Week 3

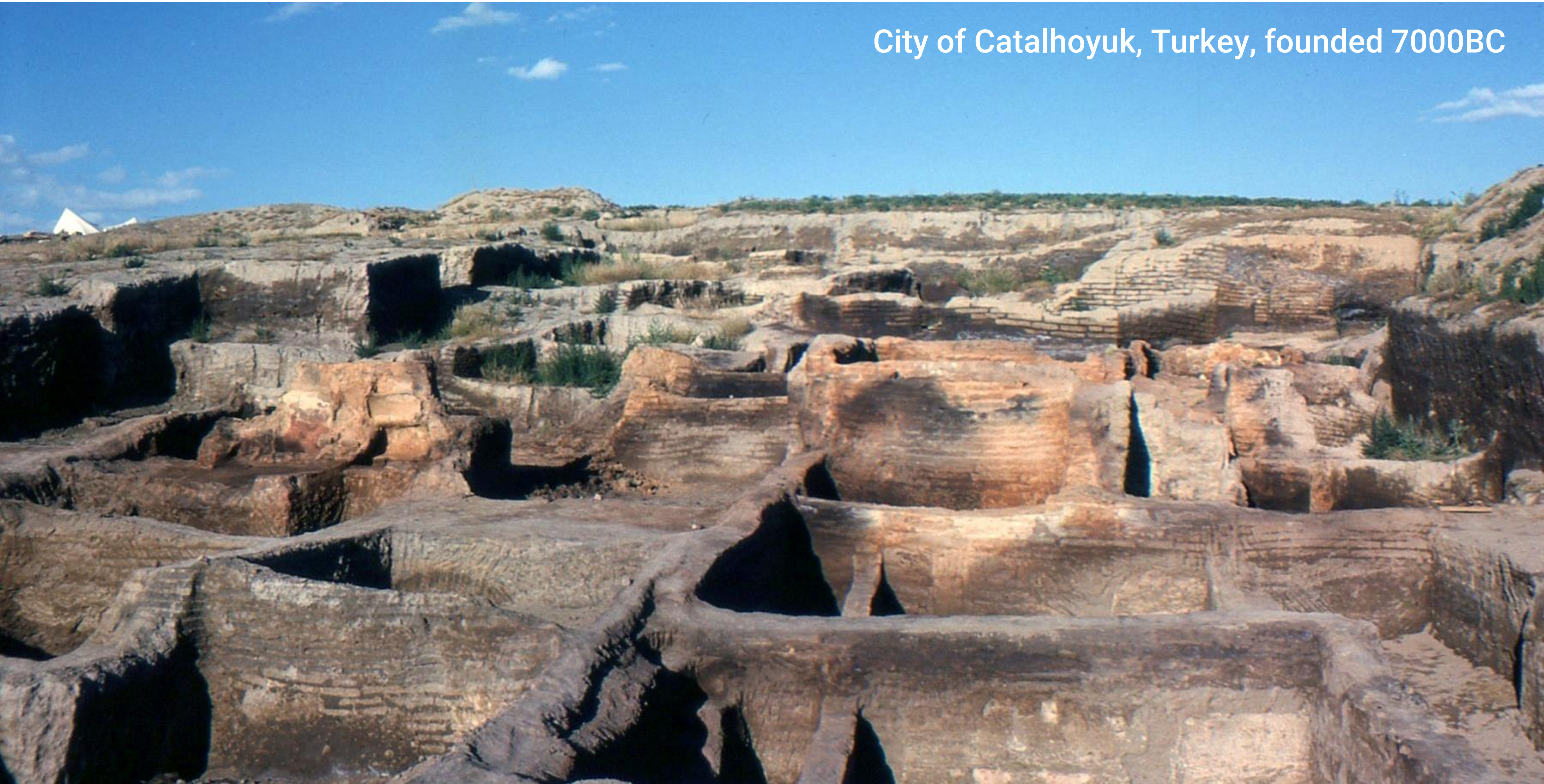
New Mobility Analytics

Paolo Santi

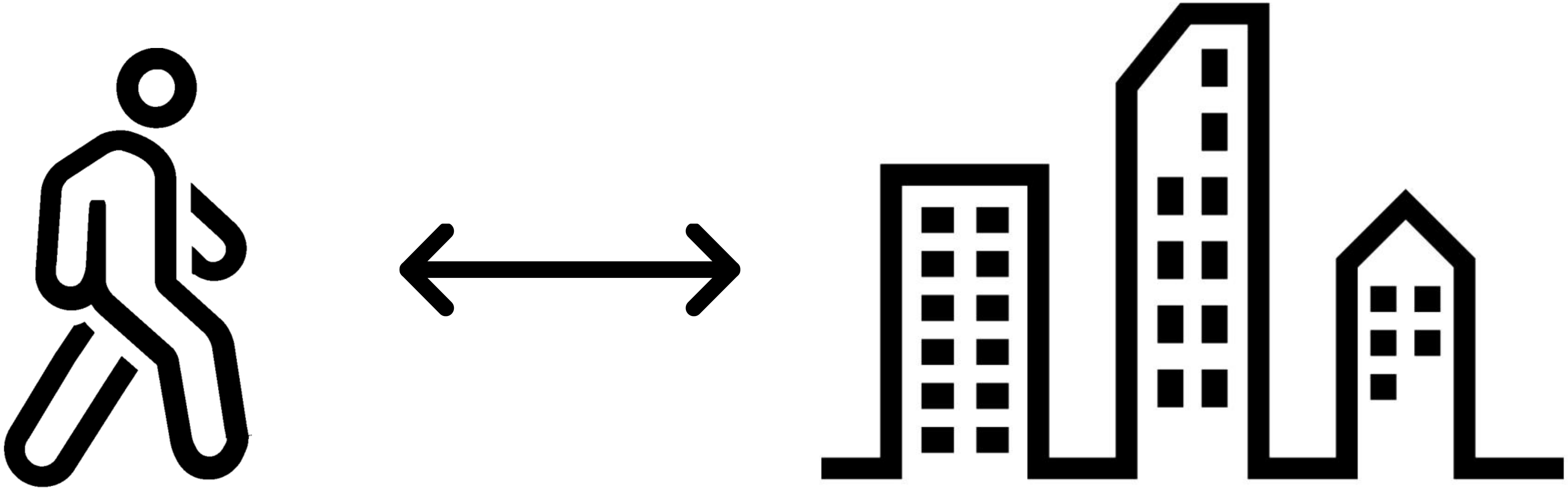
11.S951 *Senseable City: Data and Analytics* 02/18/2022

PAOLO SANTI
Principal Research Scientist
psanti@mit.edu

“Cities are the most persistent human construct” – L. Mumford



CITIES AND MOBILITY



mobility city

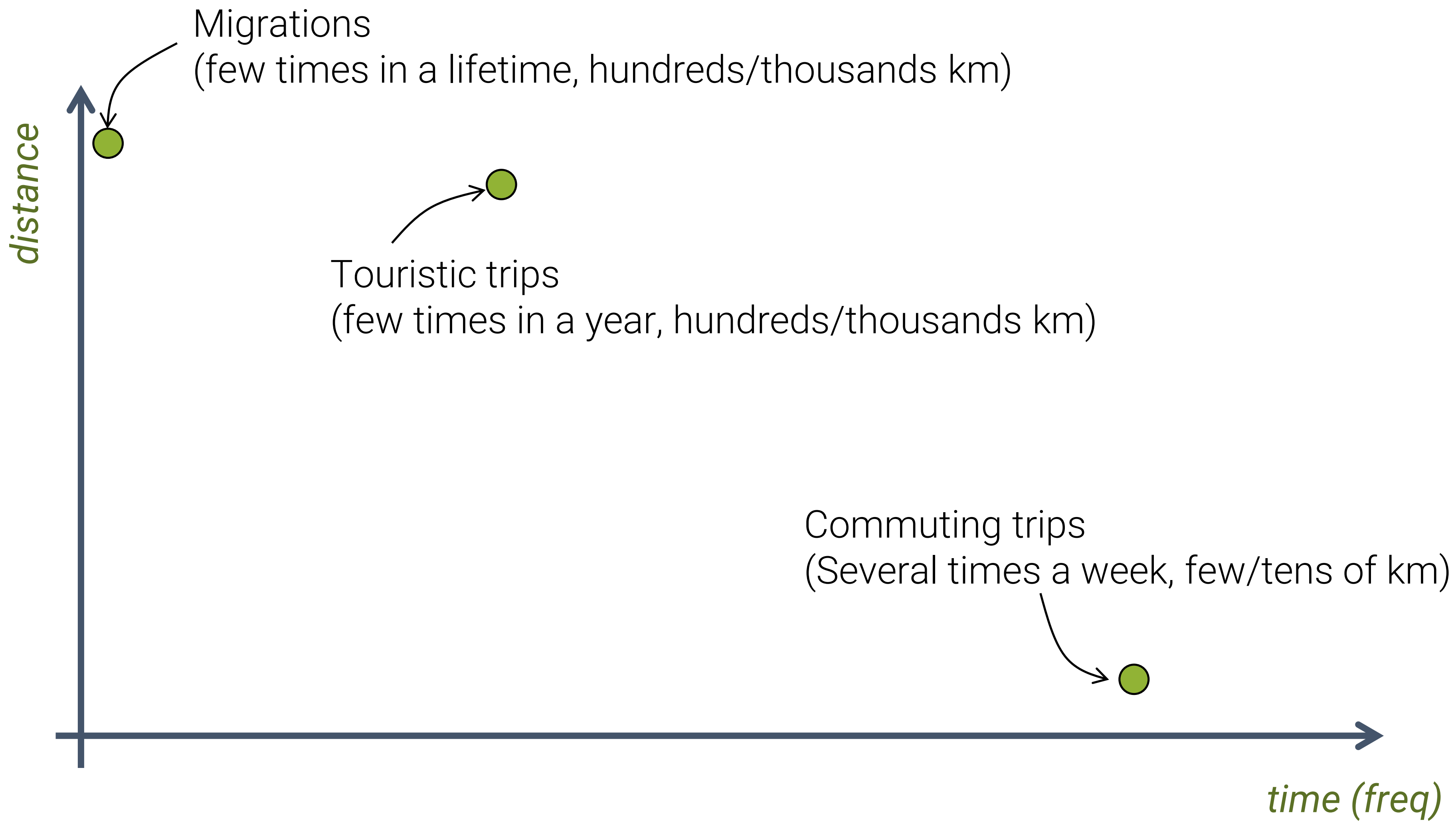

New York City – river, sea

SOME EUROPEAN CAPITAL CITIES

City	River	Sea
Amsterdam	Yes	Yes
Berlin	Yes	No
Lisbon	Yes	Yes
London	Yes	No
Madrid	No	No
Paris	Yes	No
Rome	Yes	No
Stockholm	Yes	Yes
Warsaw	Yes	No
Vienna	Yes	No

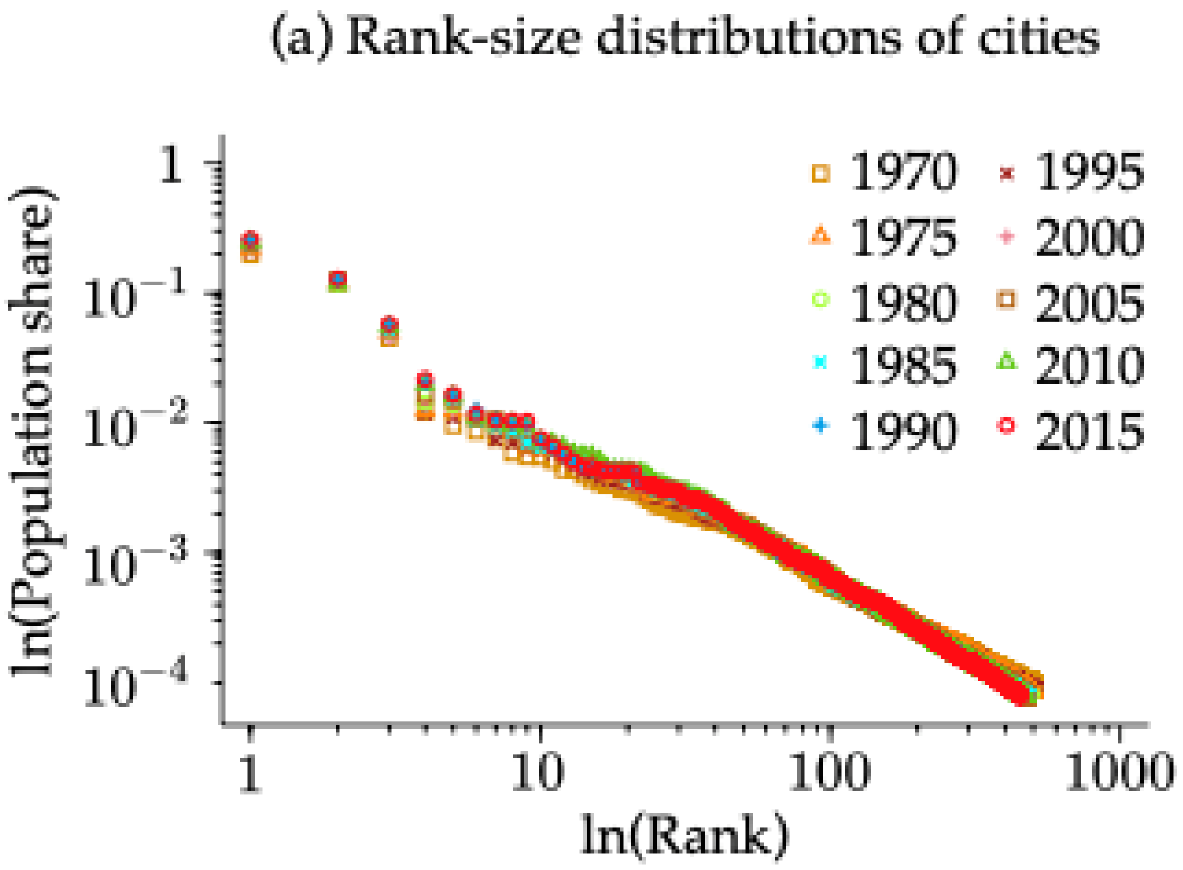
MOBILITY IN SPACE AND TIME

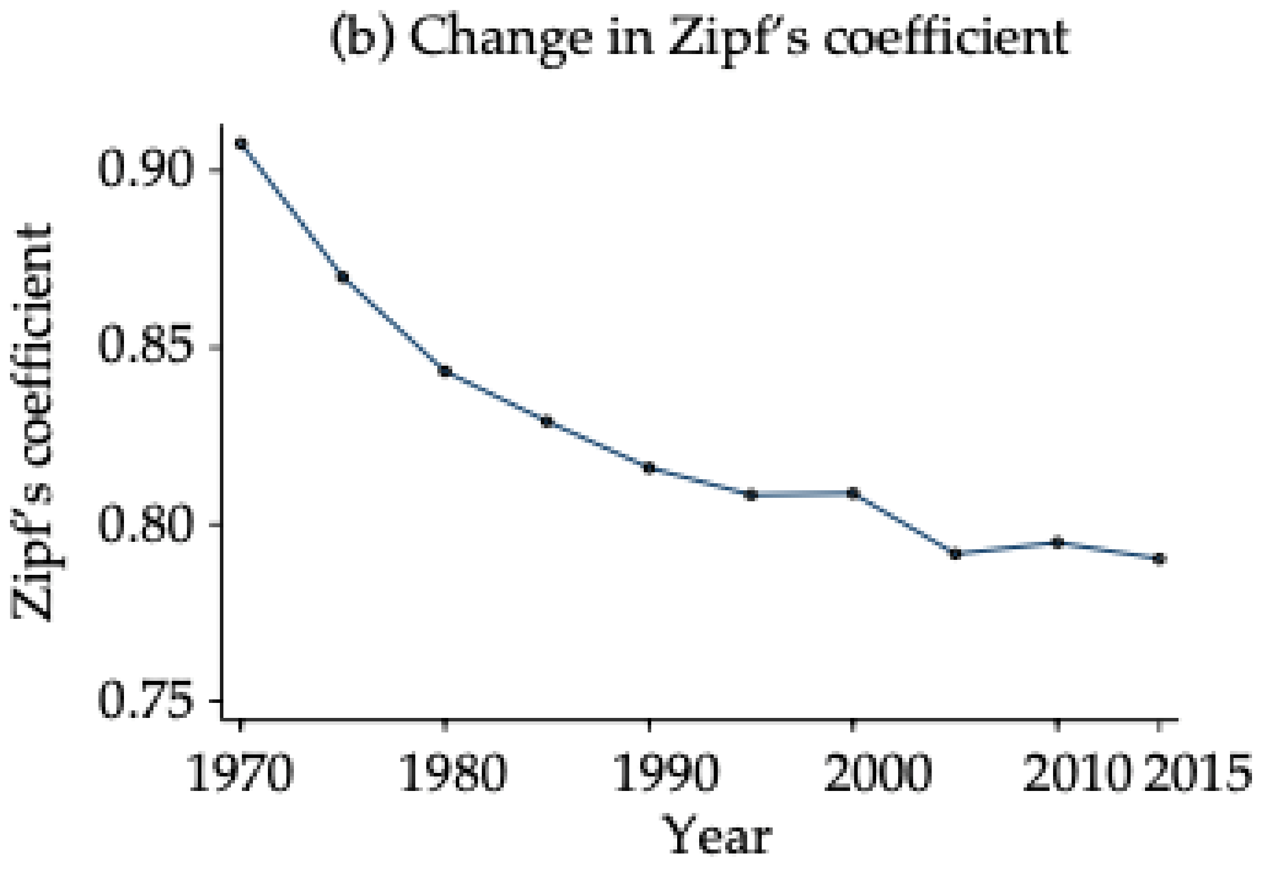
Mobility is a phenomenon that can occur at different scales in **space** and **time**



WHAT DO WE KNOW ABOUT MOBILITY?

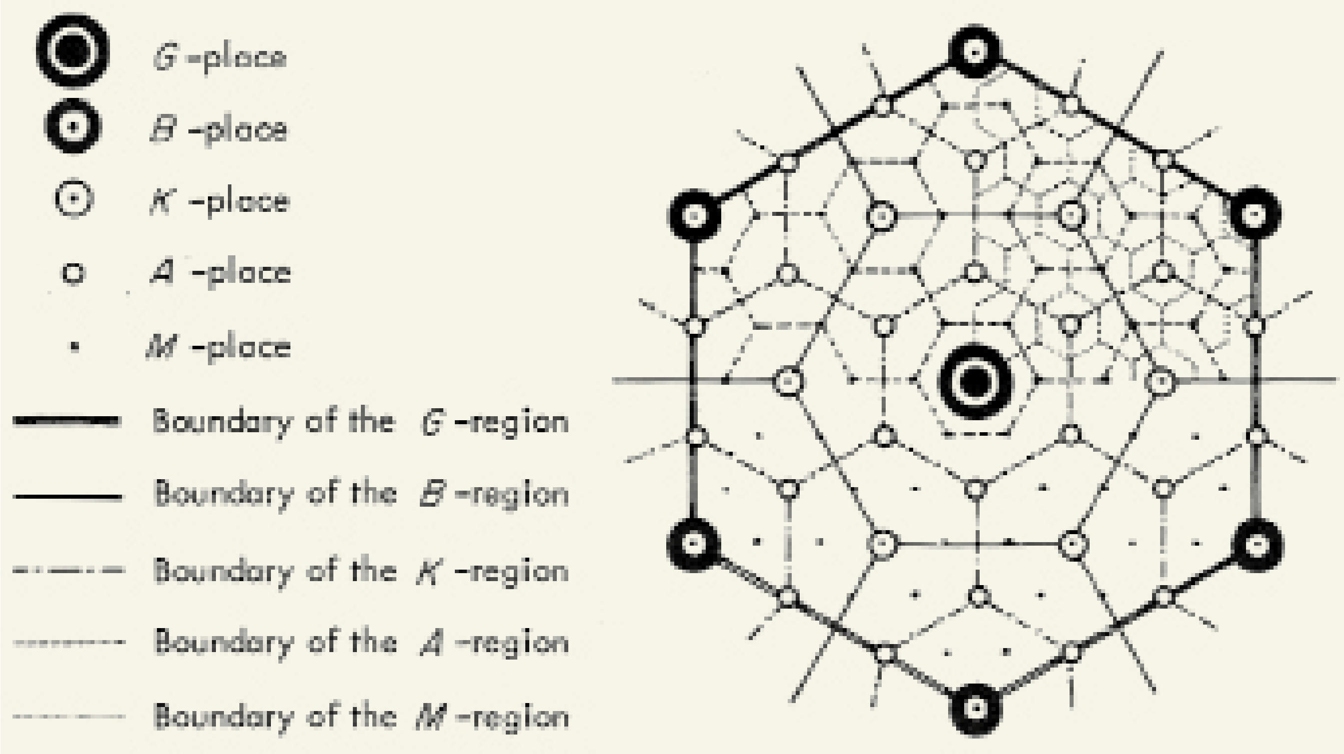
Zipf-Gibrat's law (1930s)

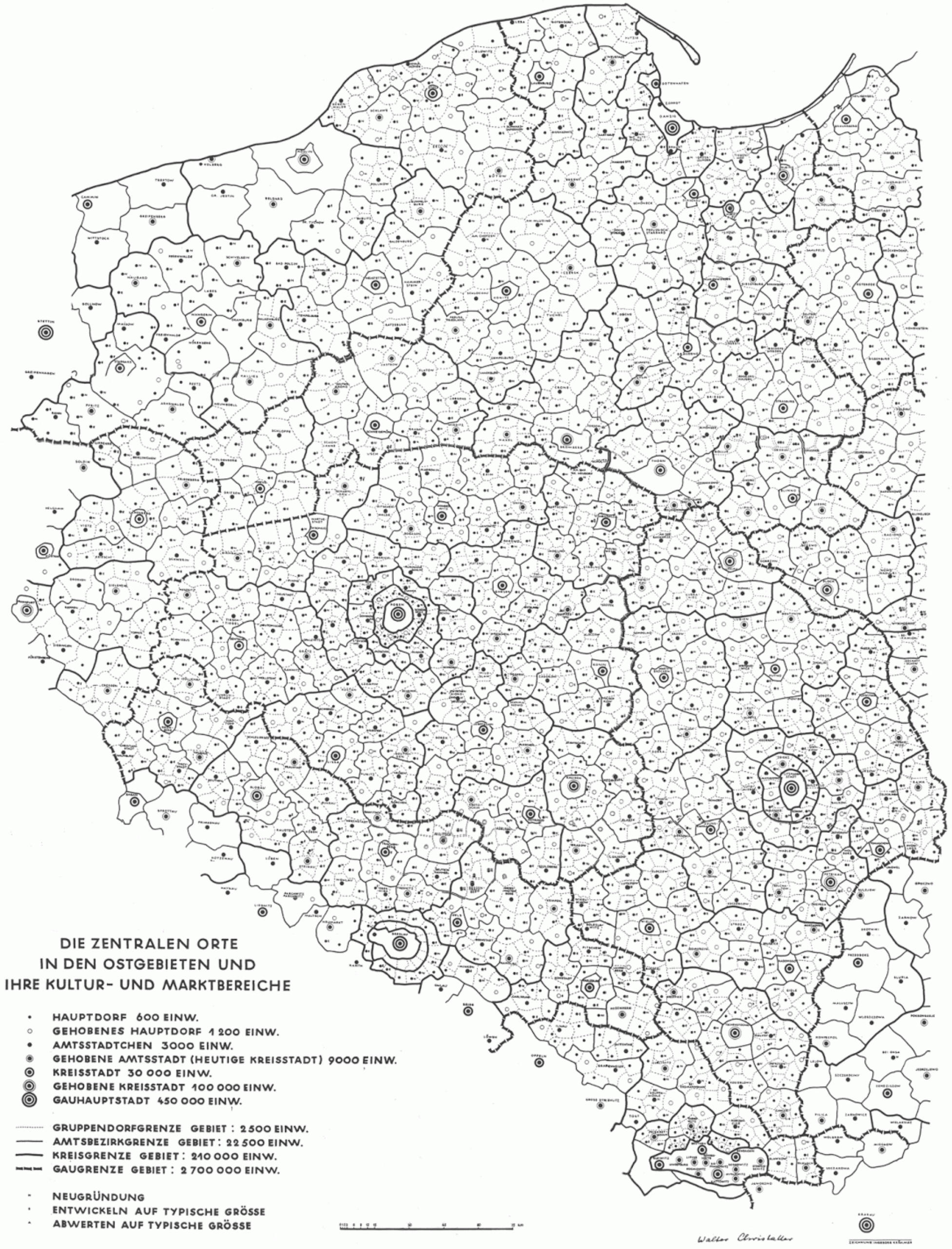




WHAT DO WE KNOW ABOUT MOBILITY?

Central Place Theory (Christaller, 1930s)





WHAT DO WE KNOW ABOUT MOBILITY?

Space syntax (Hillier, 1970s)

WHAT DO WE KNOW ABOUT MOBILITY?

