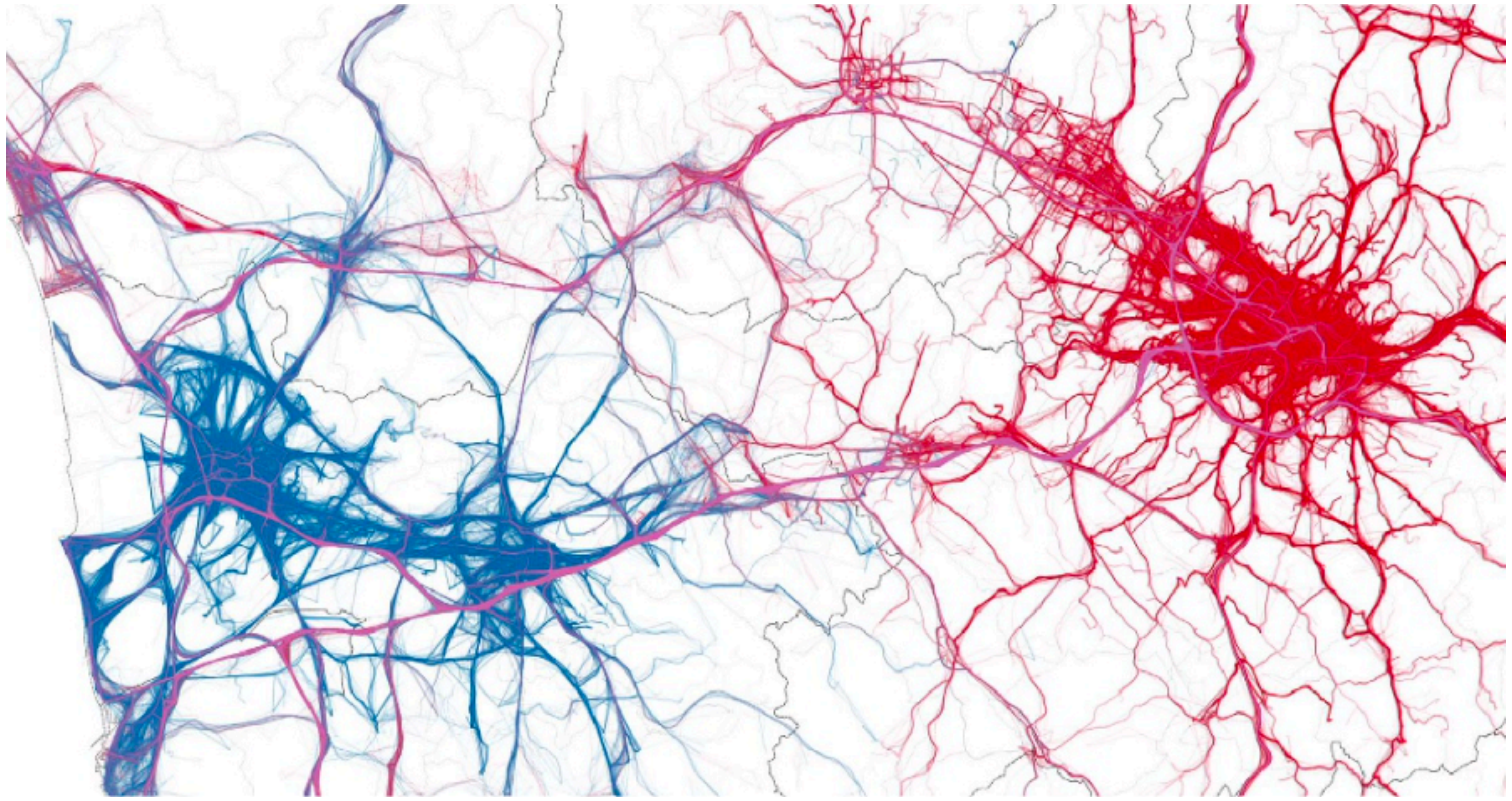
Lecture 12: Mobility patterns and predictability

Instructor: Michael Szell

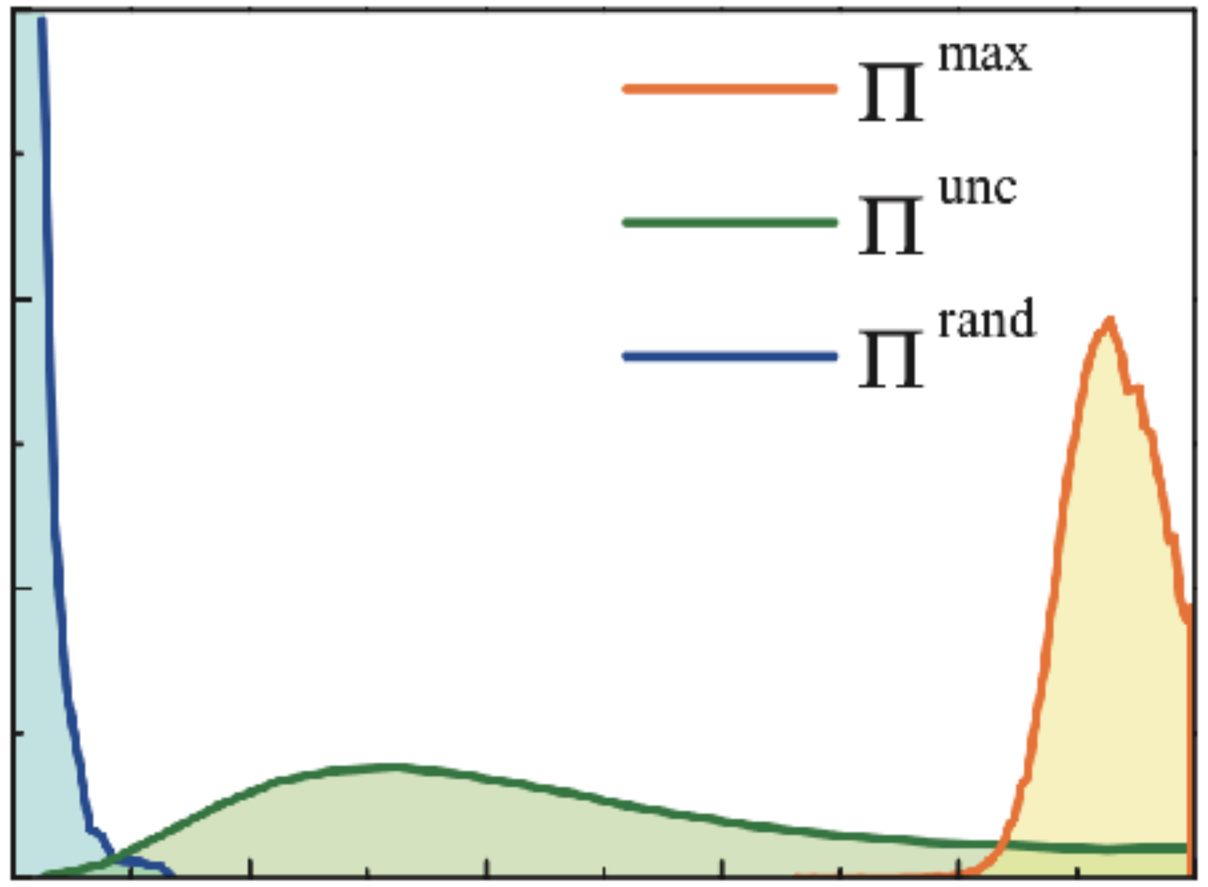
Apr 21, 2022



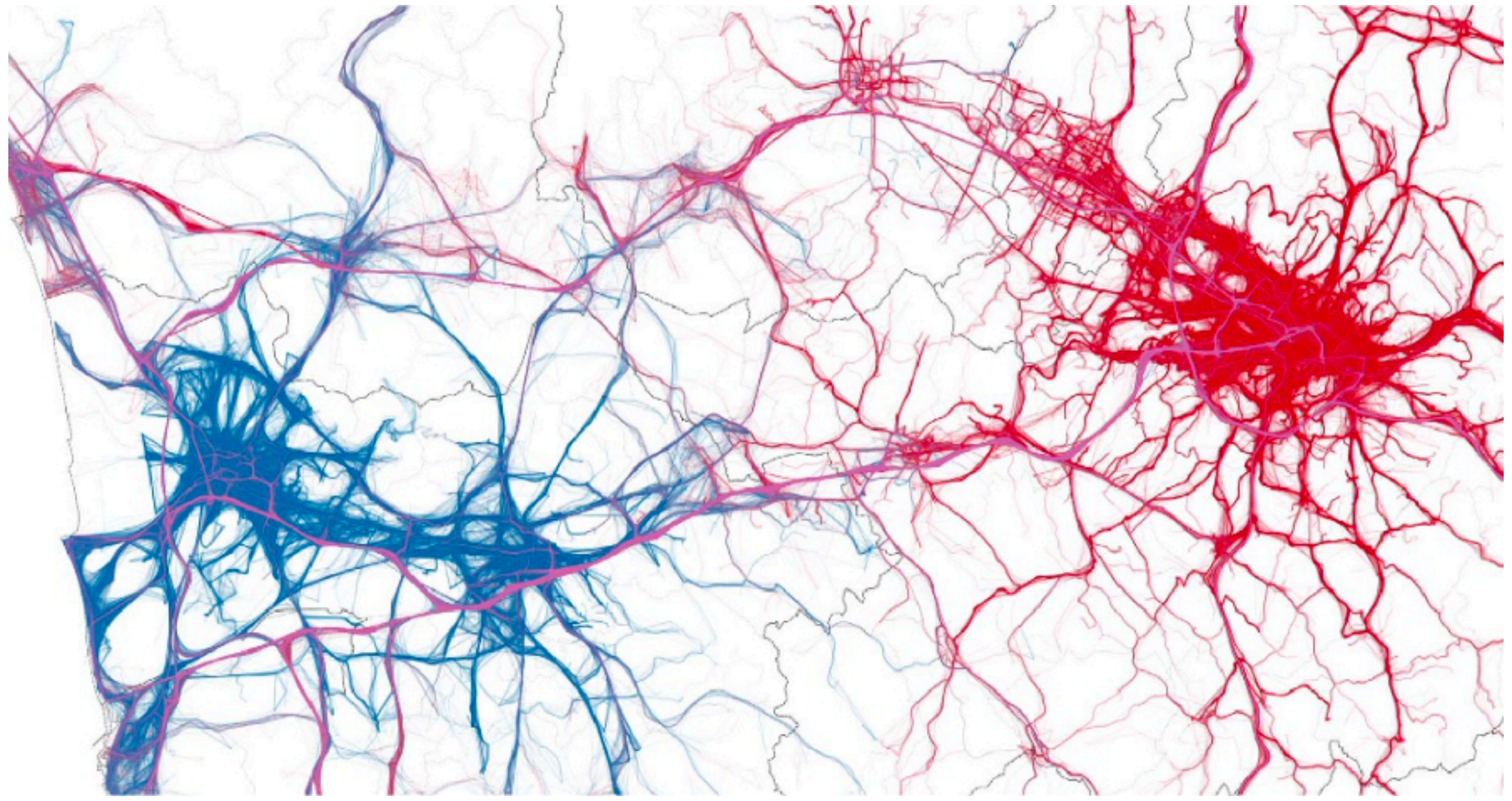
IT UNIVERSITY OF COPENHAGEN

Today you will learn about individual mobility patterns

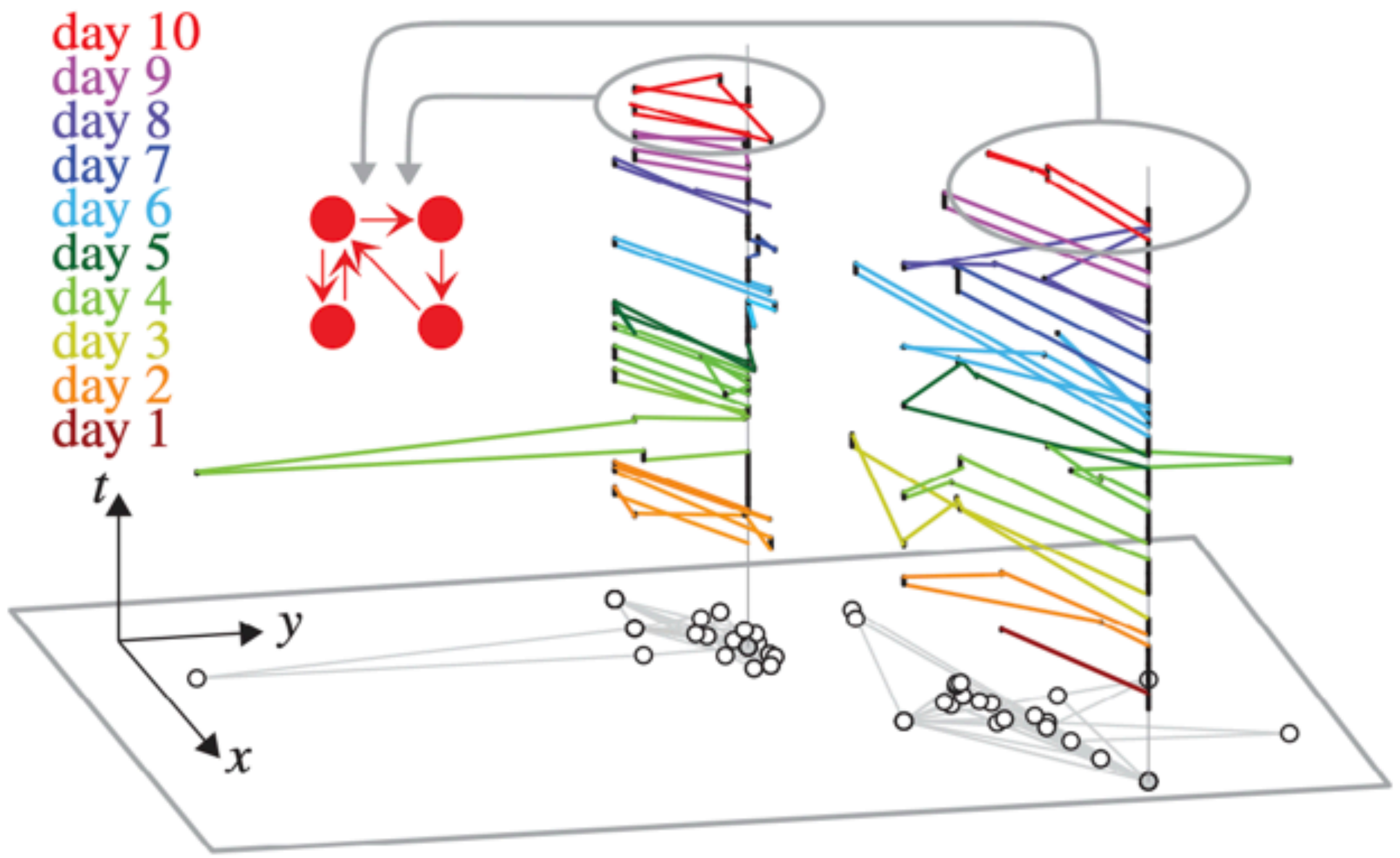
Predictability and privacy



Individual mobility models



Mobility motifs, high-res data



You can study mobility=human movements at different scales

Individual

Single-scale

Pedestrian movements
Air transport
Sea networks

Multi-scale

Intra/Inter-urban mobility
Epidemic spreading

Virtual scale

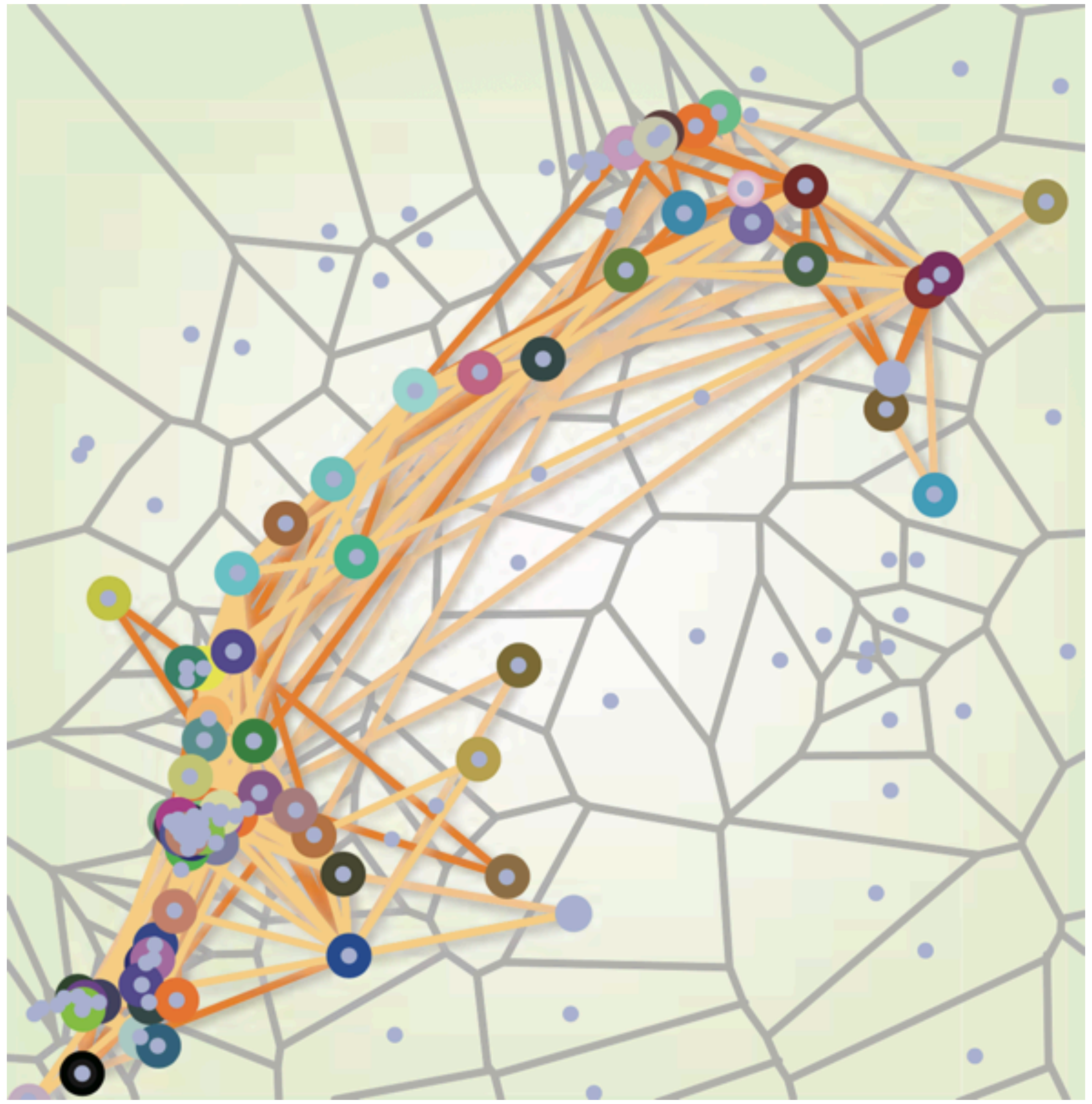
Population / Aggregate

Migration

Spatial Interaction

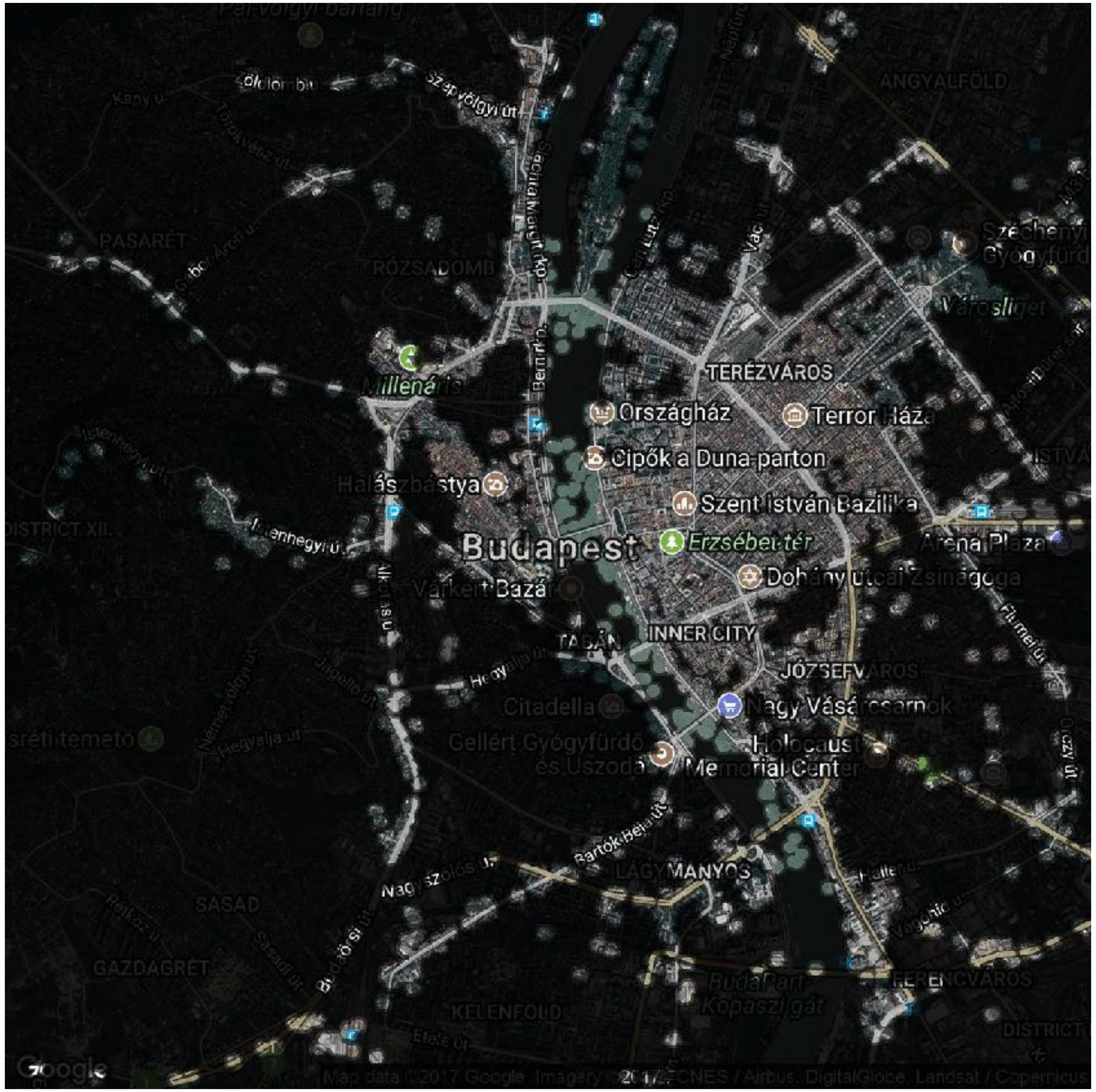
Predictability of individual trajectories

We have the option to visit many places in our lives



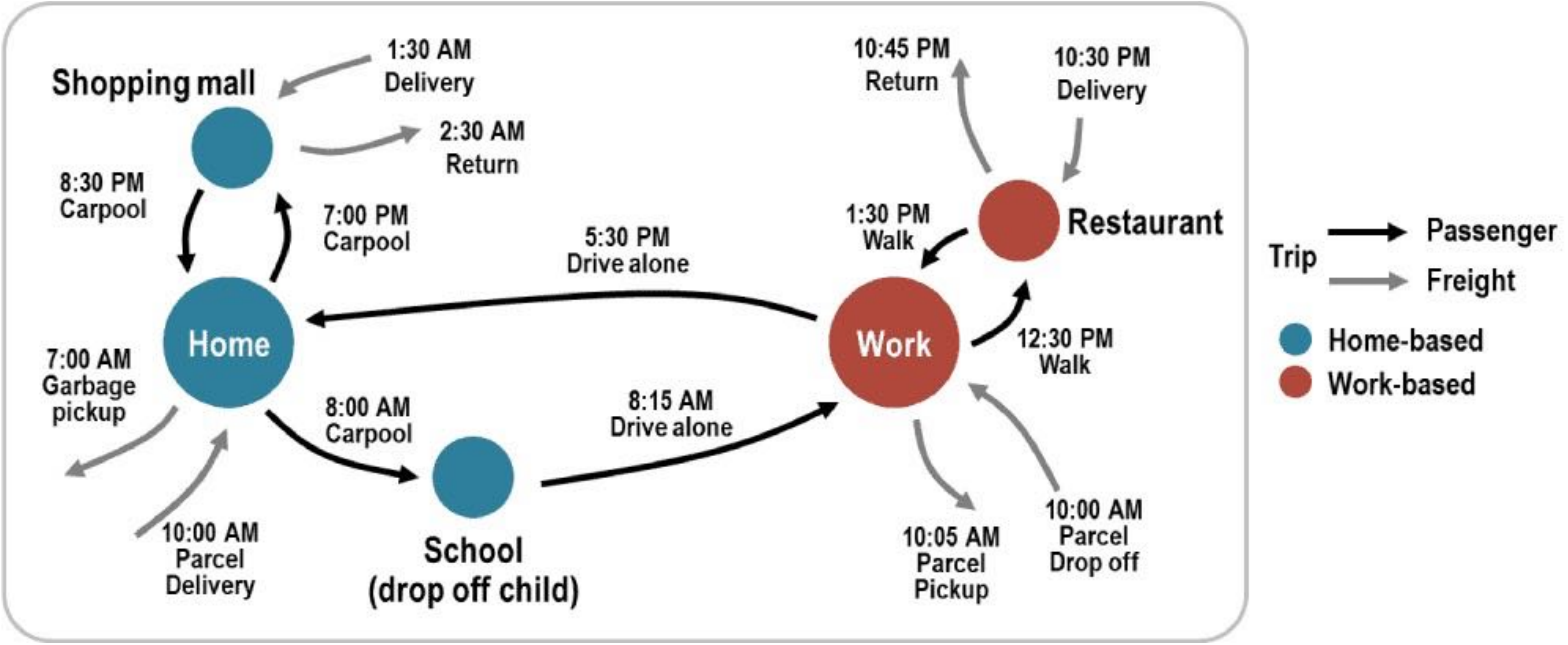
Are we actually doing it?

My own visits after living in Budapest for 2 years



https://github.com/kindofdoon/fog_of_war

Anecdotally, we have very regular lives

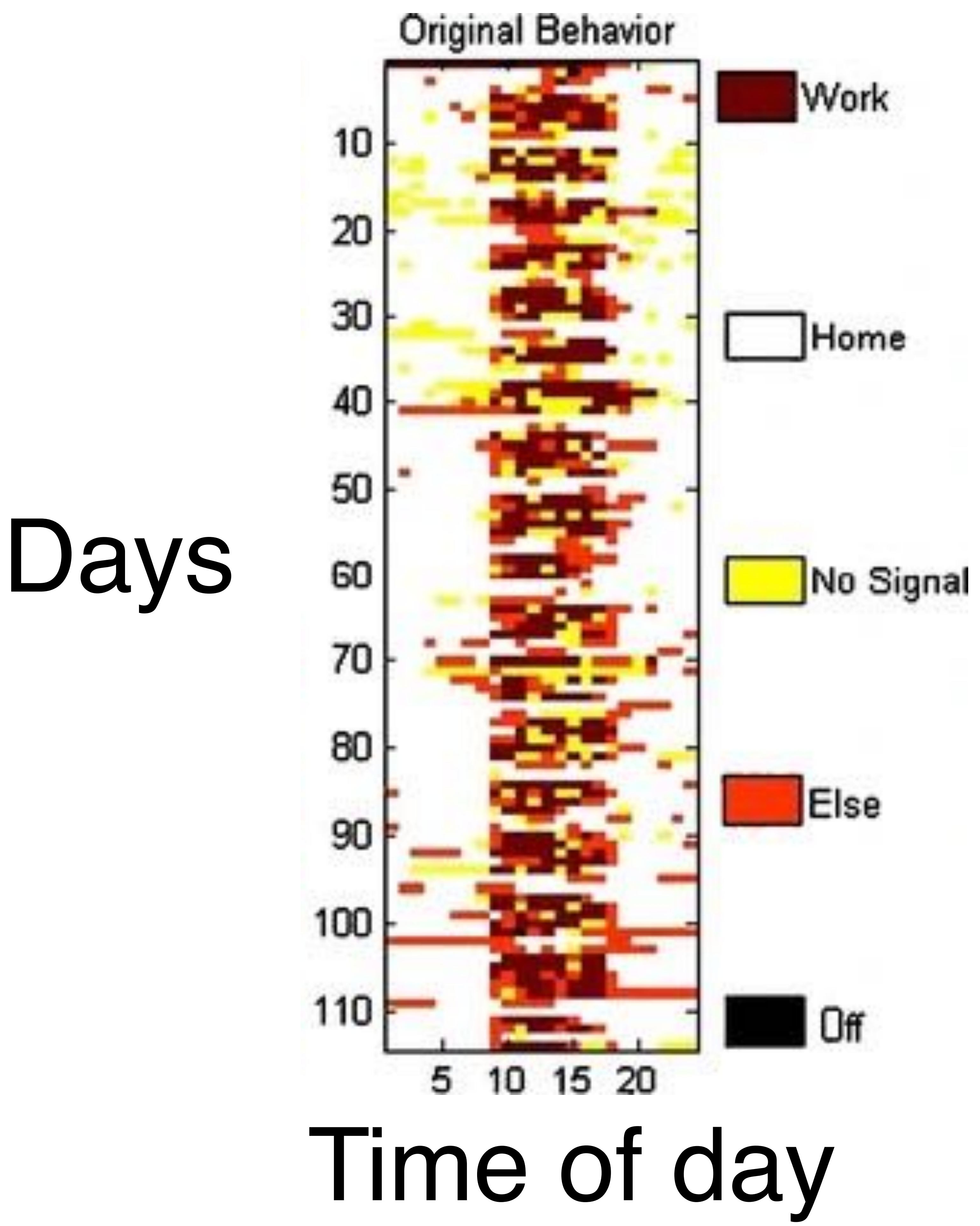


How boring is our life, and what does that mean for predicting it?

https://transportgeography.org/contents/chapter8/urban-mobility/activity-space-adult/

