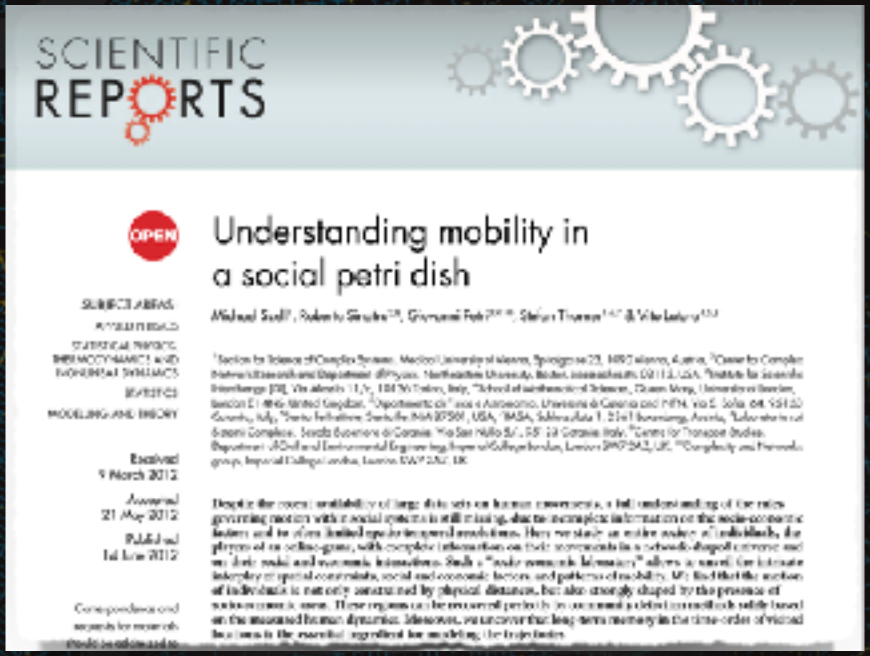
Mathematical modelling seminar

Mathematical modelling of human mobility

BME, Budapest, Sep 18th 2018

Michael Szell
@mszll

1.



SCIENTIFIC REPORTS

OPEN

Understanding mobility in a social petri dish

2.

Hungarian Acad Sci I Northeastern University I Central European University

1.

SCIENTIFIC REPORTS

OPEN

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

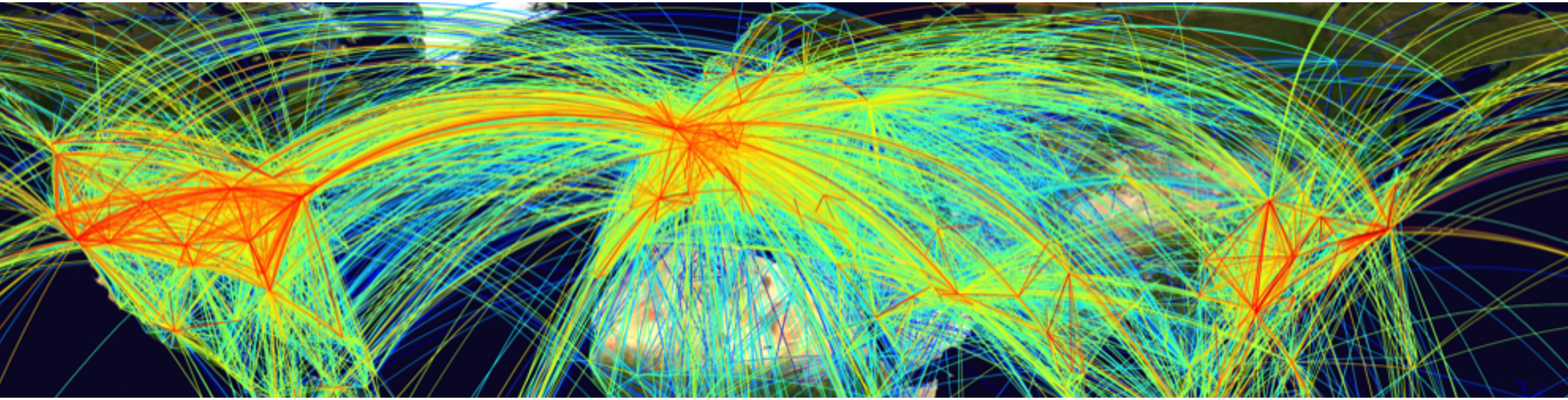
Correspondence and requests for materials should be addressed to

Michael Szell[1], Roberta Sinatra[2,8], Giovanni Petri[3,9,10], Stefan Thurner[1,6,7] & Vito Latora[4,5,8]

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

1) 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

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

- trading
- socializing
- conflicts

since 14 years

Complete data on human society!

Here: use to study mobility

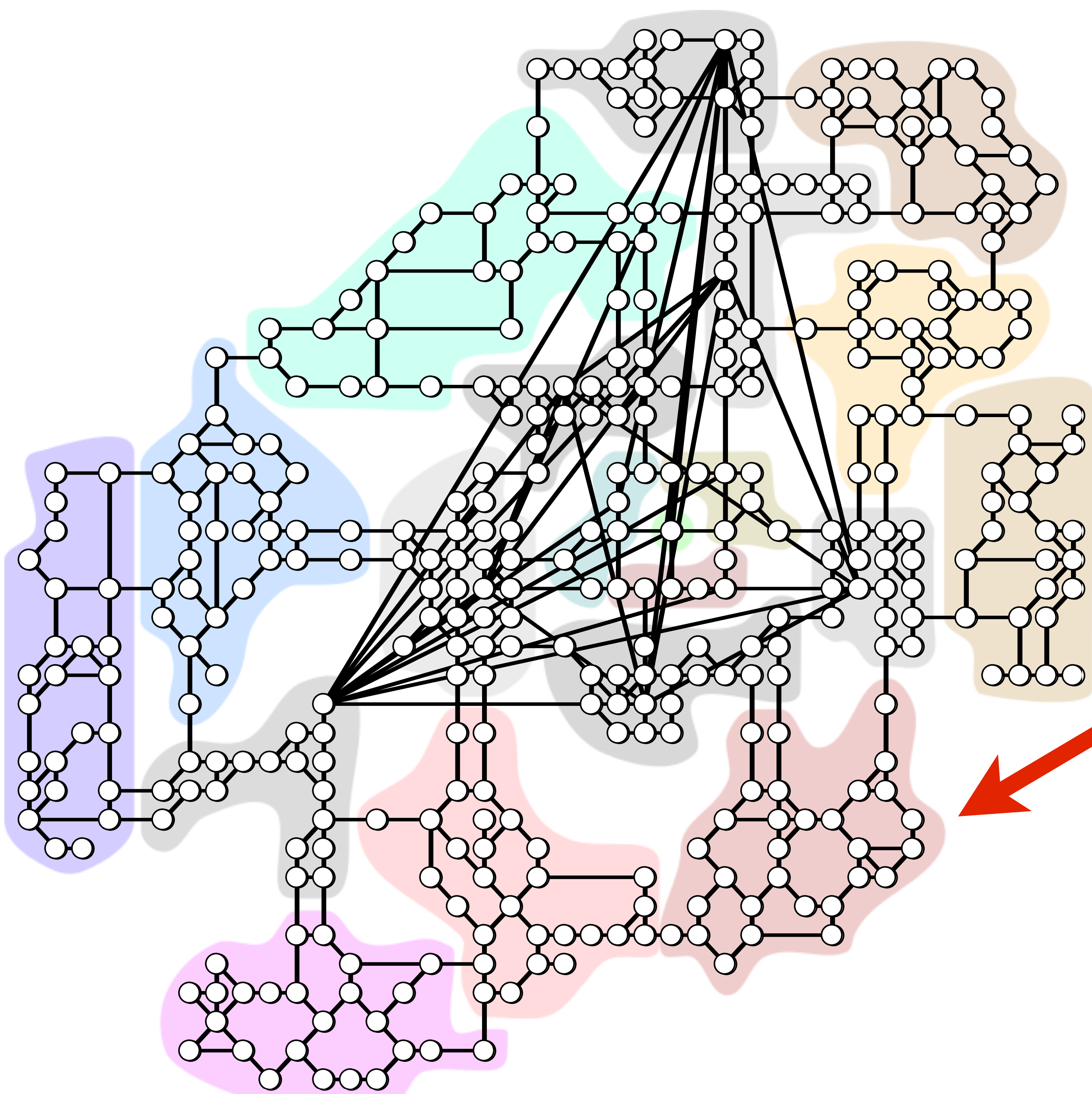
Bainbridge, Science 317, 472 (2007)
Szell and Thurner, Social Networks 32, 313-329 (2010)
Szell et al, PNAS 107, 1363-13641 (2010)

Establishing a socio-economic laboratory

www.pardus.at

The universe of the game

Network

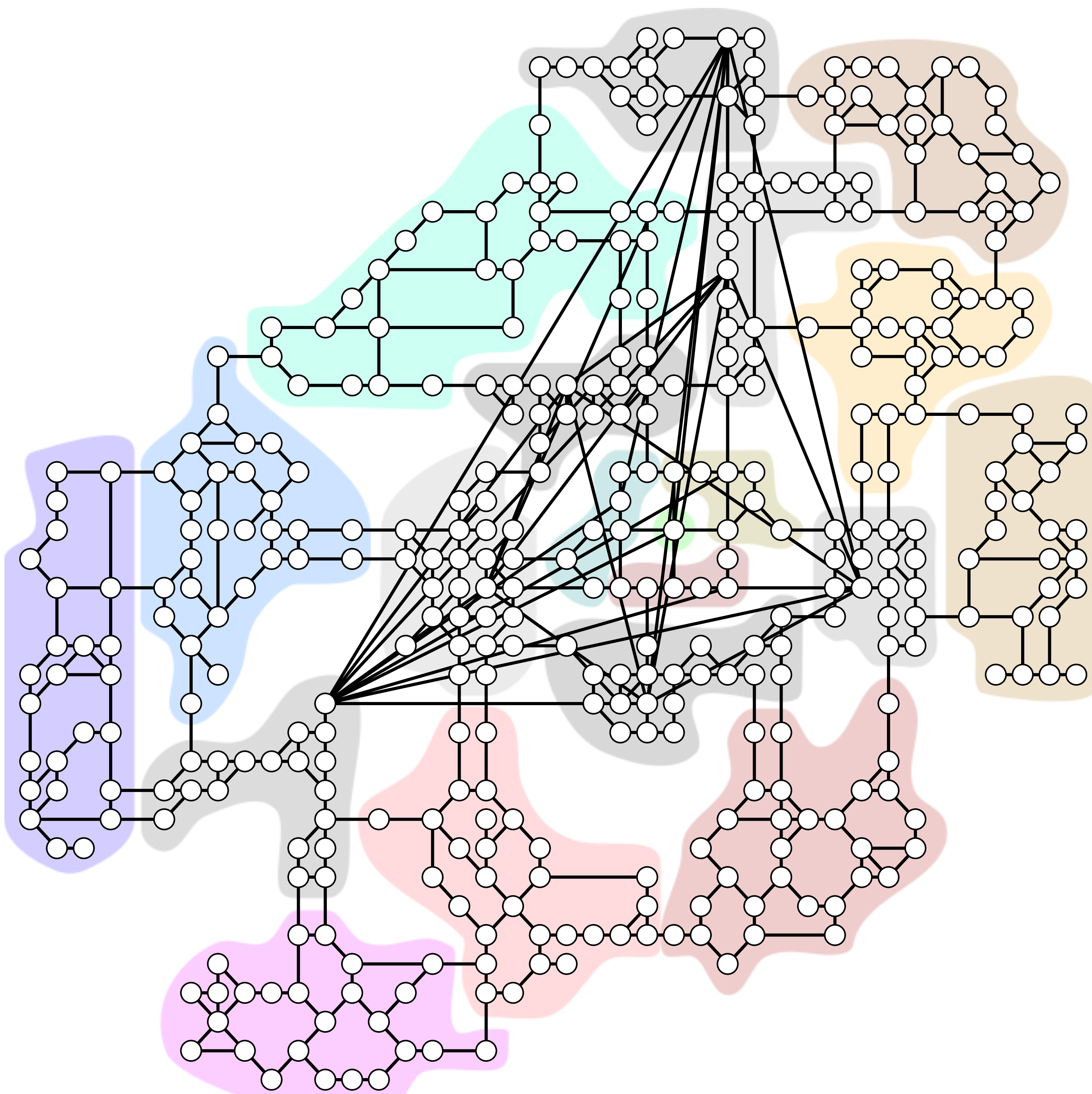


- 400 Nodes, "Cities"
- Diameter = 27
- Lattice-like

20 Clusters
"Countries"

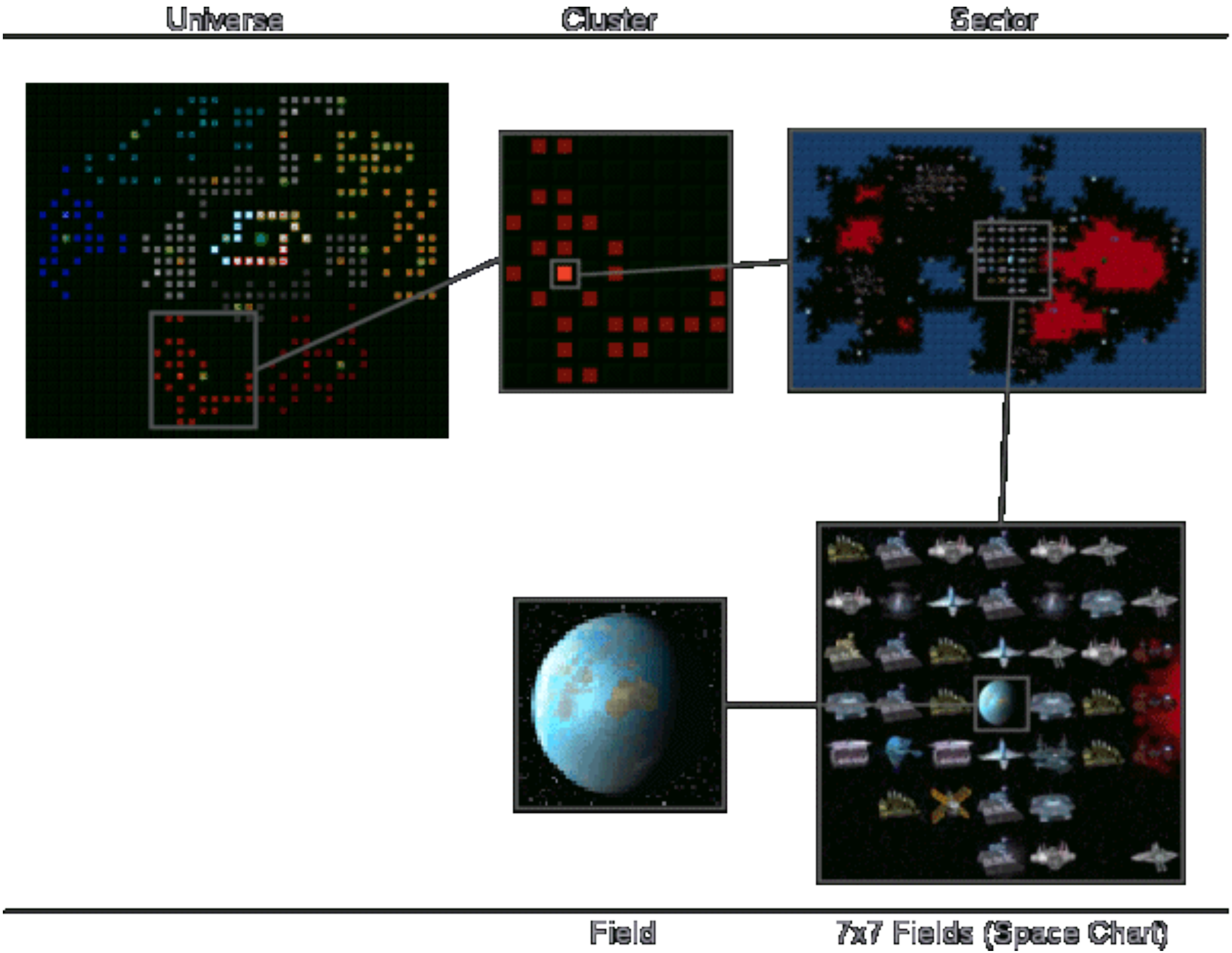
Mobility on a network

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



One mode of transportation

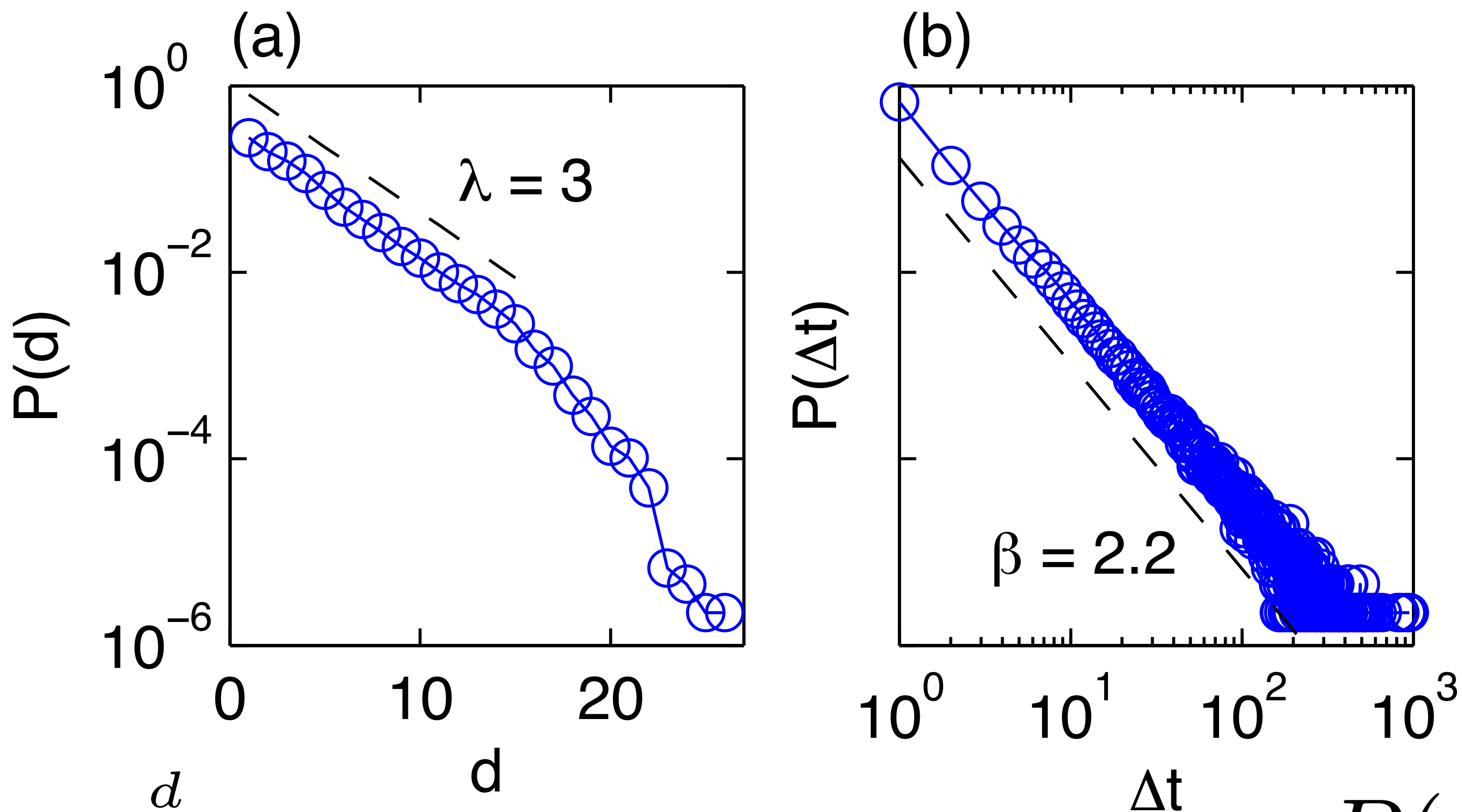
Move along links


Universe
Cluster
Sector
Field
7x7 Fields (Space Chart)