Traditional forms of data collection about mobility include:

- ✓ Census
- ✓ Travel surveys
- ✓ Small-scale travel diaries/observations
- ✓ …

CENSUS DATA

When: Collected every several years (typically 10)

Pros:

- \+ covers the entire population
- \+ exhaustive socio-economic profile of travelers

Cons:

- very high cost
- updated only every several years
- can be used only to track only long-term movement (migrations)

TRAVEL SURVEYS

When: At regular intervals (a few years); on-purpose collection

Pros:

- \+ information on travel mode/reason for travel
- \+ good socio-economic profile of travelers

Cons:

- \- high cost
- \- limited coverage (few thousands travelers at most)
- \- inaccurate information about travel habits

SMALL SCALE DIARIES/OBSERVATION STUDIES

When: on-purpose collection

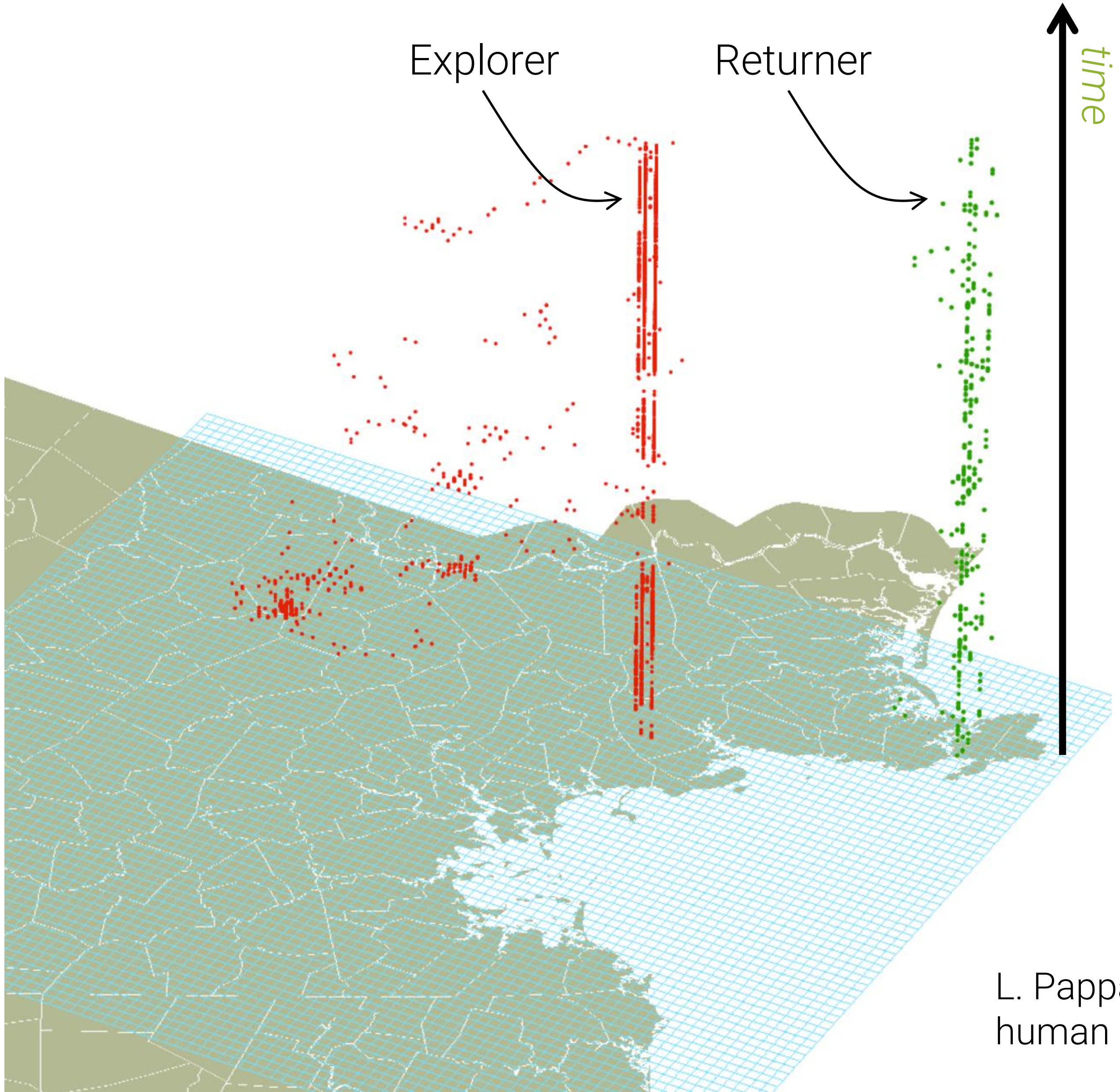
Pros:

+ information on travel mode/reason for travel
+ some socio-economic profile of travelers available

Cons:

- high cost
- very limited coverage (few hundreds travelers at most)

LARGE SCALE, HIGH RESOLUTION MOBILITY TRACKING



L. Pappalardo et al., "Returners and explorers dichotomy in human mobility", *Nature Communications*, 2015

NEW MOBILITY DATA

New forms of data collection about mobility include:

- ✓ Cell phone data
- ✓ GPS
- ✓ Flow counters
- ✓ Head counters
- ✓ …

MOBILITY AND PRIVACY

CELL PHONE DATA SETS

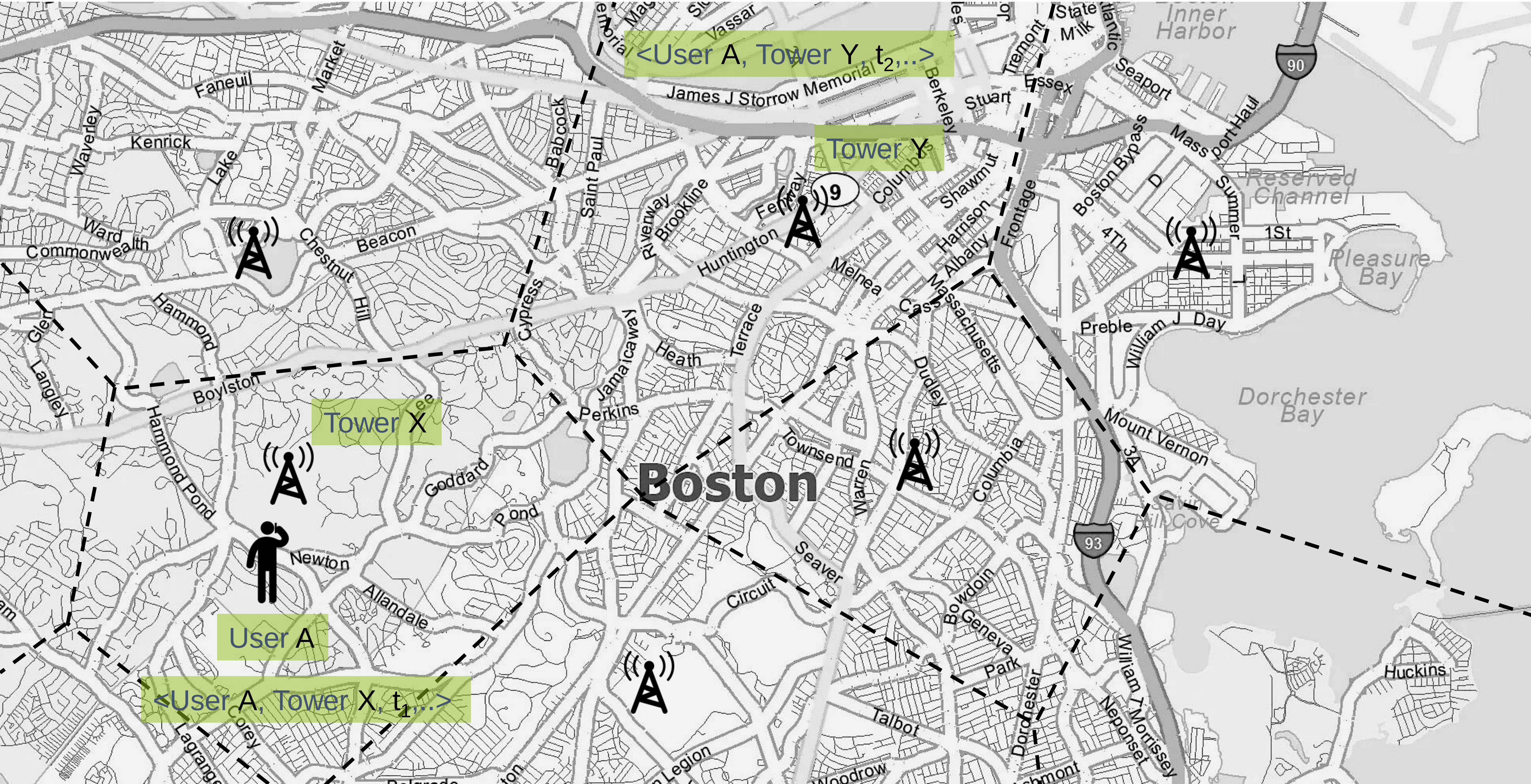
Call Detail Records

- ✓ Collected for billing purposes
- ✓ Typical content:
  - ID of caller and receiver
  - Call start time and duration
  - Call type (voice, text, etc.)
  - ID of cell tower the caller/receiver is associated with

Network signaling data

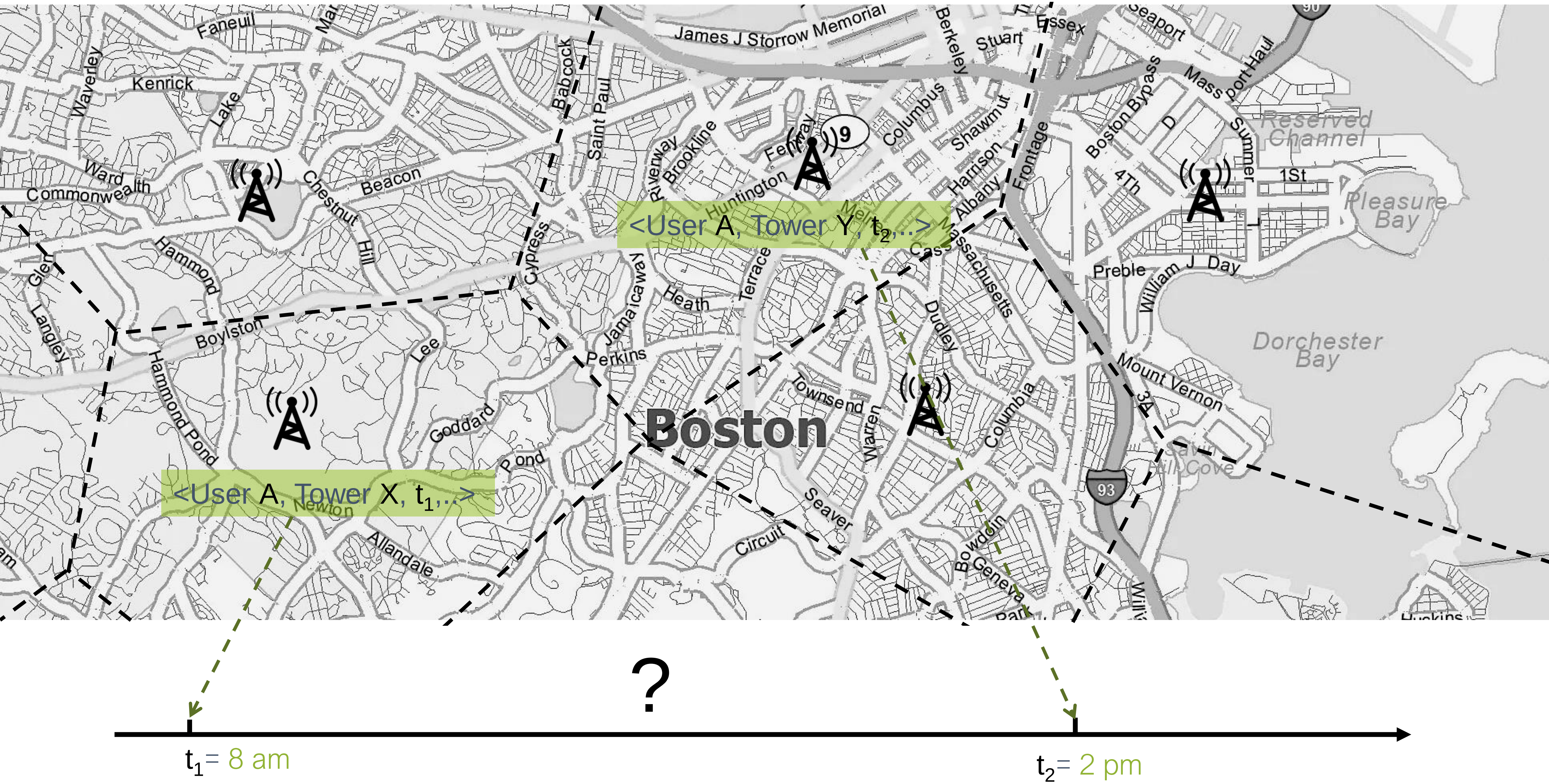
- ✓ Collected for keeping track of a user in the network
- ✓ Typical content:
  - ID of tracked user
  - Event type (includes data connections)
  - Event time
  - ID of cell tower the user is associated with

FROM CELL PHONE DATA TO MOBILITY TRACES



SPATIAL GRANULARITY

TEMPORAL GRANULARITY



CELL PHONE DATA SETS

When: opportunistic collection

Pros:

- + very good coverage (hundred thousands/million users)
- + good spatial and temporal granularity
- + record real movements of huge number of users

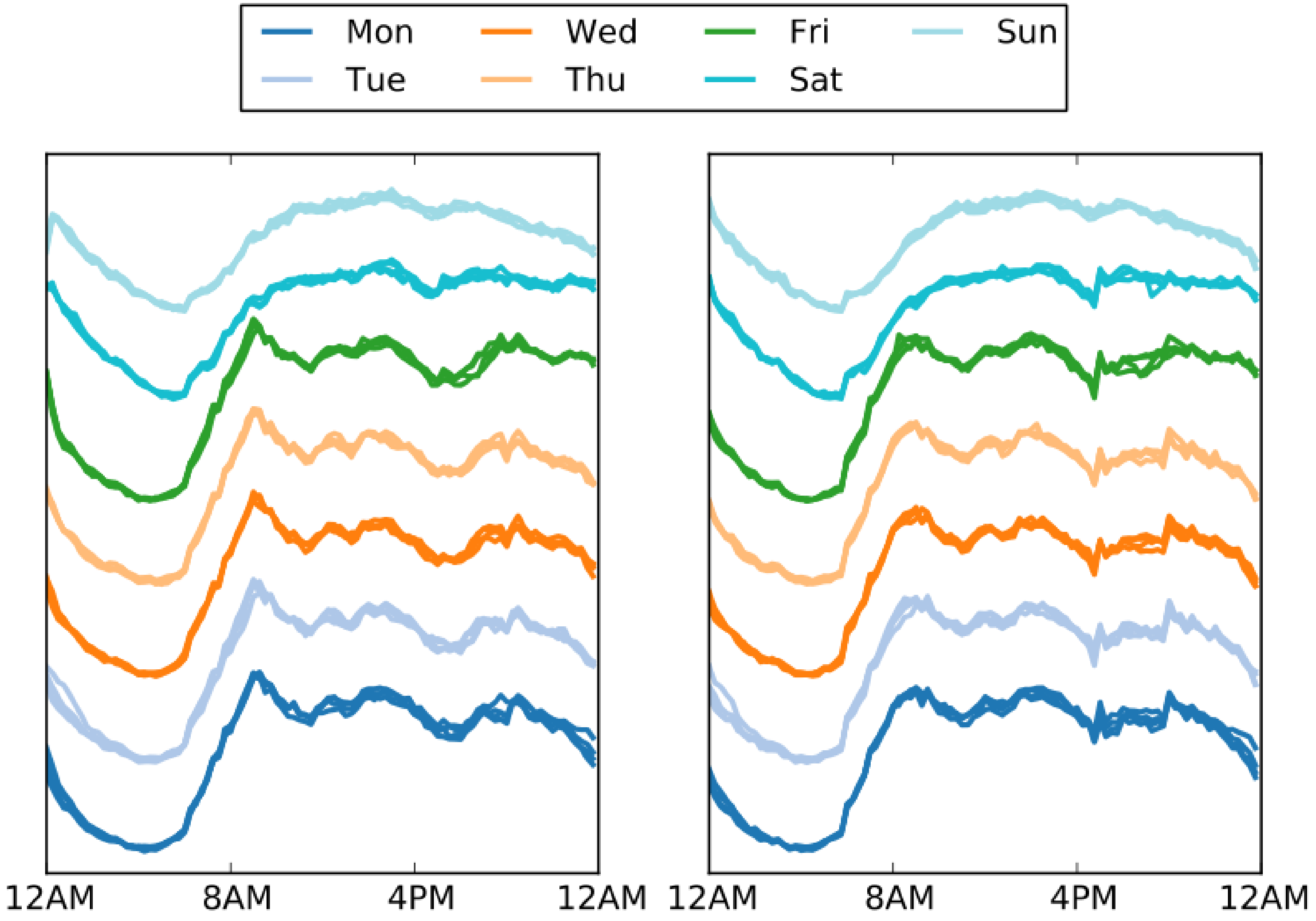
Cons:

- - difficult to obtain
- - large but non representative sample of the population
- - little/no socio-economic profile of users

CDR VS NETWORK SIGNALING DATA

feature	CDR	NetSign
spatial granularity	cell tower	cell tower
temporal granularity	Few to hundred	hundreds
home detection	Yes	Yes
work detection	Yes	Yes
Detect single movements	No	Yes
Trajectory reconstruction (cell)	No	Yes
Travel time detection	No	≃
Travel mode detection	No	≃
Travel purpose	No	No

CELL PHONE DATA AND MOBILITY PATTERNS



GPS DATA SETS

Fleet movement data

- ✓ Collected for billing/tracking purposes
- ✓ Typical content:
  - ID of vehicle
  - Trip start time and (lat,long) location
  - Trip end time and (lat,long) location
  - Trip info

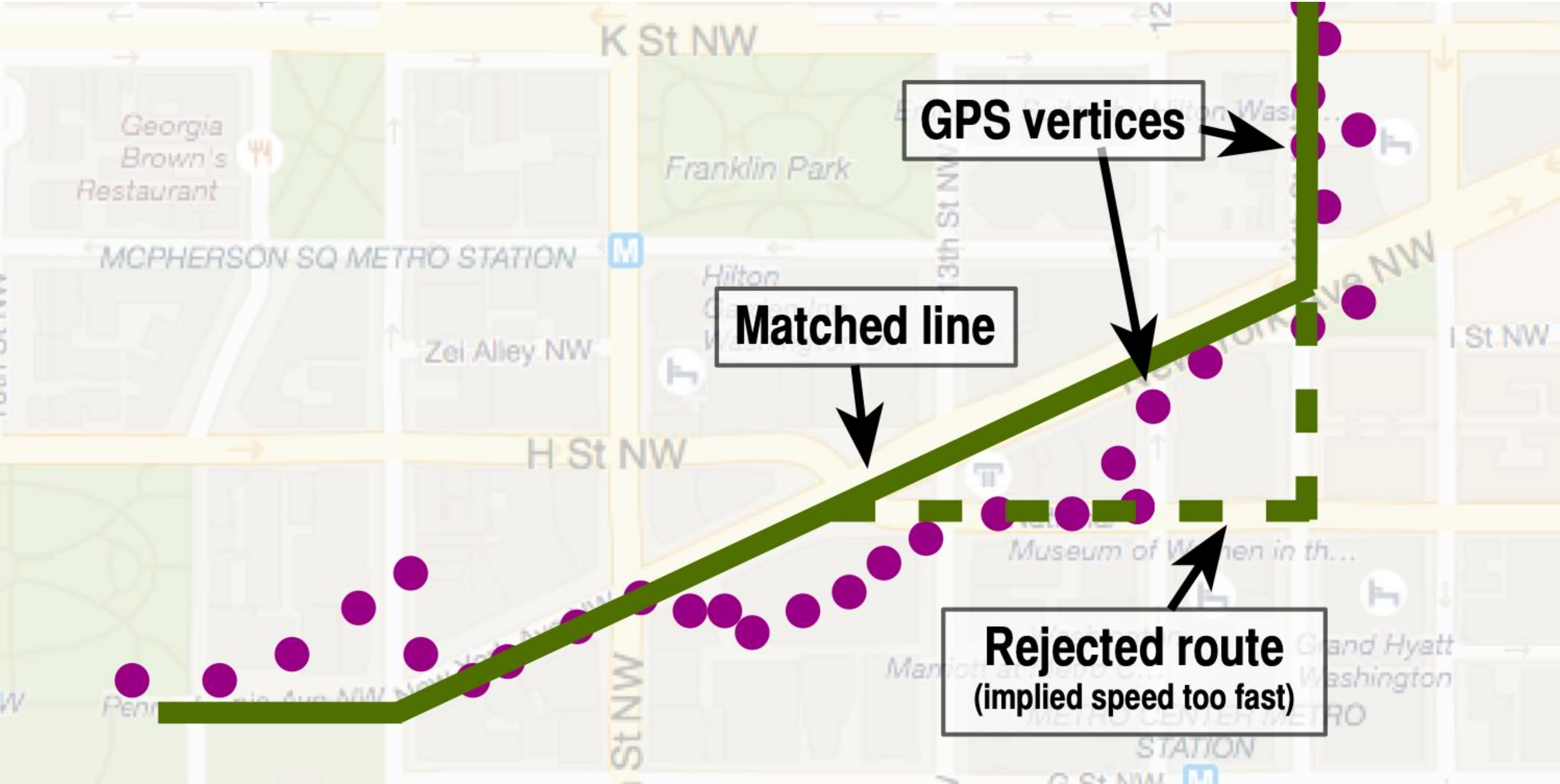
Cell-phone location-based data

- ✓ Collected by location-based apps
- ✓ Typical content:
  - User ID
  - Event type
  - Event time and (lat,long) location

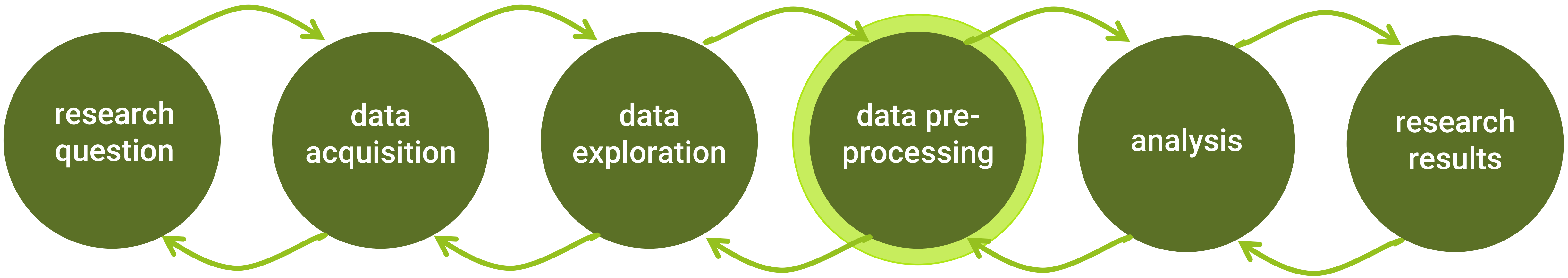
CDR VS GPS DATA

feature	CDR	GPS
spatial granularity	cell tower	location ± 10m
temporal granularity	Several mins/h	1 sec
home detection	Yes	Yes
work detection	Yes	Yes
Detect single movements	No	Yes
Trajectory reconstruction (cell)	No	Yes
Travel time detection	No	Yes
Travel mode detection	No	Yes
Travel purpose	No	≃

GPS DATA AND MAP MATCHING



DATA PROCESSING WORK FLOW

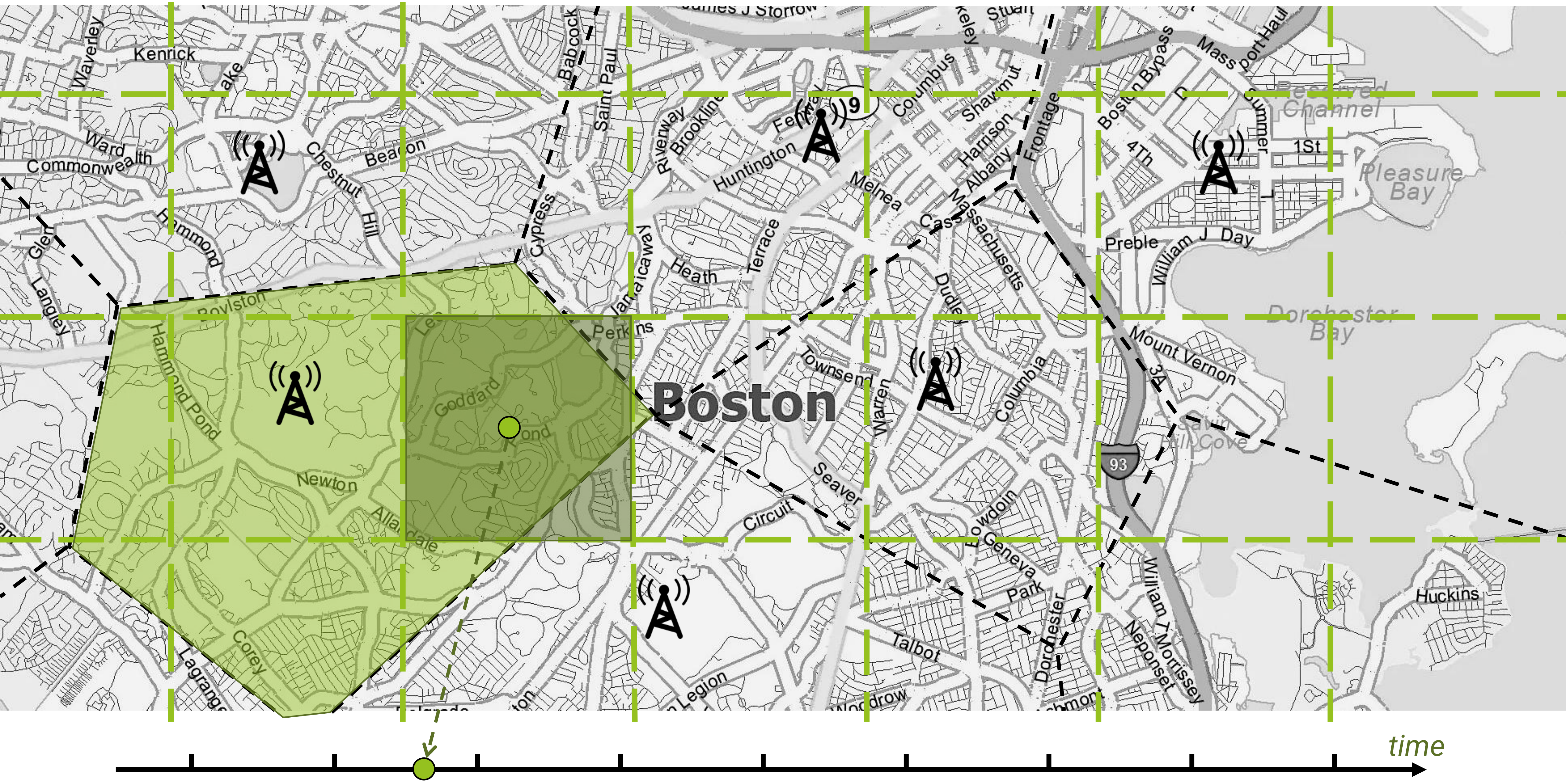


DATA PRE-PROCESSING

Typical data pre-processing steps

- ✓ Remove noisy/incomplete/inconsistent data, e.g:
  - Records where start/end of a trip is missing
  - Data points in “inconsistent” areas (water, forests, etc.)
  - Records corresponding to “impossible” trips (e.g., a trip with an excessively high speed)
- ✓ Select a subset of the original sample, e.g.:
  - Users with at least **x** CDR events in a day
  - Trips that start/end in a specific area
  - Users for which home location can be detected

SPATIO-TEMPORAL BINNING



Visitation Law

MOBILITY PATTERNS, CDR


PANIFICIO
BONCI
Le cose buone sono fatte
di cose buone...
e prodotte da persone
che hanno lo scrupolo di far
star bene chi le mangia.
Noi crediamo che la gastronomia
appartenga ancora all'agricoltura
e a coloro che riconoscono
le cose buone.
Usiamo in maniera scrupolosa
soltanto ingredienti prodotti da persone che
hanno questo ideale, mantengono e vogliono
mantenere alto il valore vero della
gastronomia e dell'alimentazione.
Noi siamo così....
SEMPLICEMENTE
NATURALI
Crediamo nella TERRA...
...grande Madre Generatrice di
ogni risorsa Alimentare.
Nei Contadini i veri EROi
del Nostro Tempo