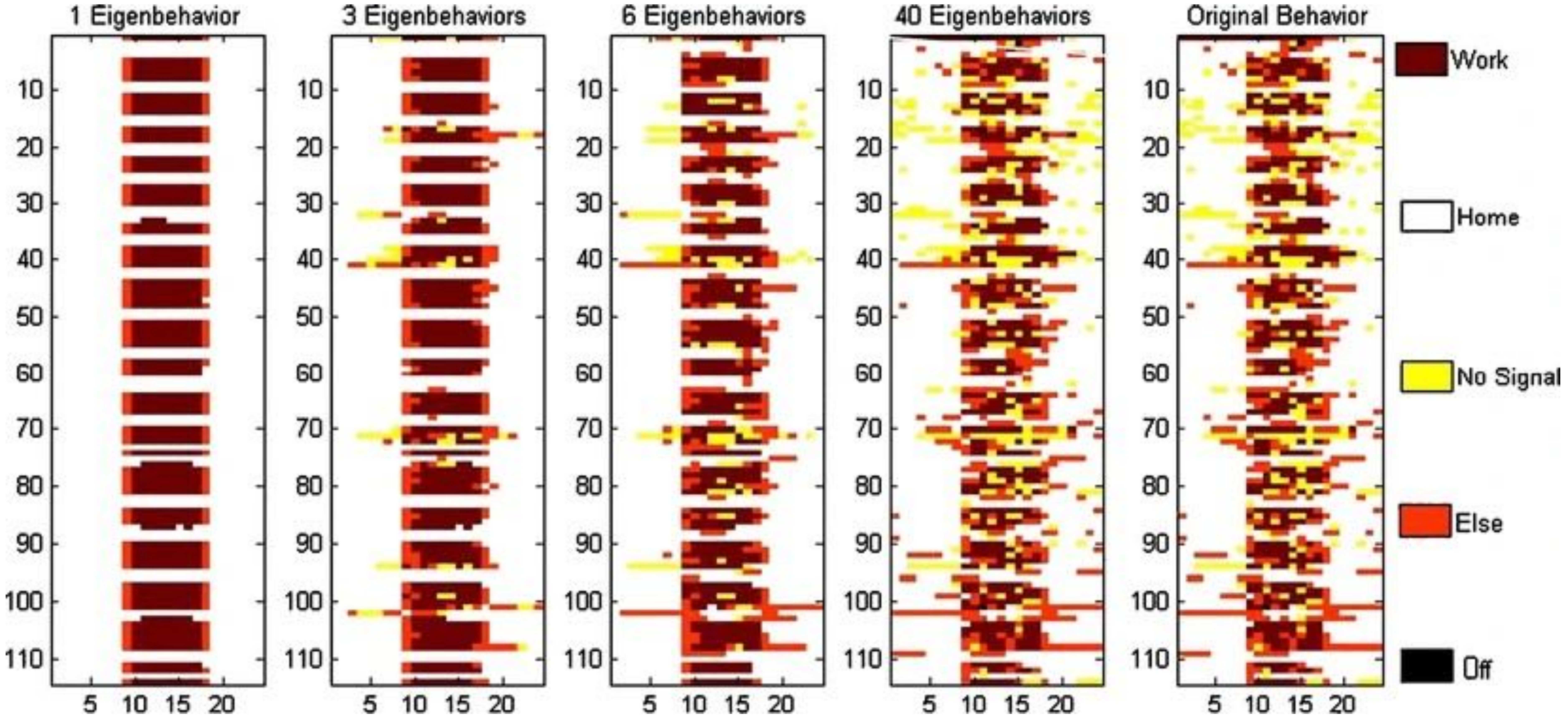
Locations of 100 students at MIT can be well simplified

2009



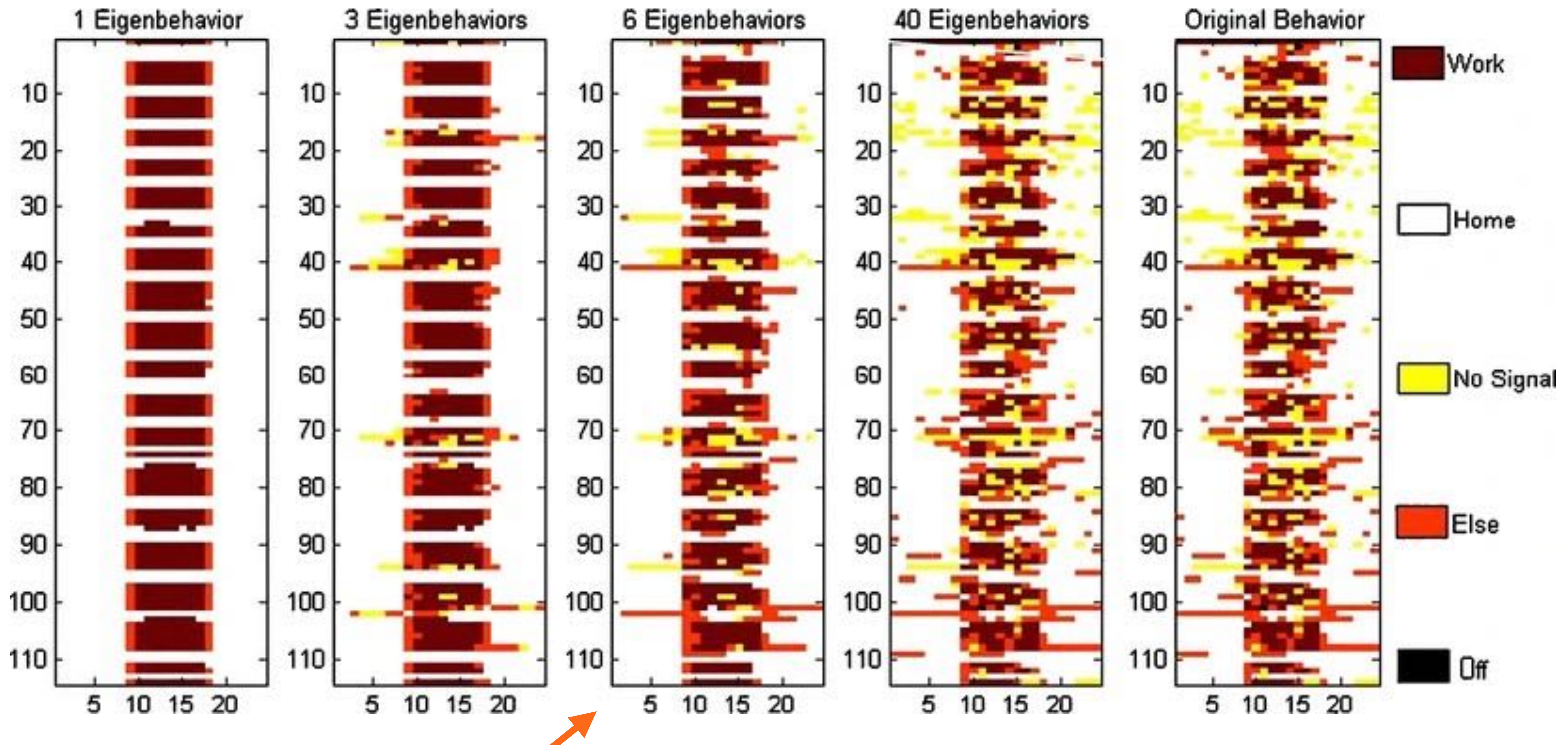
https://link.springer.com/article/10.1007/s00265-009-0739-0

Locations of 100 students at MIT can be well simplified



https://link.springer.com/article/10.1007/s00265-009-0739-0

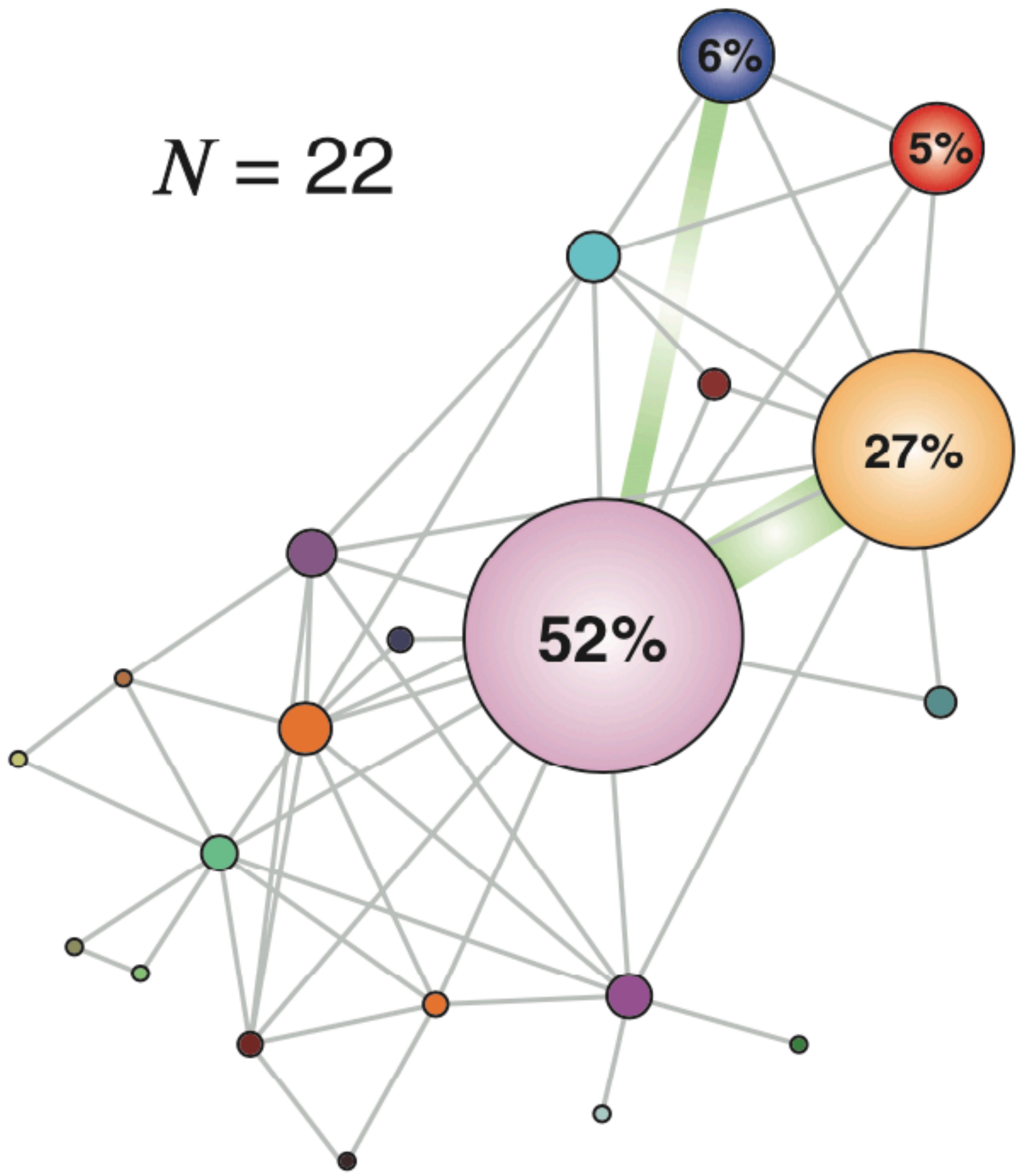
Locations of 100 students at MIT can be well simplified



This simplification predicts lab students with 96% accuracy (business students with 90%)

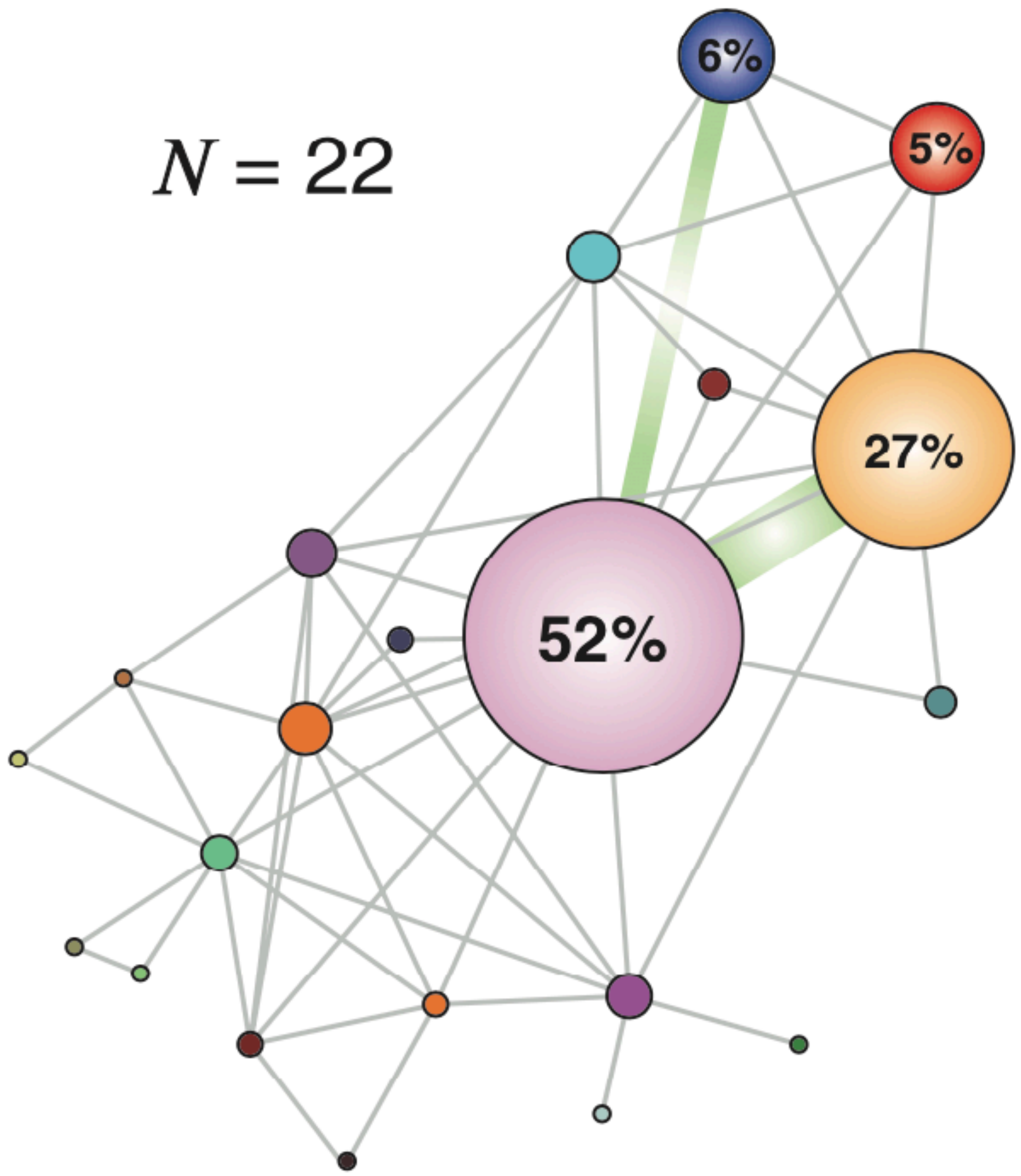
https://link.springer.com/article/10.1007/s00265-009-0739-0

For most people, most visits are to only a few places

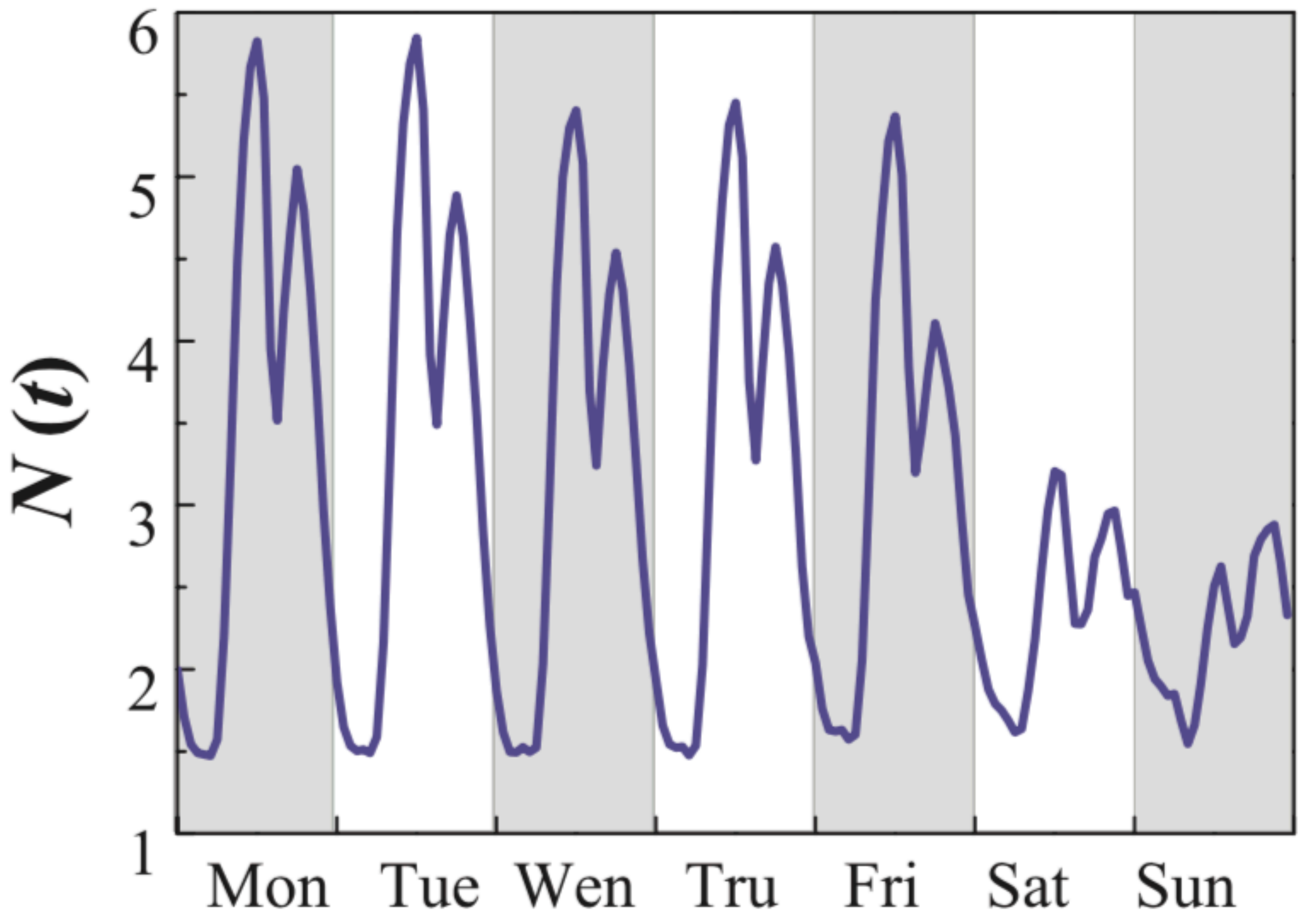


https://www.science.org/doi/abs/10.1126/science.1177170

For most people, most visits are very regular



Number of locations in each hour



How can we quantify how easy it is to predict a next location?

Entropy

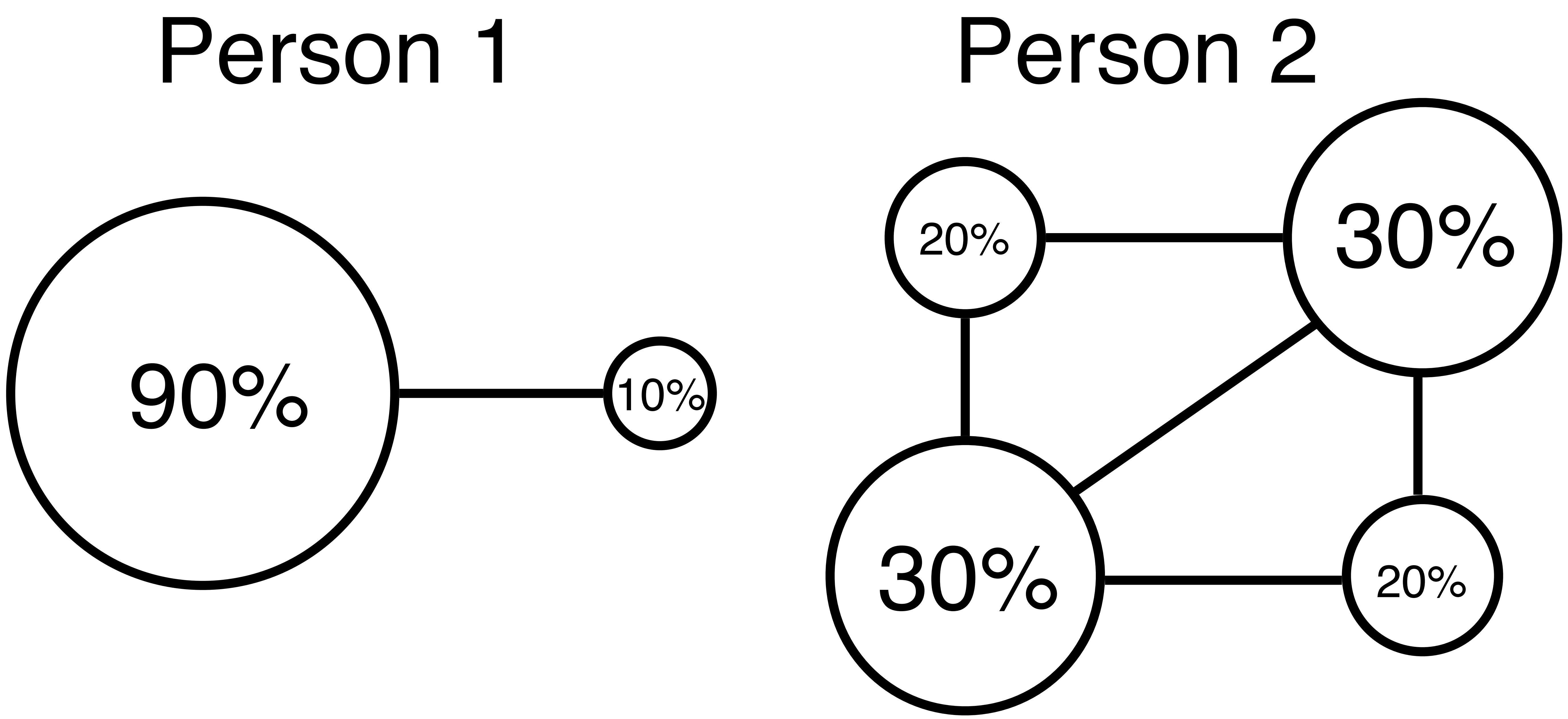
A measure for uncertainty, or randomness.

Random entropy S^{rand}: How many unique places were visited?

$$S_i^{\text{rand}} = \log_2 N_i$$

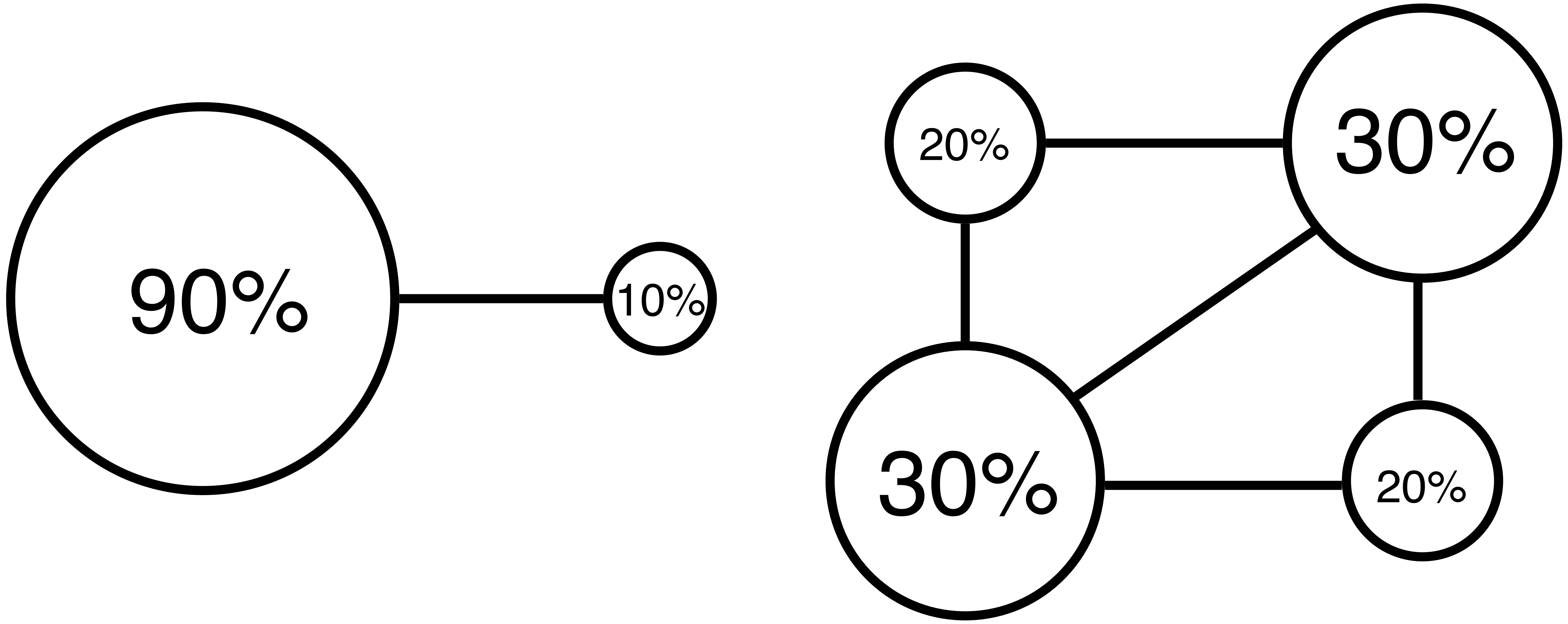
Random entropy S^{rand}: How many unique places were visited?

$$S_i^{\text{rand}} = \log_2 N_i$$



Random entropy S^{rand}: How many unique places were visited?

$$S_i^{\text{rand}} = \log_2 N_i$$

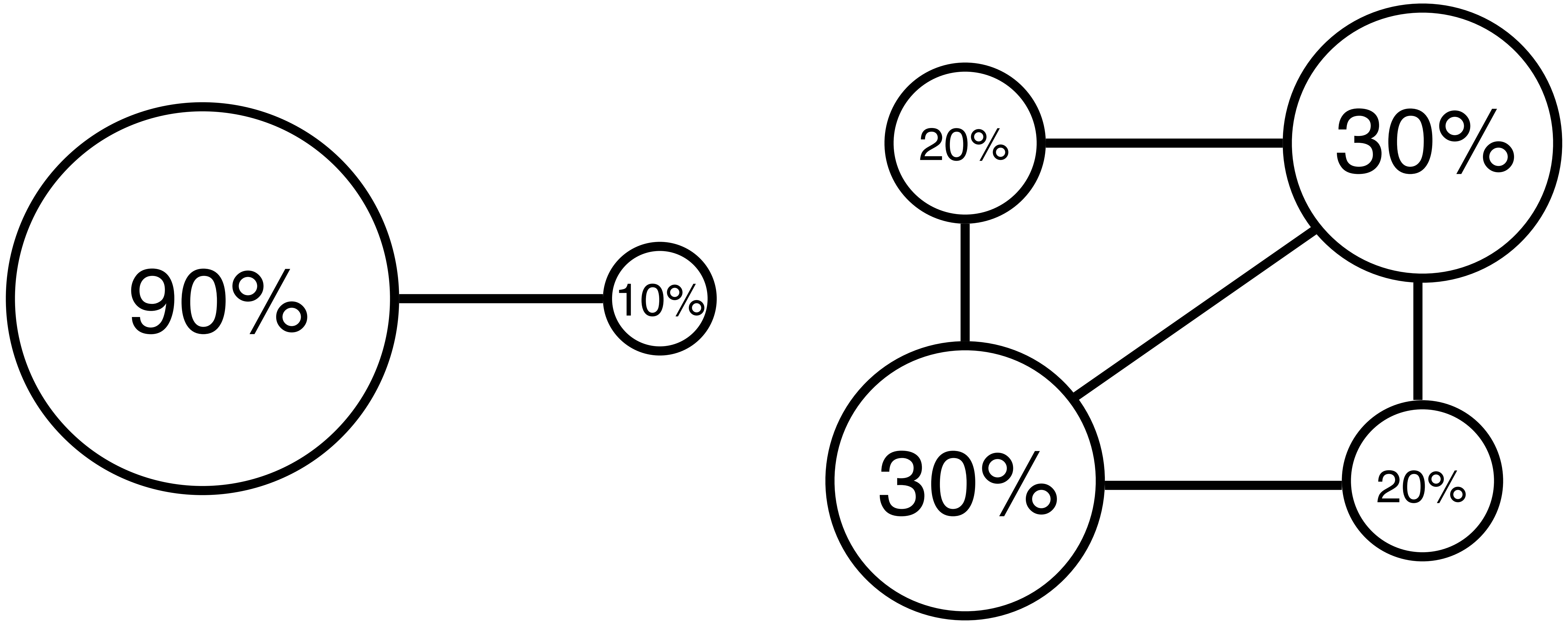
	Person 1	Person 2
		
S_i^{rand}	1	2

S^{unc}: How heterogeneous were the visitations in space?

Temporal-uncorrelated entropy

$$S_i^{\text{unc}} = -\sum_{j=1}^{N_i} p_i(j) \log_2 p_i(j)$$

Person 1 | Person 2

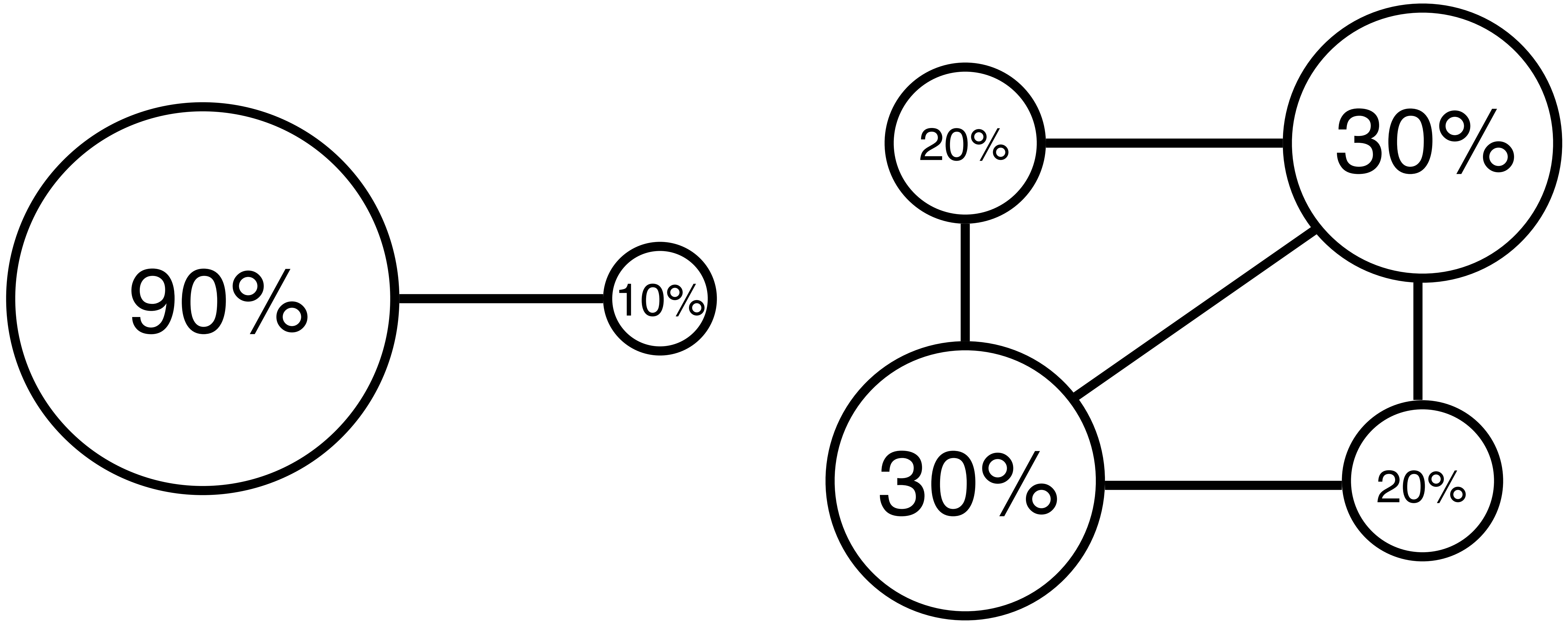


S_i^{rand}	1	2

S^{unc}: How heterogeneous were the visitations in space?

