From Individual Mobility to Urban Traffic

Marta C. Gonzalez

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

Civil and Environmental Engineering

Massachusetts Institute of Technology

The Science of Cities

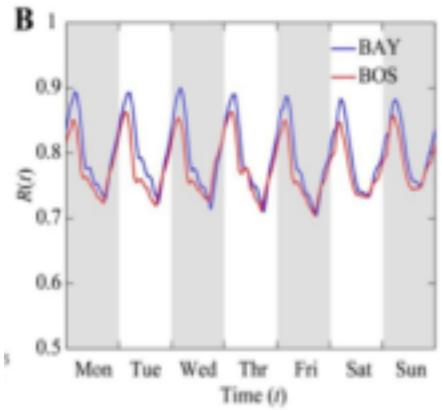
- And so a growing number of people have begun, gradually, to think of cities as problems in organized complexity -- organisms that are replete with unexamined, but obviously intricately interconnected, and surely understandable, relationships. -- **Jane Jacobs *The Life and Death of Great American Cities***

http://i.imgur.com/HhmvaQQ.jpg

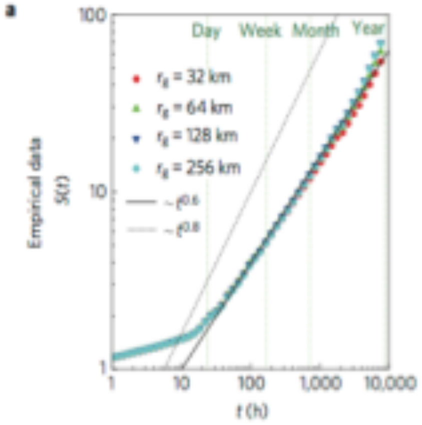
Big Data and Human Behavior

Big Data and Human Mobility

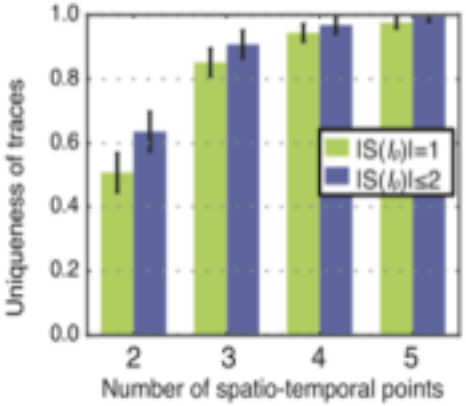
Predictable



Slow Explores



Unique



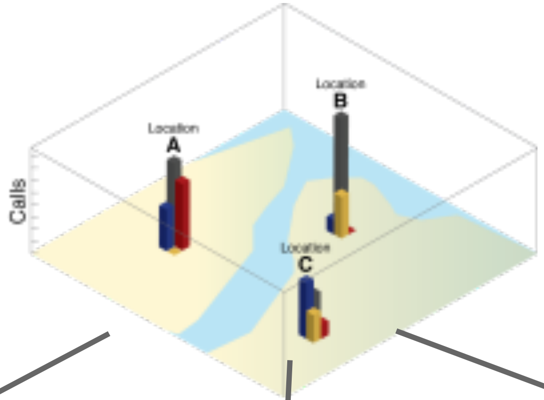
Wang, Pu, et al. "Understanding road usage patterns in urban areas." *Scientific reports* 2 (2012).

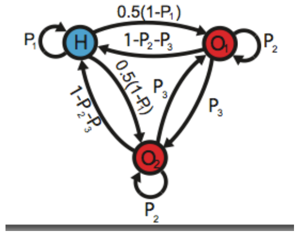
Song, Chaoming, et al. "Modelling the scaling properties of human mobility."*Nature Physics* 6.10 (2010): 818-823.

de Montjoye, Yves-Alexandre, et al. "Unique in the Crowd: The privacy bounds of human mobility." *Scientific reports* 3 (2013).

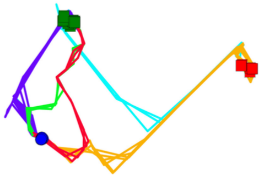
Big Data and Human Behavior

Human Mobility

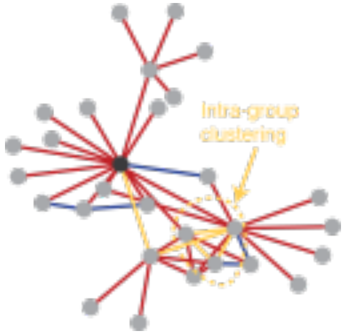




TimeGeo



Routing Behavior



Social Behavior

Overarching Goal

Opportunities:

Massive spatiotemporal information

— **millions of individuals in a given metro area**

— **long time period of observation (in months).**

Obstacles:

Massive, and passive data with lots of noise

— **anonymity of individuals**

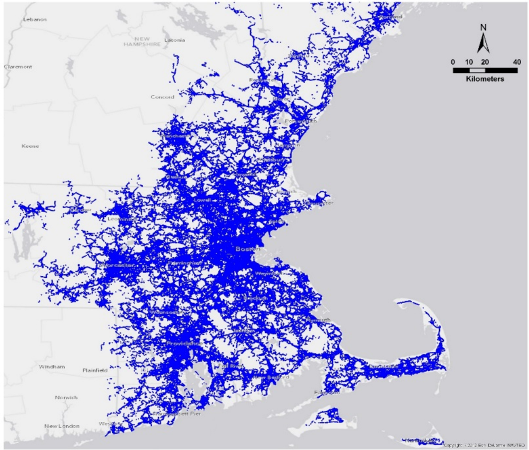
— **missing information**

— **no social demographic characteristics**

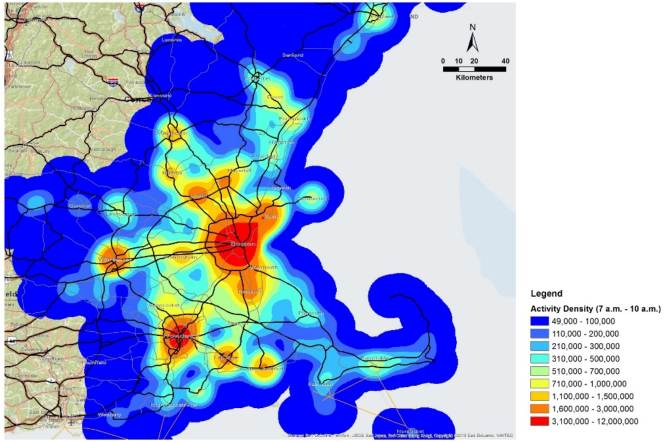
— **potentially biased sample**

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?



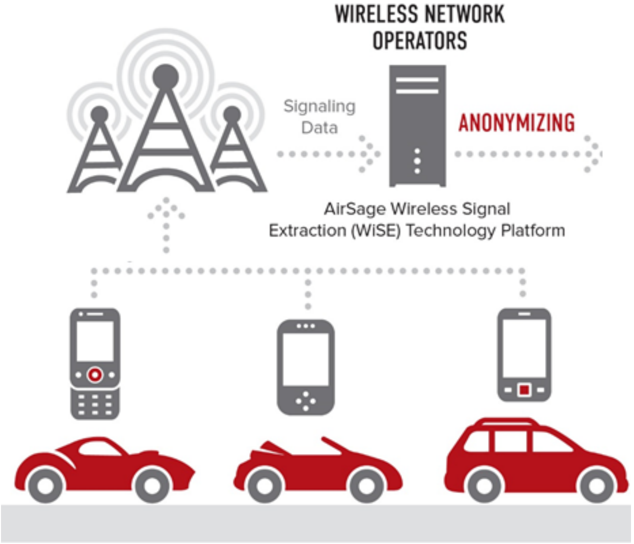
1.9 million total users observed in the 2 months, in Boston 2010.



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

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

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

Travel Survey

$200 per usable Survey

1 sample day,

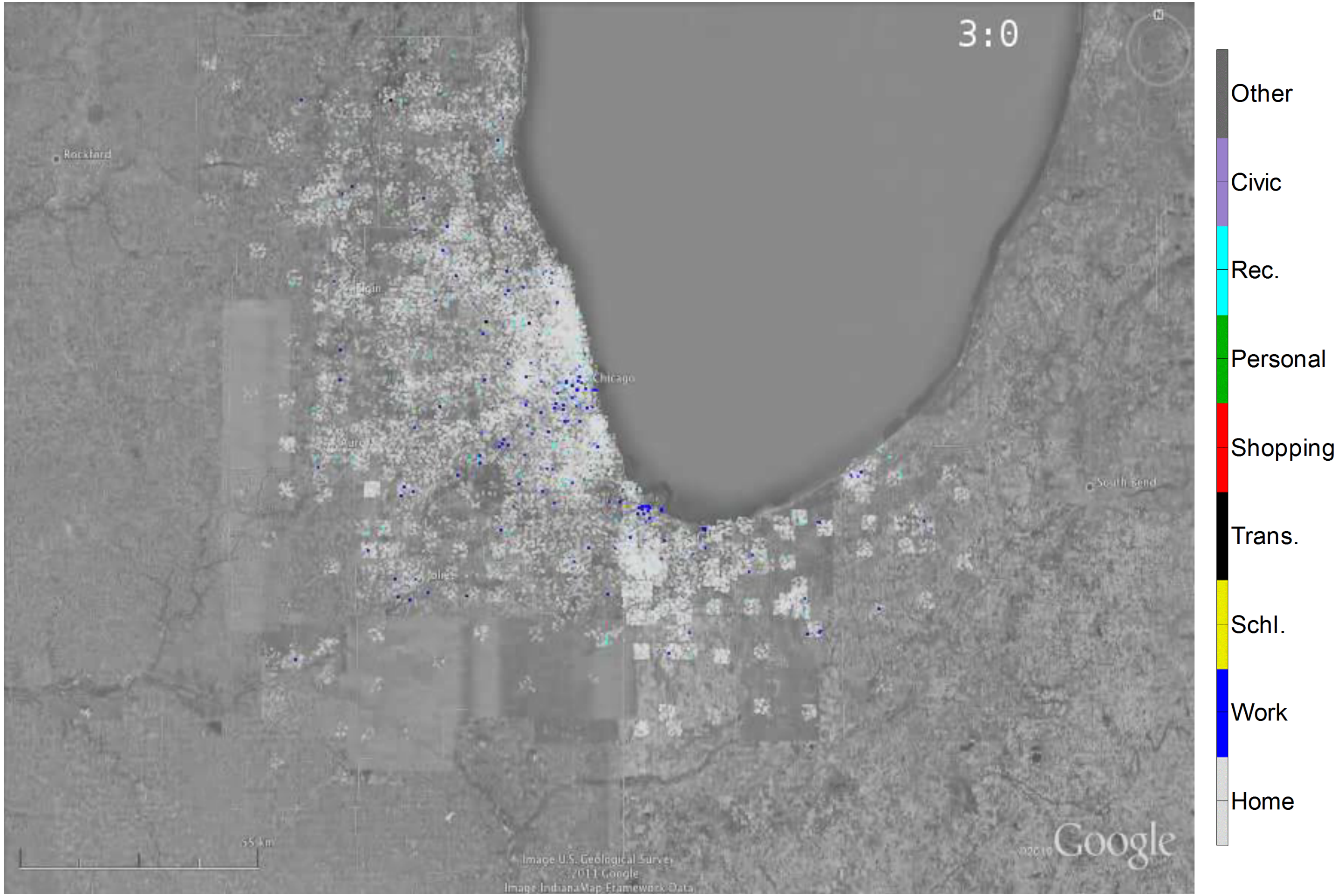
2.5×10^4 households out of 2.6×10^6

58% response rate.

