

# Mathematical modelling of human mobility

BME, Budapest, Sep 18th 2018

Michael Szell  
@mszll





# Understanding mobility in a social petri dish

SUBJECT AREAS:

APPLIED PHYSICS

STATISTICAL PHYSICS,  
THERMODYNAMICS AND  
NONLINEAR DYNAMICS

STATISTICS

MODELLING AND THEORY

Received  
9 March 2012

Accepted  
21 May 2012

Published  
14 June 2012

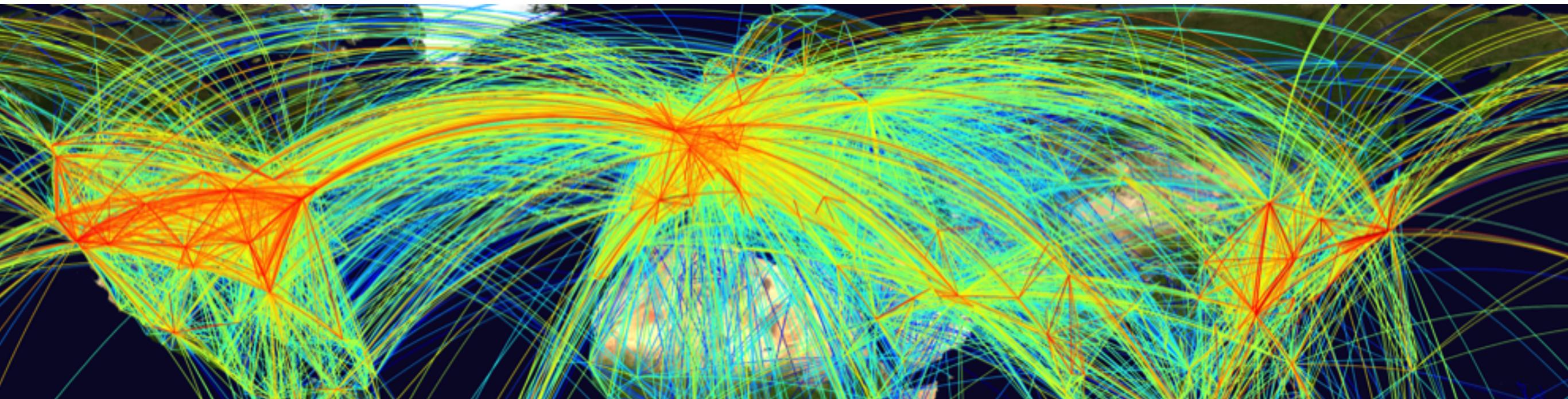
Correspondence and  
requests for materials  
should be addressed to

Michael Szell<sup>1</sup>, Roberta Sinatra<sup>2,8</sup>, Giovanni Petri<sup>3,9,10</sup>, Stefan Thurner<sup>1,6,7</sup> & Vito Latora<sup>4,5,8</sup>

<sup>1</sup>Section for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria, <sup>2</sup>Center for Complex Network Research and Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA, <sup>3</sup>Institute for Scientific Interchange (IS), Via Alassio 11/c, 10126 Torino, Italy, <sup>4</sup>School of Mathematical Sciences, Queen Mary, University of London, London E1 4NS, United Kingdom, <sup>5</sup>Dipartimento di Fisica e Astronomia, Università di Catania and INFN, Via S. Sofia, 64, 95123 Catania, Italy, <sup>6</sup>Santa Fe Institute, Santa Fe, NM 87501, USA, <sup>7</sup>IIASA, Schlossplatz 1, 2361 Laxenburg, Austria, <sup>8</sup>Laboratorio sui Sistemi Complessi, Scuola Superiore di Catania, Via San Nullo 5/i, 95123 Catania, Italy, <sup>9</sup>Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK, <sup>10</sup>Complexity and Networks group, Imperial College London, London SW7 2AZ, UK.

Despite the recent availability of large data sets on human movements, a full understanding of the rules governing motion within social systems is still missing, due to incomplete information on the socio-economic factors and to often limited spatio-temporal resolutions. Here we study an entire society of individuals, the players of an online-game, with complete information on their movements in a network-shaped universe and on their social and economic interactions. Such a “socio-economic laboratory” allows to unveil the intricate interplay of spatial constraints, social and economic factors, and patterns of mobility. We find that the motion of individuals is not only constrained by physical distances, but also strongly shaped by the presence of socio-economic areas. These regions can be recovered perfectly by community detection methods solely based on the measured human dynamics. Moreover, we uncover that long-term memory in the time-order of visited locations is the essential ingredient for modeling the trajectories.

# Why study human mobility?



- Spread of epidemics
- Urban planning, traffic management
- Crowd dynamics
- Geomarketing
- Spread of computer viruses

Brockmann and Helbing, Science 13 (2013)

Barthelemy, Phys Rep 499, 1-101 (2010)

Castellano et al, Rev Mod Phys 81, 591-646 (2009)

Pastor-Satorras and Vespignani, PRL 86, 3200-3203 (2001)

Wang et al, Science 324, 1071-1075 (2009)

Helbing et al, PRE 75, 046109 (2007)

# Measuring mobility

Large-scale datasets: mobile phones, dollar bills, subway



- Topology often a spatial network
- Predictability
- Diffusion

“Universal theory of human mobility”?

Brockmann et al, Nature 439, 462-465 (2006)

González et al, Nature 453, 779-782 (2008)

Roth et al, PLoS One 6, e15923 (2011)

Song et al, Science 327, 1018-1021 (2010)

Koelbl and Helbing, New J of Phys 5, 48 (2003)

# Data problems

- Problems:
- Often no raw data, but reconstructed positions
  - Limited to specific human activity
  - Limited spatial/temporal resolution
  - No socio-economic contexts
  - Different datasets

“Universal theory of human mobility”?

 **Not yet!**

Brockmann et al, Nature 439, 462-465 (2006)

González et al, Nature 453, 779-782 (2008)

Roth et al, PLoS One 6, e15923 (2011)

Song et al, Science 327, 1018-1021 (2010)

Koelbl and Helbing, New J of Phys 5, 48 (2003)

## Our contribution

- I) Uncover socio-economic constraints on mobility  
using **complete information on a human society**
- 2) Unveil mechanism of mobility: **Time-order**

# Establishing a socio-economic laboratory

[www.pardus.at](http://www.pardus.at)

400,000 participants live an alternative life,  
in an online society interacting with others

- trading
- socializing
- conflicts

since 14 years



B A V D O 2

## Complete data on human society!

Here: use to study mobility

Bainbridge, Science 317, 472 (2007)  
Szell and Thurner, Social Networks 32, 313-329 (2010)  
Szell et al, PNAS 107, 1363-13641 (2010)

# Establishing a socio-economic laboratory

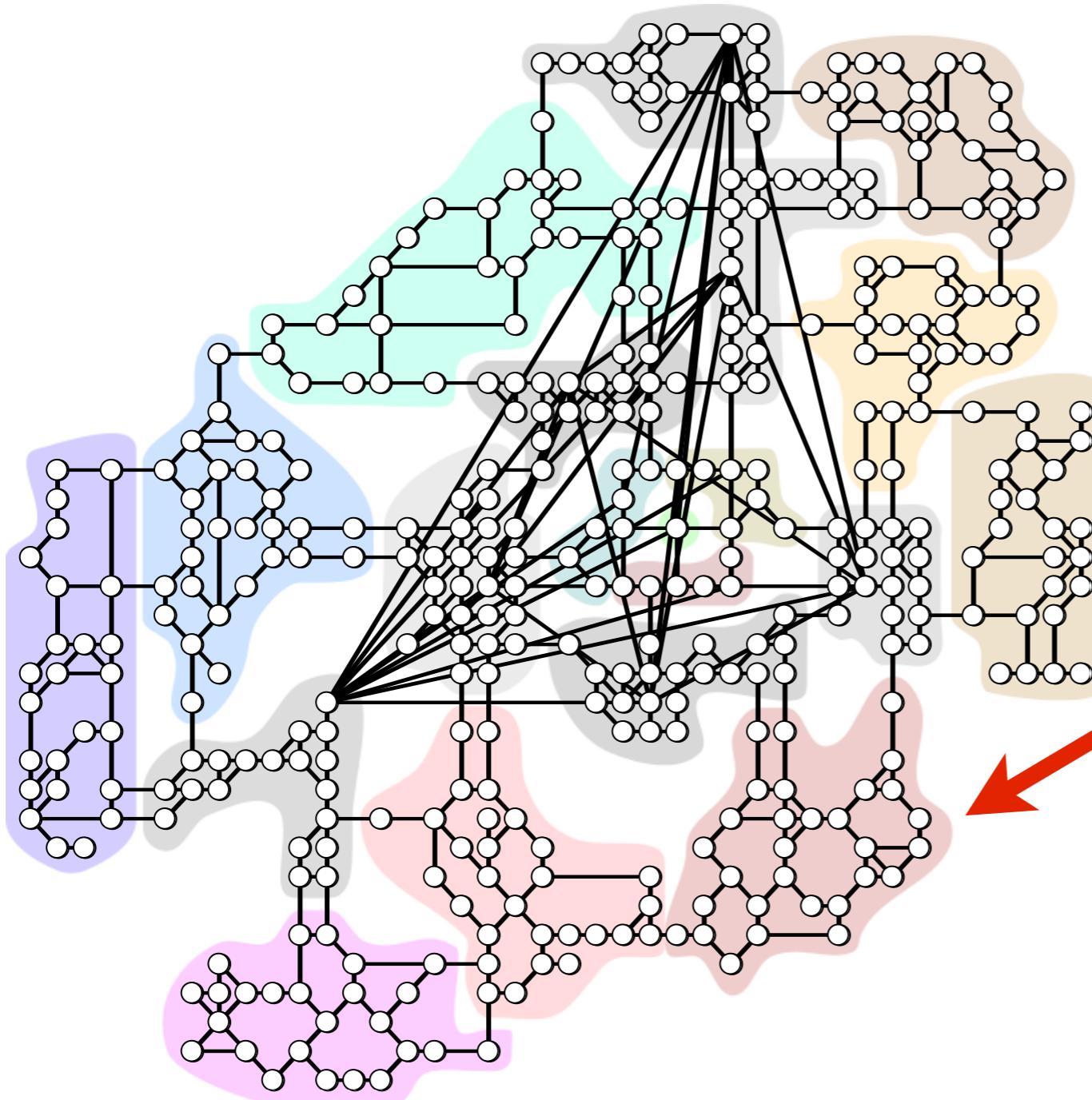
[www.pardus.at](http://www.pardus.at)