Temporal-uncorrelated entropy

$$S_i^{\mathrm{unc}} = -\sum_{j=1}^{N_i} p_i(j)\log_2 p_i(j)$$

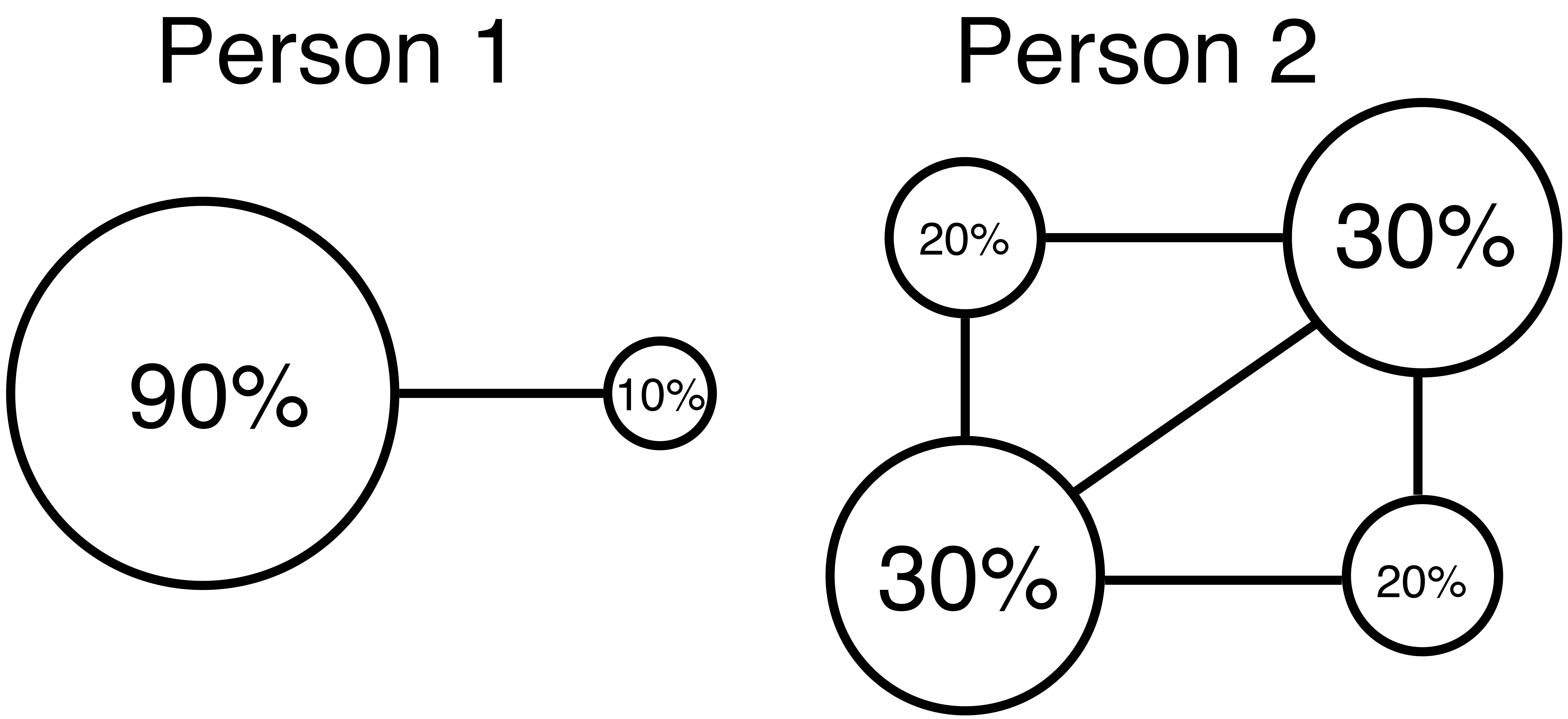


	Person 1	Person 2
S_i^{rand}	1	2
S_i^{unc}	0.469	1.971

S^{unc}: How heterogeneous were the visitations in space?

Temporal-uncorrelated entropy

$$S_i^{\text{unc}} = -\sum_{j=1}^{N_i} p_i(j)\log_2 p_i(j)$$

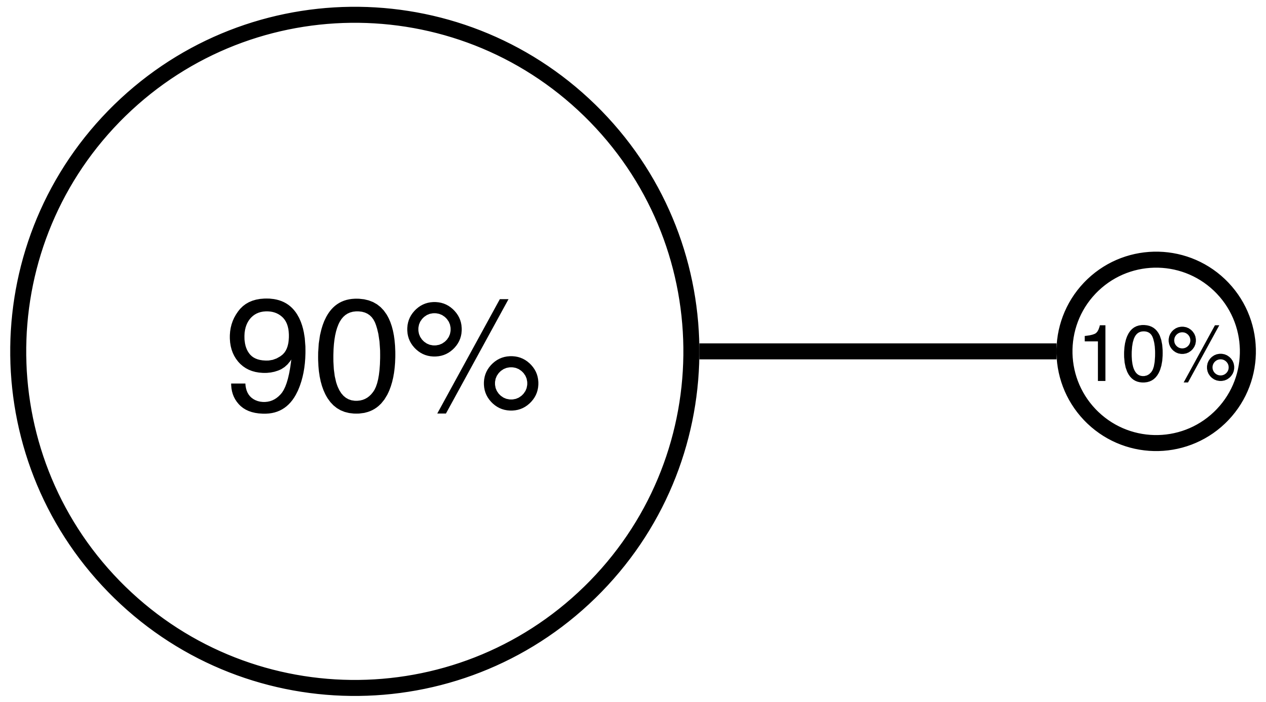


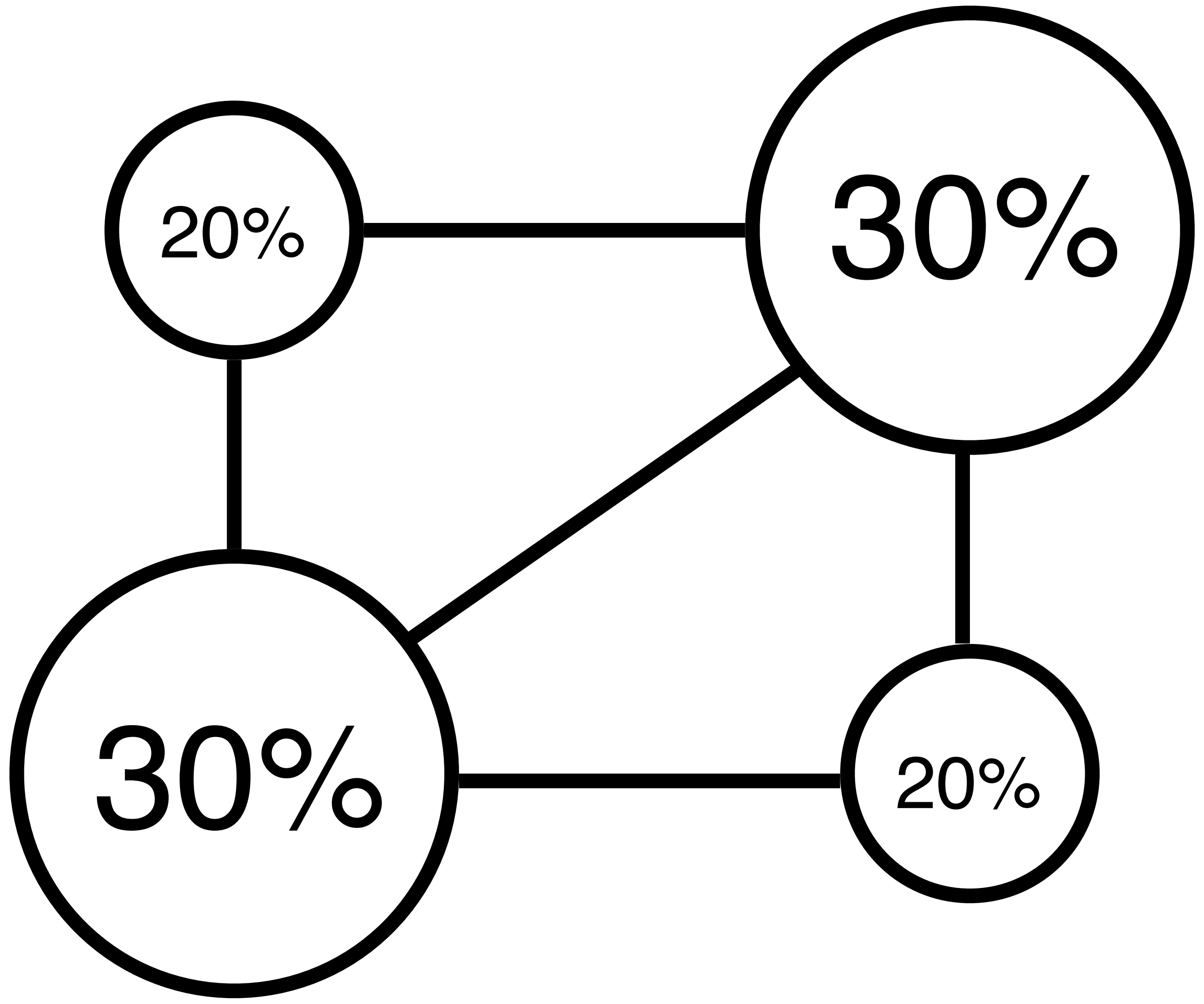
	Person 1	Person 2
S_i^{rand}	1	2
S_i^{unc}	0.469	1.971

We get the maximum value S^{unc}=S^{rand} when all locations are visited with equal probability

Entropy S: How heterogeneous were the visitations in space and time?

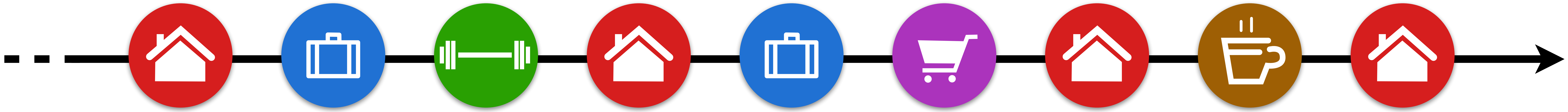
$$S_i = -\sum_{T_i' \subset T_i} P(T_i') \log_2 P(T_i')$$





	Person 1	Person 2
S_i^{rand}	1	2
S_i^{unc}	0.469	1.971

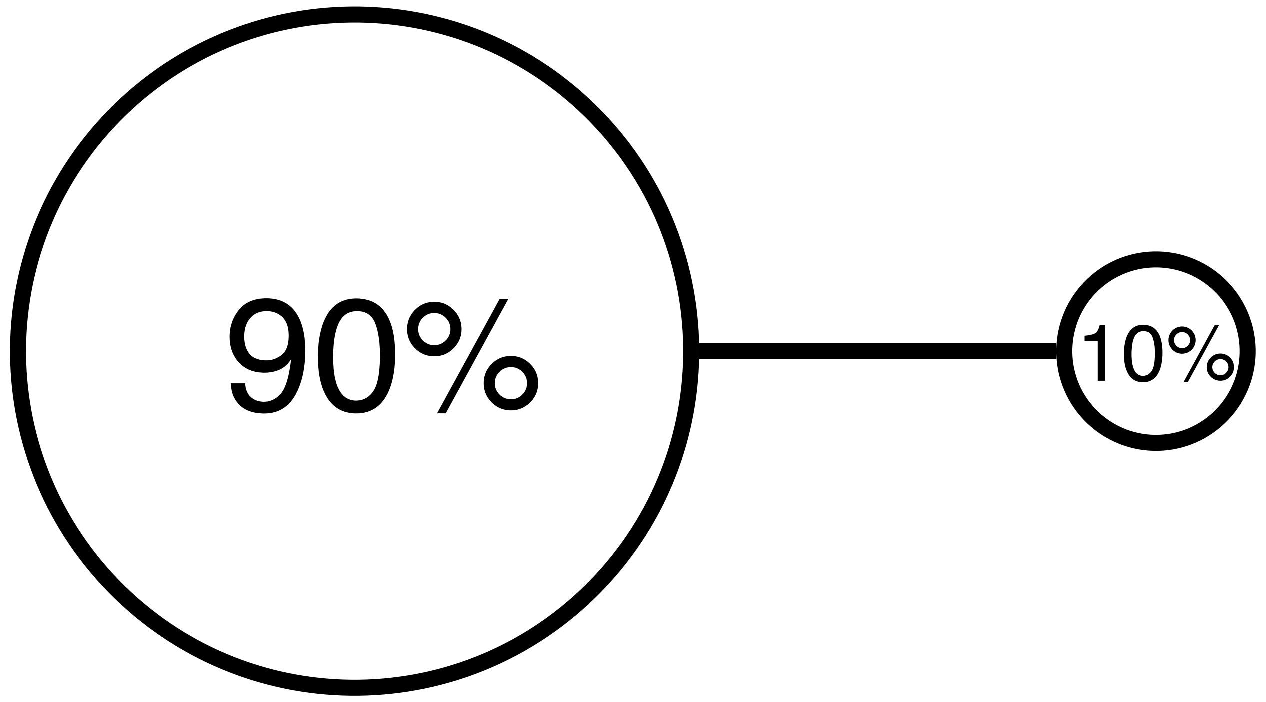
T_i



Entropy S: How heterogeneous were the visitations in space and time?

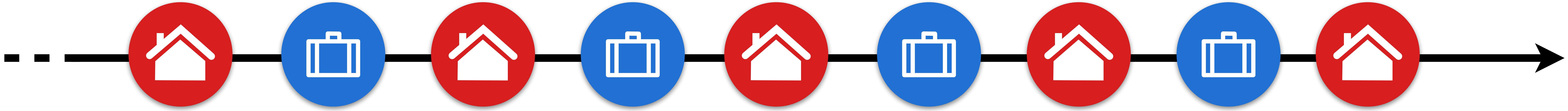
Person 1

$$S_i = -\sum_{T_i' \subset T_i} P(T_i') \log_2 P(T_i')$$



S_i^{rand}	1
S_i^{unc}	0.469
S_i	$\ll 0.469$

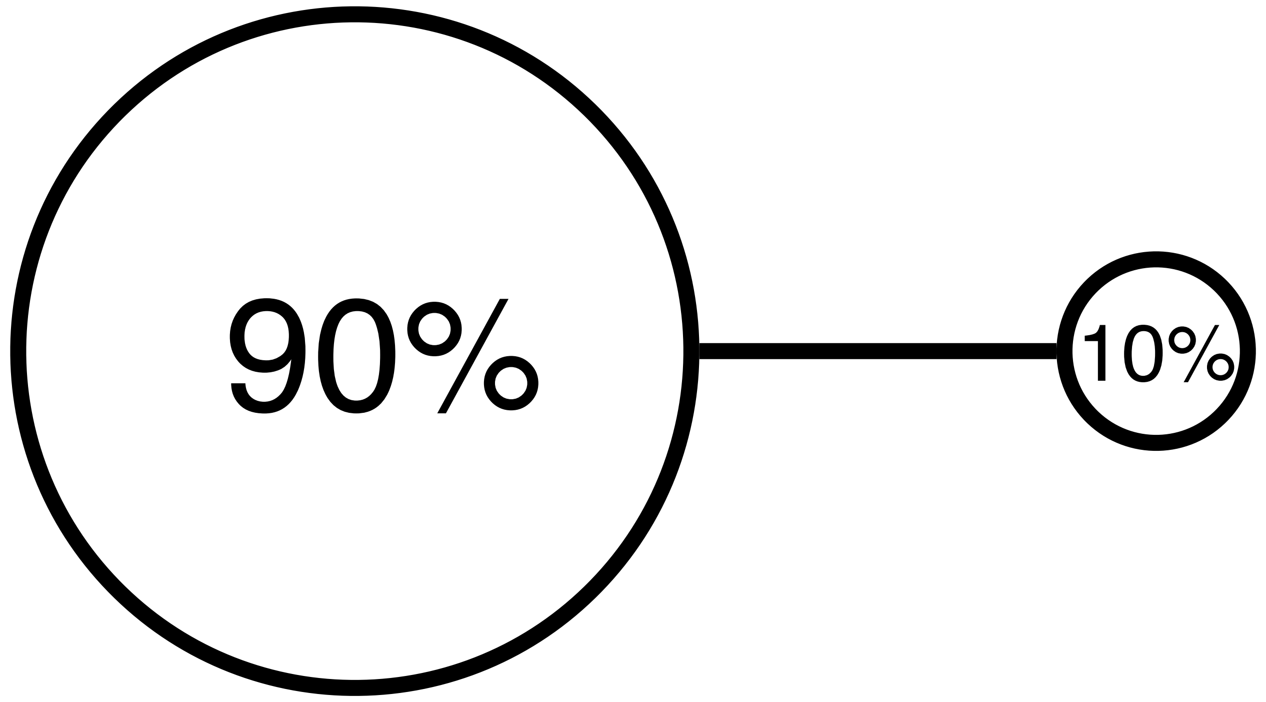
T_1



Entropy S: How heterogeneous were the visitations in space and time?

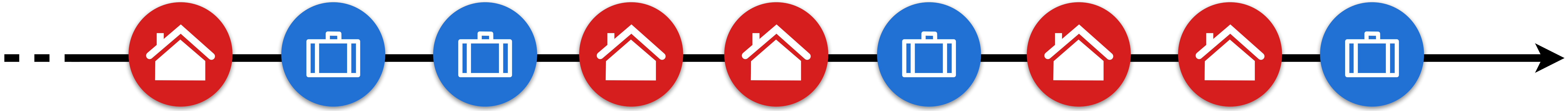
Person 1

$$S_i = -\sum_{T_i' \subset T_i} P(T_i') \log_2 P(T_i')$$



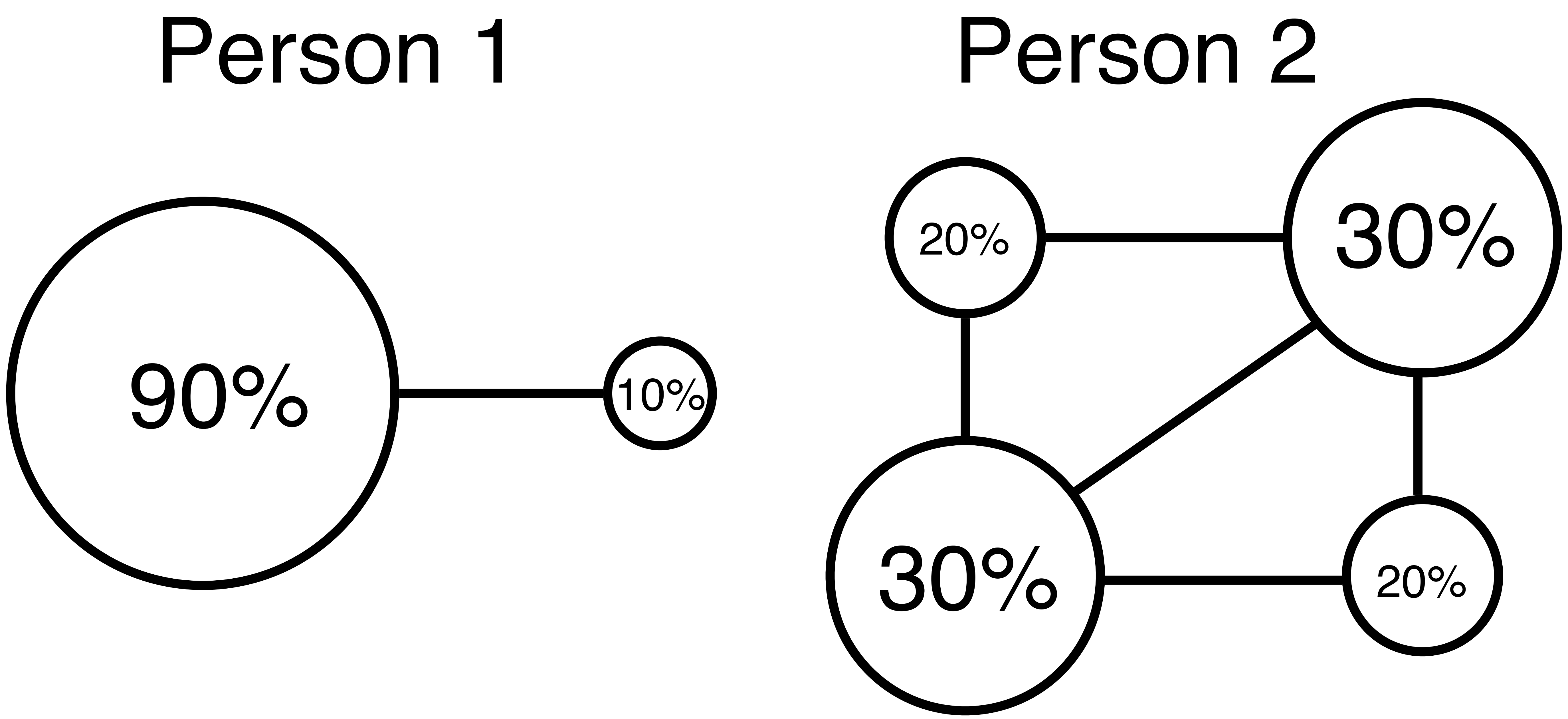
S_i^{rand}	1
S_i^{unc}	0.469
S_i	$\approx$ 0.469

T_1



Entropy S: How heterogeneous were the visitations in space and time?

$$S_i = -\sum_{T_i' \subset T_i} P(T_i') \log_2 P(T_i')$$

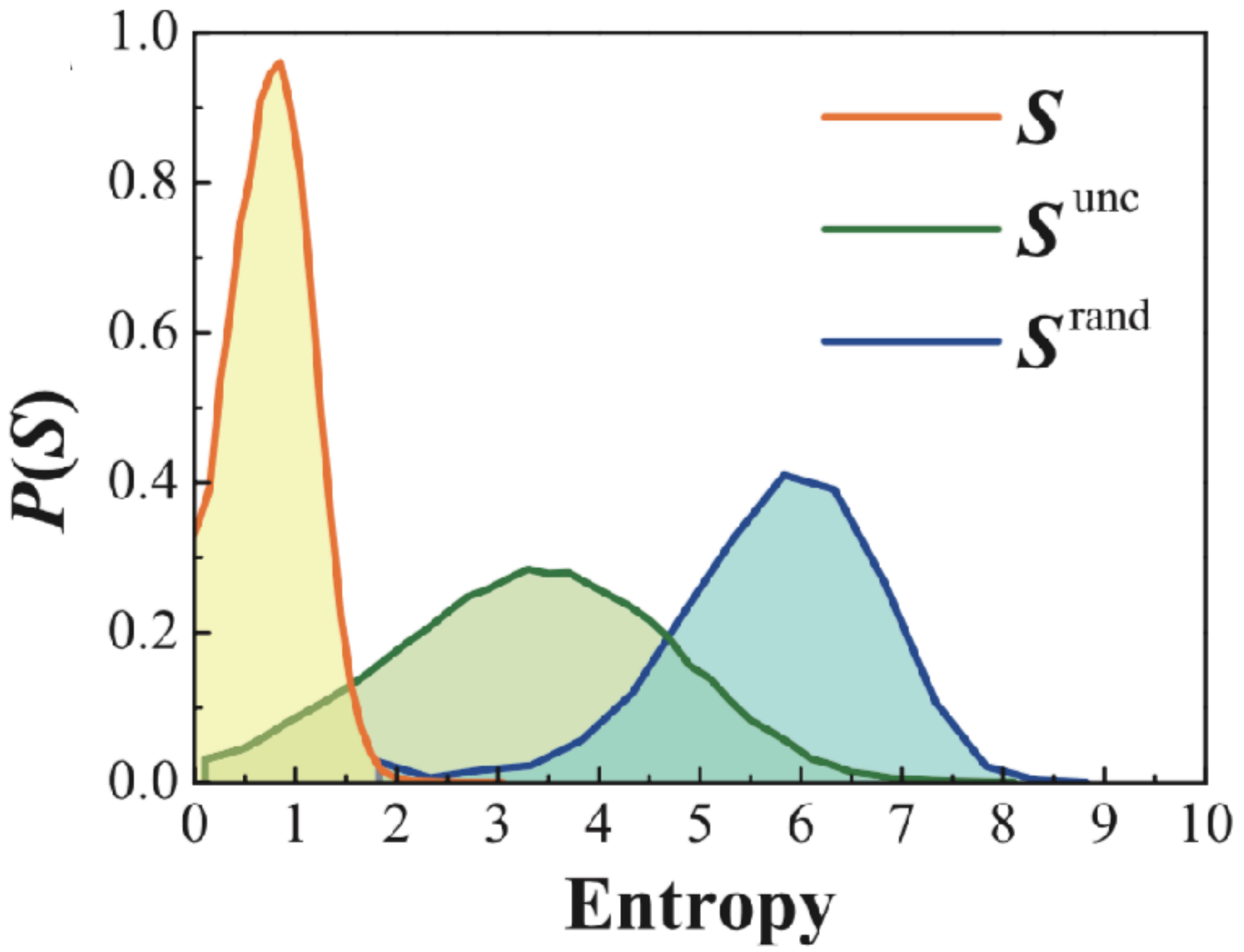


	Person 1	Person 2
S_i^{rand}	1	2
S_i^{unc}	0.469	1.971
S_i	$\leq$ 0.469	$\leq$ 1.971

We get the maximum value S=S^{unc} when the probability of the next location is independent of the current one

Accounting for more information reduces the entropy

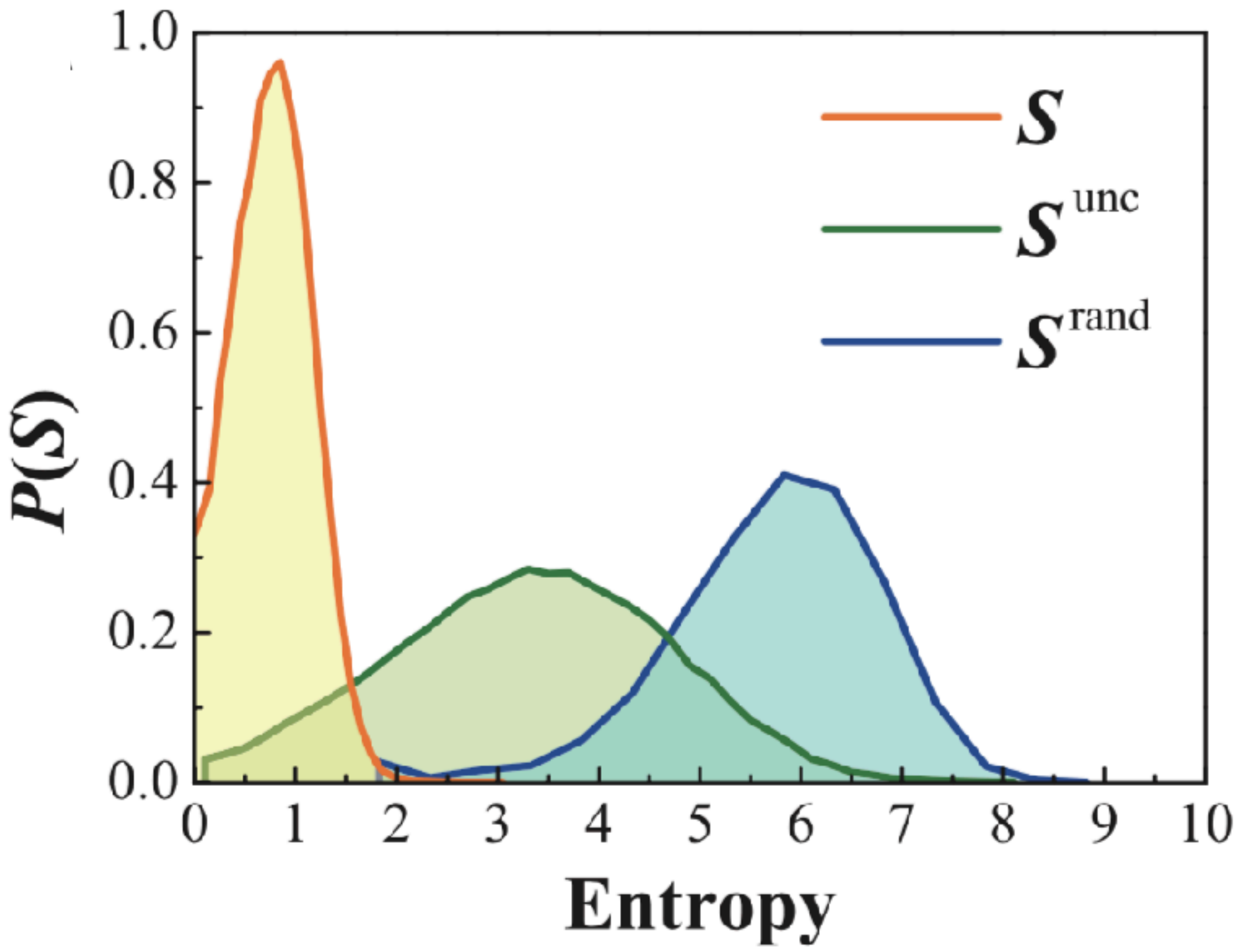
$$S \leq S^{\mathrm{unc}} \leq S^{\mathrm{rand}}$$