(3.7 calls and 17 minutes per survey)

MASSACHUSETTS TRAVEL SURVEY

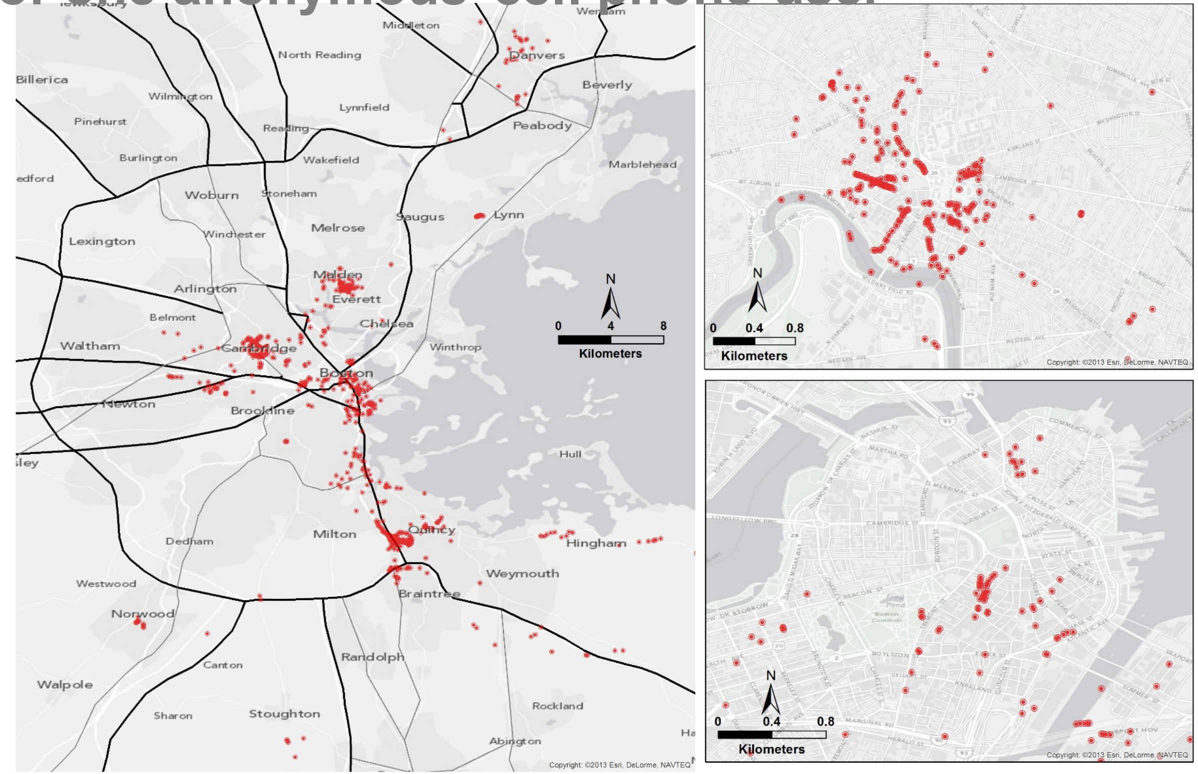
2011-2012



CHICAGO TEMPORAL ACTIVITY PATTERNS: WEEKDAY

Raw Data Description

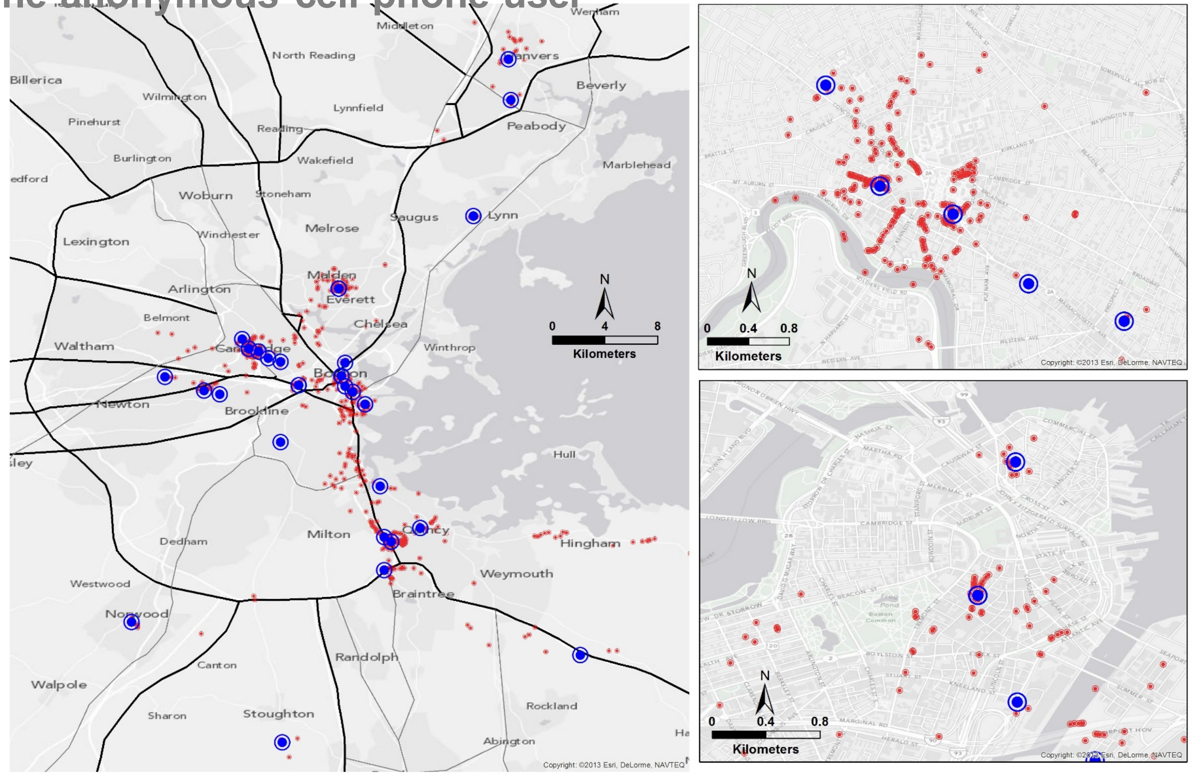
Example of one anonymous cell phone user



1776 phone records for one anonymous user in 2 months, February and March, 2010

Extraction of Daily Trajectories

Example of one anonymous cell phone user



Final agglomerated activity destination points for this user: 28 points.

Spatio Temporal Patterns

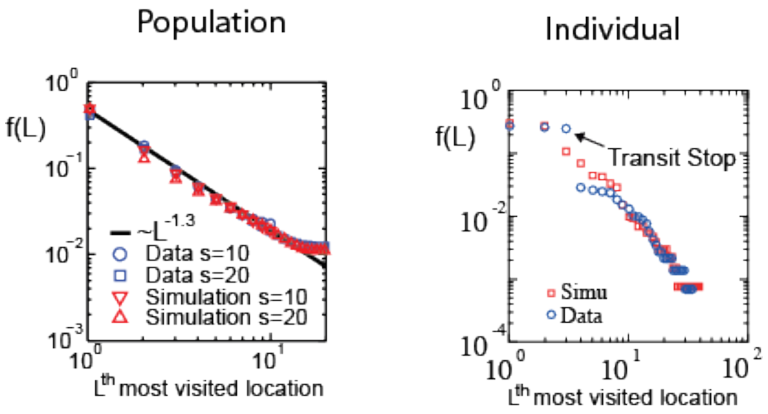
Temporal patterns

1. Activity stay duration
2. Number of trips in a day

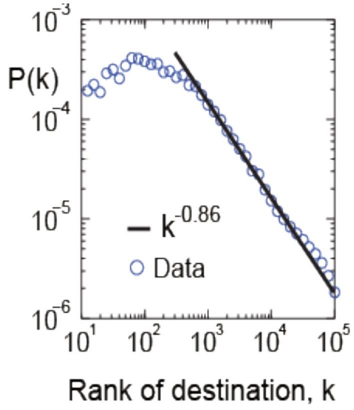
Spatial Patterns

1. Distance Traveled
2. Frequency of Visits
3. Number of locations visited per day

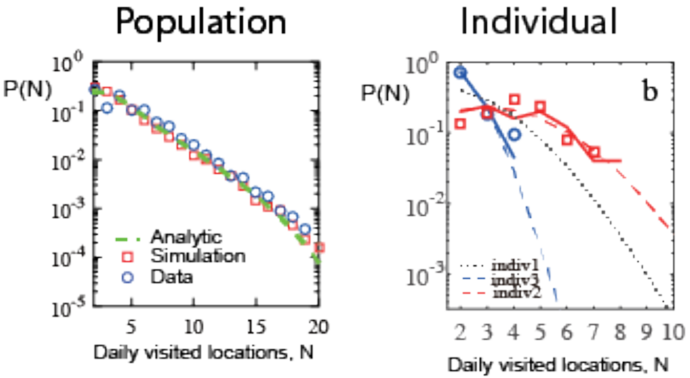
Frequency of Visits per Location



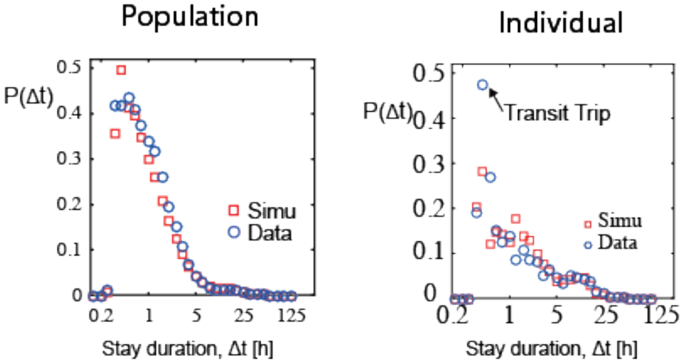
Ranking of Explorations



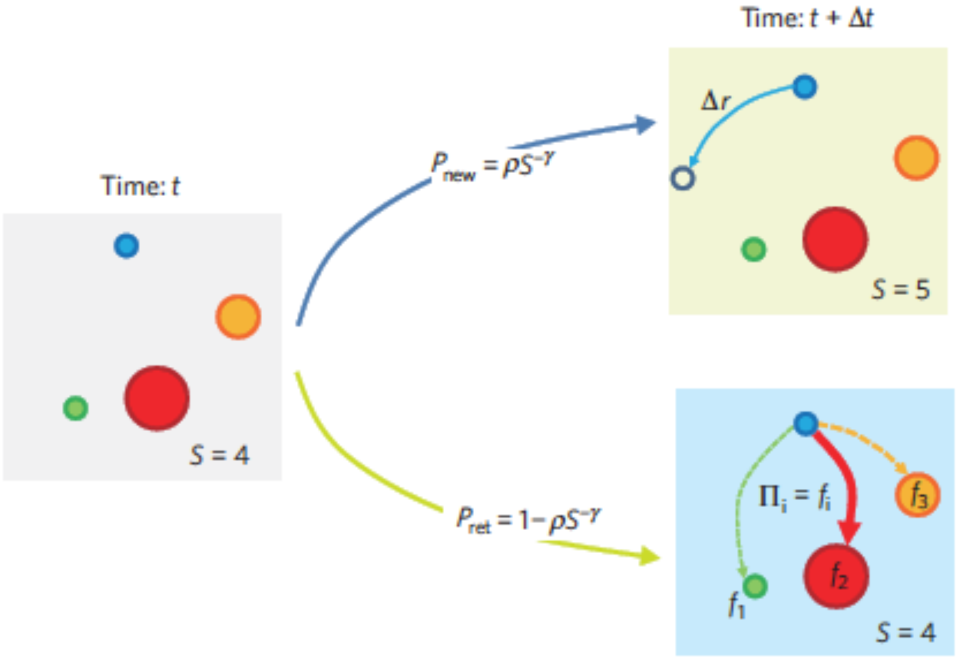
Distribution of Visited Location per Day



Distribution of stay duration



Preferential return and Exploration



Explains hierarchy of visits

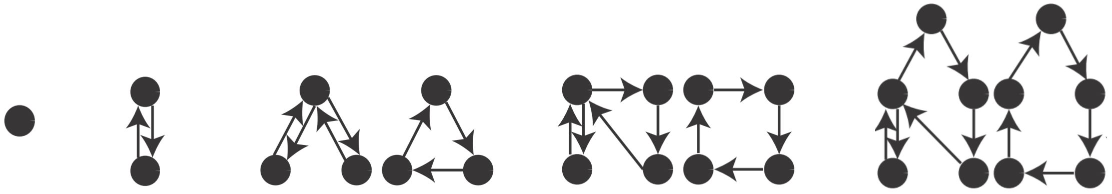
$$f(k) \approx k^{-\xi}, \xi \approx 1.2$$

Song, Koren, Wang, Barabasi, Nature Physics 2010

Universal Patterns of Individual Mobility

Daily Motifs

Statistically significant configurations of individual's travel network



✧ Nodes: visited stay regions

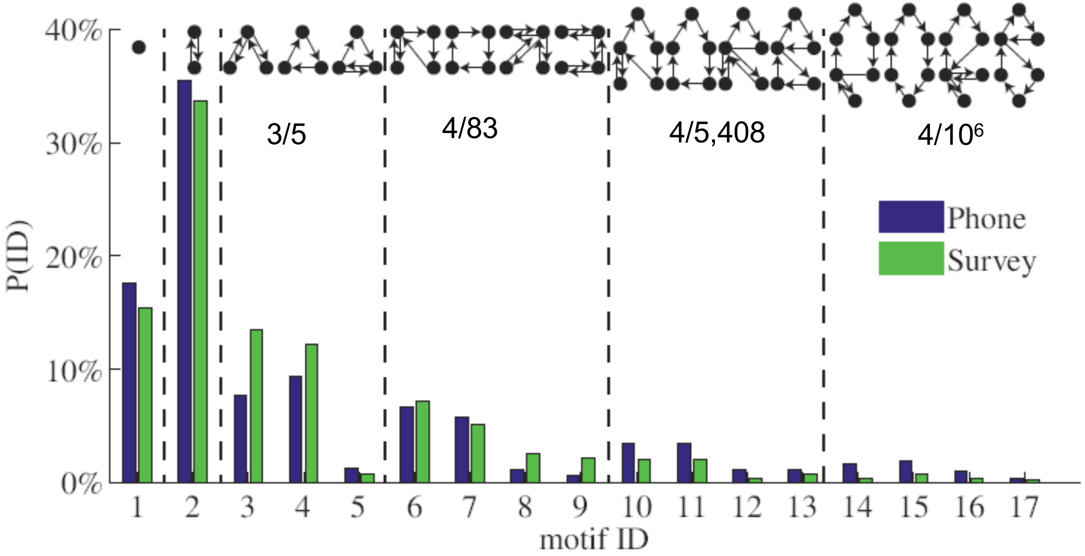
✧ Directed edges: trips between the nodes

