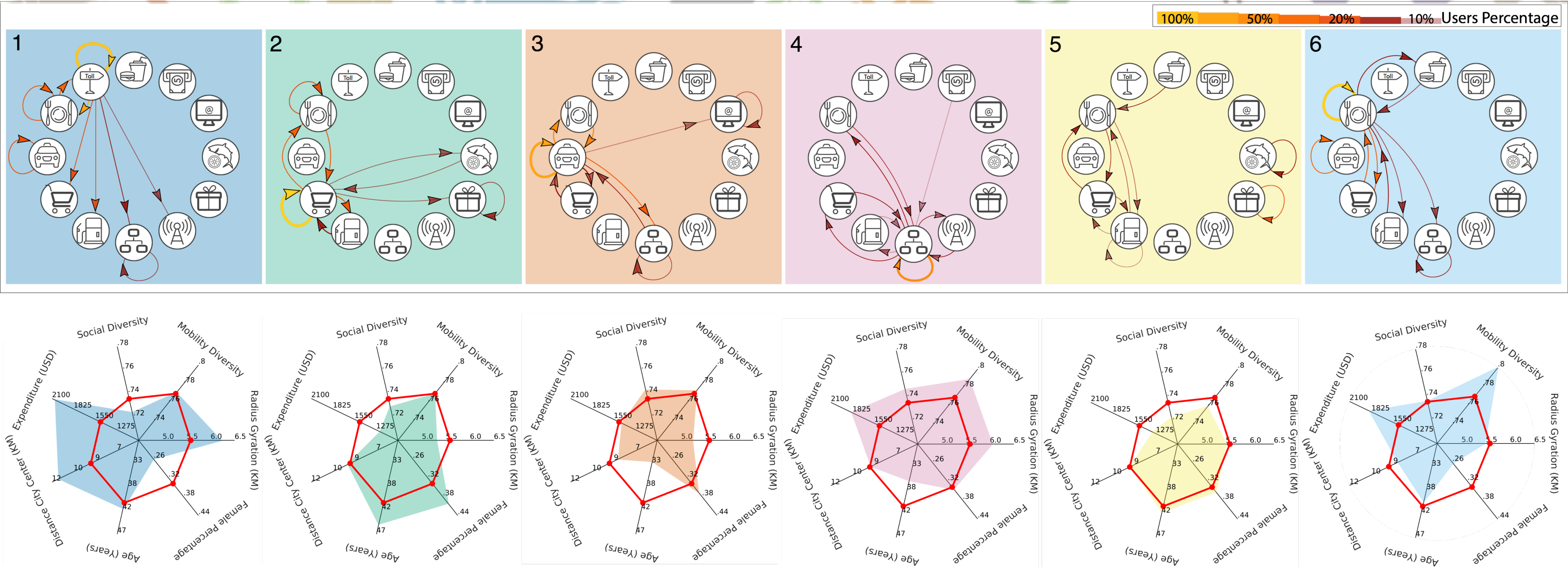
We cluster users







Lifestyles



Commuters

Households

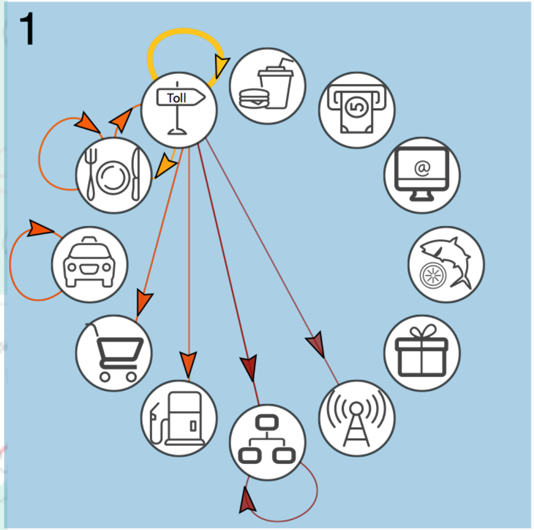
Youth 1

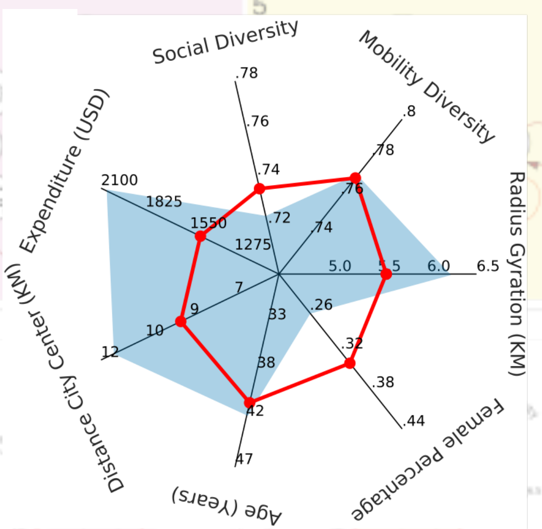
Tech users

Average