БУВДО 2

# The universe of the game

## Network

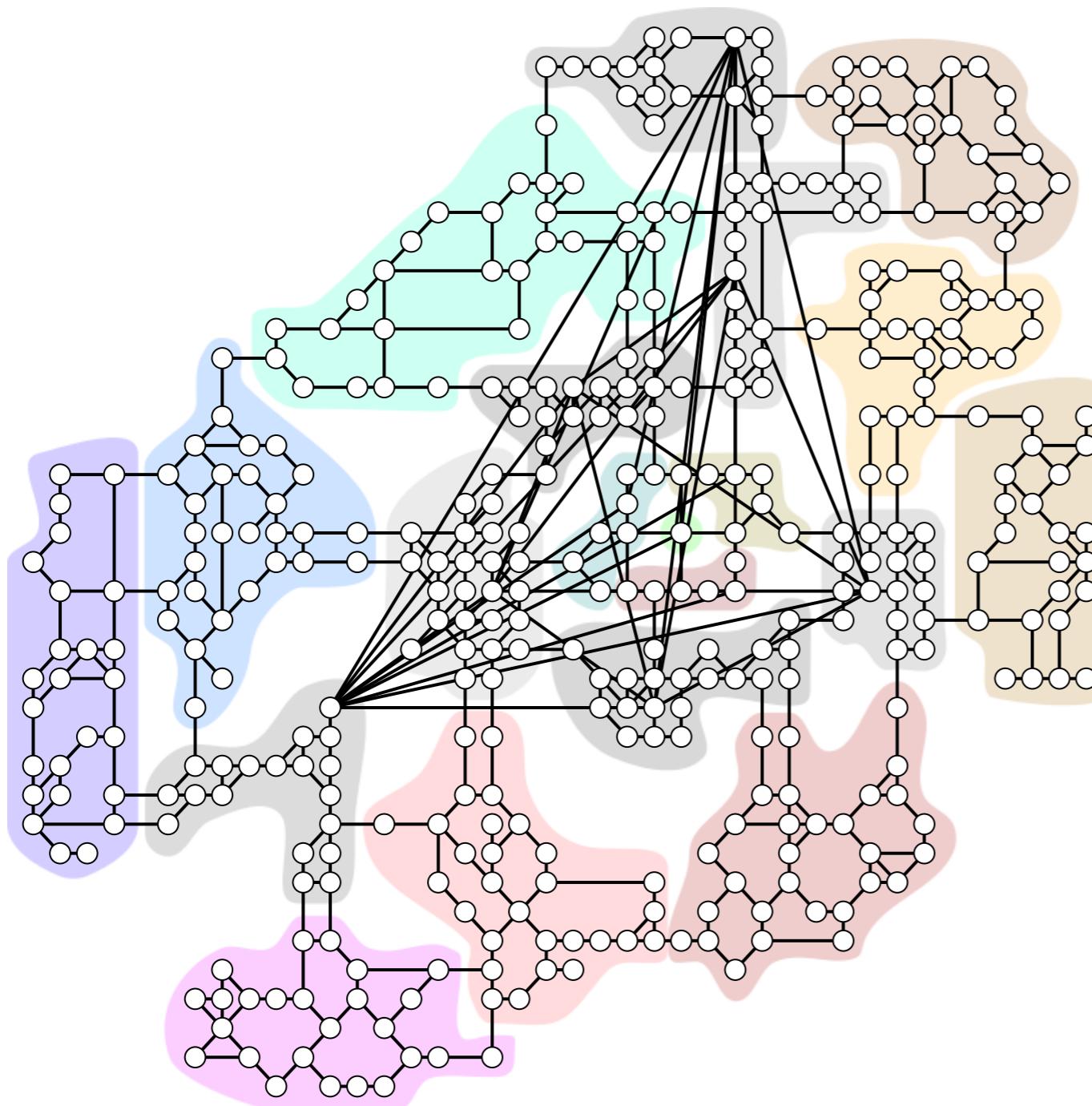


- 400 Nodes, “Cities”
- Diameter = 27
- Lattice-like

**20 Clusters  
“Countries”**

# Mobility on a network

Day-to-day mobility of active players over 1000 days

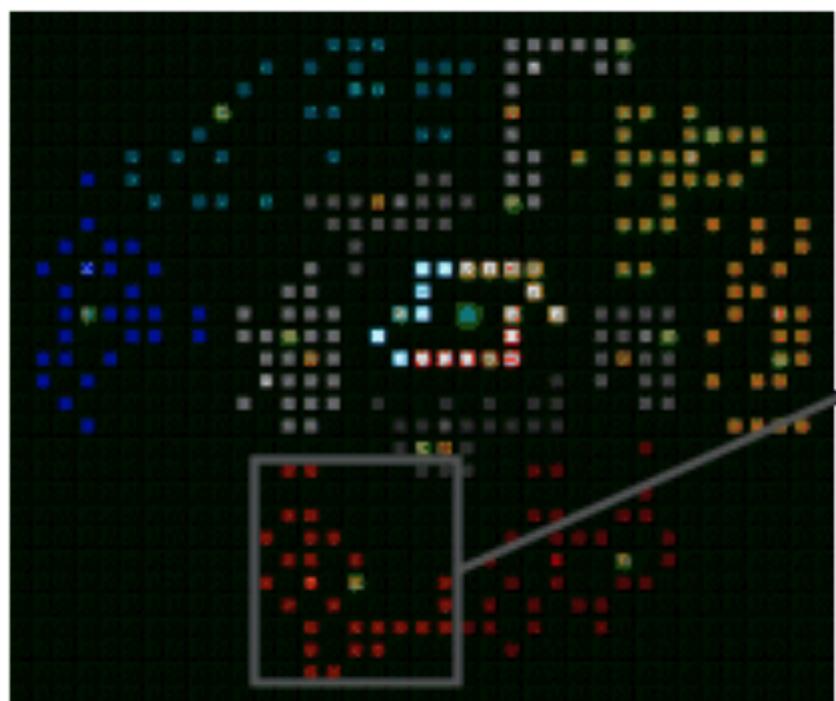


One mode of  
transportation

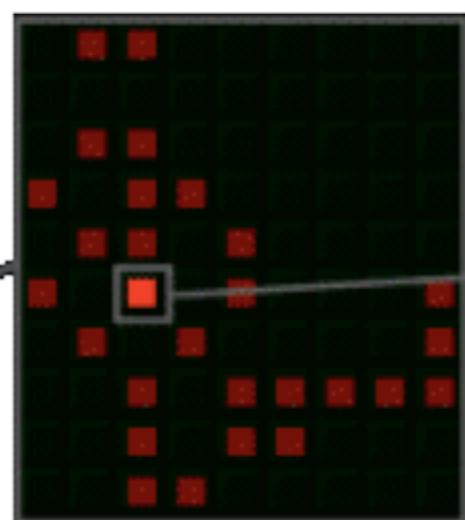


Move along links

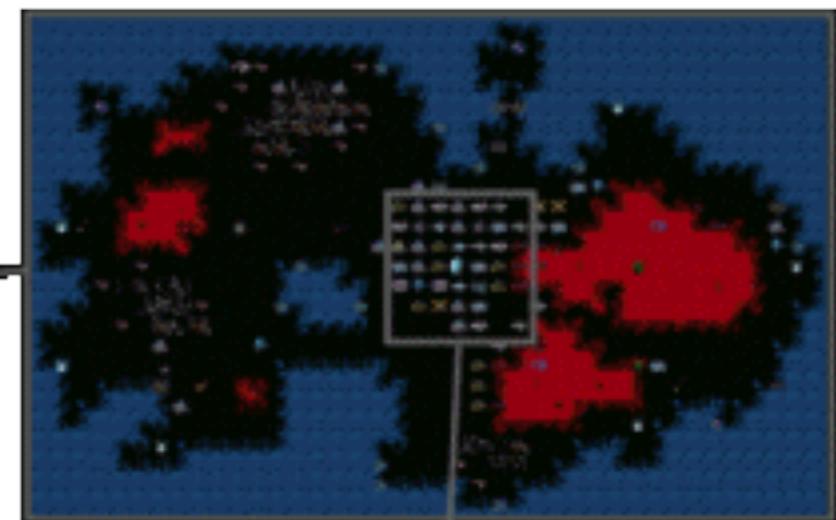
**Universe**



**Cluster**



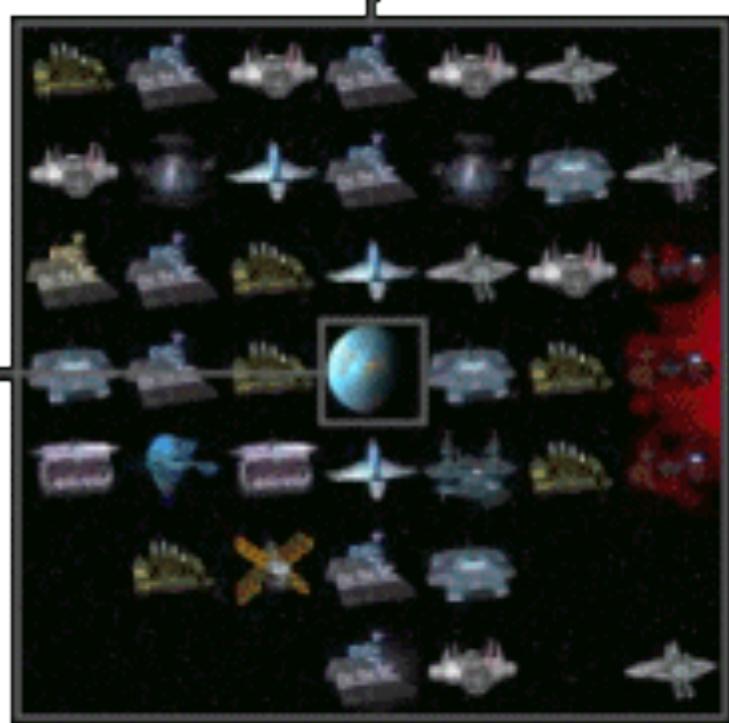
**Sector**



**Field**

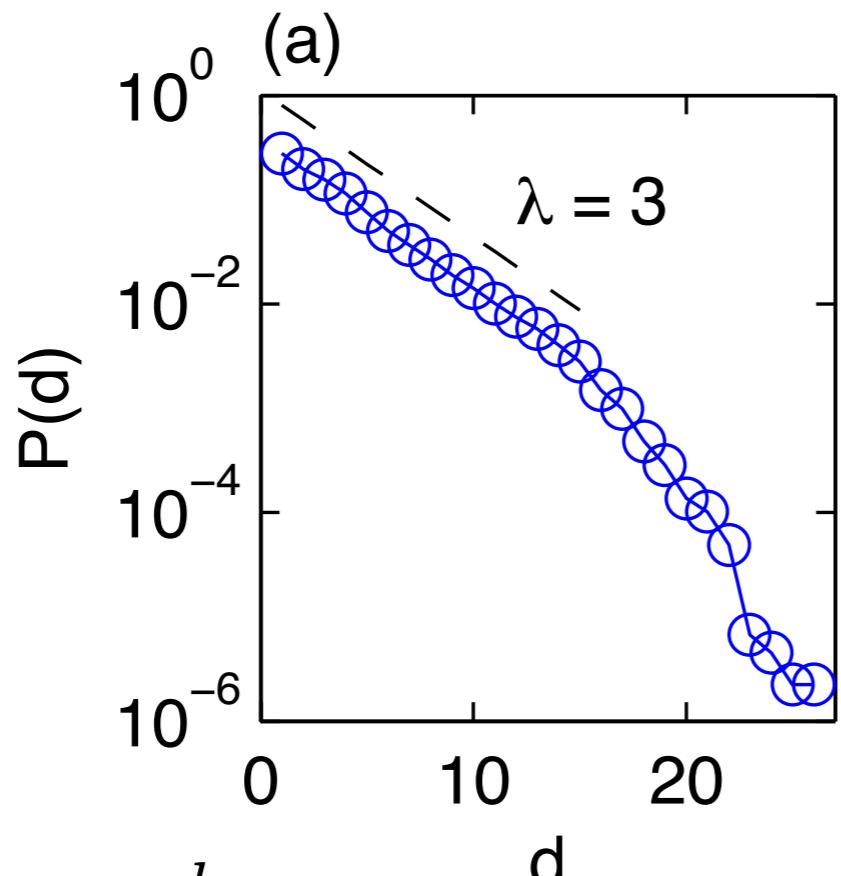


**7x7 Fields (Space Char)**



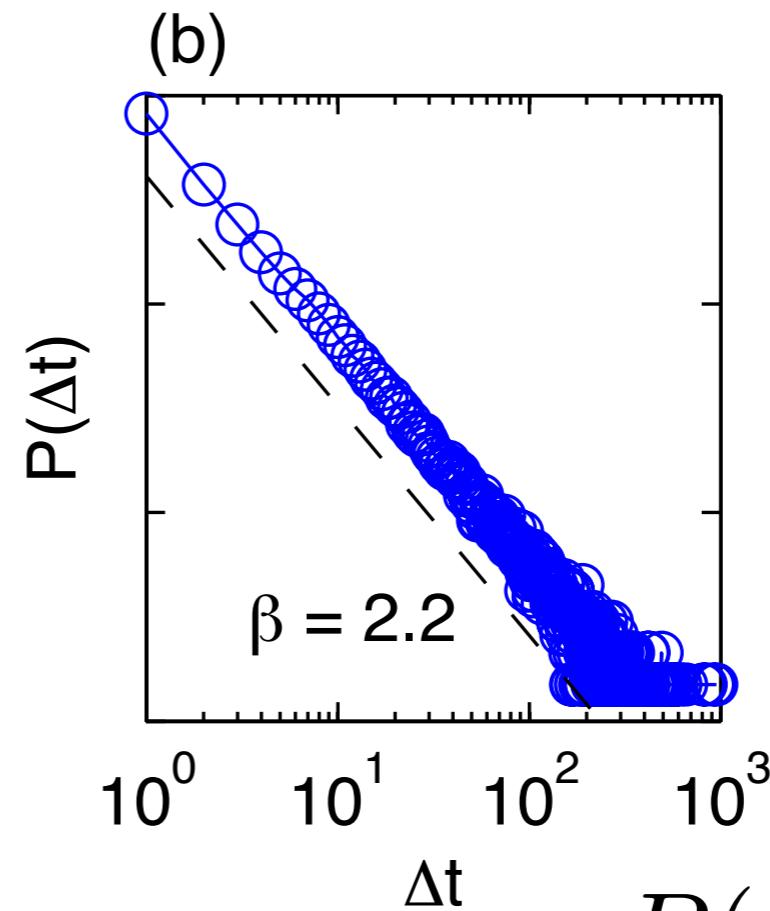
# Basic features of motion

Trip lengths  
exponential



$$P(d) \sim e^{-\frac{d}{\lambda}}$$

Waiting times  
power



$$P(\Delta t) \sim \Delta t^{-\beta}$$

Brockmann et al, Nature 439, 462-465 (2006)

Bazzani et al, J Stat, P05001 (2010)

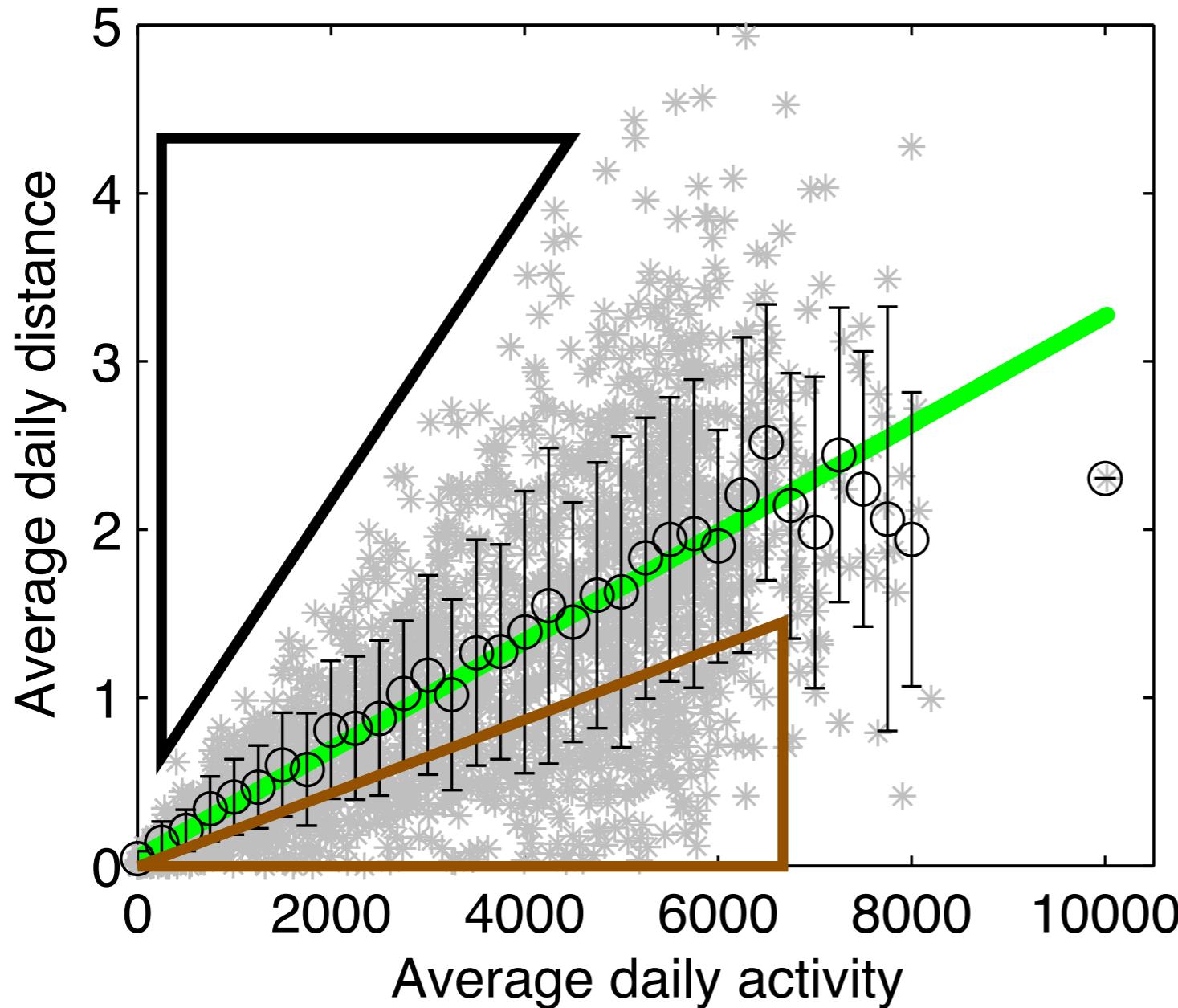
Roth et al, PLoS One 6, e15923 (2011)

Song et al, Nature Physics 6, 818-823 (2010)

Koelbl and Helbing, New J of Phys 5, 48 (2003)

Han et al, PRE 83, 036117 (2011)

# Mobility versus Activity

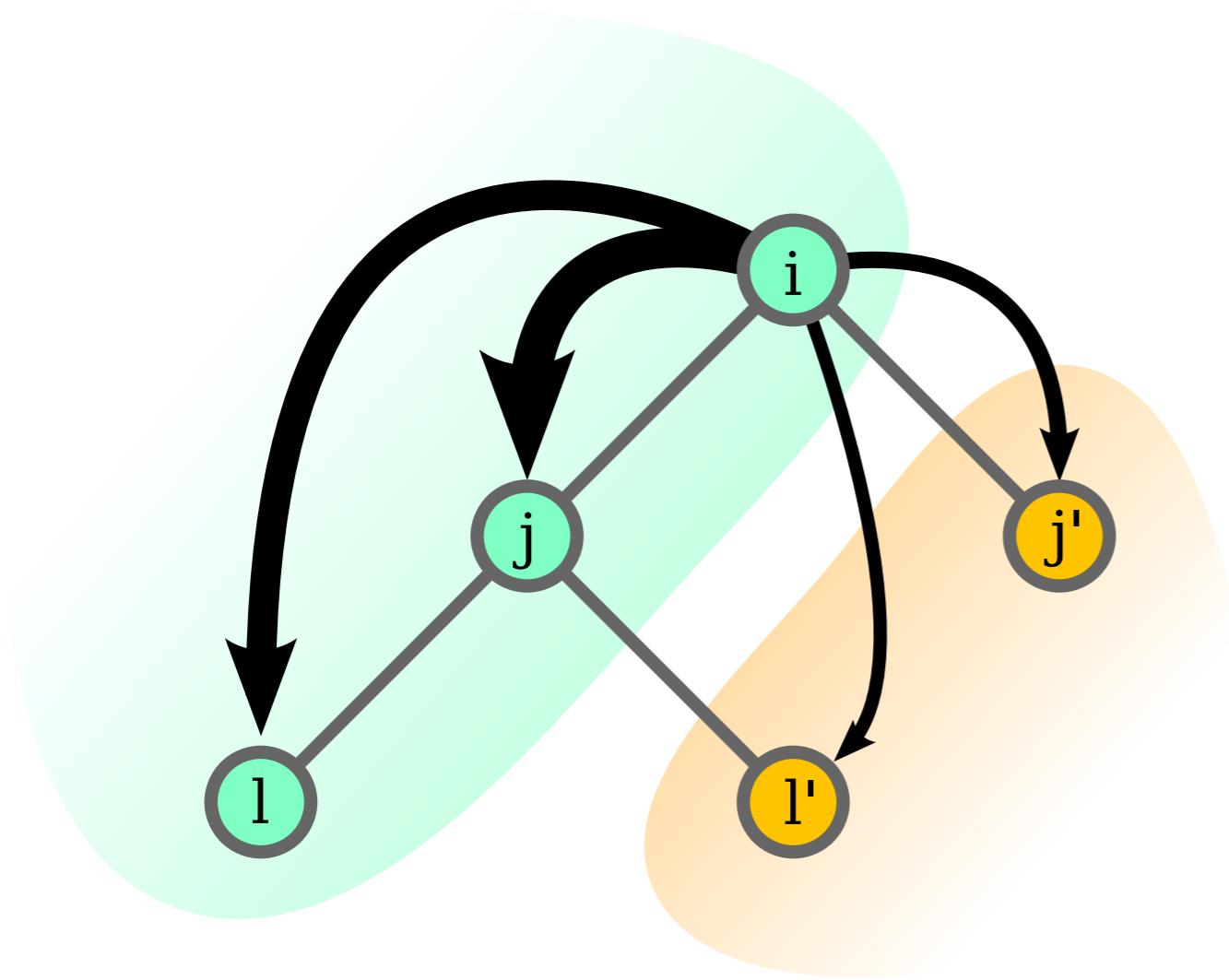


**Mobility  $\Rightarrow$  Activity**

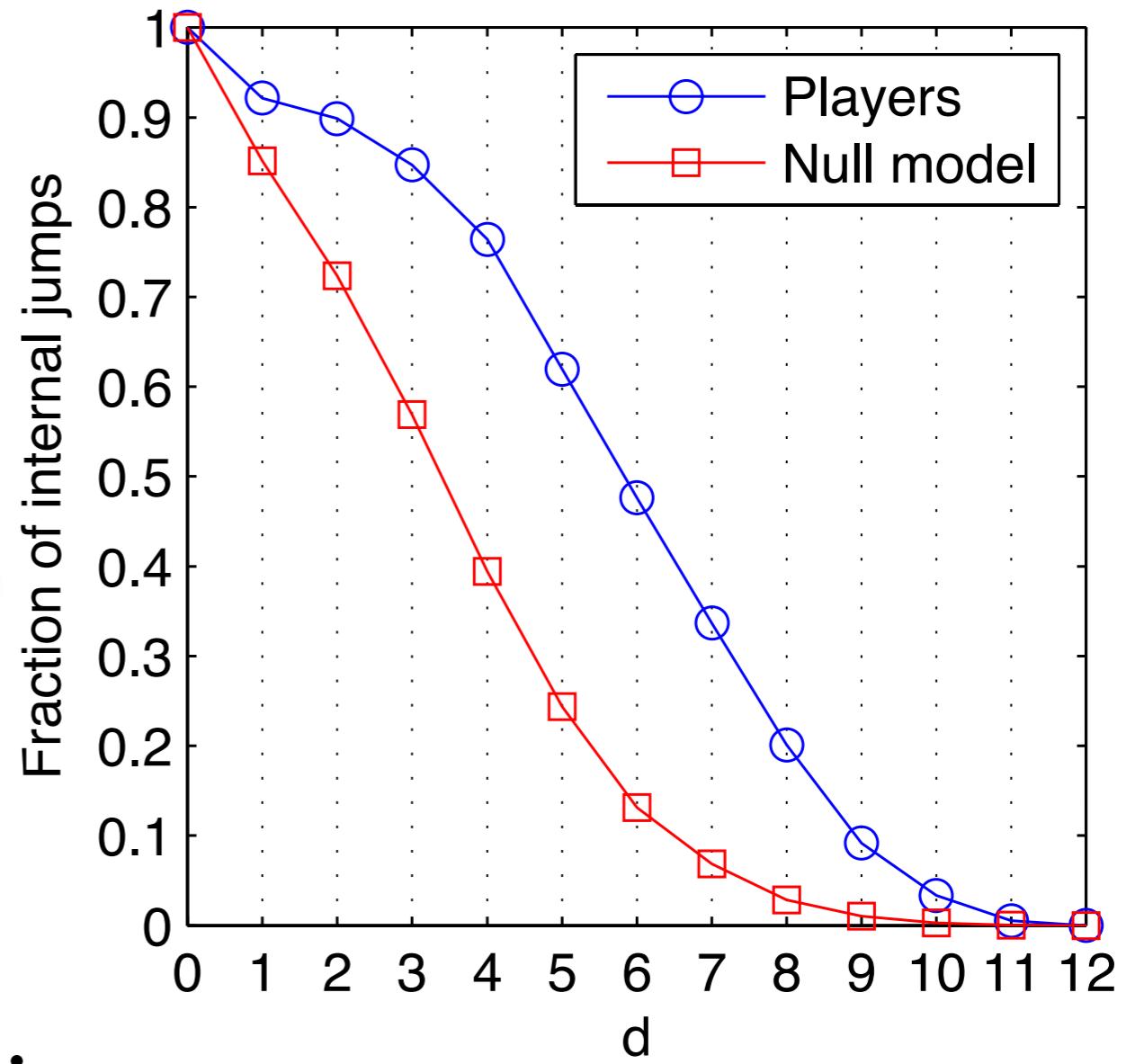
**Activity  $\not\Rightarrow$  Mobility**

**Players are highly heterogeneous**

# Do clusters influence mobility?



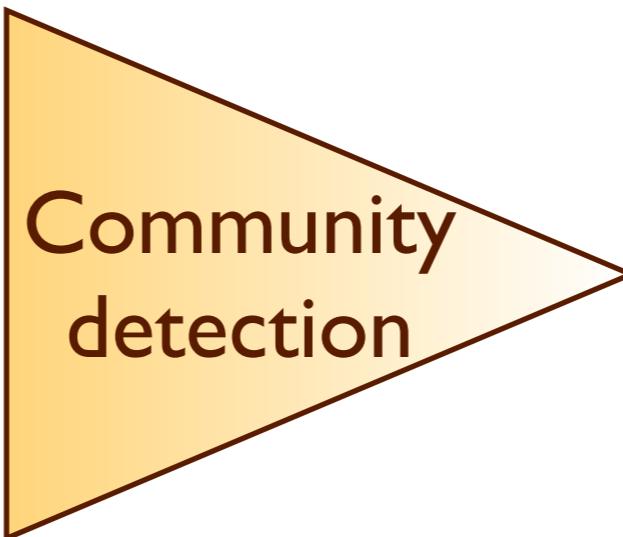
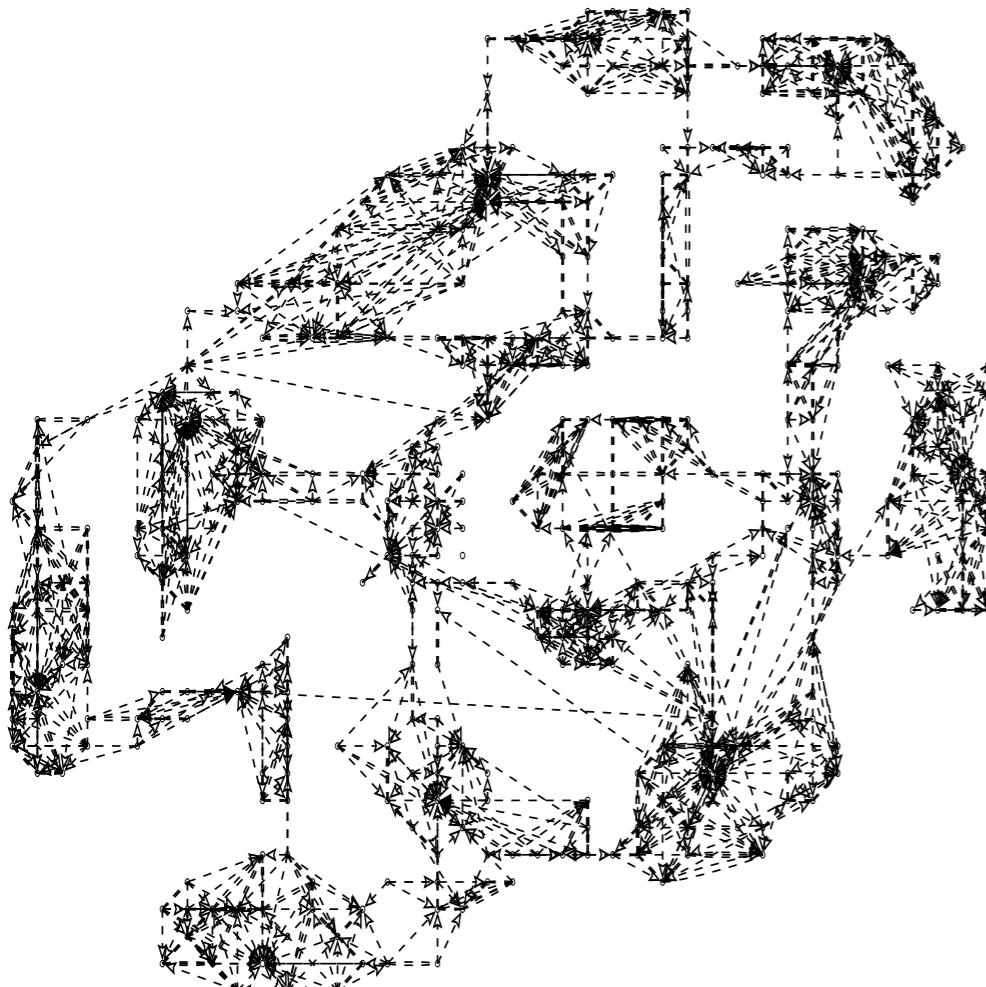
More movements within  
than between “countries”?