Universal Patterns of Individual Mobility

Daily Motifs

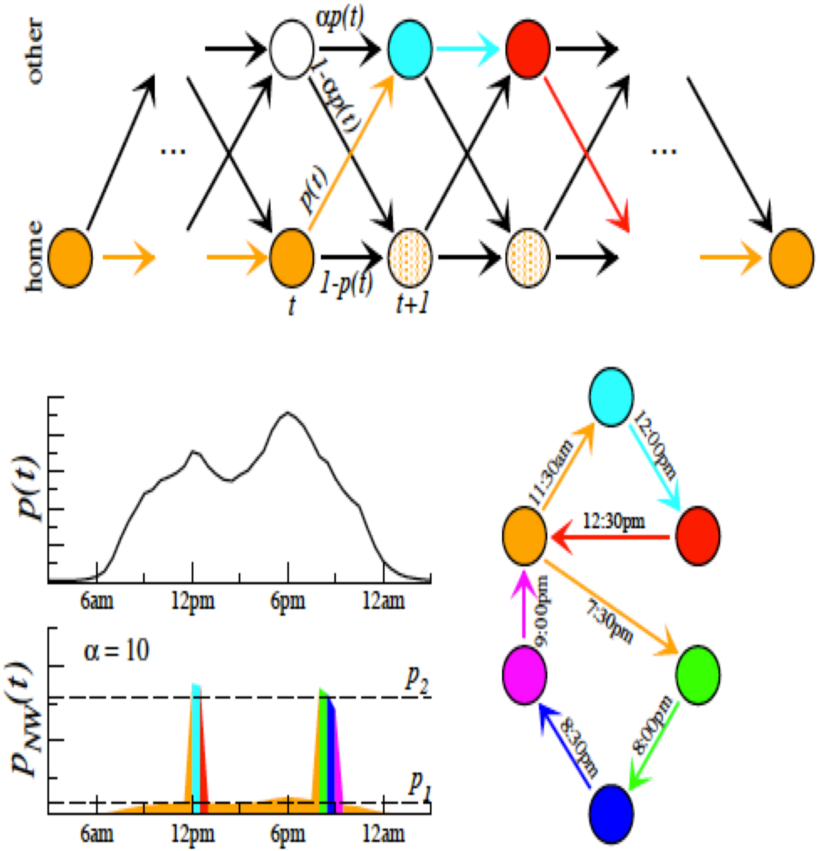
17 most frequent motifs account for over 95% of the measured daily trips.

Cell-phone data can be treated as **survey data** for analyzing human mobility.



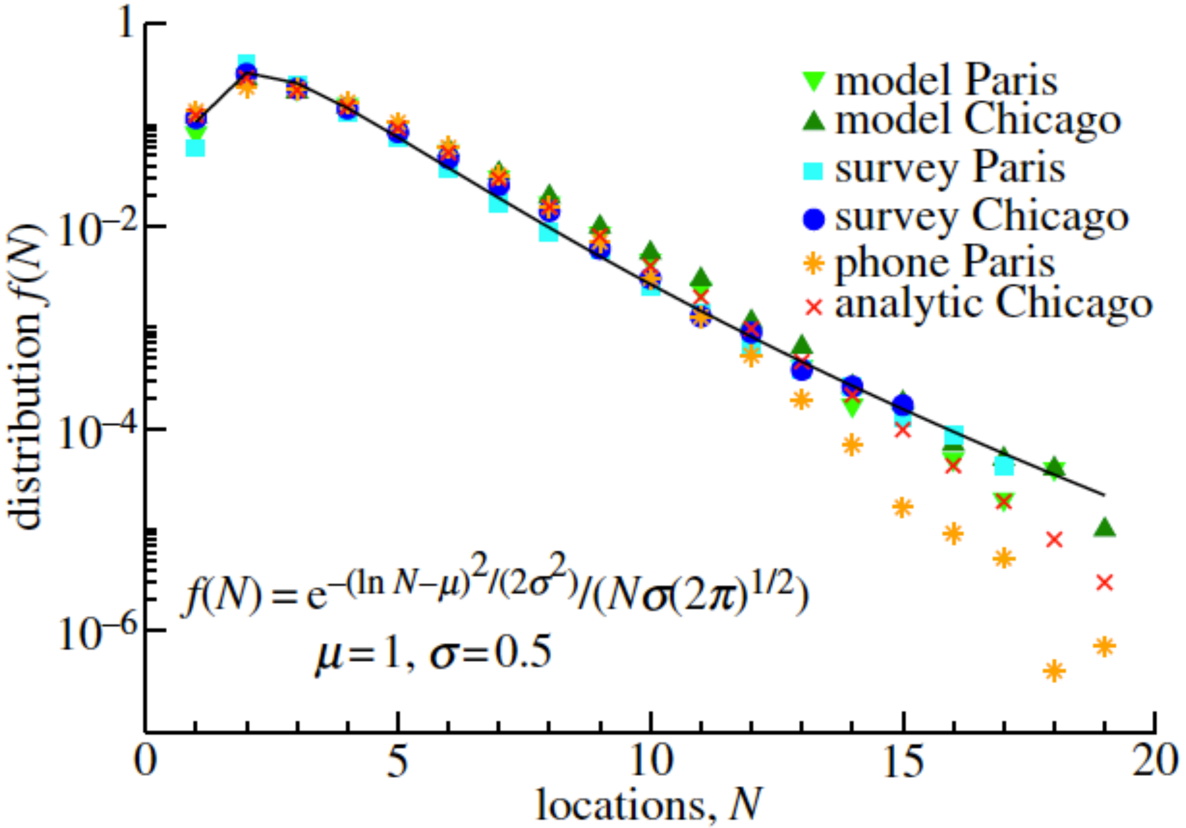
Data source: Massachusetts Travel Survey Data and Cell Phone Data in 2010.

The Perturbation Model

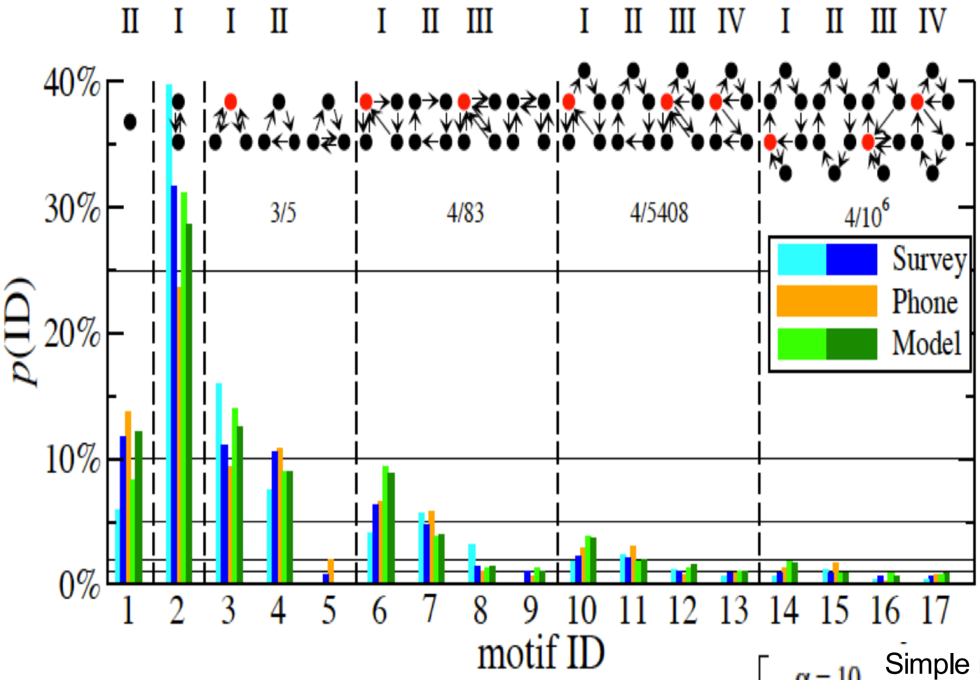


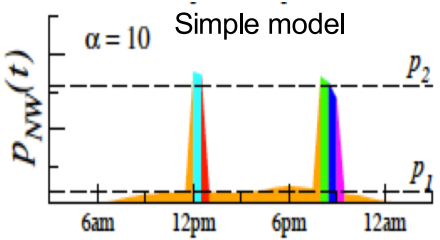
Unravelling daily human mobility motifs by *C. M. Schneider et al.*, Interface, 2013

Distr. Number of Visited Locations by the Population in 1 day.



Basic Ingredients of the Model:
Transition between Flexible Activities + Time Budget



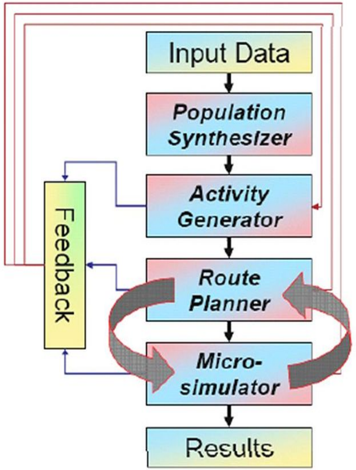


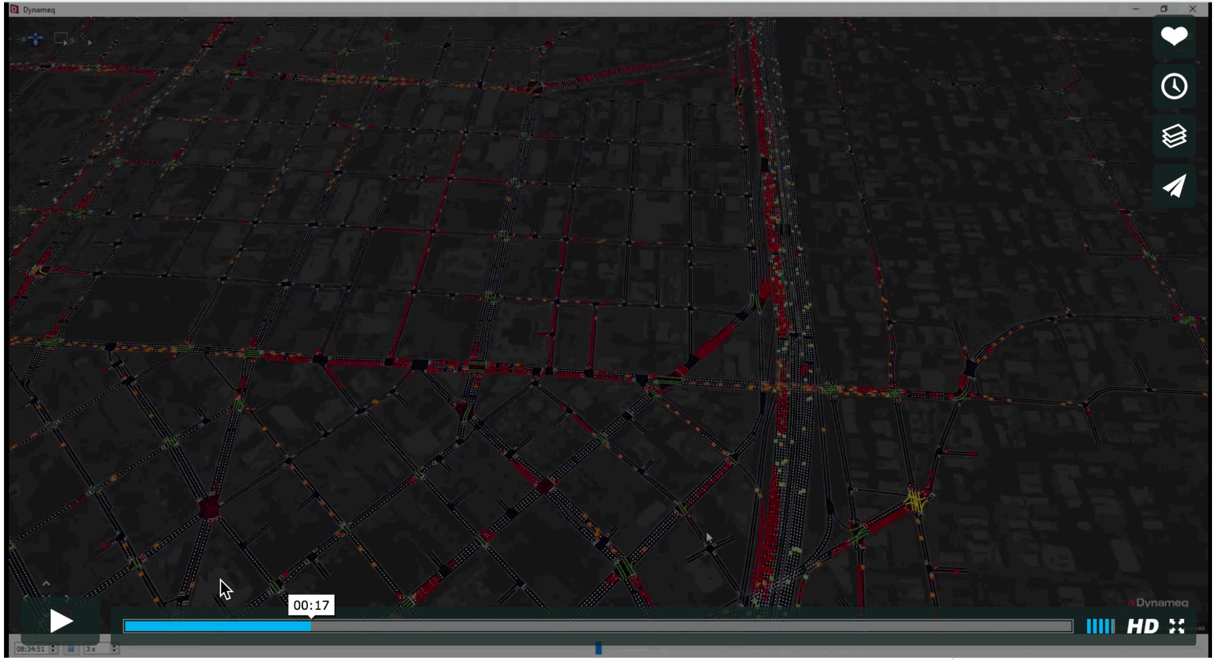
Journal of the Royal Society Interface 2013