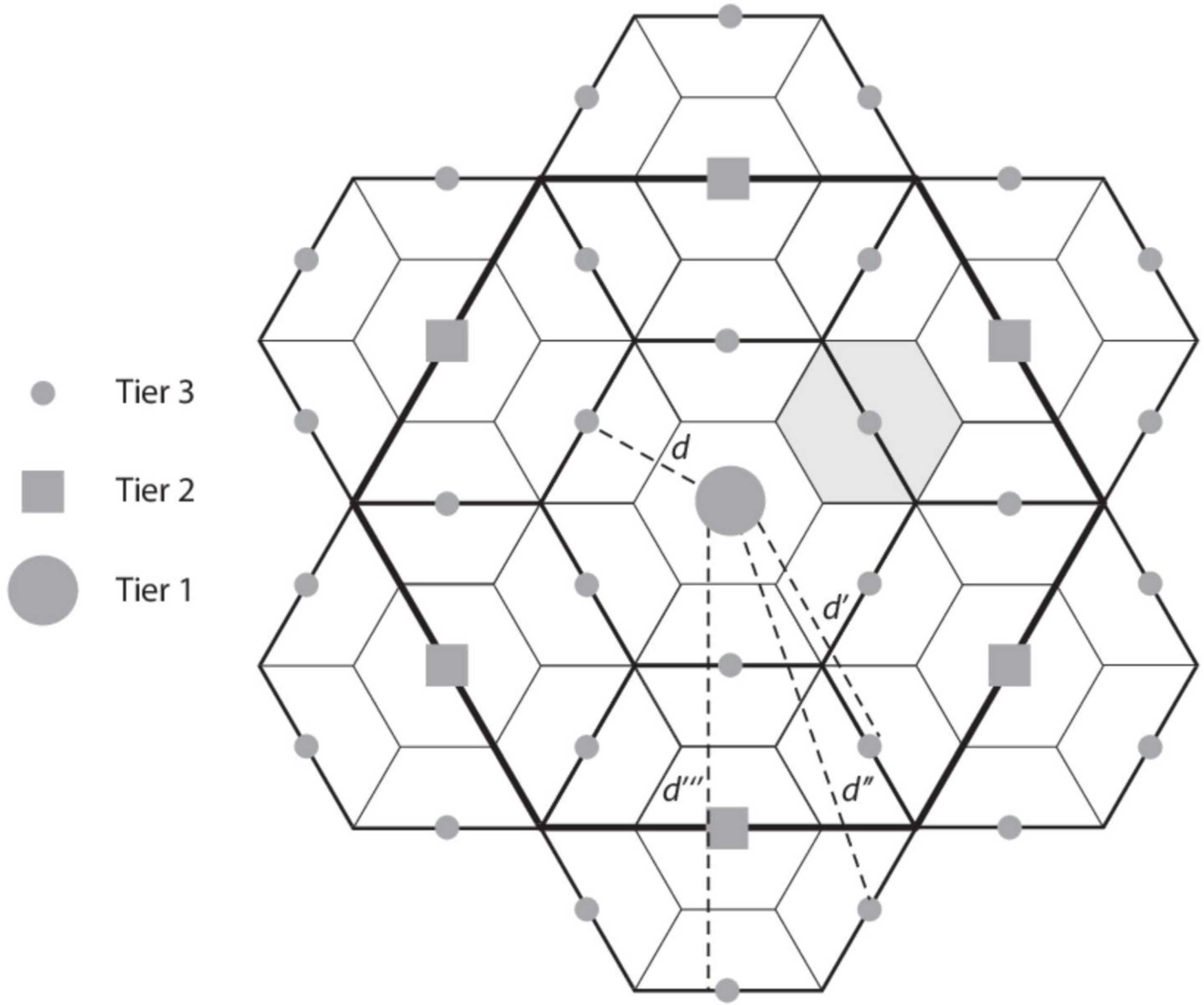
The Auto Mile – Norwood MA

THE INVERSE LAW

r - distance

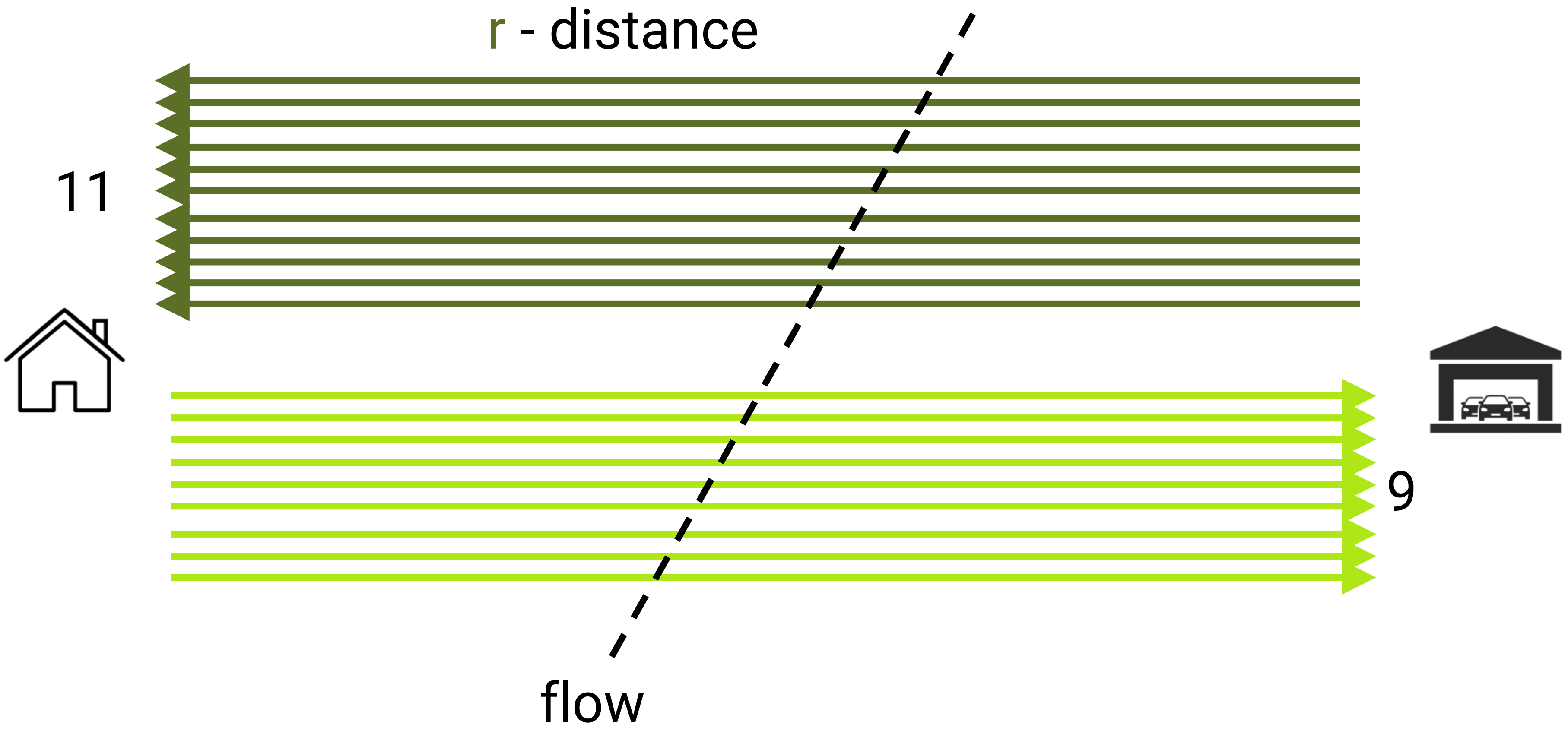
f - frequency

CENTRAL PLACE THEORY

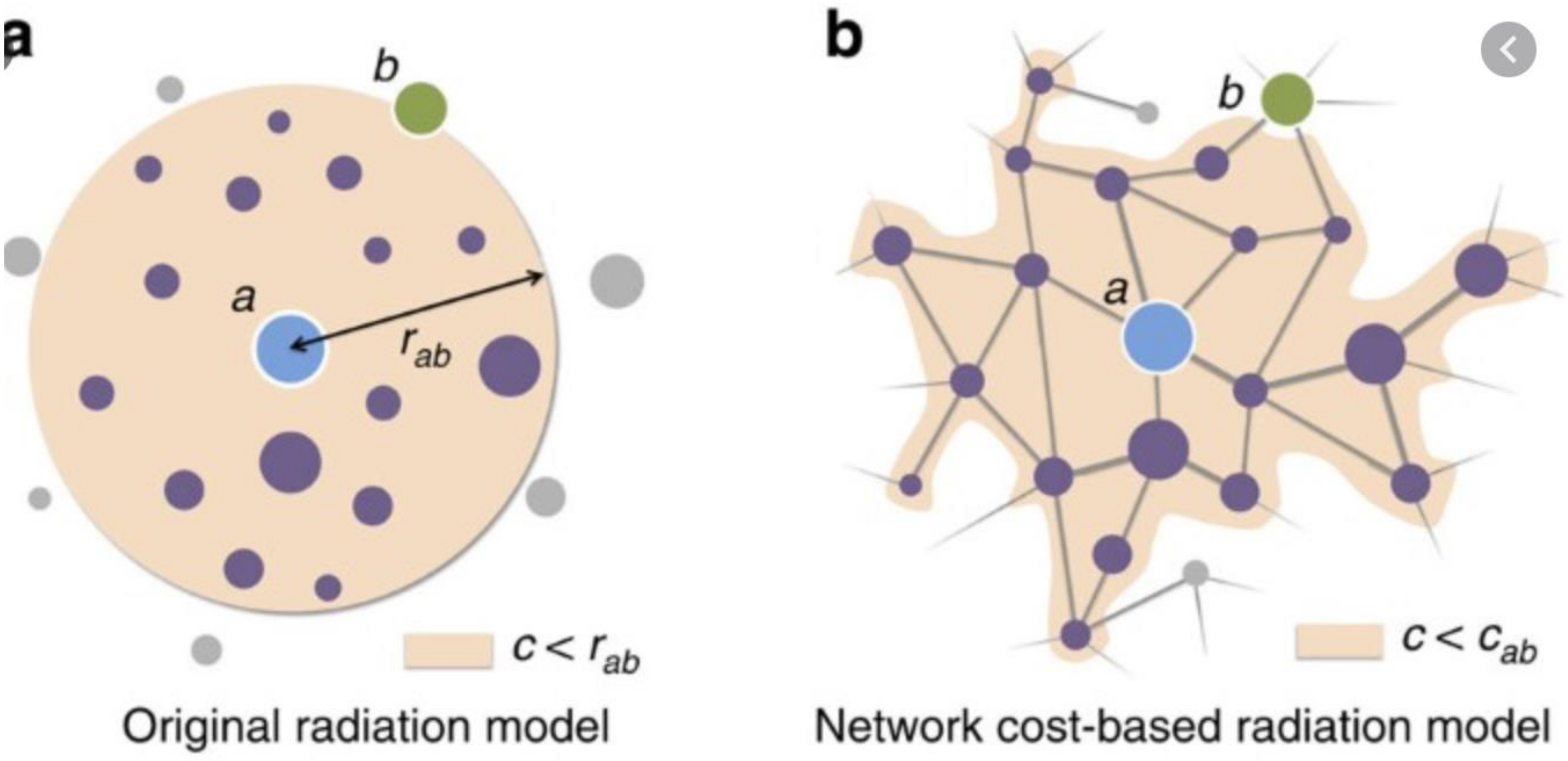


Christaller - 1930

THE INVERSE LAW: WHAT WE KNOW?

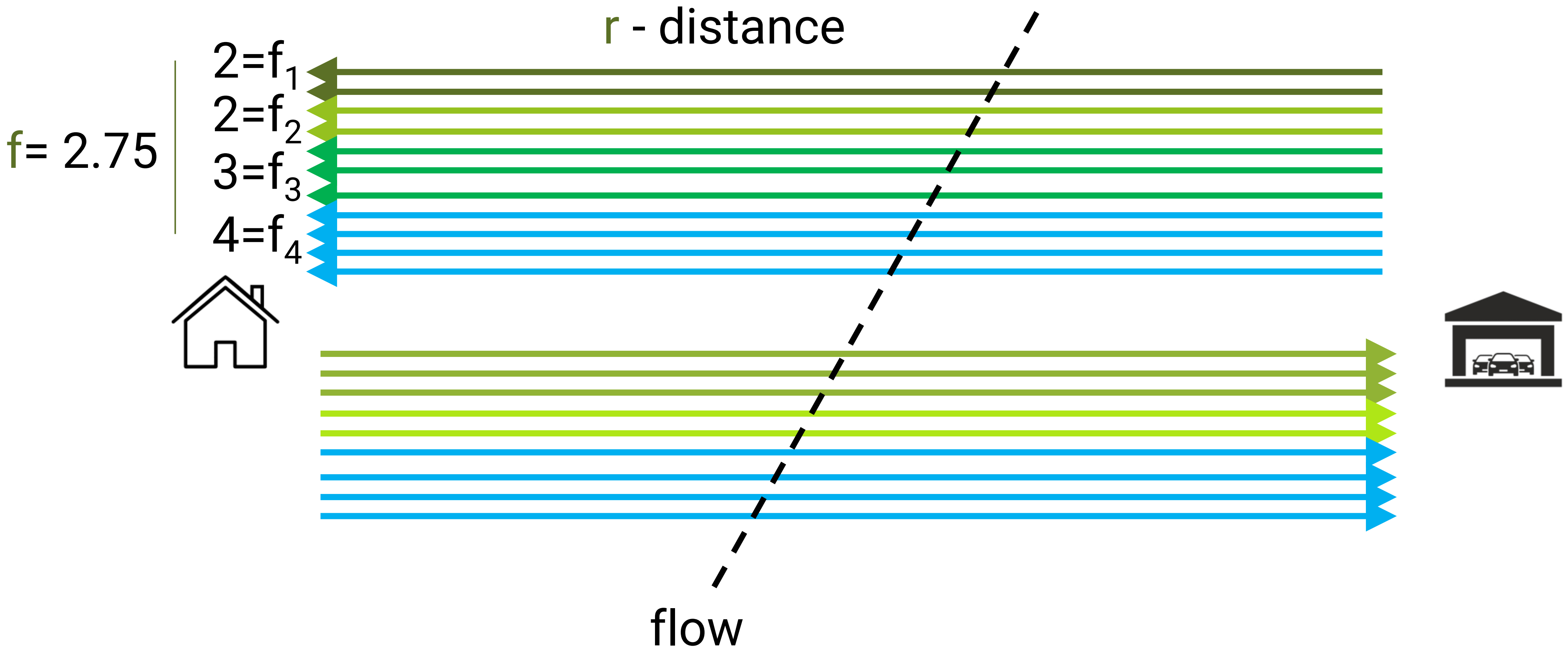


EXISTING MODELS



Radiation model – (Simini et al, Nature 2012)

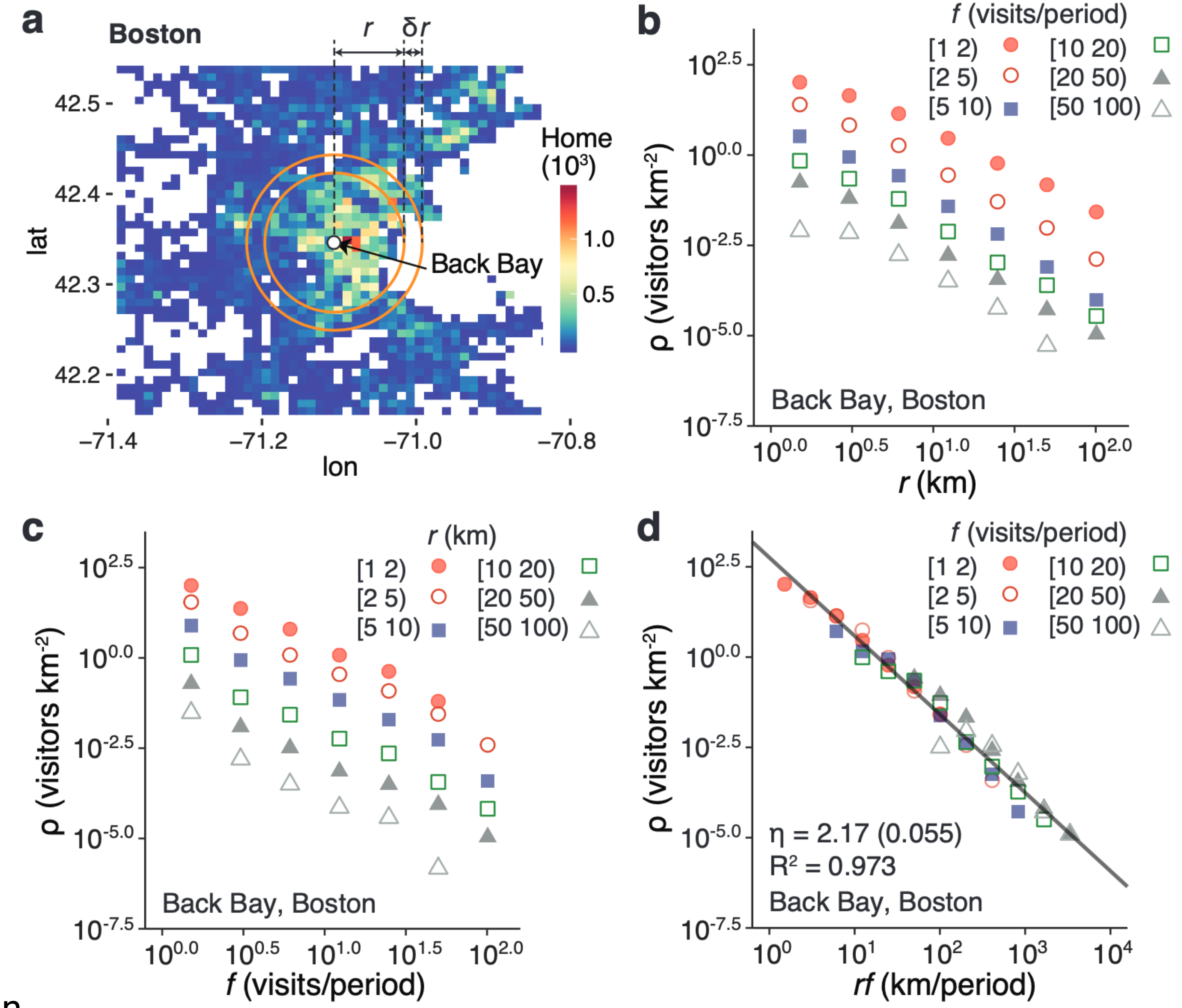
THE INVERSE LAW: WHAT ABOUT FREQUENCY?



THE INVERSE LAW IS REAL

$$\rho_i(r, f) = \frac{\mu_i}{(r f)^\eta},$$

η = 2



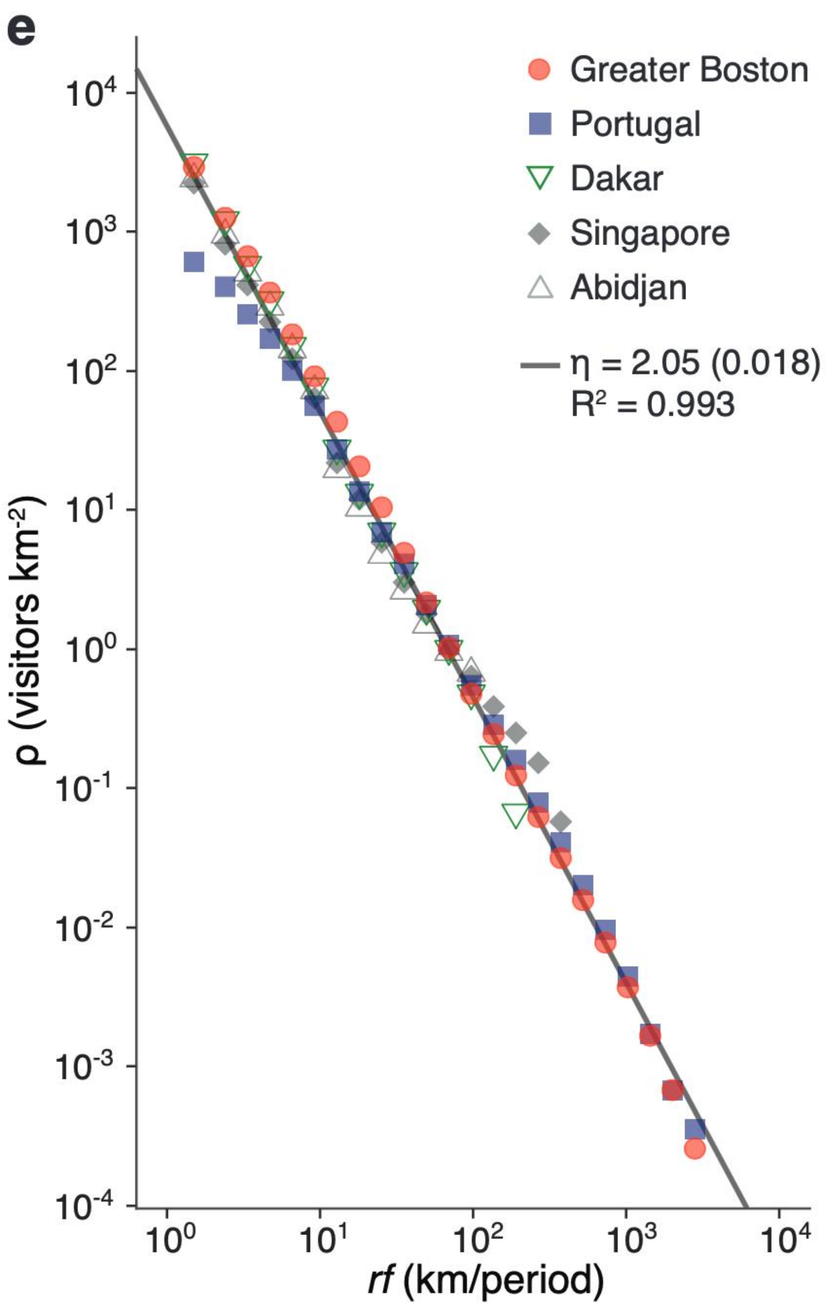
based on analysis of 1M people in Boston

..AND UNIVERSAL

$$\rho_i(r, f) = \frac{\mu_i}{(rf)^\eta},$$

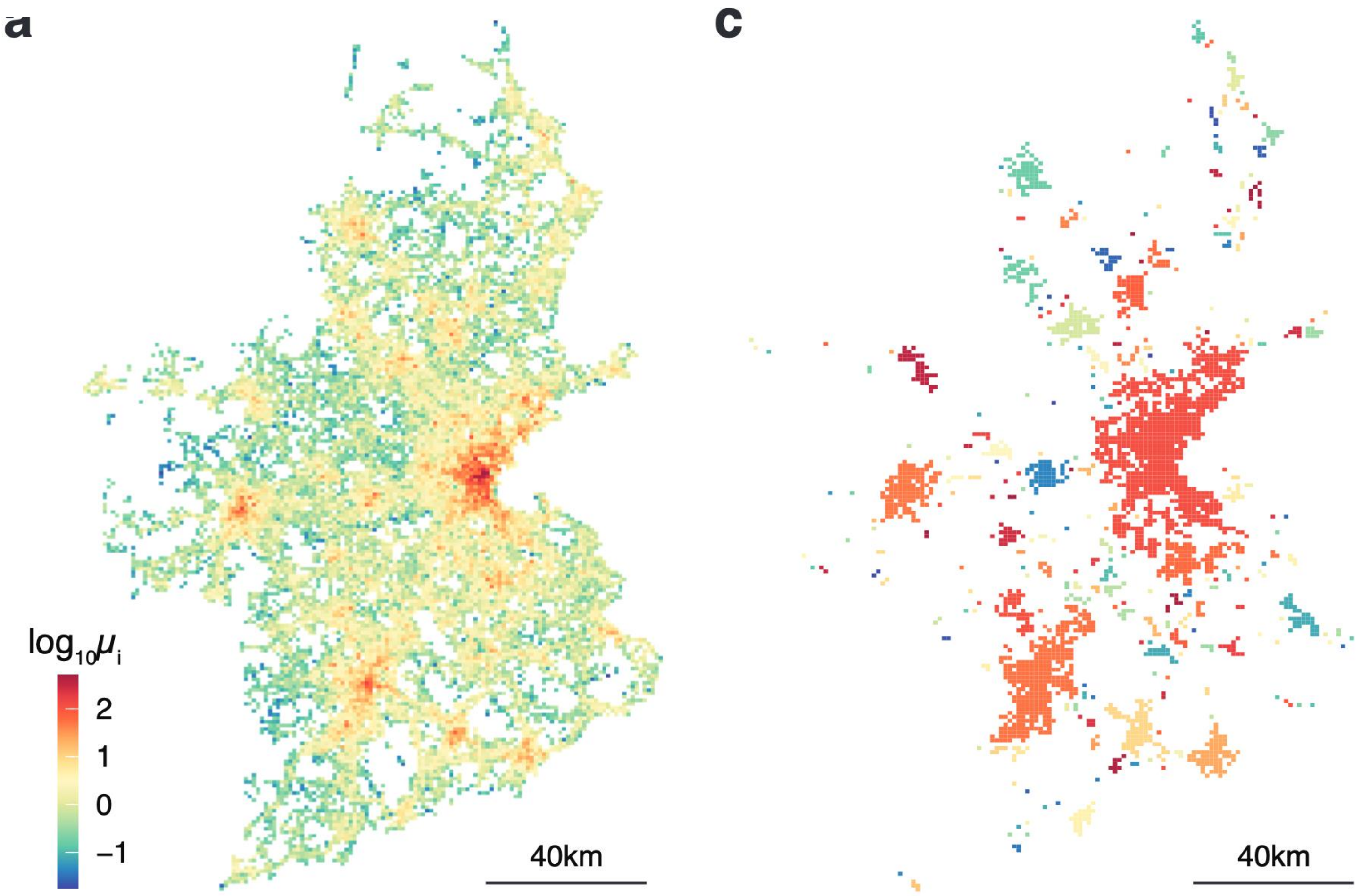
$\eta = 2$

based on analysis of 4M people in Boston, Portugal, Dakar, Abidjan, Singapore



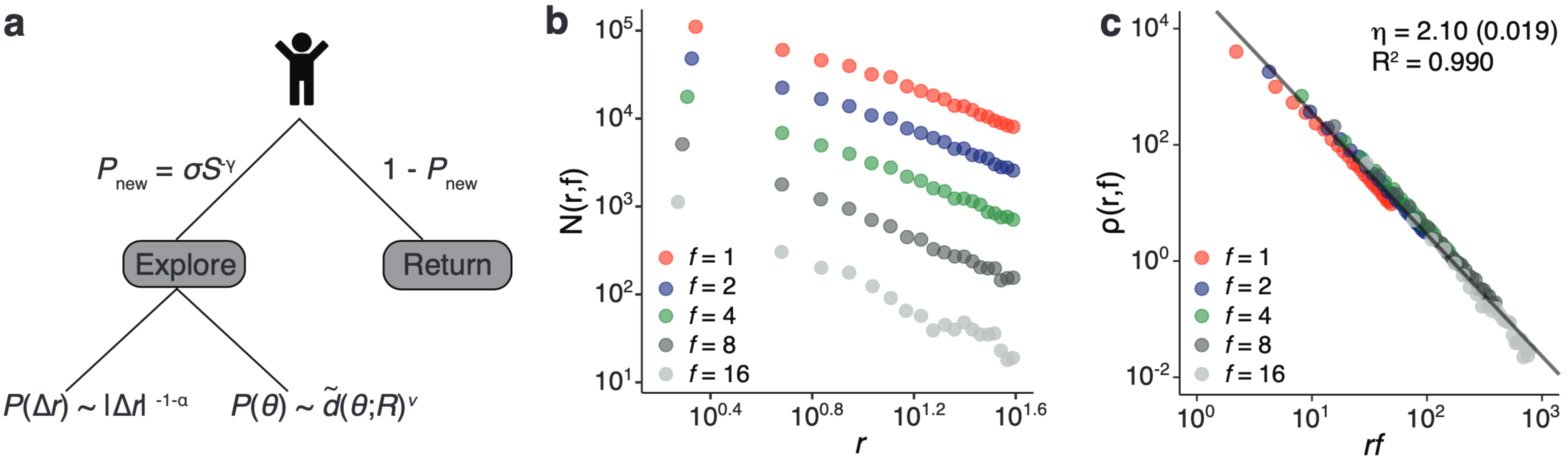
UNIVERSAL URBAN MOBILITY LAW

CENTRAL PLACE THEORY? CONFIRMED!