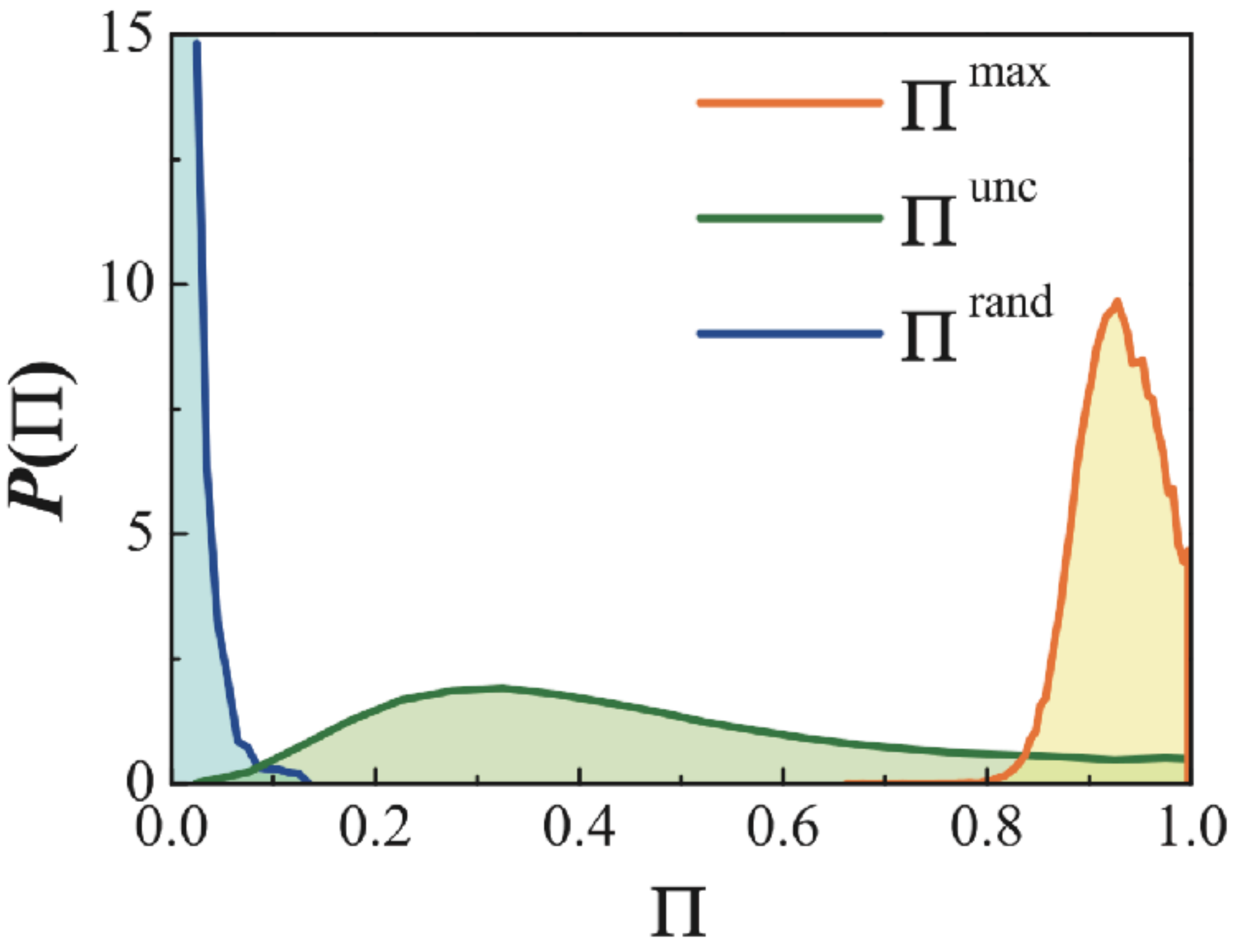


Accounting for more information increases predictability drastically

$$S \leq S^{\mathrm{unc}} \leq S^{\mathrm{rand}}$$

$$\Pi^{\mathrm{rand}} \leq \Pi^{\mathrm{unc}} \leq \Pi$$





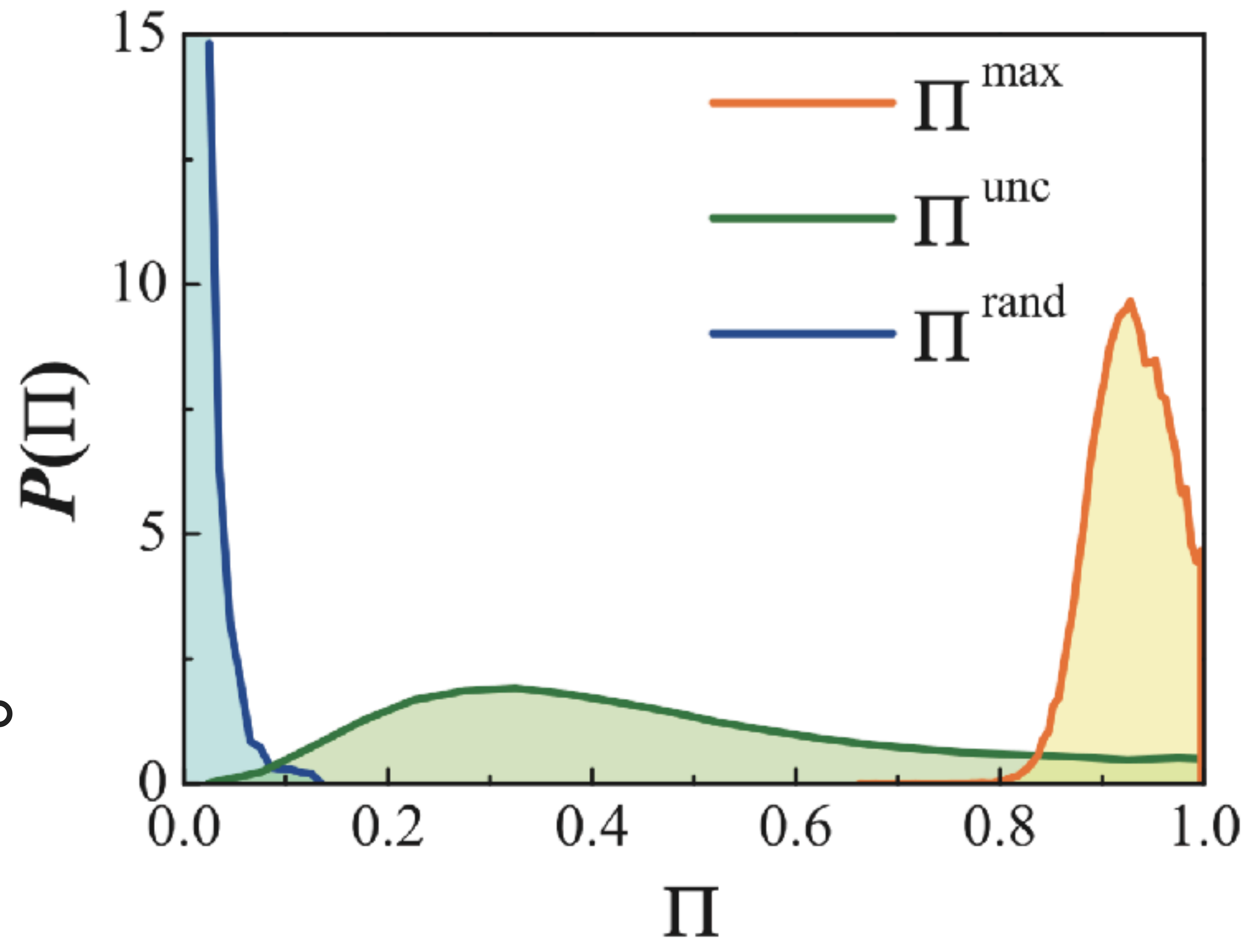
Accounting for more information increases predictability drastically

$$\Pi^{\mathrm{rand}} \leq \Pi^{\mathrm{unc}} \leq \Pi$$

A significant share of predictability is encoded in the temporal order of locations

93% of trips could be predicted

There is nobody with predictability below 80%



scikit mobility has functions for calculating entropy

$$S \leq S^{\mathrm{unc}} \leq S^{\mathrm{rand}}$$

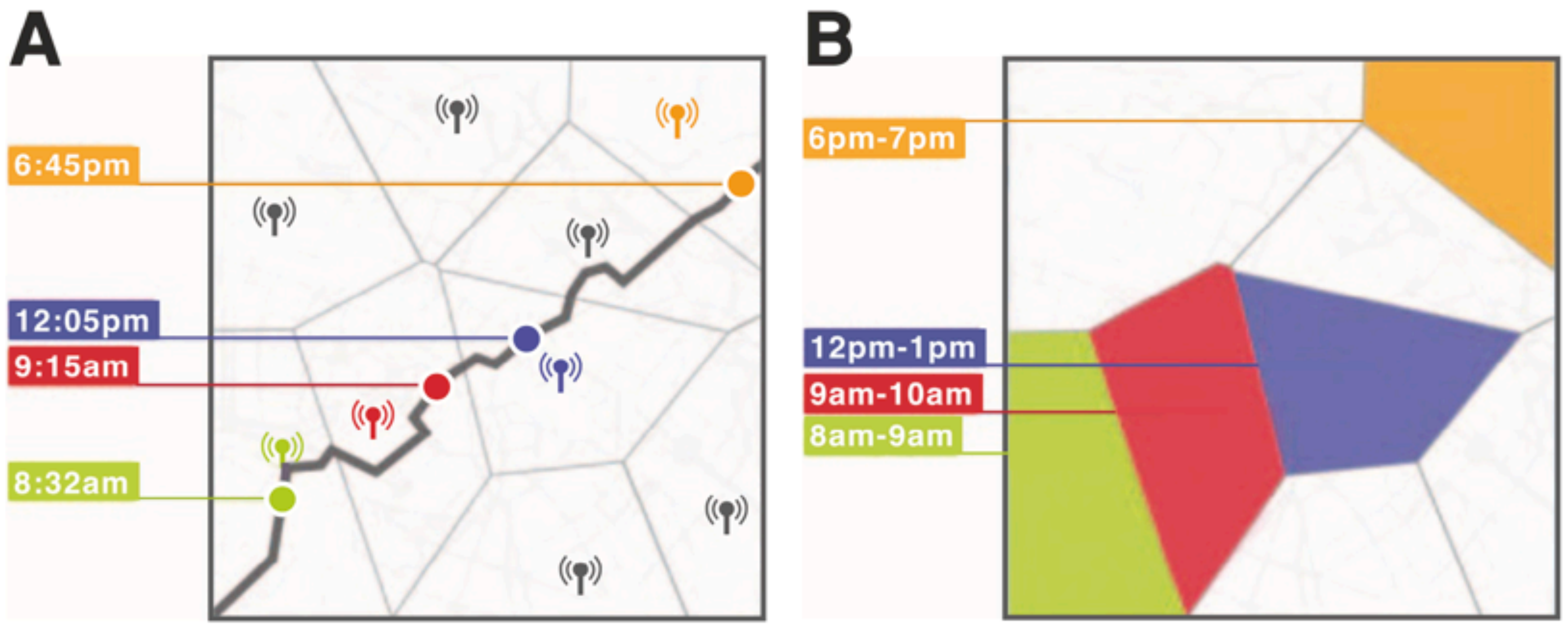
scikit mobility

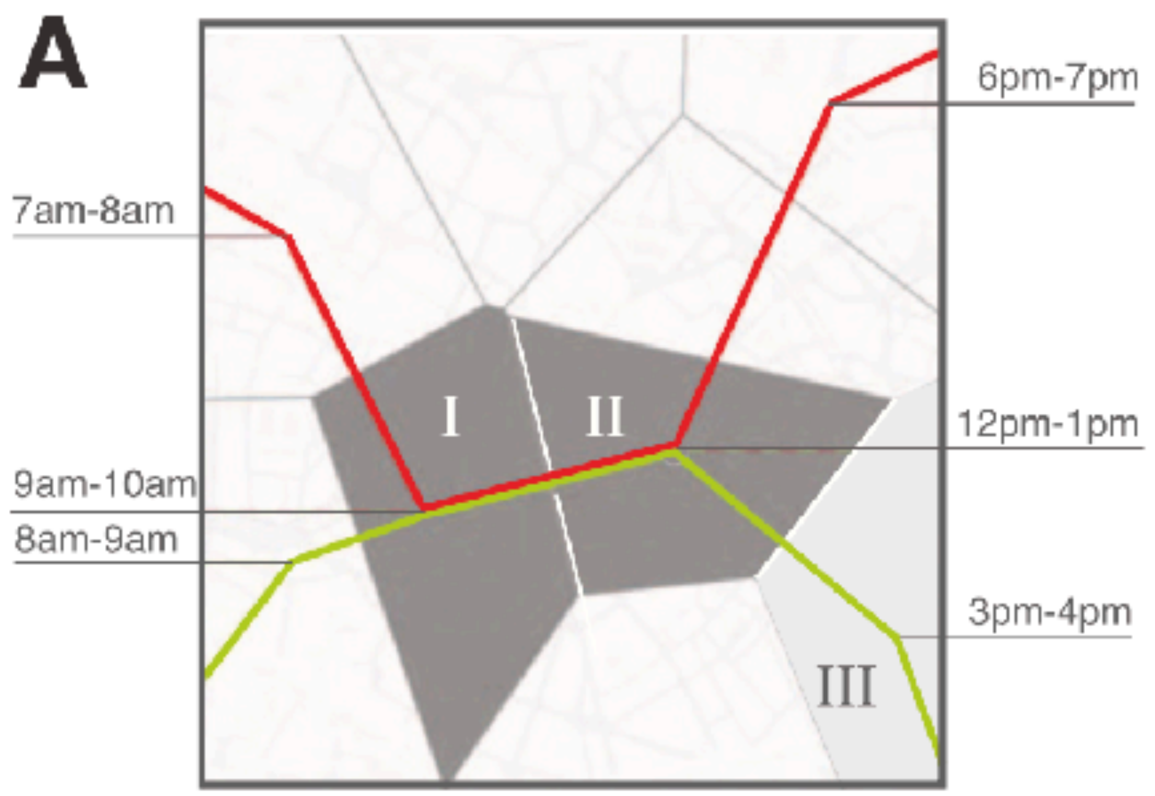
`random_entropy` (traj[, show_progress])	Random entropy.
`uncorrelated_entropy` (traj[, normalize, ...])	Uncorrelated entropy.
`real_entropy` (traj[, show_progress])	Real entropy.

High predictability means low privacy

15 months, 1.5 million people

6500 towers, 114 calls per month/user





With hourly aggregation,

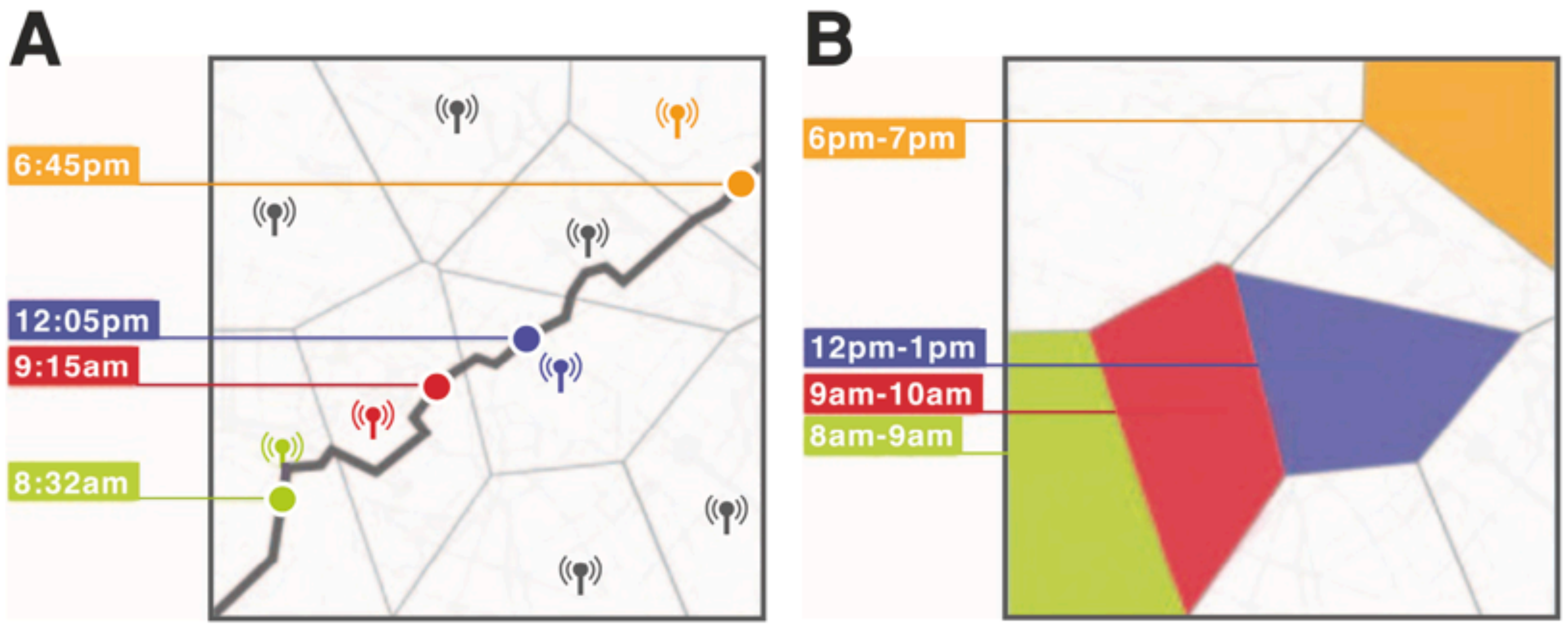
How many random spatiotemporal points are needed to identify 95% of users?

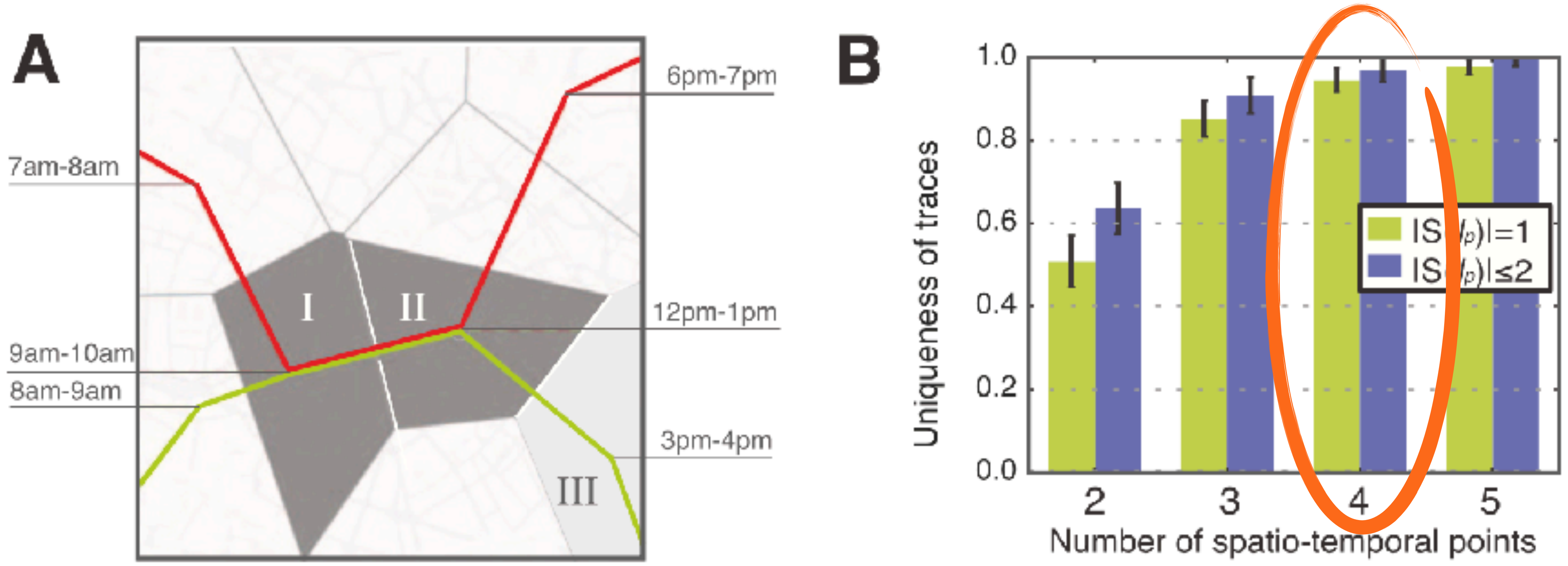
https://www.nature.com/articles/srep01376

High predictability means low privacy

15 months, 1.5 million people

6500 towers, 114 calls per month/user



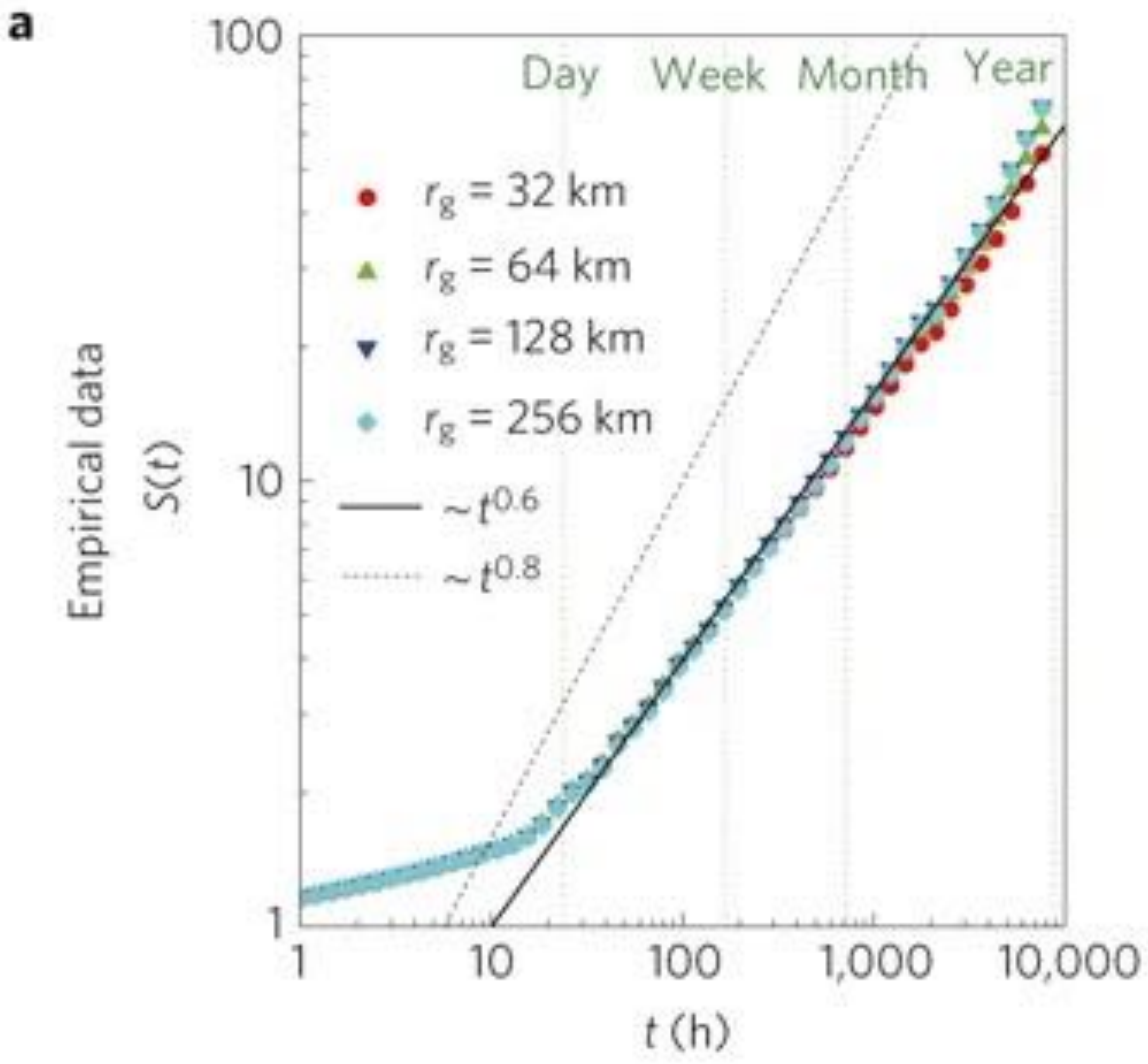


With hourly aggregation, only **4** random spatiotemporal points are needed to identify 95% of users!

https://www.nature.com/articles/srep01376

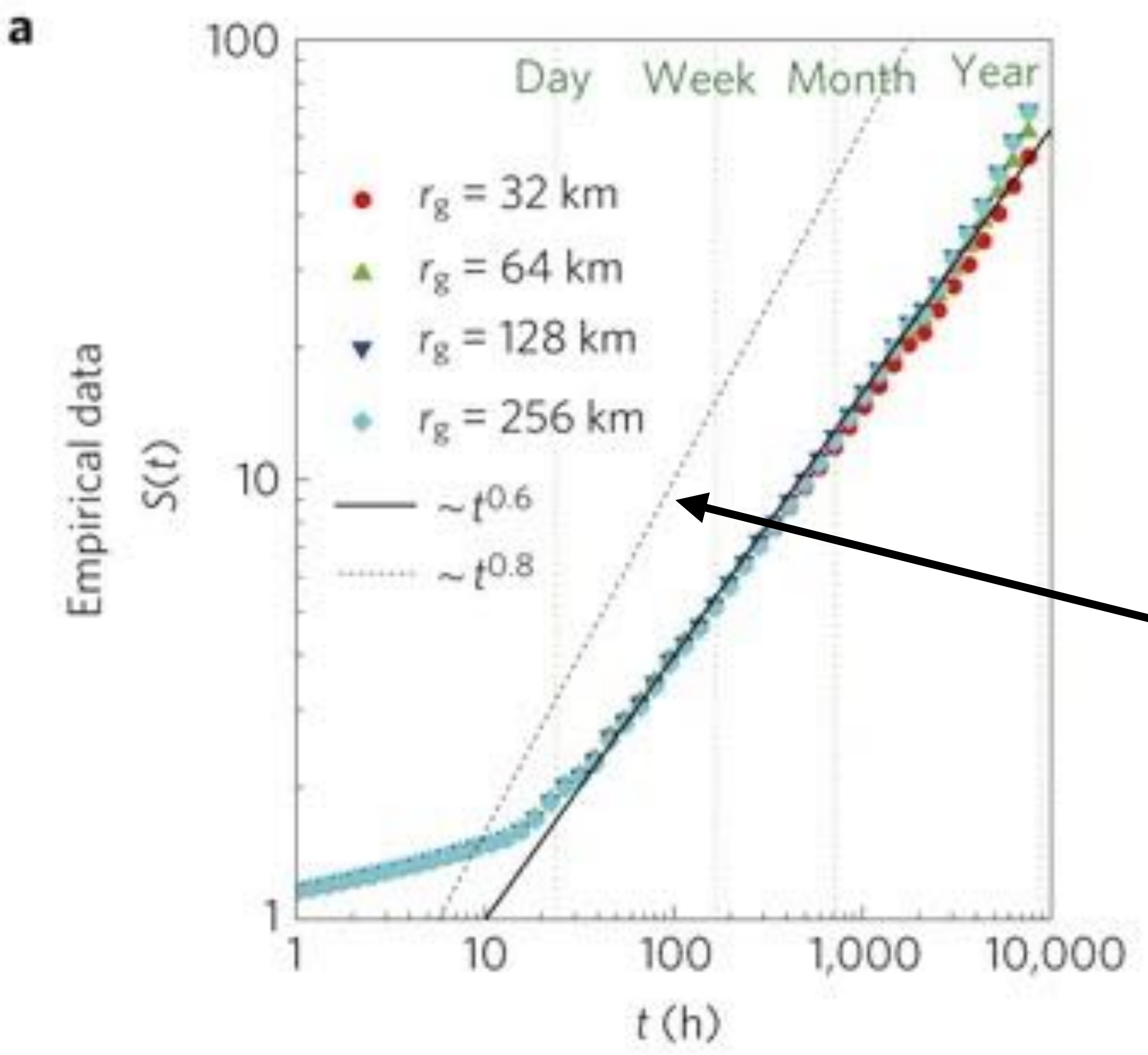
Individual mobility models

The number $S(t)$ of uniquely visited locations grows very slowly over time



https://www.nature.com/articles/nphys1760

The number $S(t)$ of uniquely visited locations grows very slowly over time

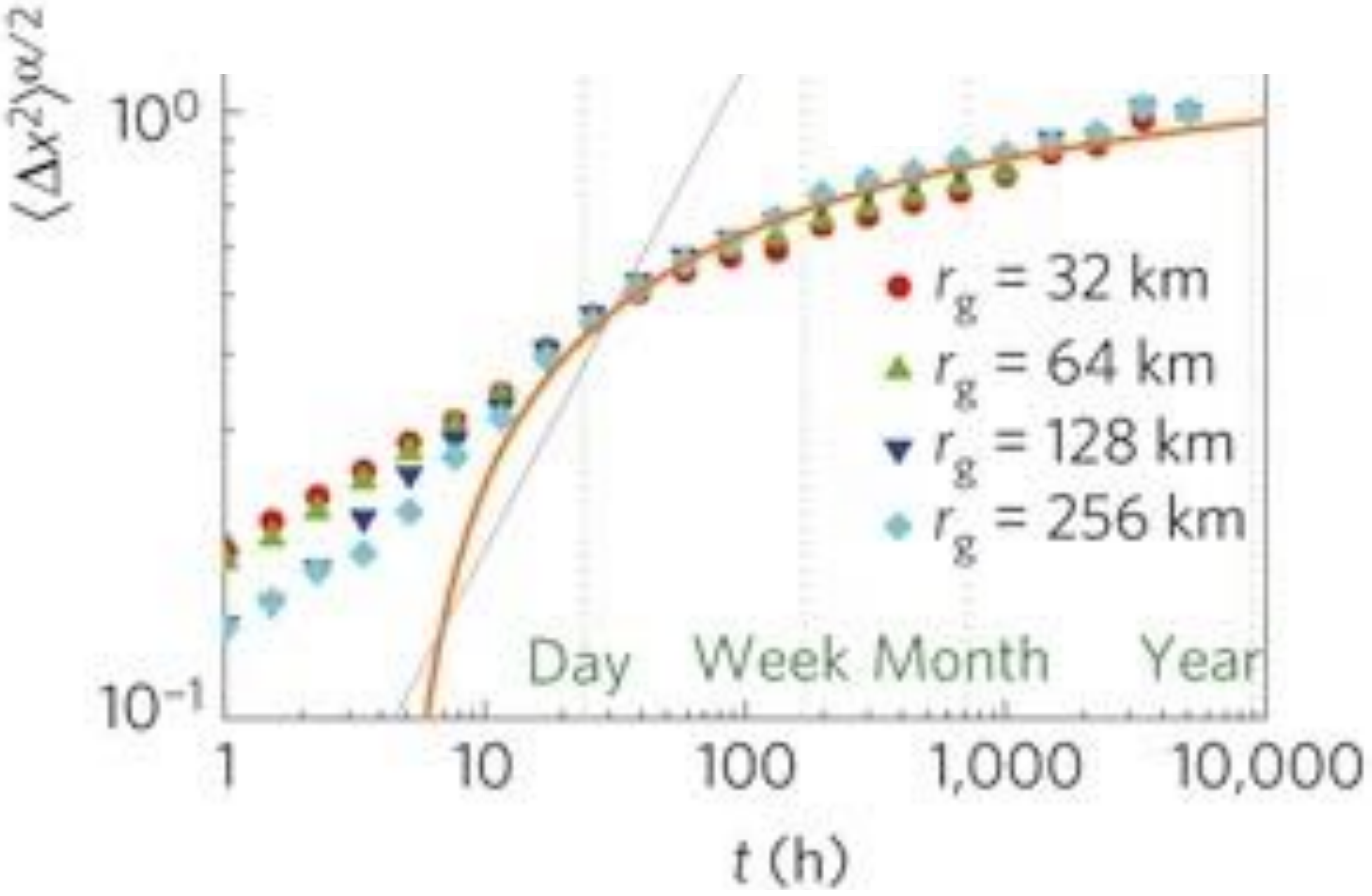


There is a decreasing tendency to visit previously unvisited locations

Incompatible with Levy-flight or CTRW

https://www.nature.com/articles/nphys1760

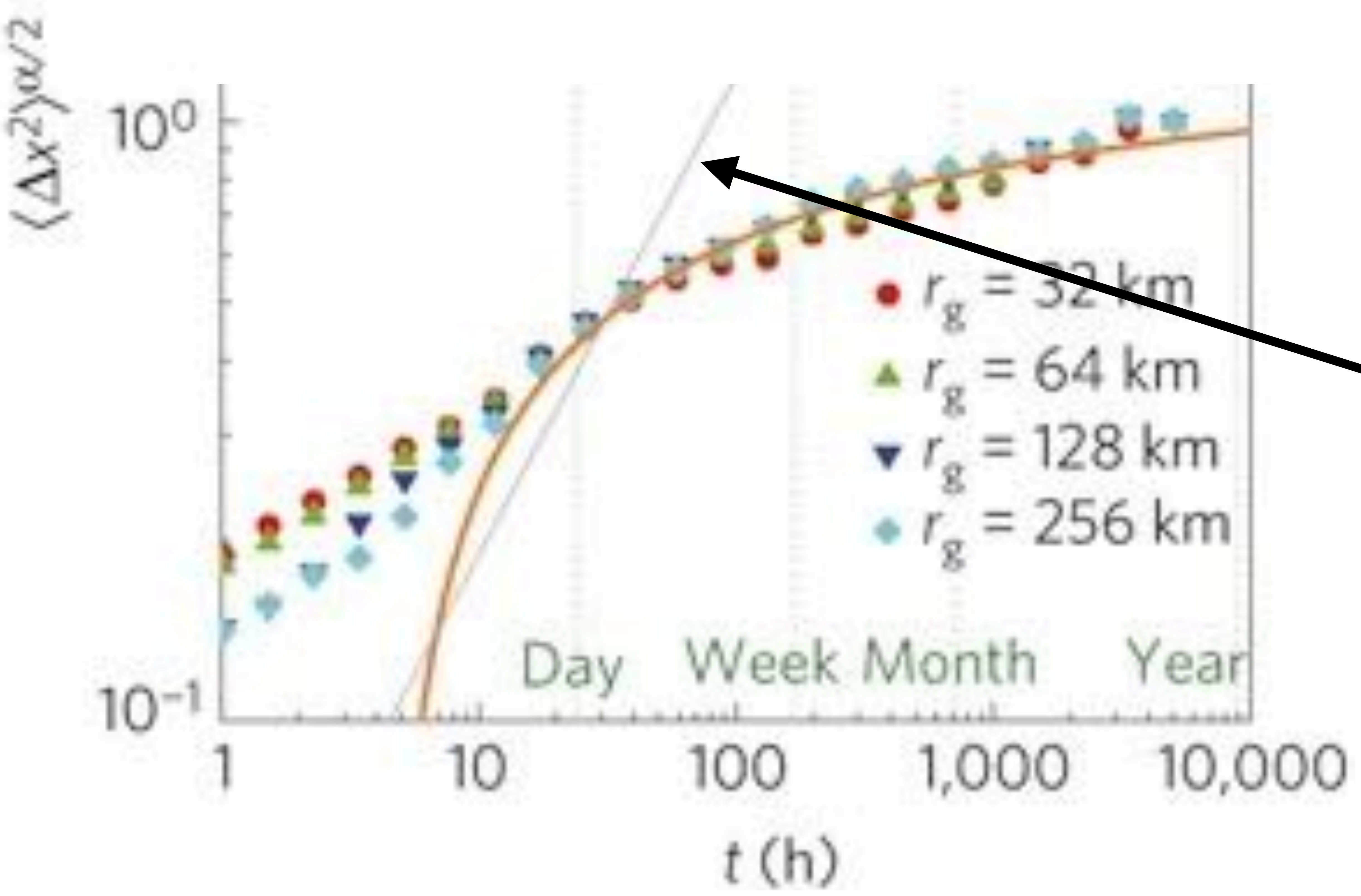
The MSD also grows ultraslowly over time