Unravelling daily human mobility motifs

Christian M. Schneider[1], Vitaly Belik[1,2], Thomas Couronné[3],
Zbigniew Smoreda[3] and Marta C. González[1,4]

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

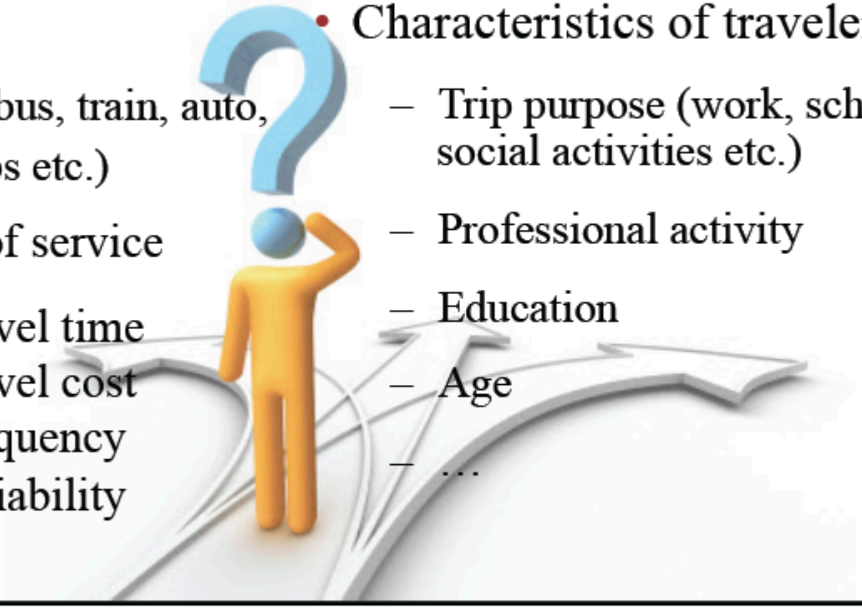




Sample Paper:
Behavioral calibration and analysis of a large-scale travel microsimulation G Flötteröd, Y Chen, K Nagel
Networks and Spatial Economics 12 (4), 481-502

Transportation demand modeling

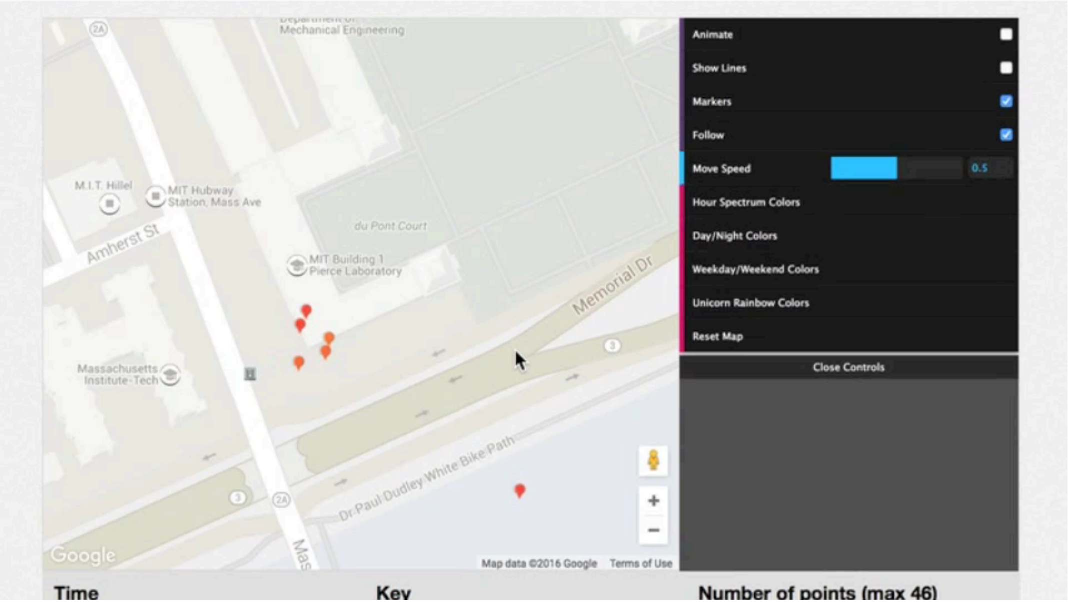
- Choice models based on the attributes of the transport alternatives and characteristics of travelers



- 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
  - …

Main unknown is Building Occupancy

OpenPath



In 2011 researchers found and Apple confirmed that iPhone and and iPad collected information related to the devise location

TimeGeo: 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

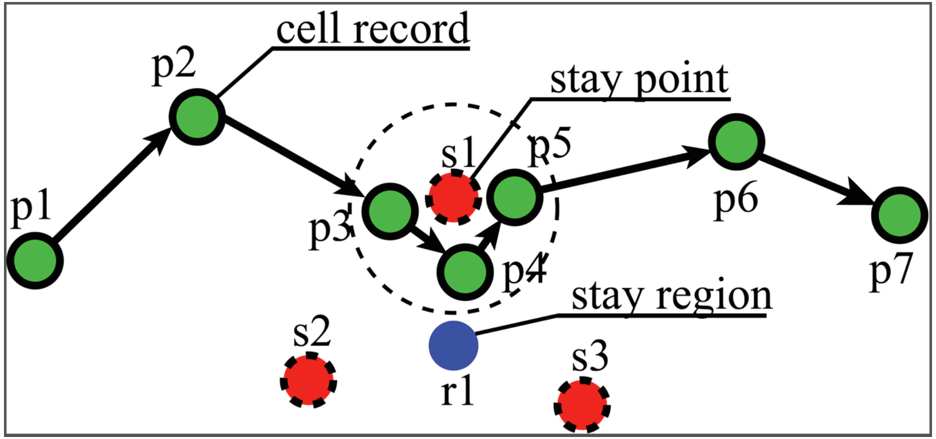
Model Individual Trajectories
(resolution 10min and 300mts radius)

Stay region extraction

From stay to stay region

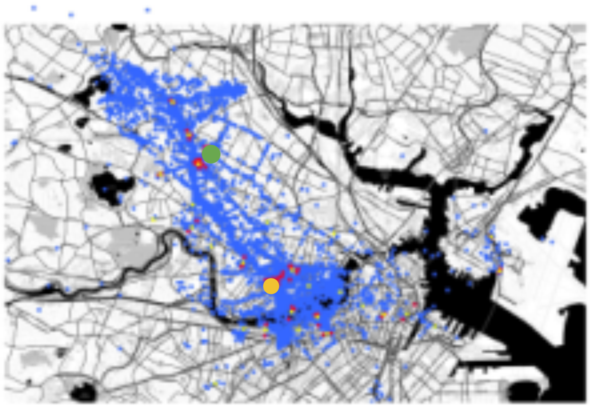
Stay Region: stays from different trajectories might represent the same location

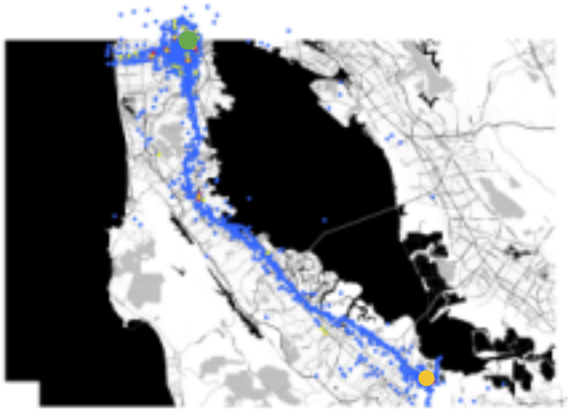
Interchangeable with “location”



Stay Extraction

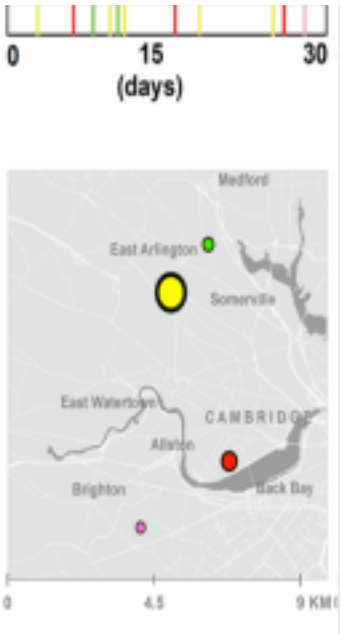
- **Home:** The stay point at which a user is observed the most between 8pm and 7am on weeknights.
- **Work:** The stay point at which a user is observed the most between 7am and 8pm on weekdays, provided this location is visited more than once per week and is more than 500m from their home location. Users are not necessarily assigned a work location.
- **Other:** All other stay locations.

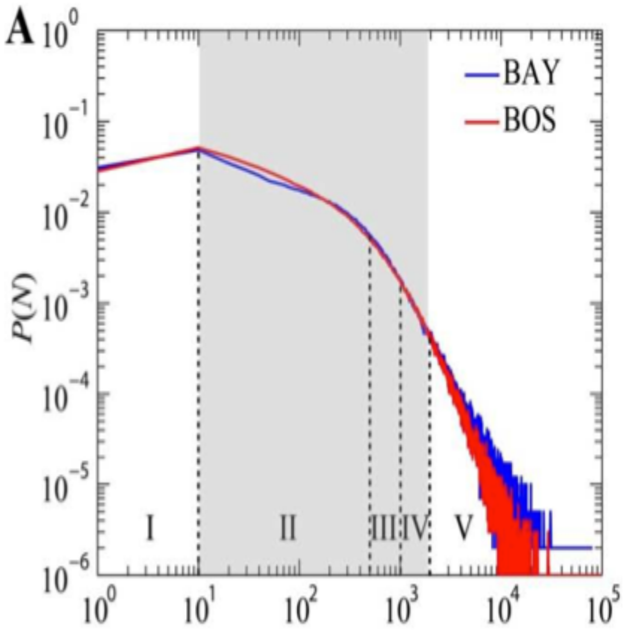




How to learn features from passive users to Model Sparse users?

Stays of Sparse User



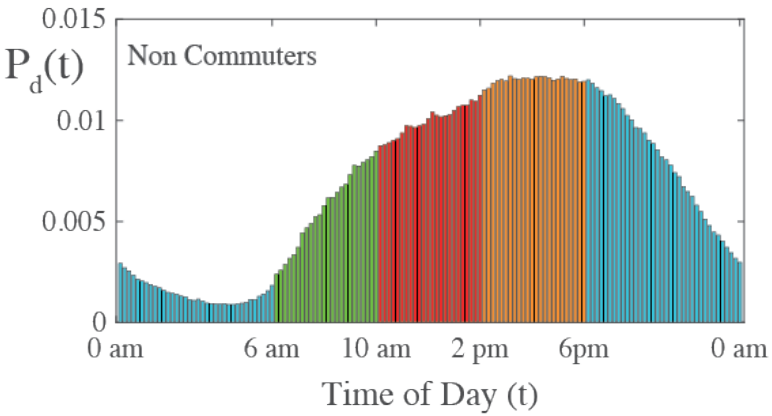


N (number of records per month)

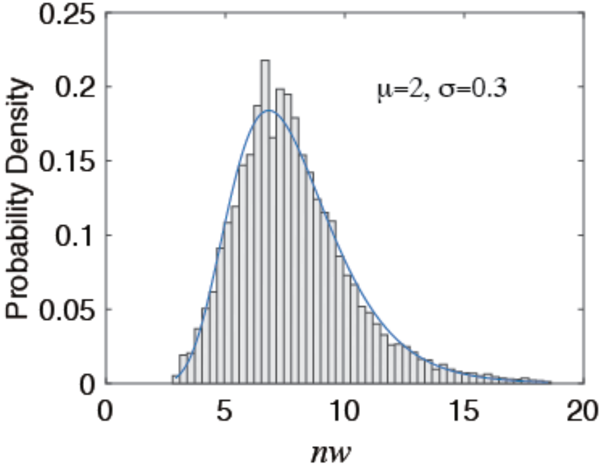
Sparse users: 10< N <30 per month

Active Users:
More than N=30 records per month

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



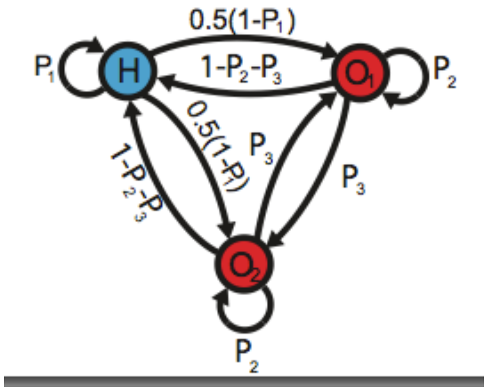
Fraction of Total Trips per time step



Weekly trips from home

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

Time independent Markov Model

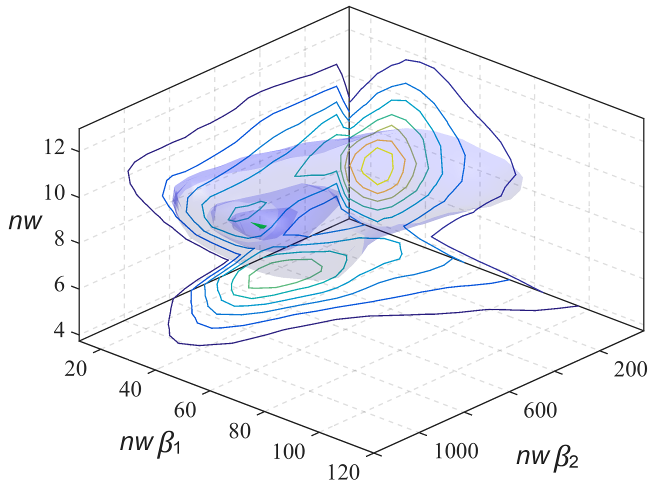


$$P_1 = 1 - P(t)$$ Moving from "home"