Christaller - 1930

PREFERENTIAL EXPLORATION



APPLICATIONS?

Traffic optimization

APPLICATIONS?

Real estate and commercial development

EPIDEMICS and MOBILITY

MOBILITY PATTERNS, PUBLIC HEALTH

INTERPLAY BETWEEN MOBILITY AND DISEASE

Articles

Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data

Ying Zhou, Renzhe Xu, Dongsheng Hu, Yang Yue, Qingquan Li, Jizhe Xia

Summary

Background Restricting human mobility is an effective strategy used to control disease spread. However, whether mobility restriction is a proportional response to control the ongoing COVID-19 pandemic is unclear. We aimed to develop a model that can quantify the potential effects of various intracity mobility restrictions on the spread of COVID-19.

Methods In this modelling study, we used anonymous and aggregated mobile phone sightings data to build a susceptible–exposed–infectious–recovered transmission model for COVID-19 based on the city of Shenzhen, China. We simulated how disease spread changed when we varied the type and magnitude of mobility restrictions in different transmission scenarios, with variables such as the basic reproductive number (R_0), length of infectious period, and the number of initial cases.

Lancet Digital Health 2020 2: e417–24

School of Public health, Shenzhen University Health Science Center, Shenzhen, China (Y Zhou PhD, D Hu PhD); Institute for Advanced Study (R Xu MSc) and Guangdong Key Laboratory for Urban Informatics, Department of Urban Informatics (Y Yu PhD,

KING PAPER SERIES

HUMAN MOBILITY RESTRICTIONS AND THE SPREAD OF THE NOVEL CORONAVIRUS (2019-NCOV) IN CHINA

Hanming Fang
Long Wang
Yang Yang

Working Paper 26906
http://www.nber.org/papers/w26906

www.nature.com/scientificreports

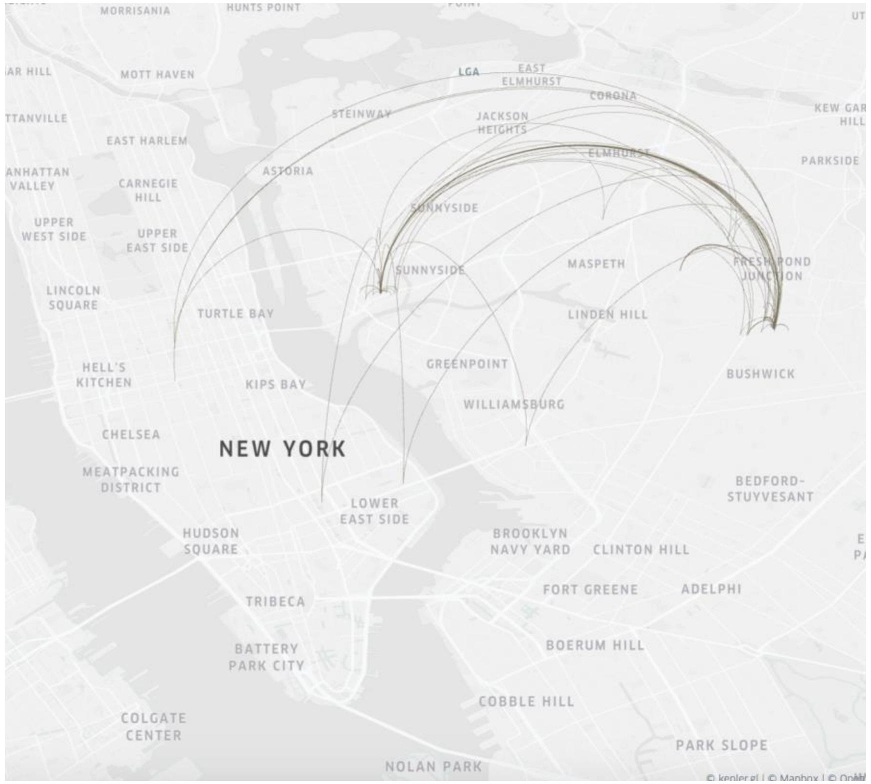
SCIENTIFIC REPORTS
natureresearch

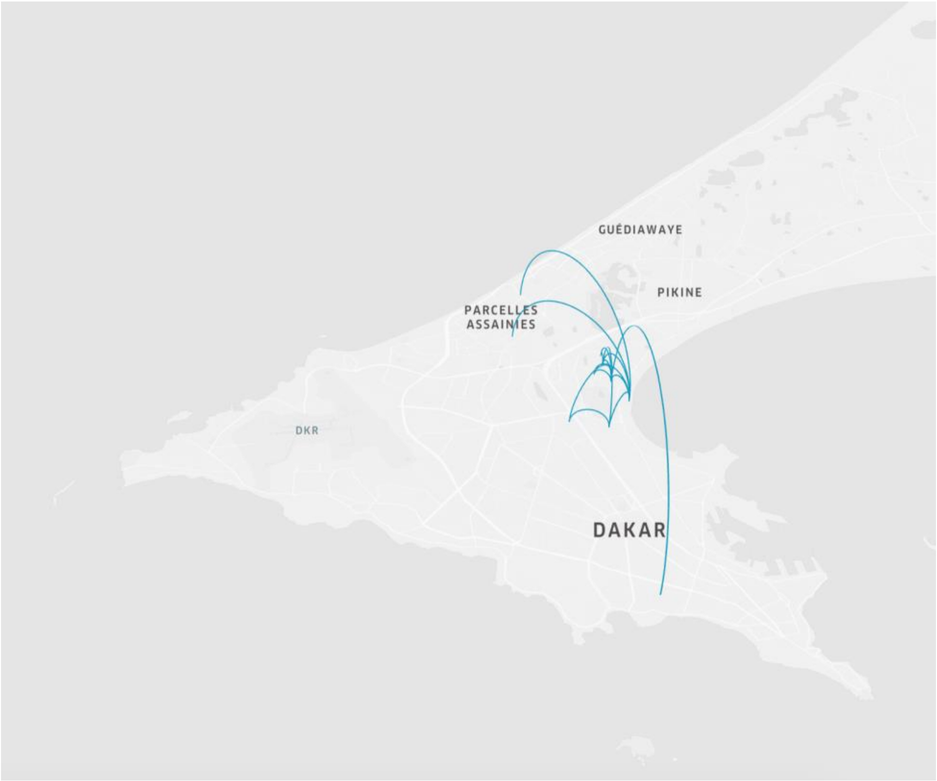
terplay between human mobility and mosquito borne diseases in urban environments

Emanuele Massaro[1*], Daniel Kondor[2] & Carlo Ratti[2,3]

SEIR EPIDEMIC MODELING

10,000 agents follow the real trajectories of mobile phone users in New York City and Dakar, Senegal





SEIR EPIDEMIC MODELING

As they encounter infected agents, they become exposed

Susceptible

β: Probability of becoming exposed to the disease over the course of one day
proportional to number of infected people within a given radius of you

Exposed

α: Probability of becoming infected on any given day after exposure
= 1/average latency period

Infected

γ: Probability of recovery on any given day after becoming infectious
= 1/average infection length

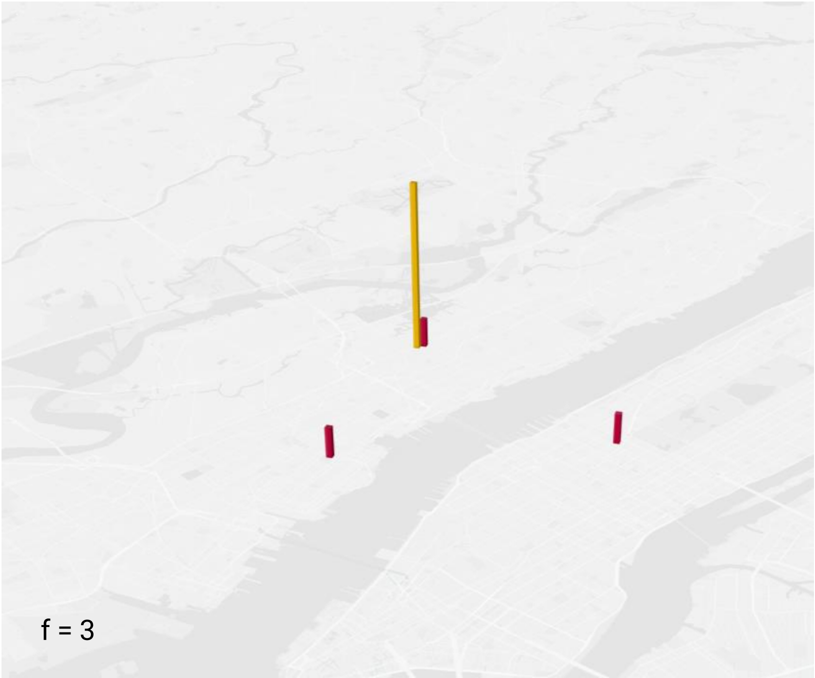
Recovered

R_0 = 3.58, Chen 2020

RESTRICTING R AND F

Within this framework, we can restrict radius of travel ***r***...

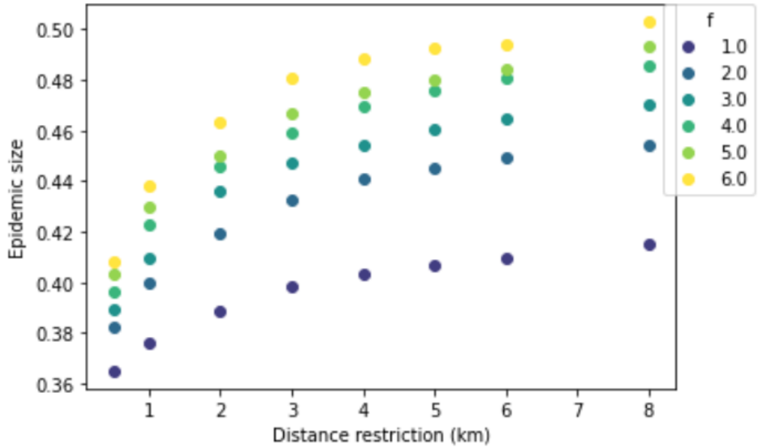
and frequency of return ***f***

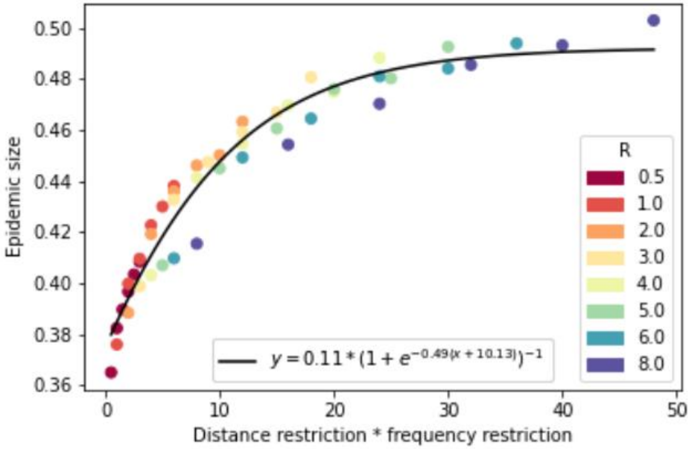


How does epidemic size after 30 days change with ***r*** and ***f***?
How does spatial diffusion of disease after 30 days change with ***r*** and ***f***?