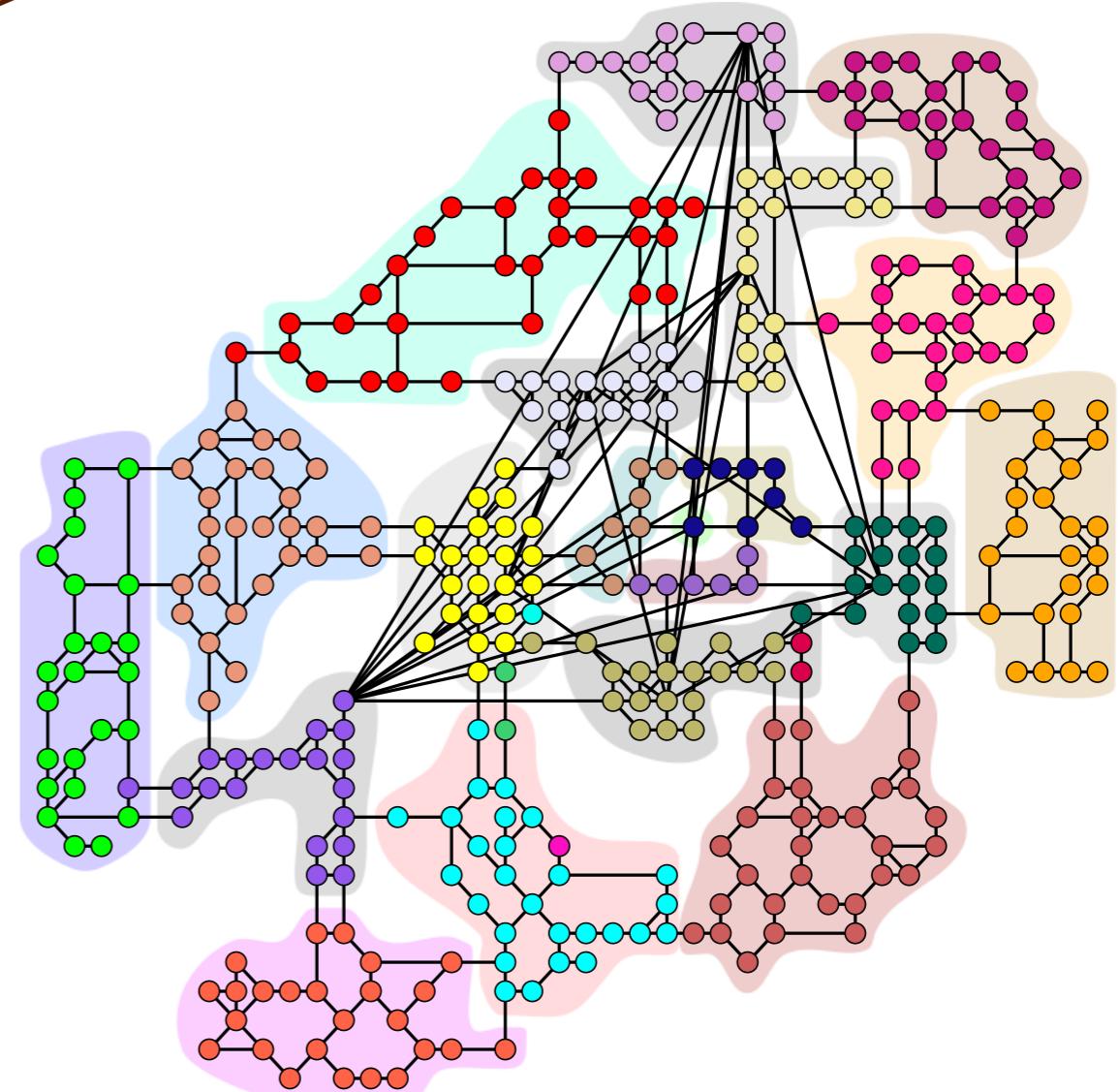
YES

# Mobility reveals socio-economic clusters

Movement data



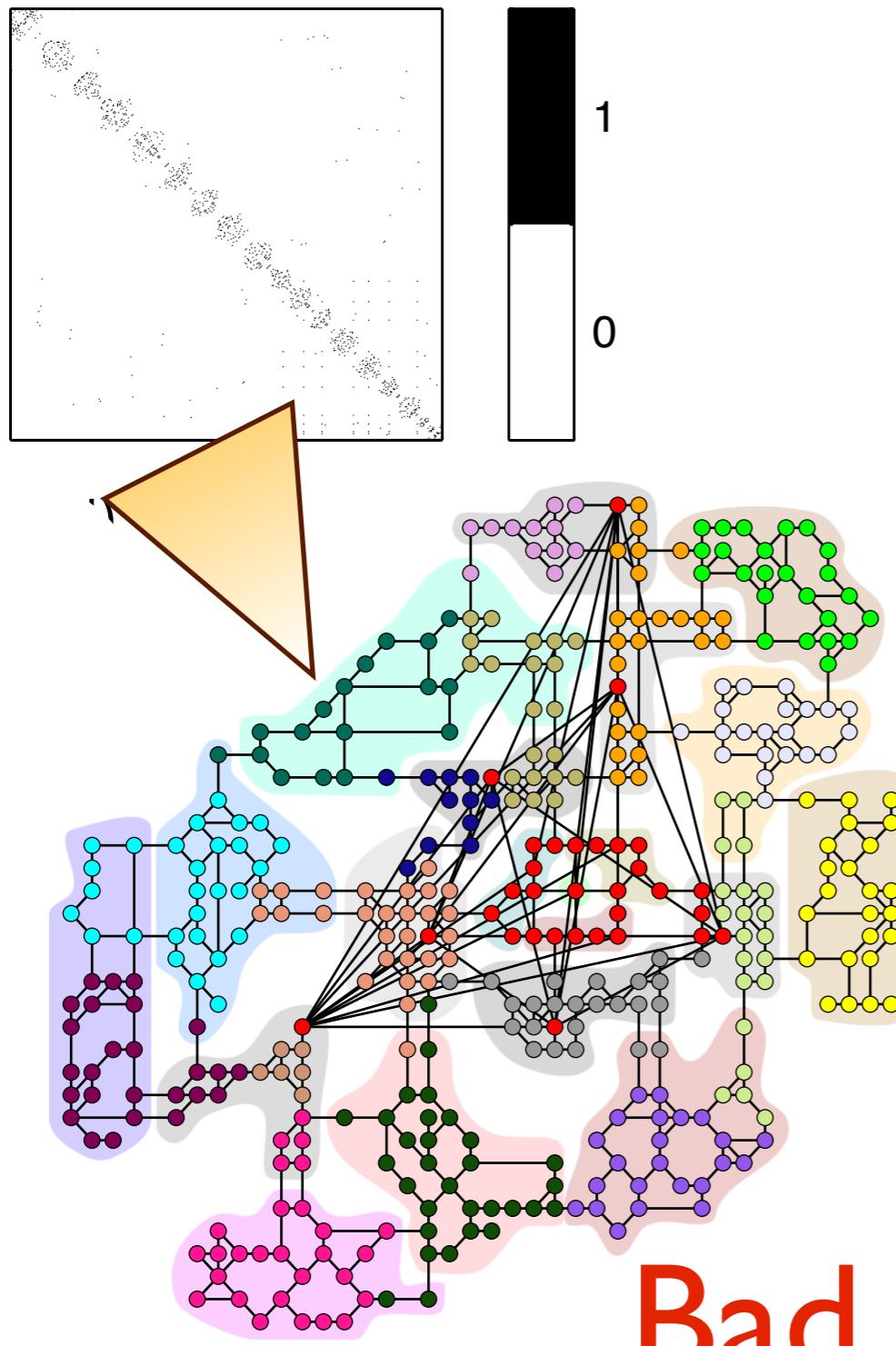
Clusters



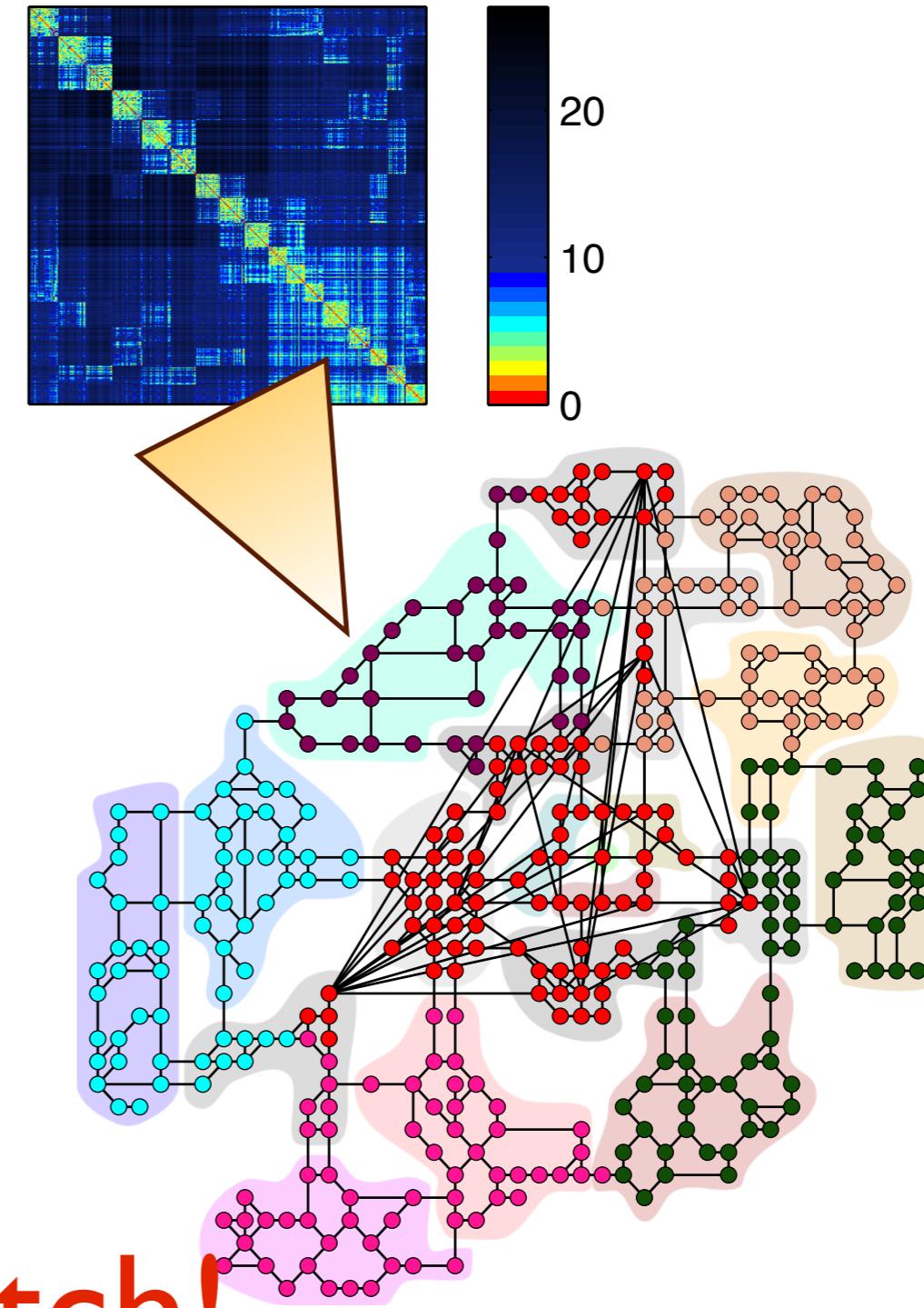
(almost) Perfect match!

# For comparison: Communities from topology

Adjacency A



Distance D

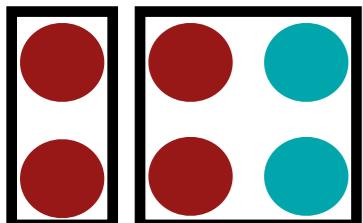


Bad match!

# Measures for quantifying matching of partitions

Many measures that quantify the match between two partitions  $X$  and  $Y$  build on these 4 numbers:

- 3     $n_{11}$     number of pairs of elements in the same community under both  $X$  and  $Y$
- 4     $n_{00}$     number of pairs of elements not in the same community under both  $X$  and  $Y$
- 4     $n_{01}$     number of pairs of elements not in the same community under  $X$  but in the same community under  $Y$
- 4     $n_{10}$     number of pairs of elements in the same community under  $X$  but not in the same community under  $Y$

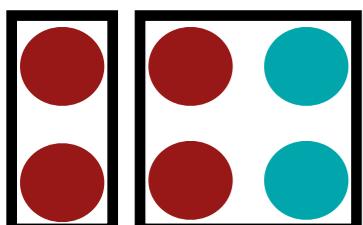


$X$     Colors  
 $Y$     Boxes

# Measures for quantifying matching of partitions

Many measures that quantify the match between two partitions  $X$  and  $Y$  build on these 4 numbers:

- 3     $n_{11}$     number of pairs of elements in the same community under both  $X$  and  $Y$
- 4     $n_{00}$     number of pairs of elements not in the same community under both  $X$  and  $Y$
- 4     $n_{01}$     number of pairs of elements not in the same community under  $X$  but in the same community under  $Y$
- 4     $n_{10}$     number of pairs of elements in the same community under  $X$  but not in the same community under  $Y$



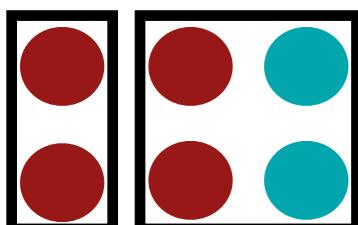
0.47    **Rand index**     $\mathcal{R} = \frac{n_{00} + n_{11}}{n_{00} + n_{11} + n_{01} + n_{10}}$

is the accuracy  
of the pair  
classification

# Measures for quantifying matching of partitions

Many measures that quantify the match between two partitions  $X$  and  $Y$  build on these 4 numbers:

- 3     $n_{11}$     number of pairs of elements in the same community under both  $X$  and  $Y$
- 4     $n_{00}$     number of pairs of elements not in the same community under both  $X$  and  $Y$
- 4     $n_{01}$     number of pairs of elements not in the same community under  $X$  but in the same community under  $Y$
- 4     $n_{10}$     number of pairs of elements in the same community under  $X$  but not in the same community under  $Y$



0.43

Fowlkes-  
Mallows  
index

$$\mathcal{F} = \sqrt{\frac{n_{11}}{n_{11} + n_{10}} \cdot \frac{n_{11}}{n_{11} + n_{01}}}$$

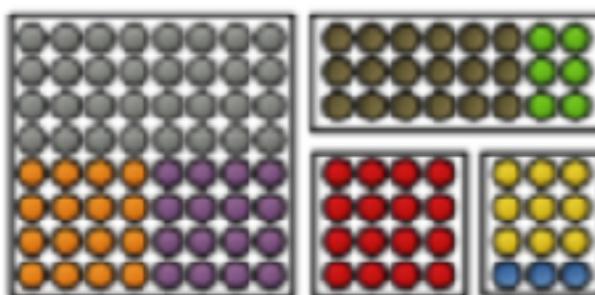
# Measures for quantifying matching of partitions

Problem: If you reshuffle the data randomly, you wont get 0, but a baseline  $> 0$  for the random expectation.

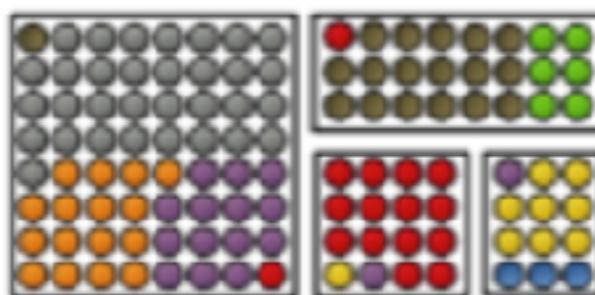
To correct for this baseline, there is an **adjusted Rand index**

$$\mathcal{ARI} = \frac{n_{00} + n_{11} - \mathbf{E}[n_{00} + n_{11}]}{n_{00} + n_{11} + n_{01} + n_{10} - \mathbf{E}[n_{00} + n_{11}]}$$

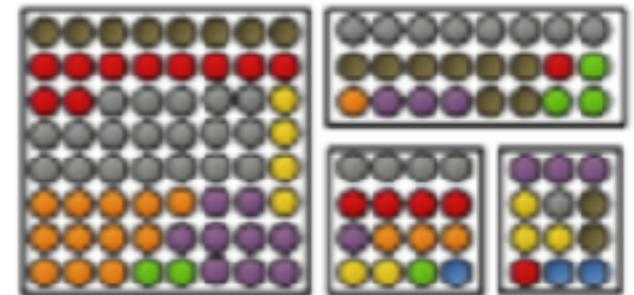
where  $\mathbf{E}[n_{00} + n_{11}]$  are the values expected by chance.  
This index can be negative.



$$\mathcal{ARI} = 1.00$$



$$\mathcal{ARI} = 0.88$$



$$\mathcal{ARI} = 0.03$$

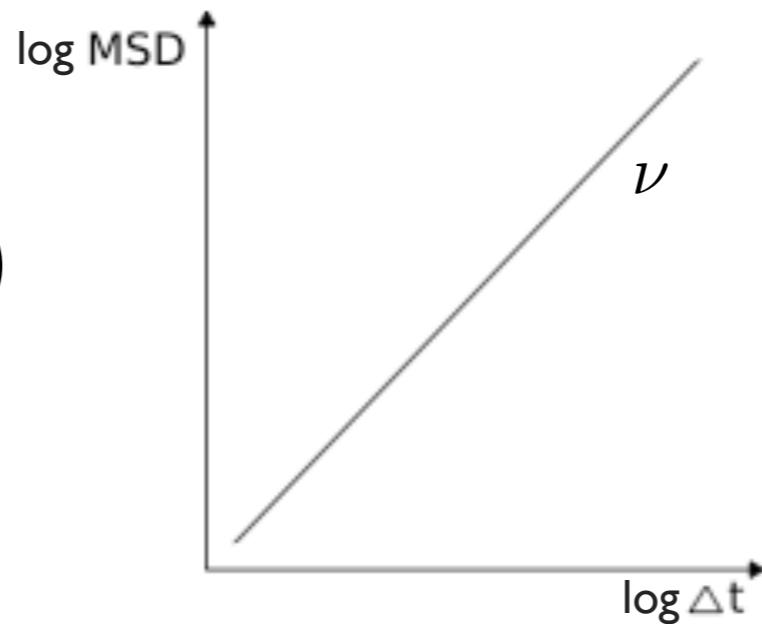
# Diffusion and MSD

## Mean Squared Displacement

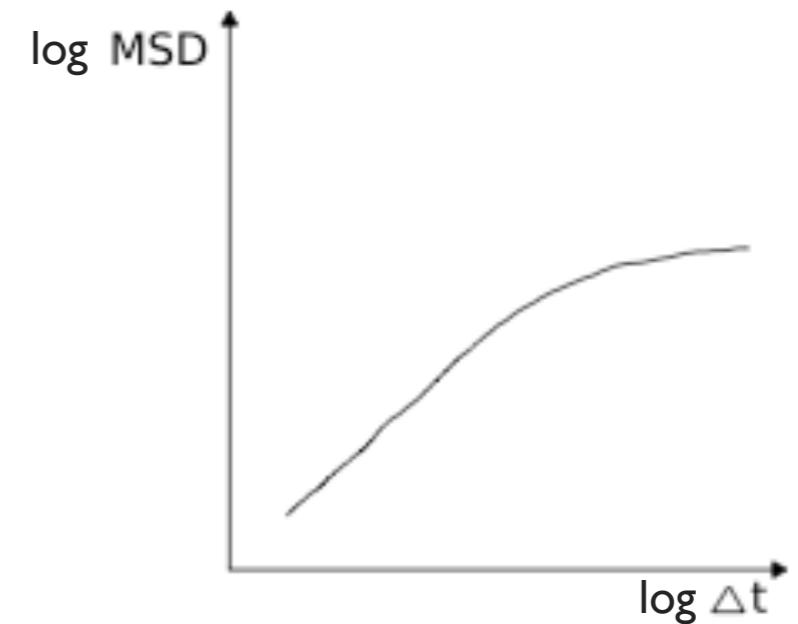


finite universe

$$\nu = \lim_{t \rightarrow \infty} \frac{d}{dt} (\text{MSD})$$



$$\text{MSD} \equiv \langle (x - x_0)^2 \rangle = \frac{1}{N} \sum_{n=1}^N (x_n(t) - x_n(0))^2$$