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

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

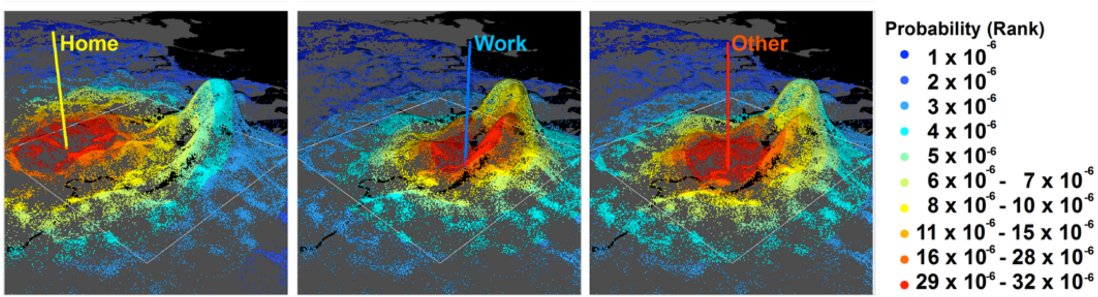
Spatio Temporal features from passive data

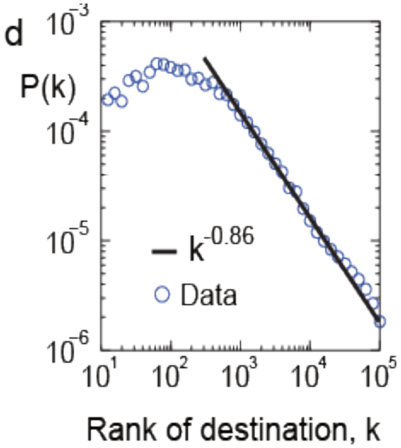


β_1 Generates shorter stays in the flexible location state.

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

Ranking of POIs to select New Destination

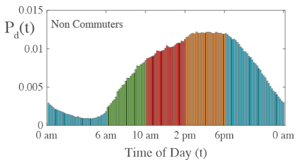




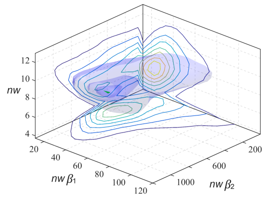
The Model

Features Extracted from data of Active Users

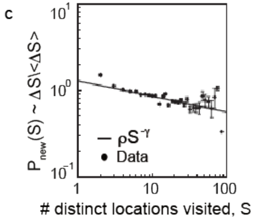
Global Trip prob.



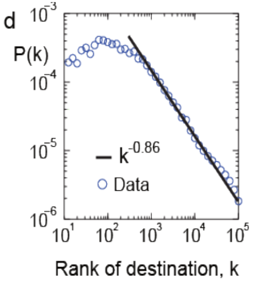
Individual Mobility Rates



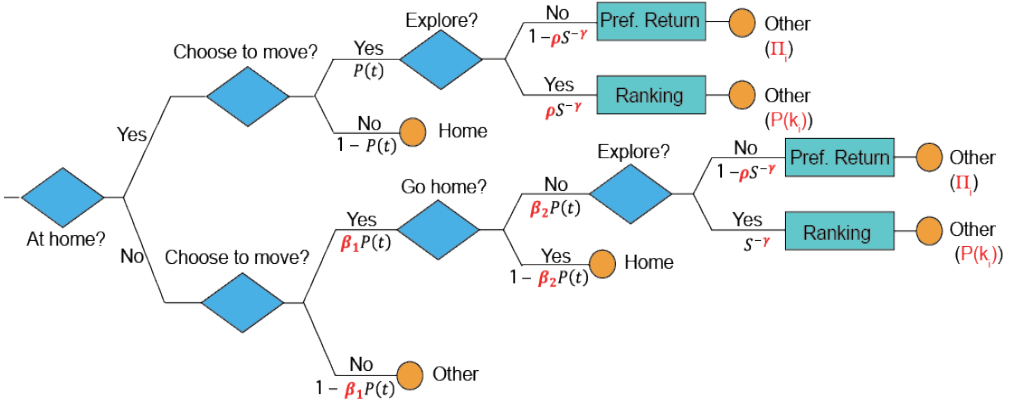
Preferential Return



Ranking of Explorations



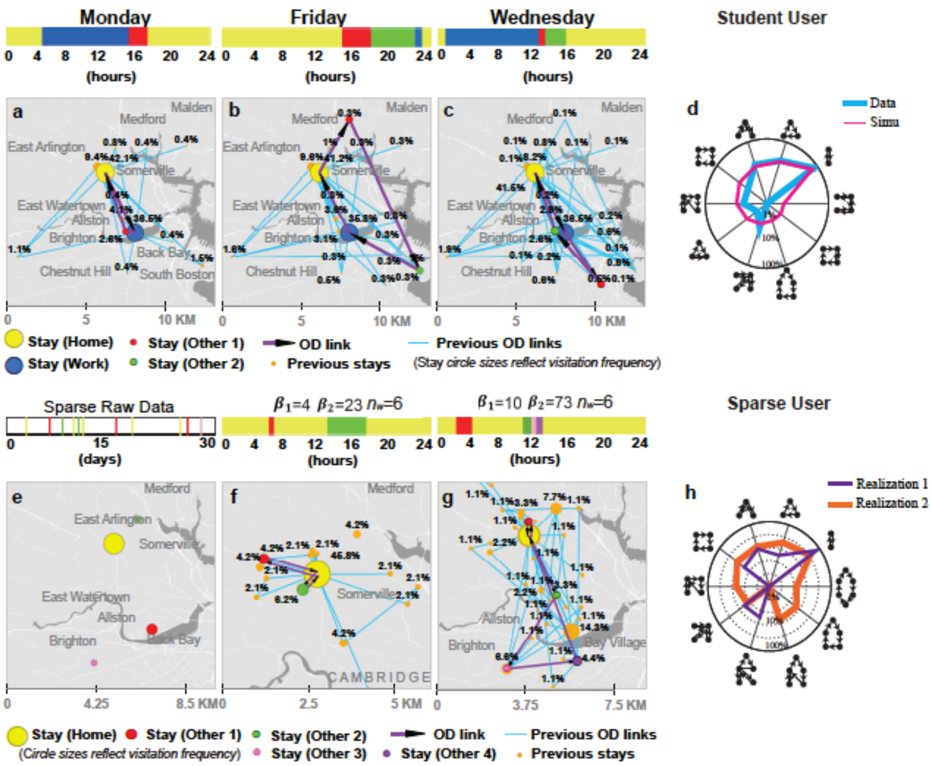
Flowchart of the Model



Marta C. Gonzalez

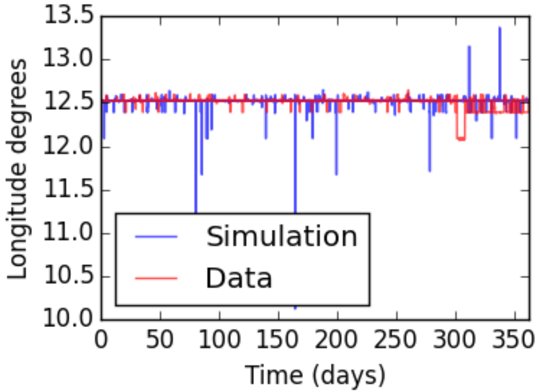
Models Results

Modeled Trajectories

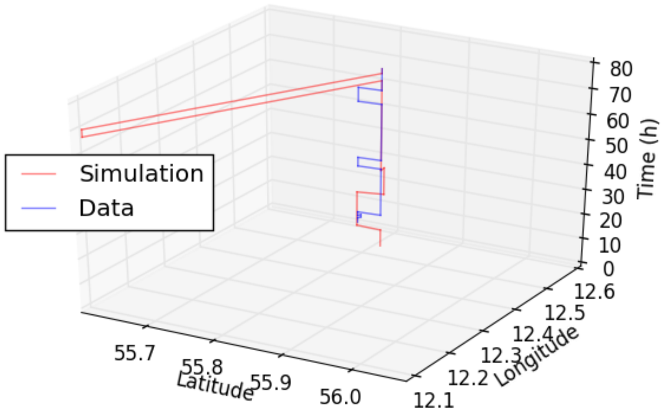


TimeGeo: Individual 1

364-day simulation



3-day simulation



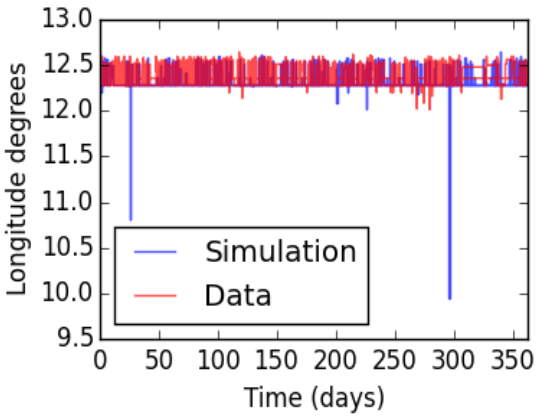
Similarity Data vs. Synthetic Trajectory

Day 1- 28.5% Day 2- 54% Day 3- 65.3%

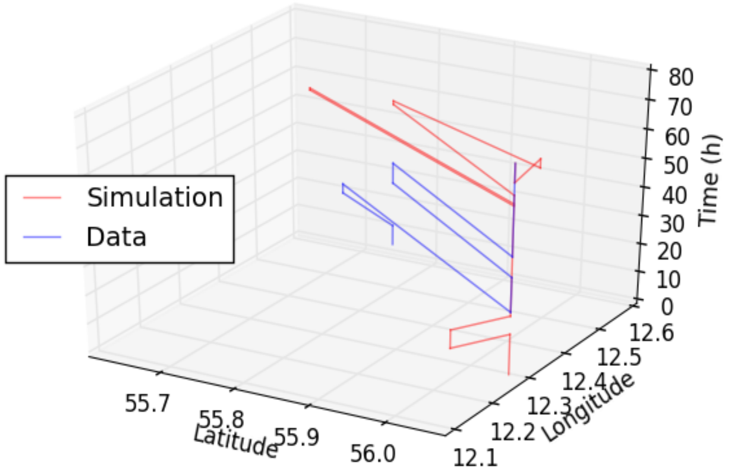
Overall- 53.7% (363 days)