Basic features of motion

Trip lengths — exponential

Waiting times — power

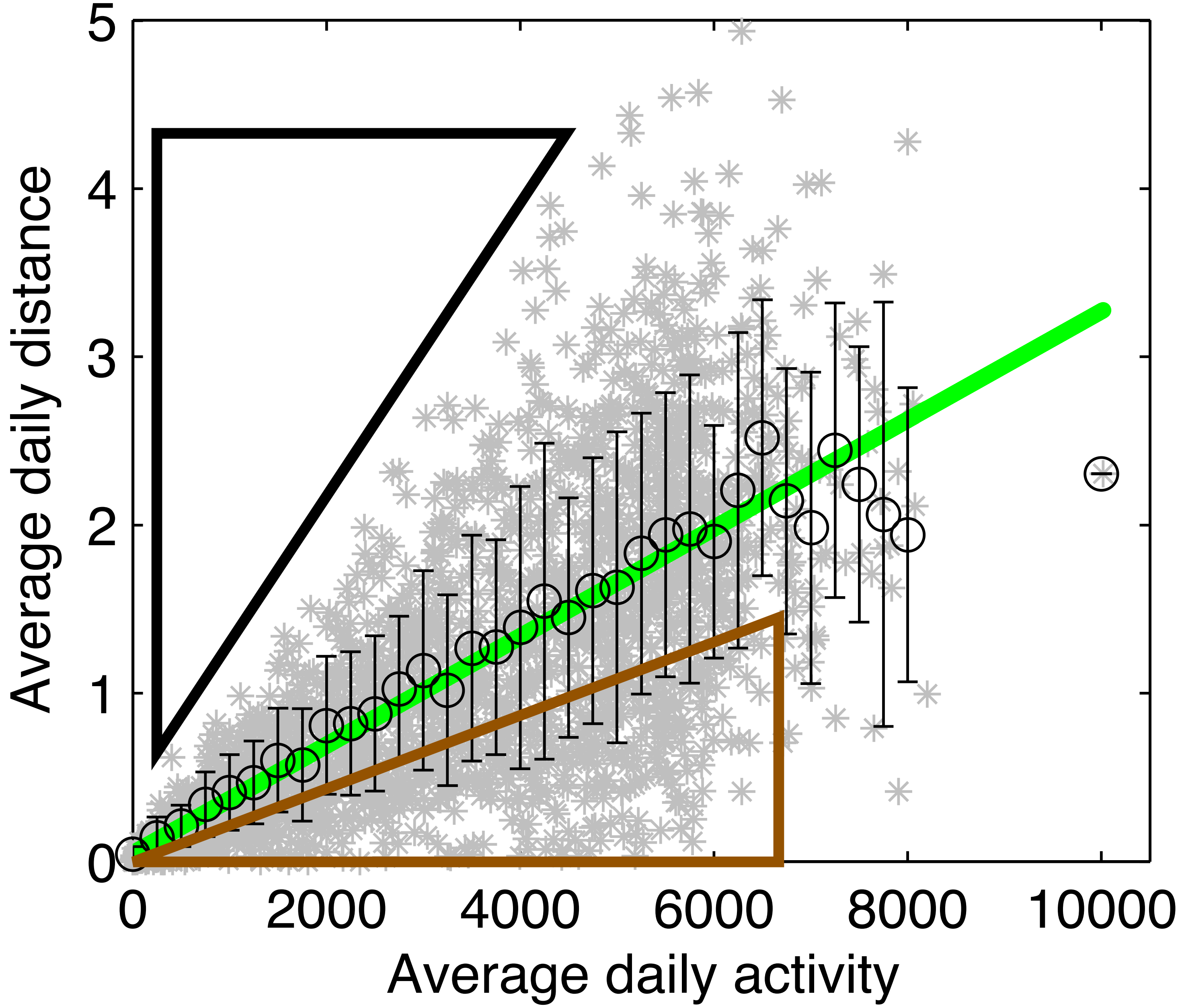


$$P(d) \sim e^{-\frac{d}{\lambda}}$$

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

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

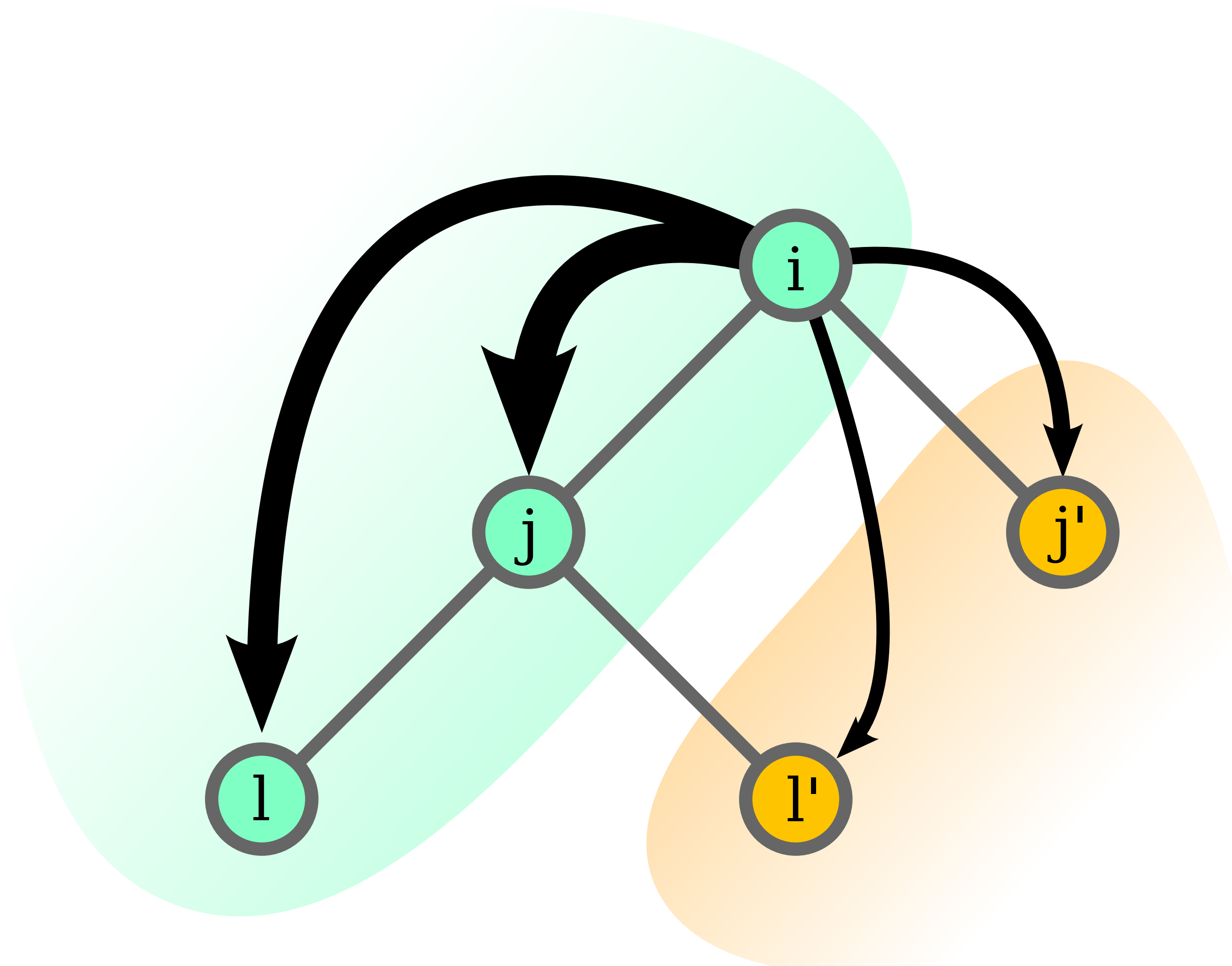


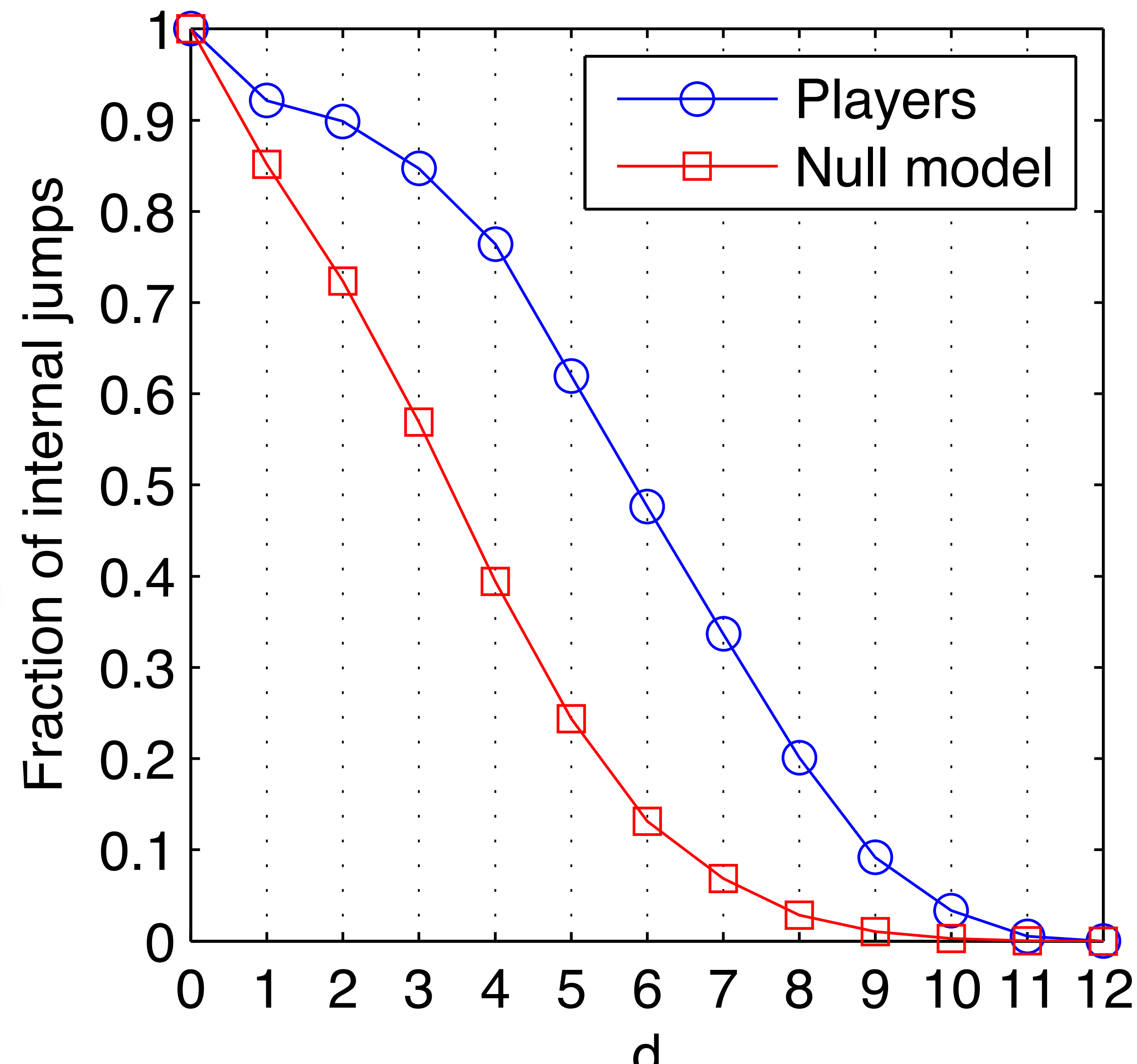
Mobility $\Rightarrow$ Activity

Activity $\not\Rightarrow$ Mobility

Players are highly heterogeneous

Do clusters influence mobility?





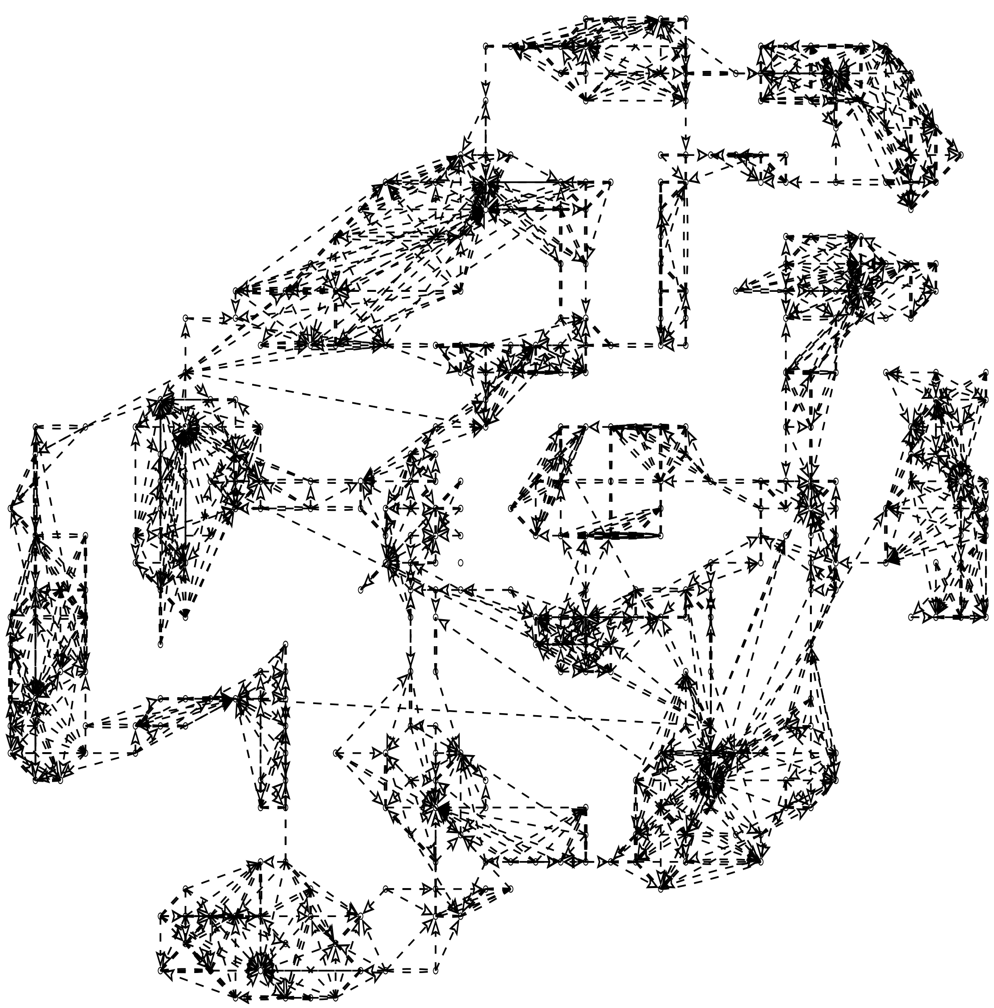
More movements within than between "countries"?