https://www.nature.com/articles/nphys1760

The MSD also grows ultraslowly over time

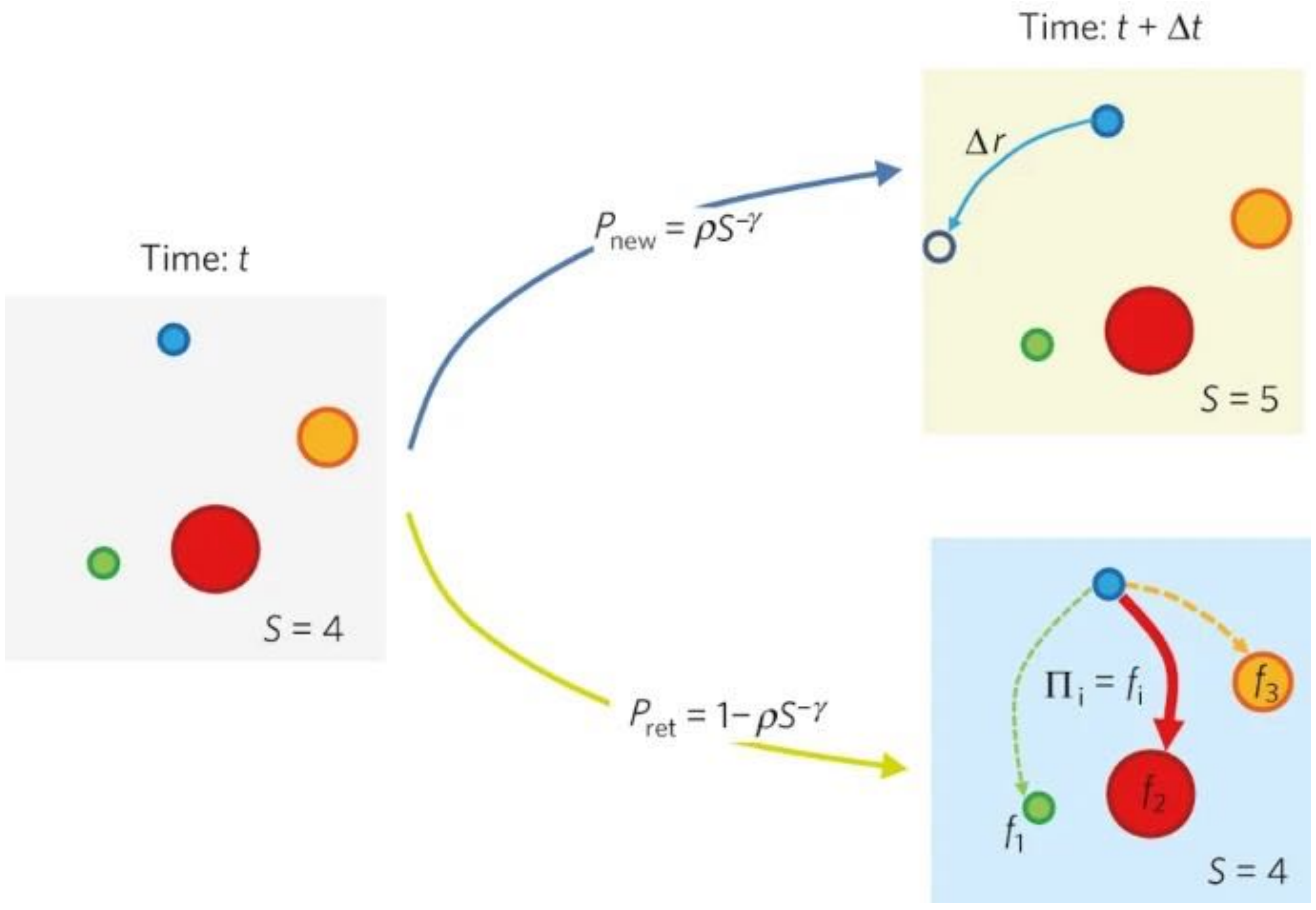
humans have a tendency to return home on a daily basis, suggesting that simple diffusive processes, which are not recurrent in two dimensions, do not offer a suitable description of human mobility



Incompatible with Levy-flight or CTRW

https://www.nature.com/articles/nphys1760

This can be fixed by preferential return: Often people return to old places



Explore a new place

Return to an old place

https://www.nature.com/articles/nphys1760

When preferential return is too simple

1) Time-ordered memory
2) Returners-Explorers

Problem 1 with preferential return: Walkers get stuck in often visited places

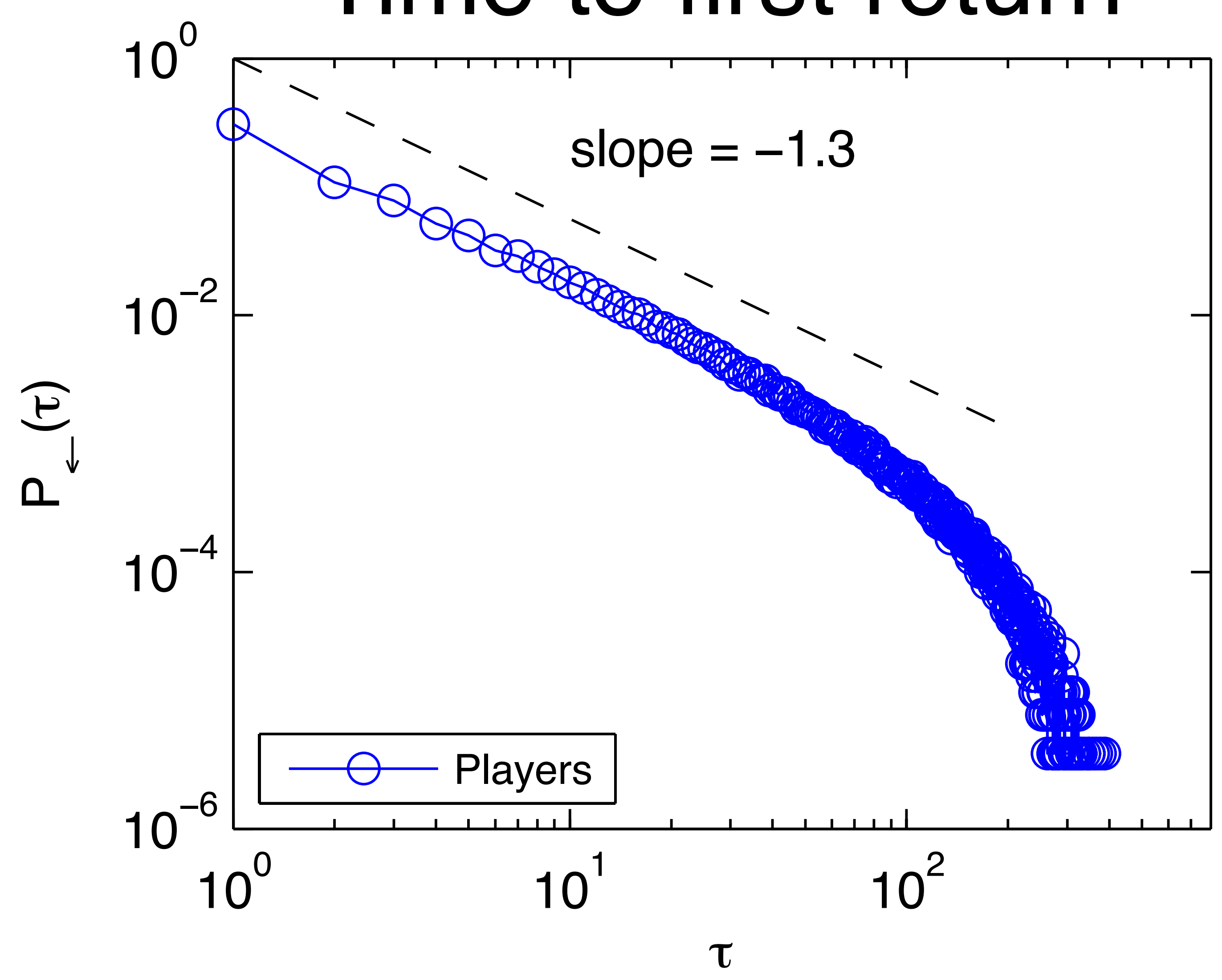
Because of $\Pi_i = f_i$, I will prefer to return to an often visited place, and I will get "stuck" in my first visited places.

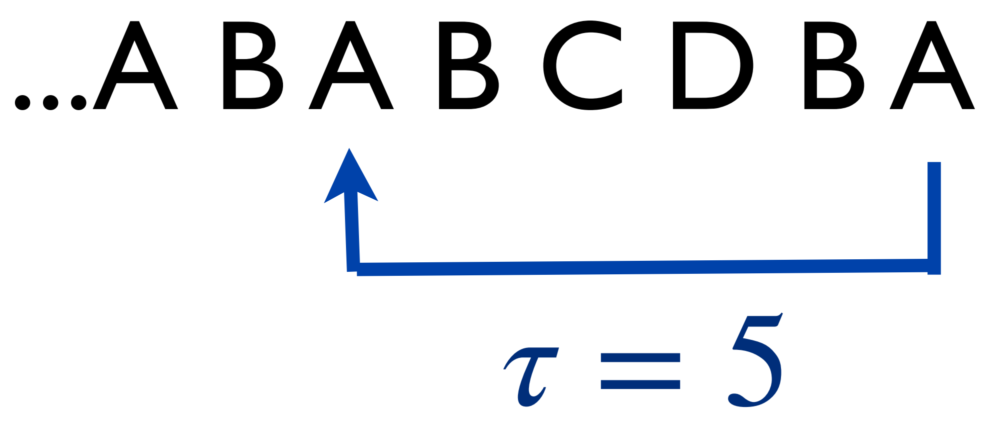
Problem 1 with preferential return: Walkers get stuck in often visited places

Because of $\Pi_i = f_i$, I will prefer to return to an often visited place, and I will get "stuck" in my first visited places.

New idea: I want to return to a *recently* visited place, not necessarily to an often visited one.

Time to first return

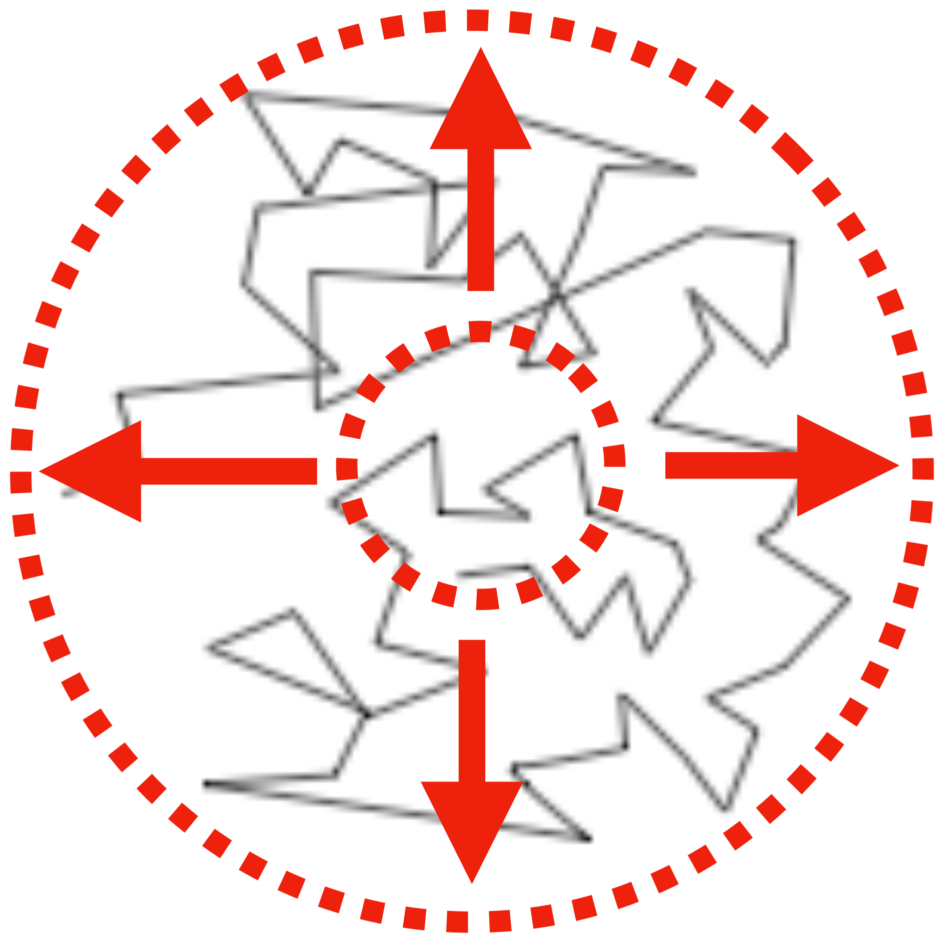




Long-time memory

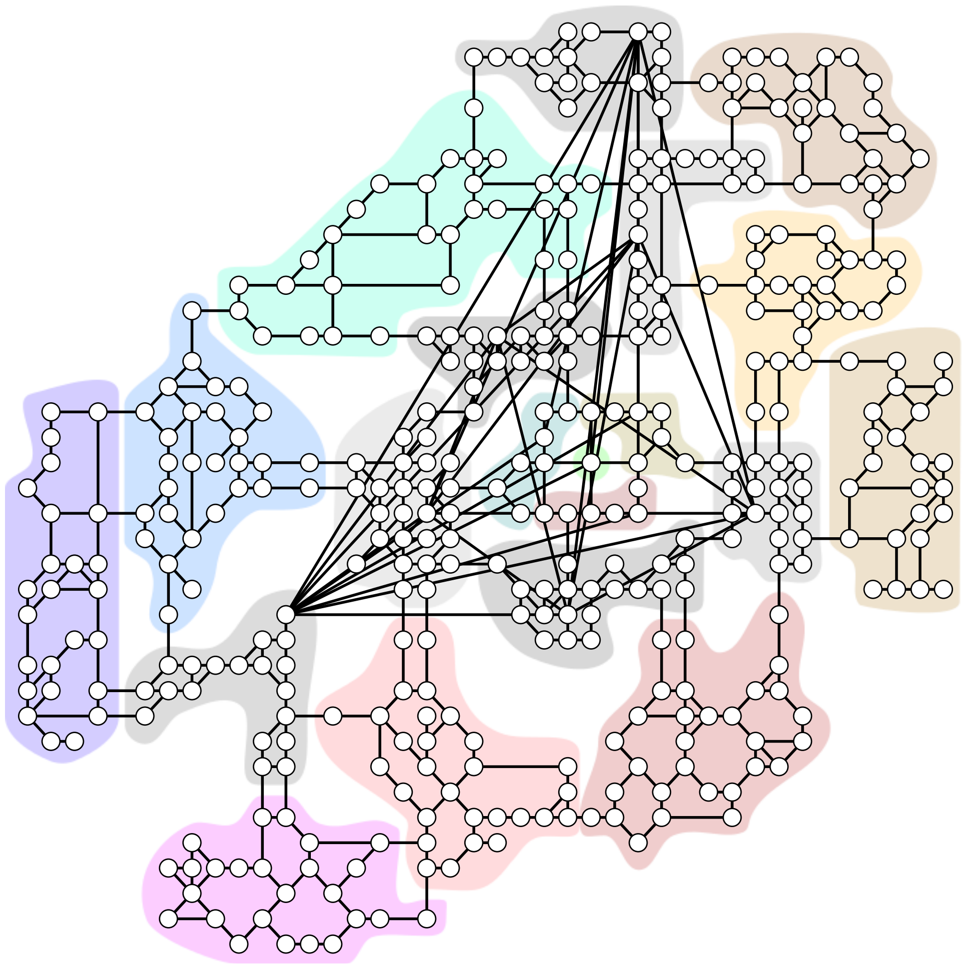
$$P_{\hookleftarrow}(\tau) \sim \tau^{-1.3}$$

Problem 1 with preferential return: Walkers get stuck in often visited places

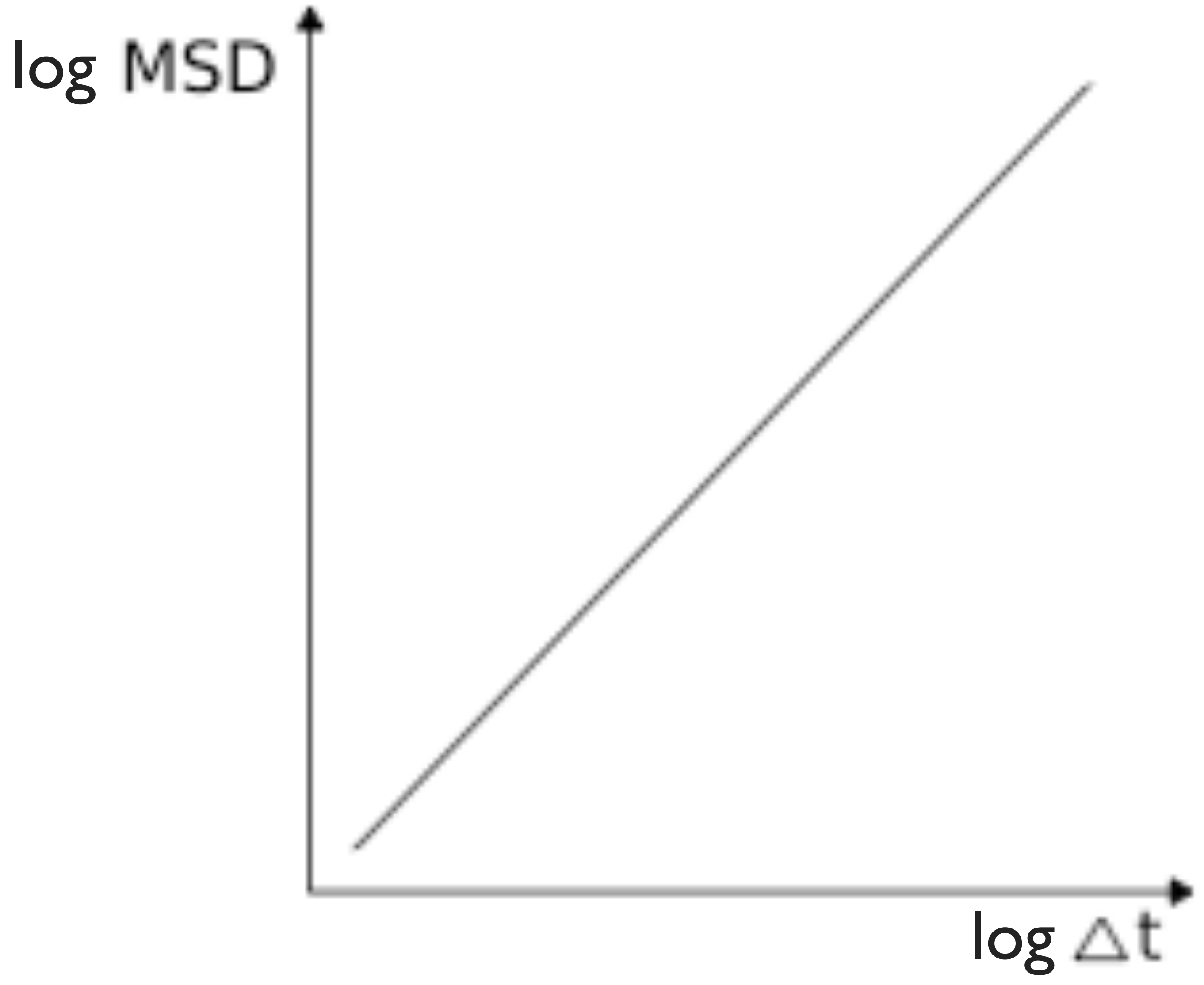


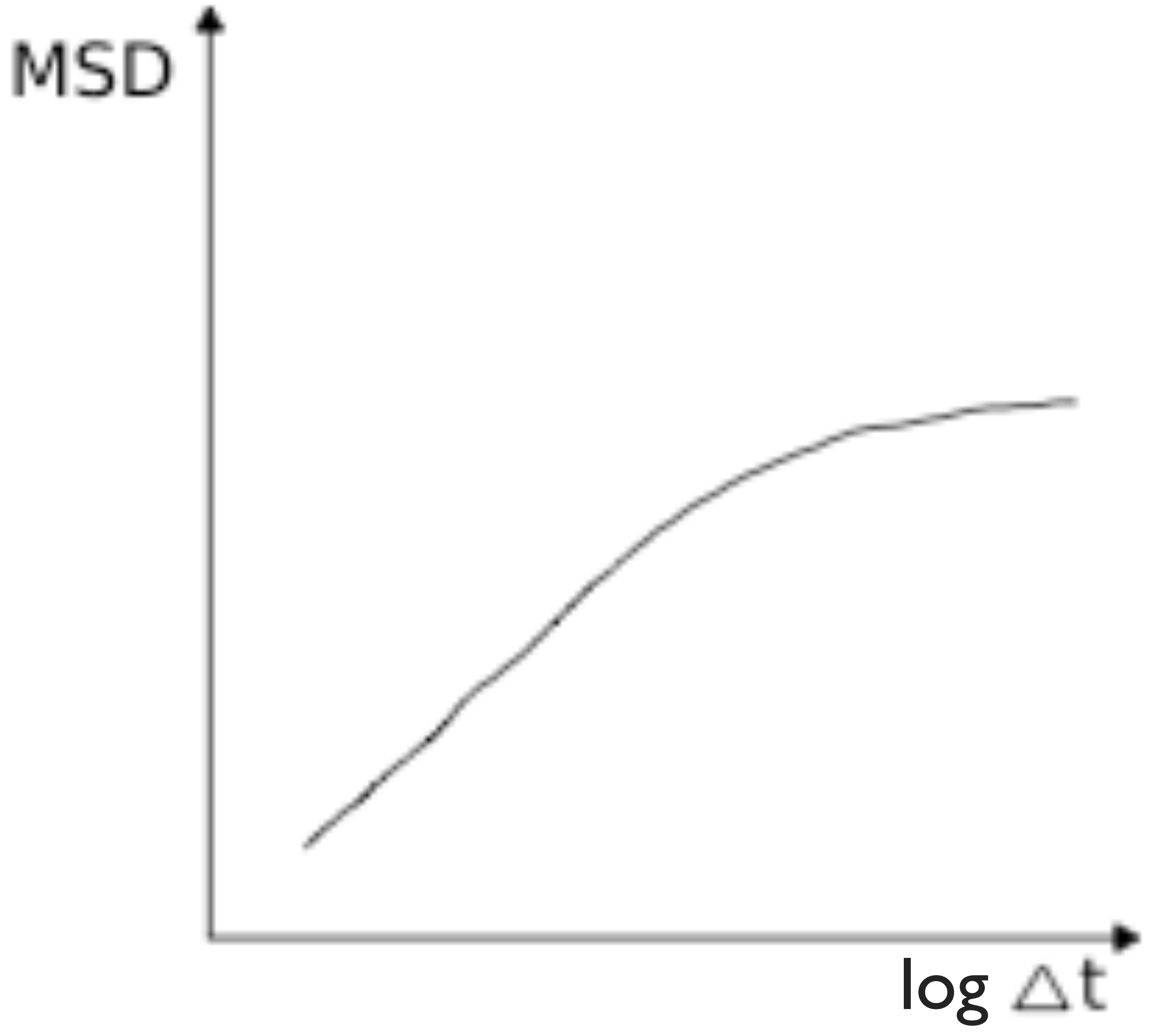
Problem 1 with preferential return: Walkers get stuck in often visited places

2012



We are in a finite universe,
so the MSD should saturate





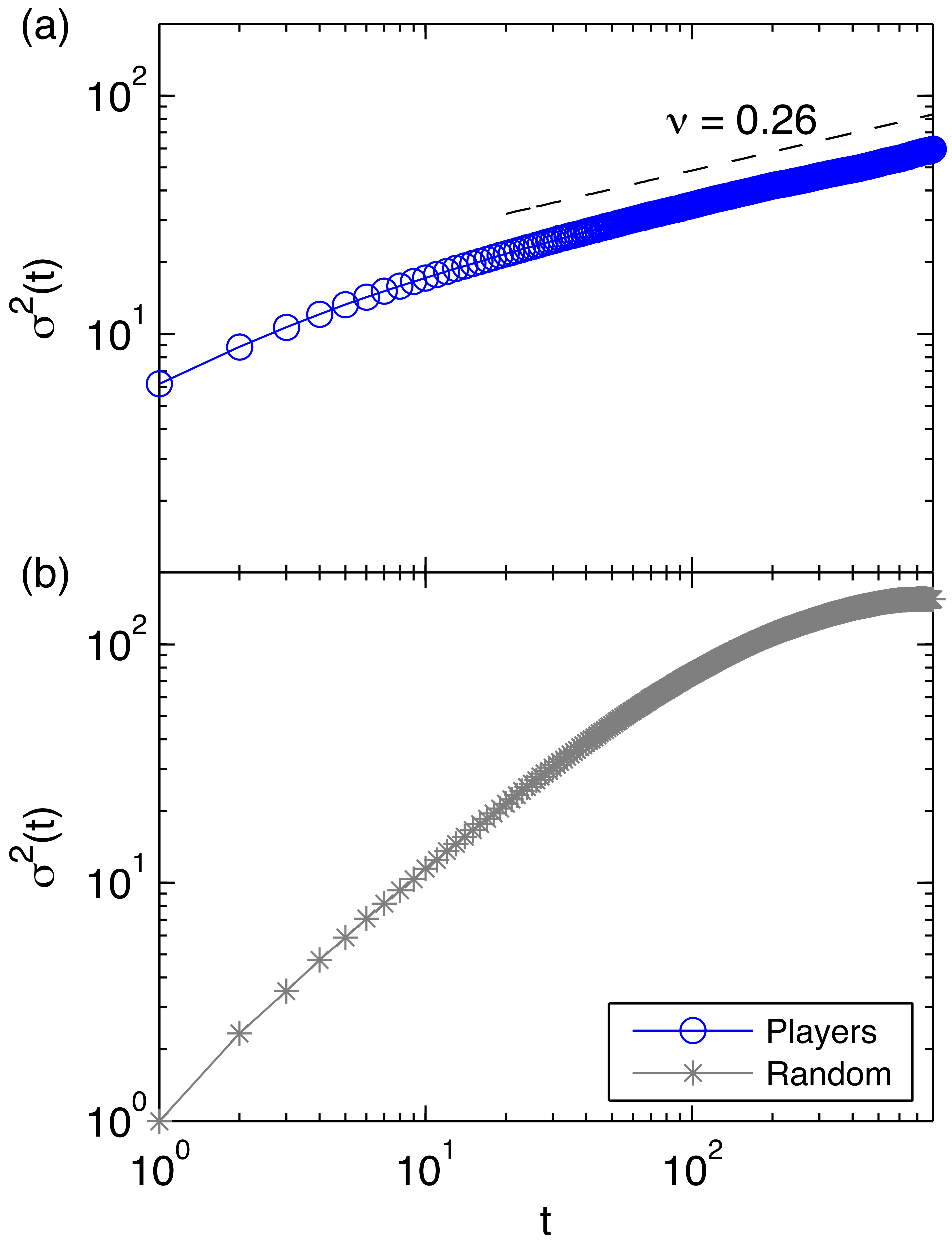
Problem 1 with preferential return: Walkers get stuck in often visited places

$\nu = 0.26 < 1$

Subdiffusive

Random walkers

have $\nu = 1$



Players

Models

Problem 1 with preferential return: Walkers get stuck in often visited places

In a markov process, the transition probability p_{ij} depends only on the present state:

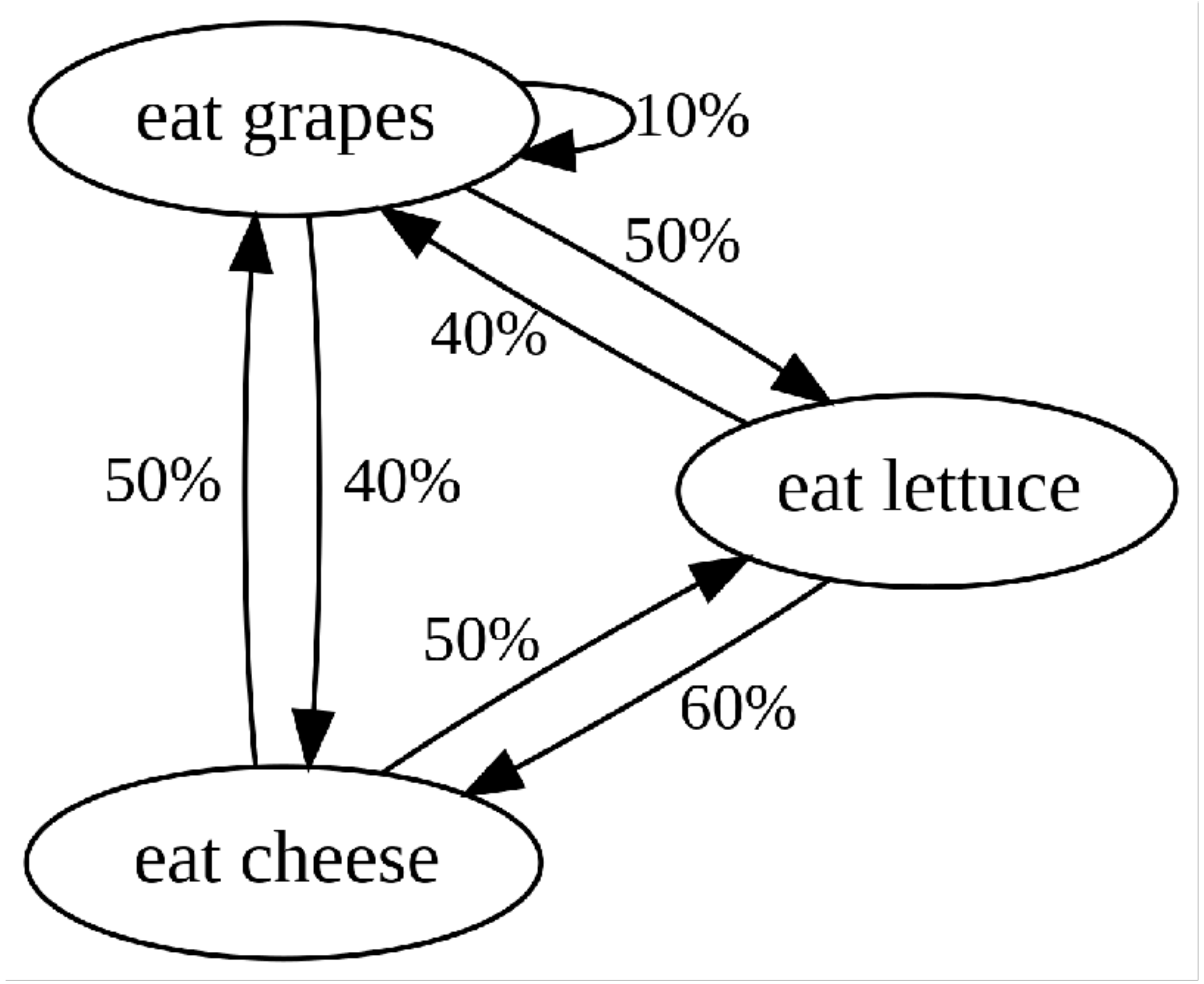
$$p_{ij} = P(X_{n+1} = j \,|\, X_n = i)$$

https://en.wikipedia.org/wiki/Markov_chain

Problem 1 with preferential return: Walkers get stuck in often visited places

In a markov process, the transition probability p_{ij} depends only on the present state:

$$p_{ij} = P(X_{n+1} = j \mid X_n = i)$$



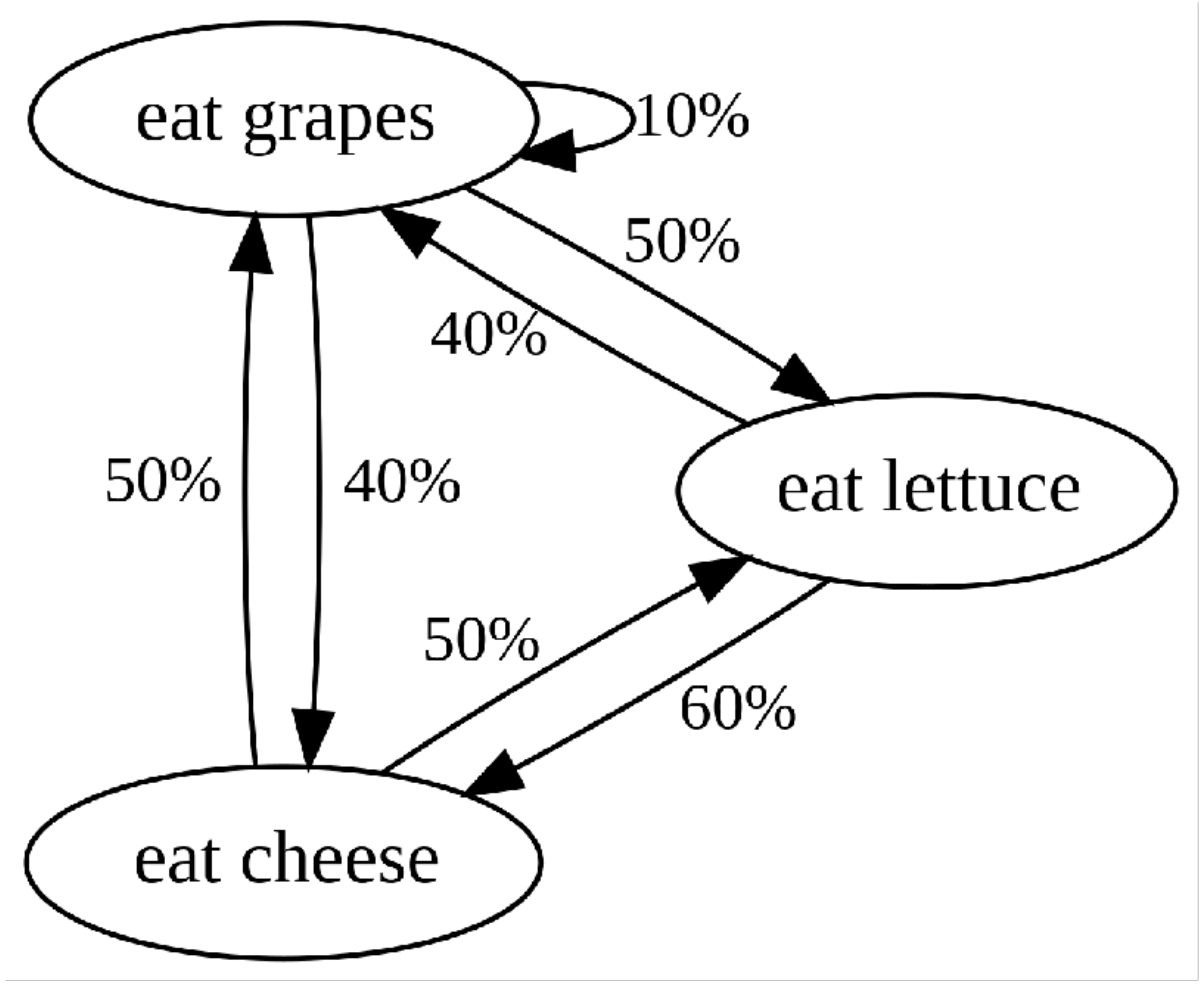
https://en.wikipedia.org/wiki/Markov_chain

Problem 1 with preferential return: Walkers get stuck in often visited places

In a markov process, the transition probability p_{ij} depends only on the present state:

$$p_{ij} = P(X_{n+1} = j \,|\, X_n = i)$$

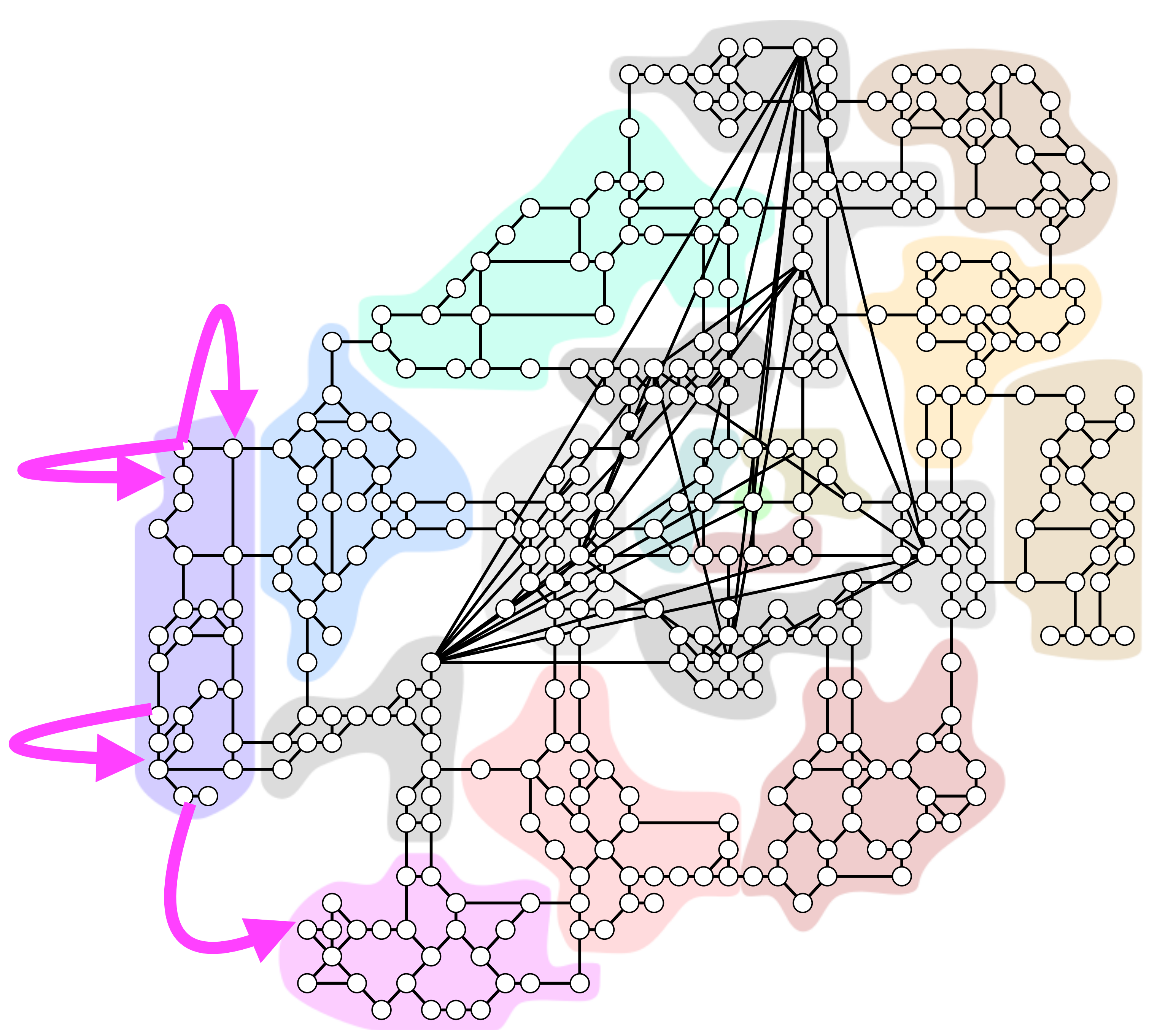
Markov-property (memorylessness):
There is no memory.



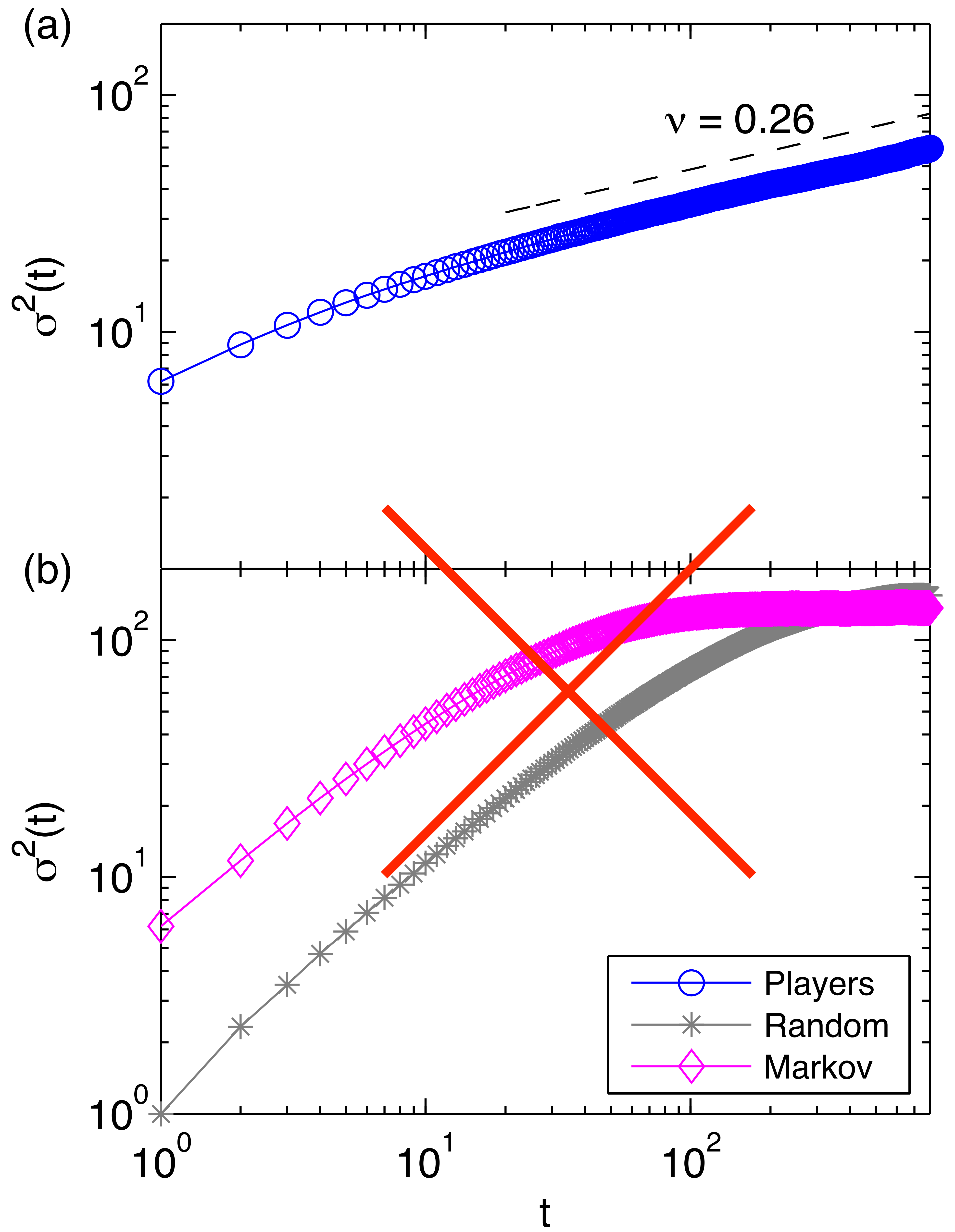
https://en.wikipedia.org/wiki/Markov_chain

Problem 1 with preferential return: Walkers get stuck in often visited places

400x400 = 160,000



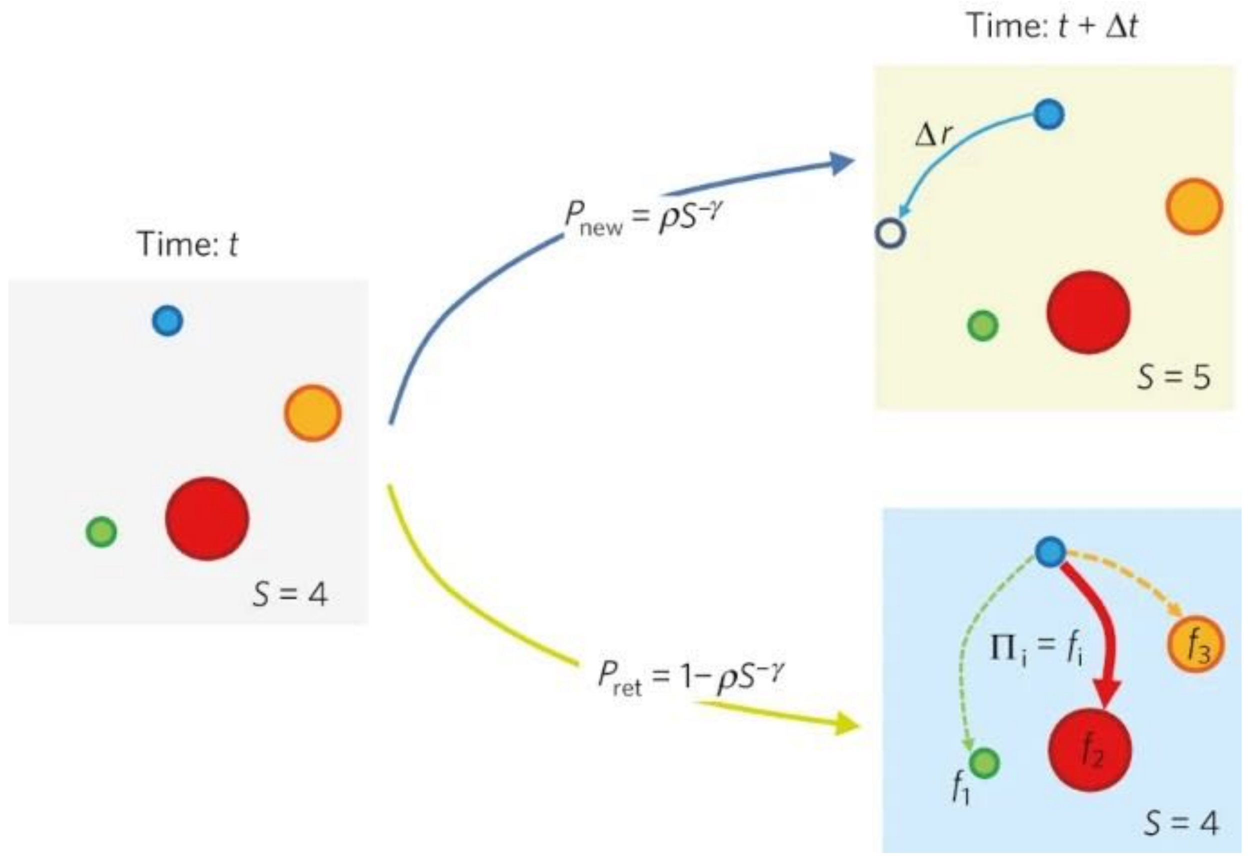
Markov model: Use ALL day-to-day transition probabilities



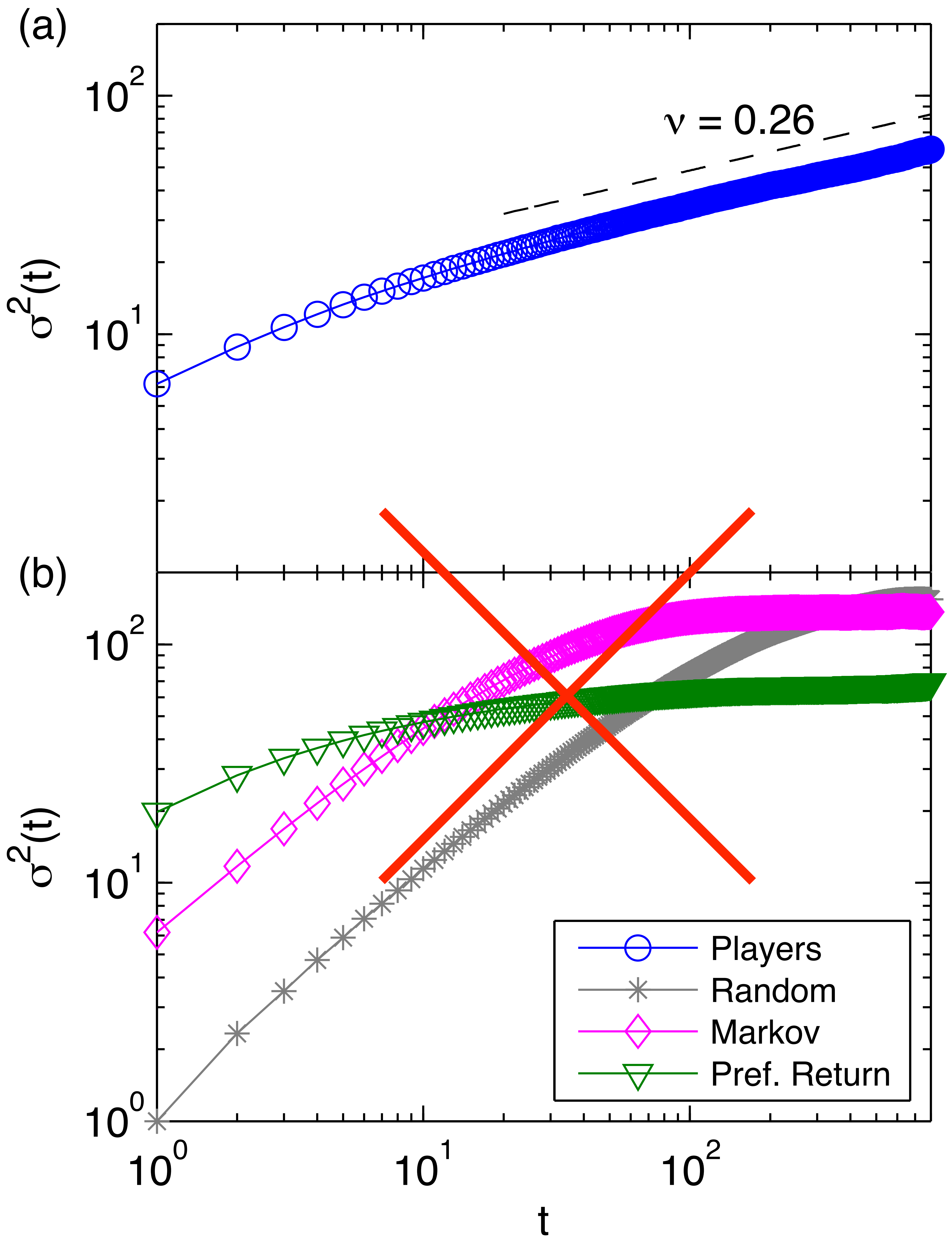
Players

Models

Problem 1 with preferential return: Walkers get stuck in often visited places



Preferential return
to often visited places



Players

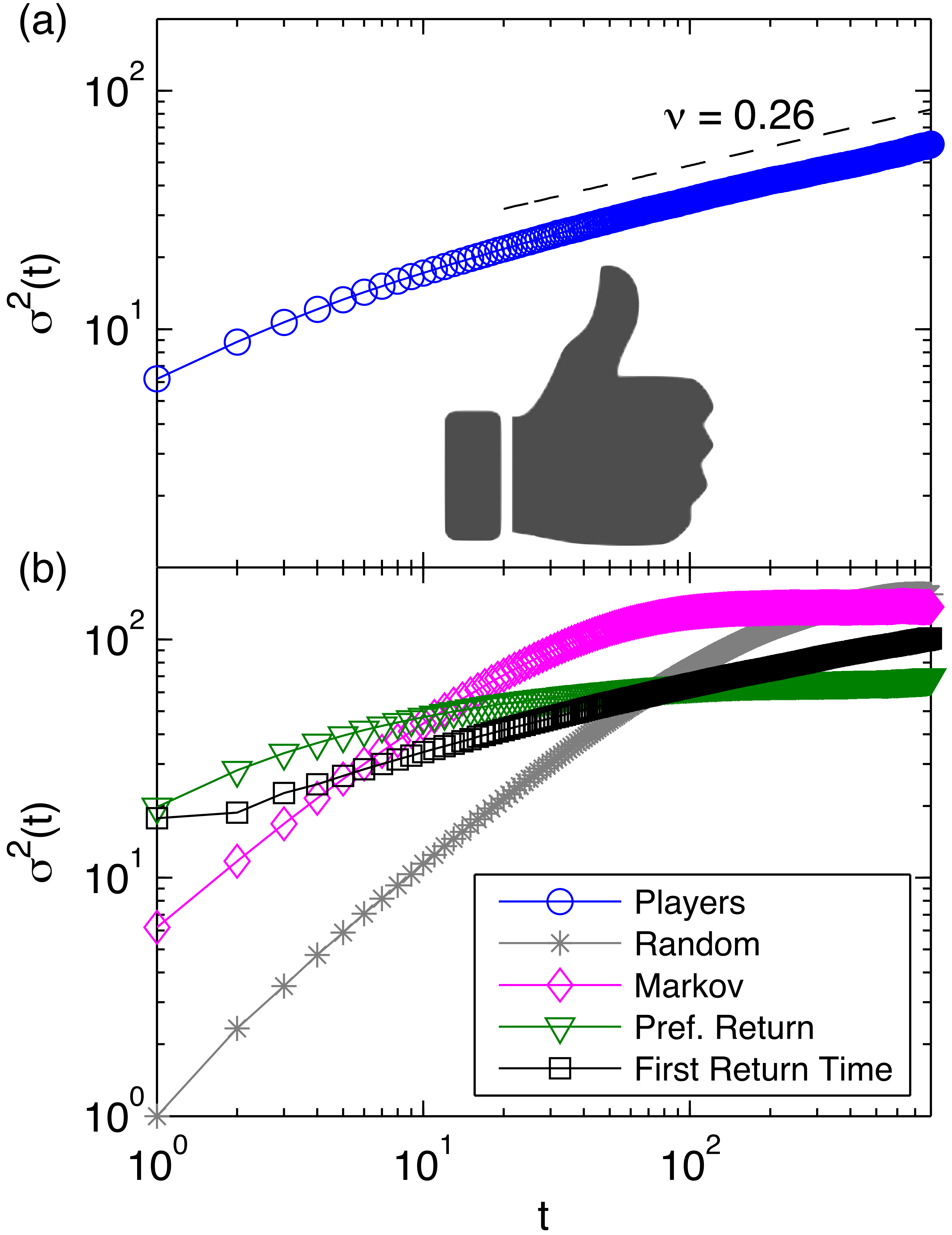
Models

Problem 1 with preferential return: Walkers get stuck in often visited places

...A B A B C D B A

5

Time-Ordered Memory (TOM):
Return to recently visited places

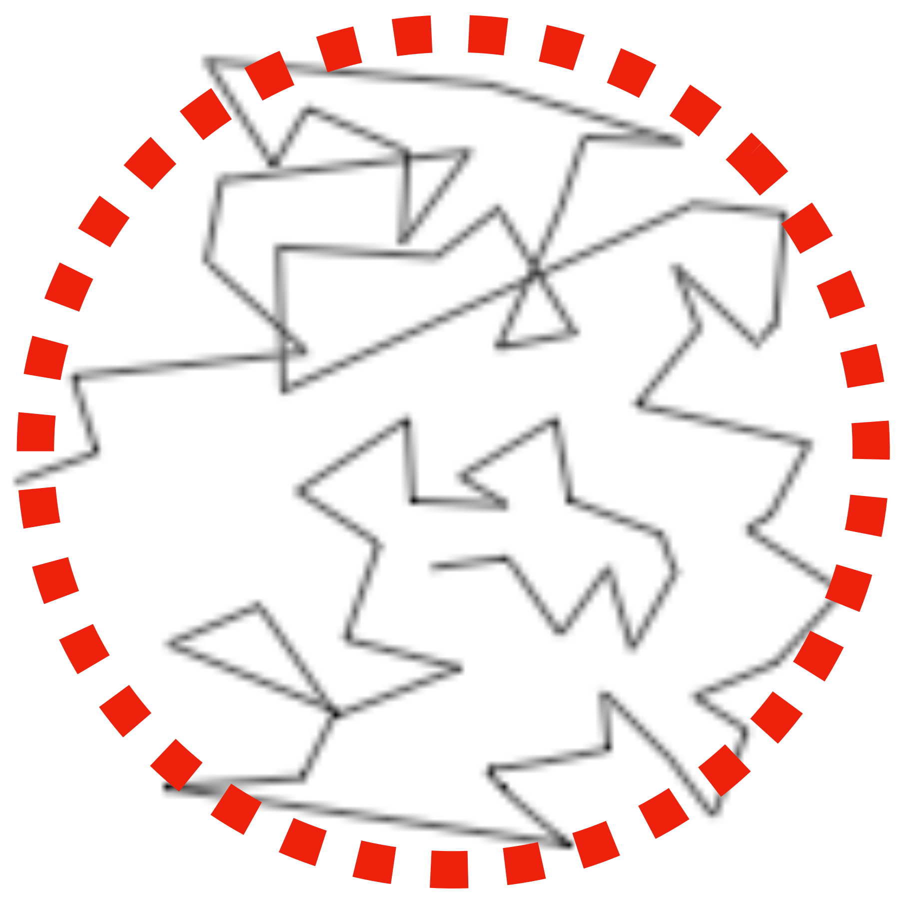


Players

Models

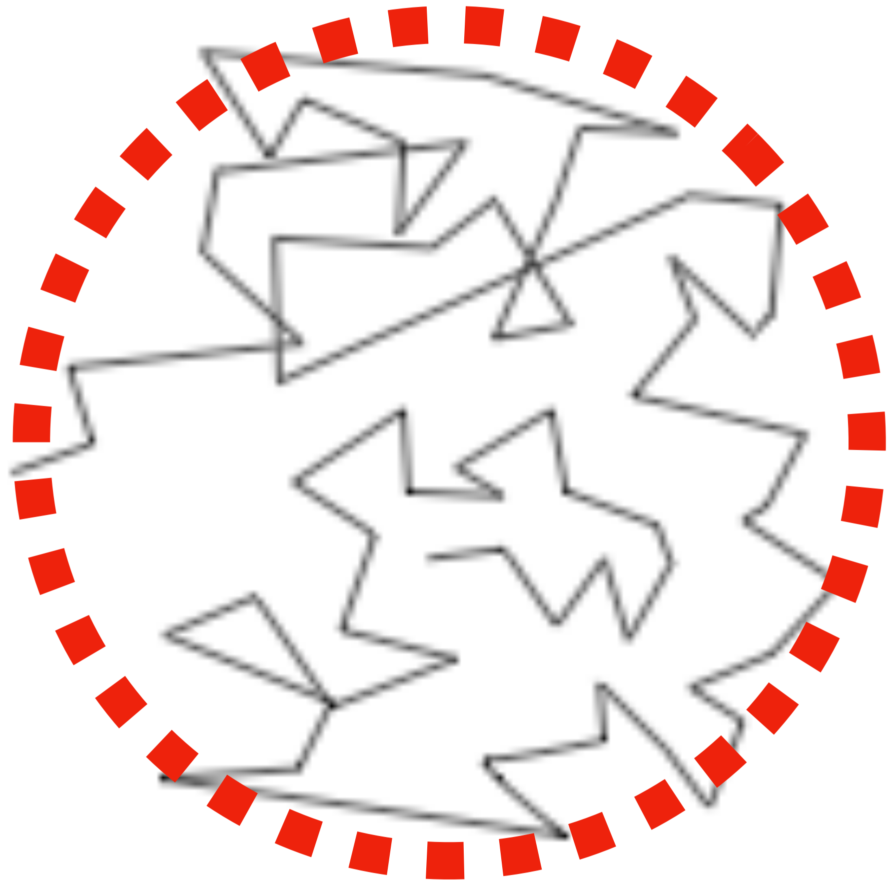
Problem 2 with preferential return: Cannot explain explorers

$$r_g = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |r_i - r_{\mathrm{cm}}|^2}$$