TimeGeo: Individual 2

364-day simulation



3-day simulation



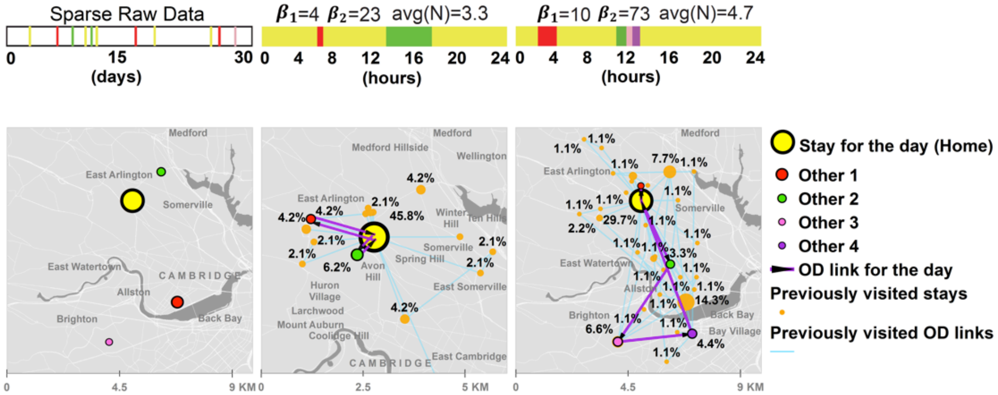
Similarity Data vs. Sim. in each day

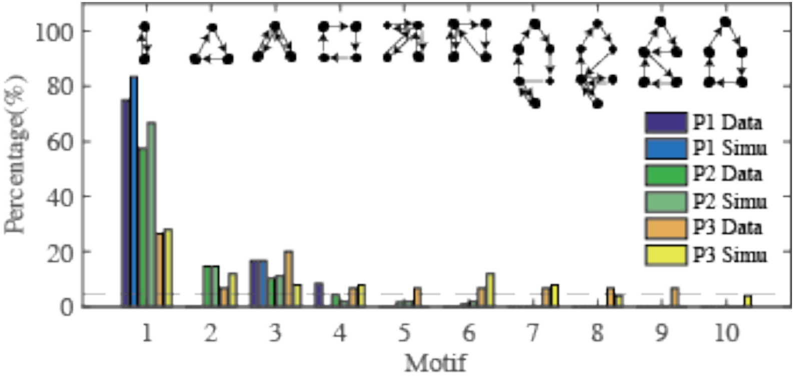
Day 1- 58% Day 2- 70.8% Day 3- 79.8%

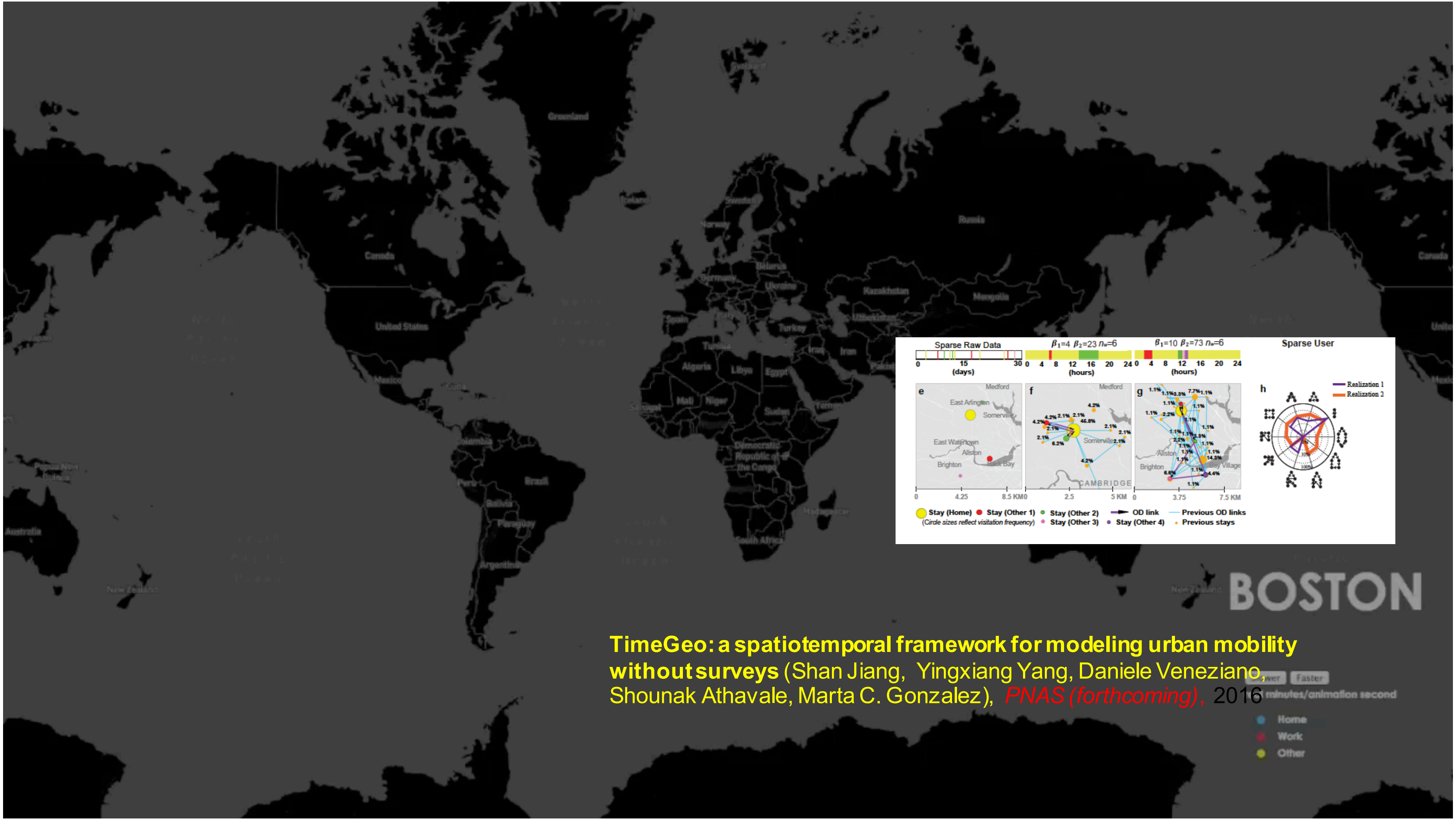
Overall- 46.9% (363 days)

TimeGeo Sparse Users

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

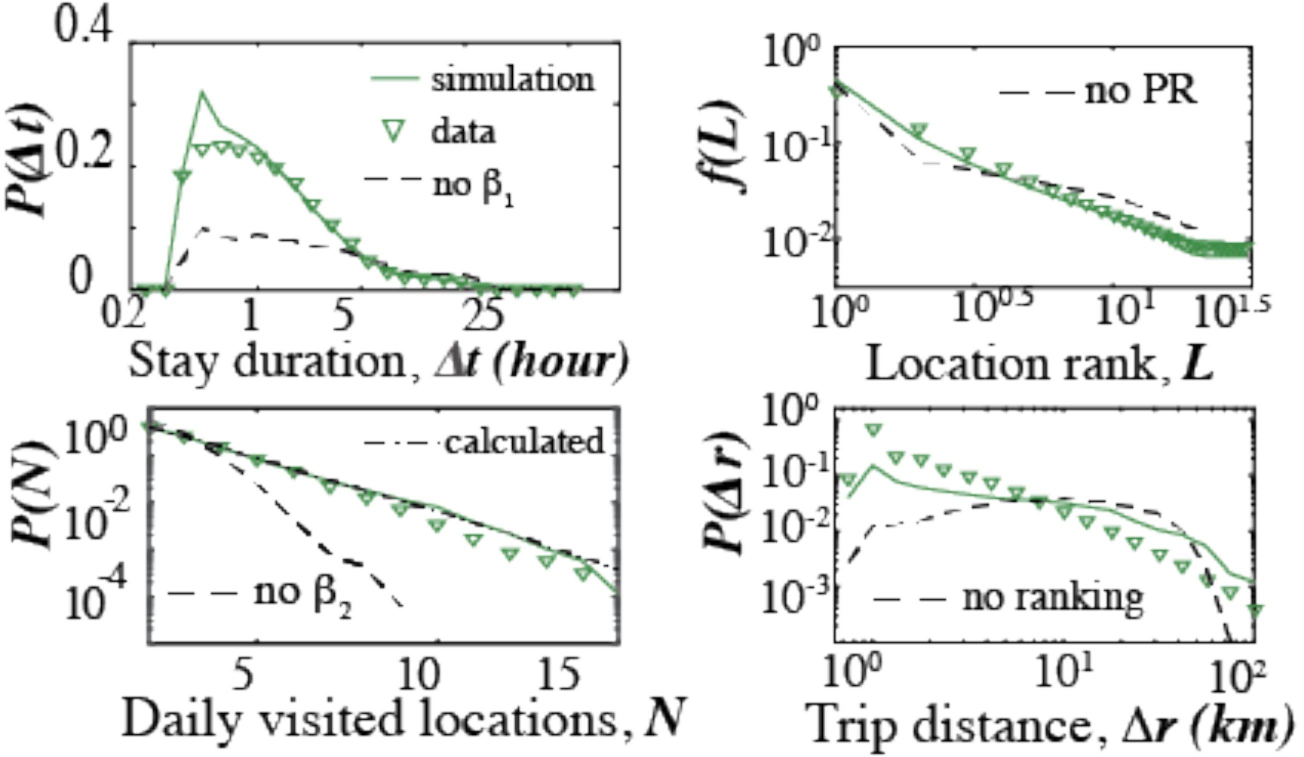




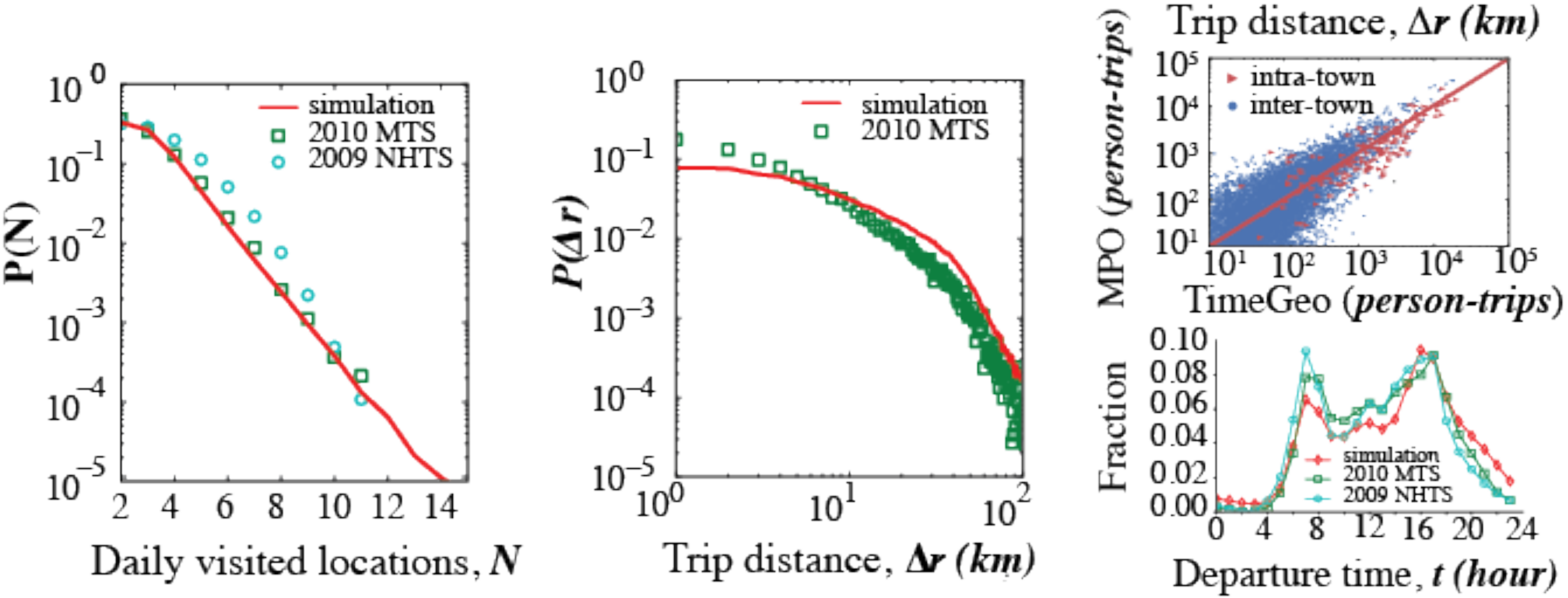


TimeGeo: a spatiotemporal framework for modeling urban mobility without surveys (Shan Jiang, Yingxiang Yang, Daniele Veneziano, Shounak Athavale, Marta C. Gonzalez), *PNAS (forthcoming)*, 2016

TimeGeo Individual Patterns



Comparison with travel demand models based on travel surveys



Conclusions

Mainstream models require sociodemographic information from costly manual surveys, which are small in sample sizes and updated in low frequency

We presented an individual mobility framework, TimeGeo, that extracts features from passively collected
data sources.

The model is able to generate individual trajectories in high spatial resolution with interpretable mechanisms, capturing heterogenoeus individual choices.