YES

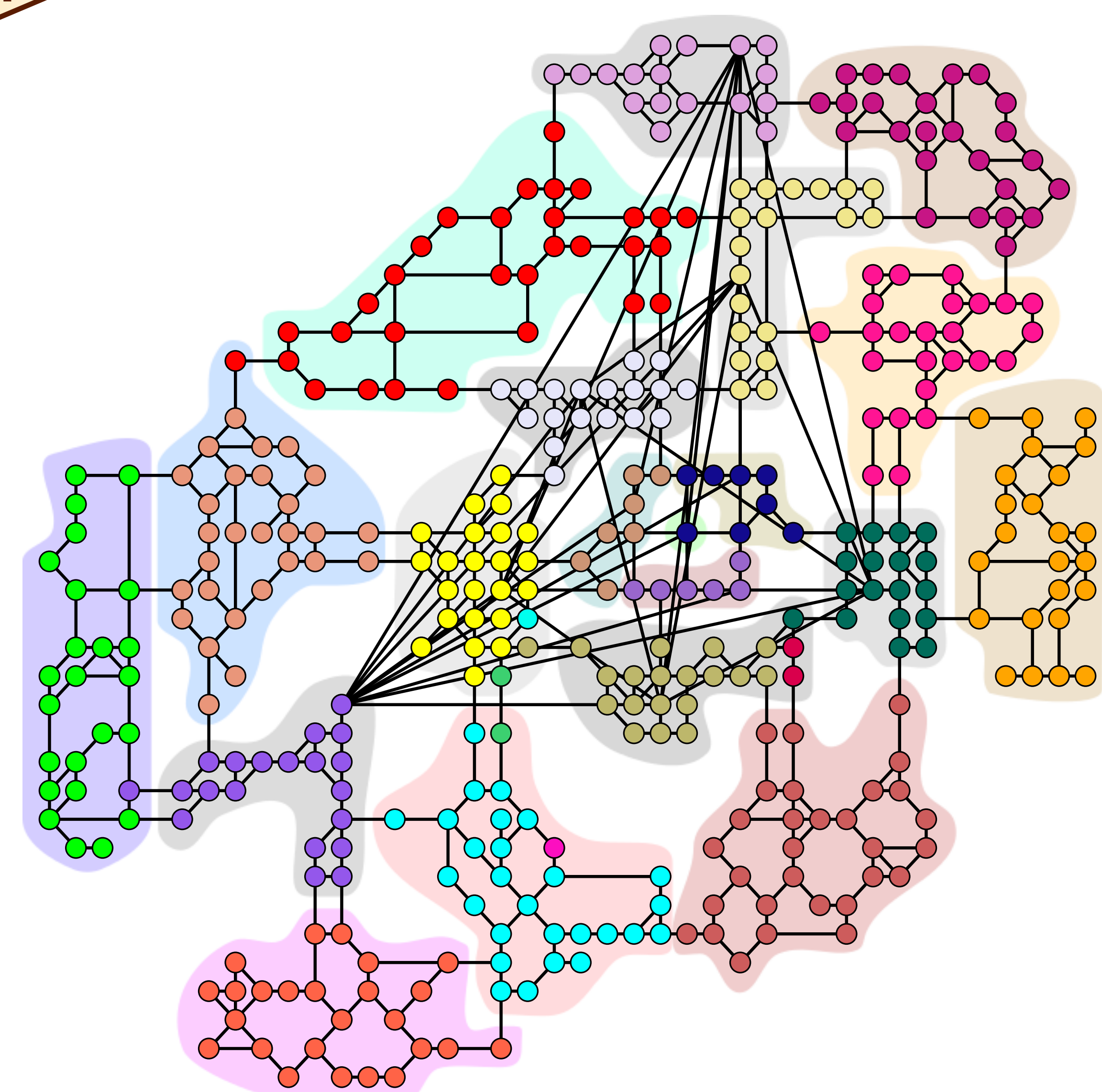
Mobility reveals socio-economic clusters

Community detection

Movement data

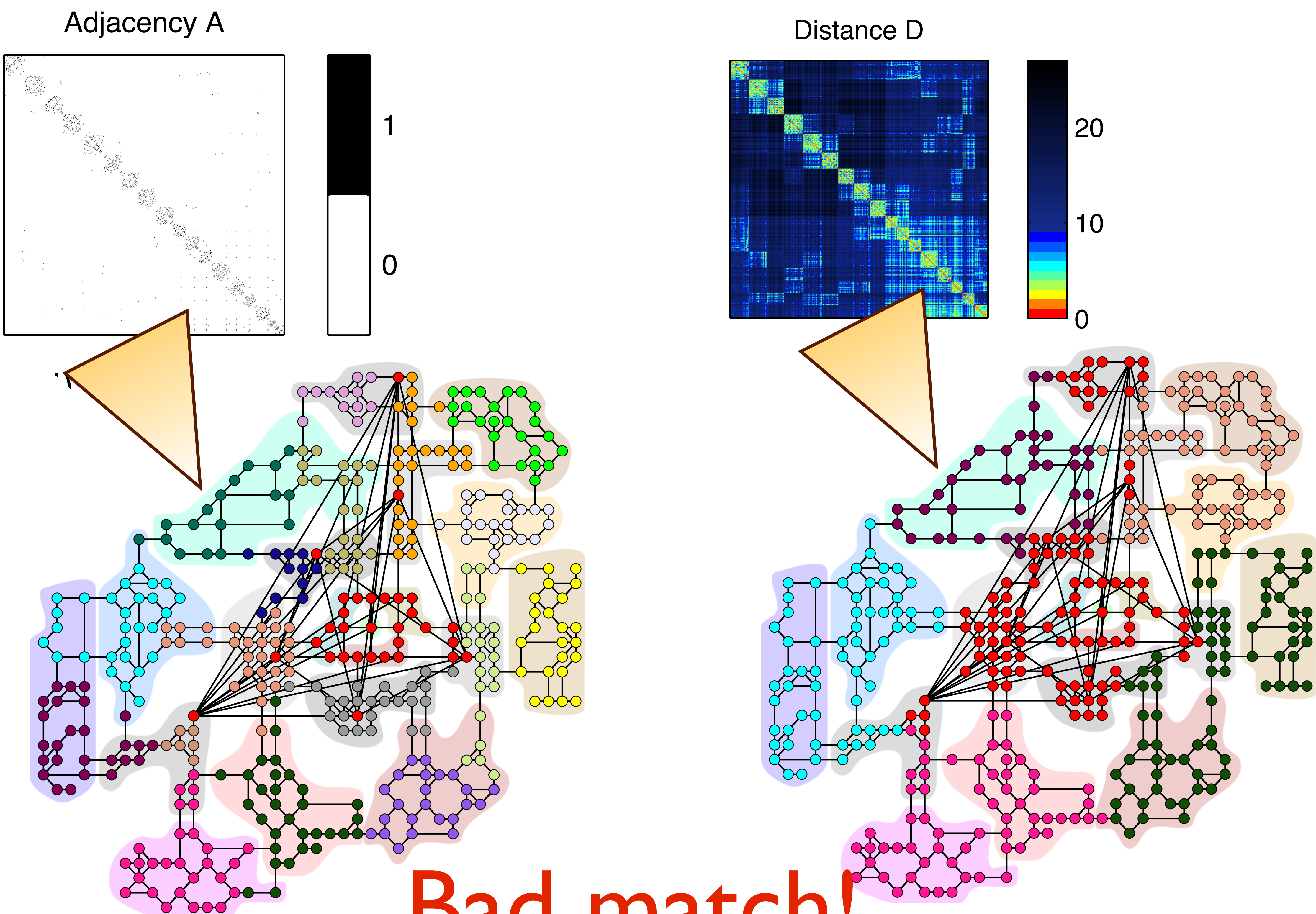


Clusters



(almost) Perfect match!

For comparison: Communities from topology

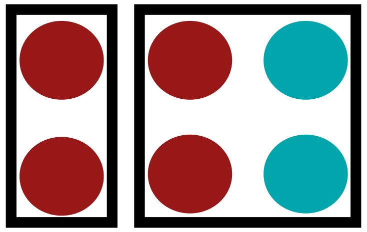


Bad match!

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

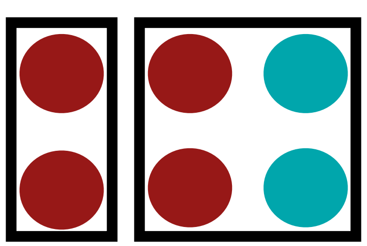
Many measures that quantify the match between two partitions X and Y build on these 4 numbers:

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

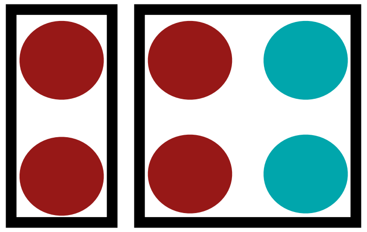


0.47 Rand index $\mathcal{R} = \dfrac{n_{00} + n_{11}}{n_{00} + n_{11} + n_{01} + n_{10}}$ is the accuracy of the pair classification

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

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

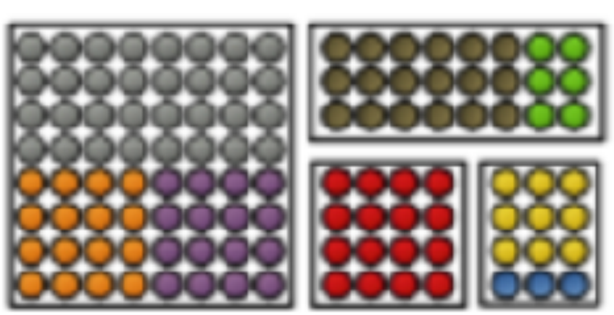
Measures for quantifying matching of partitions

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

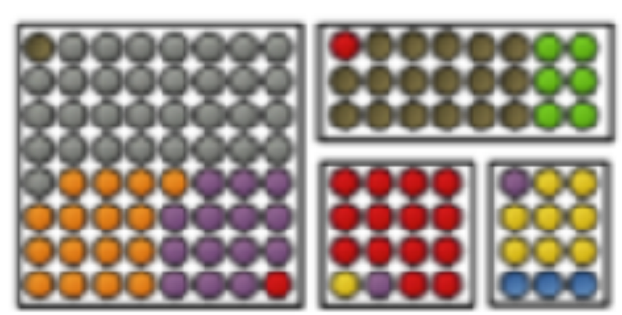
To correct for this baseline, there is an adjusted Rand index

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

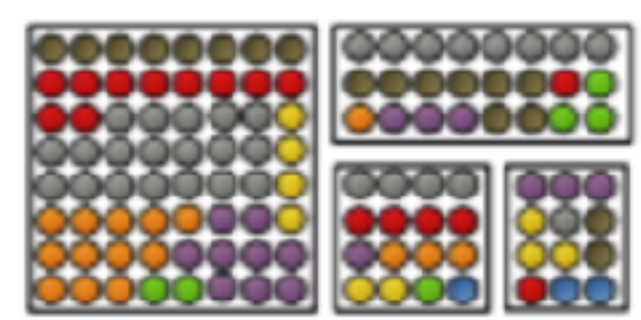
where $\mathbf{E}[n_{00} + n_{11}]$ are the values expected by chance. This index can be negative.