expect finite size effect

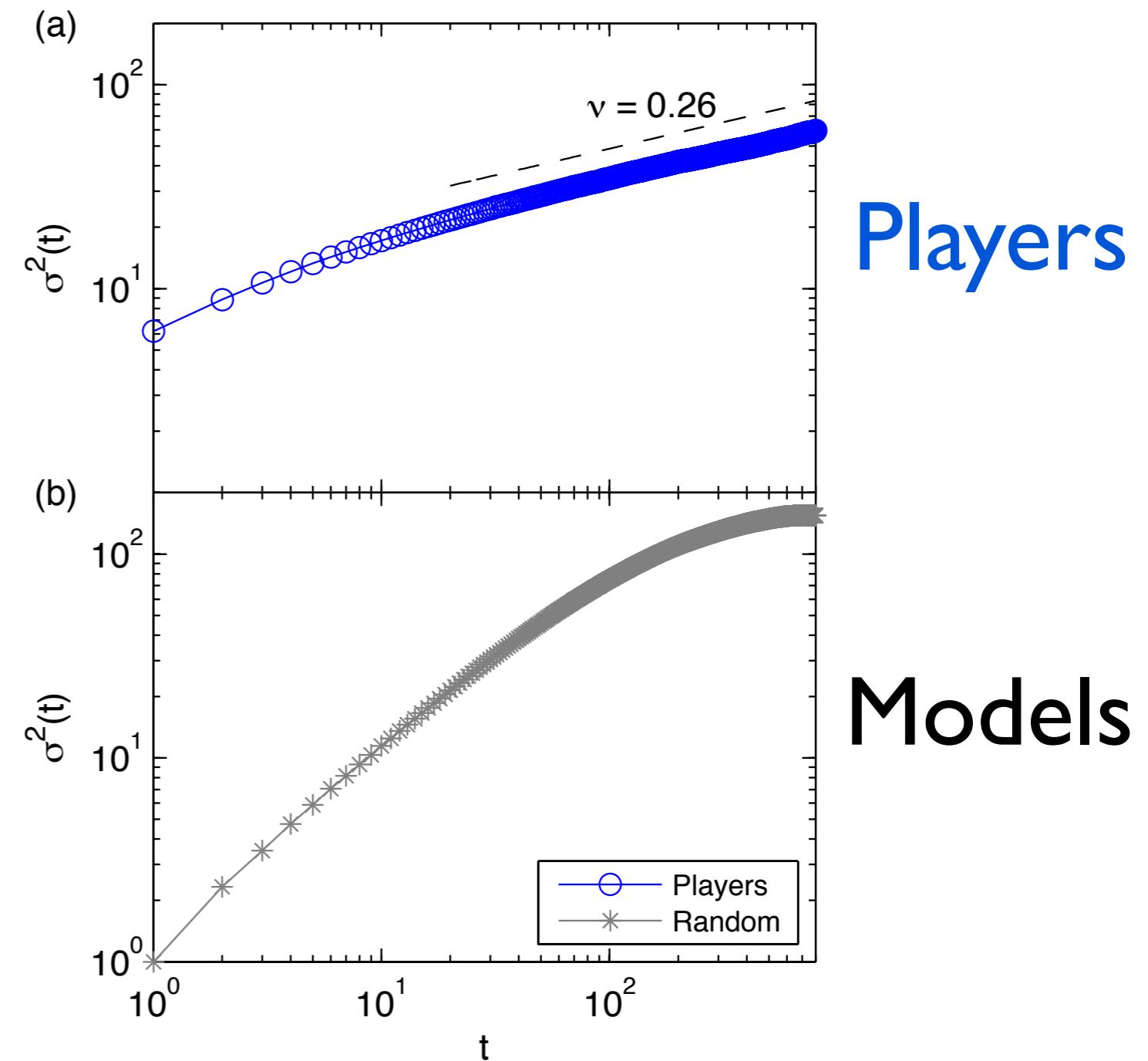
# Anomalous diffusion

$\nu = 0.26 < 1$

**Subdiffusive**

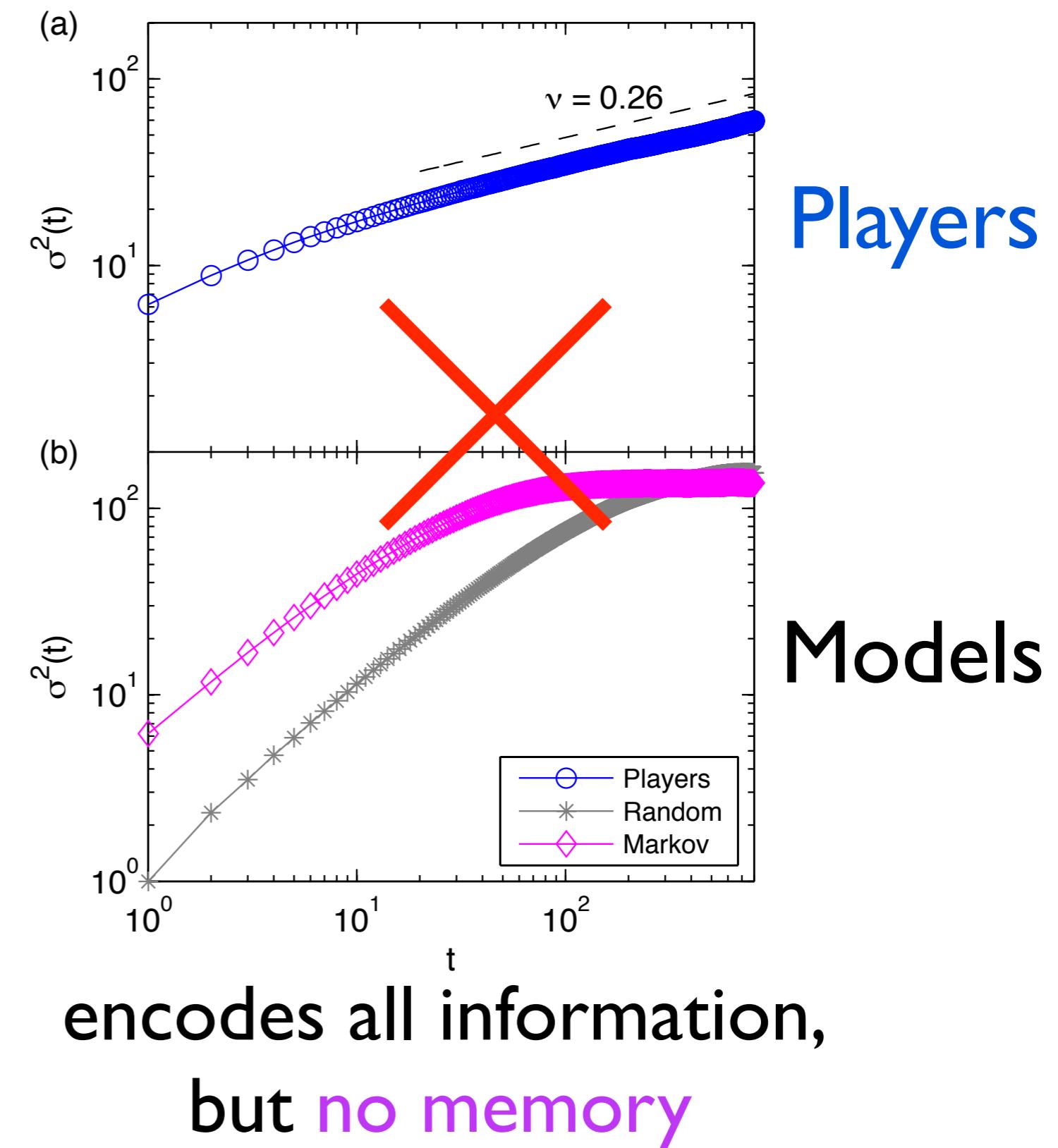
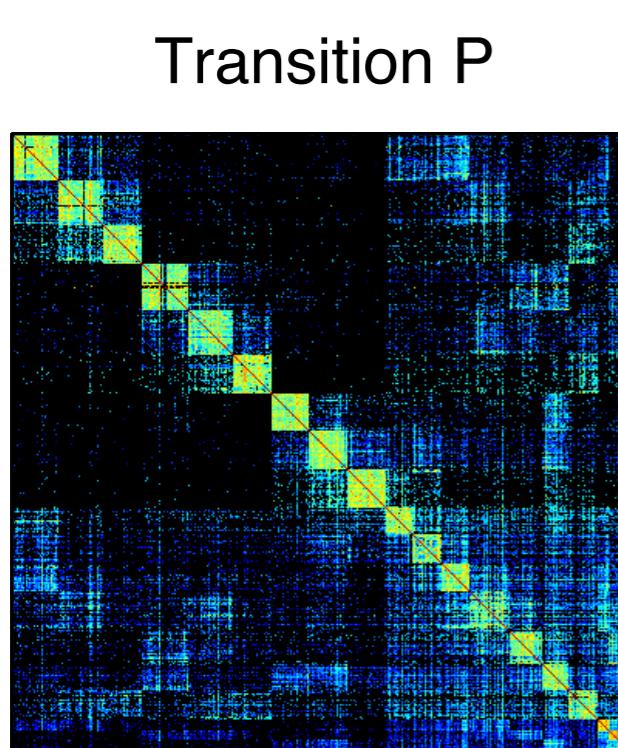
Random walkers

have  $\nu = 1$



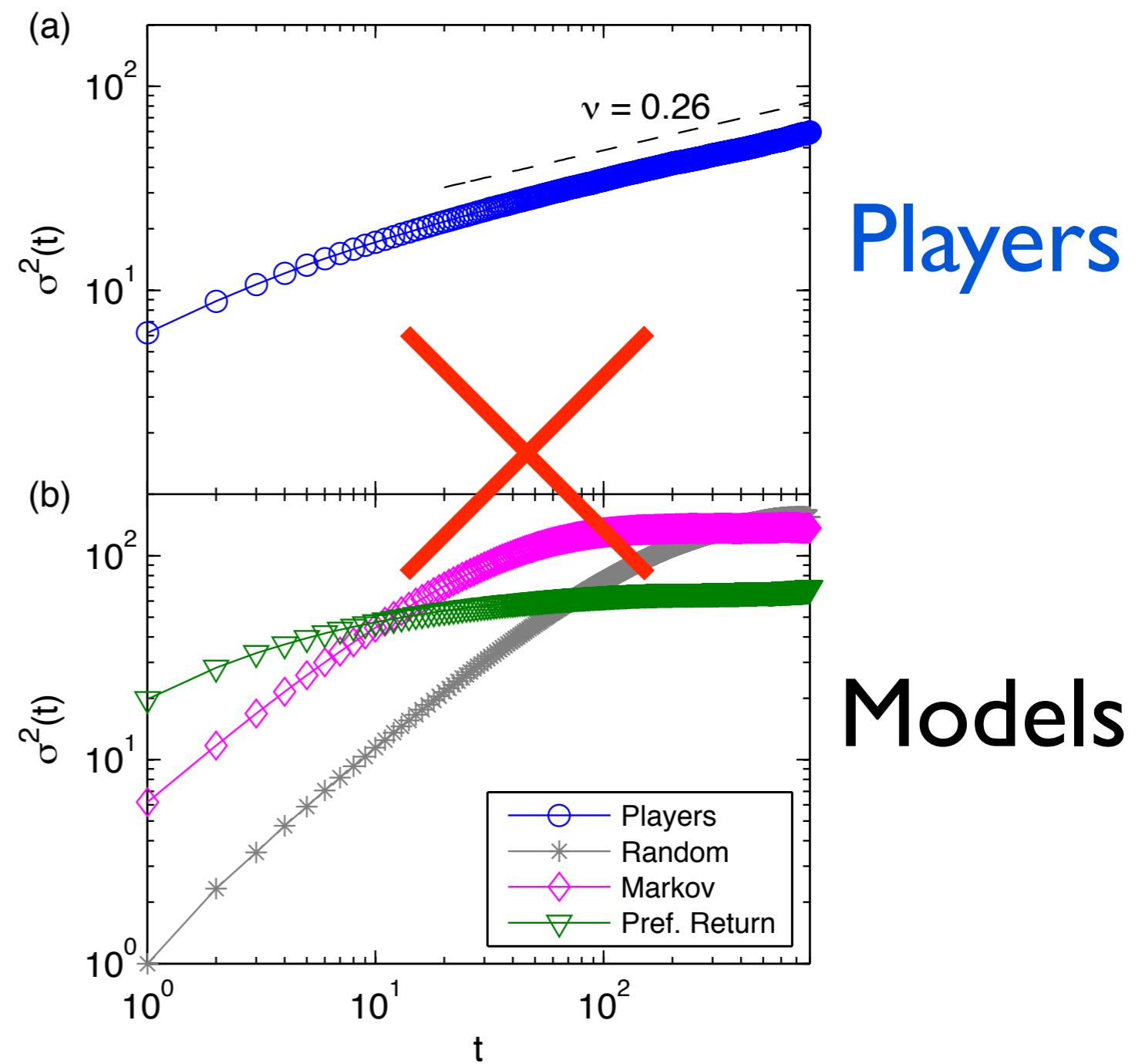
# Model I: Markov

- **Markov** use all day-to-day transitions



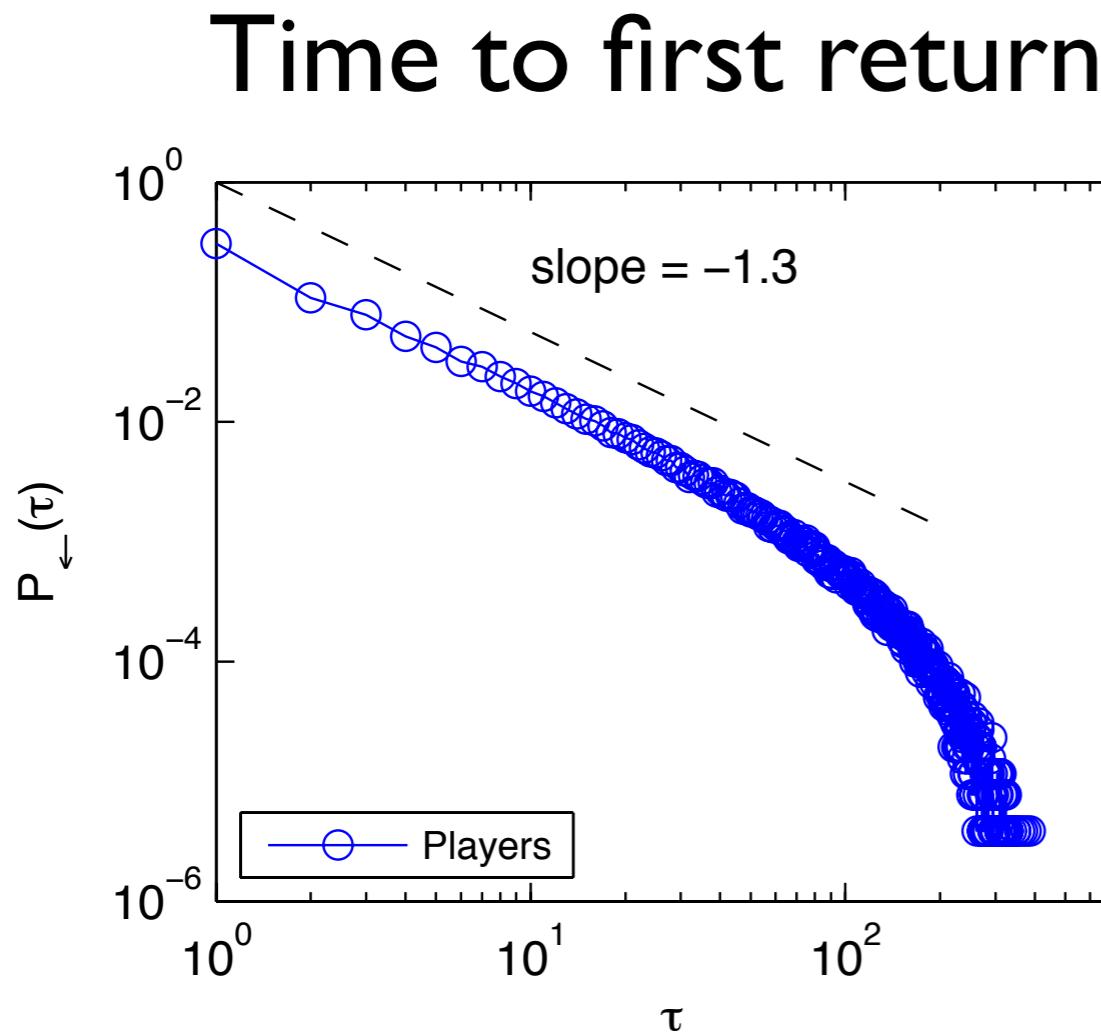
## Model II: Preferential return

- **Markov** use all day-to-day transitions
- Preferential return to often visited places



# What is the essential ingredient?

## Order of visitations!



...A B A B C D B A

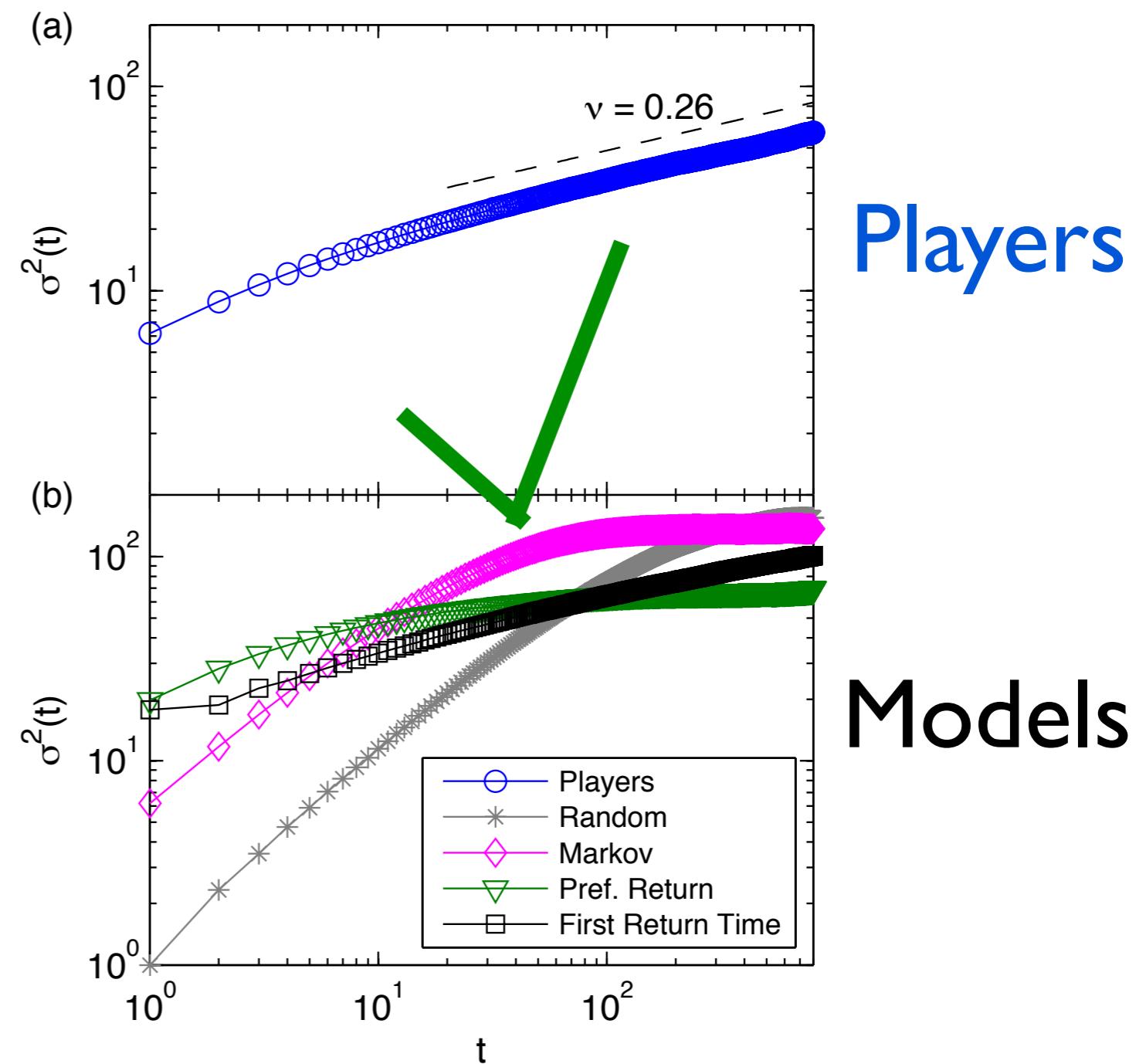


Long-time memory

$$P_{<}(\tau) \sim \tau^{-1.3}$$

# Our Model: First return

- **Markov** use all day-to-day transitions
- Preferential return to often visited places
- First return to recently visited places



## Our Model: First return

I) Draw waiting time  $\Delta t$  from  $\Delta t^{-\beta}$

$$\beta = 2.2$$

2a) With probability  $\nu$

Return to visited location with  
first return time  $\tau$  from  $\tau^{-\alpha}$

$$\nu = 0.9$$

$$\alpha = 1.3$$

2b) With probability  $1 - \nu$

Explore new location at  
distance  $d$  from  $e^{-\frac{d}{\lambda}}$

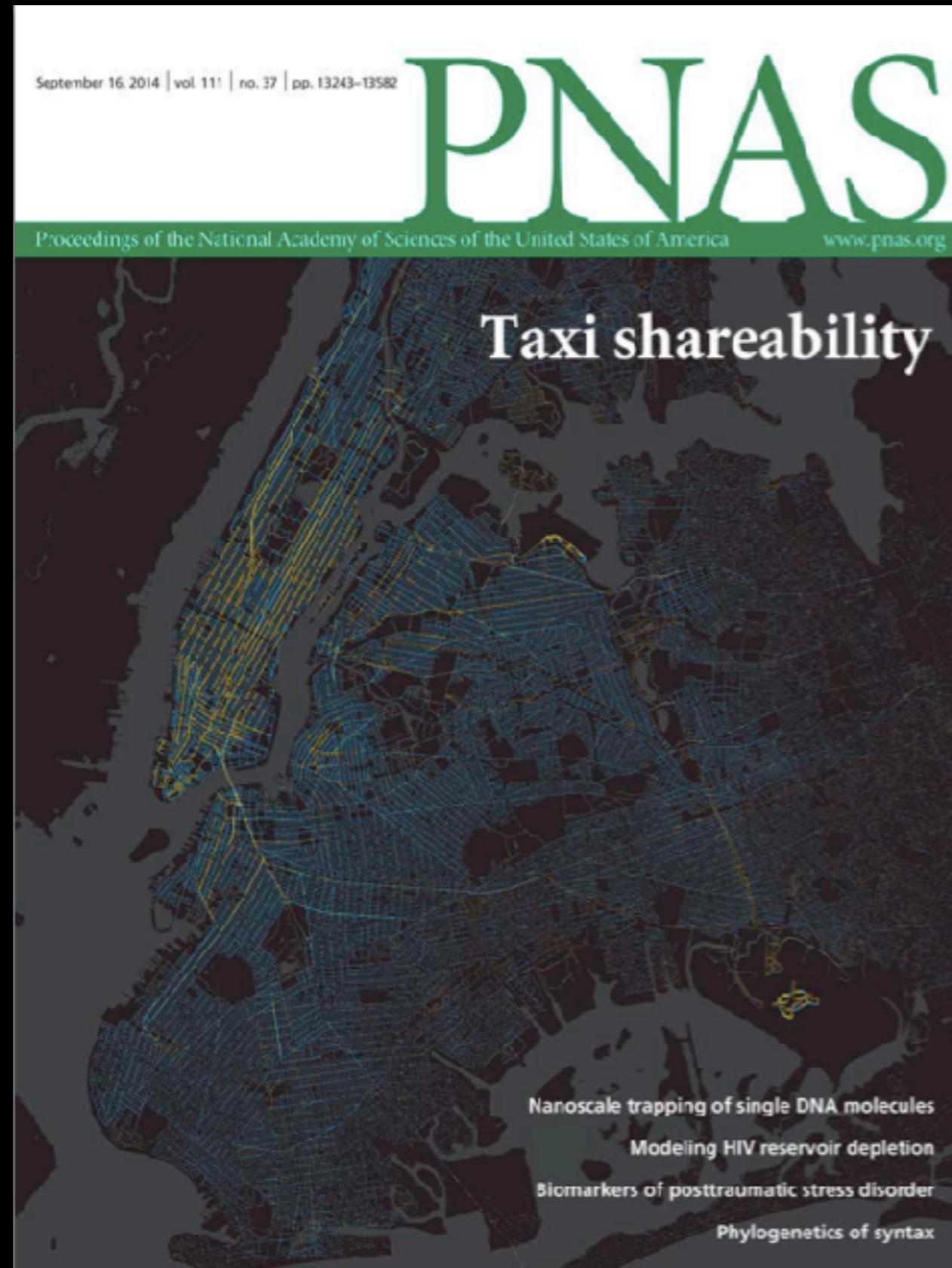
$$\lambda = 3$$

4 measured  
parameters

# Mobility in a socio-economic laboratory

- Socio-economic laboratory with complete information on a **human society**
- Movements reveal **socio-economic constraints**
- Driving mechanism: **Time-order** of visitations,  
**First Return Model**

2.