It can be flexibly adapted to input data with various resolutions, and extended for various modeling purposes

Coupling Human Mobility and Social Behavior

Collaborators

- Carlos Herrera-Yague - PhD Student
- Christian M. Schneider - Postdoc
- Marta C. Gonzalez – Advisor

INTERFACE

JRS Interface (2015)

rsif.royalsocietypublishing.org

Mobility and Social Behavior

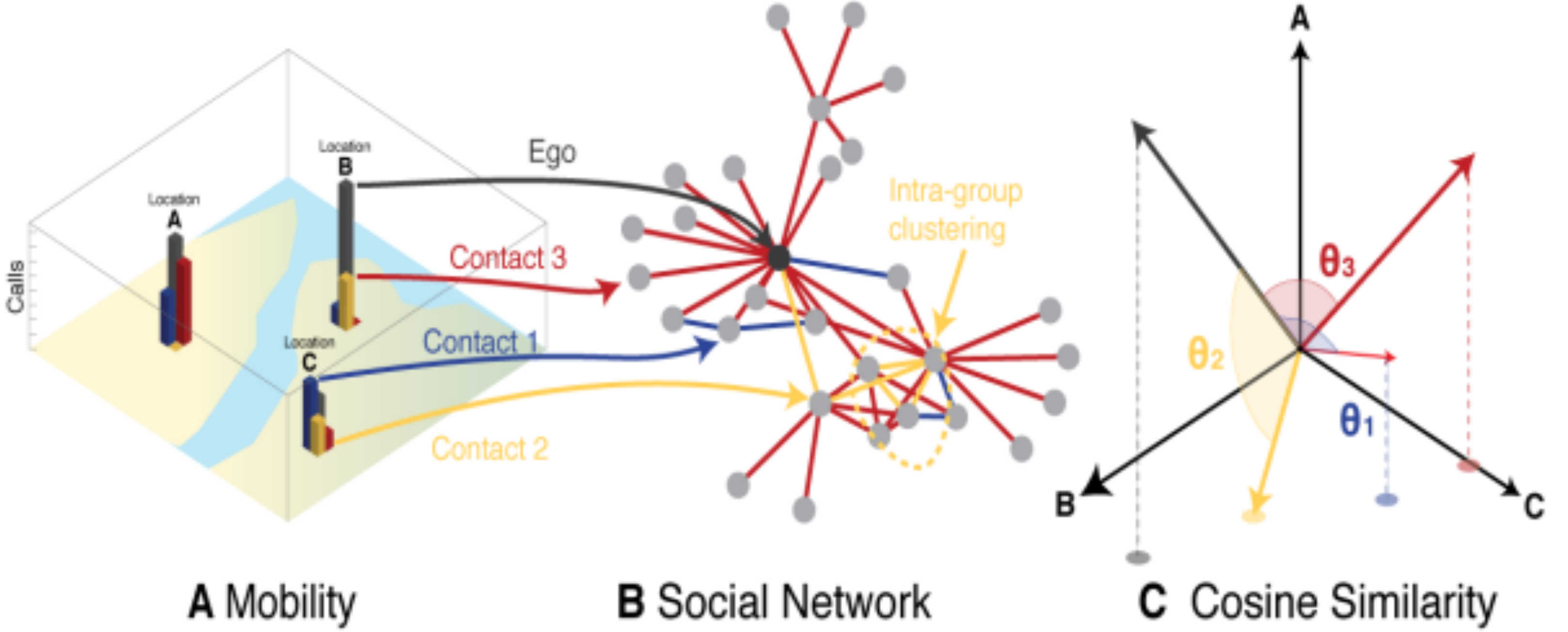
Geography in social networks:

- Users who live near each other are far more likely to be friends (Liben-Nowell 2005)
- Geographic proximity can improve link prediction in networks (Wang 2011)
- Roughly 15-30% of trips are taken for social purposes (Cho 2011, Grabowicz 2013)
- Predictions of a user's movement are improved by information about the movement of their friends (Domenico 2012)

Open Questions:

- How do we measure mobility similarity within urban areas where distance is less important?
- How much of a user's visitation patterns can we hope to reconstruct from the movement's of their social contacts?
- Can we contextualize social relationships by looking at features of movement?

Mobility and Social Behavior



Mobility and Social Behavior

Mobility Similarity

- vi and vj are location vectors for nodes i and j.
- Takes values between -1 and 1.
- Accounts for visit frequency.
- Similarity not inflated by many 0 elements.
- Controls for differences in call volume.

$$\cos\theta_{i,j} = \frac{\mathbf{v_i} \cdot \mathbf{v_j}}{|\mathbf{v_i}||\mathbf{v_j}|}$$

Predictability

- Location vectors of contacts form a vector subspace.
- Project a user's location vector onto the subspace.
- Compare projection to actual vector.
- A upper bound on how much of a user's visitation patterns can be reconstructed from the visits of their contacts

$$\mathbf{A} = [v_{f_1}, v_{f_2}, \ldots, v_{f_n}] \text{ where } f_i \text{ are contacts of a user } j$$

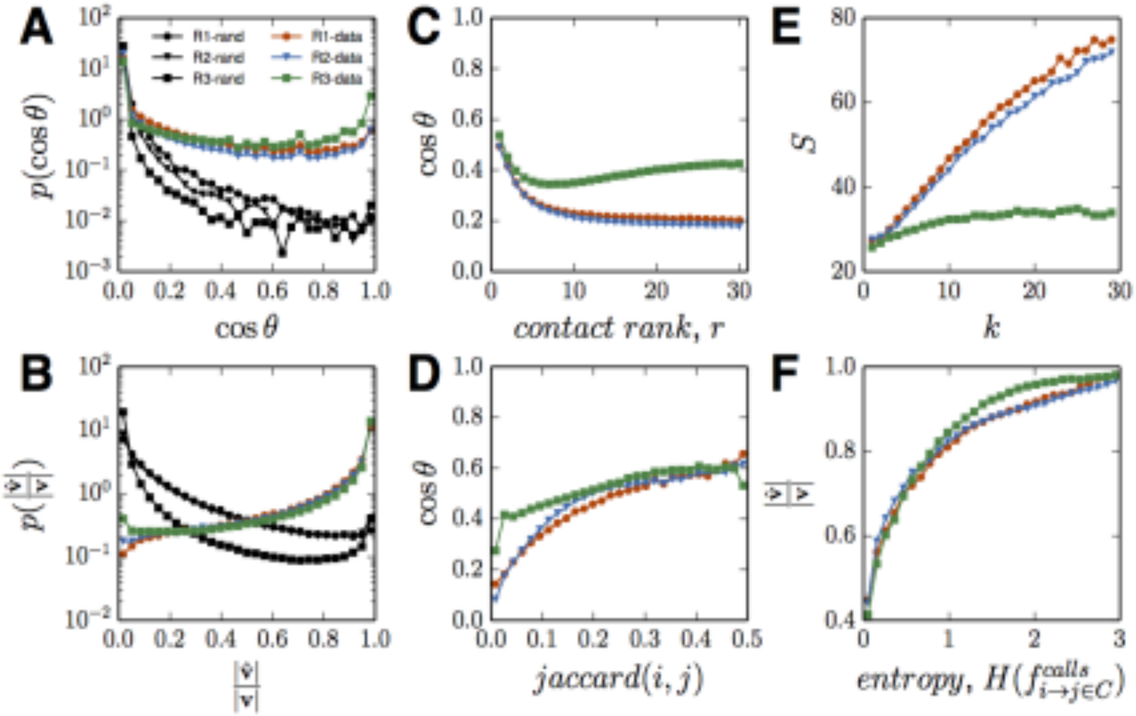
$$\mathbf{B} = [q_1, \ldots, q_{|F|}] \text{ where } \mathbf{B} \text{ is an orthonormal basis of } \mathbf{A}$$

$$\hat{\mathbf{v}} = \sum_{i=1}^{|F|} \langle q_i, \mathbf{v} \rangle q_i$$

$$predictability = \frac{|\hat{\mathbf{v}}|}{|\mathbf{v}|}$$

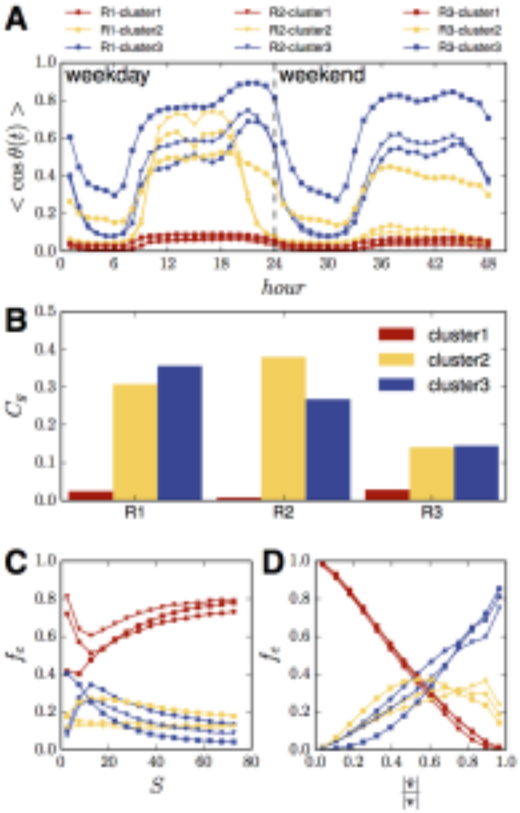
Mobility and Social Behavior

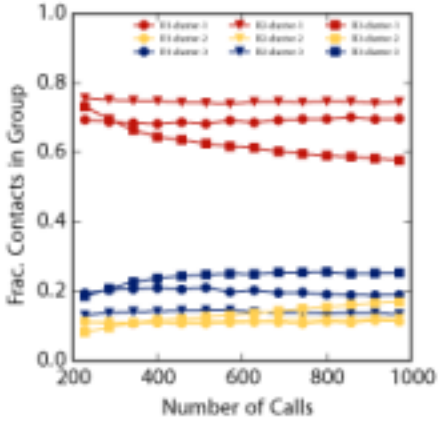
- Measurements made in three cities R1, R2, and R3 and two countries (R1-2, R3)
- Users are far more similar to and predictable by social contacts than strangers.
- Tie strength is positively correlated with mobility similarity.
- Social explorers are geographic explorers



Mobility and Social Behavior

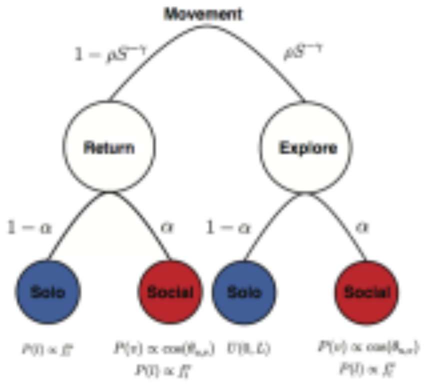
- Measure mobility similarity over time
- Cluster edges using k-means (paint edges)
- High intra-group clustering coefficient
- Composition of ego-network is correlated with mobility behaviors.

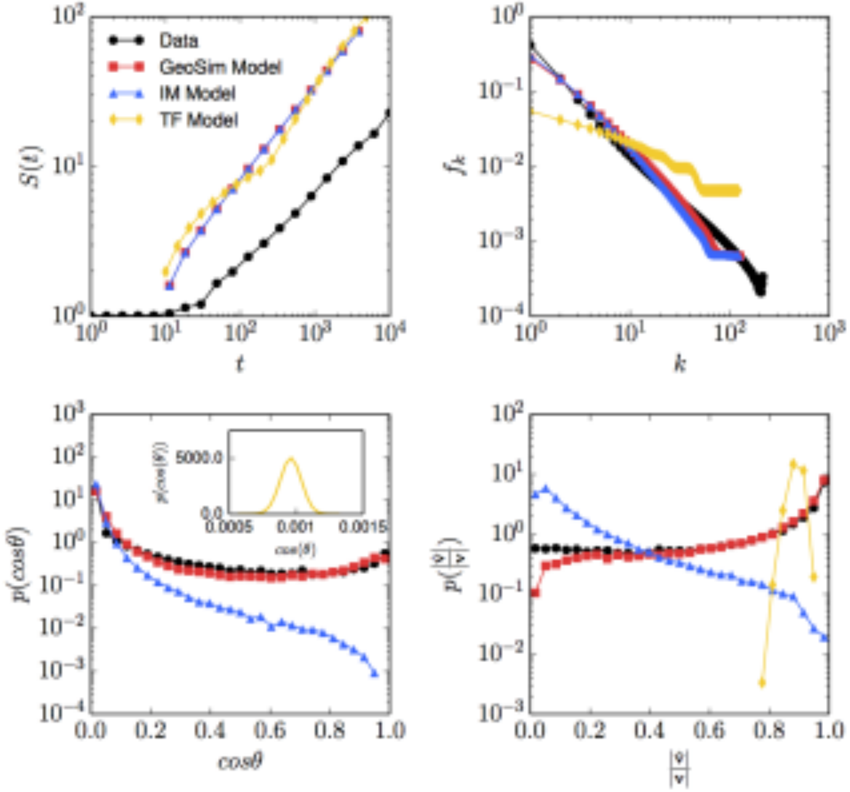




Mobility and Social Behavior

- Extend mobility model introduced by Song et al. Nature Physics 2010 to include social behavior.
- Alpha determines the strength of social forces for each user.
- Best guess at alpha's distribution is exponential with a mean of 0.3 inline with previous findings that 30% of trips are social in nature (Cho 2011).