Dinners

Lifestyle

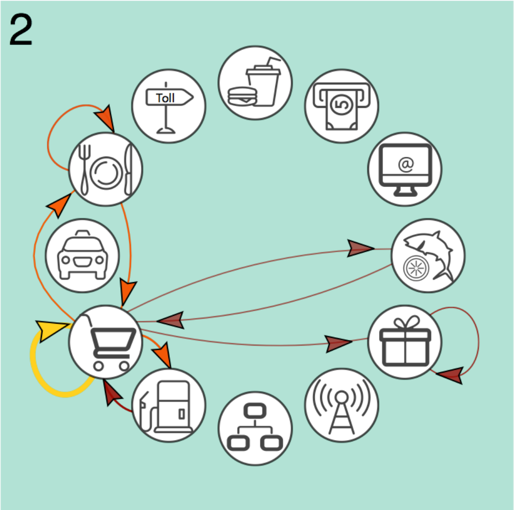


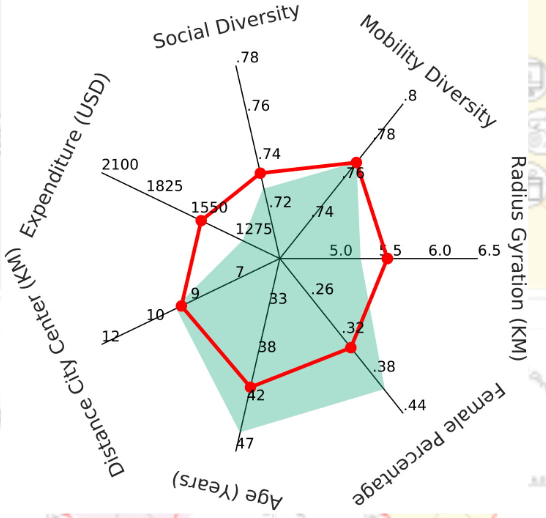


Commuters:

High expenditure, Living far from the city center,
High radius of gyration, Low social diversity.