# Urban transportation is facing serious issues

Pollution



Traffic jams



Fatalities



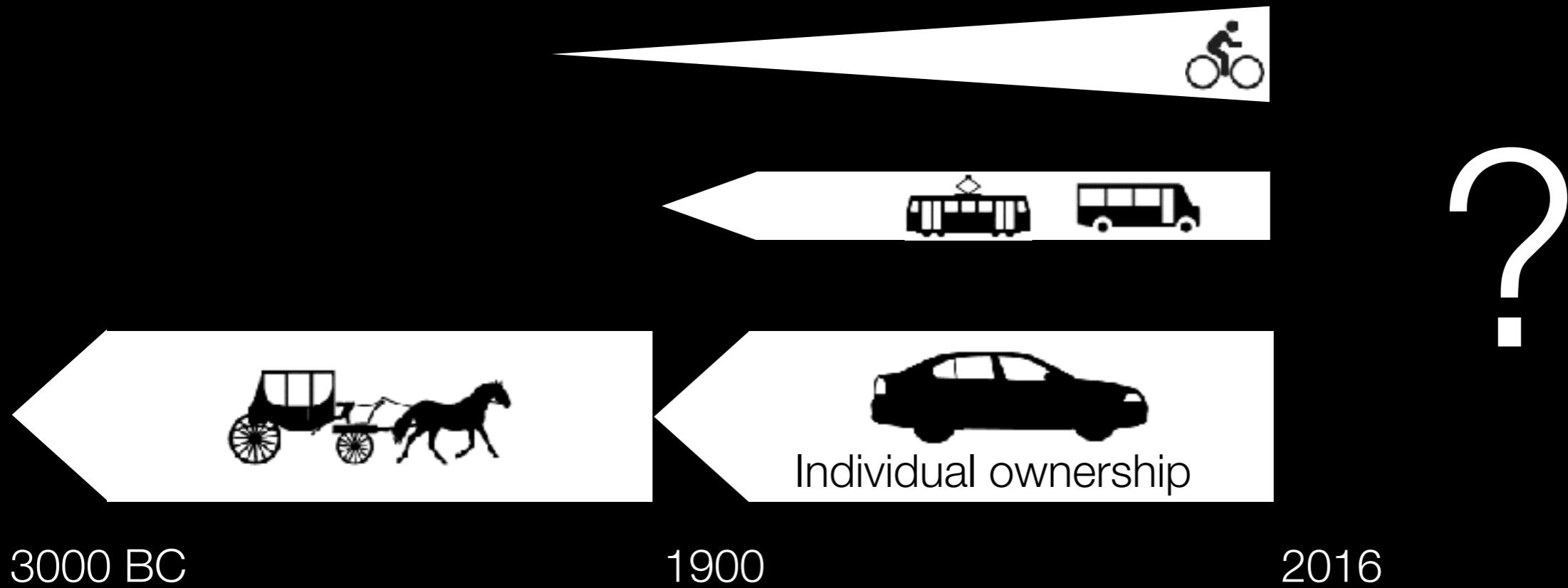
3.7 Mio.

1.2 Mio.

How efficient is urban transport?

We don't know.

# Urban transportation is at a turning point



# Urban transportation is changing

## Vehicle sharing



Ownership → Accessibility

Botsman, R. and Rogers, R. HarperCollins, New York (2010)  
Eckhardt, G.M. and Bardhi, F. Harvard Business Review (2015)

# Urban transportation is changing

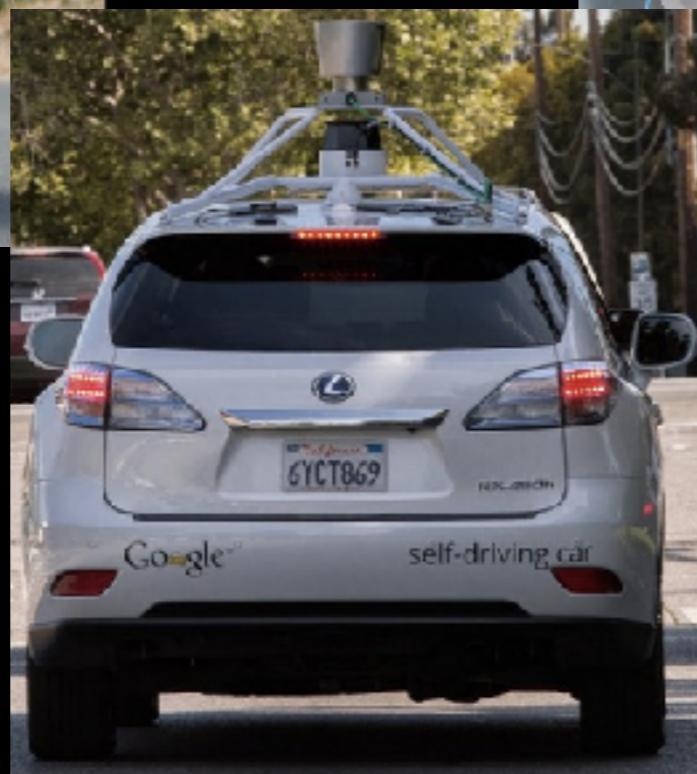
Algorithms and mobile technology  
to improve existing services



Ownership → Accessibility

# Urban transportation is changing

## Self-driving cars



Lack of driver further motivates giving up ownership

# The taxi system is inefficient



High emissions, cost, waiting times

# Real-time data sets are available

Can we use realtime data and mobile technology to

- 1) Quantify existing taxi systems
- 2) Design an improved system?



NYC  
13,500 cabs

# Real-time data sets are available

Can we use realtime data and mobile technology to

- 1) Quantify existing taxi systems
- 2) Design an improved system?



Goals

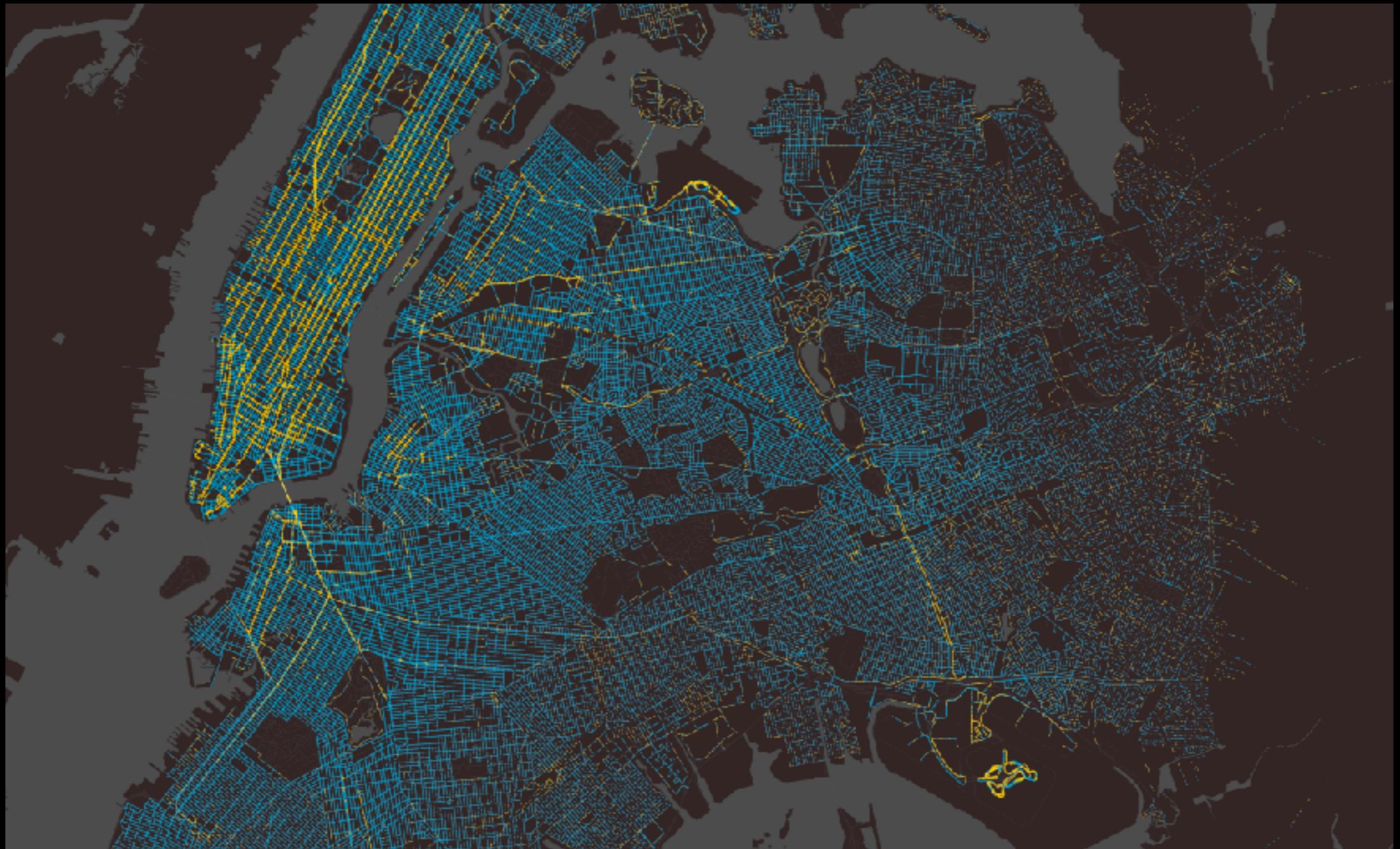
- more efficient
- less emissions
- affordable alternative