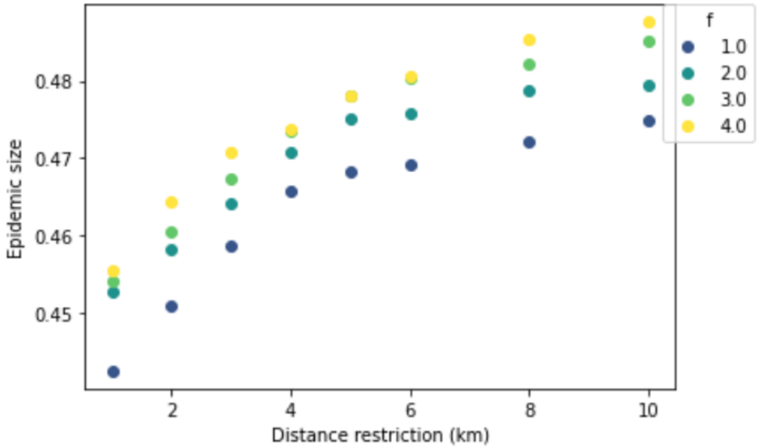
RESULTS: EPIDEMIC SIZE

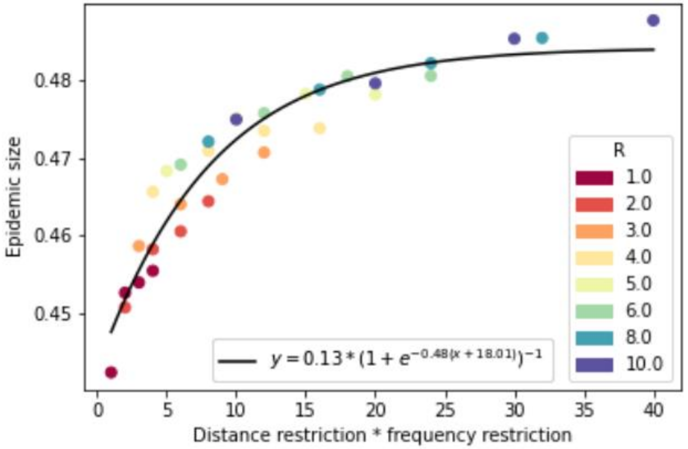
New York City



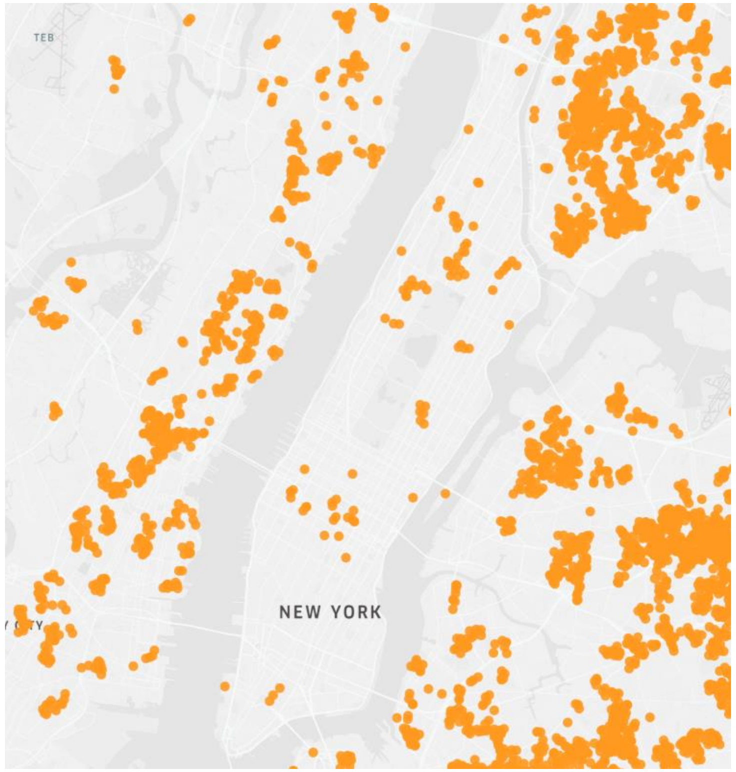


Dakar

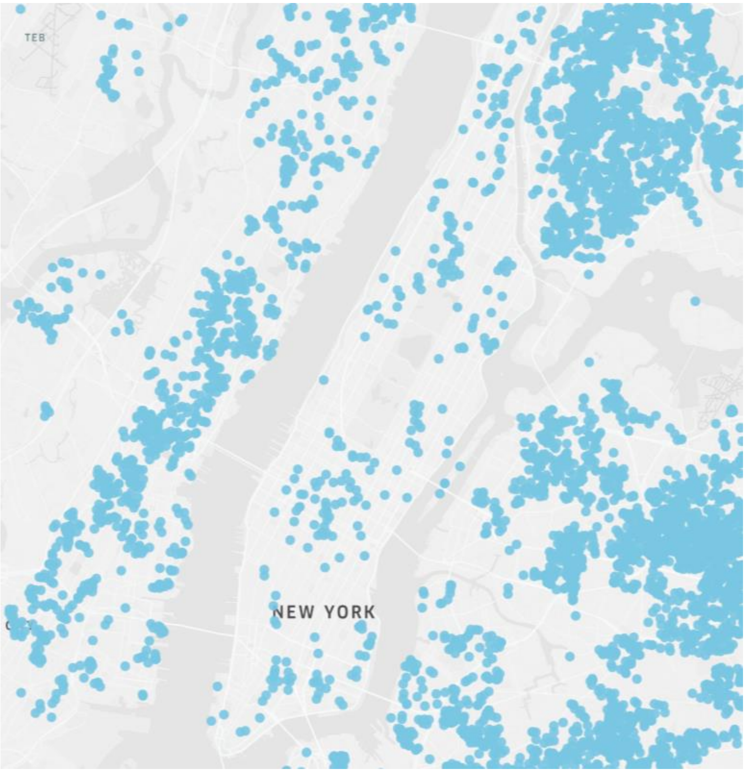




RESULTS: DISEASE DIFFUSION

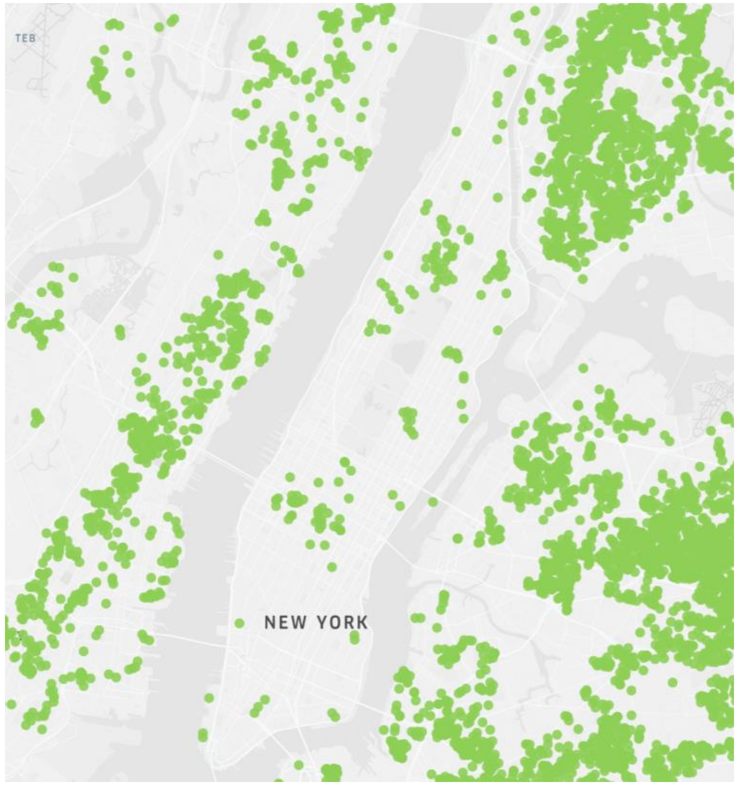


r = .5, *f* = 1
*r***f* = .5

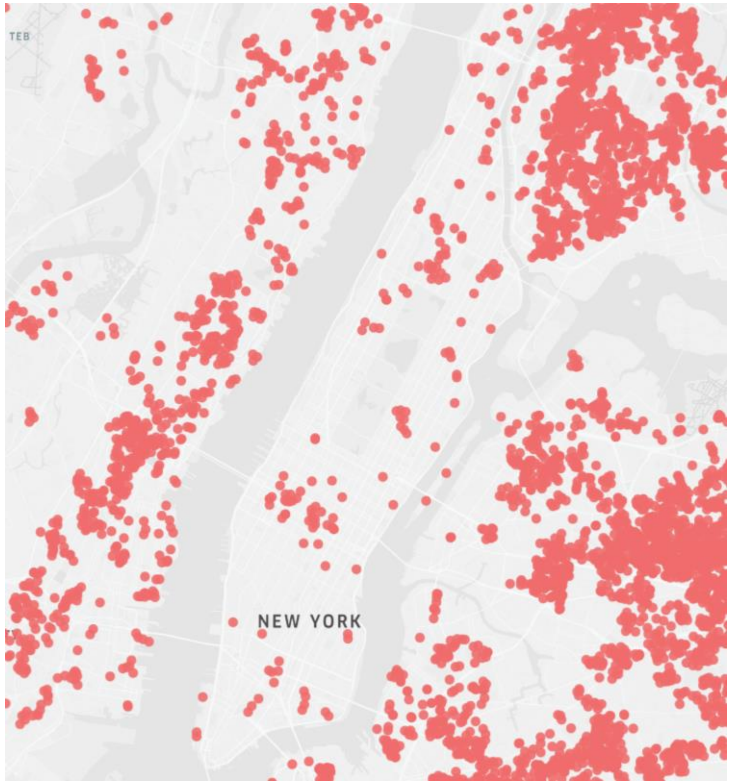


r = 6, *f* = 6
*r***f* = 36

RESULTS: DISEASE DIFFUSION

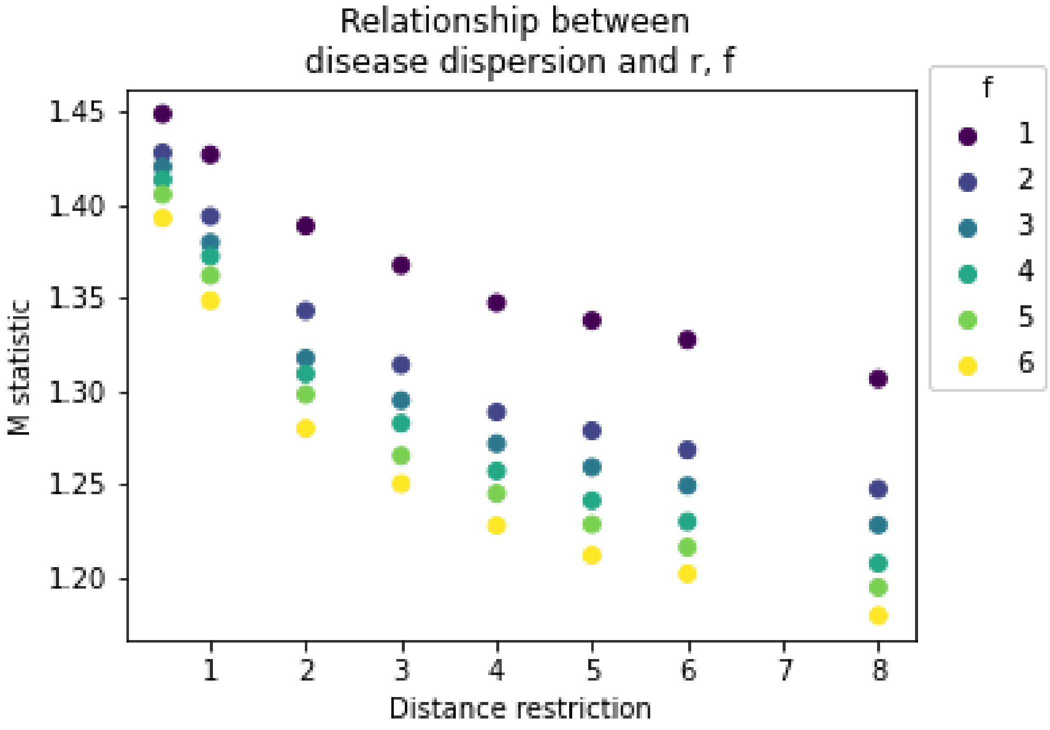


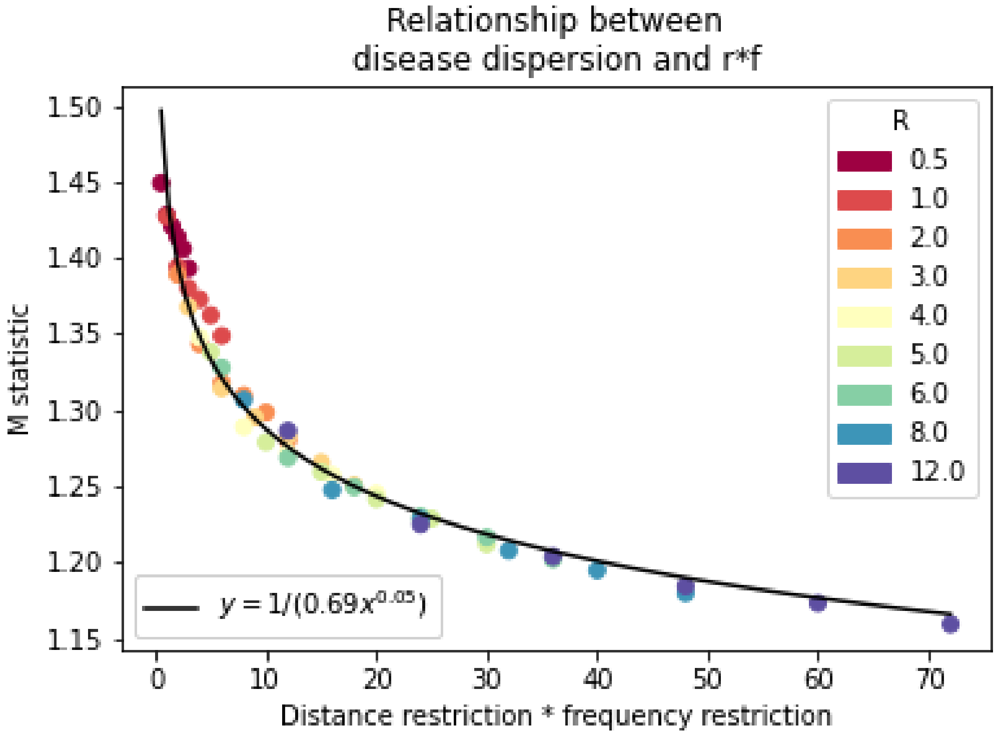
$r = 6, f = 2$
$r*f = 12$



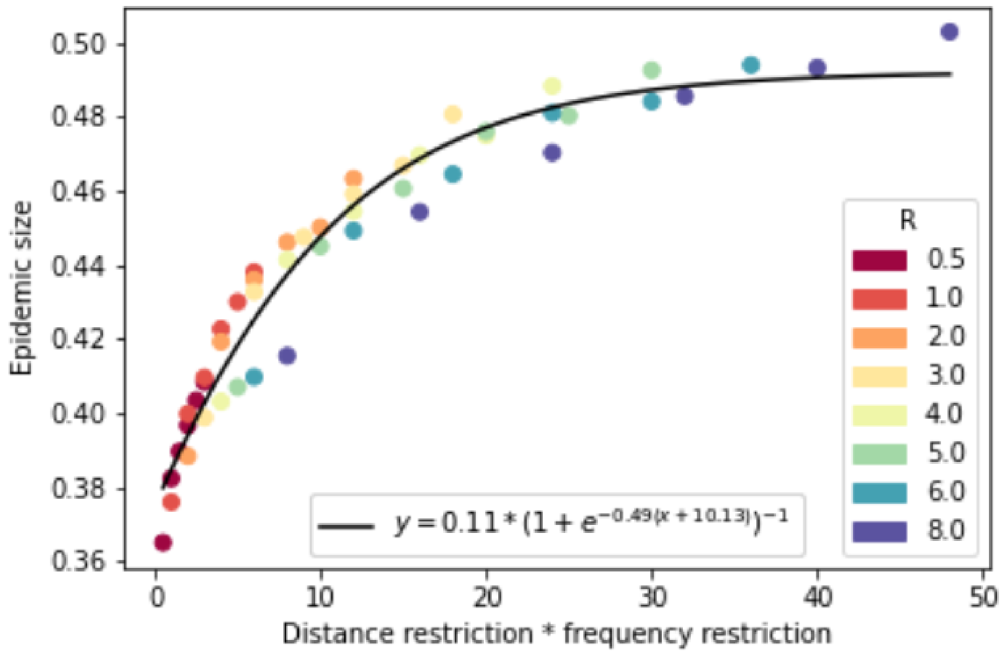
$r = 3, f = 4$
$r*f = 12$

RESULTS: DISEASE DIFFUSION

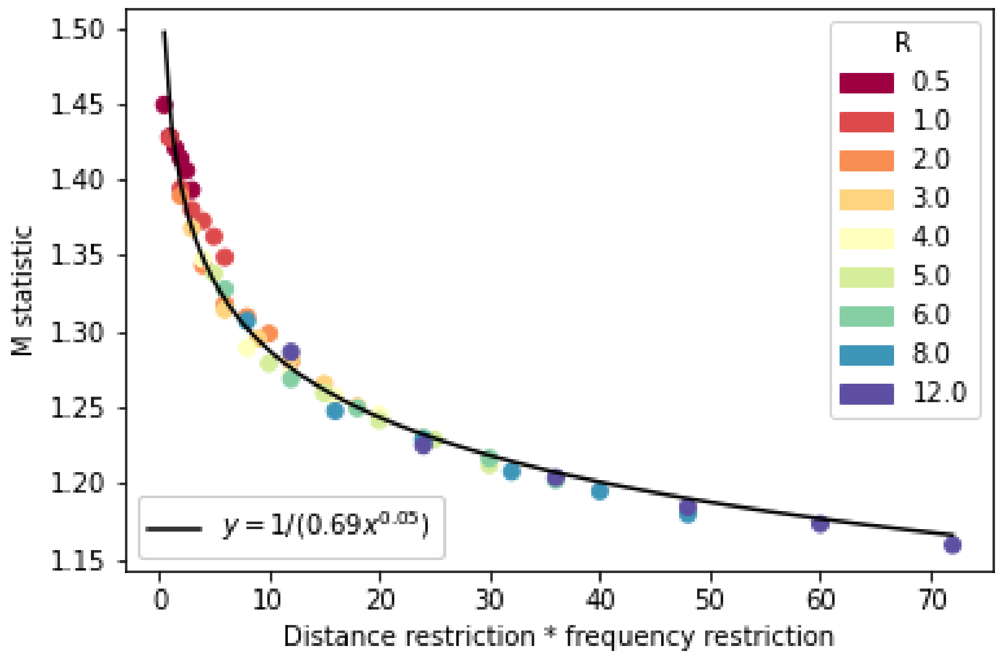




RESULTS



Epidemic size
Distance restriction * frequency restriction
R
0.5
1.0
2.0
3.0
4.0
5.0
6.0
8.0
$y = 0.11 * (1 + e^{-0.49(x+10.13)})^{-1}$

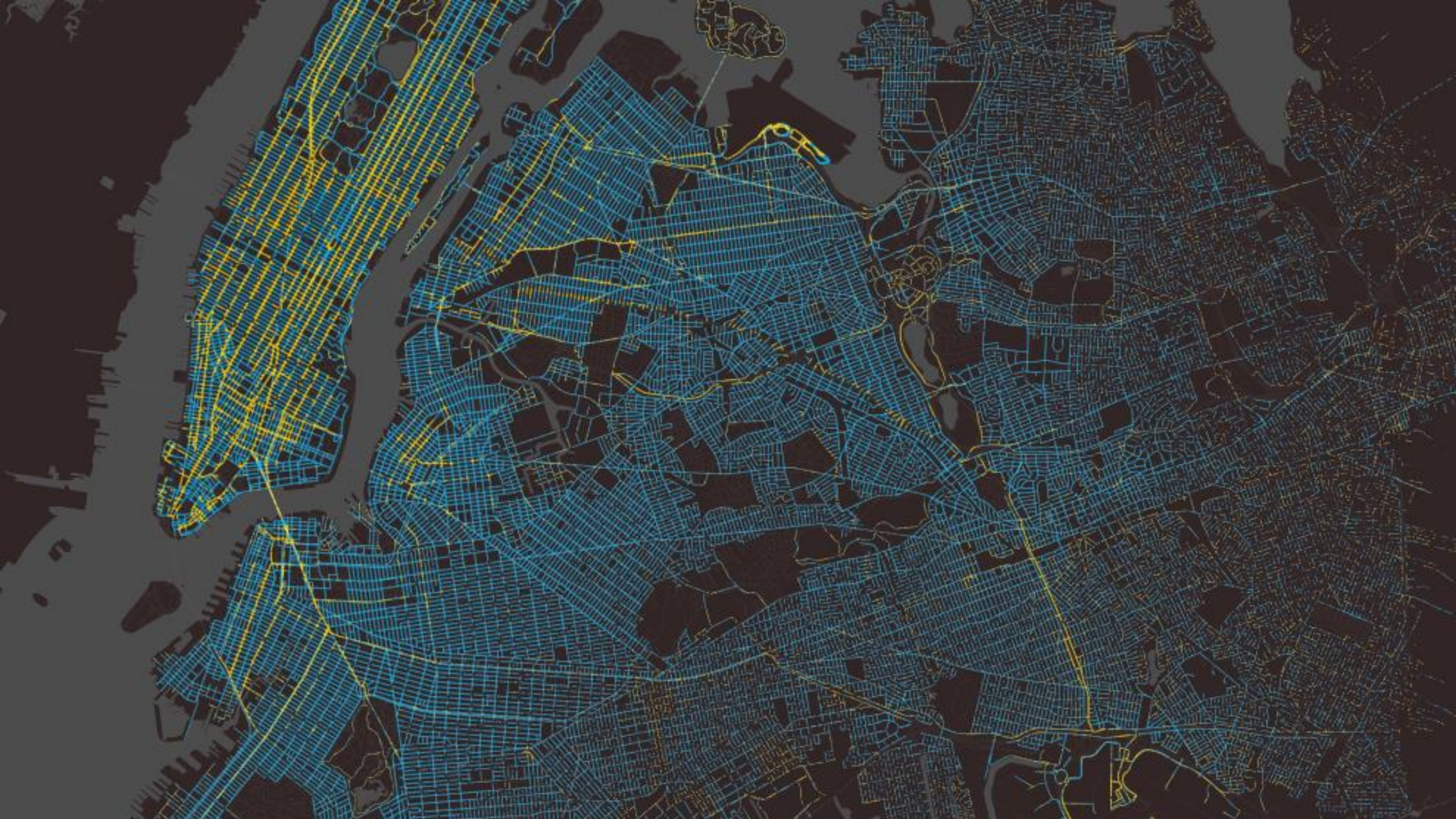


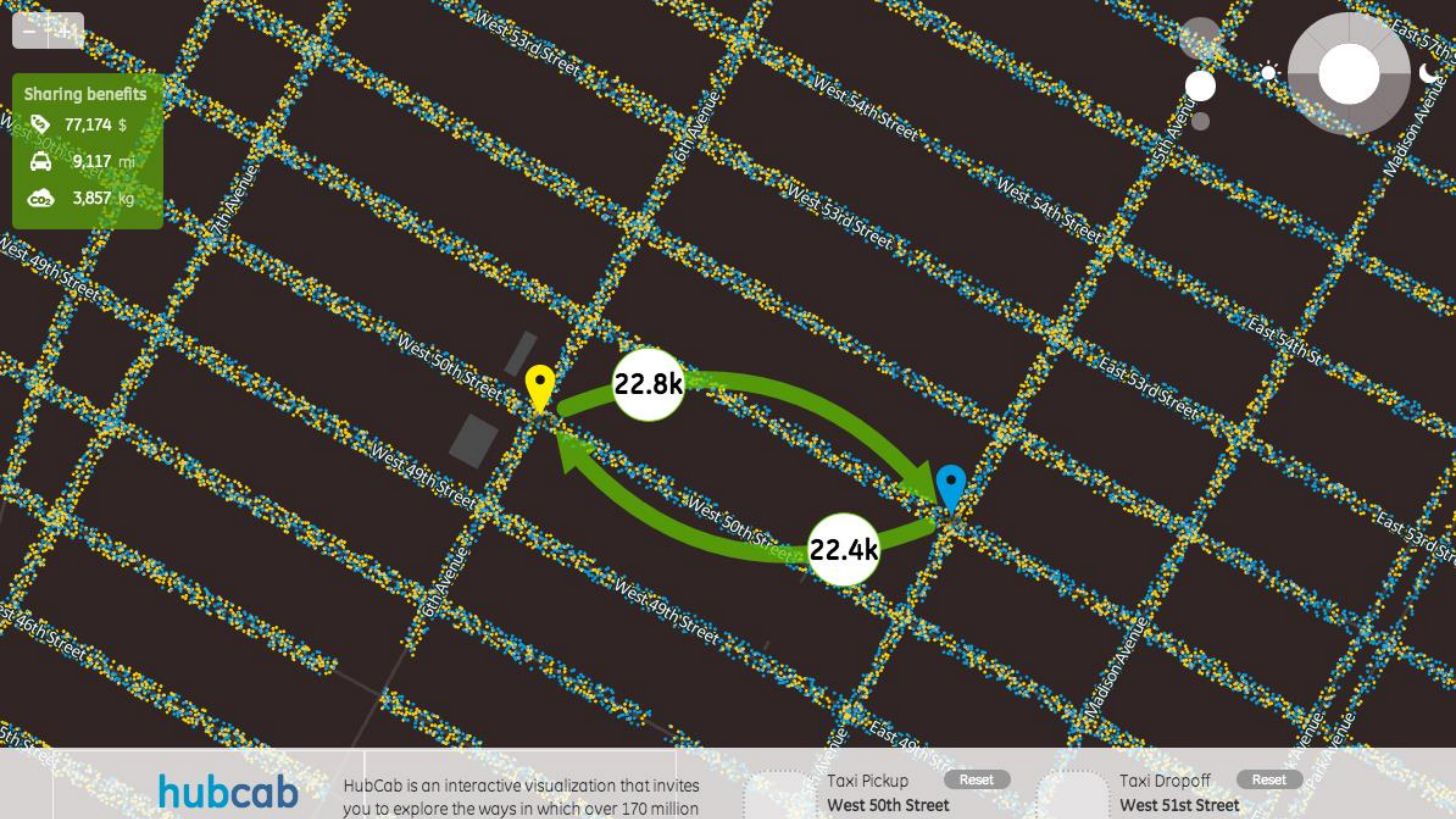
M statistic
Distance restriction * frequency restriction
R
0.5
1.0
2.0
3.0
4.0
5.0
6.0
8.0
12.0
$y = 1/(0.69x^{0.05})$

HubCab

MOBILITY, URBAN PATTERNS

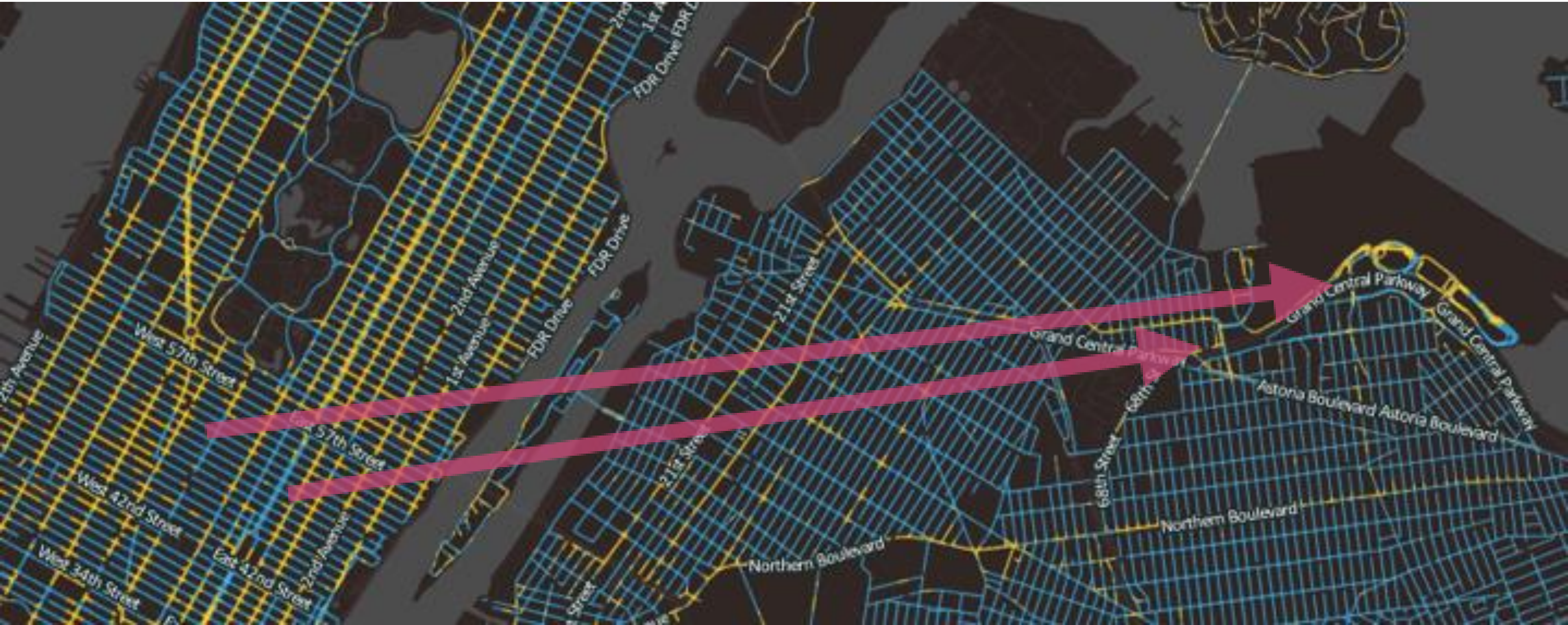

W 23 St
BN85
8P51



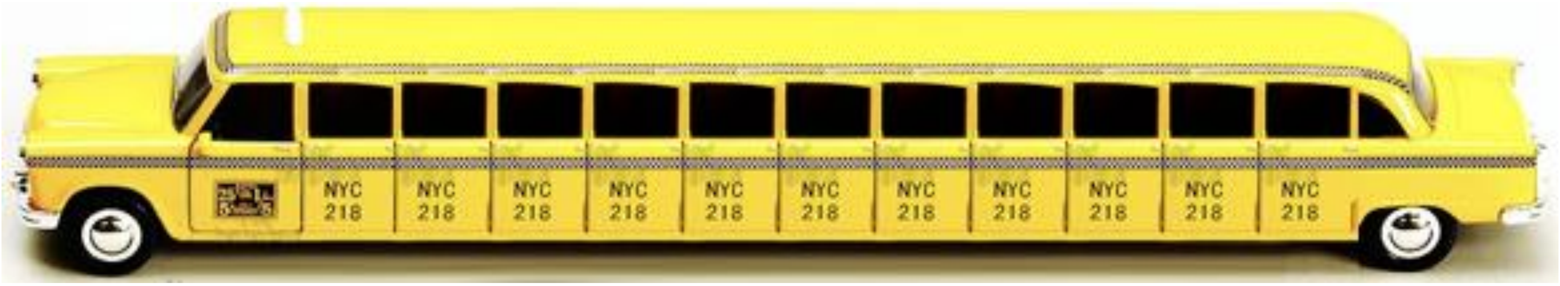


Sharing benefits
77,174 $
9,117 mi
3,857 kg
West 53rd Street
West 54th Street
West 53rd Street
West 54th Street
East 57th
Madison Avenue
5th Avenue
6th Avenue
7th Avenue
West 49th Street
West 50th Street
22.8k
West 49th Street
West 50th Street
22.4k
East 54th St
East 53rd Street
East 53rd Str
West 49th Street
6th Avenue
Madison Avenue
Park Avenue
East 49th Str
hubcab
HubCab is an interactive visualization that invites
you to explore the ways in which over 170 million
Taxi Pickup
Reset
West 50th Street
Taxi Dropoff
Reset
West 51st Street

SHARING TWO TRIPS



…AND EVEN MORE TRIPS!

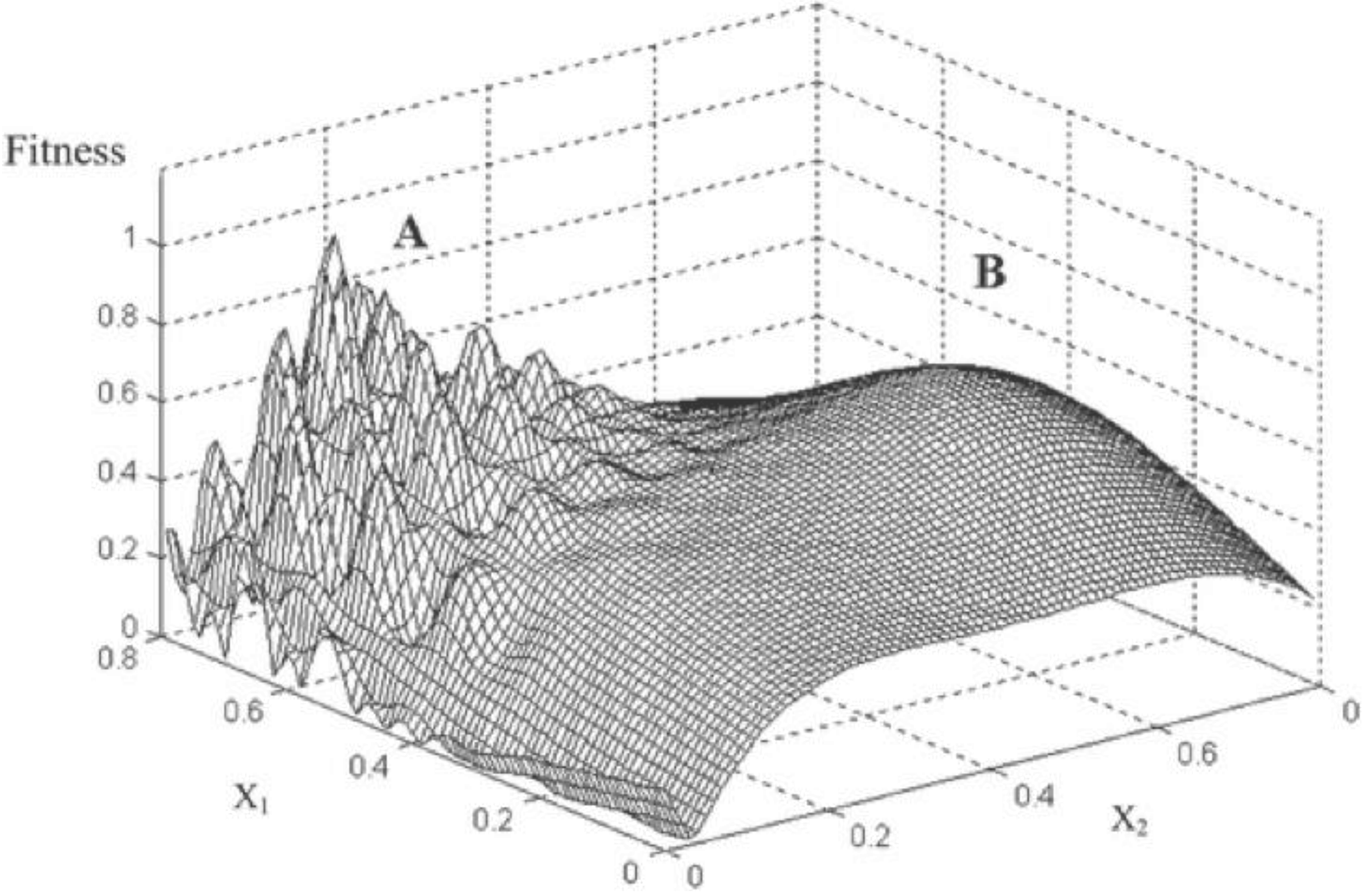


LEGACY APPROACH

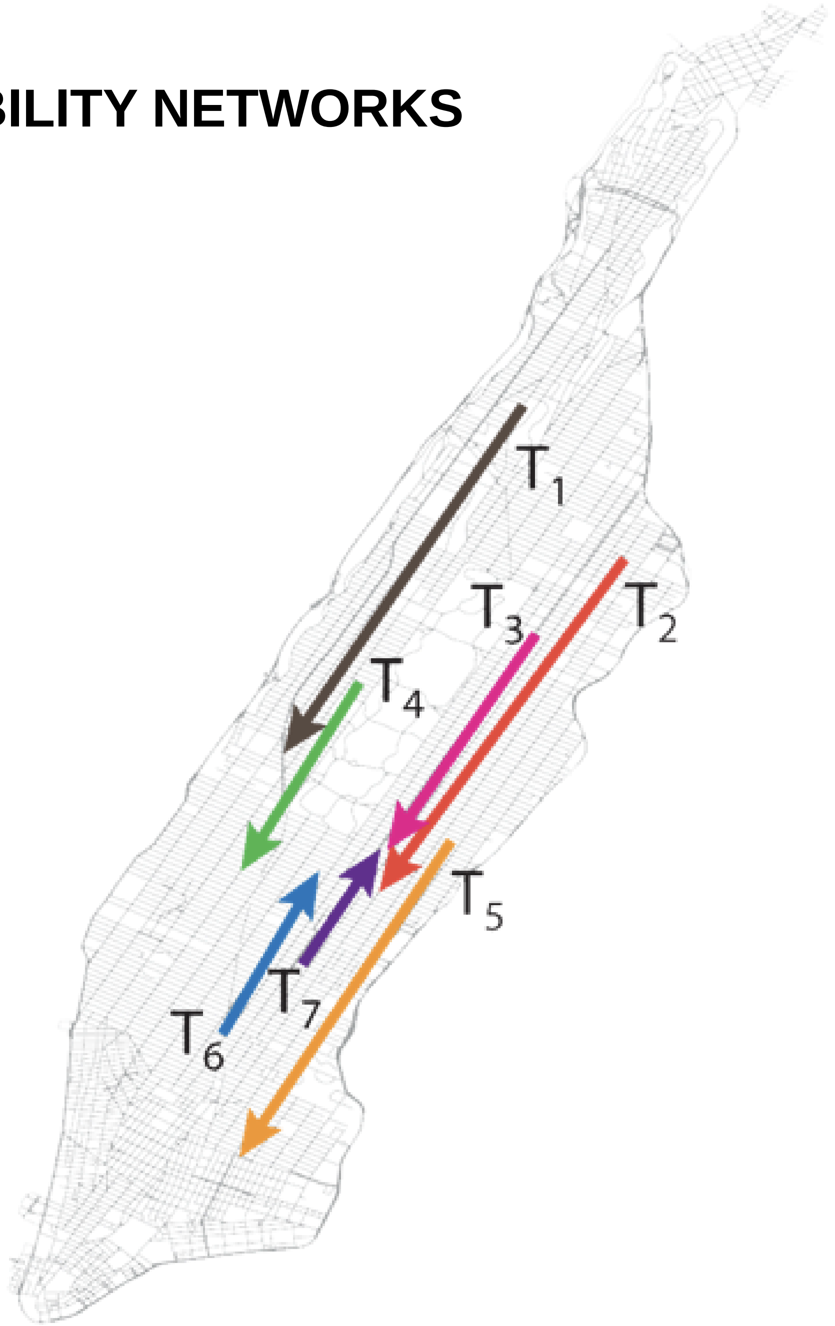
Variation of Traveler Salesman Problem

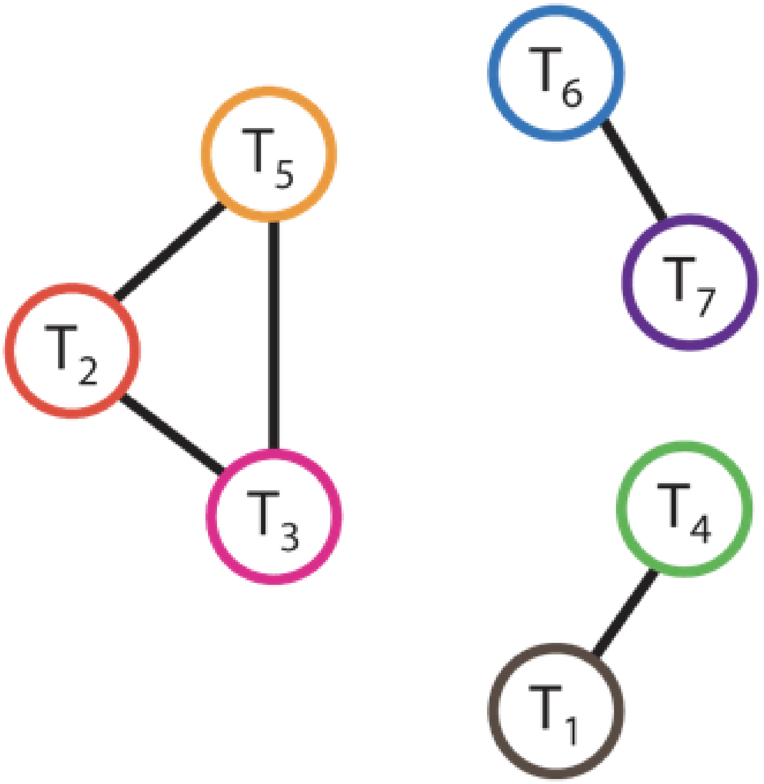
$$
\begin{aligned}
\min \sum_{r\in\Omega} c_r y_r + \sum_{d'\in D'} c_{d'} x_{d'} \\
\text{s.t.,} \sum_{d'\in D'_d} x_{d'} = 1, \quad \forall d \in D \\
\sum_{d'\in M_s} x_{d'} \leq 1, \quad \forall s \in \{S : |M_s| > 1\} \\
\sum_{r\in\Omega} \alpha_r^p y_r \geq 1, \quad \forall p \in P \\
EDC x_{dem(p')} - \sum_{r\in\Omega} \gamma_r^{p'} y_r \geq 0, \quad \forall p' \in P' \\
y_r \in \{0,1\}, \quad \forall r \in \Omega \\
x_{d'} \in \{0,1\}, \quad \forall d' \in D'
\end{aligned}
$$