Understanding individual routing behavior

Lima, A, Stanojevik R., Papagianaki D., Rodriguez P. & Gonzalez, M. C

INTERFACE

rsif.royalsocietypublishing.org

Massachusetts Institute of Technology

JRS Interface *(2016)*

Motivations

It is the natural next step in understanding human mobility.

Route choice is a fundamental step of traffic prediction, the task of transforming a set of travel demand (OD matrix) into flows and travel times.

The assumption that “people choose the minimum cost path”, although widely accepted in academic and commercial enviroments, has little empirical support.

How do people navigate in the city?

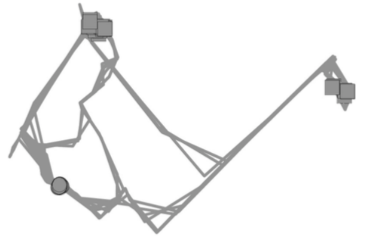
We analyse 1,5 M GPS trajectories, driven by a set of individuals within four major European cities during a period of 18 months.

How many routes a driver uses typically.

If the routes performed by users are "optimal".

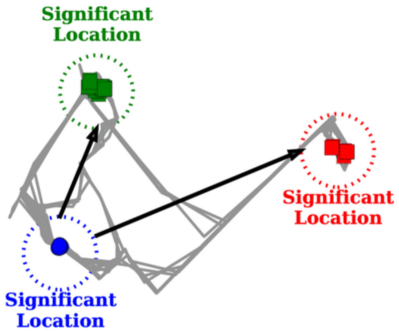
Whether some routes are predominant over others.

From trajectories to route choices



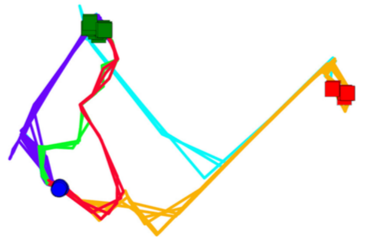
Each trajectory is composed by an arbitrary number of points, every N seconds.

We cluster each driver's source / destination points into a set of **significant locations**, here shown as dotted circles.



We group trajectories by source-destination pair into **routine trips**, here shown as black arrows.

Finally we further cluster the trajectories in each routine trip into a set of **route choices**, color coded in figure.



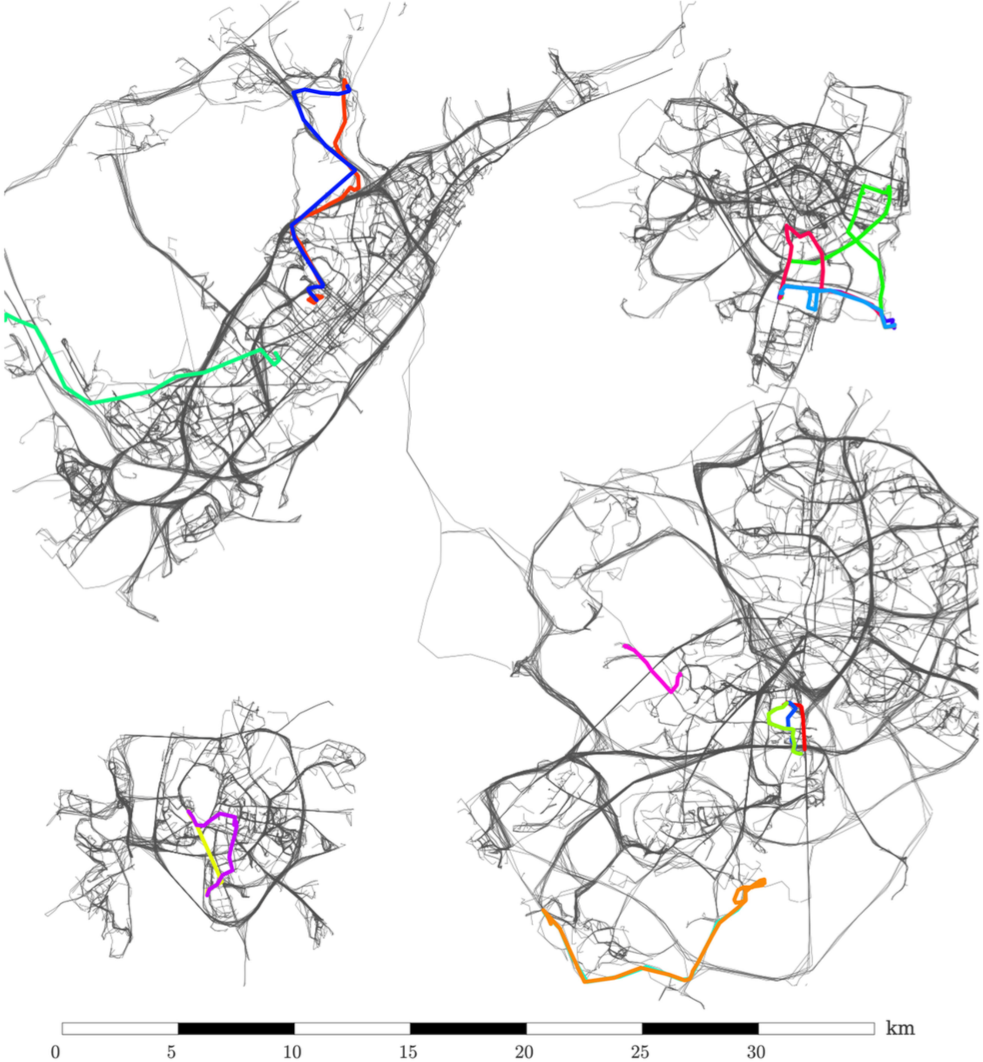
From *messy* trajectories to route choices

Clustering algorithms typically require the number of clusters to be specified. We instead use non-parametric algorithms, like MeanShift and DBSCAN.

Trajectories have an heterogeneous number of points (even on the same routes, because of traffic jams, delays, …). It is not trivial to compare them.
We used Dynamic Time Warping to establish a matching between the two sets of trajectory points.

This methodology is agnostic of the underlying urban network. It can be used to transform unstructured location sequences into route choices between significant locations in any city.

$$W(A_i, B_j) = d(a_i, b_j) + \min \begin{cases} W(A_{i+1}, B_{j+1}) \\ W(A_{i+1}, B_j) \\ W(A_i, B_{j+1}) \end{cases}$$



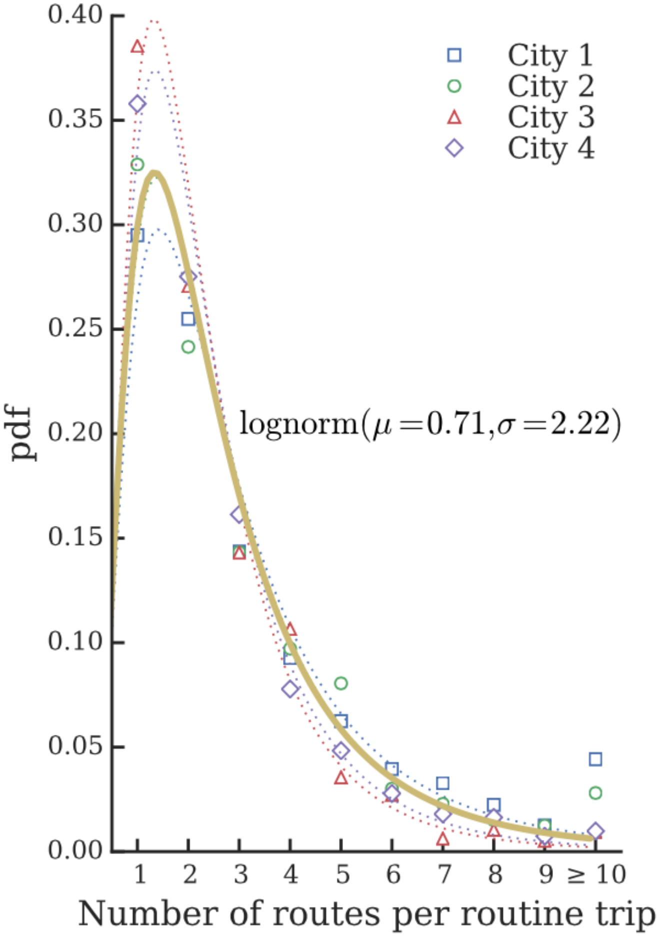
km
0
5
10
15
20
25
30

Most people use few routes, despite the total period under consideration is 18 months.

We compared user trips to trips returned by Google Directions API, which accounts for distance and traffic conditions.

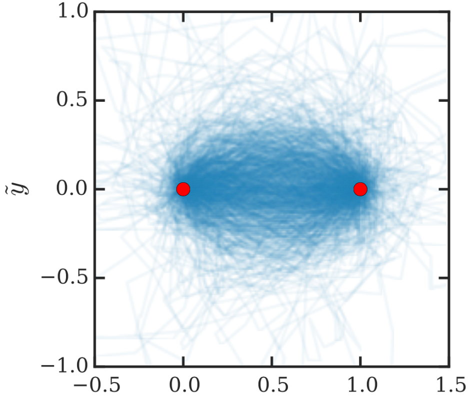
We found that **53% of the preferred trips ever used are not optimal**.

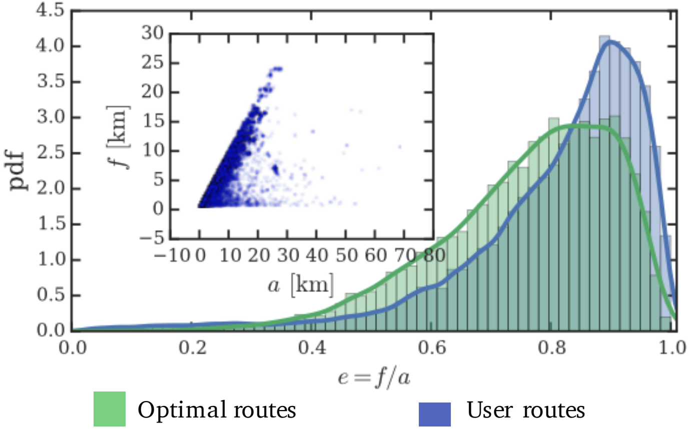
And the more often people travel between two locations, the more likely is for them to have a **preferred route**.

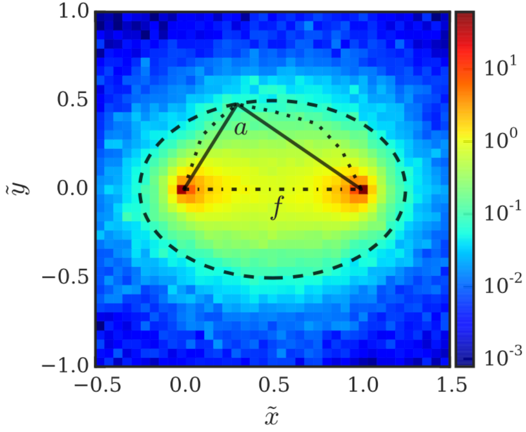


The boundaries of human routes

- We rototranslated and scaled every trajectory to the same reference system, having source (0, 0) and destination (1, 0).
- 95% of the positions are contained within an ellipse of high-eccentricity.
- Eccentricity measures us how much the user is away from the ideal straight location.







flickr: rogermeyer

Take away messages - Recap

Drivers often do not choose the shortest path.

Regardless of the urban network, they drive within an high-eccentricity ellipse, with foci as source / destination.

For recurring trips, a dominant route is preferred, and some alternative routes are occasionally taken.

This set of behavioural rules can be used to inform realistic models of routing behaviour that are not based on minimum-cost assumptions.


LANET 2017
HOME
ABOUT
SPEAKERS
PROPOSE A MINISYMPOSIUM
CONTRIBUTIONS & REGISTRATION
LANET BOARD
ORG. COMMITTEE
CONTACT
LANET 2017
1ST LATIN AMERICAN CONFERENCE ON COMPLEX NETWORKS (25-29 SEPTEMBER 2017, PUEBLA, MEXICO)

ALEX ARENAS

URV
Tarragona
Spain

S. BOCCALETTI

CNR & Italian Embassy
Tel Aviv
Israel

MARIO CHÁVEZ

CNRS
Paris
France

ERNESTO ESTRADA

University of Strathclyde
Glasgow
U.K.

MARTA C. GONZÁLEZ

MIT
Boston
USA

BRUNO GONÇALVES

New York University
New York
USA

CRISTINA MASOLLER

UPC
Terrassa
Spain

J.F.F. MENDES

University of Aveiro
Aveiro
Portugal

RONALDO MENEZES

Florida Institute of Tech.
Melbourne (FL)
USA

ADILSON MOTTER

Northwestern University
Evanston
USA

MASON A. PORTER

UCLA
Los Angeles
USA

JOSE S. ANDRADE JR

Univ. Federal do Ceará
Fortaleza
Brazil