NYC

13,500 cabs

# Step 1: Analyze data

NYC taxi trips in 2011



13500 cabs

150 million trips ~400.000 per day

Pickups

Dropoffs

# Step 1: Analyze data

Pickups

Dropoffs

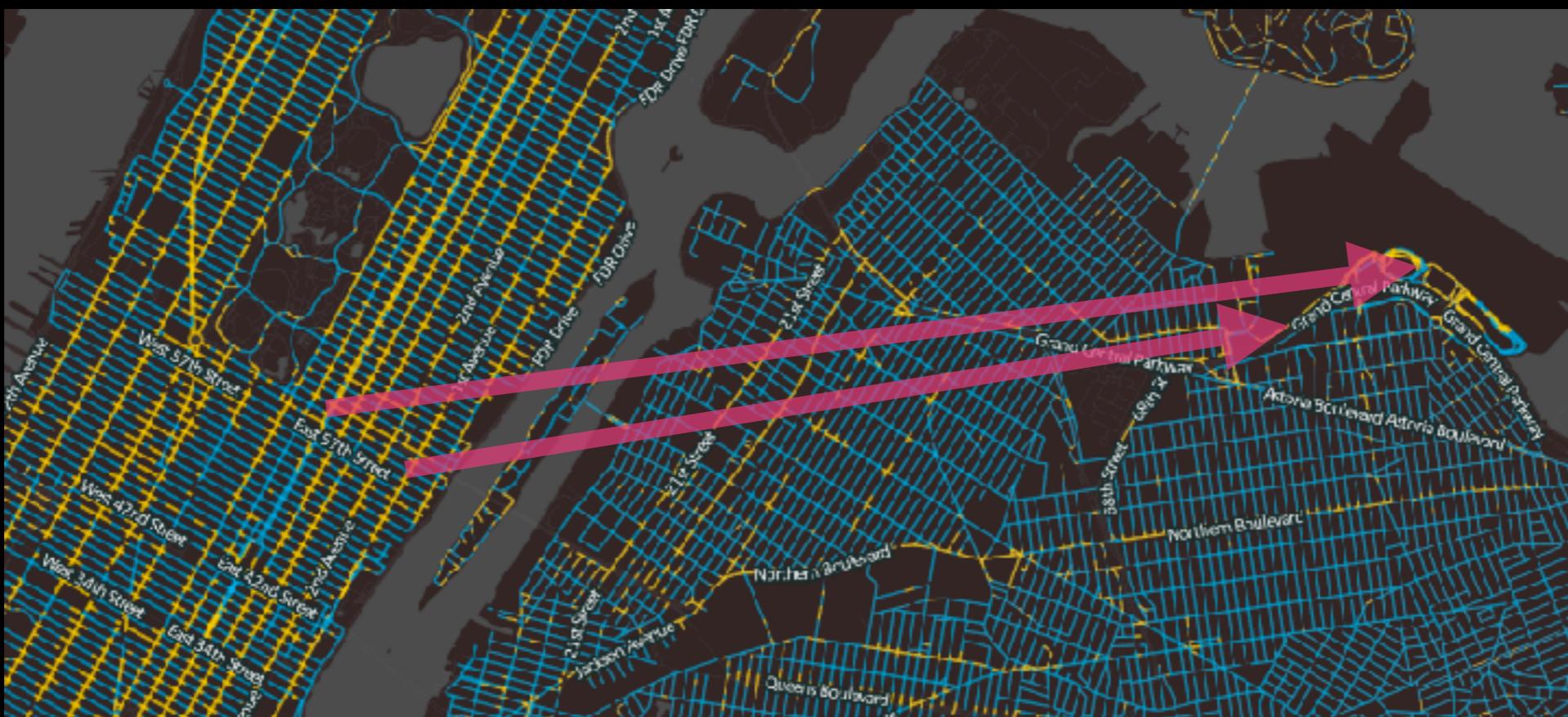


# Trips can be combined



# Step 2 : A new dispatch process

Combine 2 trips



# Step 2 : A new dispatch process

Combine **k** trips



# Street network



Extraction from  
OpenStreetMap

9000 street  
segments

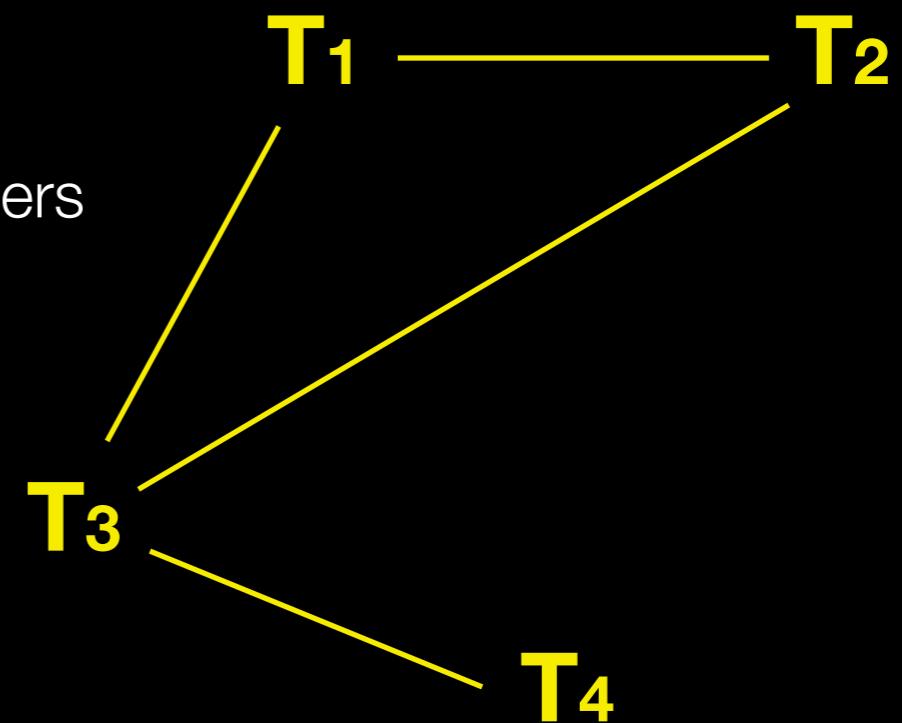
Matching GPS-  
coordinates

# Shareability network strategy

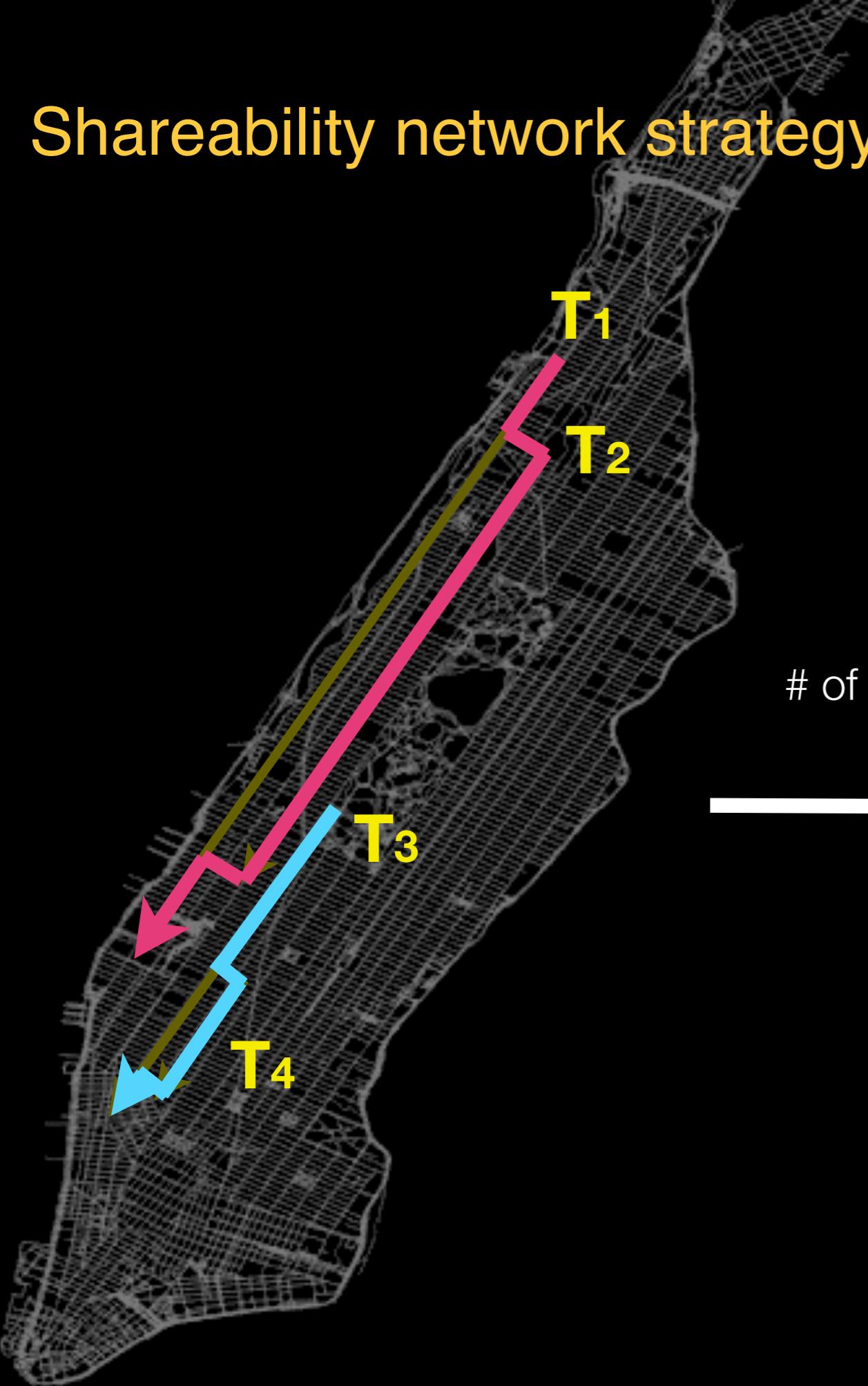


Mapping

# of passengers  
 $k = 2$

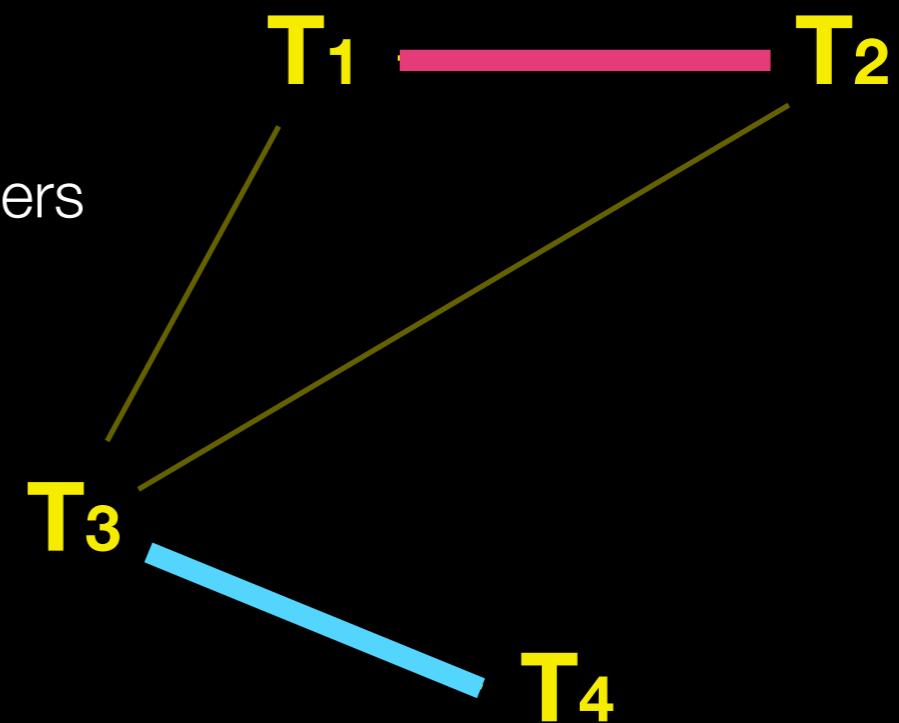


# Shareability network strategy



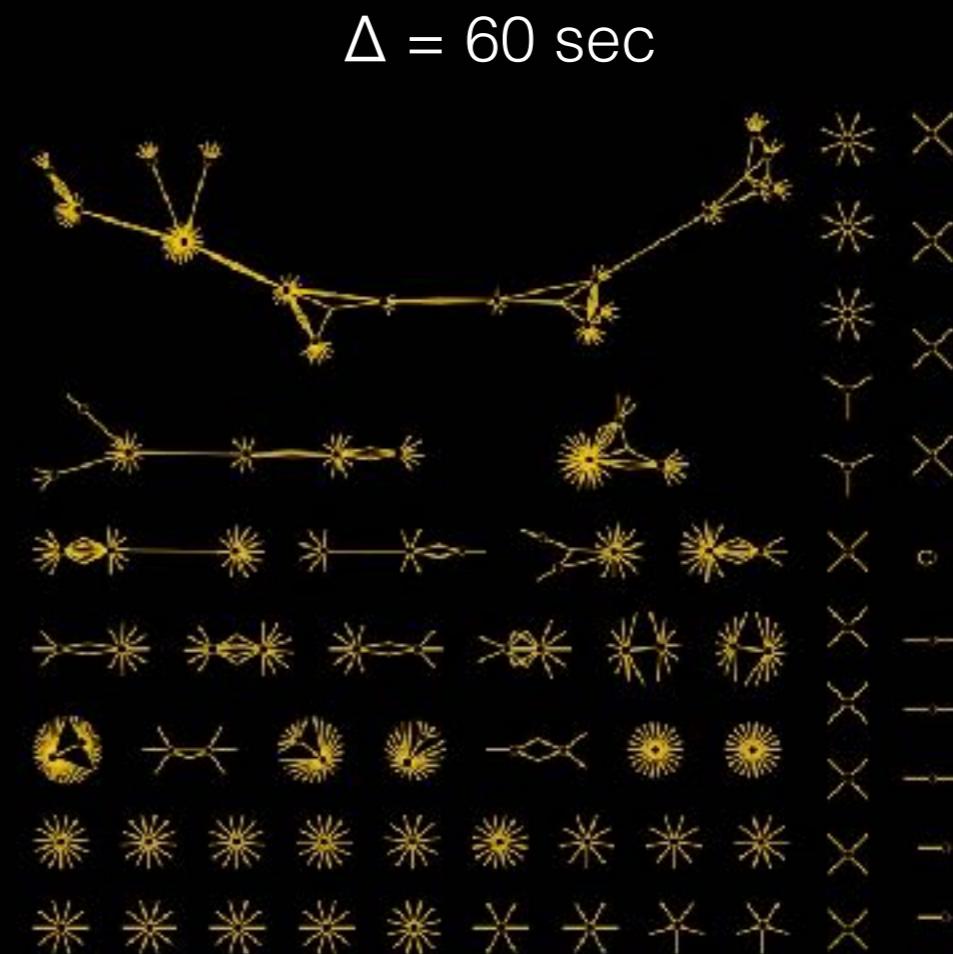
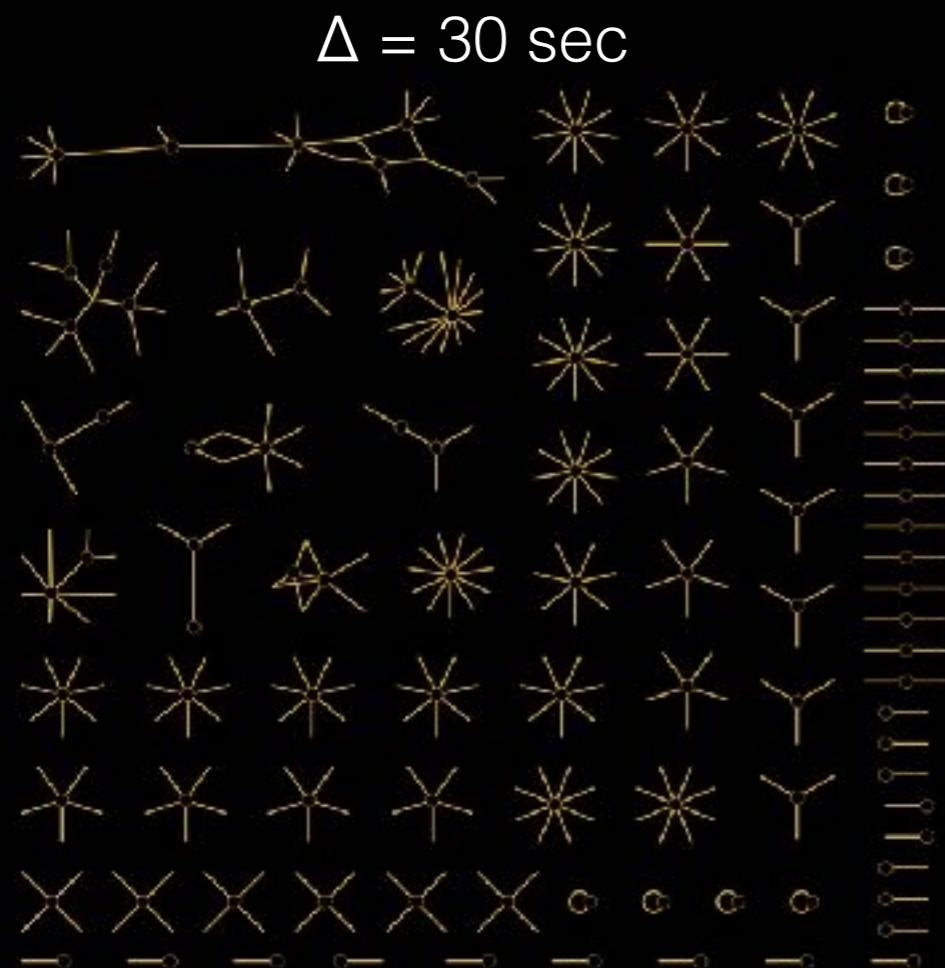
Maximum matching

# of passengers  
k = 2



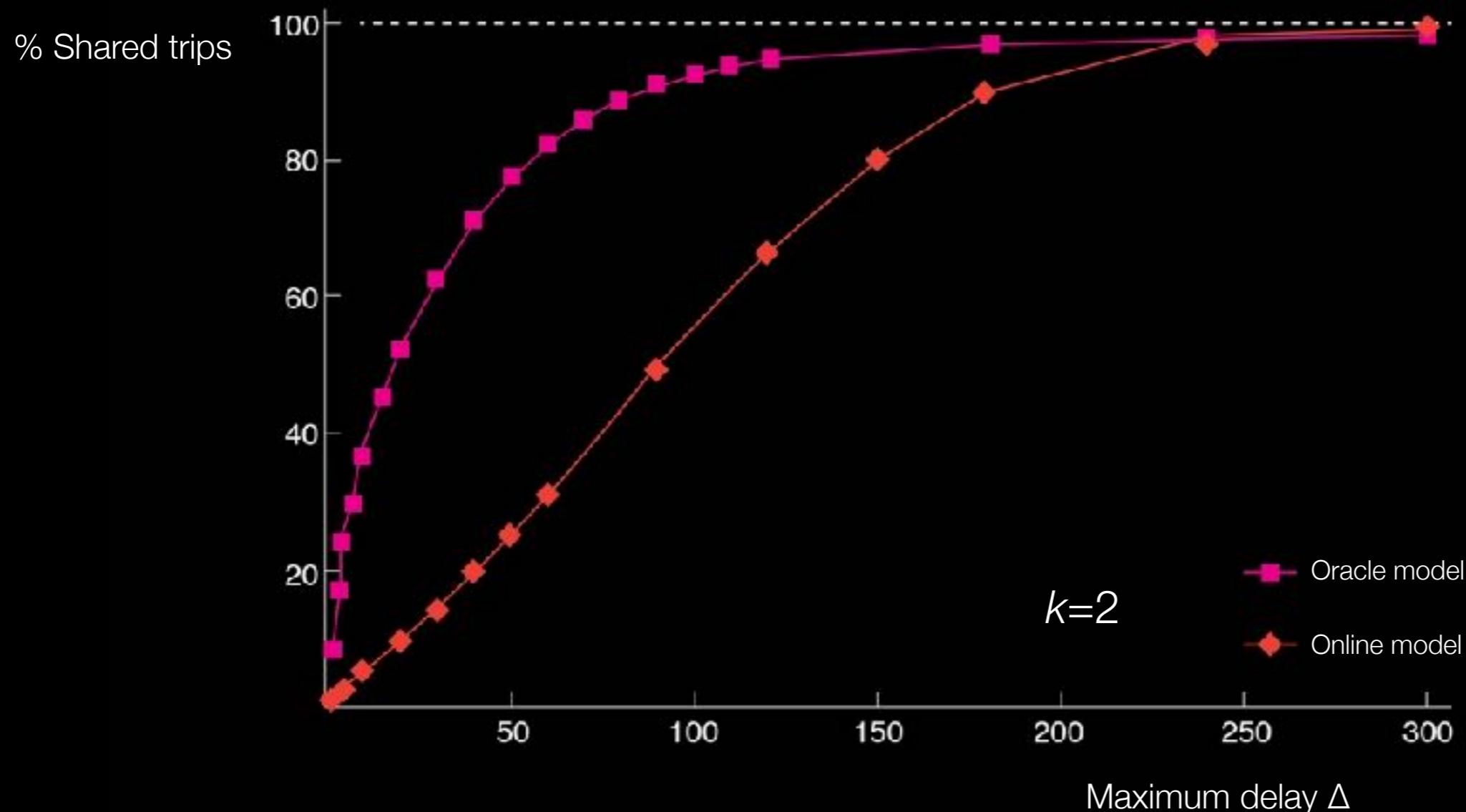
# Shareability network densification

Maximum time delay  $\Delta$



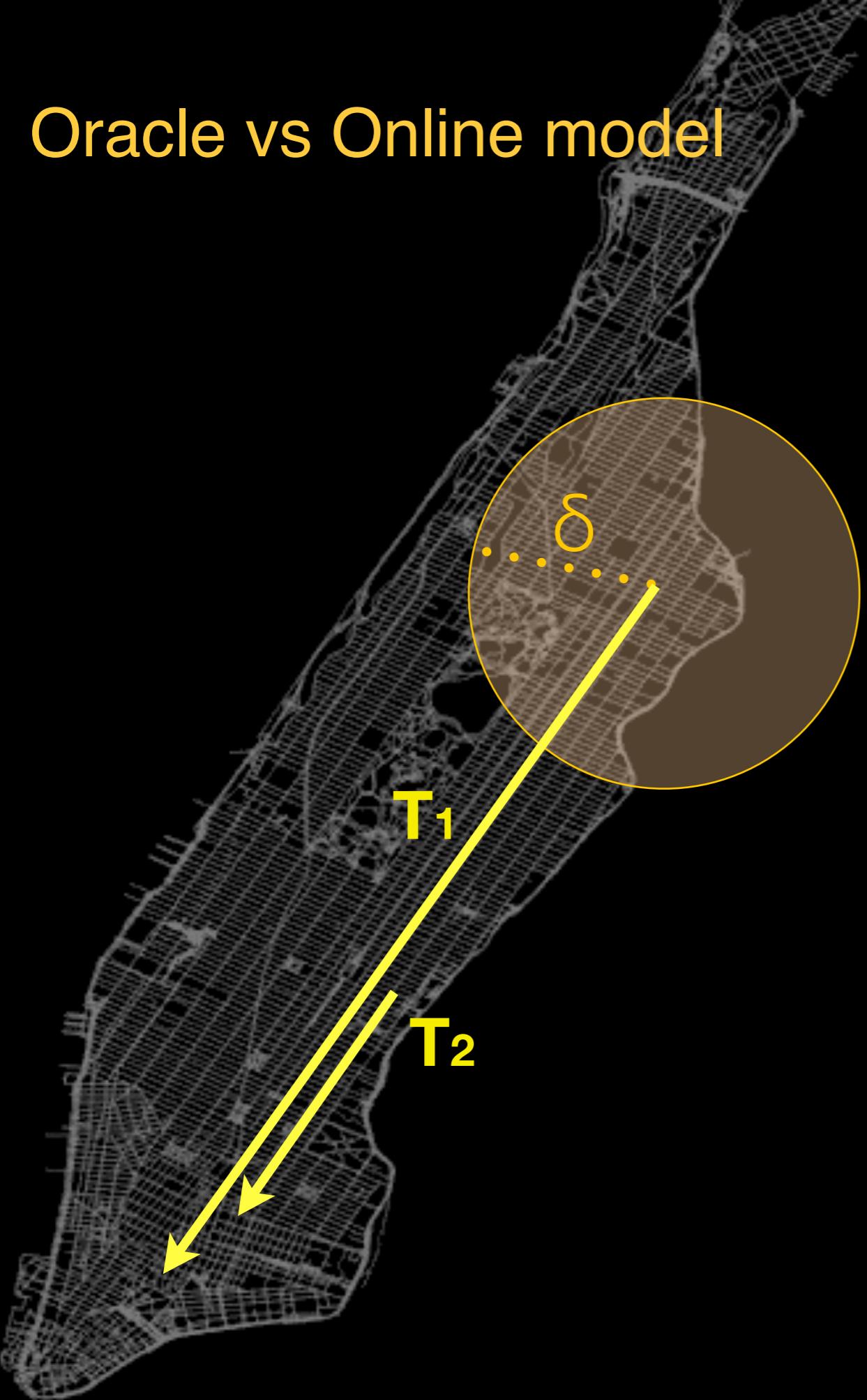
More tolerance = denser network = more sharing opportunities

# The majority of trips is shareable!



The majority of trips is sharable with minimal passenger inconvenience

# Oracle vs Online model

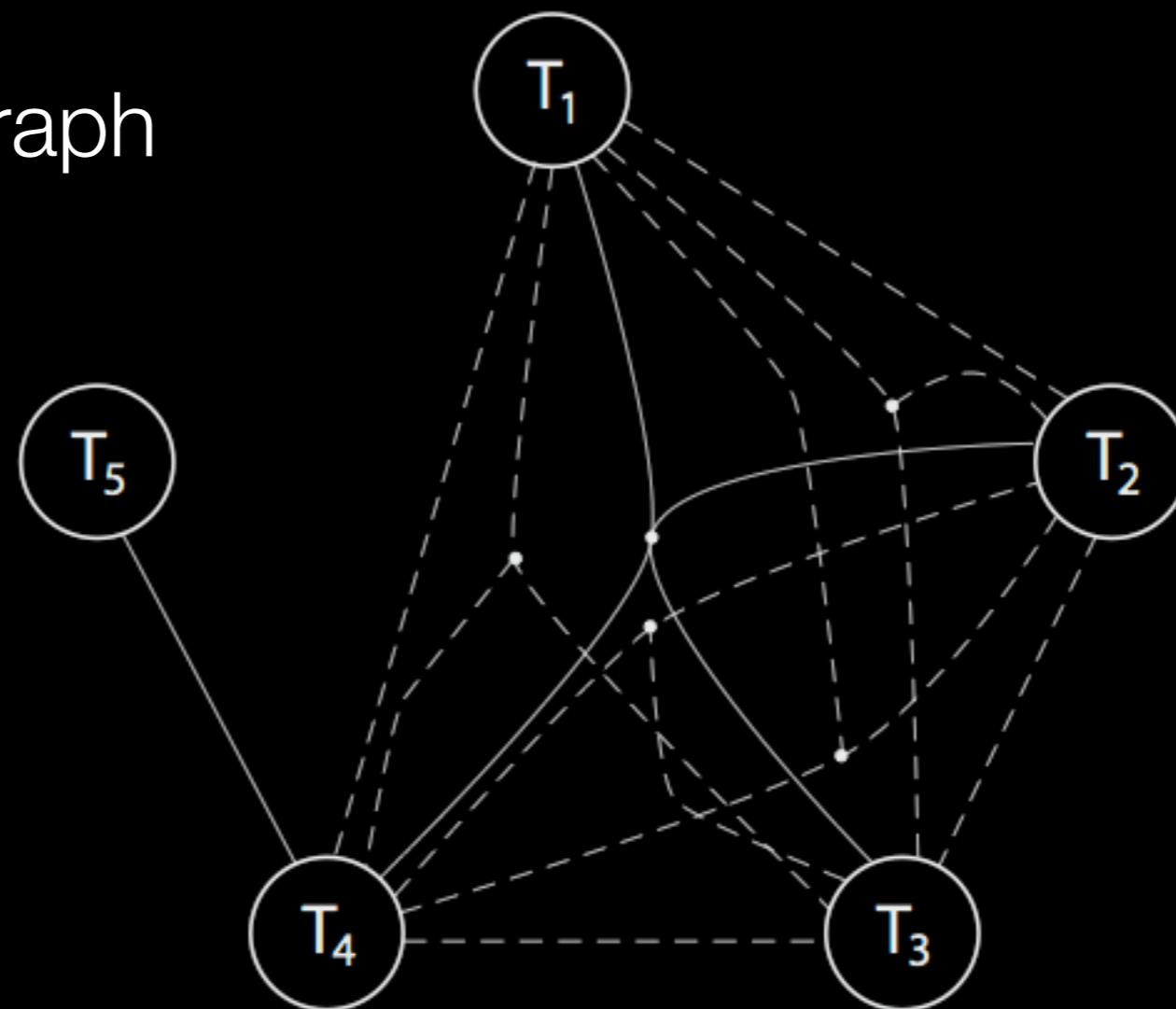


Oracle: omniscient

Online: realistic,  
constrained by time  
window  $\delta = 1 \text{ min}$

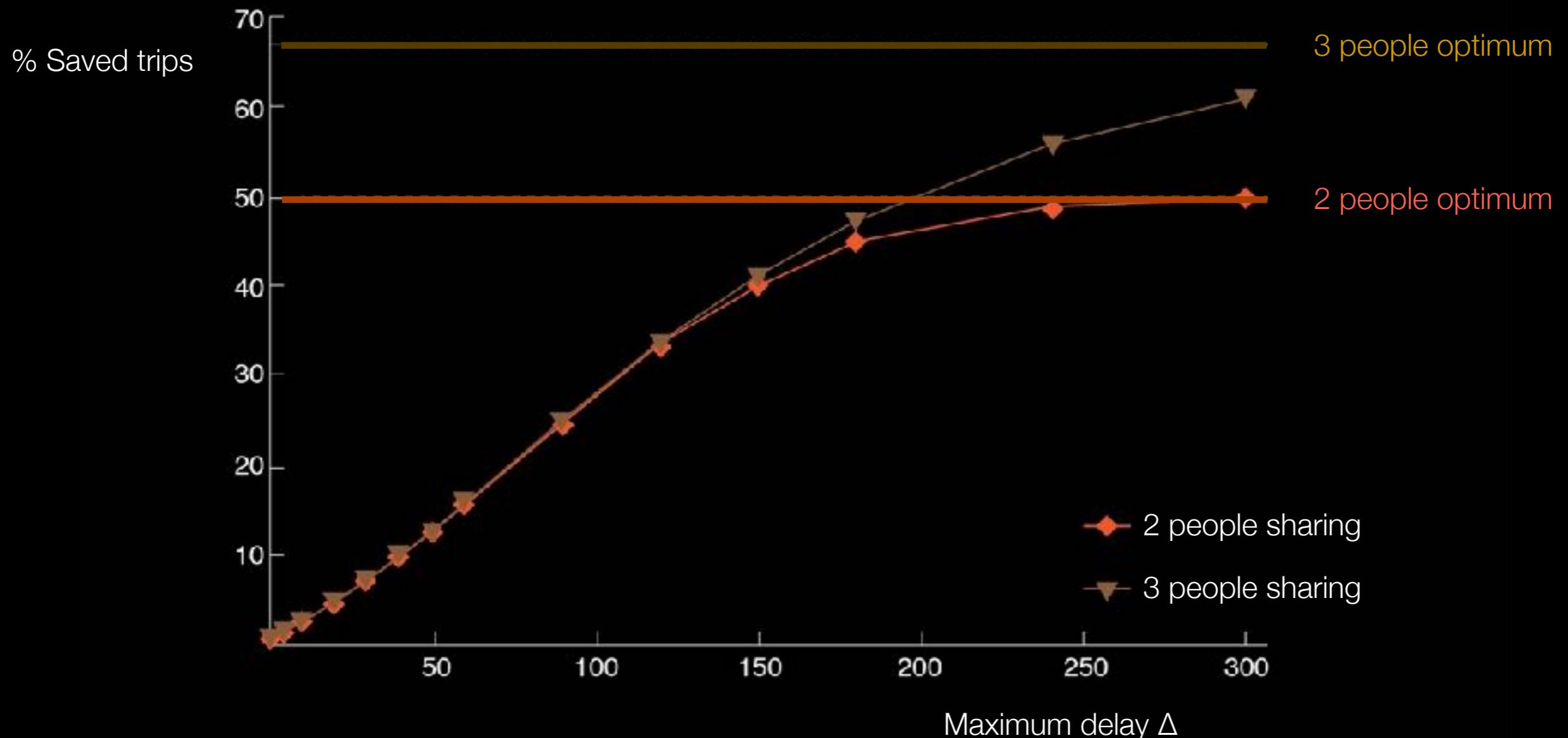
Approach can be extended to higher dimensions