$\mathcal{ARI} = 1.00$



$\mathcal{ARI} = 0.88$



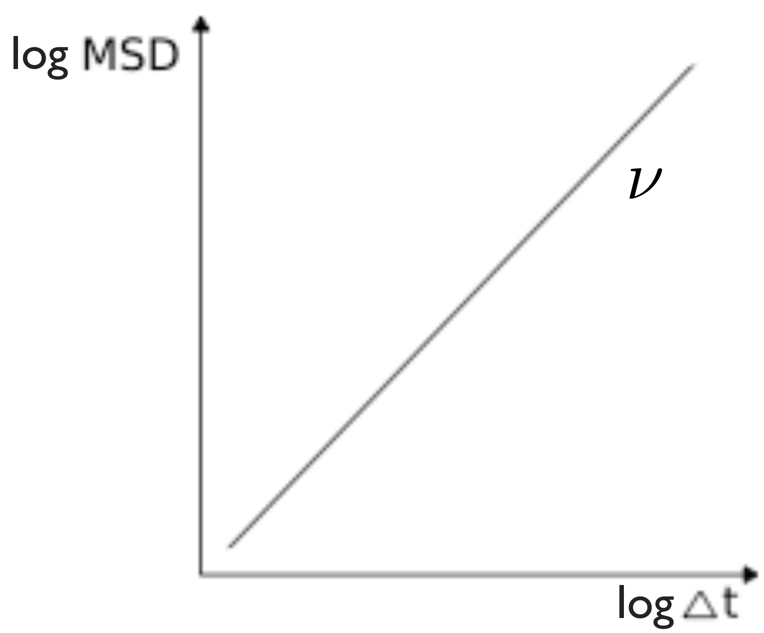
$\mathcal{ARI} = 0.03$

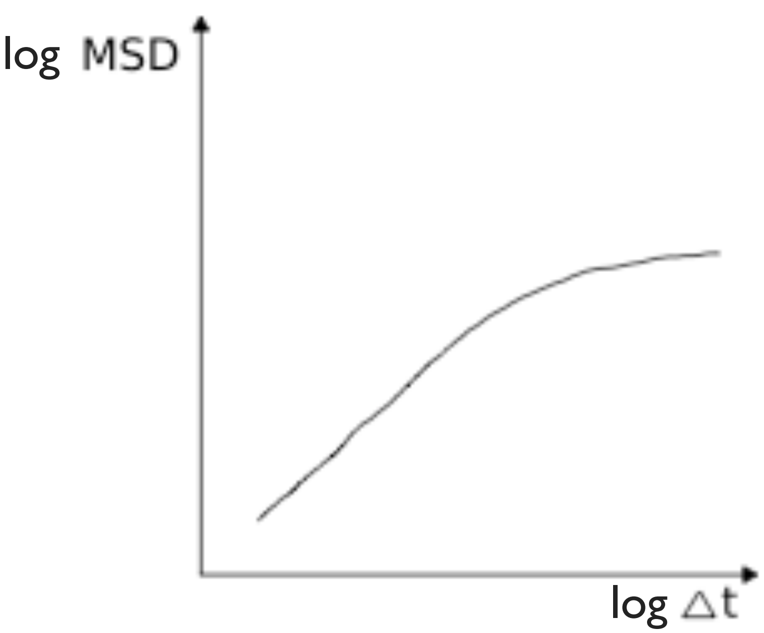
Diffusion and MSD

Mean Squared Displacement

finite universe

$$\nu = \lim_{t \to \infty} \frac{d}{dt}(\mathrm{MSD})$$





expect finite size effect

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

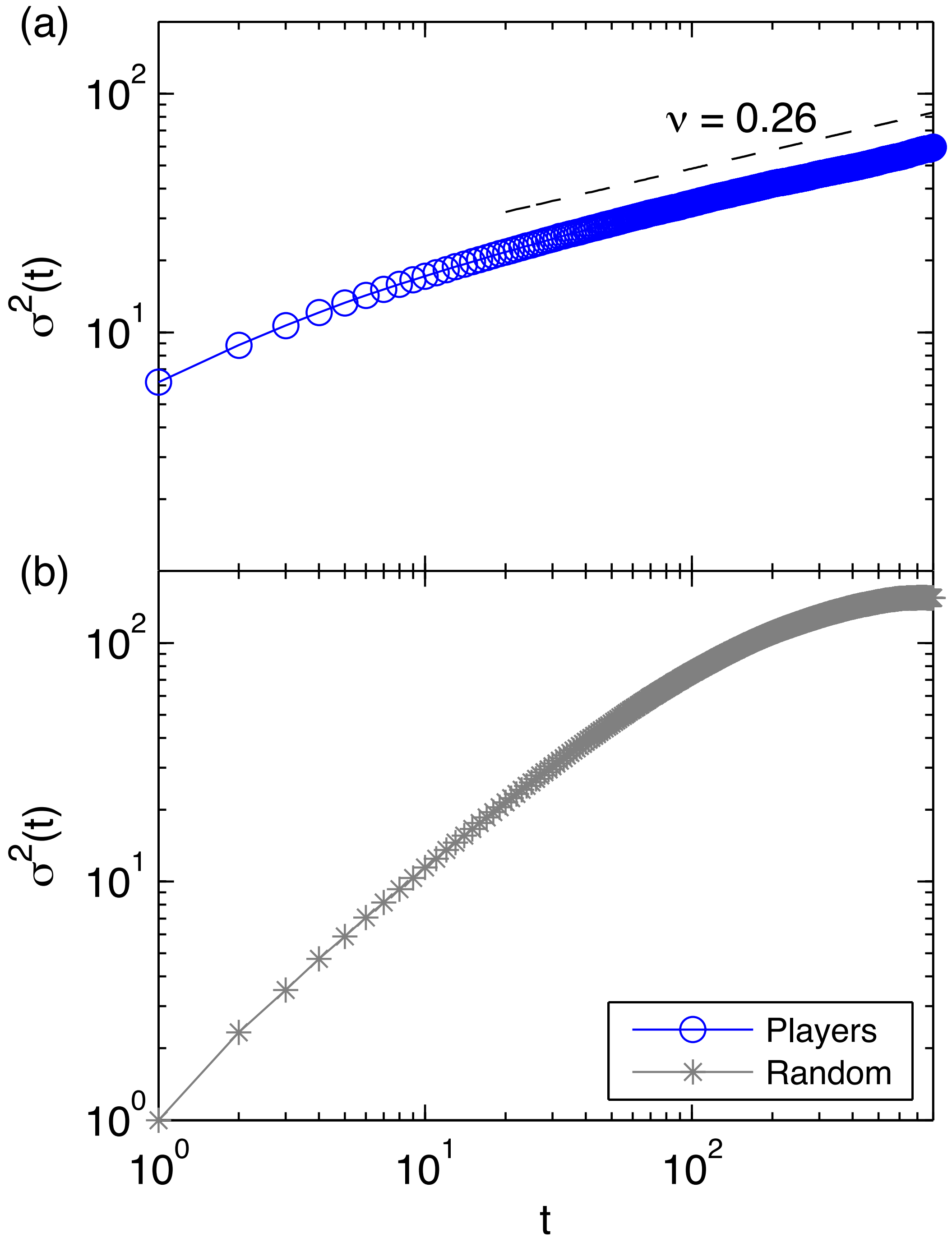
Anomalous diffusion

$\nu = 0.26 < 1$

Subdiffusive

Random walkers

have $\nu = 1$

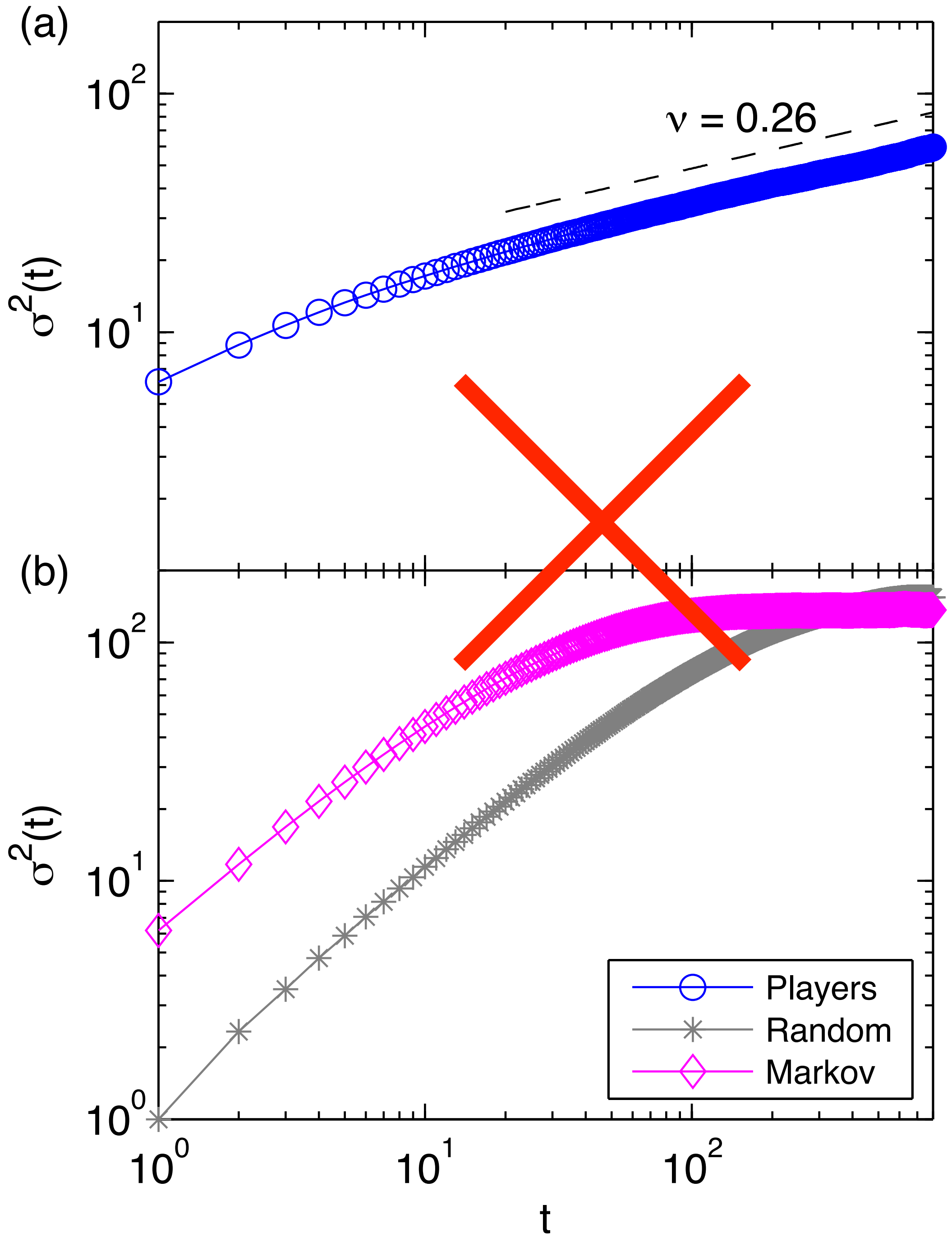


Players

Models

Model I: Markov

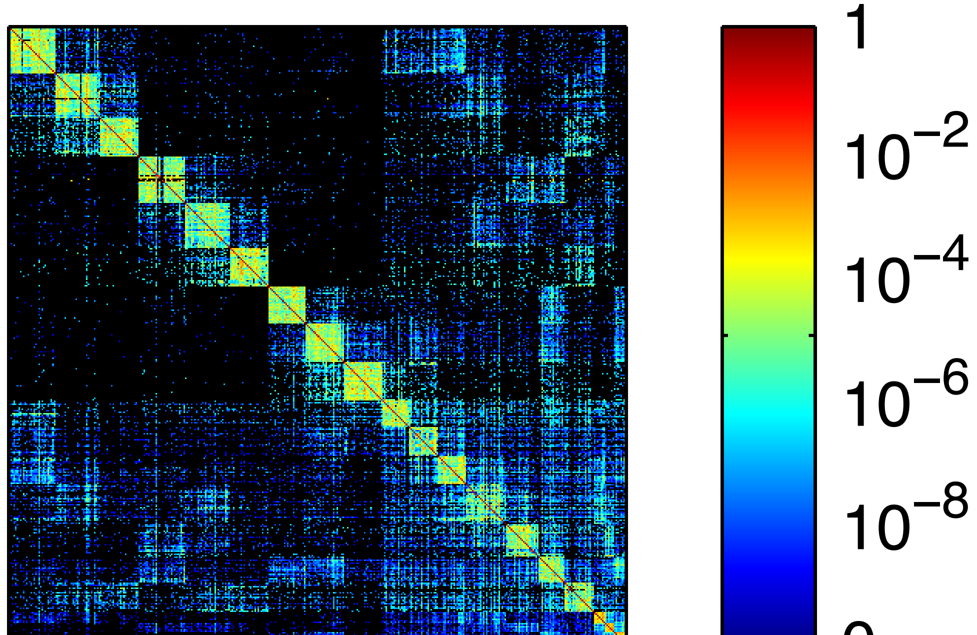
- Markov use <u>all</u> day-to-day transitions



Players

Models

Transition P

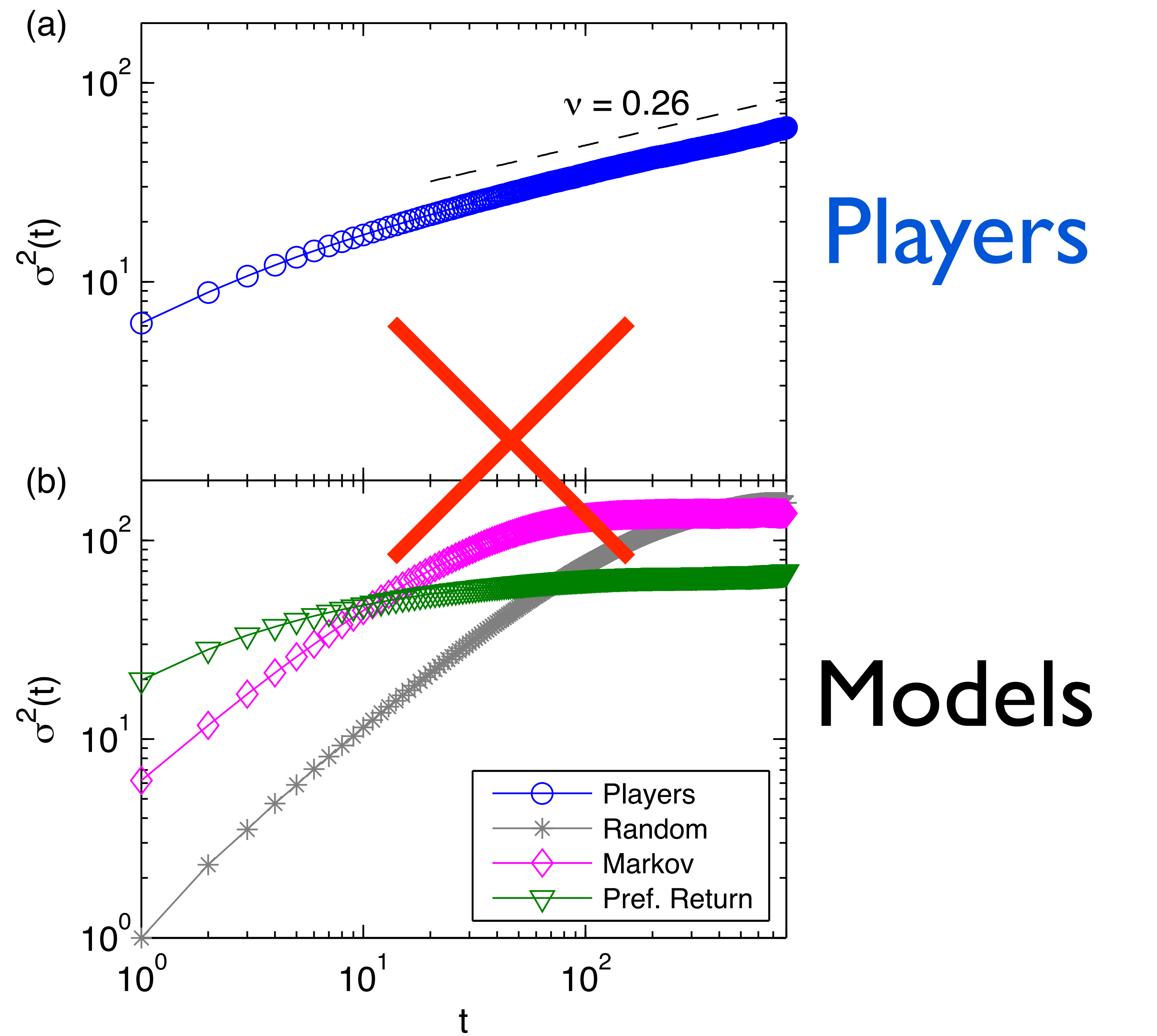


encodes all information, but no memory

Model II: Preferential return

- Markov use all day-to-day transitions

- Preferential return to <u>often</u> visited places

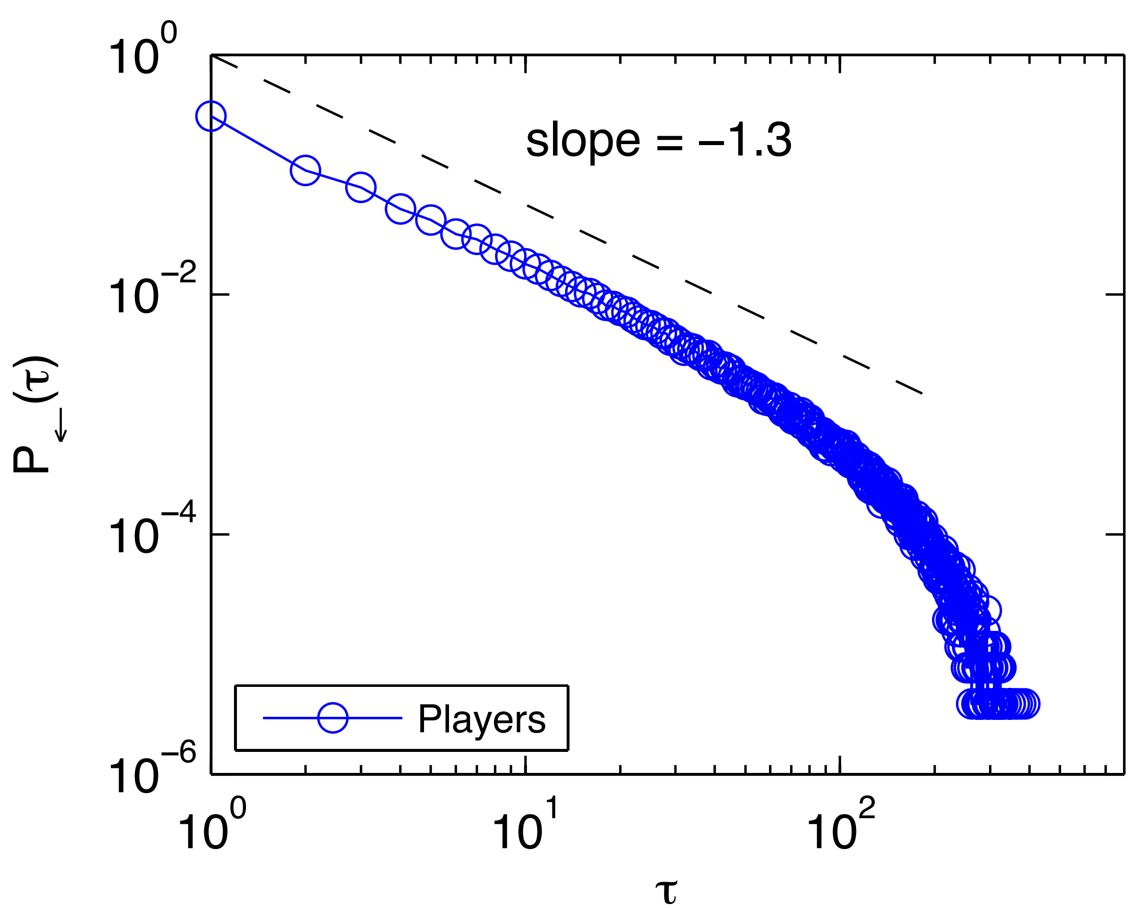


Song et al, Nature Physics 6, 818-823 (2010)

What is the essential ingredient?

Order of visitations!

Time to first return



...A B A B C D B A

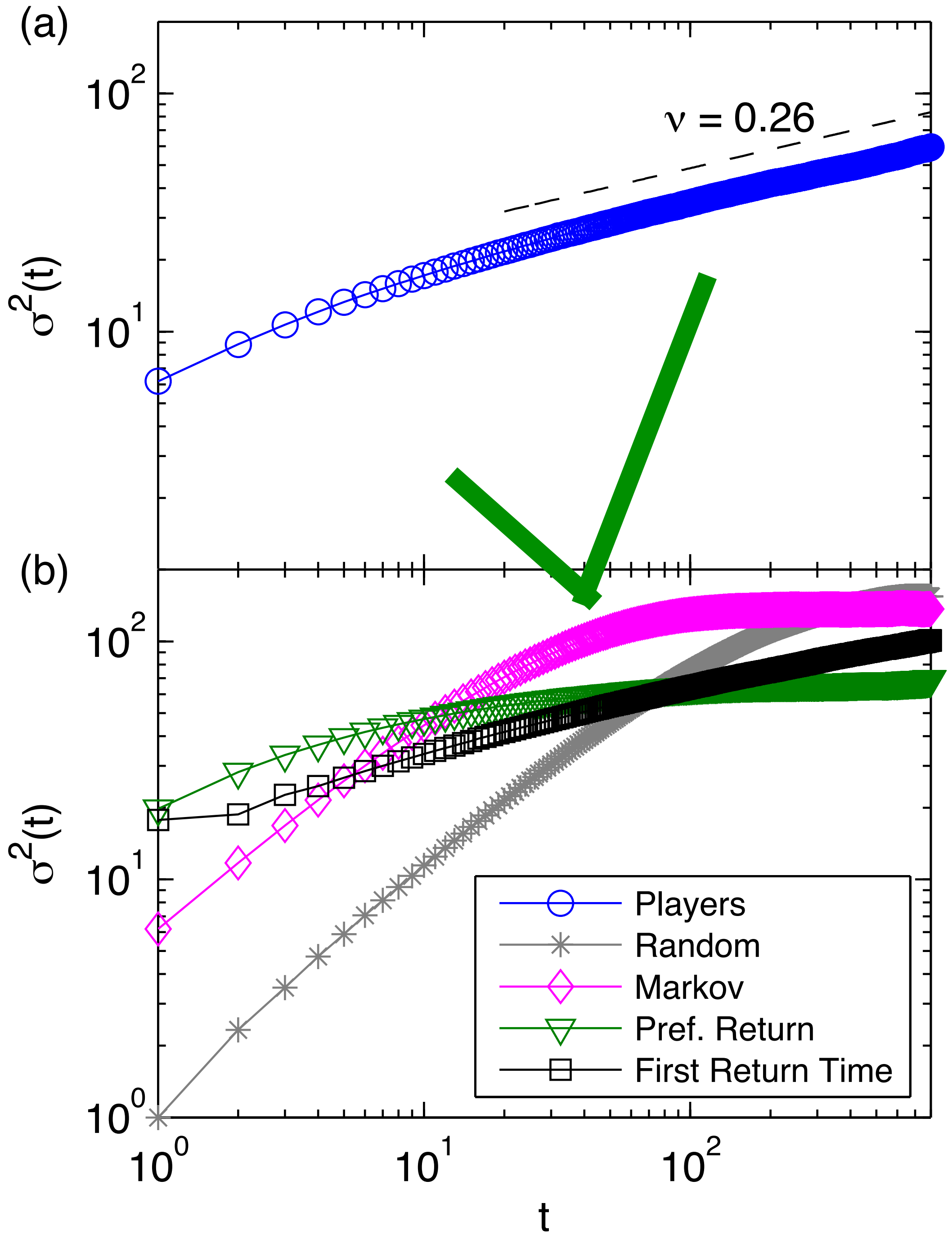
5

Long-time memory

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

Our Model: First return

- Markov use all day-to-day transitions
- Preferential return to often visited places
- First return to <u>recently</u> visited places



Players

Models

Our Model: First return

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

2a) With probability ν

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

2b) With probability $1-\nu$

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

$\beta = 2.2$

$\nu = 0.9$

$\alpha = 1.3$

$\lambda = 3$

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.

Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S.H. and Ratti, C. PNAS 111 (2014)

Urban transportation is facing serious issues

Pollution

3.7 Mio.

Traffic jams

Fatalities

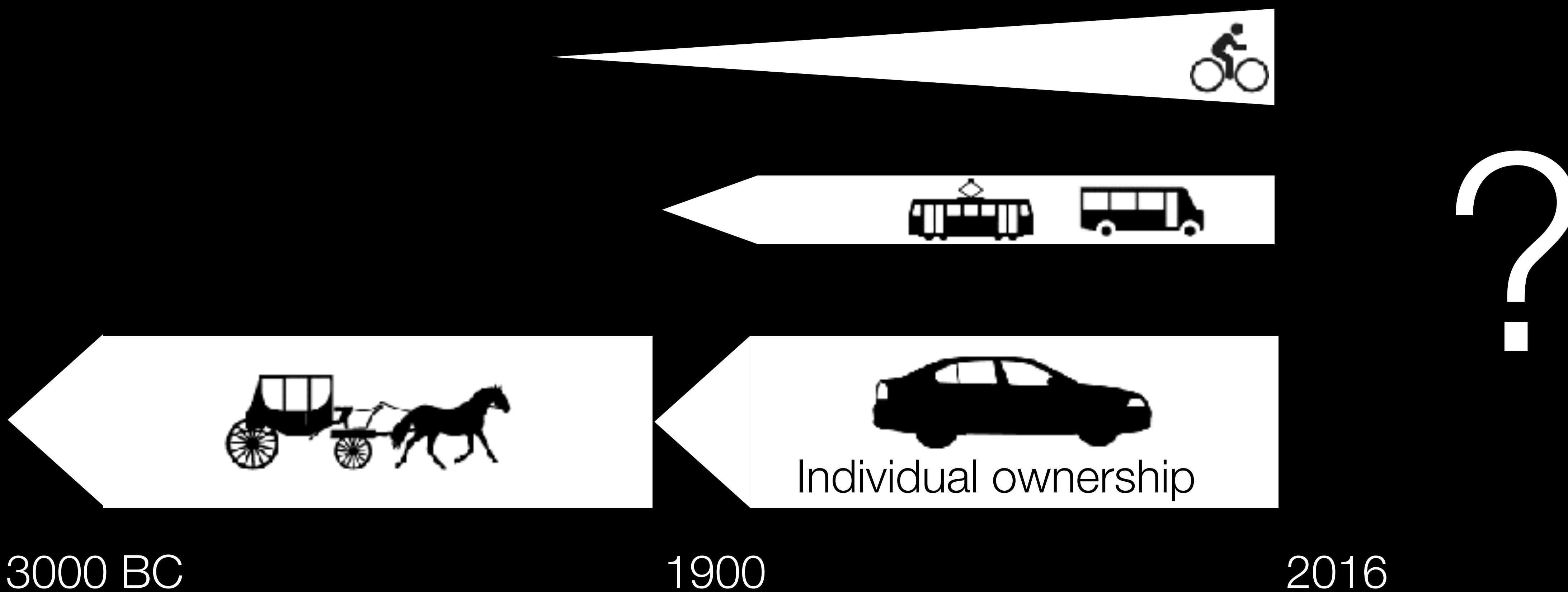
1.2 Mio.

How efficient is urban transport?

We don’t know.

Caiazzo et al, Atmosph Env 79 (2013)
WHO statistics (2011)

Urban transportation is at a turning point



Warner, S.B. Harvard Univ. Press (1978)

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

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

Urban transportation is changing

Self-driving cars

Lack of driver further motivates giving up ownership

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

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

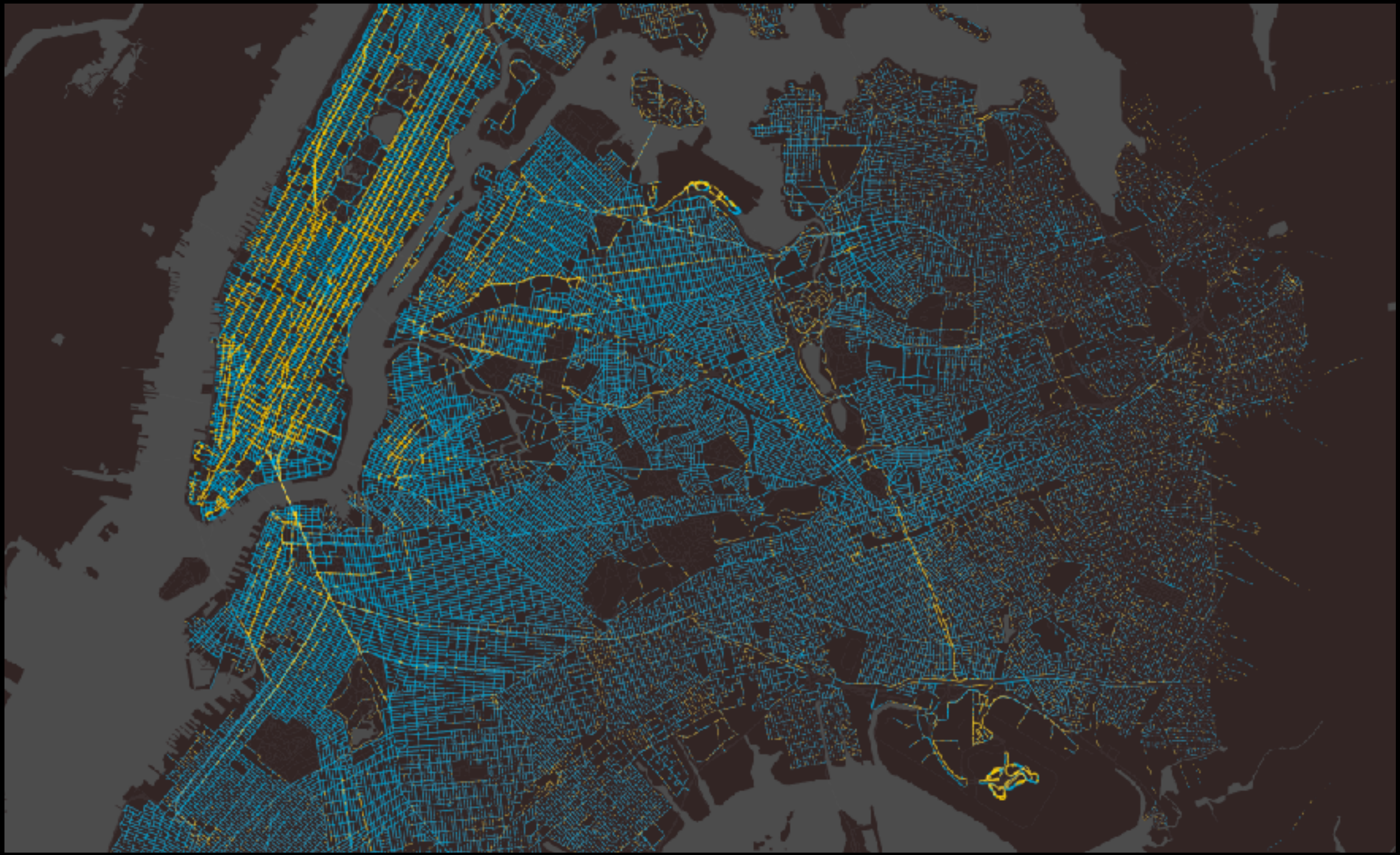
1) Quantify existing taxi systems
2) Design an improved system?

Goals
- more efficient
- less emissions
- affordable alternative

NYC
13,500 cabs

Step 1: Analyze data

NYC taxi trips in 2011

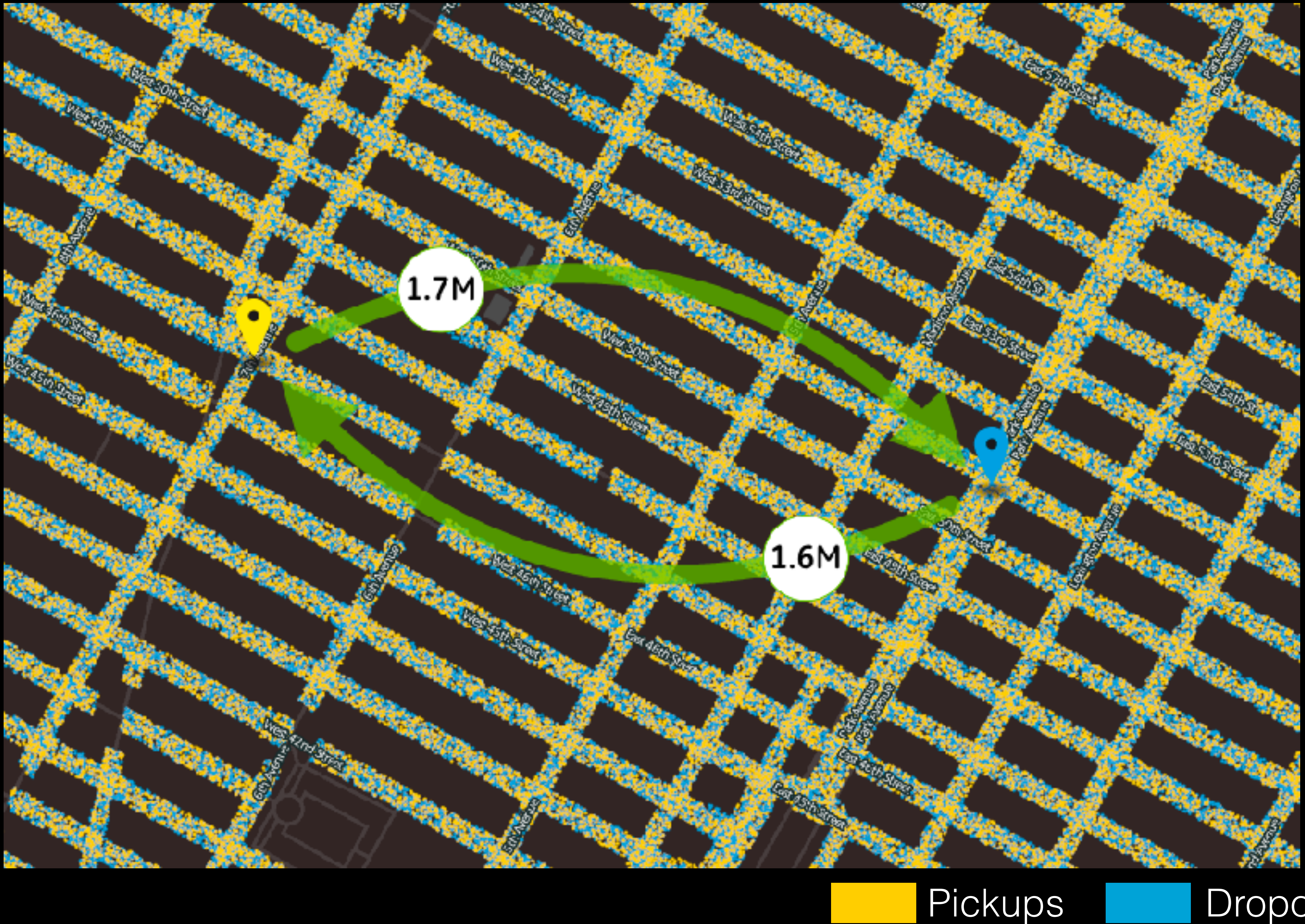


13500 cabs
150 million trips ~400.000 per day

Pickups Dropoffs

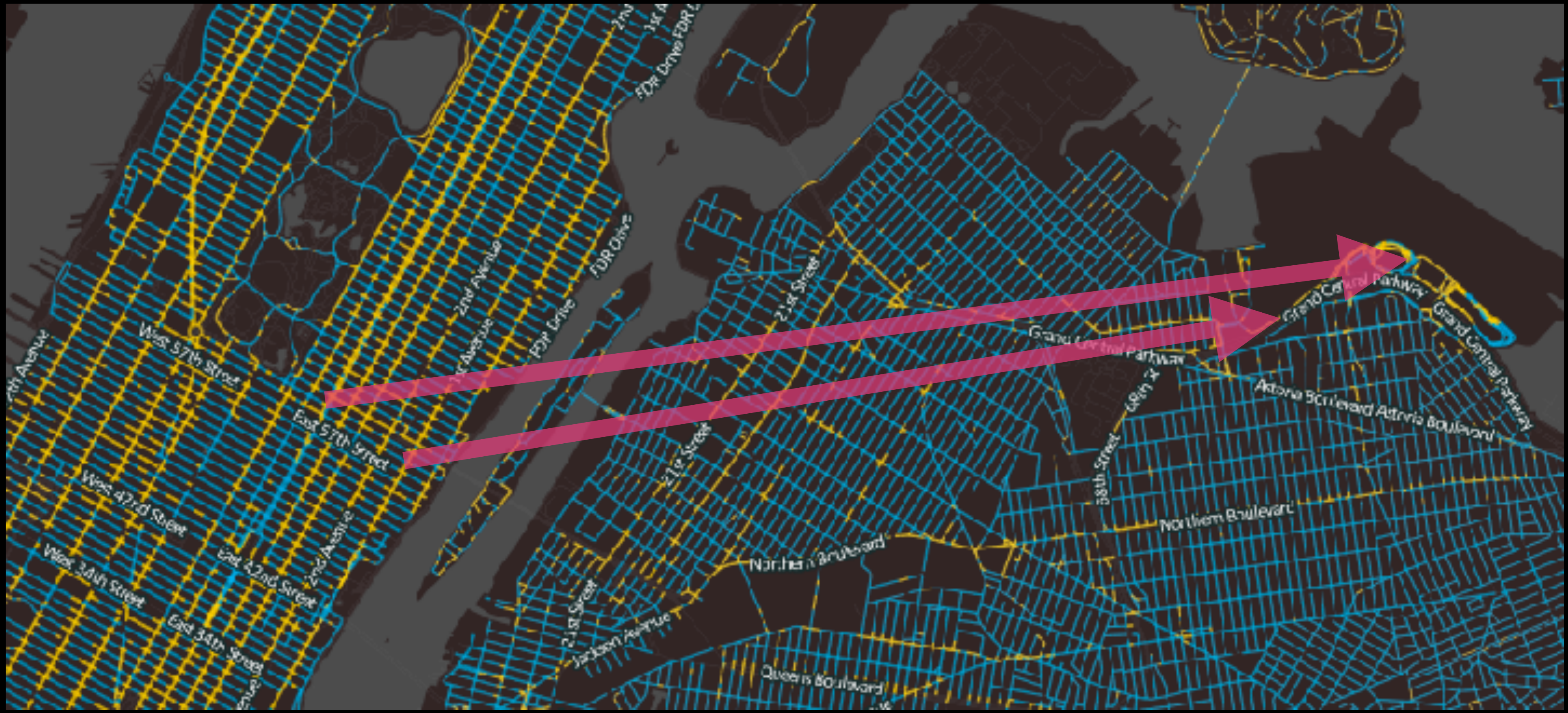
Step 1: Analyze data

Trips can be combined



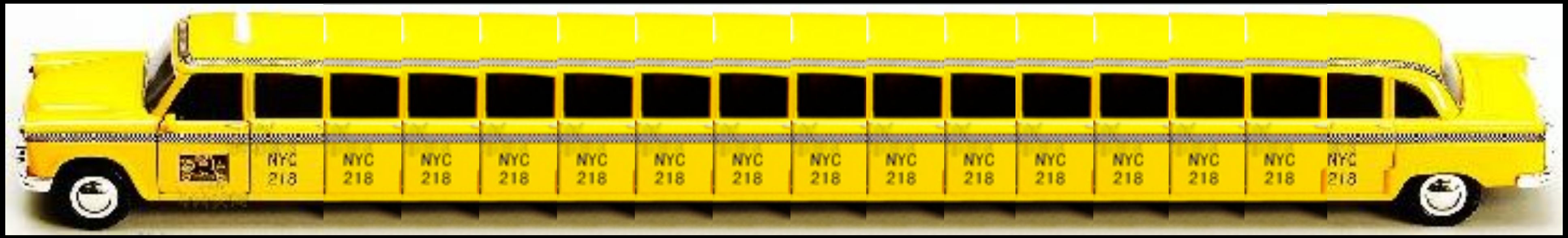
Step 2 : A new dispatch process

Combine 2 trips



Step 2 : A new dispatch process

Combine k trips



Oestreicher-Singer and Sundararajan, MIS Quarterly 36 (2012)