Problem 2 with preferential return: Cannot explain explorers

$$r_g = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |r_i - r_{cm}|^2}$$

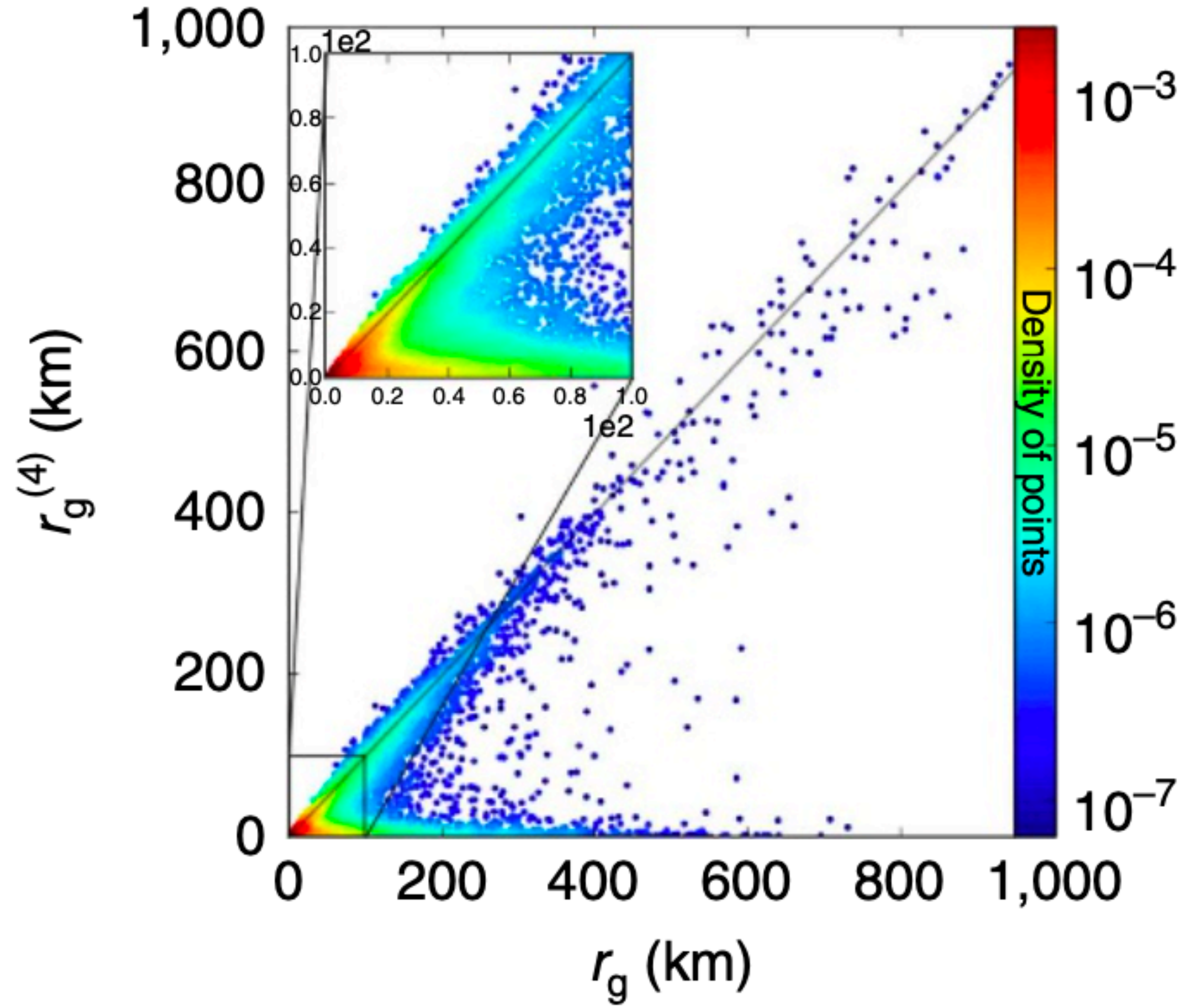


$$r_g^{(k)} = \sqrt{\frac{1}{n_k}\sum_{i=1}^{n_k} |r_i - r_{cm}|^2}$$

k-radius of gyration = radius of gyration but just for the top k visited places

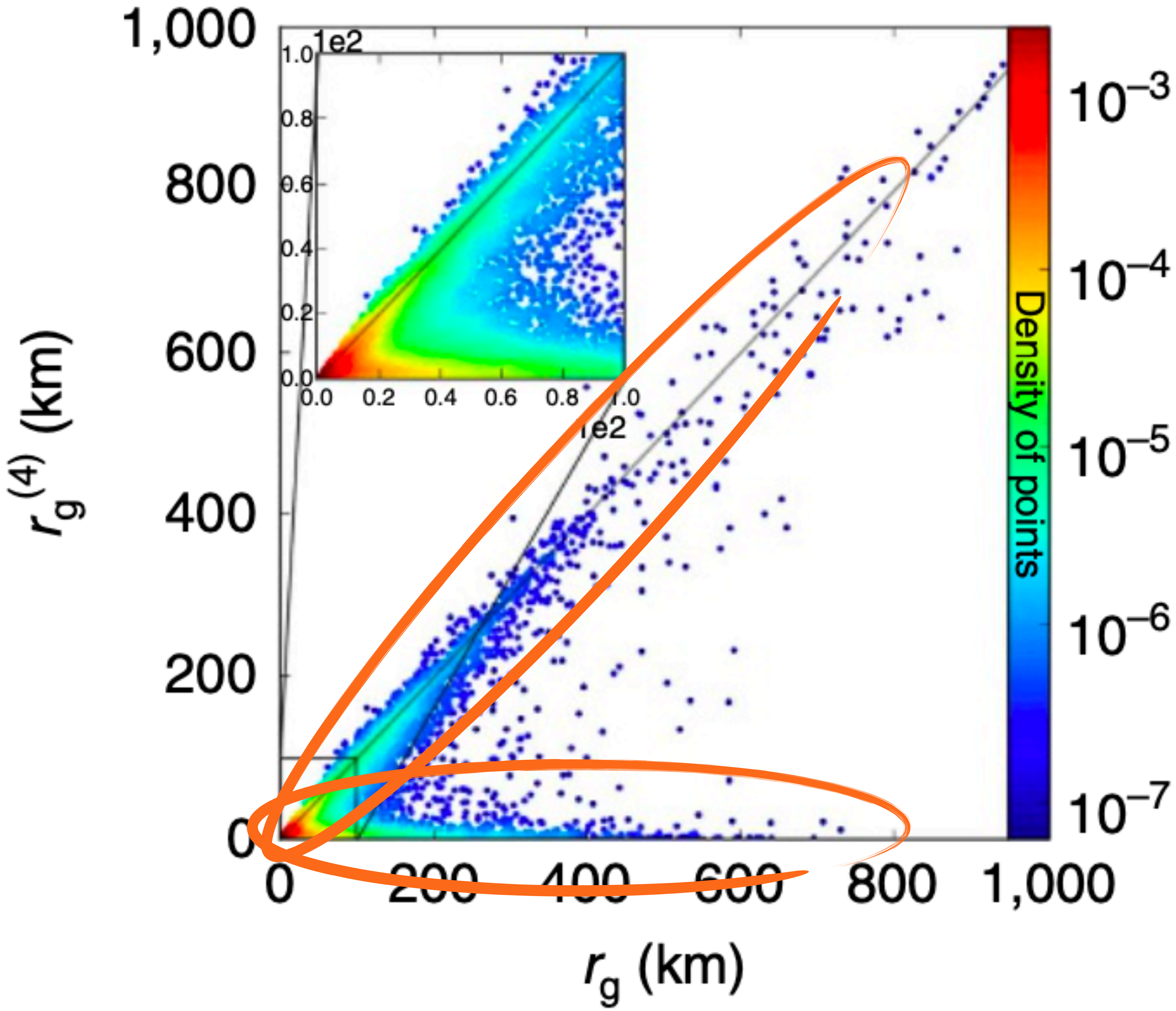
https://www.nature.com/articles/ncomms9166

Problem 2 with preferential return: Cannot explain explorers



https://www.nature.com/articles/ncomms9166

Problem 2 with preferential return: Cannot explain explorers

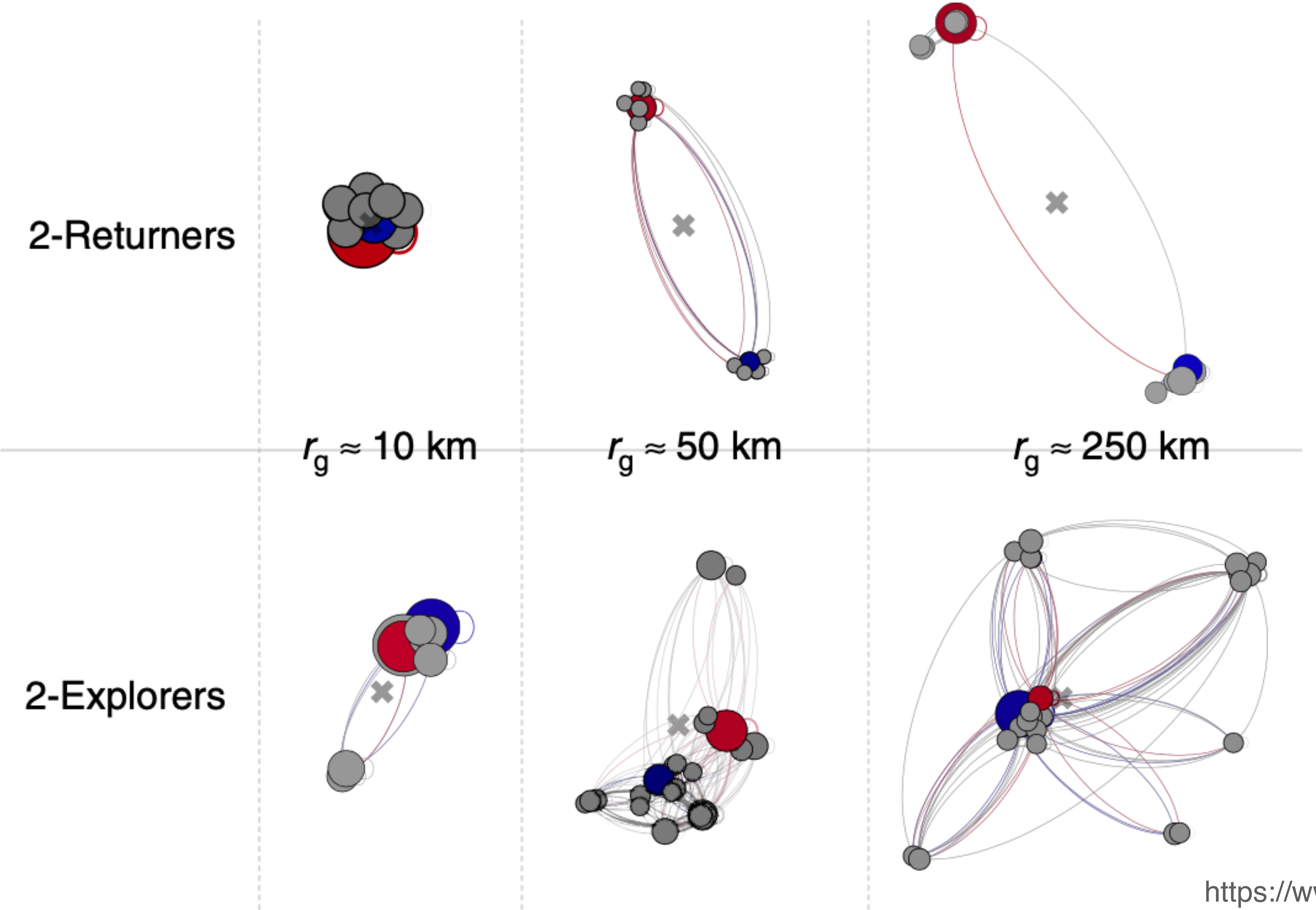


k-returners: Their mobility is well-approximated by their k most visited places

k-explorers: We cannot reduce their mobility to their top k places

https://www.nature.com/articles/ncomms9166

Problem 2 with preferential return: Cannot explain explorers



https://www.nature.com/articles/ncomms9166

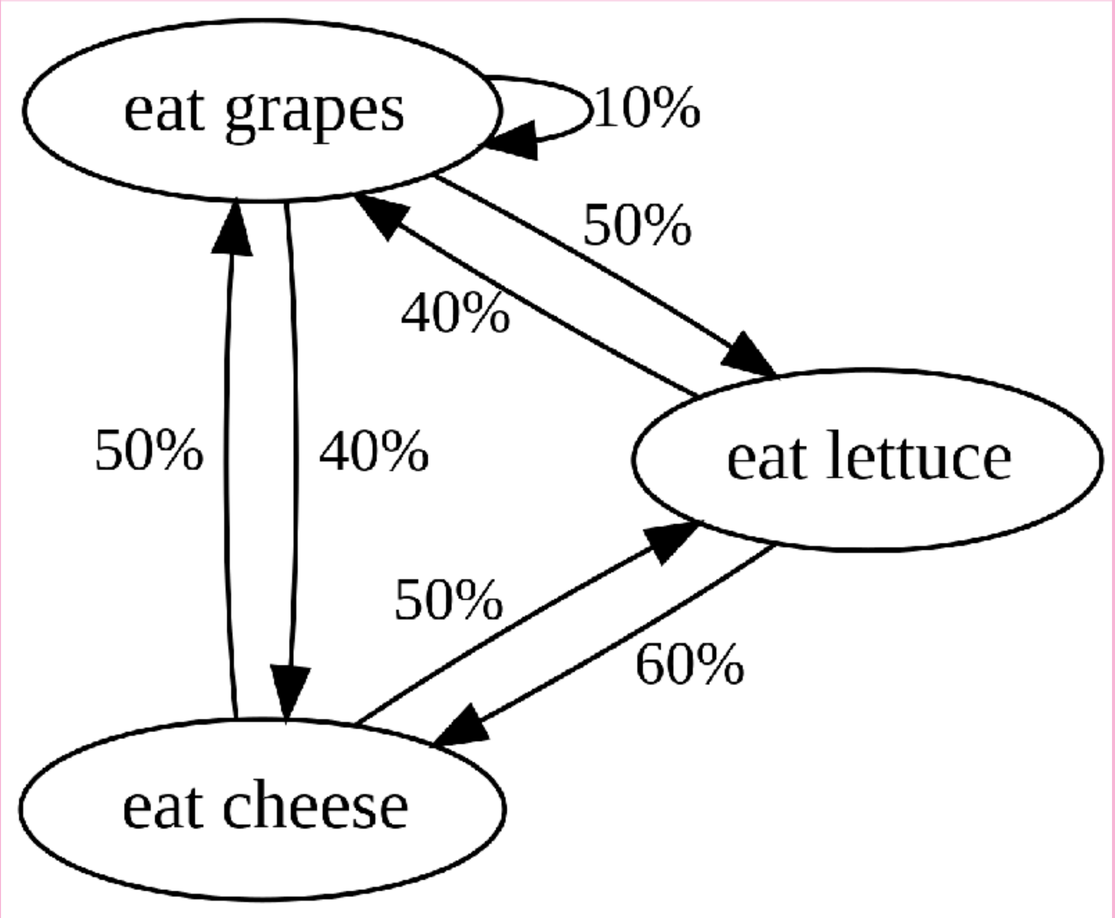
Problem 2 with preferential return: Cannot explain explorers

The EPR model (Exploration and Preferential Return) extends preferential return with a more elaborate exploration phase.

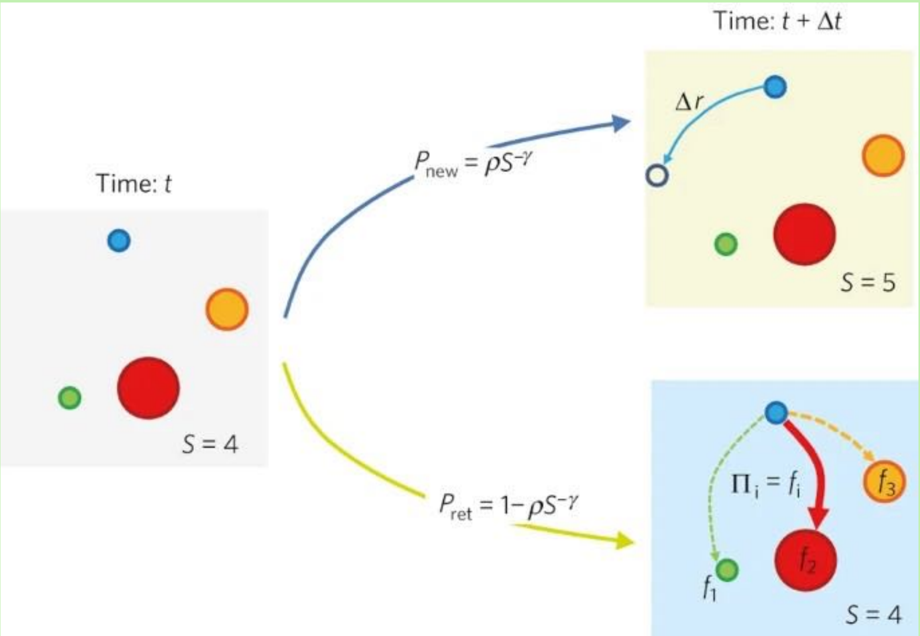
https://www.nature.com/articles/ncomms9166

Summary of individual mobility models

Random walker



Markov Process

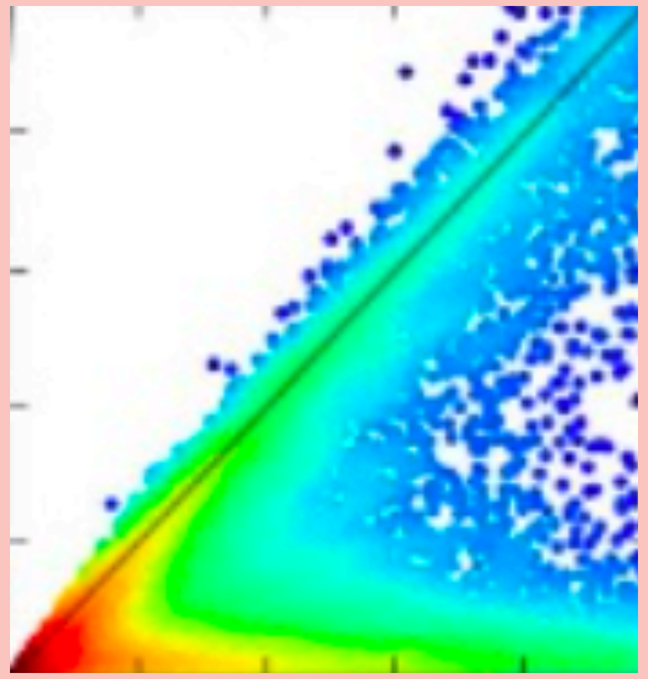


Preferential return

...A B A B C D B A

5

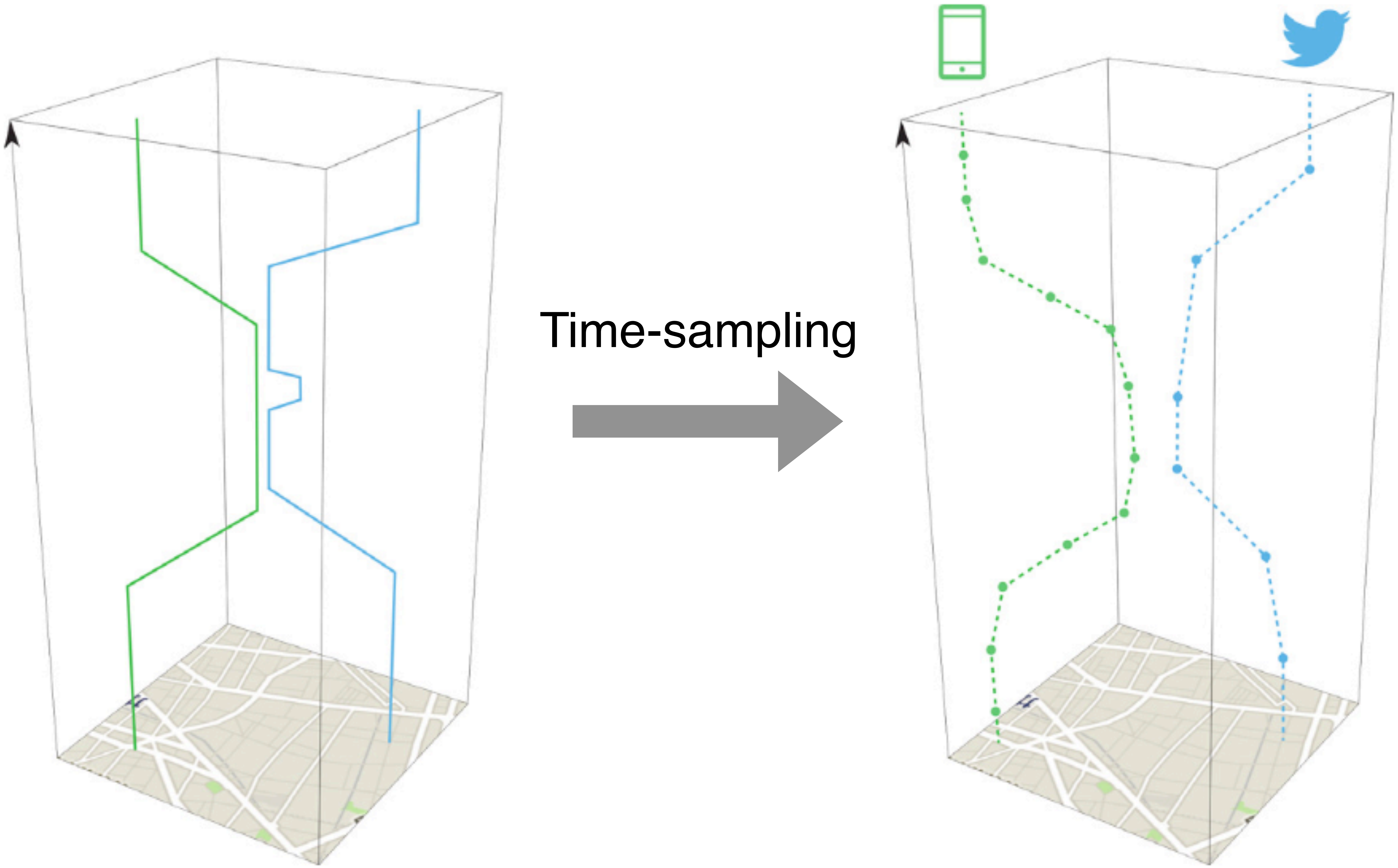
Time-Ordered Memory (TOM)



Explorers+Returners (EPR)

Motifs and hi-res data

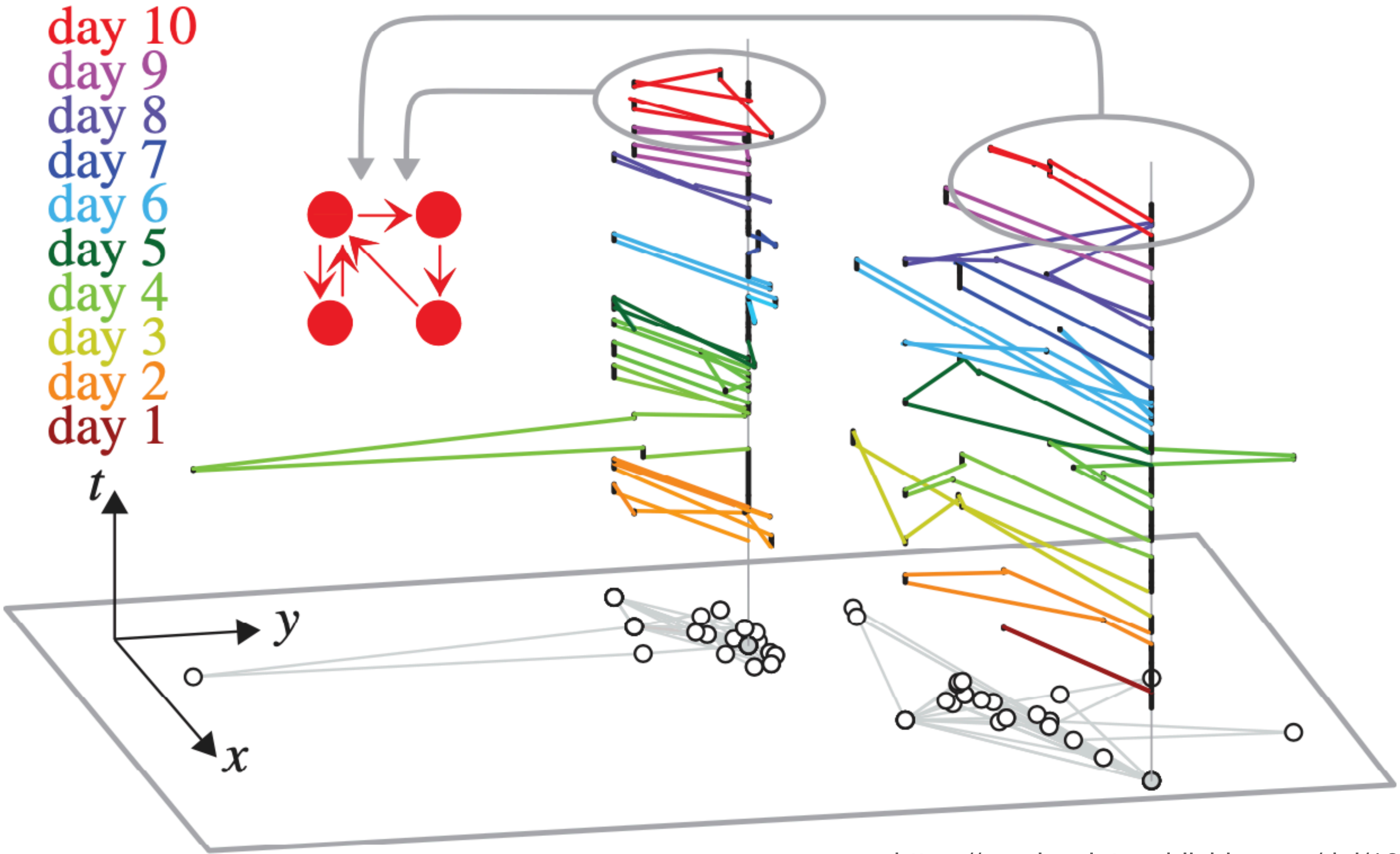
The cube of time geography shows mobility in space-time



Hägerstrand (1970)

We can now measure daily patterns: motifs

2013

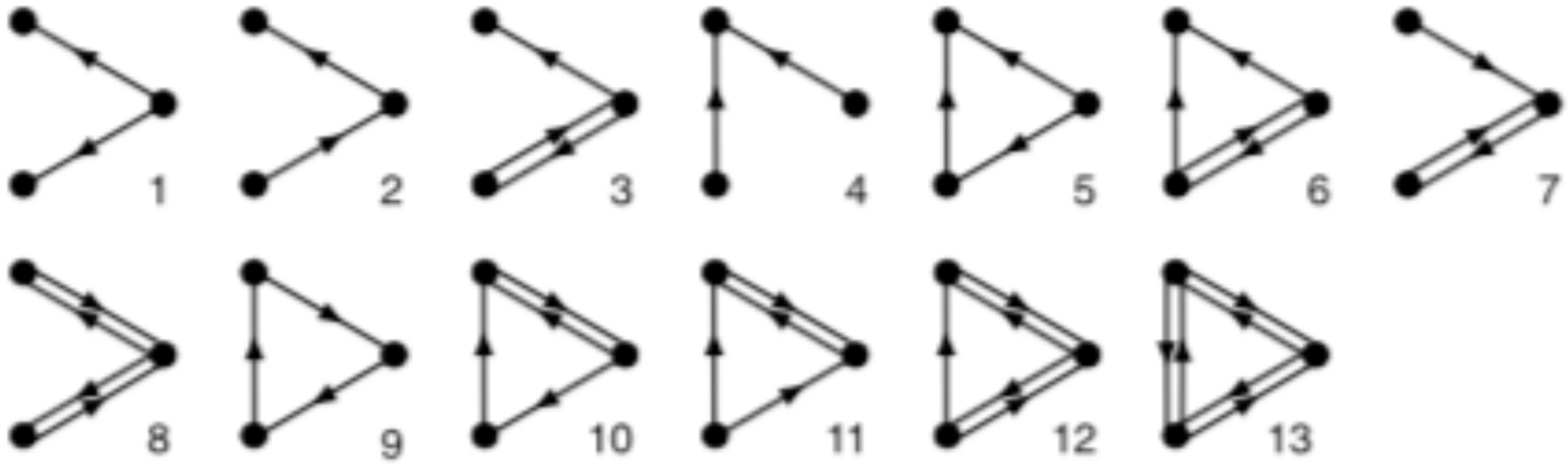


https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0246

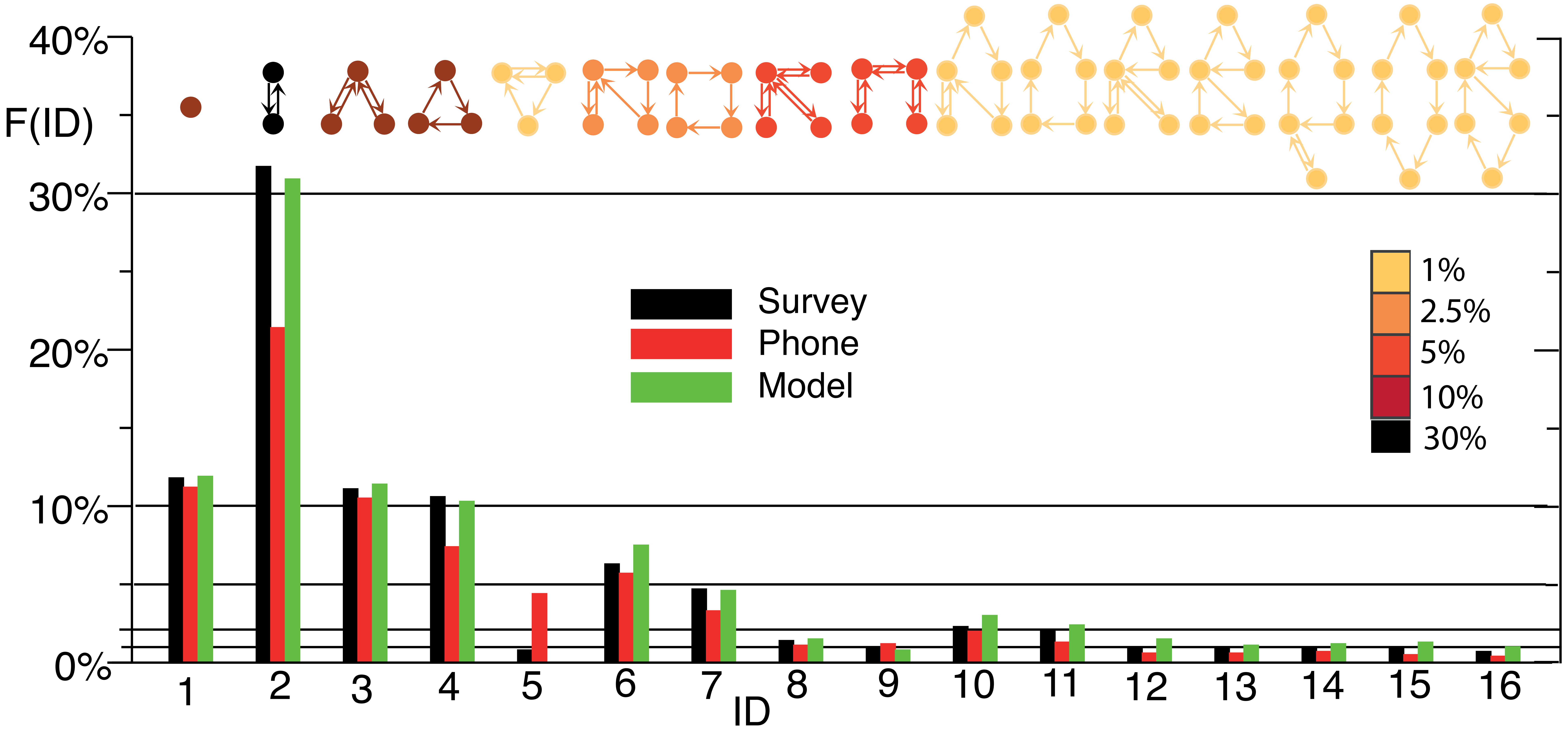
We can now measure daily patterns: motifs

A network motif is a recurrent and statistically significant sub-graph.