Lifestyle



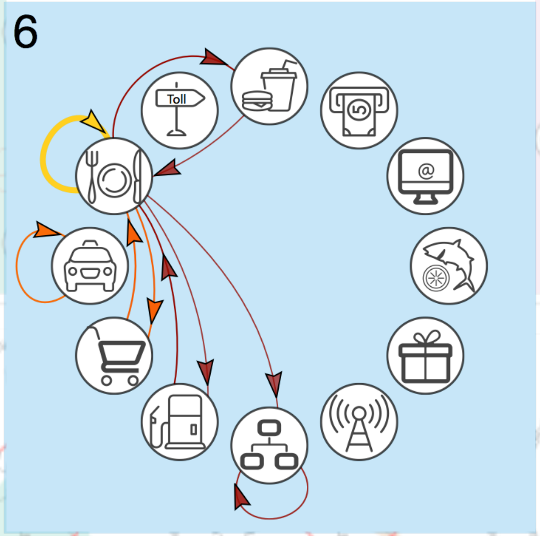


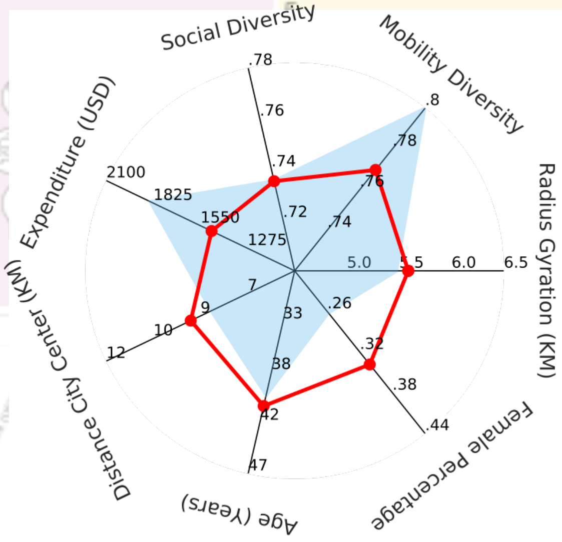
Households:

Low Expenditure, Older Age, Low Social Mobility, High Female Component

Commuting | Households | Younger 1 | Younger 2 | Average | Dinner-Out

Lifestyle





Dinners:

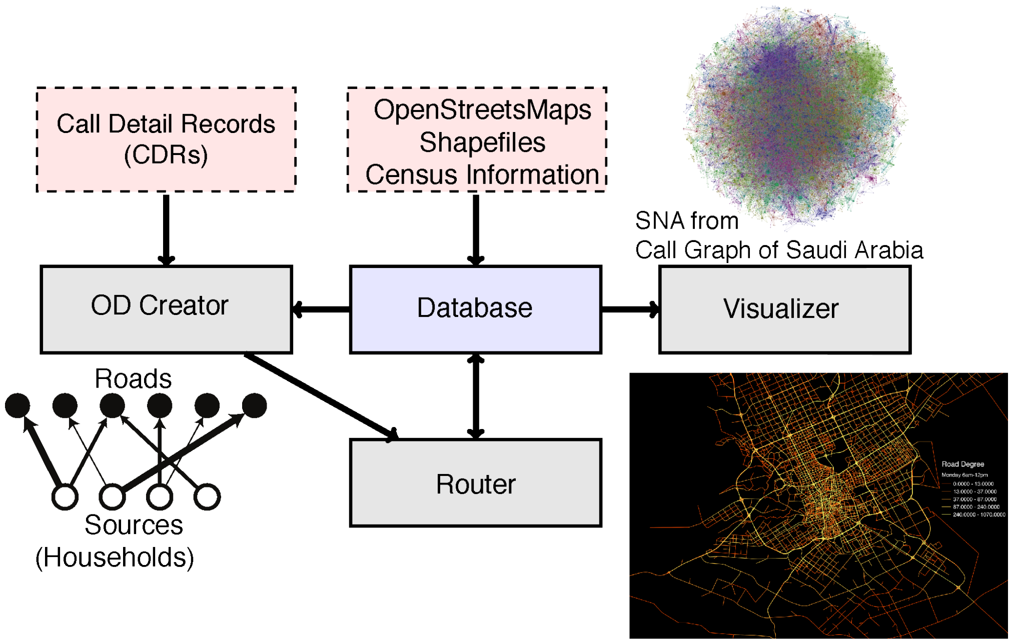
High Expenditure, Avr. Social Diversity
and High Mobility,

Commuting

Housel

Dinner-Out

Towards Computational Urban Science



Riyadh road attractors

- Today we can measure behavior from digital traces and could use ICT to better plan cities with them.
- Finding universal patterns is the first step, looking for mechanisms and heterogeneities come next.
- Complex system approaches give us an unique opportunity to find the way for that.