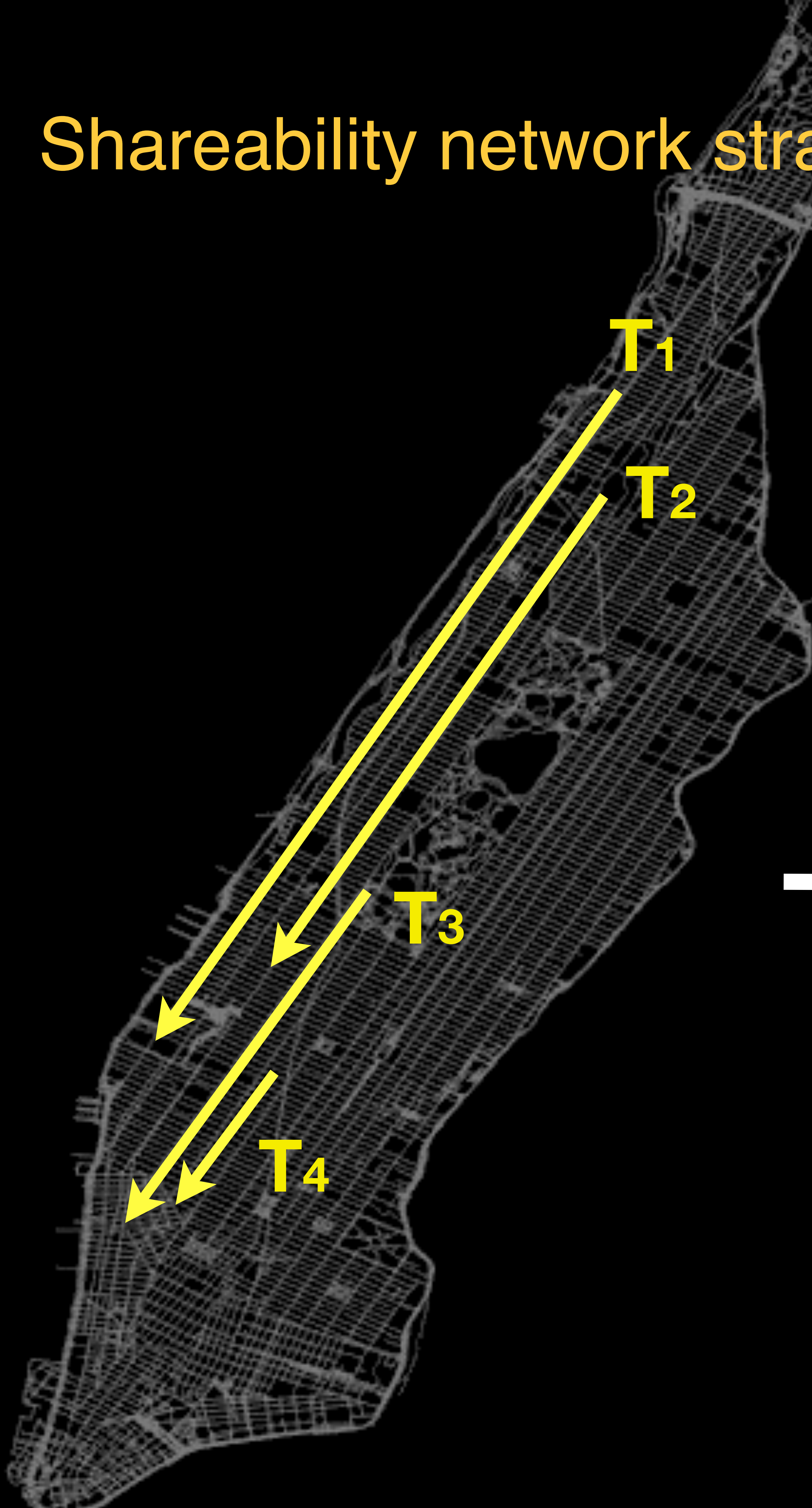
Street network

Extraction from OpenStreetMap

9000 street segments

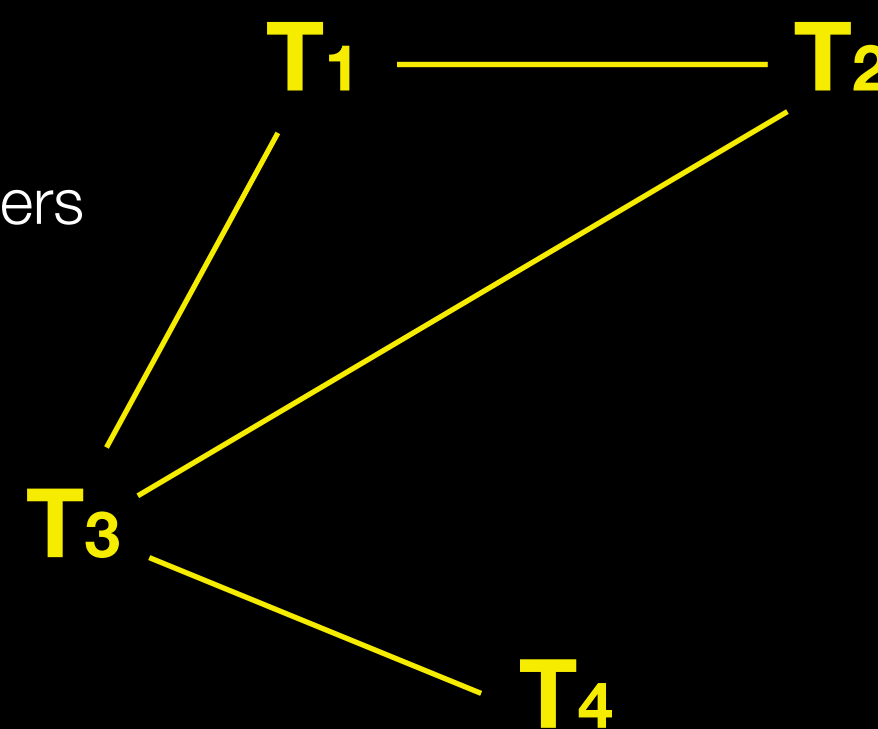
Matching GPS-coordinates

Shareability network strategy



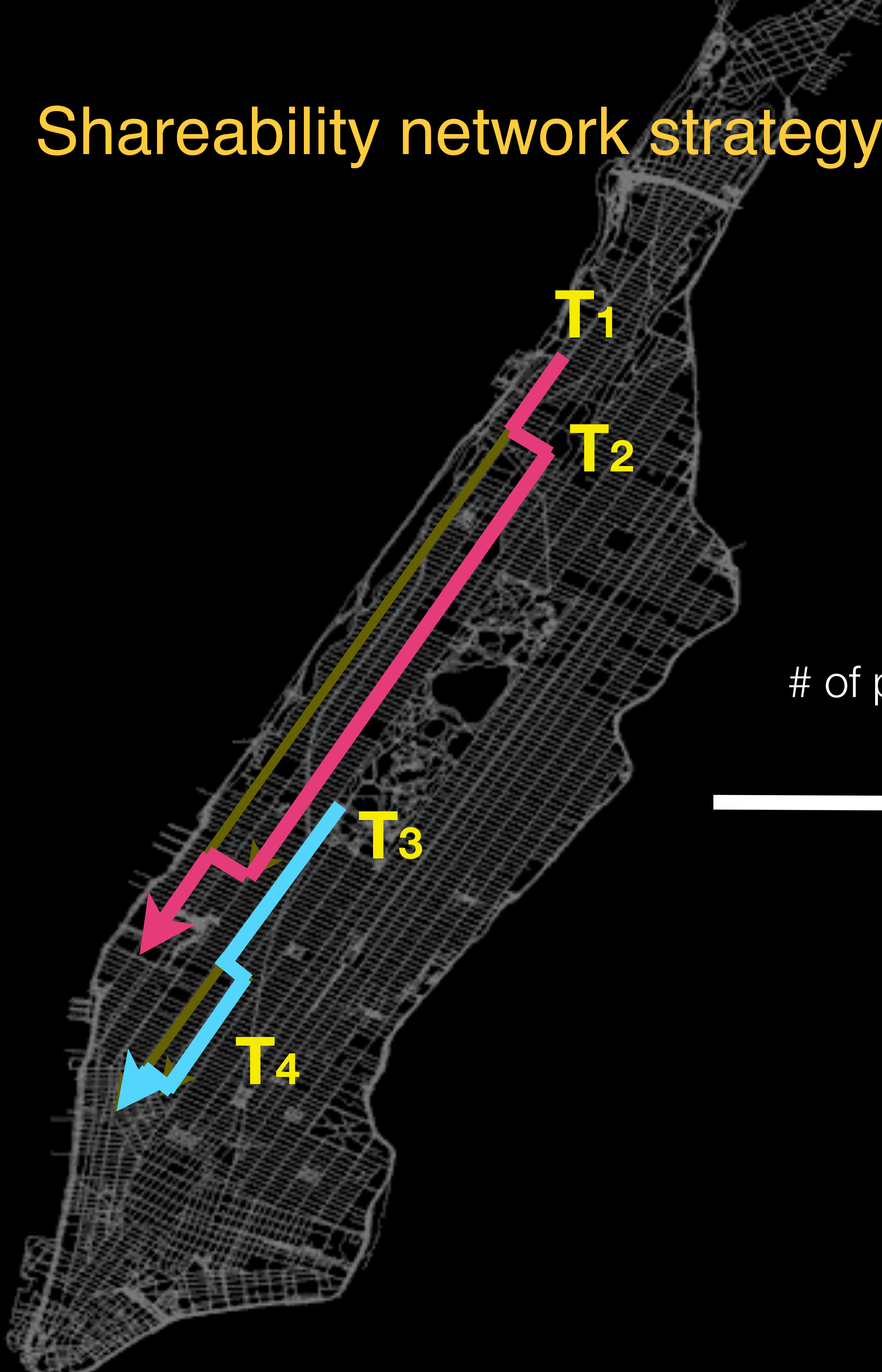
Mapping

# of passengers
k = 2



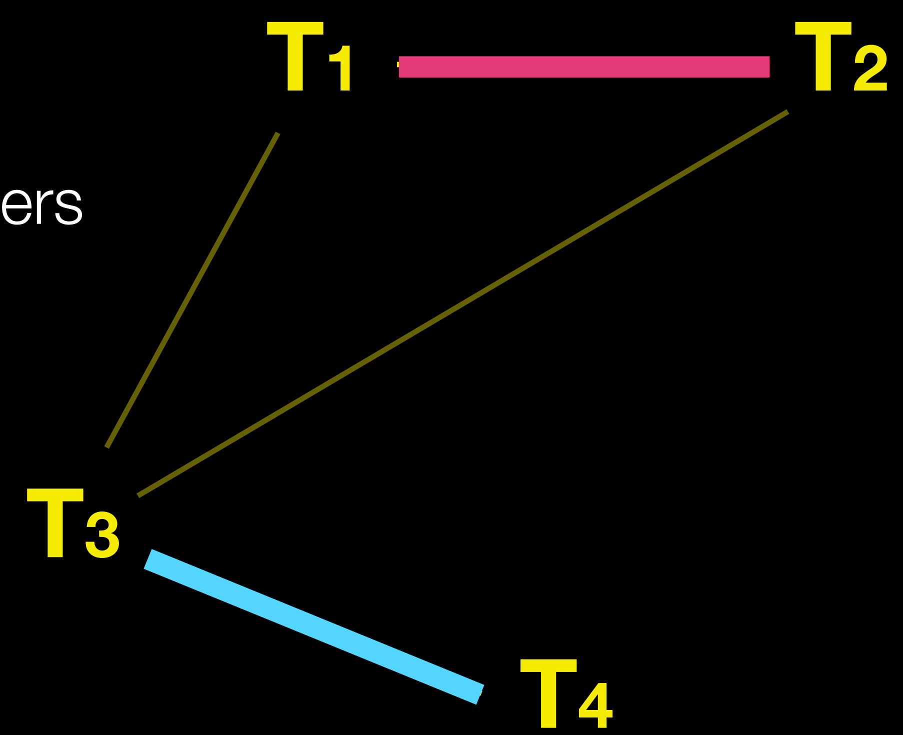
Chandra and Halldorsson, J Alg 39 (2001)

Shareability network strategy



# of passengers
k = 2

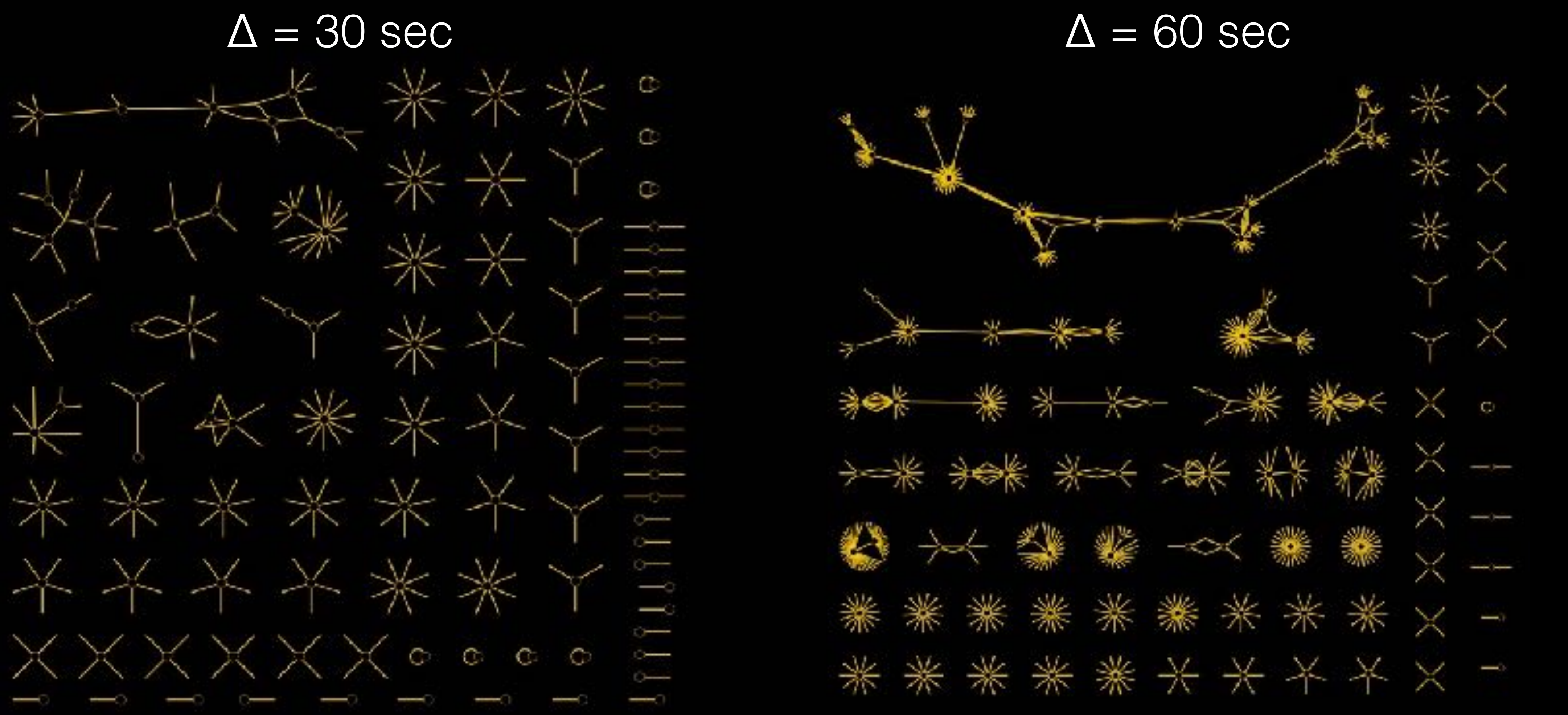
Maximum matching



Chandra and Halldorsson, J Alg 39 (2001)

Shareability network densification

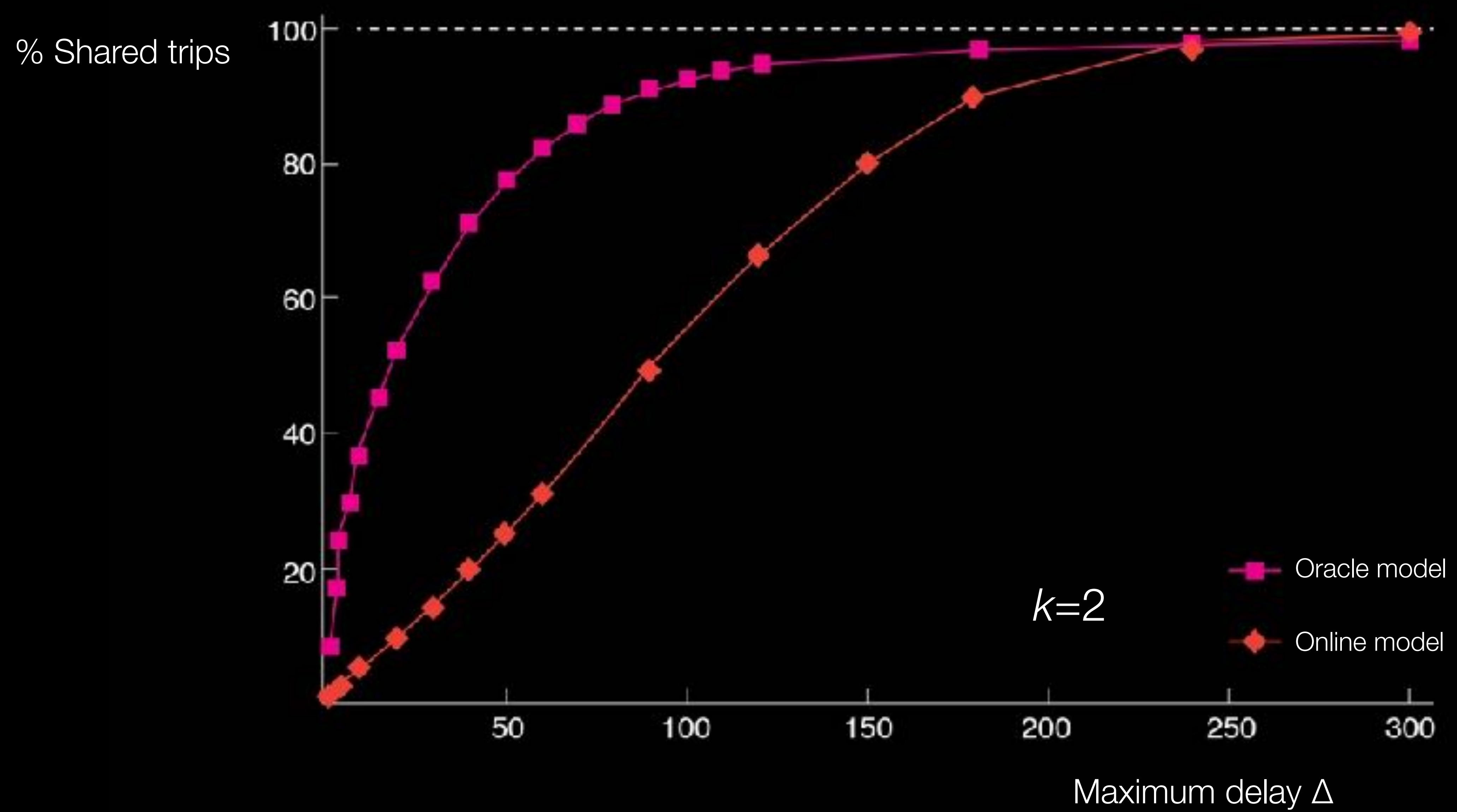
Maximum time delay Δ



More tolerance = denser network = more sharing opportunities

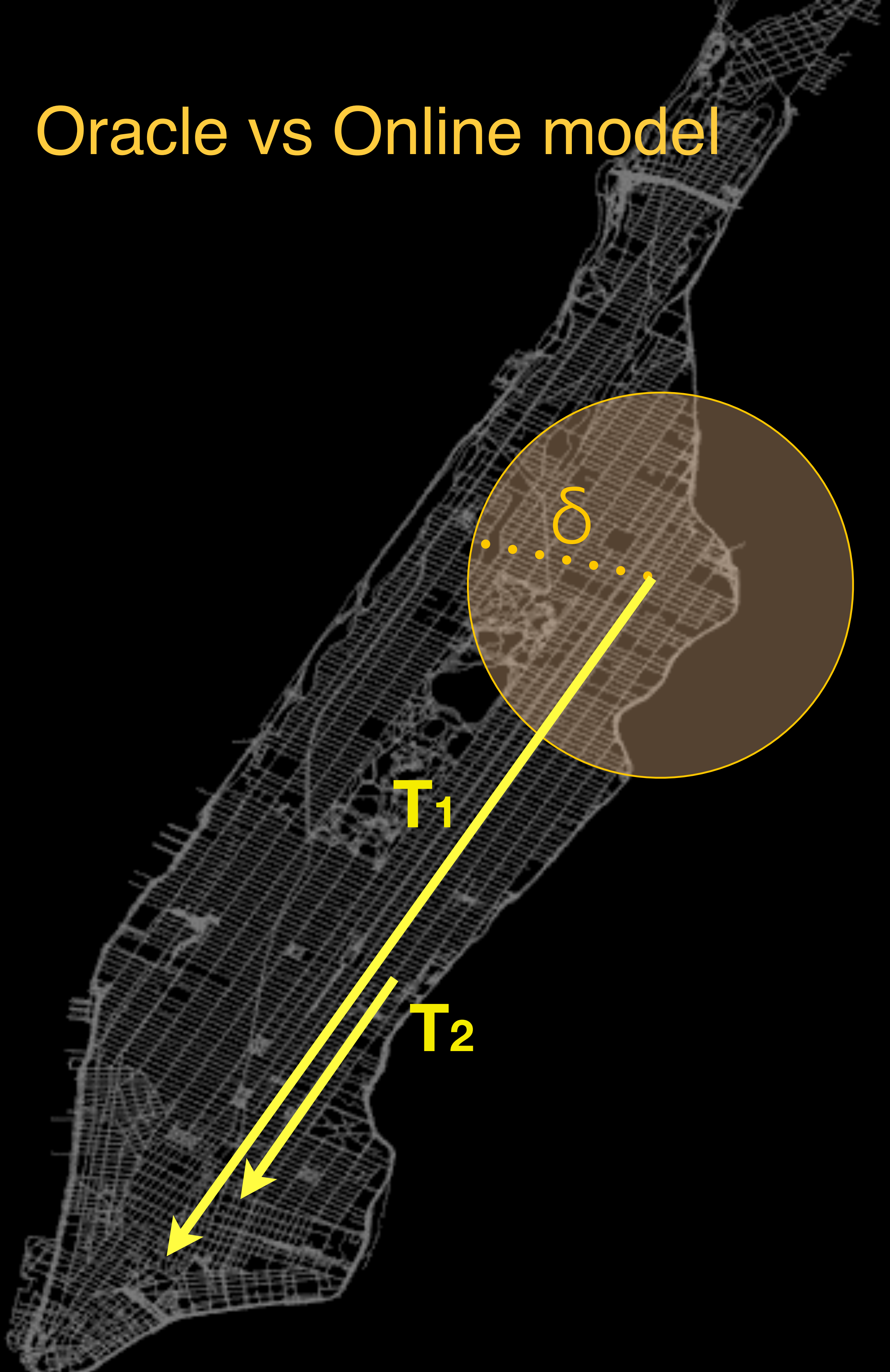
Krings et al, EPJ Data Sci 1 (2012)

The majority of trips is shareable!



