

Consiglio Nazionale  
delle Ricerche

# Individual mobility laws and models

# Understanding the laws of individual human mobility

- is there a typical traveling distance?
- can we profile individuals according to their mobility behavior?
- to what extent are humans predictable?
- are there typical mobility motifs?

# Modelling individual human mobility

- What determines the decision to start a trip?
- What determines the choice of the destination?
- What determines the decision to come back home or to explore new locations?
- Can we generate realistic individual trajectories?

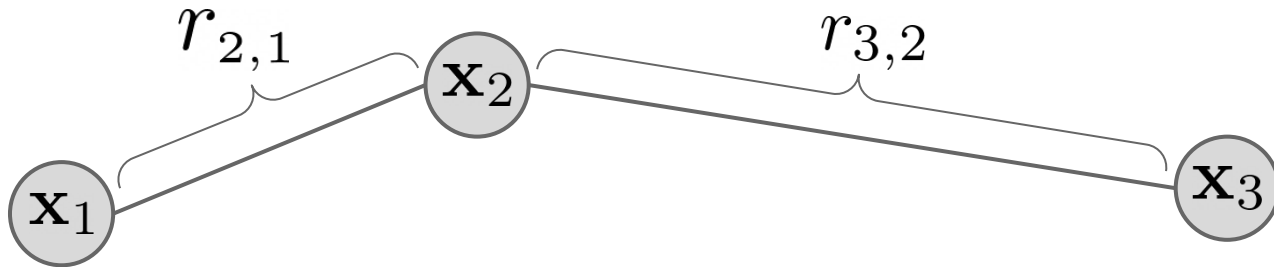
**Distances**

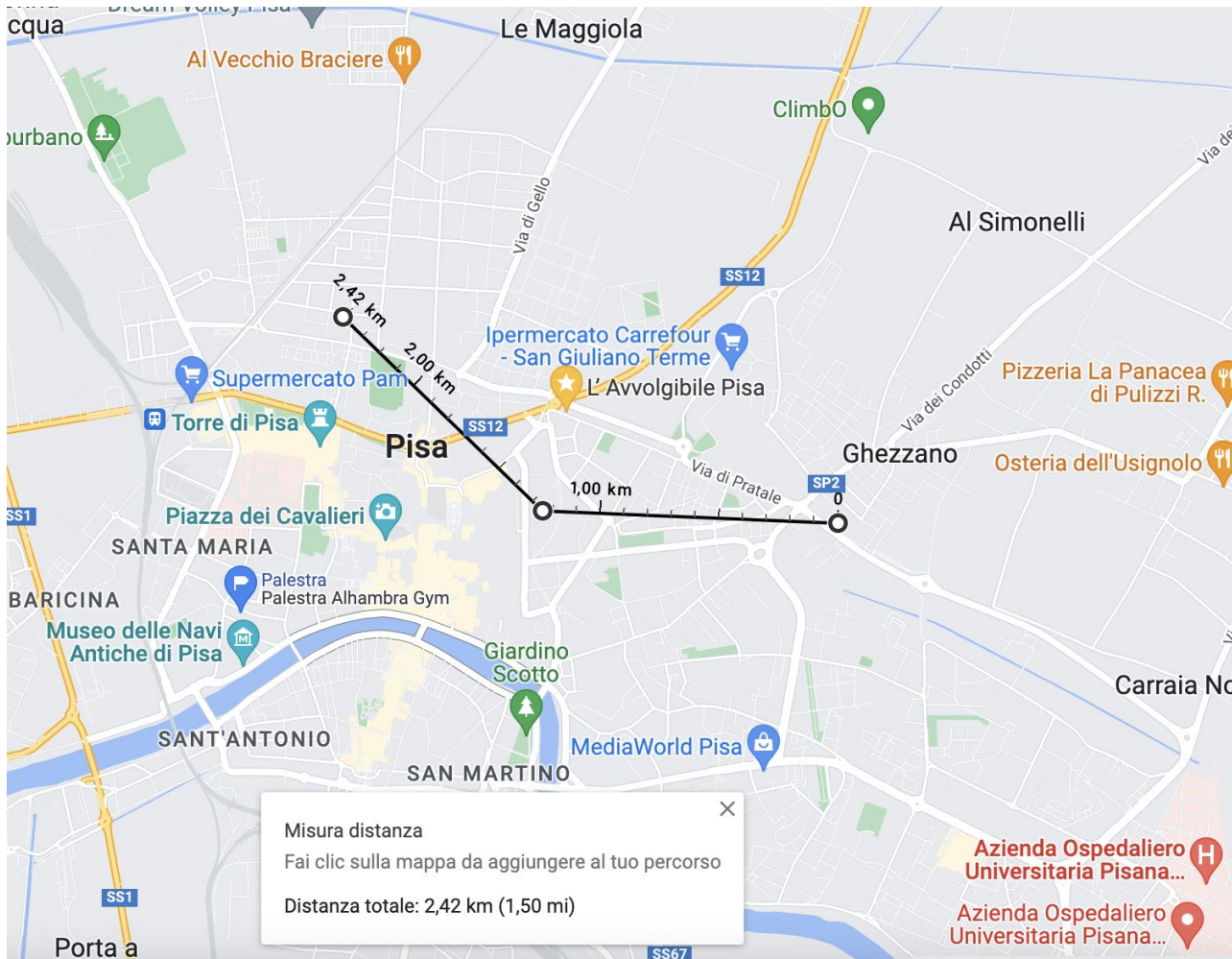
# Travel distance (jump length)