Motifs are usually small: 3-5 nodes



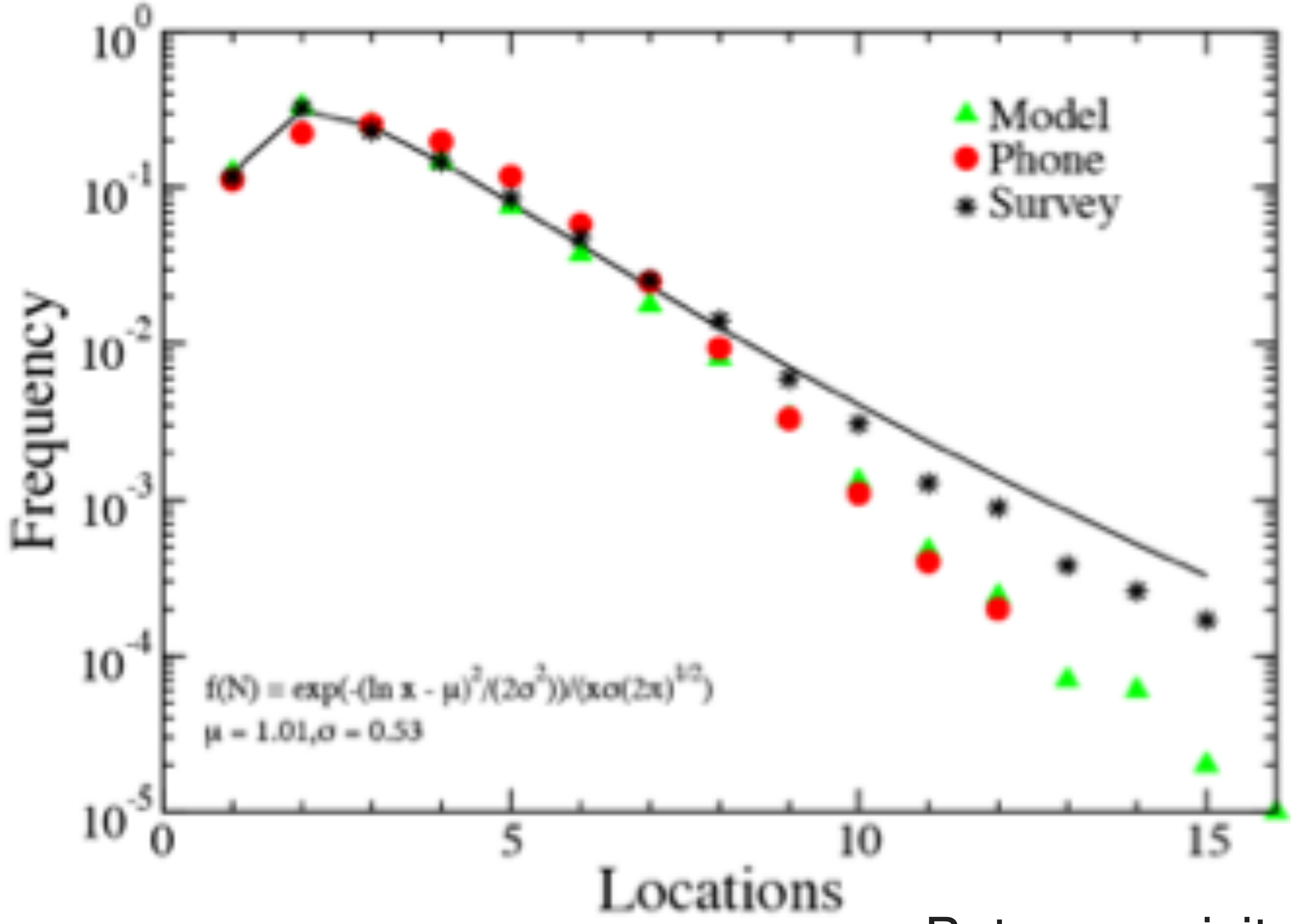
90% of trips can be described by only these 16 patterns



The patterns are the same in different cities.

https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0246

Most people do not visit more than 5 places per day



But some visit a lot

https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0246

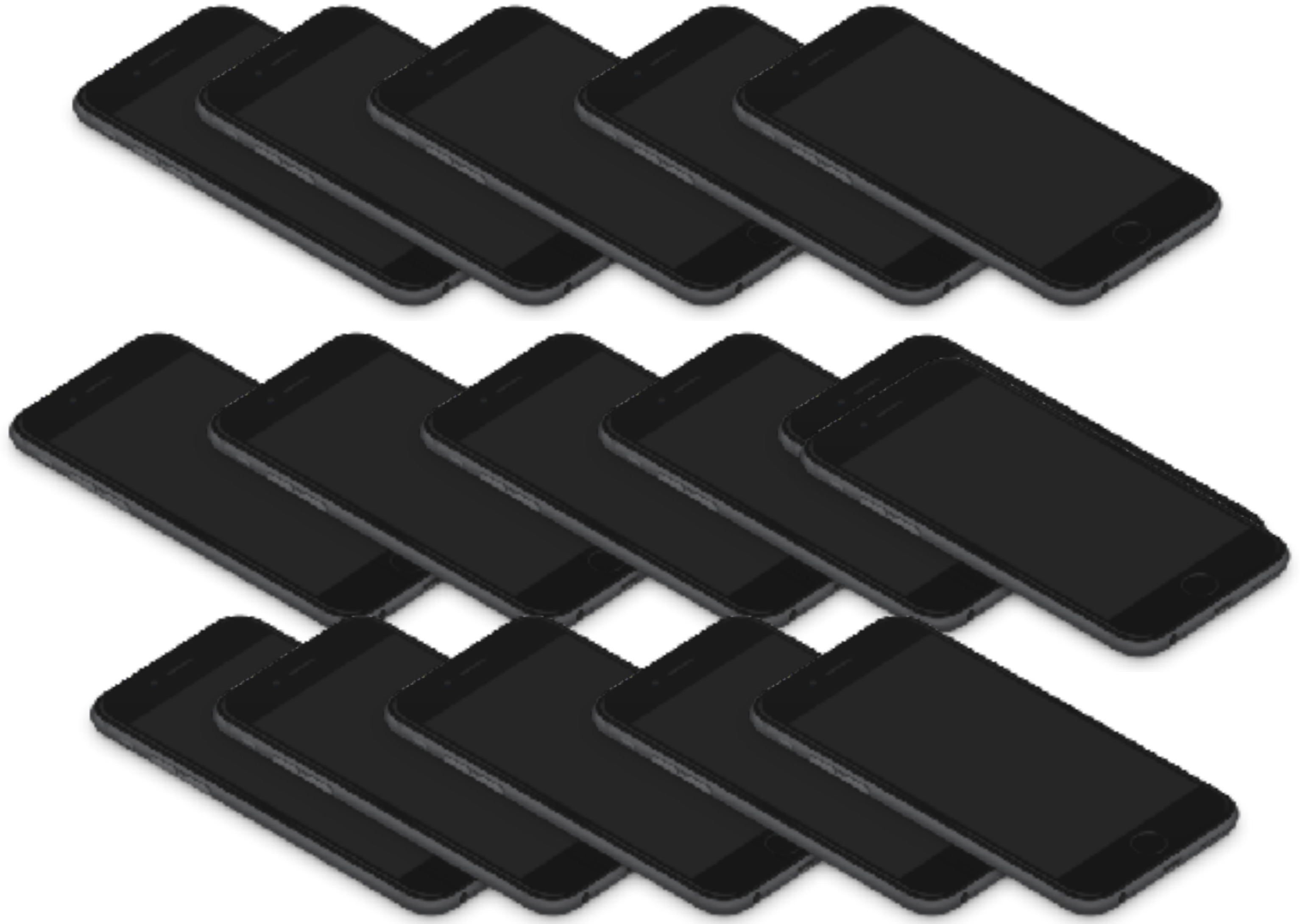
What about our social relations?

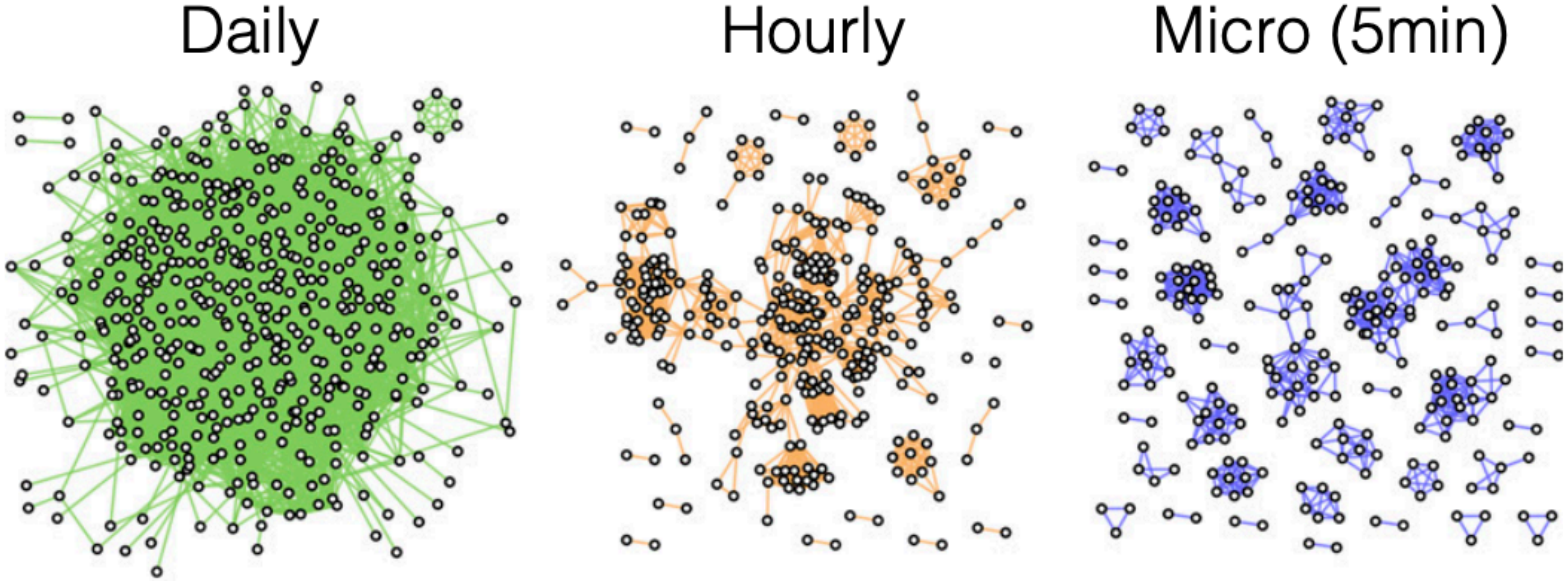
Do we meet the same people over and over?

Copenhagen network study

Tracking 1000 students at DTU

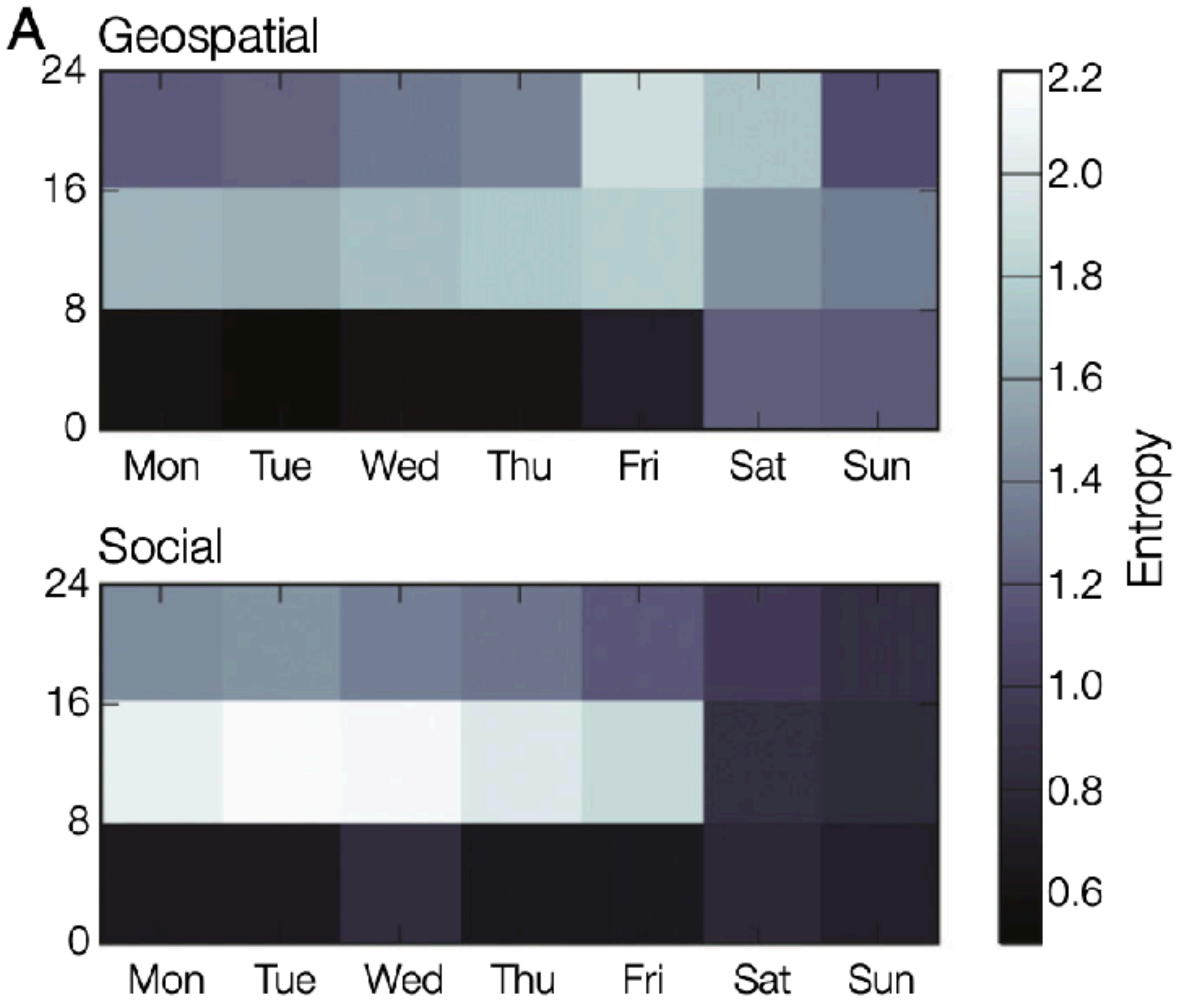
Hi-res data over years





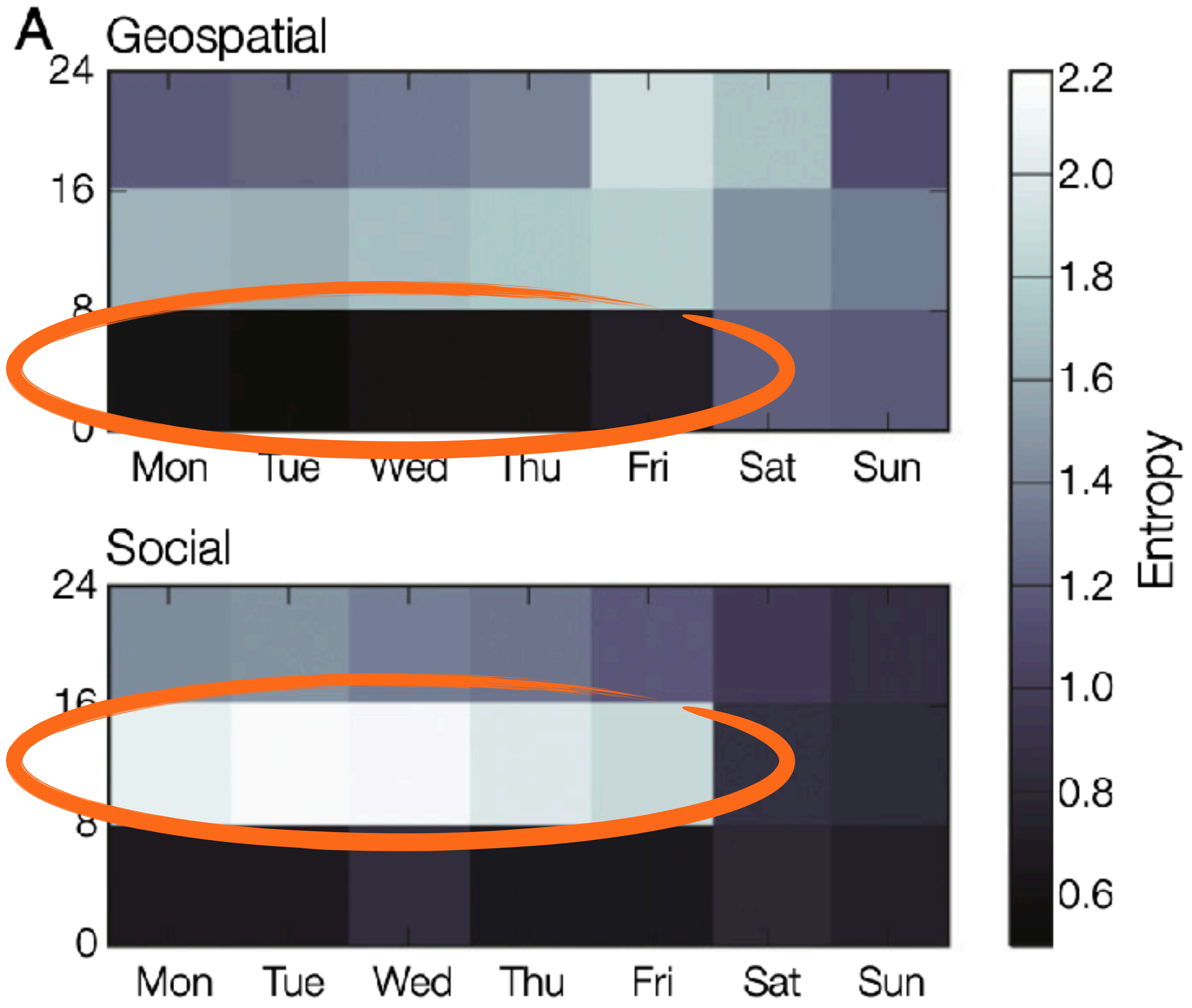
https://www.pnas.org/doi/10.1073/pnas.1602803113

Individual geospatial and social entropy are different



https://www.pnas.org/doi/10.1073/pnas.1602803113

Individual geospatial and social entropy are different

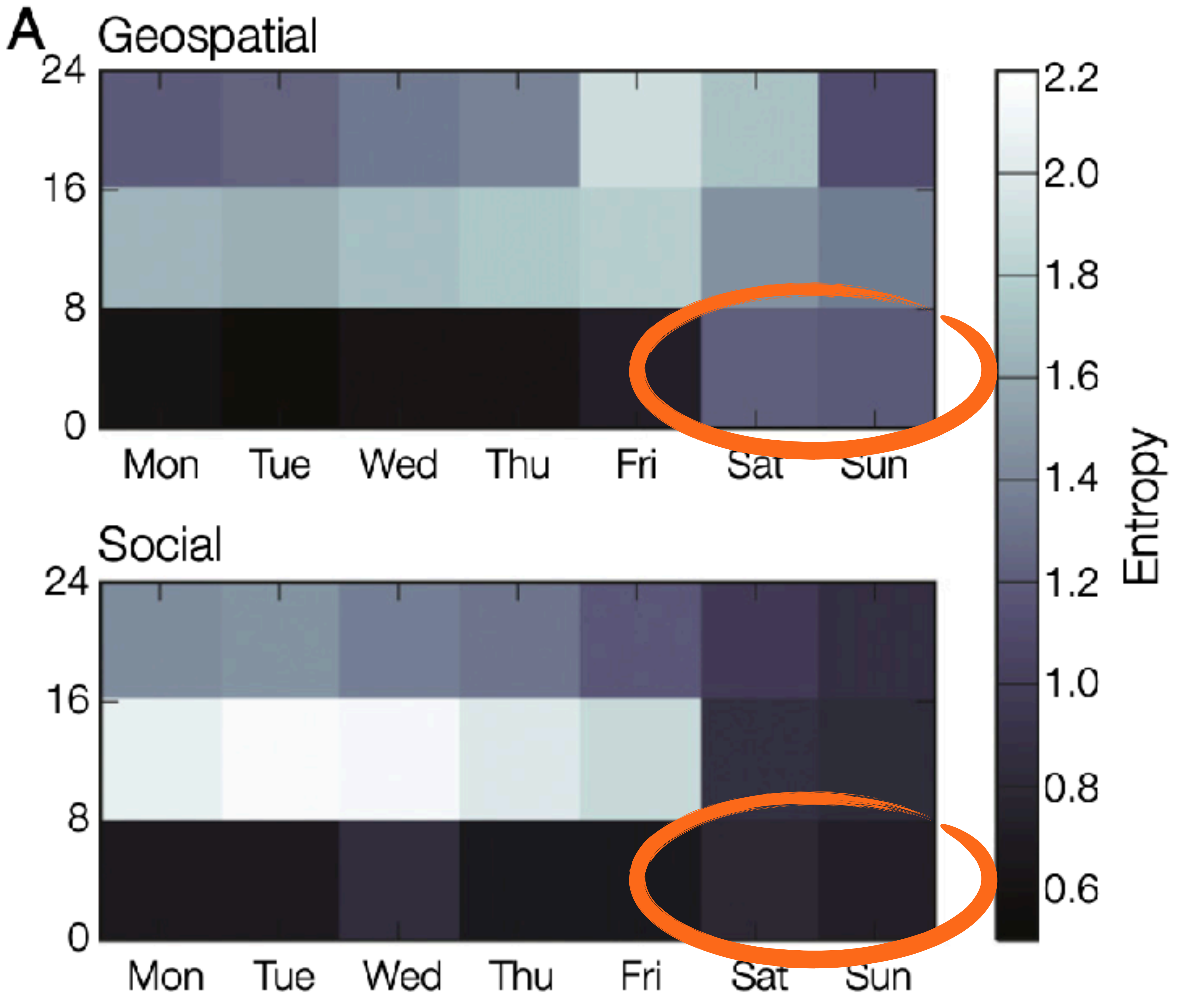


Simple mobility behavior Mo-Fr night (sleeping at home)

Complex social behavior Mo-Fr day (meeting many students)

https://www.pnas.org/doi/10.1073/pnas.1602803113

Individual geospatial and social entropy are different

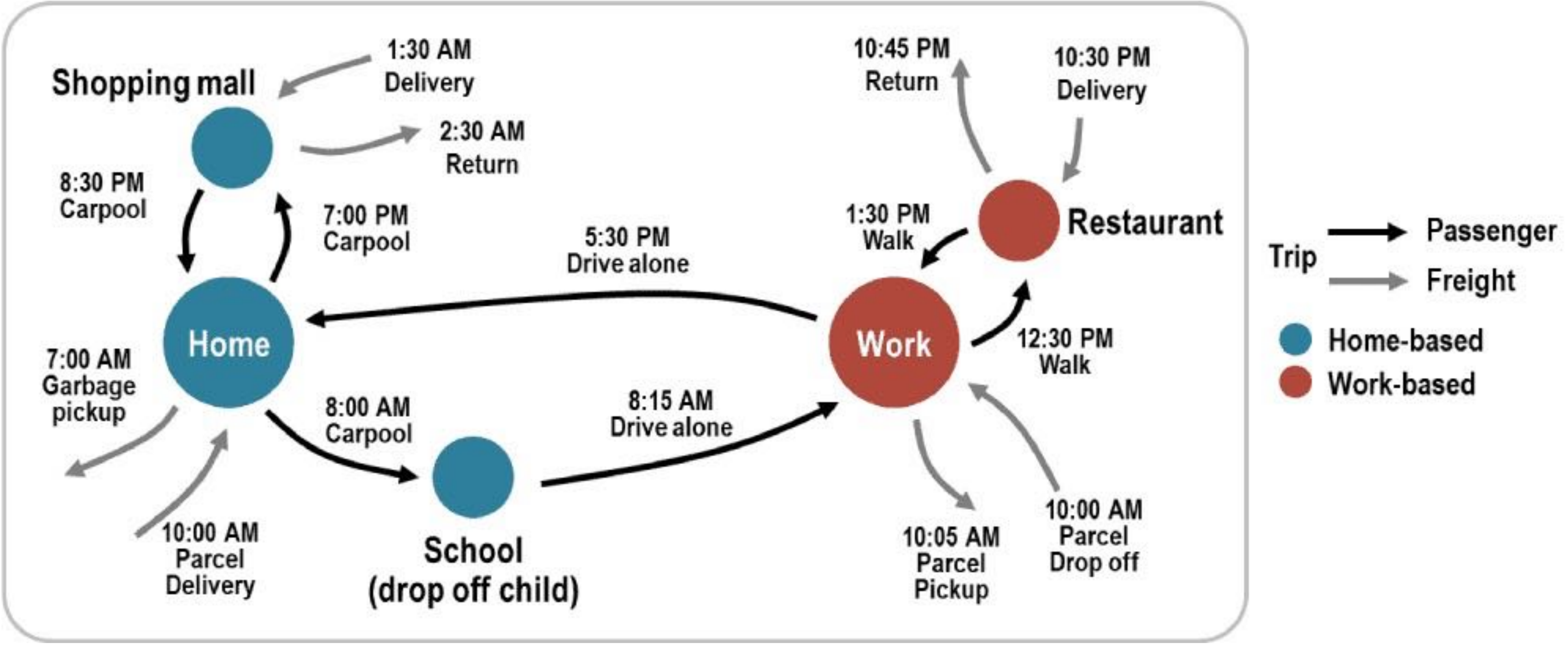


Exploratory mobility behavior Sa-Su night (party)

Conservative social behavior Sa-Su night (same friends)

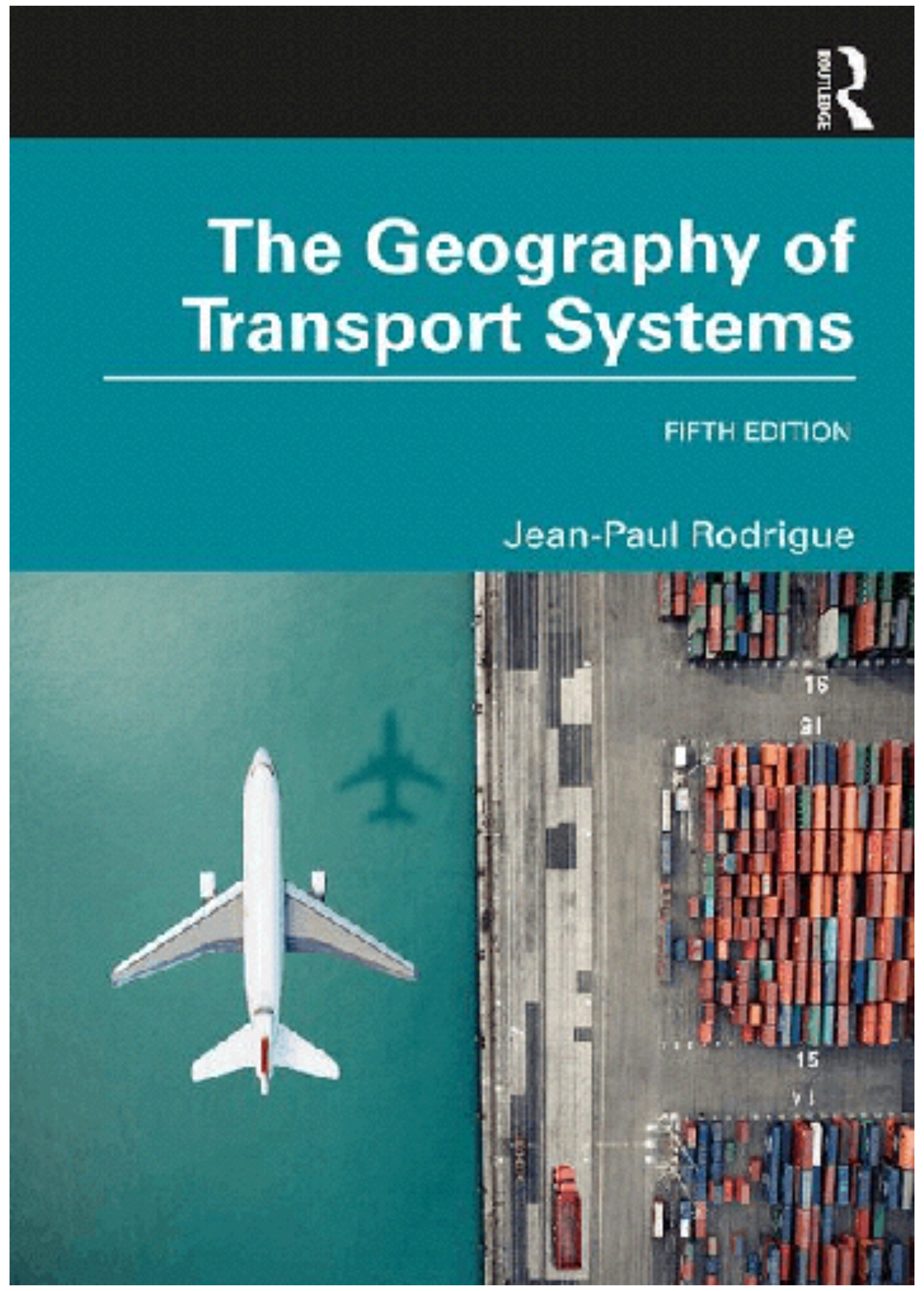
https://www.pnas.org/doi/10.1073/pnas.1602803113

We can show with data that we have very regular lives



But also: A few people have "exciting" lives

Sources and further materials for today's class



Limits of Predictability in Human Mobility

Chaoming Song,[1,2] Zehui Qu,[1,2,3] Nicholas Blumm,[1,2] Albert-László Barabási[1,2*]

Returners and explorers dichotomy in human mobility

Luca Pappalardo[1,2,3,4], Filippo Simini[4,5,6], Salvatore Rinzivillo[1], Dino Pedreschi[1,2], Fosca Giannotti[1] & Albert-László Barabási[3,6,7]

Unique in the Crowd: The privacy bounds of human mobility

Yves-Alexandre de Montjoye[1,2], César A. Hidalgo[1,3,4], Michel Verleysen[2] & Vincent D. Blondel[2,5]

Understanding mobility in a social petri dish

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

Modelling the scaling properties of human mobility

Chaoming Song[1,2†], Tal Koren[1,2†], Pu Wang[1,2†] and Albert-László Barabási[1,2,3*]

Unravelling daily human mobility motifs

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

Fundamental structures of dynamic social networks

Vedran Sekara[a], Arkadiusz Stopczynski[a,b], and Sune Lehmann[a,c,1]

https://alexandrakapp.blog/2022/03/14/privacy-preserving-techniques-and-how-they-apply-to-mobility-data/