The majority of trips is sharable with minimal passenger inconvenience

Oracle vs Online model

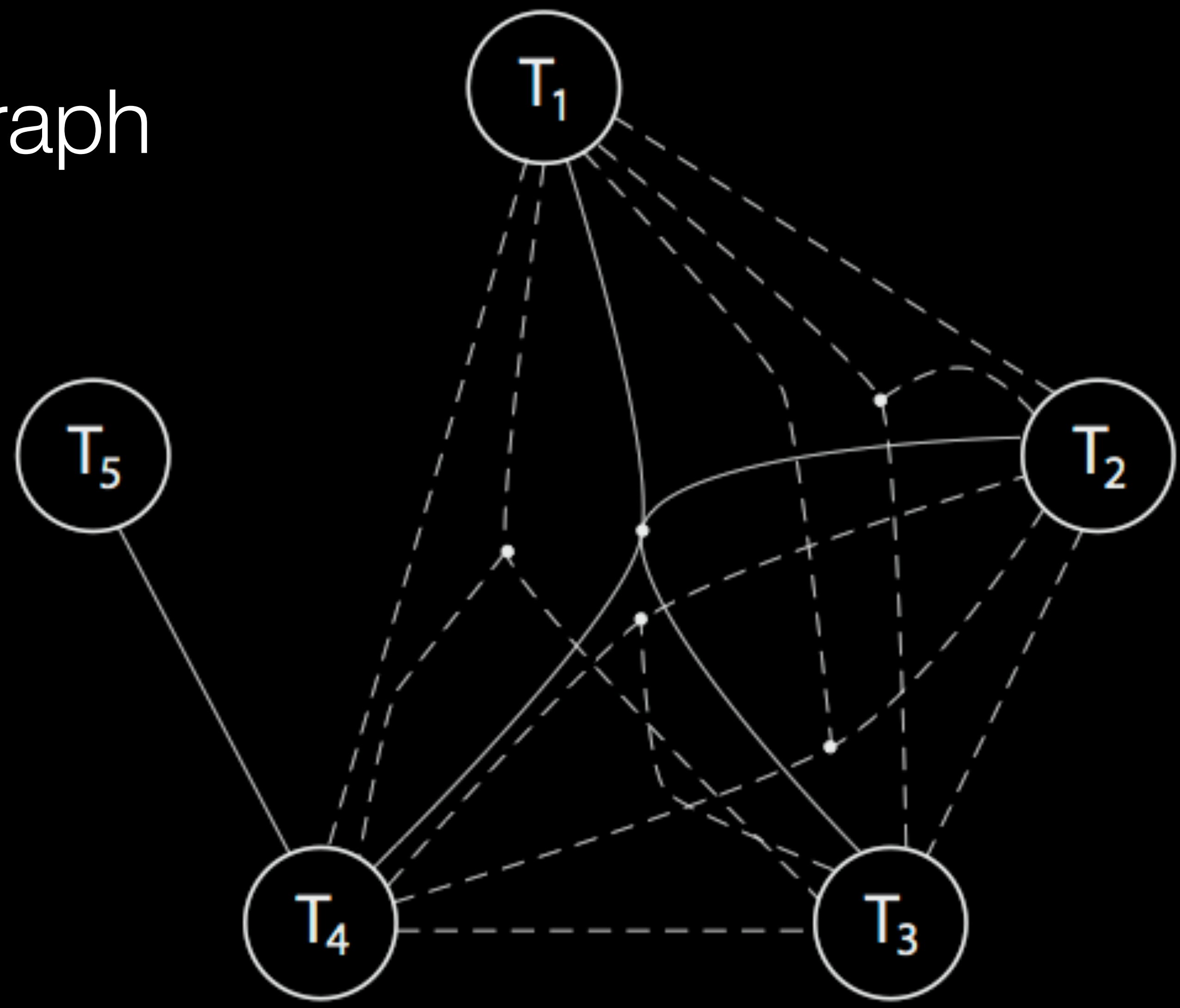


Oracle: omniscient

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

Approach can be extended to higher dimensions

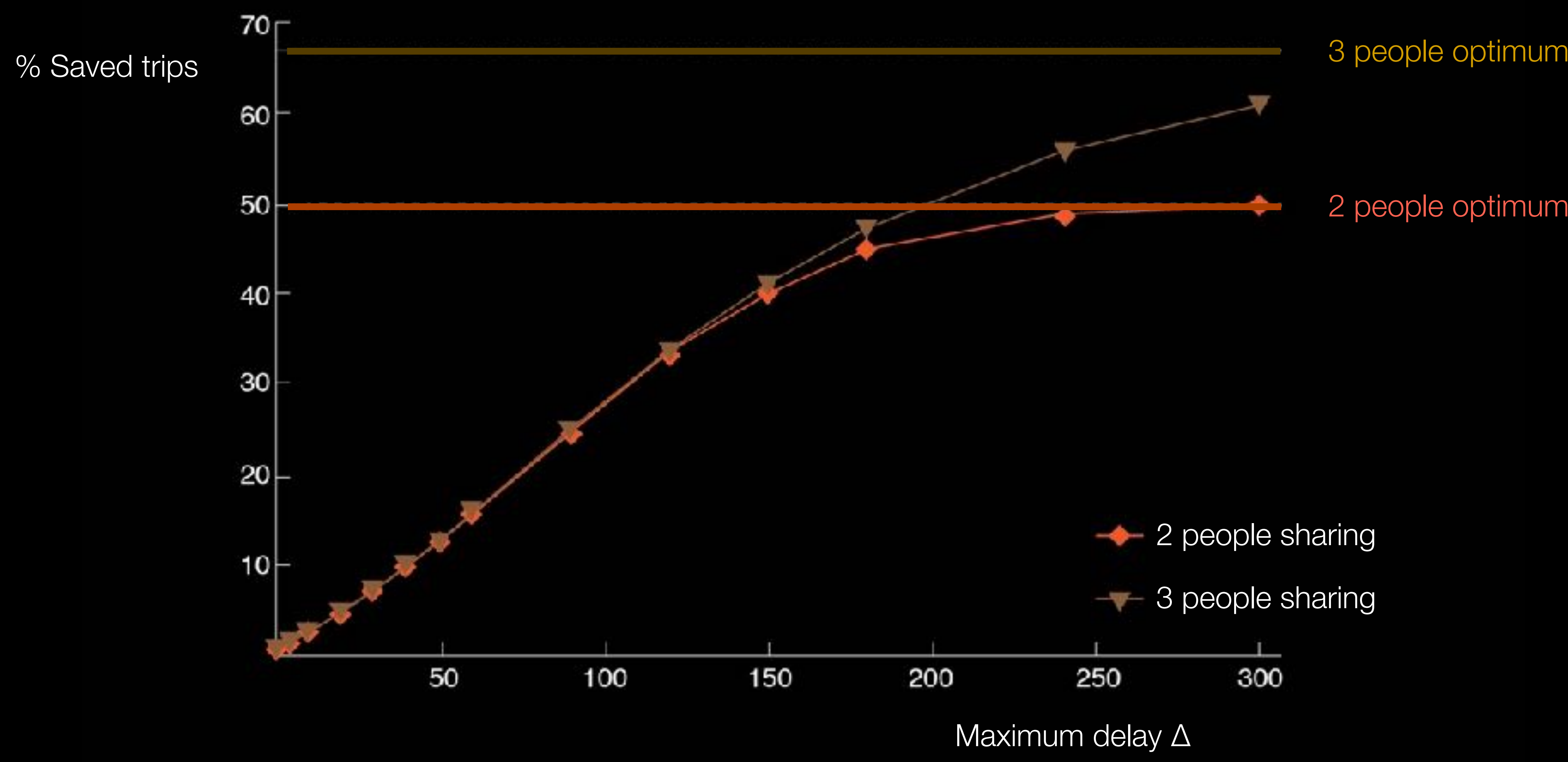
Trip hypergraph



Approximation in O(n log n)

Chandra and Halldorsson, J Alg 39 (2001)

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

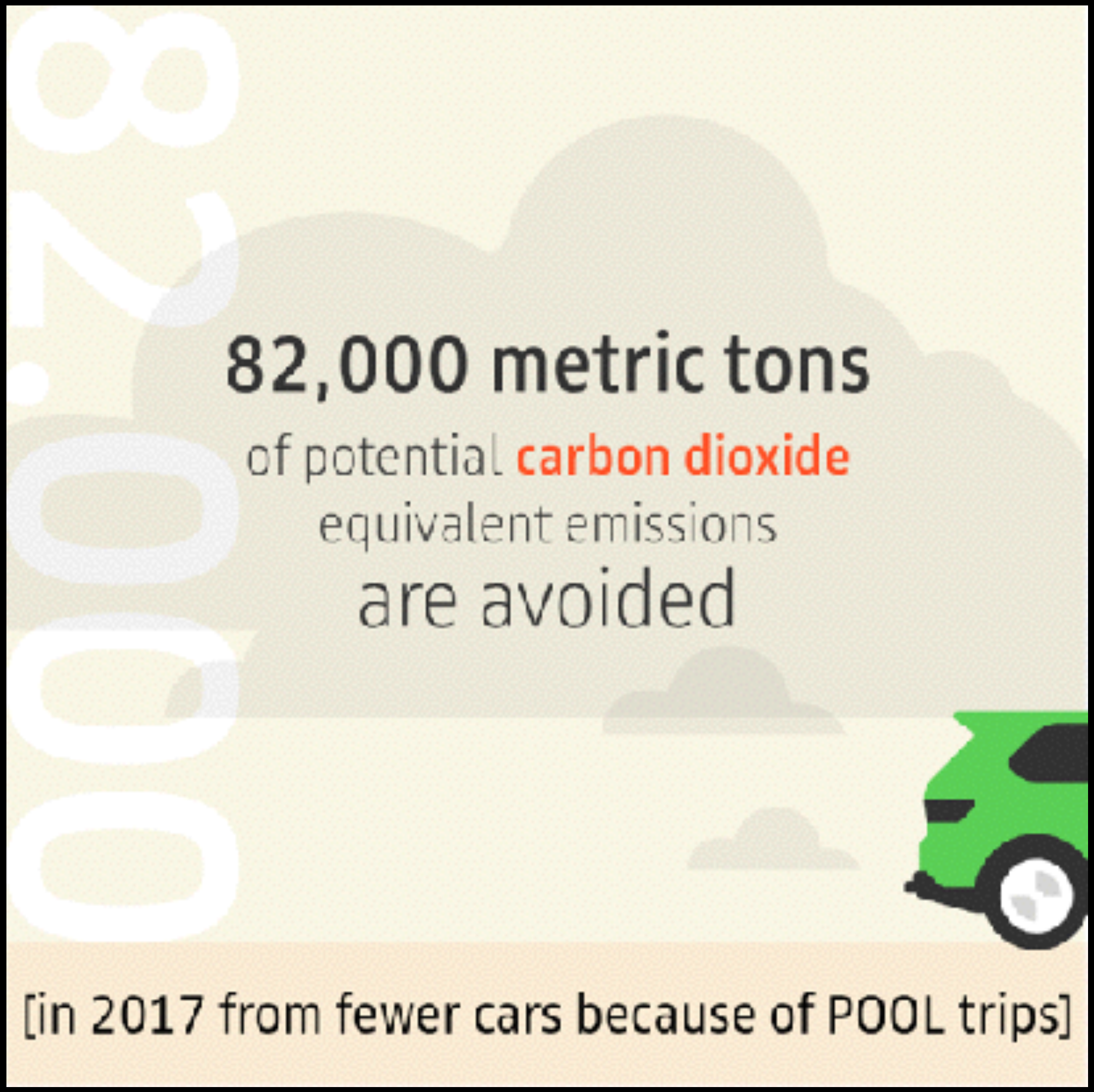
Online tool hubcab for interactive exploration of sharing benefits

http://www.hubcab.org

Zoom into the data, flow exploration, time selection

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

Lopez et al., PNAS 111.51, E5488 (2014)
Santi, Resta, Szell, Sobolevski, Strogatz, Ratti, PNAS 111.51, E5489 (2014)

3.

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Scaling Law of Urban Ride Sharing

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

Received: 31 October 2016
Accepted: 18 January 2017
Published: 06 March 2017

Sharing rides could drastically improve the efficiency of car and 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.

Tachet, Sagarra, Santi, Resta, Szell, Strogatz, Ratti, Scientific Reports 7 (2017)