Thank you!

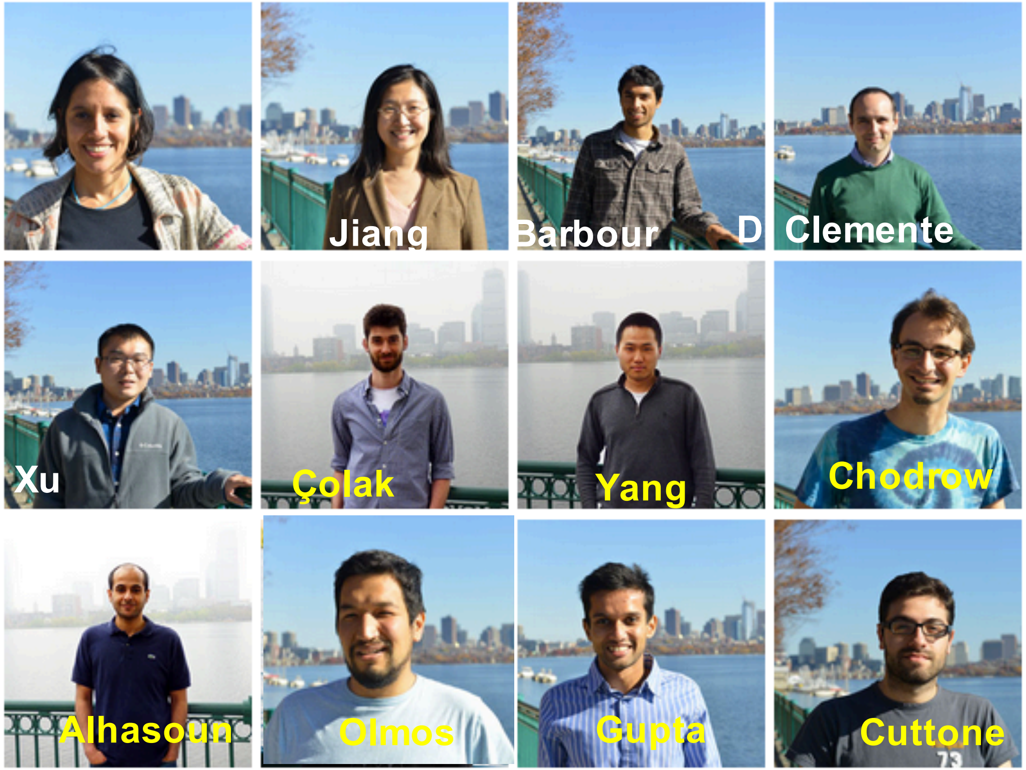
Berkeley
UNIVERSITY OF CALIFORNIA

Massachusetts
Institute of
Technology

http://humnetlab.mit.edu

HUMNET

Questions?