UNSTRUCTURED SEARCH SPACE

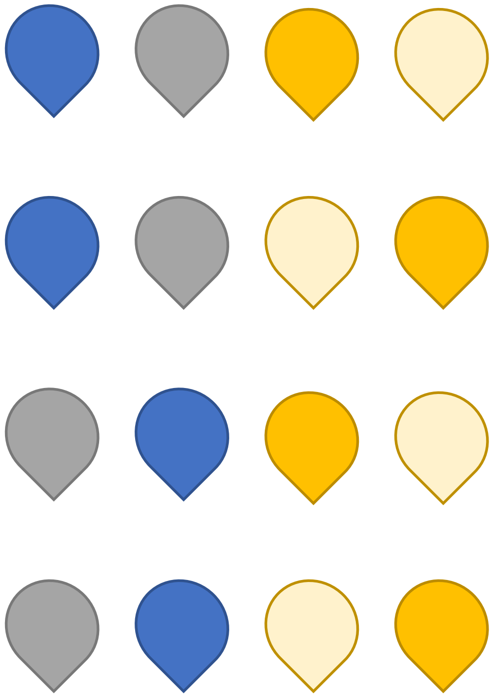


SHAREABILITY NETWORKS

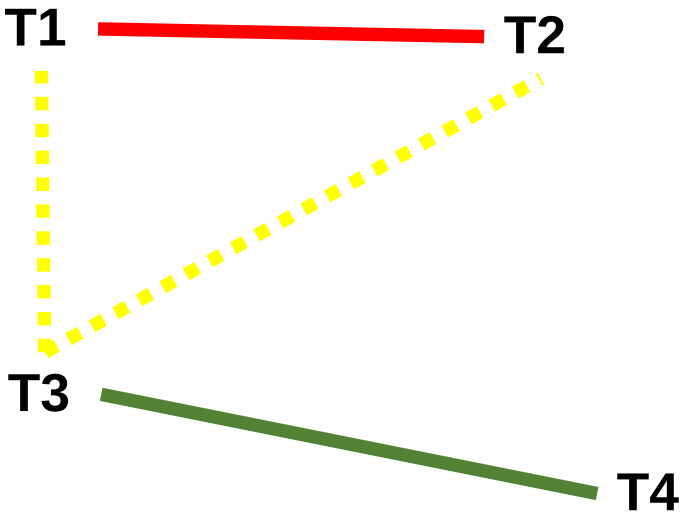




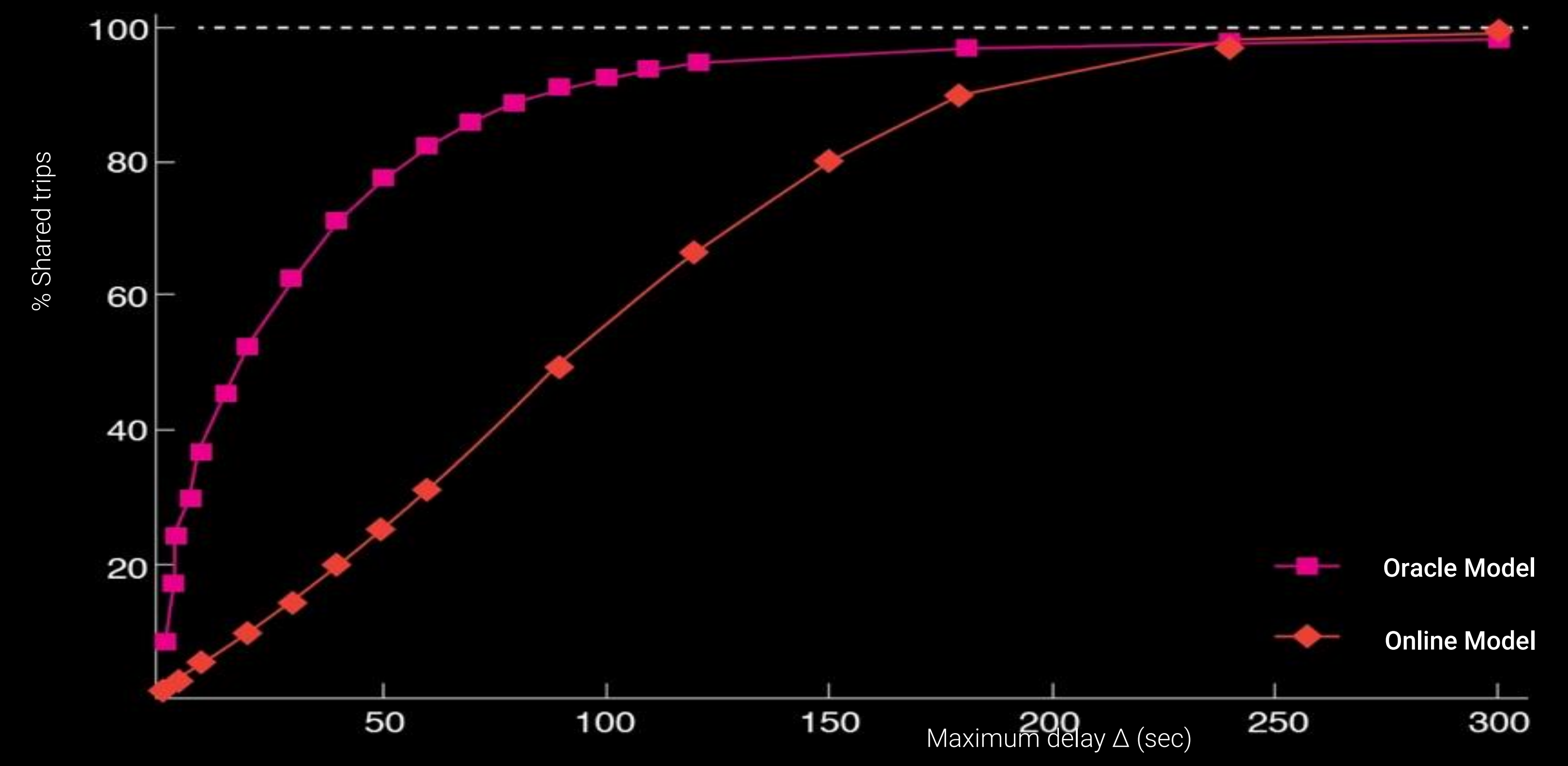
SHAREABILITY CONDITION



MAXIMUM MATCHING

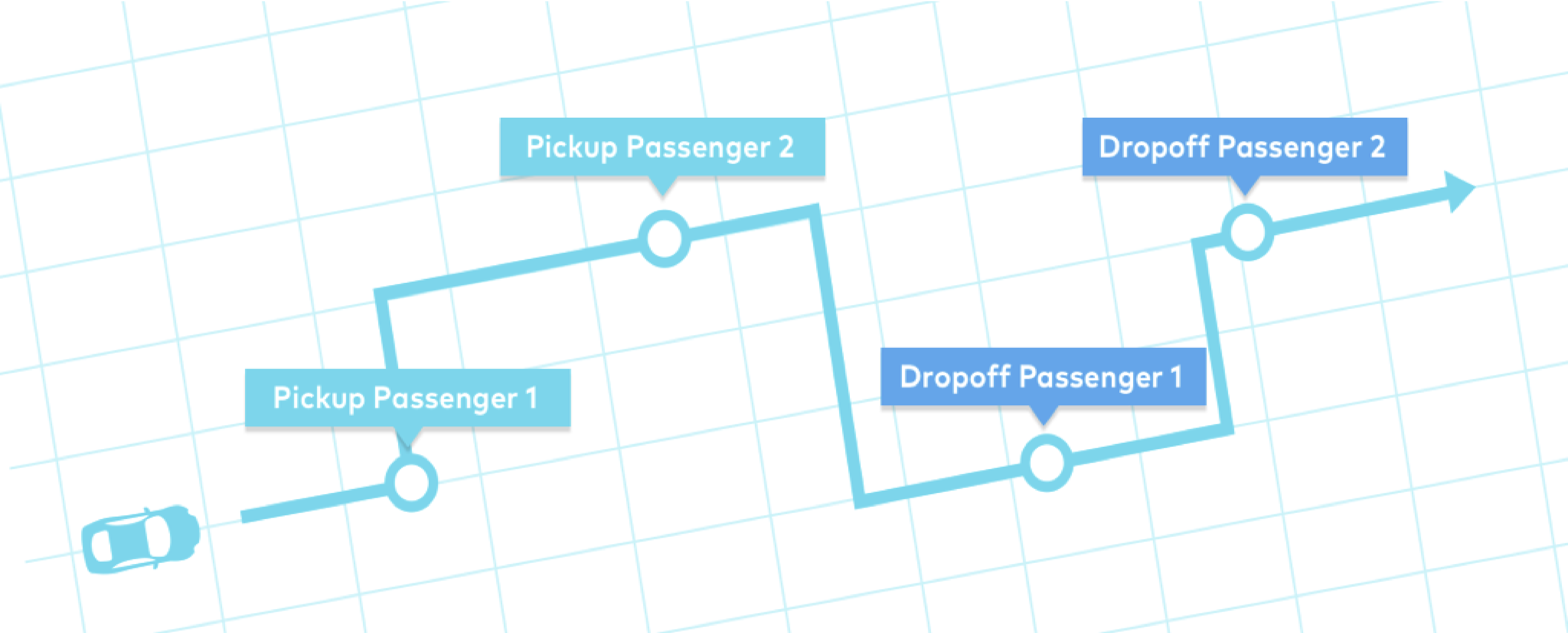


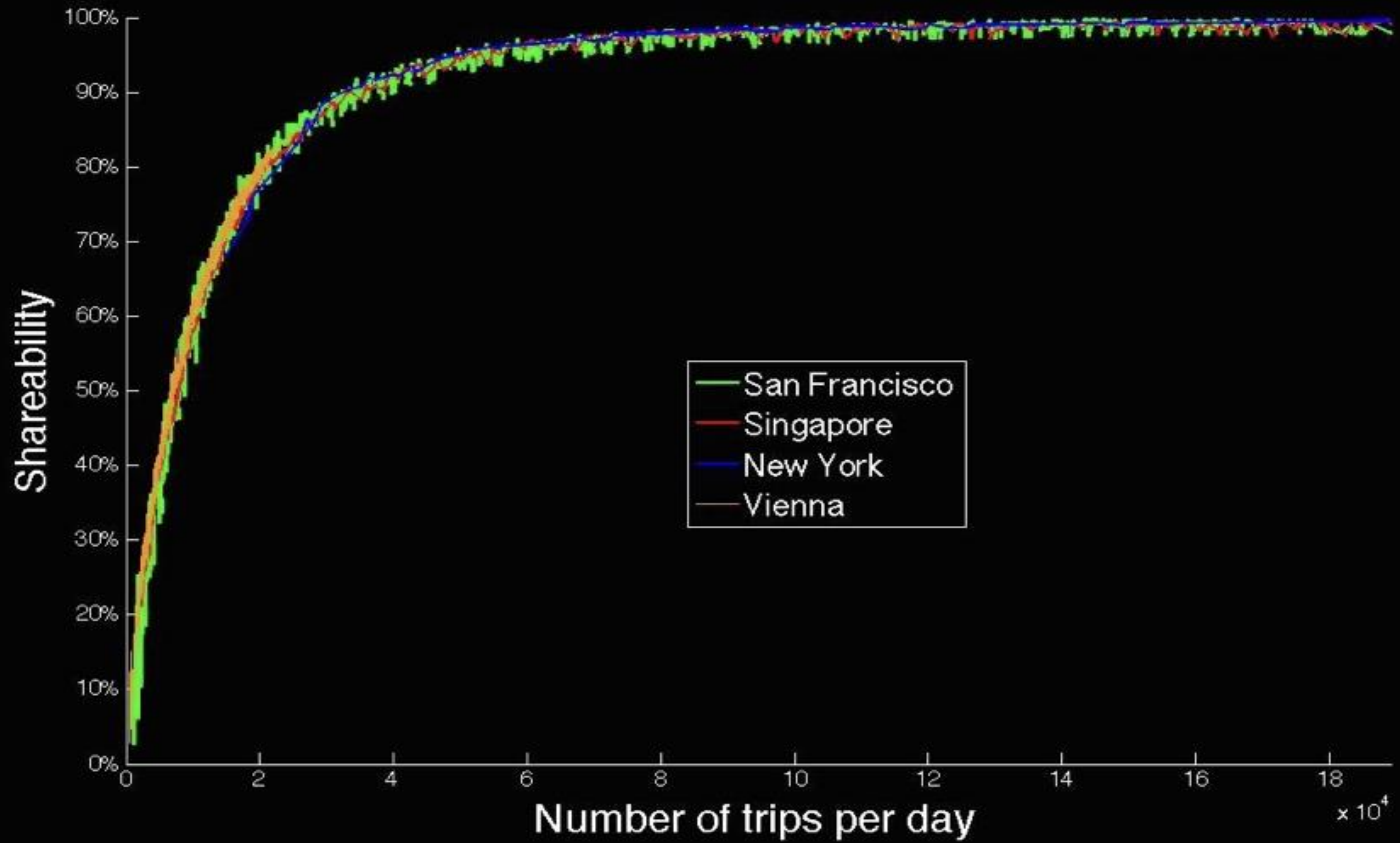
SHAREABILITY RESULTS



P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, C. Ratti,"Quantifying the Benefits of Vehicle Pooling with Shareability Networks",*Proc. National Academy of Science*, Vol. 111, n. 37, pp. 13290-13294, 2014

SHAREABLE CITIES




Shareability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
2
4
6
8
10
12
14
16
18
x 10^4
Number of trips per day
San Francisco
Singapore
New York
Vienna

CAN WE MODEL SHAREABILITY?

Input

Trip generation rate **λ**
Average car speed **v**
Delay tolerance **Δ**
City area **A**