Trip hypergraph



Approximation in  $O(n \log n)$

## From 2 to 3 people sharing



In NYC, benefits of 3 people sharing  
increase only slightly

## Consequences of trip-sharing

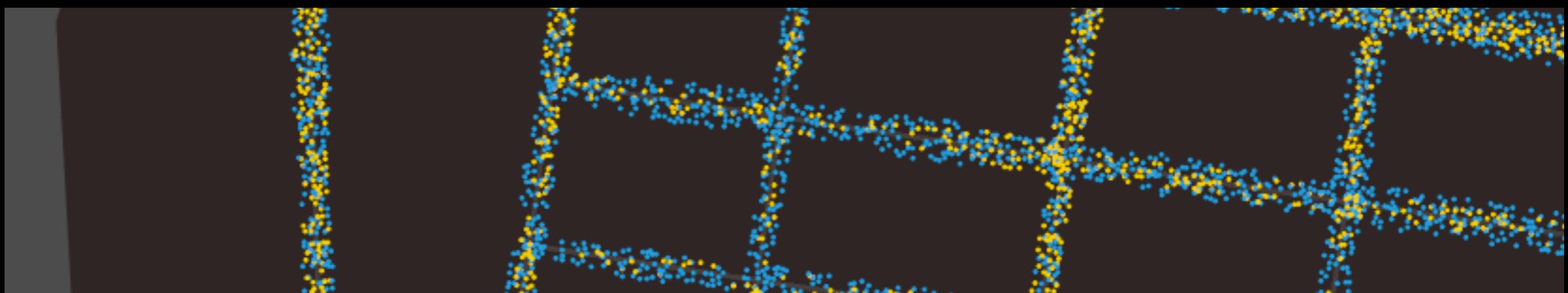
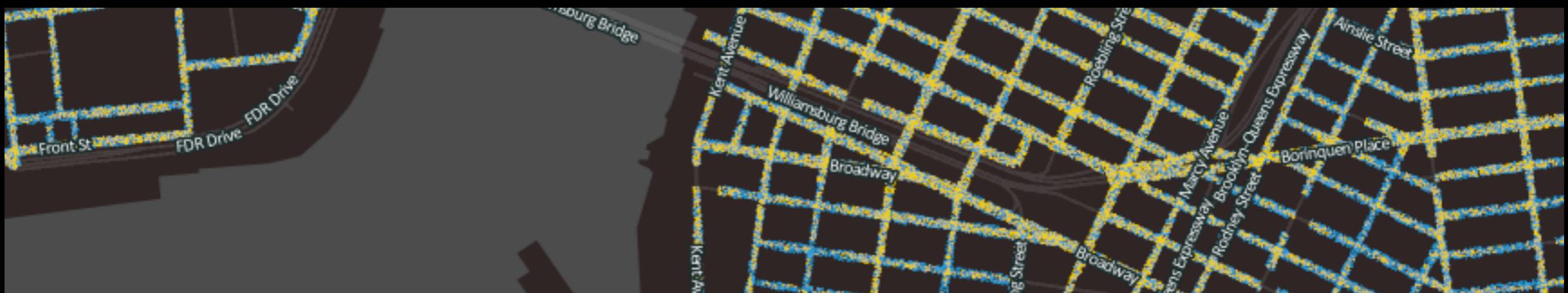
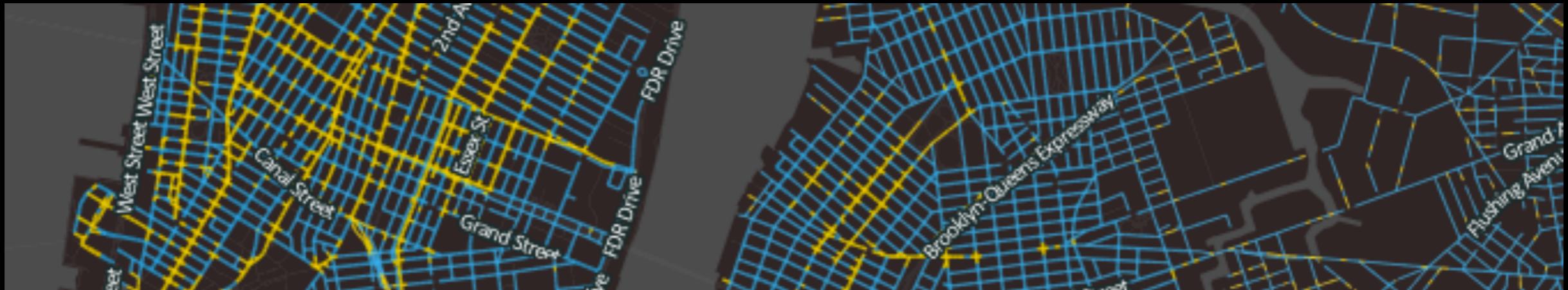
- More efficient use of vehicles: 40% decrease
- Decreased pollution and deaths
- Blurring line between private and public transportation
- New design of fare system

# Online tool hubcab for interactive exploration of sharing benefits



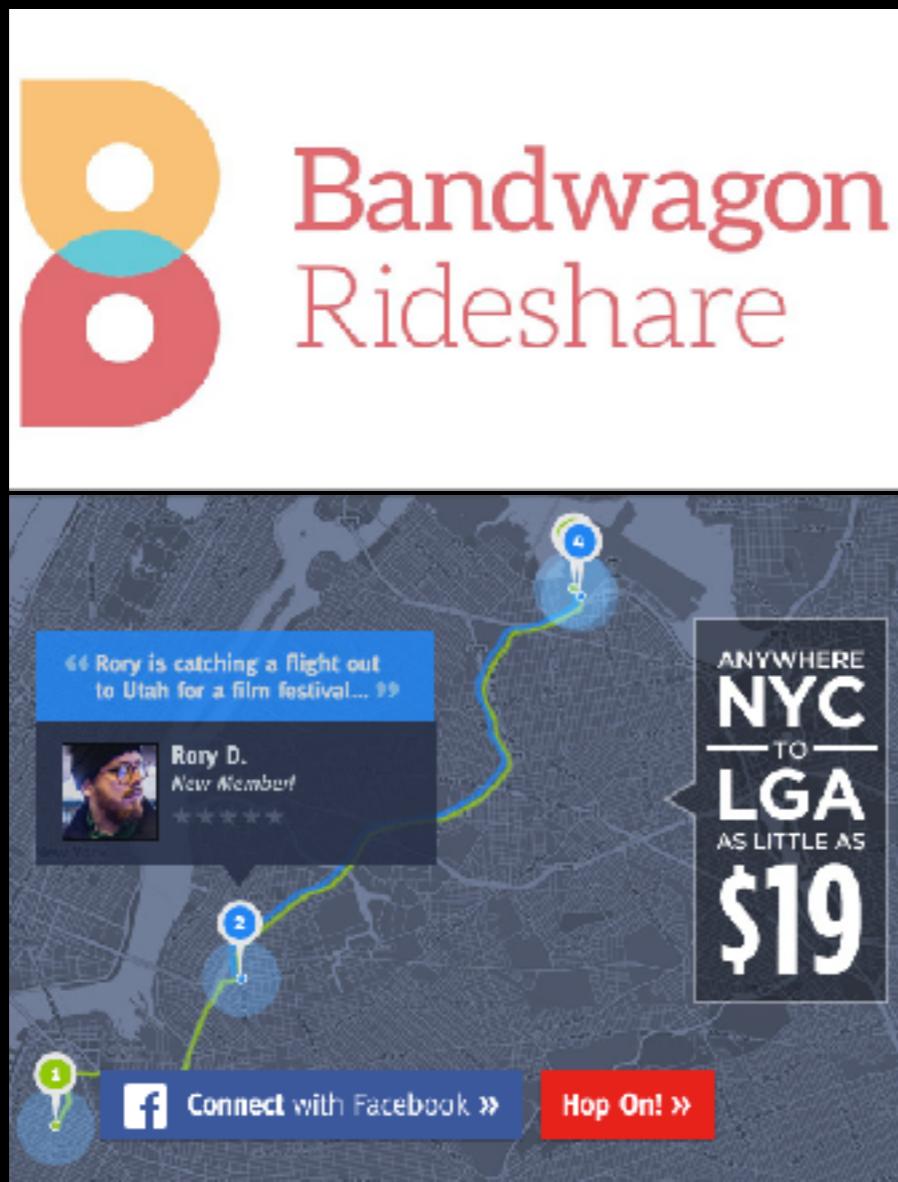
<http://www.hubcab.org>

# Zoom into the data, flow exploration, time selection

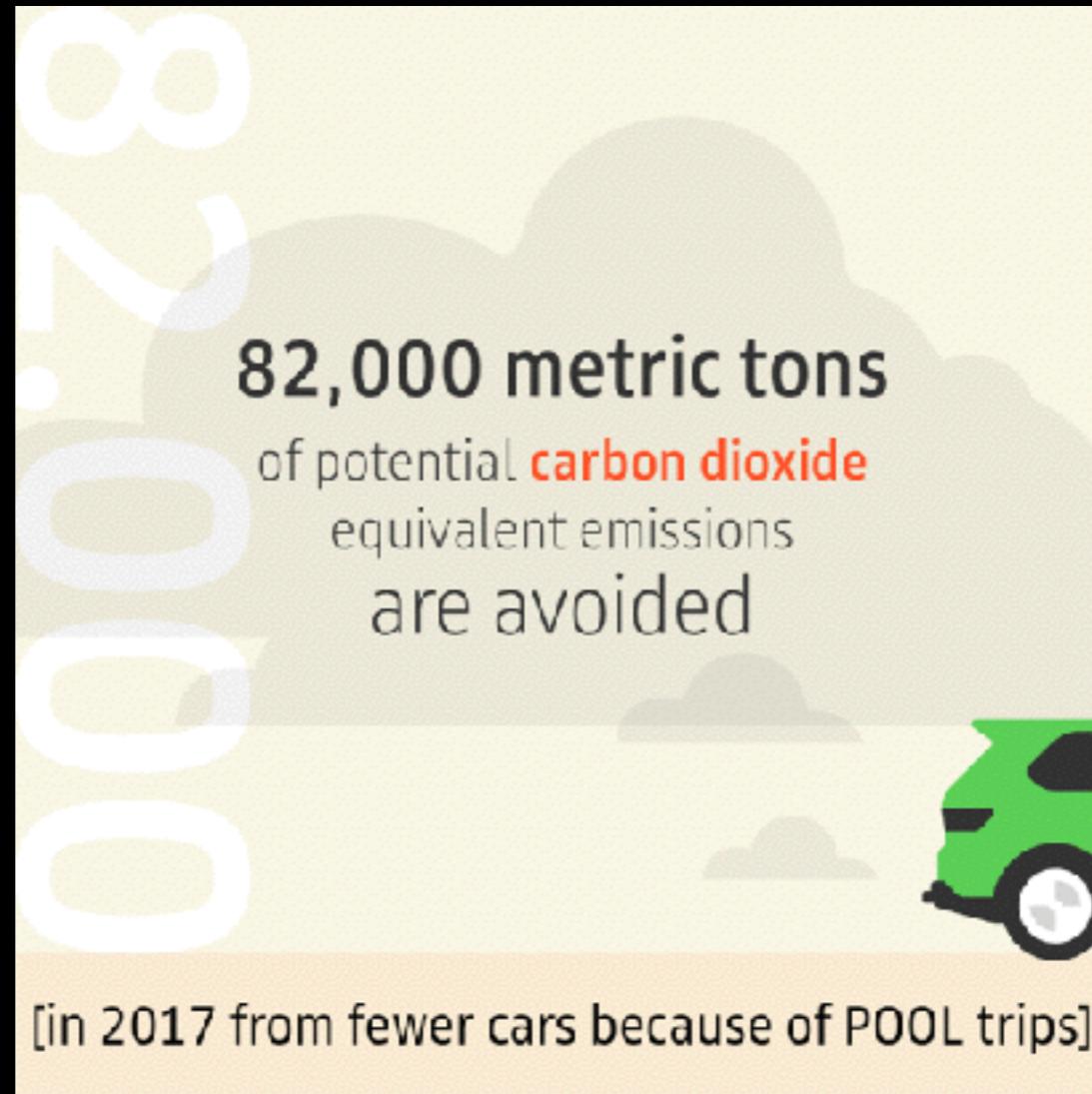


  Pickups        Dropoffs

# Trip-sharing is now implemented!



# Trip-sharing is now implemented!



Source: <https://www.uber.com/blog/earth-day-2018/>

# Beware of unintended consequences!

Cars become more attractive



Cities invest less in public infrastructure



MORE cars on the road = MORE problems

Sustainable implementation of sharing  
requires understanding systemic effects

## 3.

# SCIENTIFIC REPORTS

www.nature.com/scientificreports/

OPEN

## Scaling Law of Urban Ride Sharing

R. Tachet<sup>1</sup>, O. Sagarra<sup>1,2,3</sup>, P. Santi<sup>4</sup>, G. Resta<sup>5</sup>, M. Szell<sup>1,5</sup>, S. H. Strogatz<sup>6</sup> & C. Ratti<sup>1</sup>

Sharing rides could drastically improve the efficiency of urban taxi transportation. Unleashing such potential, however, requires understanding how urban parameters affect the fraction of individual trips that can be shared, a quantity that we call shareability. Using data on millions of taxi trips in New York City, San Francisco, Singapore, and Vienna, we compute the shareability curves for each city, and find that a natural rescaling collapses them onto a single, universal curve. We explain this scaling law theoretically with a simple model that predicts the potential for ride sharing in any city, using a few basic urban quantities and no adjustable parameters. Accurate extrapolations of this type will help planners, transportation companies, and society at large to shape a sustainable path for urban growth.

Received: 31 October 2016

Accepted: 18 January 2017

Published: 06 March 2017

Mobility of people and goods has been vital to urban life since cities emerged more than 7,000 years ago<sup>1</sup>. Indeed, the success, prosperity, and viability of cities are directly related to the effectiveness of their mobility systems<sup>2</sup>. However, due to fixed schedules, limited coverage, and low quality of travel experience, public transportation systems accommodate only a fraction of the urban mobility demand<sup>3</sup>. The rest is satisfied by private vehicles and taxis, inefficient transportation modes that move only 1.3 passengers per vehicle on average<sup>4,5</sup>, causing the road congestion observed in most cities worldwide, with immense economic and societal costs<sup>6</sup>. Enhancing transportation efficiency is a key to rendering sustainable the urban growth predicted for the coming years<sup>7</sup>.

The emerging sharing economy<sup>8–10</sup> promises to improve the efficiency of individual, on-demand transportation. Bridging the gap between shared but inflexible public transportation and flexible but not shared private transportation, novel services such as those provided by Uber<sup>TM</sup>, Lyft<sup>TM</sup>, and ZipCar<sup>TM</sup> can significantly contribute to reducing road congestion and emissions. But the realizability of such potential benefits depends on the answer to a fundamental methodological question: How compatible – in space and time – are individual mobility patterns?

While recent literature<sup>11–13</sup> has unveiled spatial and temporal regularity of individual mobility patterns, very little is known about their mutual similarity. In a previous study<sup>14</sup>, we introduced the notion of a shareability network to quantify the spatial and temporal compatibility of individual trips. The nodes in the network represent trips, and links between them mark trips that can be shared. Two trips are defined to be shareable if they would incur a sharing delay of no more than 5 minutes, relative to a single ride (see *Supplementary Information*). Let the shareability metric  $\delta$  denote the fraction of individual trips that can be shared. We found<sup>14</sup> that taxi trips in New York City offer a shareability well above 95% for  $\delta = 5\text{min}$ , and that  $S$  increases rapidly with the number of trips available for sharing.

But that previous study<sup>14</sup> left a key question unanswered: Might the results be peculiar to New York City? There was good reason to suspect so given that New York is singular in several respects, namely, its large population, its small geographical area, and its enormous density of taxi traffic. In what follows, we study ride shareability in three other major world cities — San Francisco, Singapore, and Vienna — for which extensive data is available. Although these cities differ greatly from each other and from New York City in their traffic characteristics, population size, and geographical area, we find they all obey the same empirical law governing the potential for ride sharing. To the best of our knowledge, the existence of such a seemingly universal law has not been reported before. We explain the mechanism underlying this law of ride sharing using a simple mathematical model. The model's prediction accounts for more than 90% of the variance in the data, and does so without any adjustable parameters. What is important here is the generality of the law, as well as its rapidly saturating shape, because together they imply that ride sharing could have a large beneficial impact in virtually any city, not just New York City.

### Results

Let  $C$  be a city,  $\Omega(C)$  its spatial domain,  $|\Omega(C)|$  its area,  $v(C)$  the average traffic speed in  $C$  and  $\lambda$  the average number of trips per hour with both endpoints in  $\Omega(C)$ . Figure 1(a) shows that the computed curve of shareability

<sup>1</sup>Senseable City Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>2</sup>Complexity Lab, Barcelona, Universitat de Barcelona, 08028 Barcelona, SPAIN. <sup>3</sup>ORIBIA Data Research, 08012 Barcelona, SPAIN.

<sup>4</sup>Instituto di Informatica e Telematica del CNR, 56124 Pisa, ITALY. <sup>5</sup>Hungarian Academy of Sciences, Centre for Social Sciences Országos Széchényi 30, 1014 Budapest, HUNGARY. <sup>6</sup>Department of Mathematics, Cornell University, Ithaca, NY 14853, USA. Correspondence and requests for materials should be addressed to R.T. (email: rtachet@mit.edu).