Alumni: Toole, Herrera-Yague, Alexander, Lima, Sturt, Desu, Grestle

Postdoc: Schneider, Wang, Belik, Halu

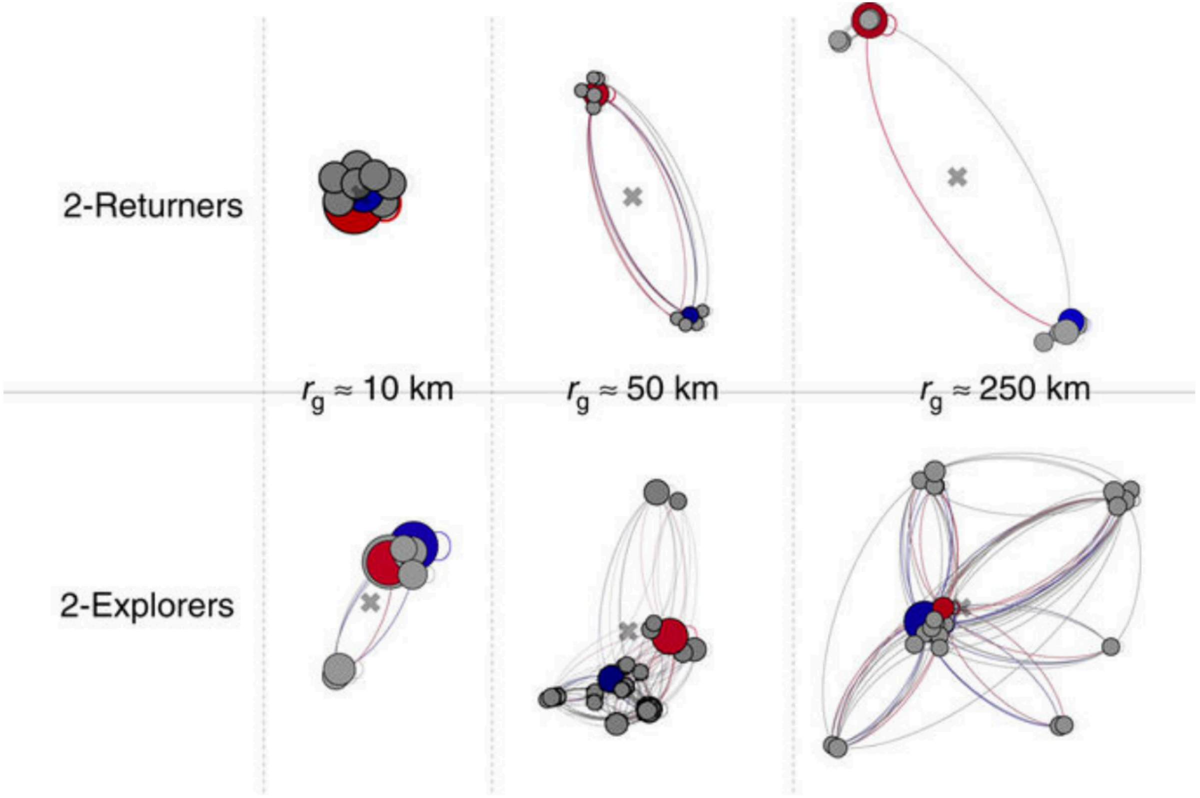
Summary

A city's congestion fingerprint is related to measurable characteristics, namely a ratio of total demand to total supply (Γ).

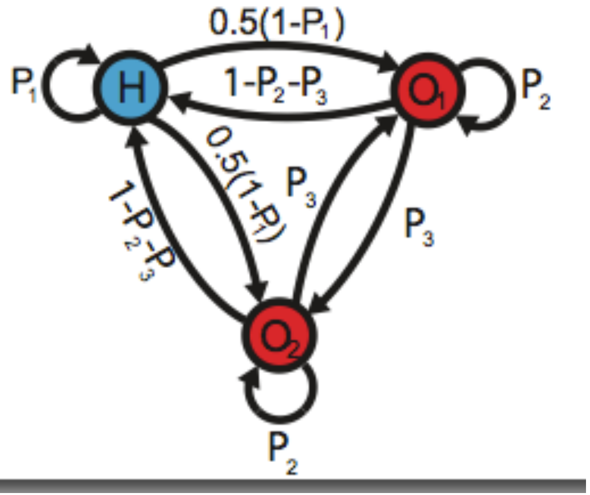
Lower λ will moderate the magnitude of benefits and losses while realizing most of possible benefits, making resulting policies fairer and easier to implement

Çolak, S., Lima, A. & Gonzalez, M. C., ***Understanding congested travel in urban areas***, accepted and to appear in *Nature Communications (2016)*

Heterogeneity 1: Returners vs. Explorers



Markov Mobility Model



rhythm(t)

$$P_1 = 1 - P(t)$$ Staying at "home"

$$P_2 = 1 - \beta_1 P(t)$$ Staying in current errand

$$P_3 = \beta_1 P(t) \beta_2 P(t)$$ Moving from current errand to a new current errand