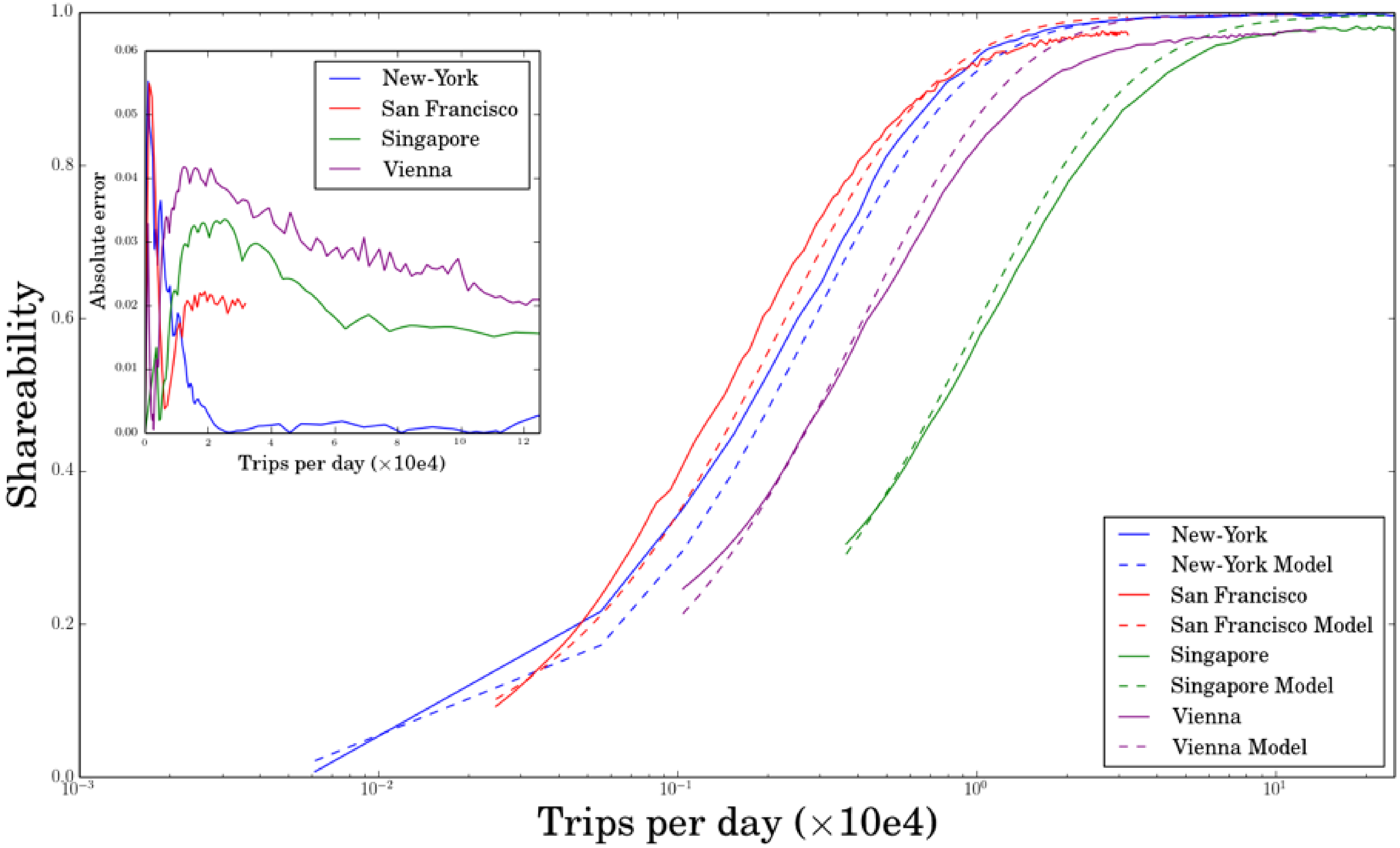
Output

Percentage **s** of shareable trips

$$s \propto \frac{\lambda \cdot \Delta^3 \cdot v^2}{A}$$

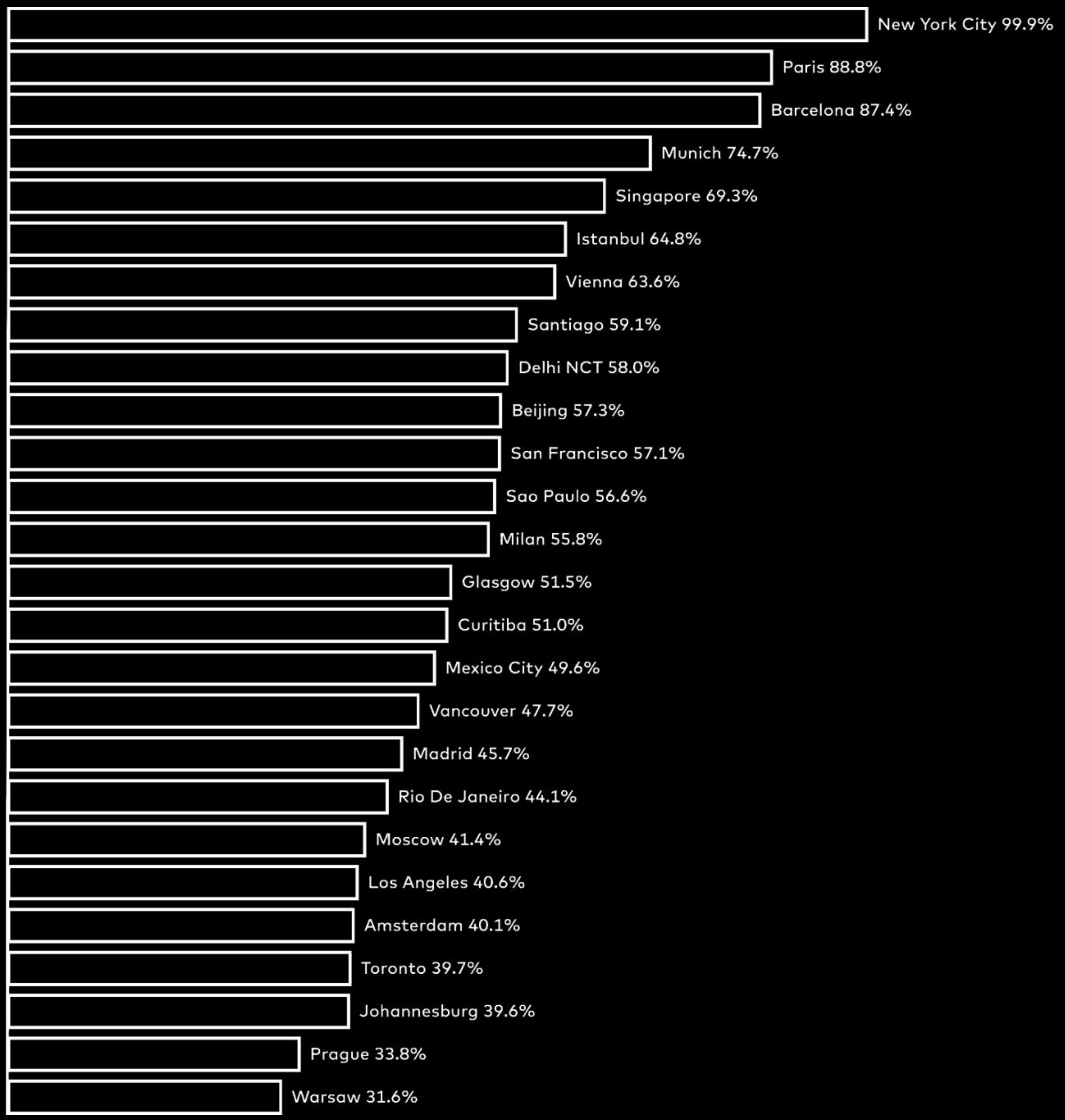
Shareability shadow

SHAREABILITY MODEL ACCURACY



SHAREABLE CITIES

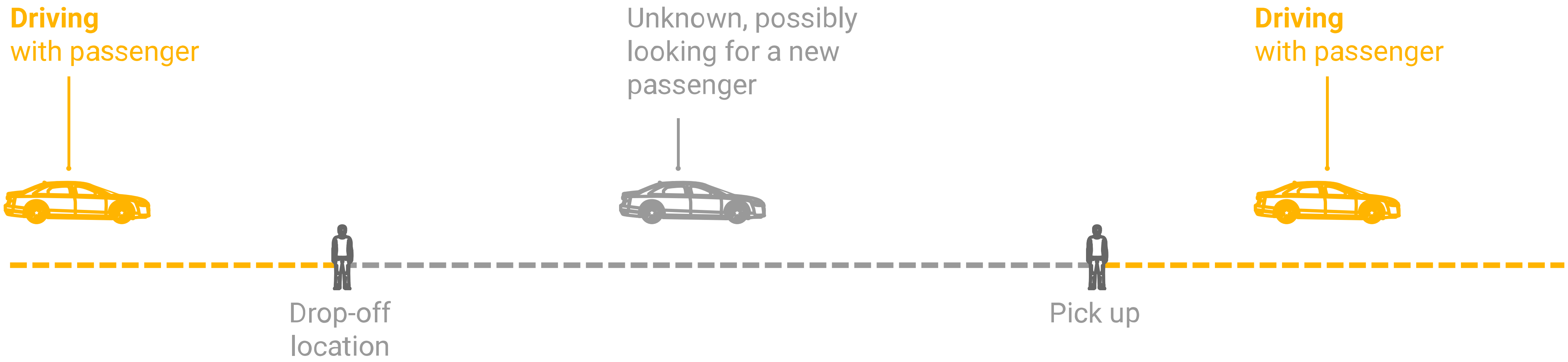
Project in collaboration
with UBER



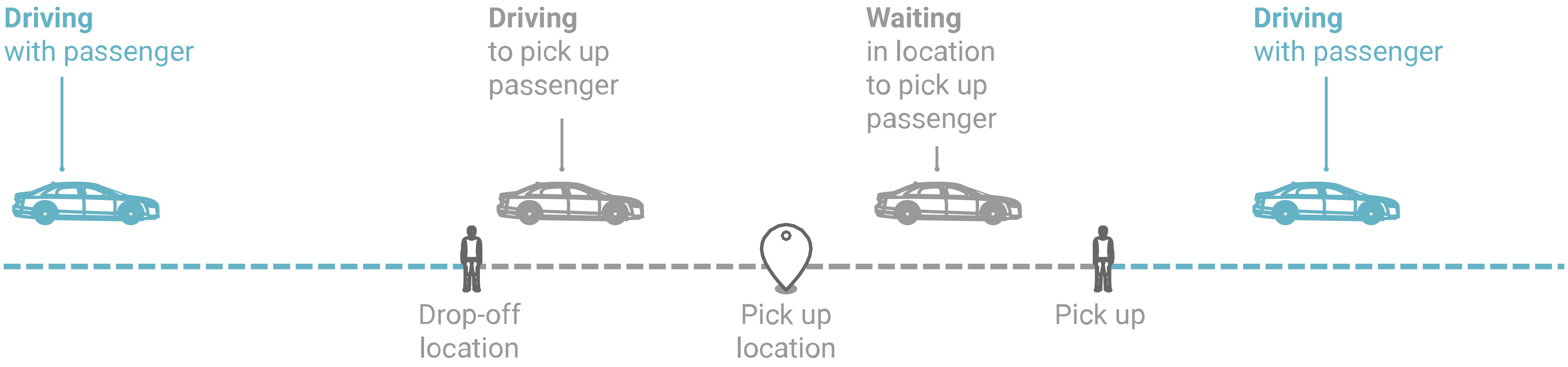
Minimum Fleet

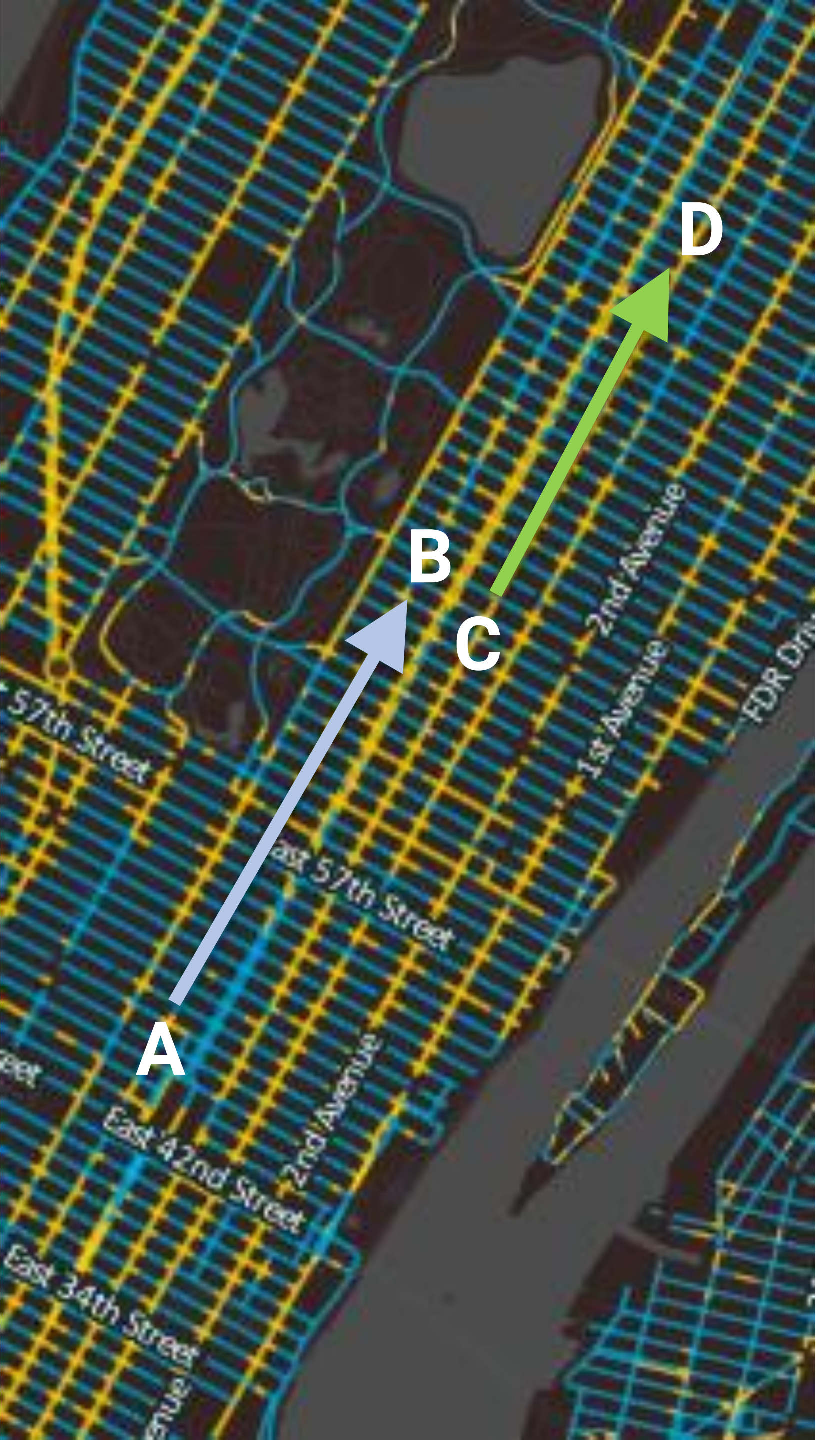
MOBILITY, OPTIMIZATION

CURRENT TAXI SITUATION

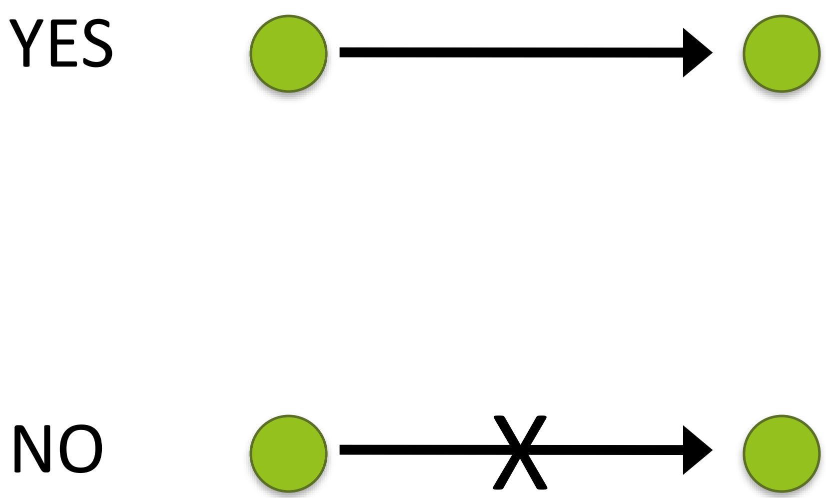


MINIMUM FLEET NETWORK MODEL

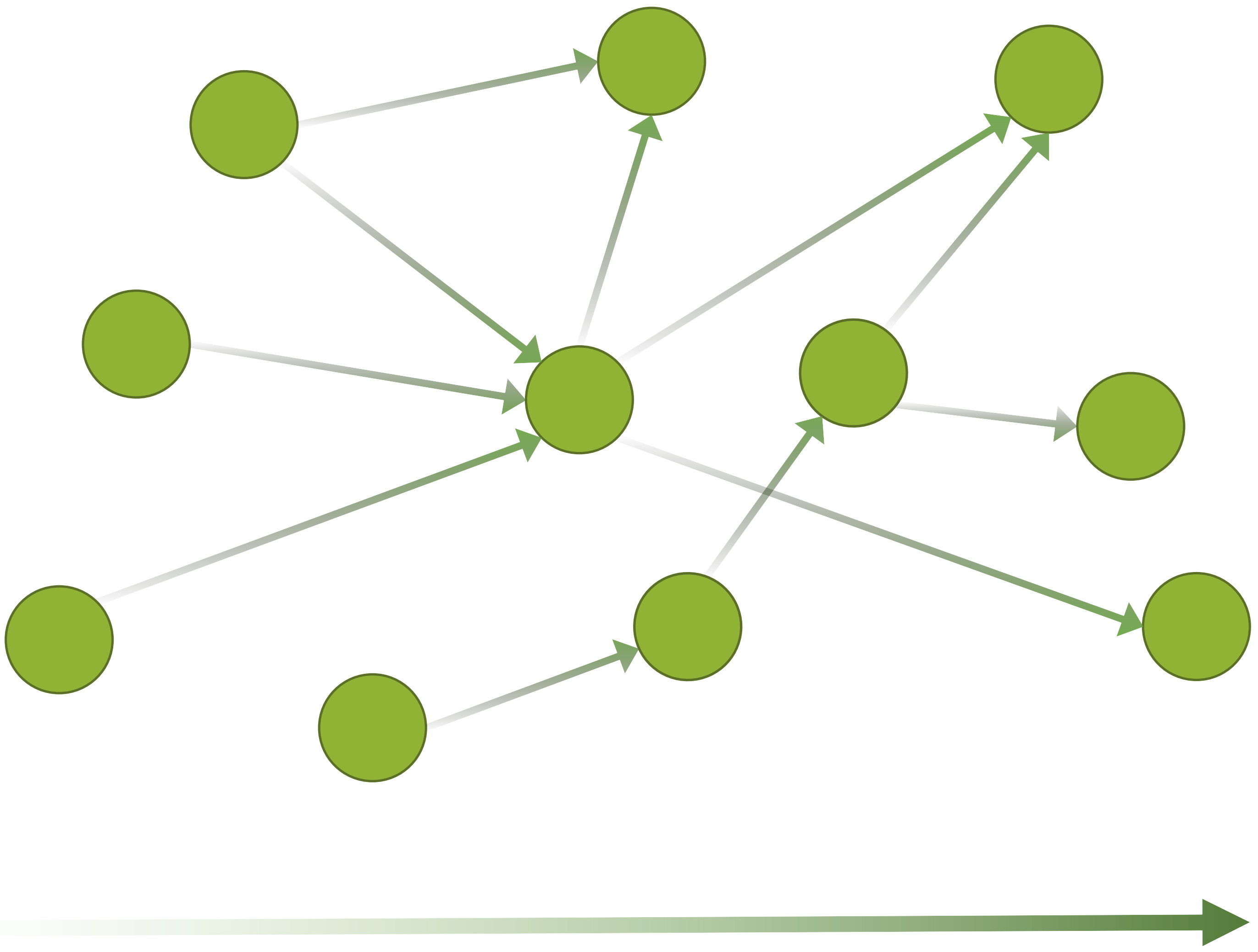




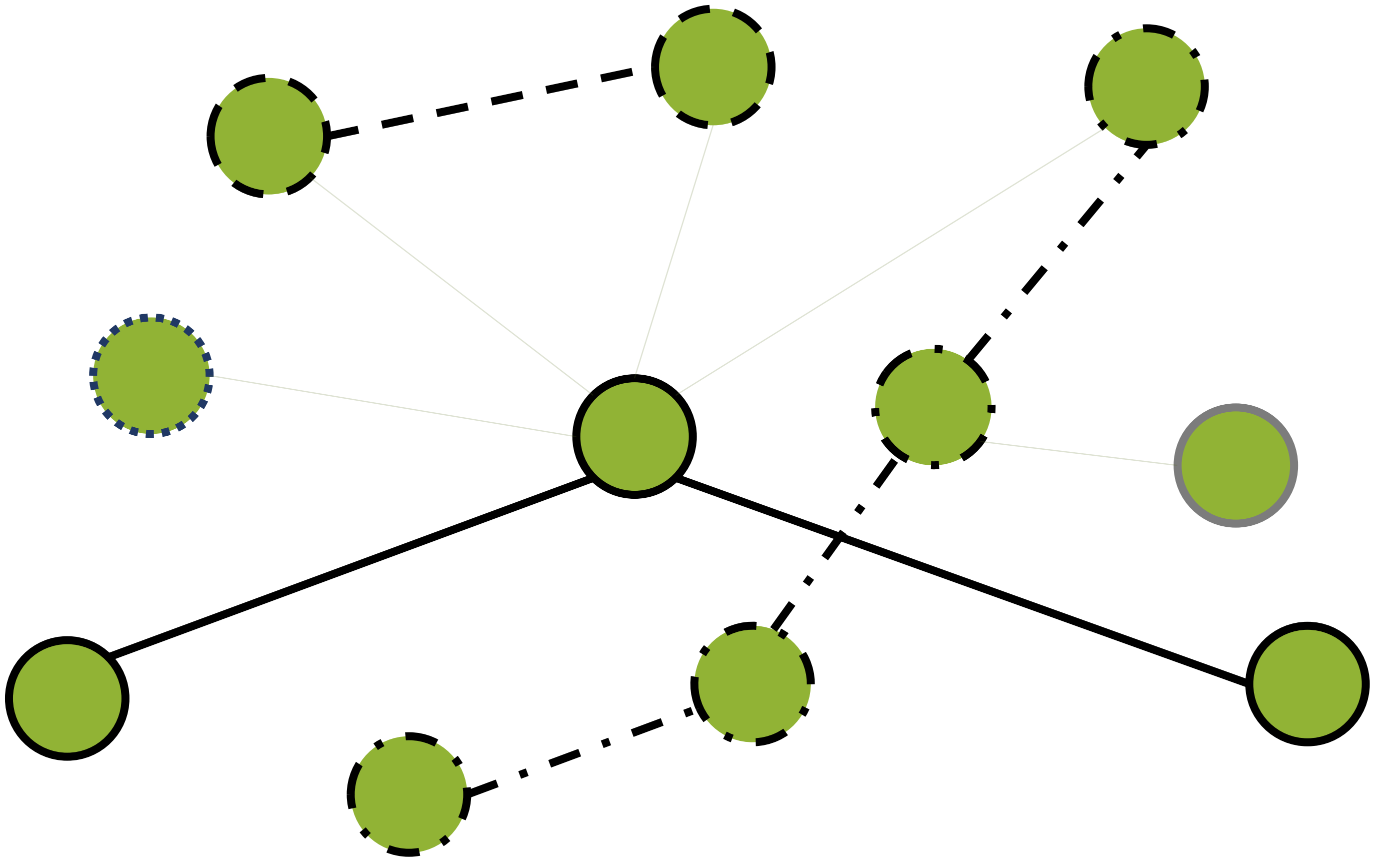
Can a vehicle dropping a passenger at B reach C before trip C➞D starts?



COVER ALGORITHM



SHAREABILITY NETWORK



https://doi.org/10.1038/s41586-018-0095-1

Addressing the minimum fleet problem in on-demand urban mobility

M. M. Vazifeh[1]*, P. Santi[1,2], G. Resta[2], S. H. Strogatz[3] & C. Ratti[1,4]

Information and communication technologies have opened the way to new solutions for urban mobility that provide better ways to match individuals with on-demand vehicles. However, a fundamental unsolved problem is how best to size and operate a fleet of vehicles, given a certain demand for personal mobility. Previous studies[1–5] either do not provide a scalable solution or require changes in human attitudes towards mobility. Here we provide a network-based solution to the following 'minimum fleet problem', given a collection of trips (specified by origin, destination and start time), of how to determine the minimum number of vehicles needed to serve all the trips without incurring any delay to the passengers. By introducing the notion of a 'vehicle-sharing network', we present an optimal computationally efficient solution to the problem, as well as a nearly optimal solution amenable to real-time implementation. We test both solutions on a dataset of 150 million taxi trips taken in the city of New York over one year[6]. The real-time implementation of the method with near-optimal service levels allows a 30 per cent reduction in fleet size compared to current taxi operation. Although constraints on driver availability and the existence of abnormal trip demands may lead to a relatively larger optimal value for the fleet size than that predicted here, the fleet size remains robust for a wide range of variations in historical trip demand. These predicted reductions in fleet size follow directly from a reorganization of taxi dispatching that could be implemented with a simple urban app; they do not assume ride sharing[7–9], nor require changes to regulations, business models, or human attitudes towards mobility to become effective. Our results could become even more relevant in the years ahead as fleets of networked, self-driving cars become commonplace[10–14].

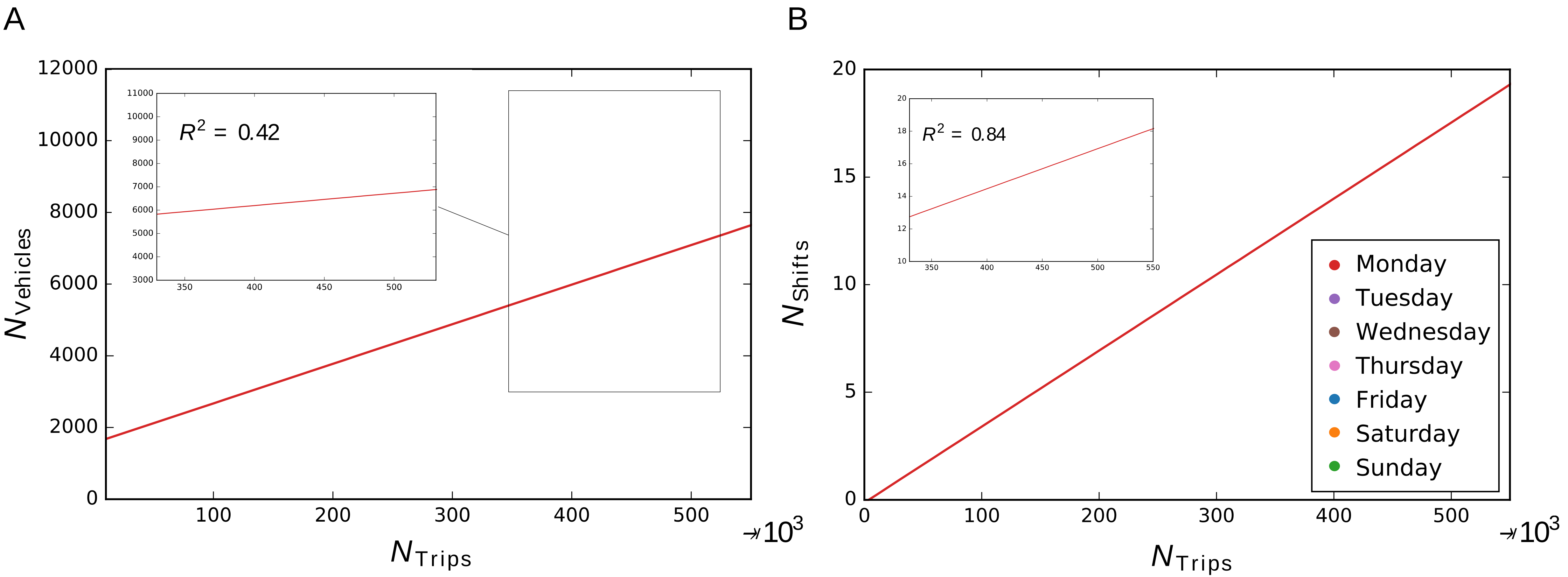
Two trends—the rise of the autonomous and connected car, and the emergence of a 'sharing economy'[10,11] of transportation—seem poised to revolutionize the way personal mobility needs will be addressed in cities. The way current modes of transportation such as the private car, taxi or bus operate will be challenged and increasingly replaced by personalized, on-demand mobility systems operated by vehicle fleets, similar to what companies like Uber and Lyft offer. If such trends con-

In what follows, we solve the 'minimum fleet problem' for the general case of on-demand mobility, and show that its solution for a specific case—taxi trips—could lead to breakthroughs in operational efficiency. To the best of our knowledge, no publicly available solution currently exists to address this minimum fleet-size problem at the urban scale for on-demand mobility in both private and public sectors. On the one hand, accurate methods based on mathematical programming (as traditionally used in the design of transportation systems[1–5,9]) can handle only a few thousand trips or vehicles at most, which is well below the hundreds of thousands or even millions of trips or vehicles routinely operating in large cities. On the other hand, city-scale studies[17] are obtained using a model of transportation based on aggregated mobility data and Euclidean spatial assumptions, and hence lack the resolution necessary to estimate the urban-scale benefits of vehicle sharing accurately.

We start from the notion of the shareability network introduced in ref. [7], which did not focus on the dispatching of vehicles. The type of shareability network introduced here is profoundly different from the type studied previously: it models the sharing of vehicles, whereas previous networks[7–9] modelled the sharing of rides. The main methodological contribution of this Letter is to show how this vehicle-sharing network can be translated into an exact formulation of the minimum fleet problem as a minimum path cover problem on directed graphs, thus establishing a connection to the rich applied mathematics and computer science field of graph algorithms. Besides revealing a structural property of vehicle-sharing networks, this connection allows the derivation of computationally efficient algorithms for optimal vehicle deployment and dispatching. Although optimally solving the minimum fleet size problem requires offline knowledge of daily mobility demand, in the following we also present a near-optimal, online version of the algorithm that can be executed in real time knowing only a small amount of the trip demand.

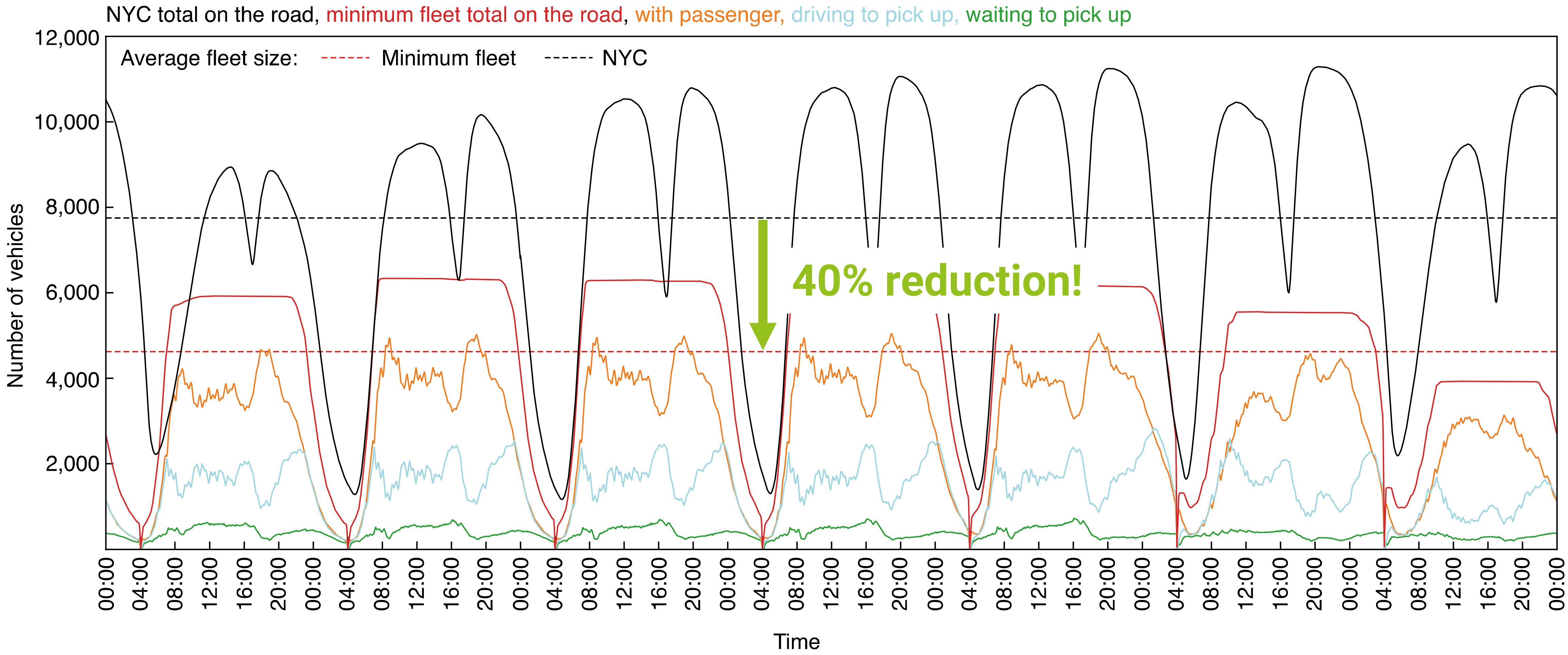
We are given a collection $\mathcal{T}$ of individual trips representing a portion of urban mobility demand during a certain time interval, such as a day. Each trip $T_i \in \mathcal{T}$ is defined as a tuple $(t_i^p, t_i^d, l_i^p, l_i^d)$ where t_i^p represents the desired pick-up time, l_i^p the pick-up location, t_i^d the drop-off

MFS VS. #TRIPS



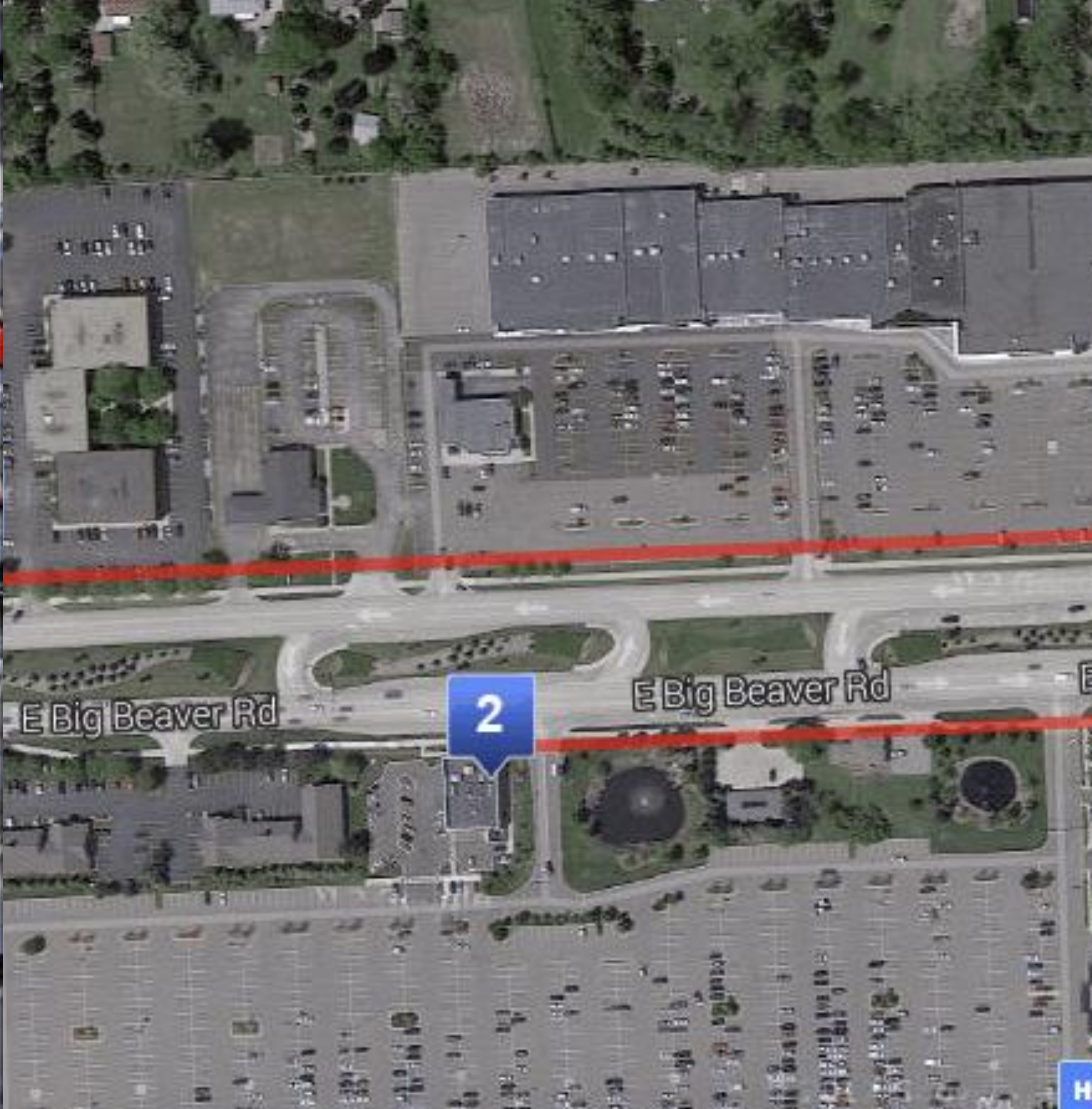
A
B
$R^2 = 0.42$
$R^2 = 0.84$
N_{Vehicles}
N_{Shifts}
N_{Trips}
$\times 10^3$
12000
10000
8000
6000
4000
2000
0
11000
10000
9000
8000
7000
6000
5000
4000
3000
350
400
450
500
550
20
18
16
15
14
12
10
5
100
200
300
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday

COMPARISON VS. NY TAXI



Unparking

PARKING, AUTONOMOUS VEHICLE


E Big Beaver Rd
2
E Big Beaver Rd

MOTIVATION

Typical car is parked for over 95% of its lifetime

In busy time of day, cars spend up to 30% of driving time to look for parking

Los Angels County facts:

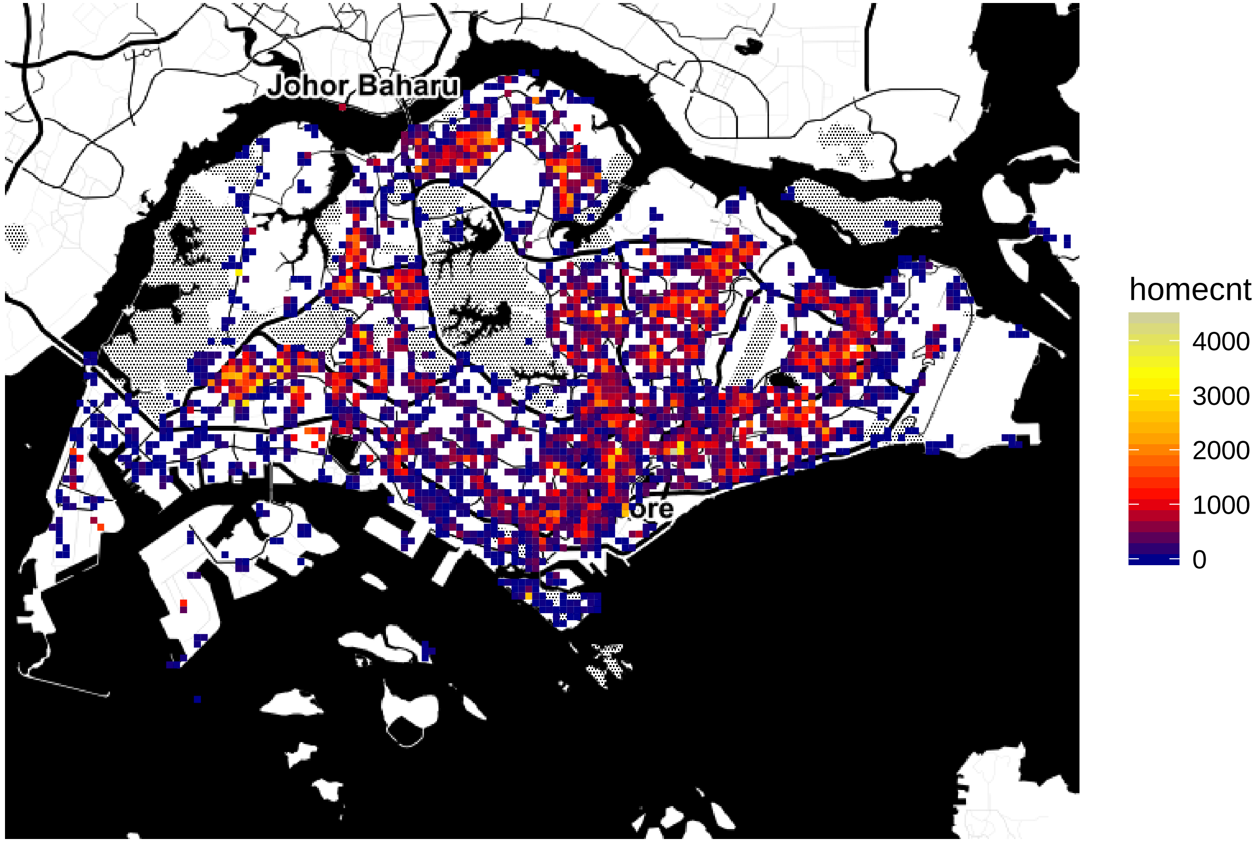
- **9.8 M** people; **5.6 M** cars; **18.6 M** parking spaces (*data from 2010*)
- **140 sq miles** of roads; **200 sq miles** of parking (14% of total incorporated area)

Can the trend toward **shared** and **autonomous mobility** helps solving parking issues?