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/km²

San Fran 1,500 cabs

6,659/km²

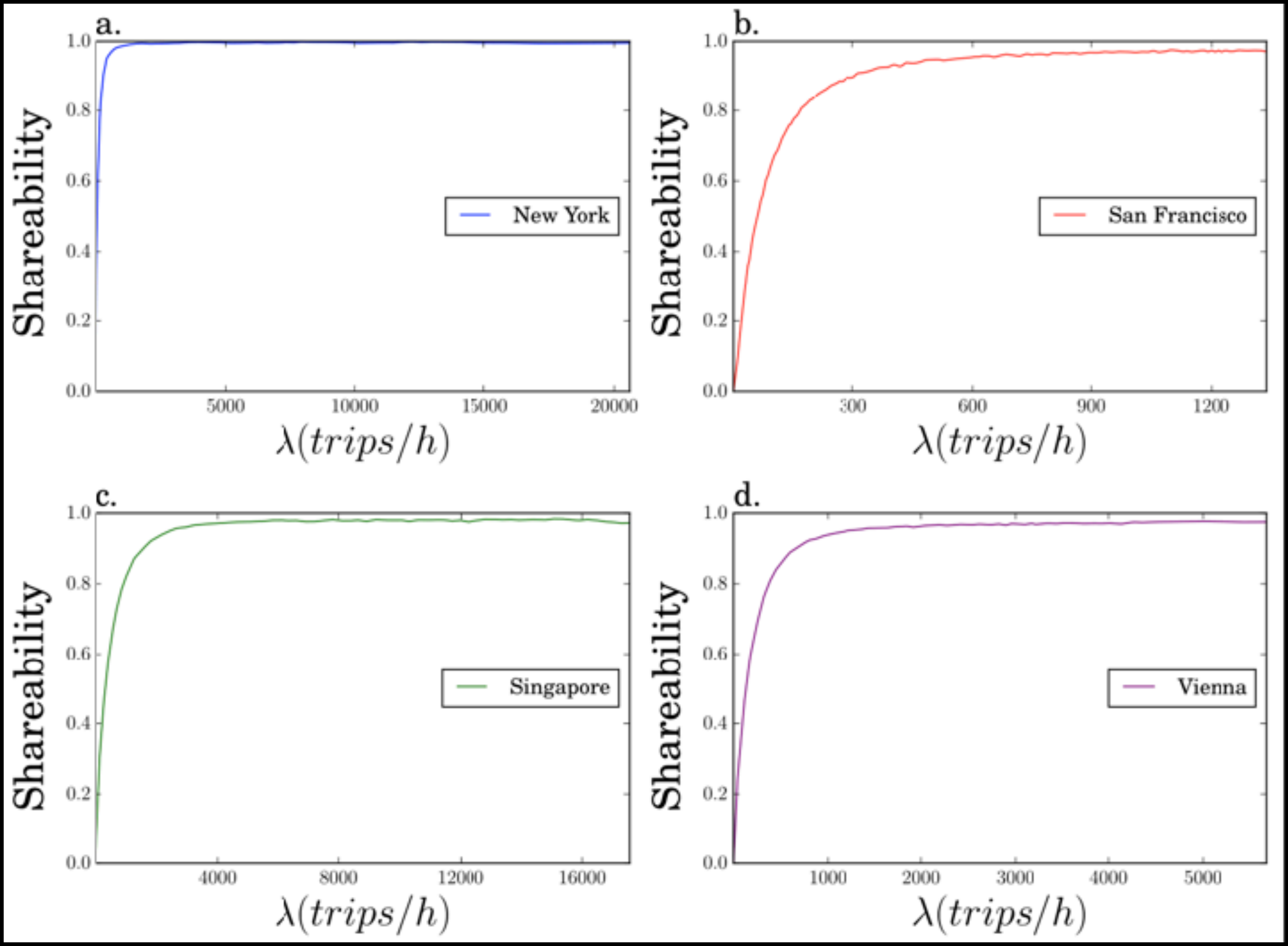
Singapore 26,000 cabs

7,988/km²

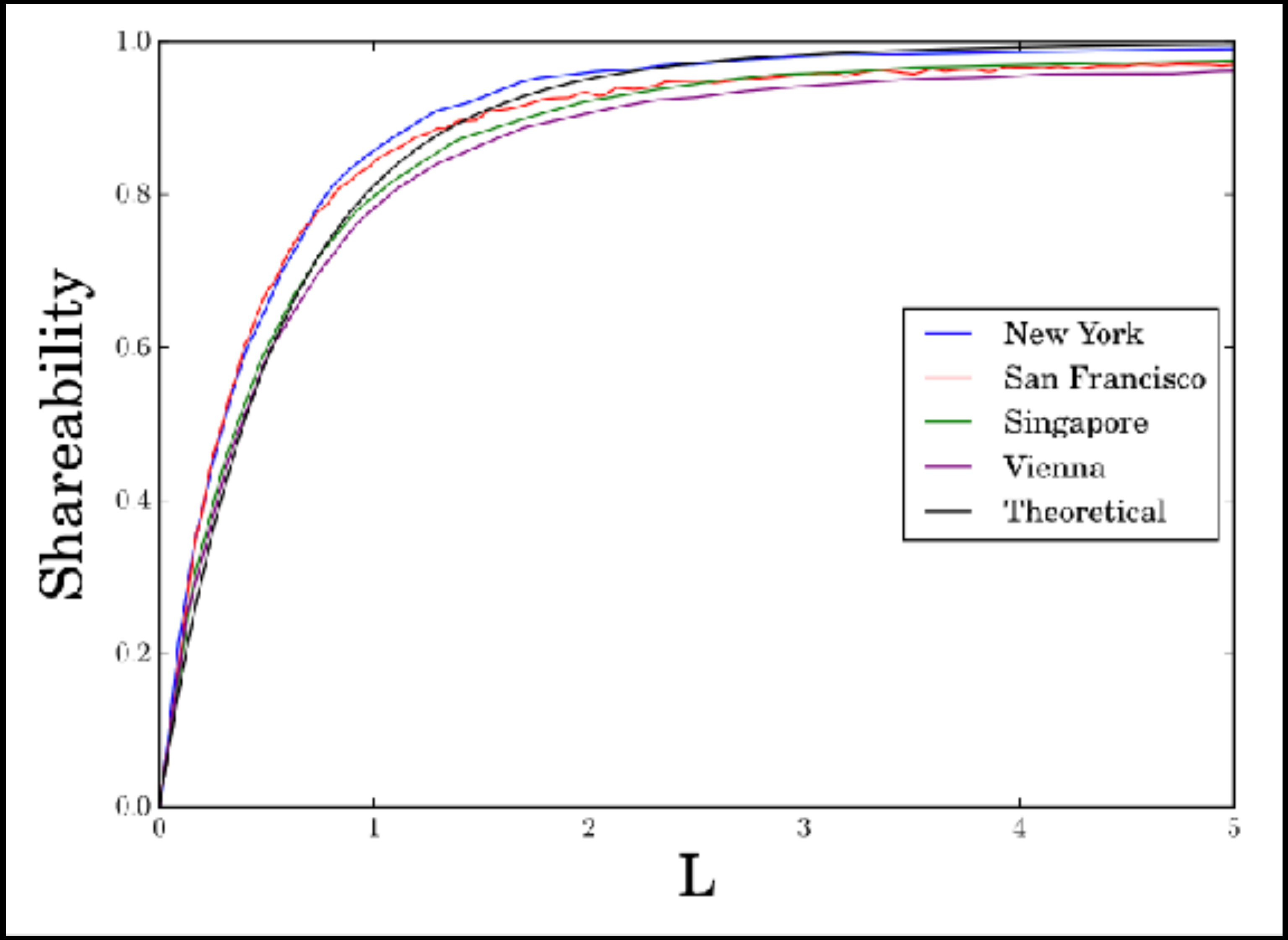
Vienna 5,000 cabs

4,002/km²

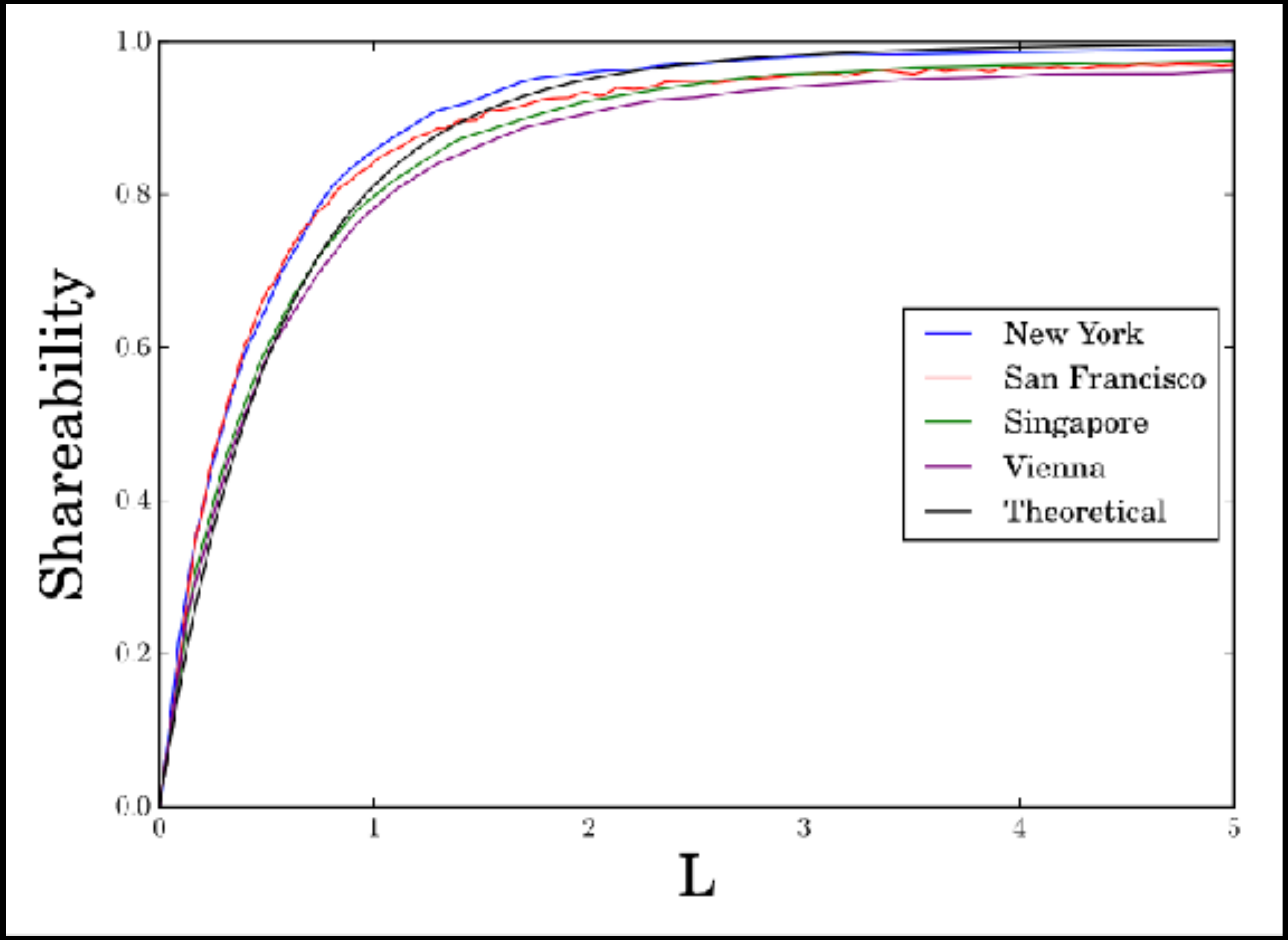
Different cities have similar shareability curves



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

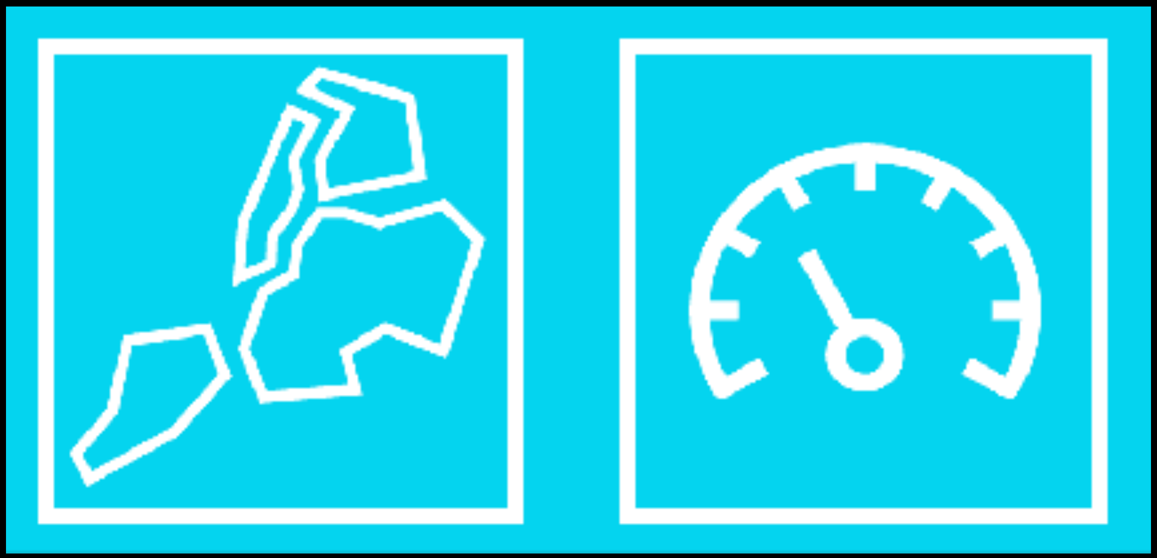
$$L = \lambda \Delta^3 \frac{v^2}{|\Omega|}$$

City area

Rescaling through L

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

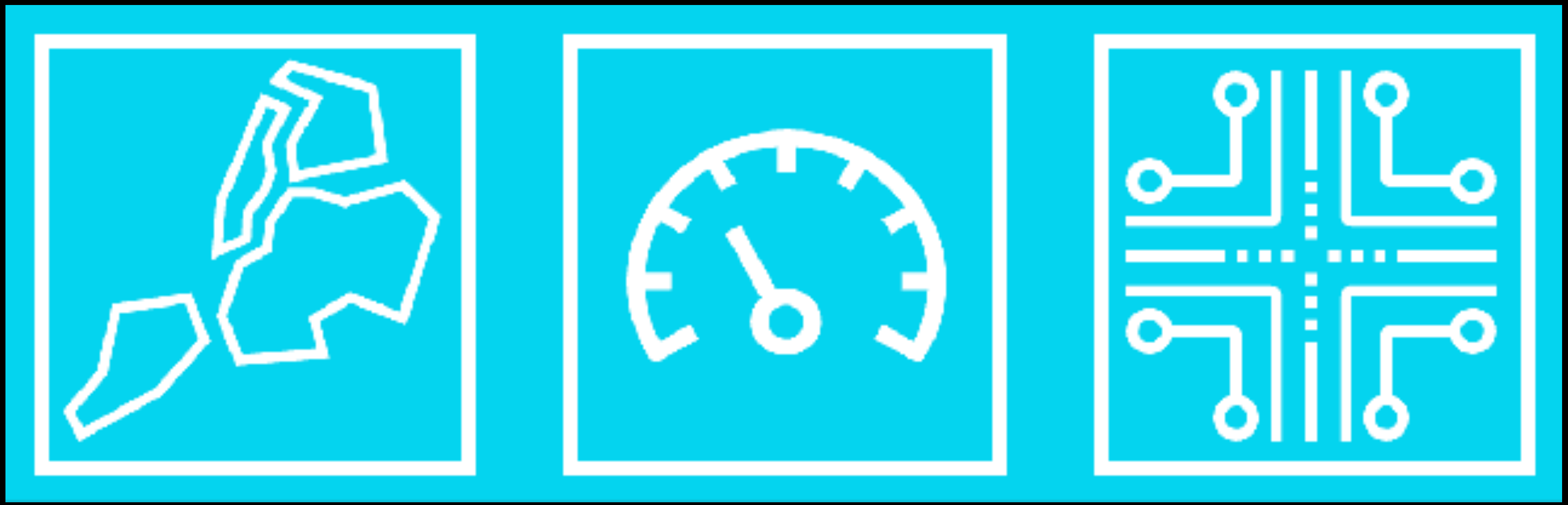
Avg speed



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

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

Trip density λ [T^{-1}] $\nearrow$

Max delay Δ [T] $\nearrow$

Avg speed v [L / T] $\nearrow$

City area $|\Omega|$ [L^2] $\searrow$

L is a ratio between two timescales

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

Tolerable delay time

vs

Expected waiting time

$$\frac{\Delta}{t_{\mathrm{wait}}}$$

L is a ratio between two timescales

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

Tolerable delay time $\quad \Delta$

vs $\quad \overline{}$

Expected waiting time $\quad t_{\mathrm{wait}} = \frac{1}{\lambda} \times \frac{|\Omega|}{(v\Delta)^2}$

L is a ratio between two timescales

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

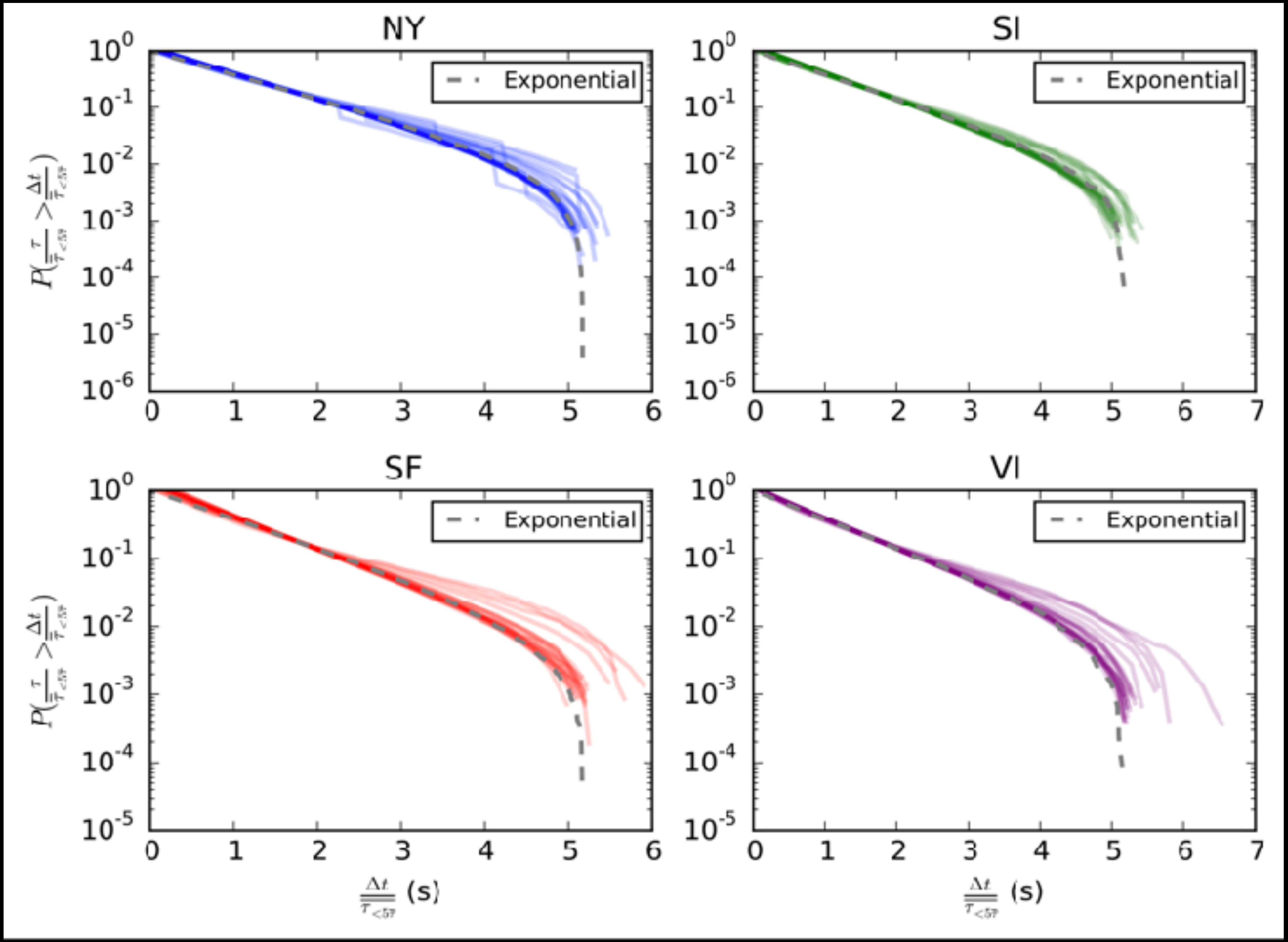
Characteristic time for new trip

Tolerable delay time $\qquad \Delta$

vs $\qquad \overline{}$

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

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

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

Characteristic scale of a vicinity

Sure, cities are different…

…but ride-sharing works well almost everywhere

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

<u>Research</u>

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

<u>Visualizations</u>

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

Michael Szell
@mszll
michael.szell@gmail.com
http://michael.szell.net

Hungarian Acad Sci I Northeastern University I Central European University