"bursts" rate

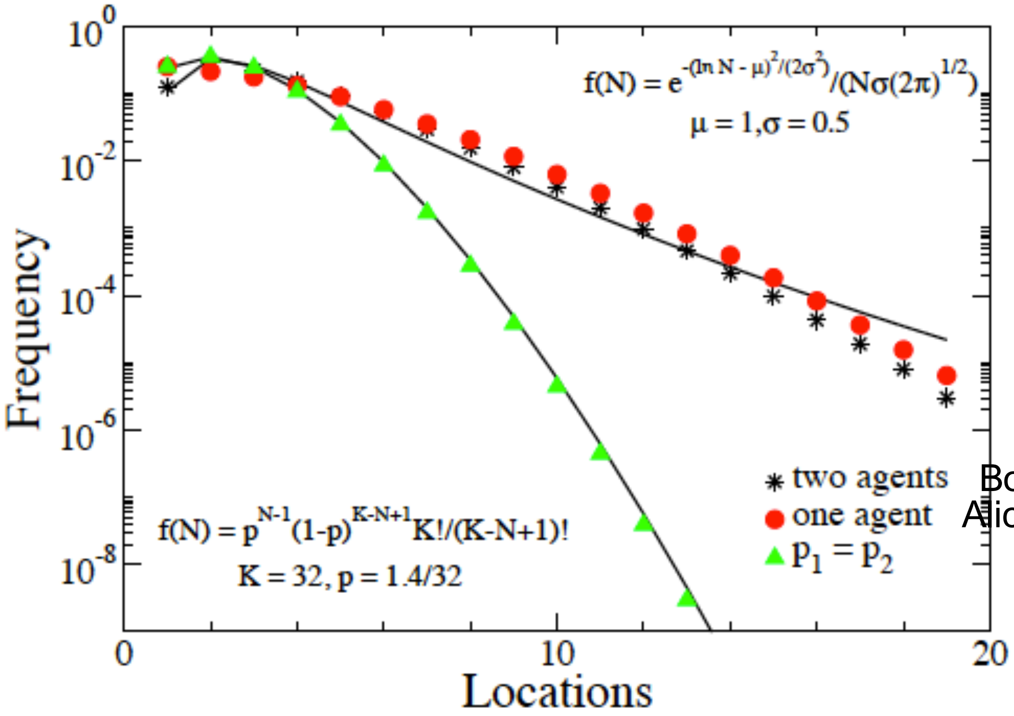
exploration rate

Generates shorter stays in the errand states.

Generates different number of activities in a row per **active** cycle

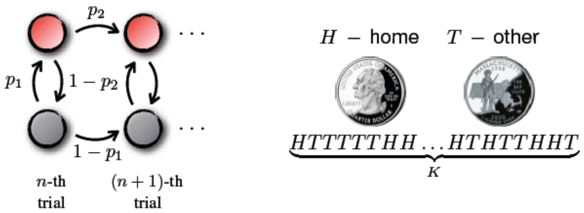
$$\frac{P(O_1 \to O_2)}{P(O_1 \to H)} = \frac{\beta_2 n_w P(t)}{1 - \beta_2 n_w P(t)}.$$

The rates explain the number of visited locations



Bob
Alice

Calculating number of visits



$$P(N = x) = \xi_0 \left(\prod_{t=1}^{K} \Lambda_t \right) U^{\mathrm{T}}(C_x), \; C_x \in \Omega \qquad (1)$$

$$\Omega = \{(x, i);\; x = 0, \ldots, K-1;\; i = 0, 1\} \qquad (2)$$

J. C. Fu and M. V. Koutras, Distribution Theory of Runs: A Markov Chain Approach *J. Amer. Stat. Assoc.* **89** 1050 (1994)