## New York is special - What about other cities?

NYC 13,500 cabs



San Fran 1,500 cabs



Singapore 26,000 cabs



Vienna 5,000 cabs



# New York is special - What about other cities?

NYC 13,500 cabs



25,846/km<sup>2</sup>

San Fran 1,500 cabs



6,659/km<sup>2</sup>

Singapore 26,000 cabs



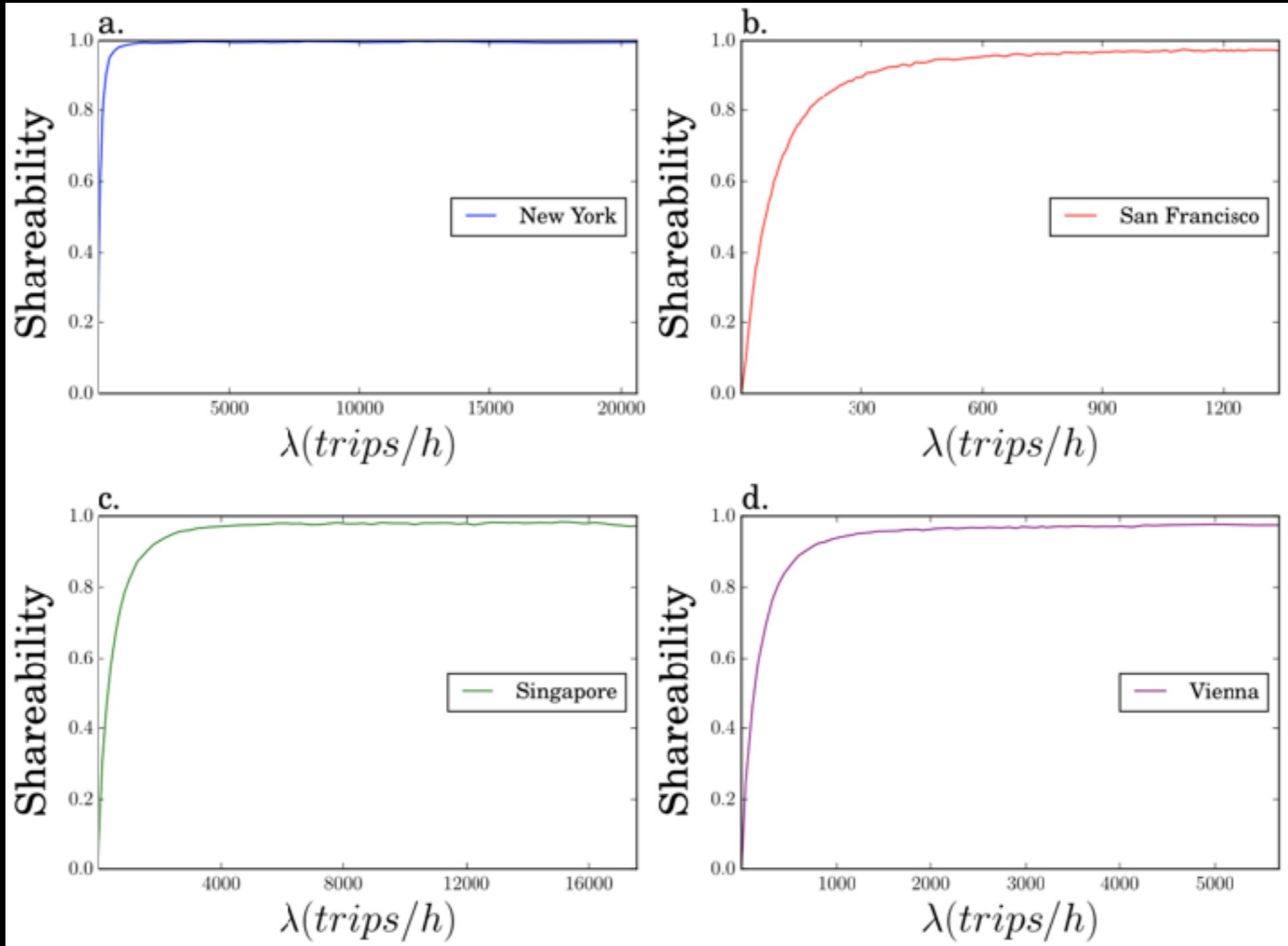
7,988/km<sup>2</sup>

Vienna 5,000 cabs

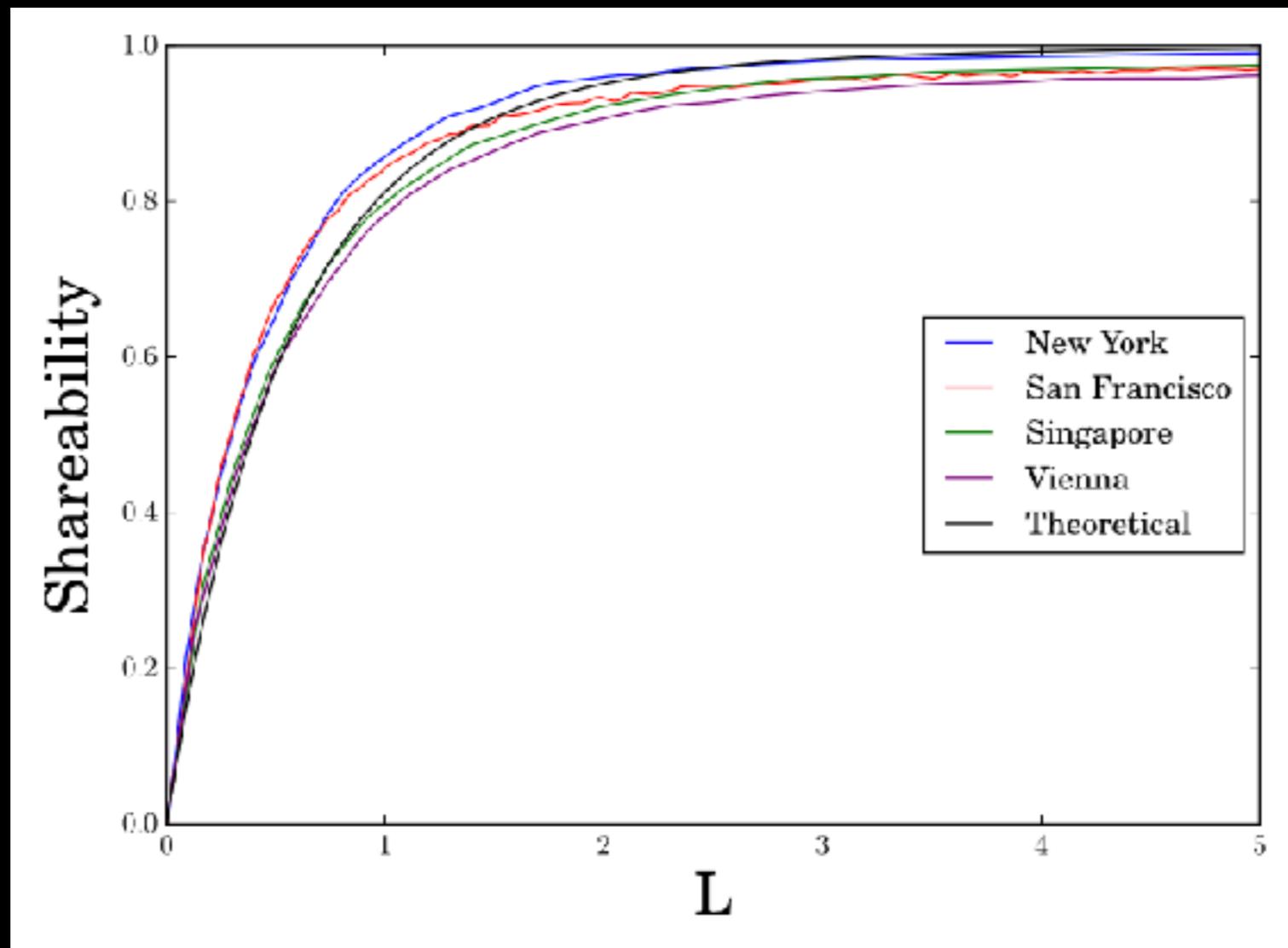


4,002/km<sup>2</sup>

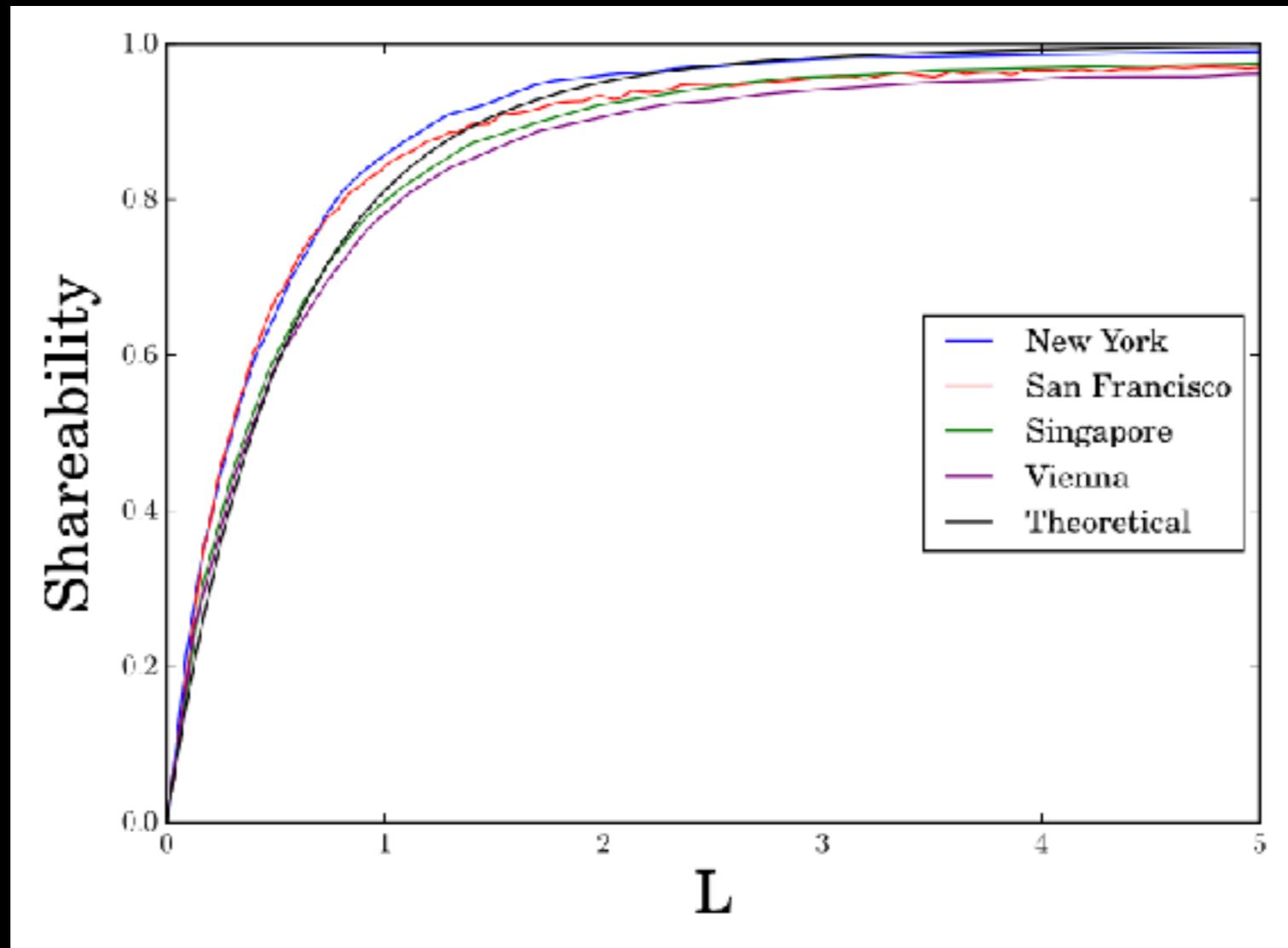
# Different cities have similar shareability curves



# All curves collapse onto a universal curve!



# All curves collapse onto a universal curve!



What is L?

# Rescaling through L

$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$

# Rescaling through L

$$L = \lambda \Delta^3 \overline{v^2}$$

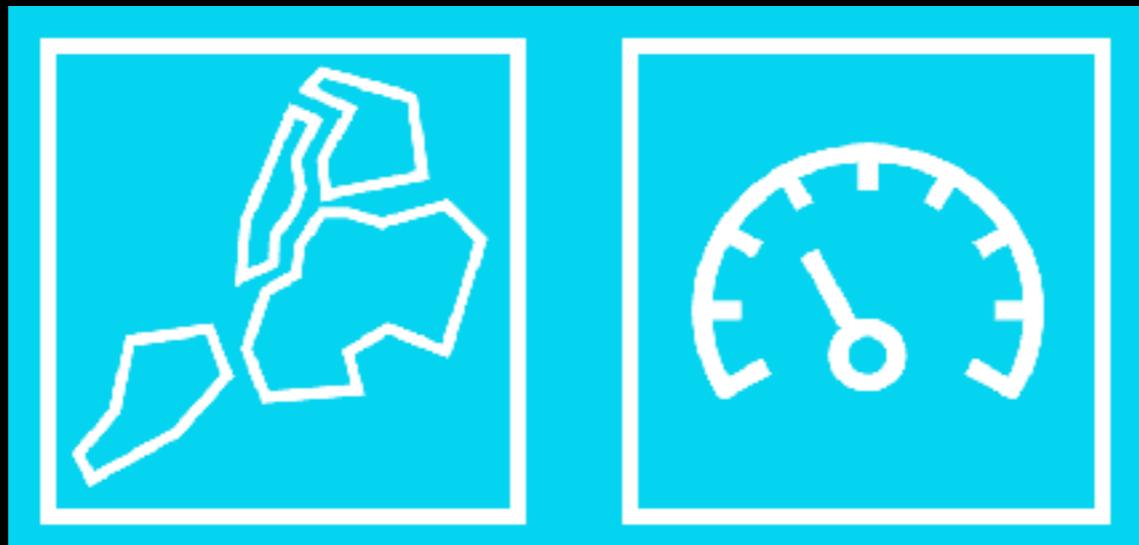
City area



## Rescaling through L

$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$

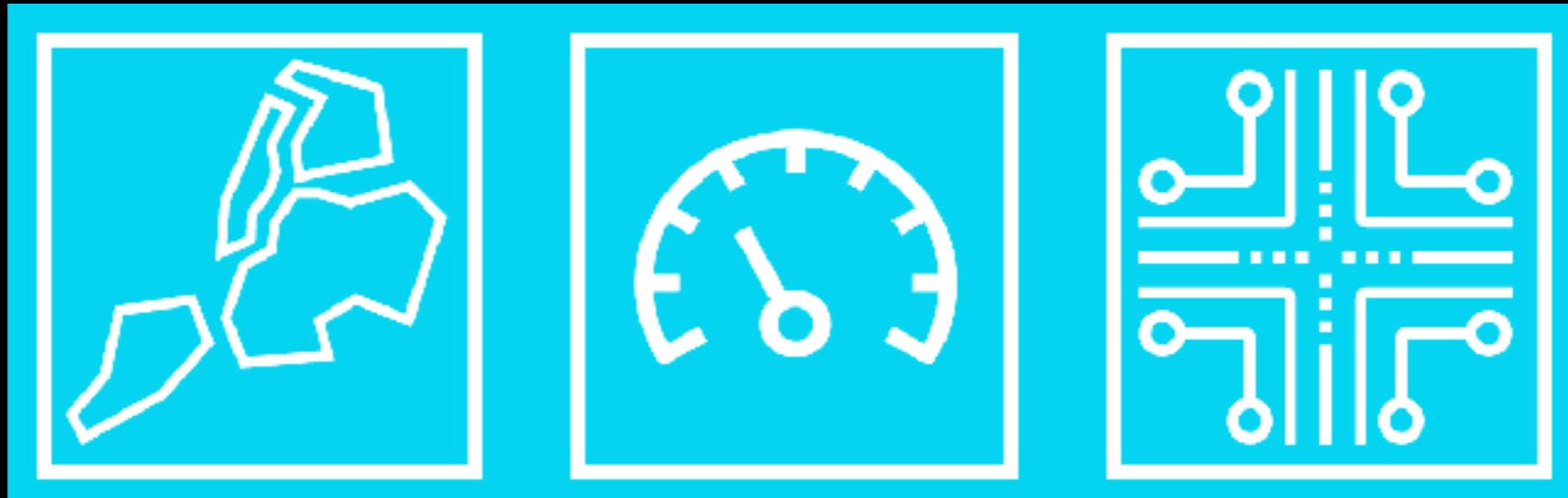
Avg speed



## Rescaling through L

$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$

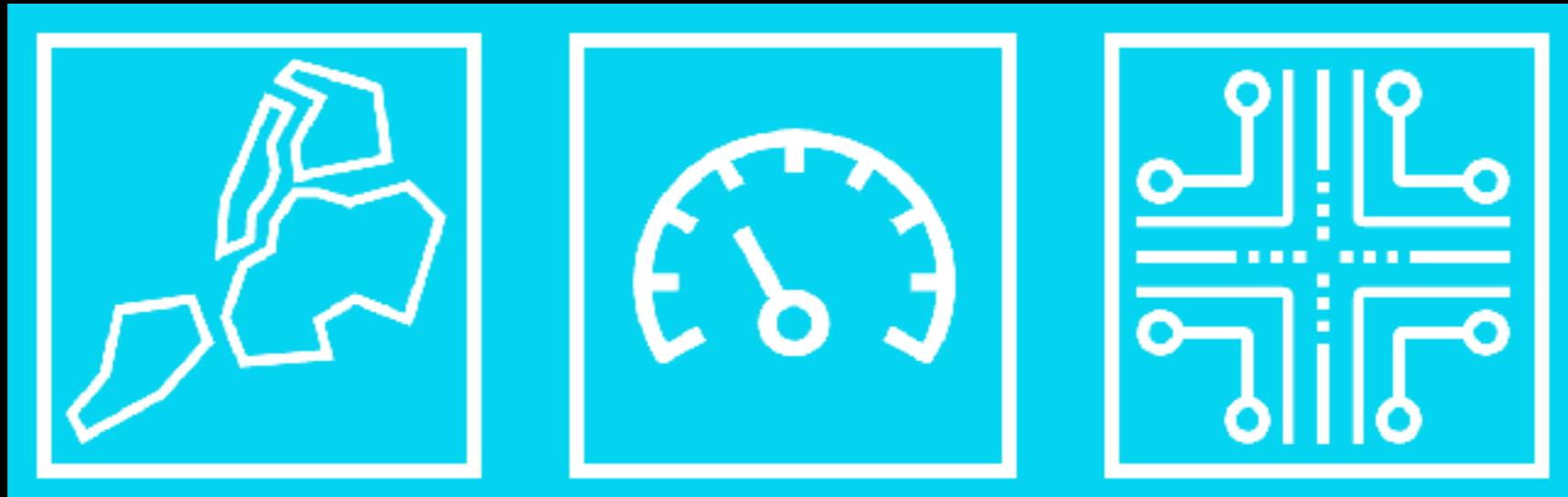
**Trip density**



# Rescaling through L

$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$

**City area      Avg speed      Trip density**



L is dimensionless

$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$



L is a ratio between two timescales

$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$

Tolerable delay time

$\Delta$

vs

Expected waiting time

$t_{\text{wait}}$

L is a ratio between two timescales

$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$

Tolerable delay time

$$\Delta$$

vs

Expected waiting time

$$t_{\text{wait}} = \frac{1}{\lambda} \times \frac{|\Omega|}{(v\Delta)^2}$$

L is a ratio between two timescales

$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$

Tolerable delay time

vs

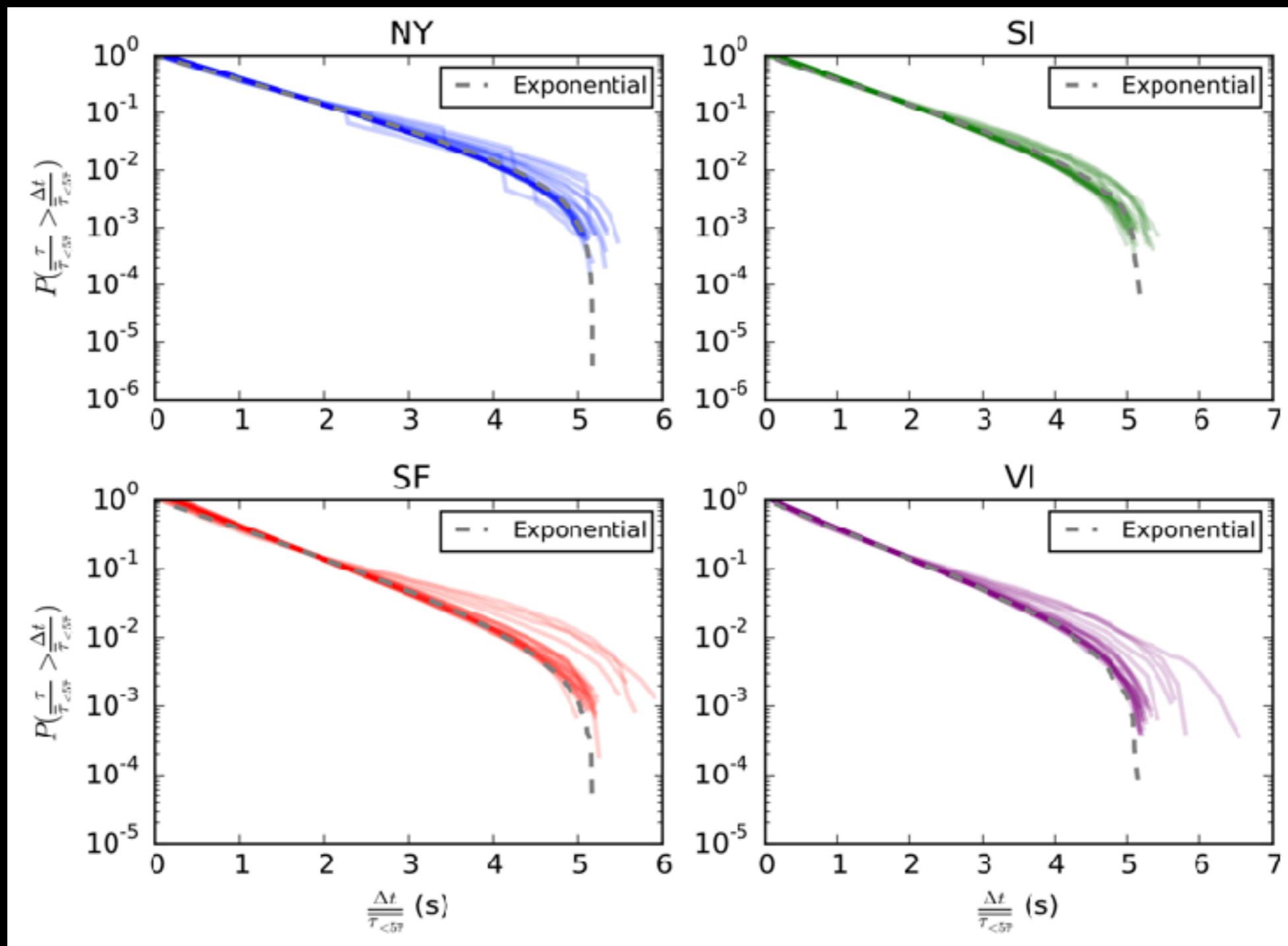
Expected waiting time

$\Delta$

**Characteristic  
time for new trip**

$$t_{\text{wait}} = \frac{1}{\lambda} \times \frac{|\Omega|}{(v\Delta)^2}$$

# Trip generation is poissonian



L is a ratio between two timescales

$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$

Tolerable delay time

vs

Expected waiting time

$\Delta$

**Characteristic  
scale of a vicinity**

$$t_{\text{wait}} = \frac{1}{\lambda} \times \frac{|\Omega|}{v \Delta^2}$$

Sure, cities are different...

...but ride-sharing works well almost everywhere

<http://senseable.mit.edu/shareable-cities/>

## Research

Paolo Santi  
Giovanni Resta  
Remi Tachet  
Oleguer Sagarra  
S. Sobolevsky  
Carlo Ratti  
Steven Strogatz

## Visualizations

Benedikt Groß  
Joey Lee  
Eric Baczkuk  
Carlo Ratti  
Andi Weiß  
Stefan Landsbeck  
Pierrick Thebault

Michael Szell

@mszell  
[michael.szell@gmail.com](mailto:michael.szell@gmail.com)  
<http://michael.szell.net>