METHODOLOGY

Use cell phone data set to estimate mobility demand (home-work commuting)

Considered four scenarios:
- private car, private parking (today)
- private car, shared parking
- shared car, shared parking
- shared autonomous car, shared parking



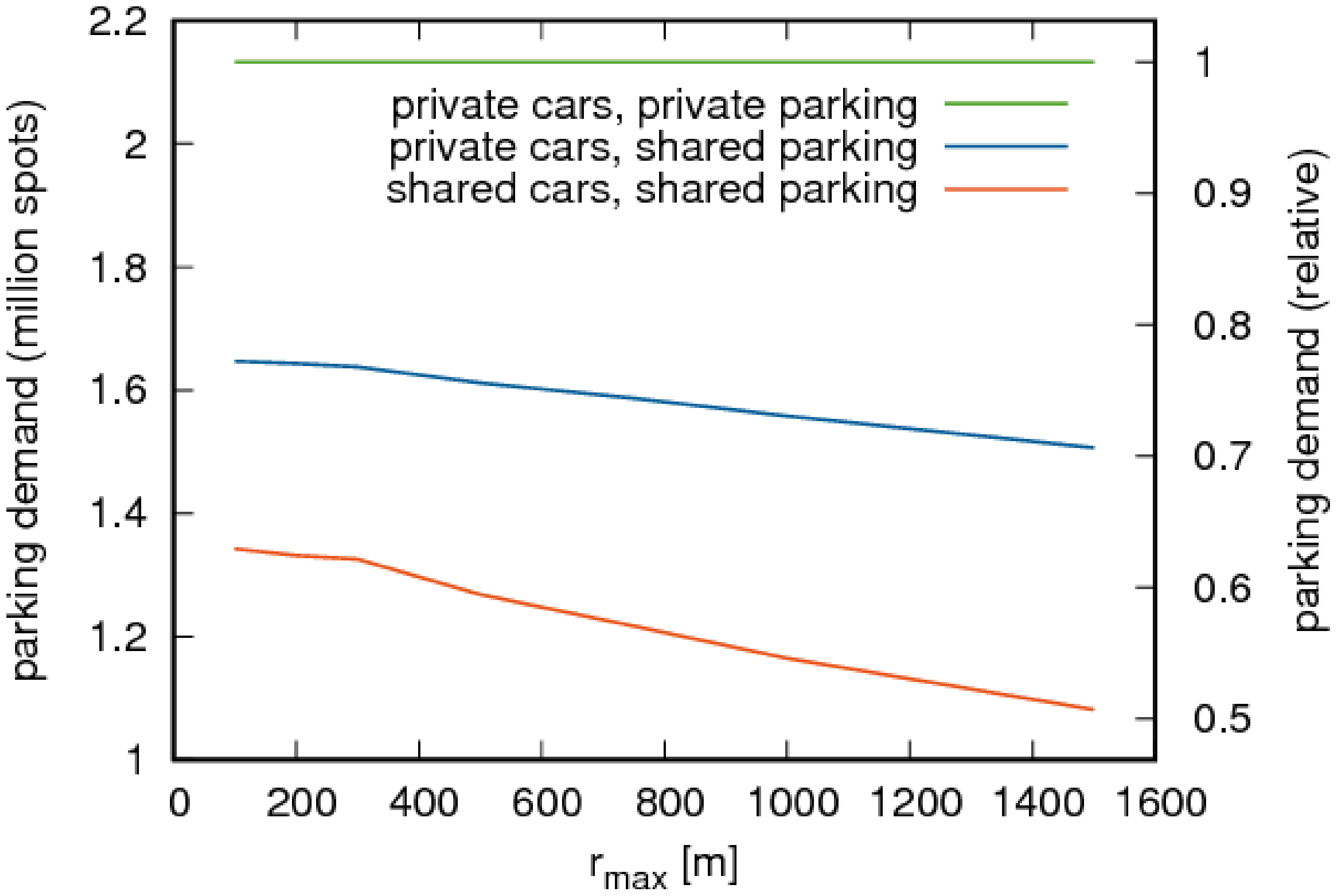
Estimated home locations in Singapore

RESULTS

40% savings when compared to reserved parking and search radius of up to **500m** (*shared cars*)

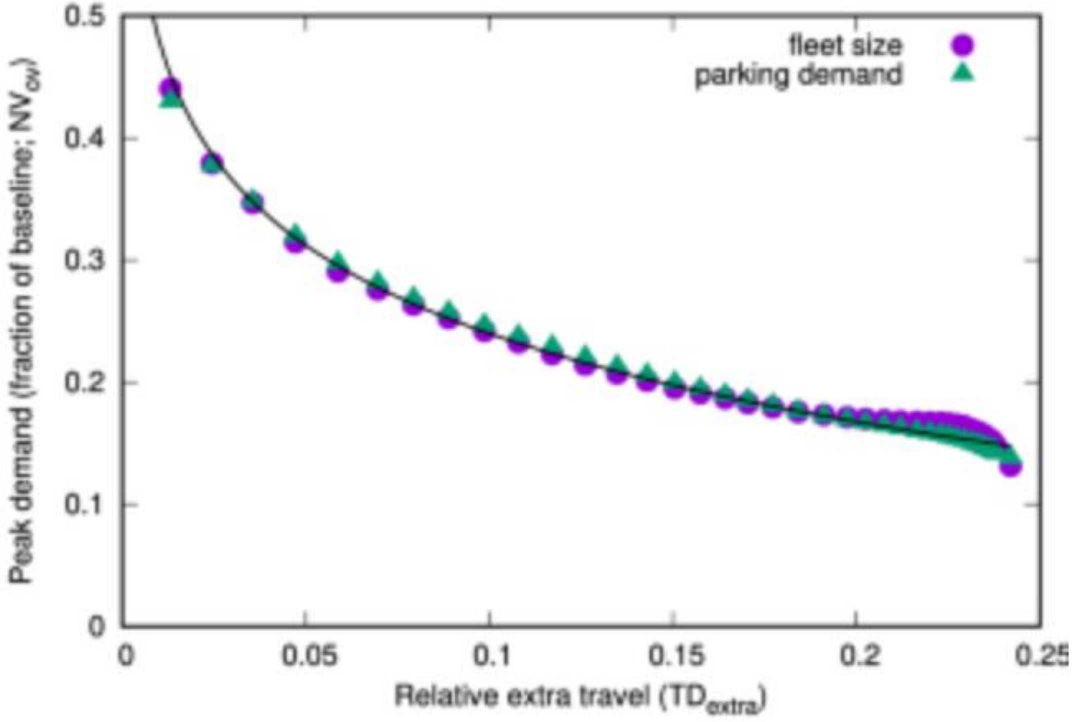
60% savings with larger search ranges (*shared autonomous cars*)

Even more notable considering that refer to **home-work commuting**



MORE GENERAL RESULTS

Figure 1



Pointiest Path

MOBILITY PATTERNS, PEDESTRIAN NAVIGATION

WHAT DO WE KNOW ABOUT PEDESTRIAN MOBILITY?

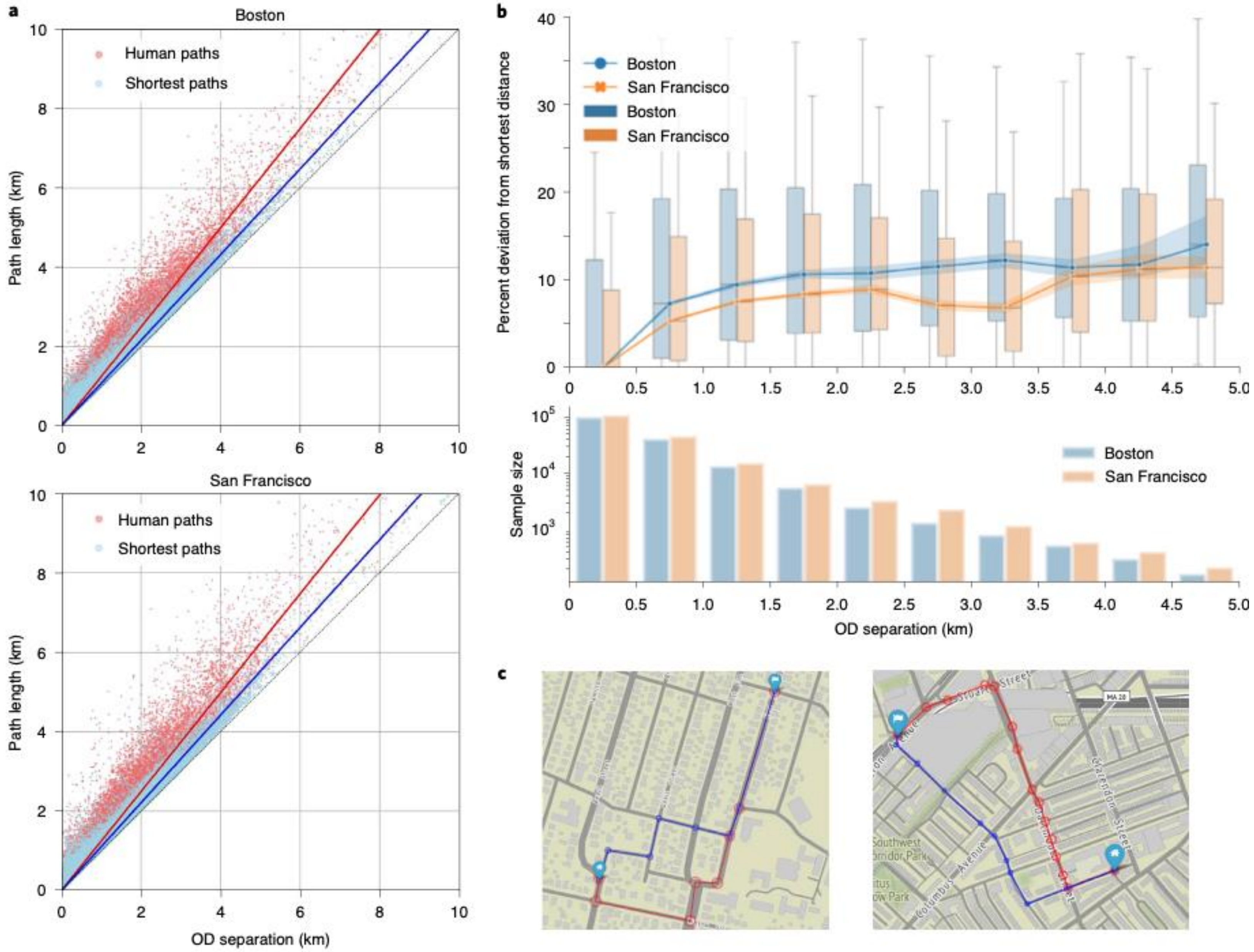
Studies so far performed in small-scale experiments, often in VR environments

- ✓ Performed in **controlled environments** to address specific research hypothesis
- ✓ **Small-scale** in both size of the environment and number of participants
- ✓ Some basic candidate mechanisms for navigation (landmarks, mental maps, etc.) identified

But what happens in the real world, e.g., pedestrian walking in a city?

REAL-WORLD PEDESTRIAN MOBILITY

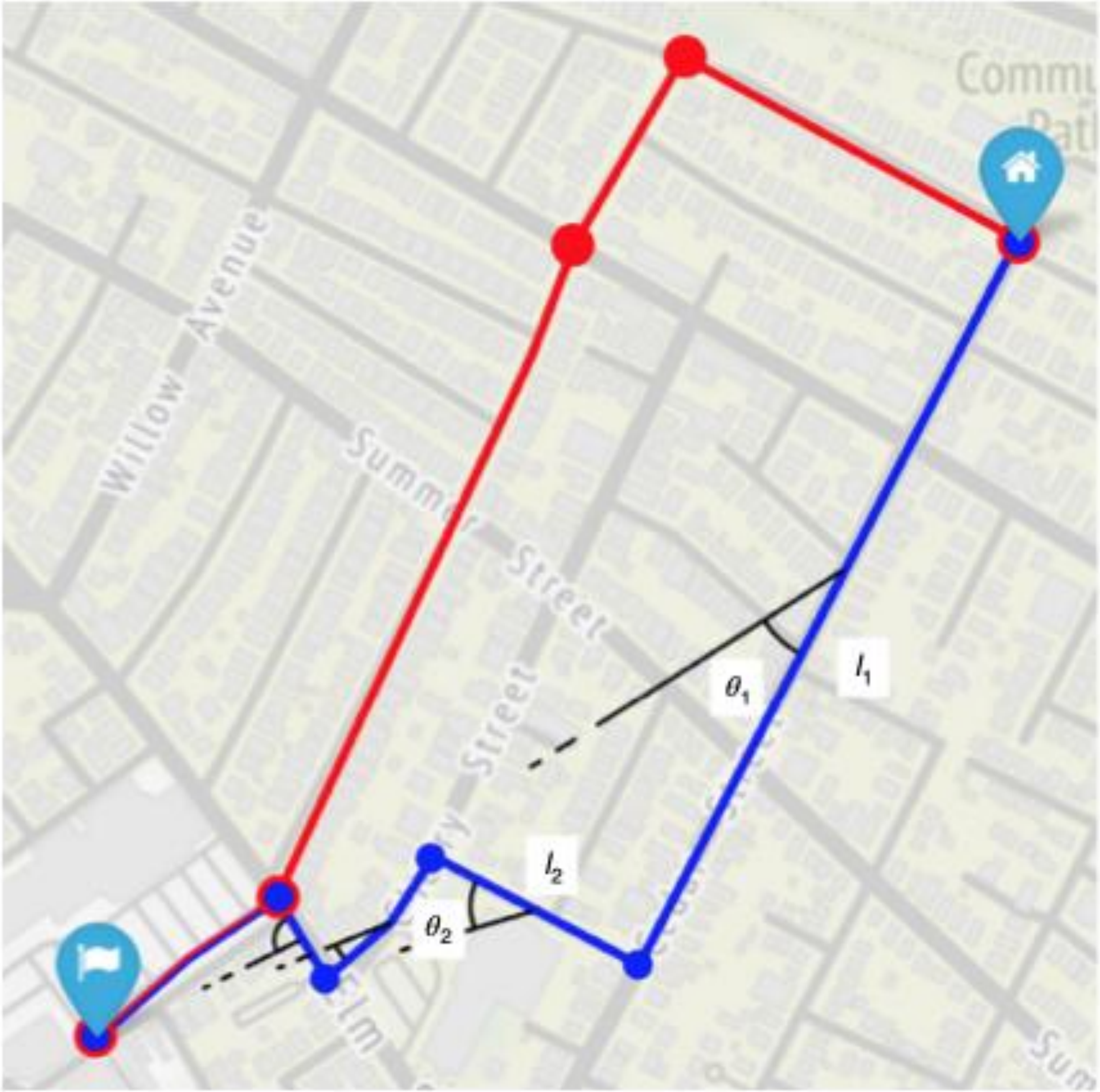
Analysis of over 100,000 pedestrian paths (GPS) in Boston and SF



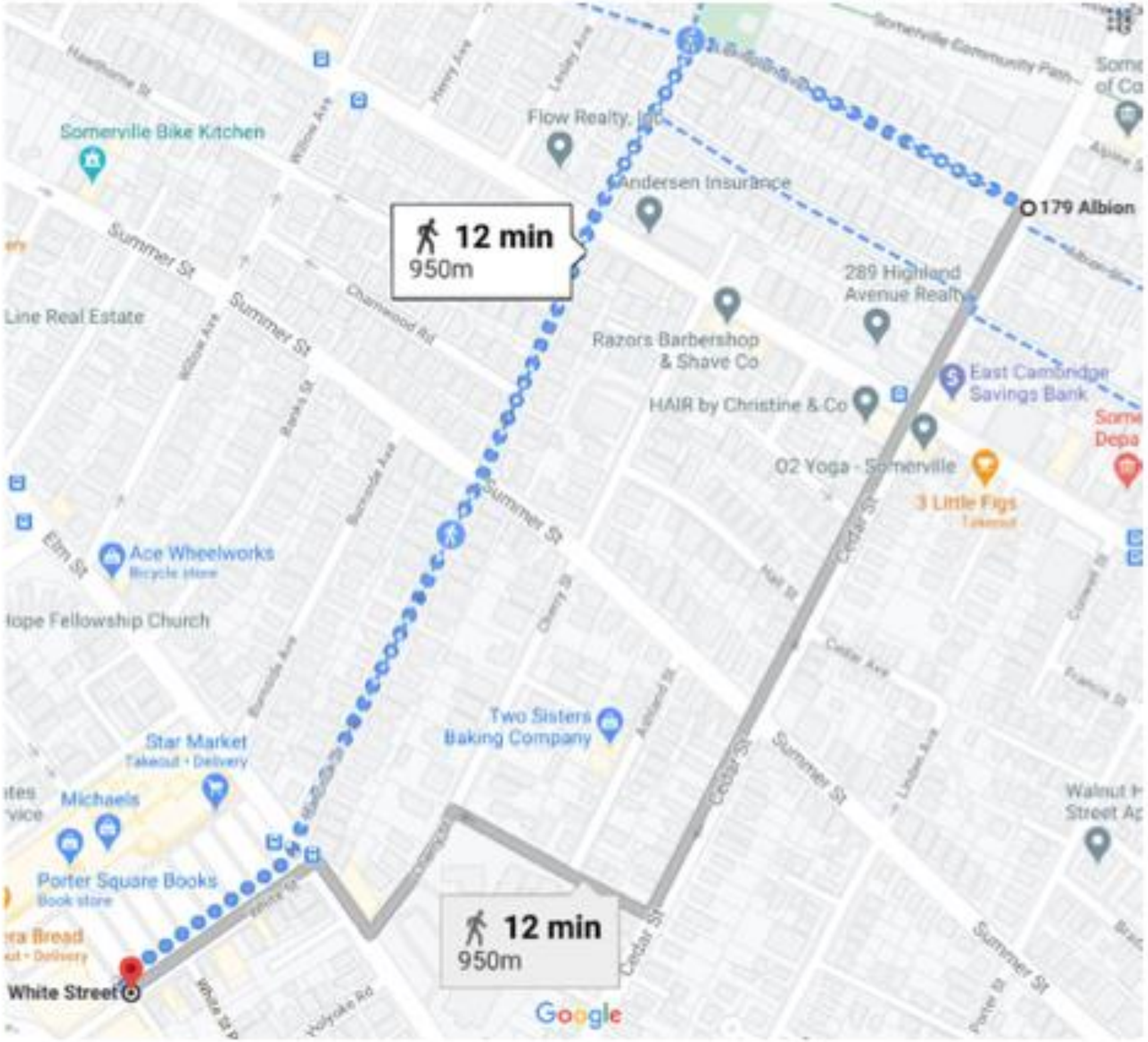
C. Bongiorno, Y. Zhou, M. Kryven, D. Theurel, A. Rizzo, P. Santi, J. Tenenbaum, C. Ratti, “Vector-based pedestrian navigation in cities”, *Nature Computational Science*, 2021.

VECTOR-BASED NAVIGATION

a



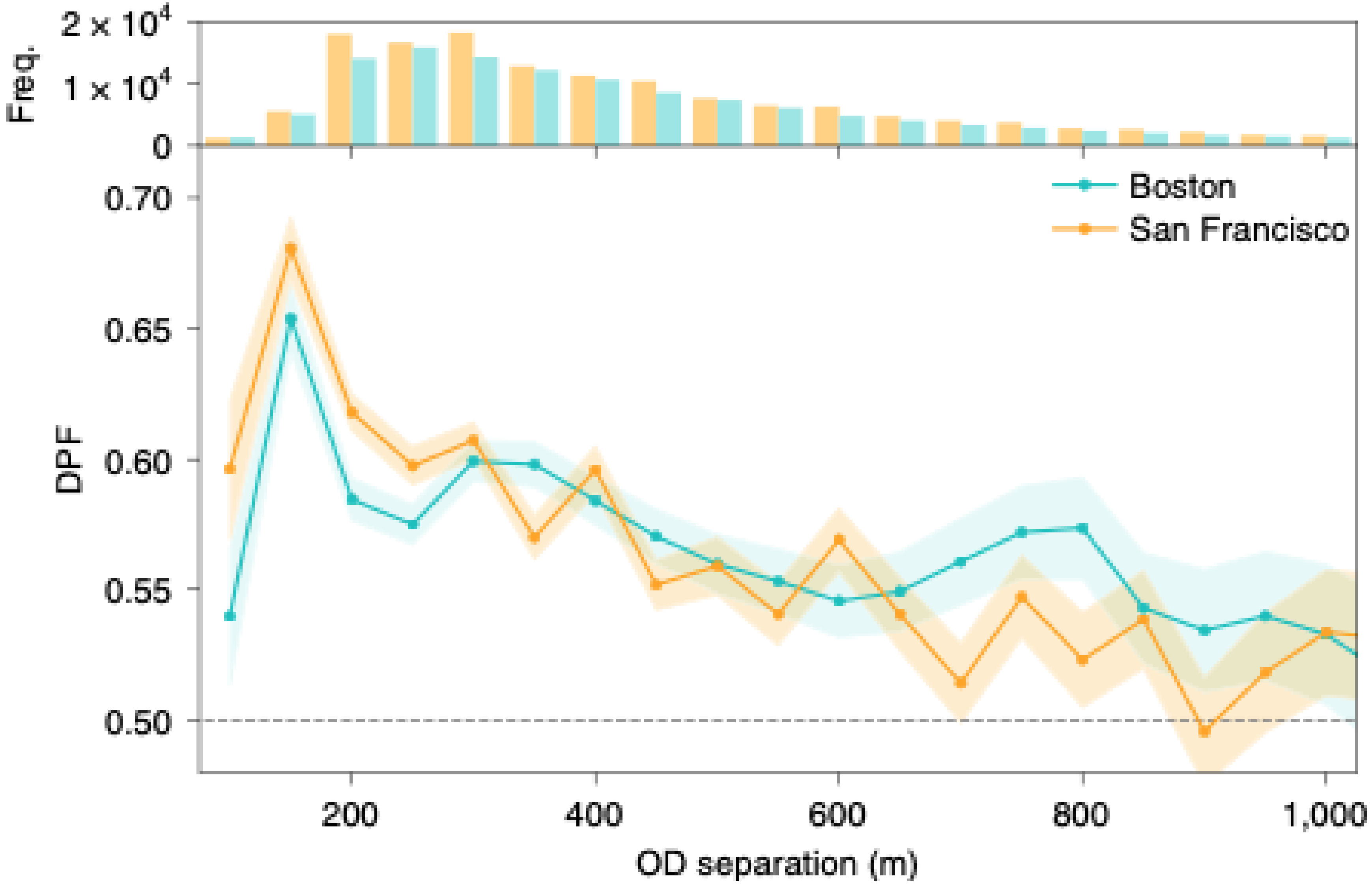
b



VECTOR-BASED MODEL

$$C_{\mathrm{dir}}[\mathcal{P}] = \sum_{S_i \in \mathcal{P}} c(\theta_i, l_i),$$

$$c(\theta_i, l_i) = \mathrm{e}^{\mathcal{N}(\log(|\theta_i| l_i), \sigma)}$$



Thank you!

QUESTIONS?