Distance between  
two consecutive locations  
visited by a moving object

$$r = \left| \mathbf{X}_2 - \mathbf{X}_1 \right|$$

Earth distance





Misura distanza  
Fai clic sulla mappa da aggiungere al tuo percorso  
Distanza totale: 2,42 km (1,50 mi)

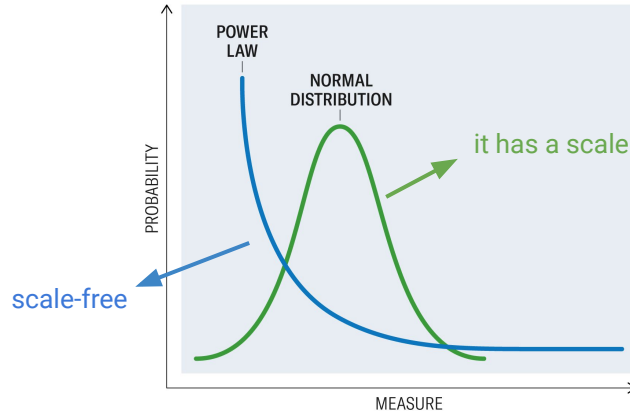
# Travel distance probability

$P(r)$  = probability of finding a trip of length  $r$

## A Pareto Distribution vs. a Gaussian Curve

A normal distribution (i.e., a Gaussian curve) is bell-shaped, whereas a Pareto distribution (i.e., power law) is shaped like a hockey stick with long tails.

What's the shape of this distribution?



# Tracking of dollar bills

The screenshot shows the homepage of WheresGeorge.com. At the top, the URL 'www.WHERESGEORGE.COM' is displayed. Below it is a navigation menu with links for Home, Top 10, Public Forum, Tools/Fun, Join/Login, Store, FAQs/About, and Contact Us. A 'Logon' link is also present. The main heading reads 'Welcome to Where's George?' followed by 'To get started tracking your bills, please select below'. Two circular buttons are shown: 'I found a Where's George? bill, and I want to see where it has been' and 'I want to enter and track my own Where's George? dollar bills'. Below these are two columns of explanatory text. The bottom of the page features a statistics bar with the following data: 309,182,332 Total Bills Entered, \$1,662,473,399 Total Dollar Value, 1,117 Bills Entered Today, and 118 Hits Today. The footer includes copyright information for 2022 WheresGeorge.com LLC and mentions 'CURRENCY TRACKING PROJECT'.

www.WHERESGEORGE.COM

Home Top 10 Public Forum Tools/Fun Join/Login Store FAQs/About Contact Us

Logon

### Welcome to Where's George?

To get started tracking your bills, please select below

**I found a Where's George? bill, and I want to see where it has been**

**I want to enter and track my own Where's George? dollar bills**

By entering this bill, you will help maintain the tracking of the journey this bill has made. Additionally, you will get back a list of all the cities, states, and countries where your bill has been recorded, as well as travel time and distance along its journey.

You can start and enter your own bills to track! You will need to **Register** before entering your own bills so you can be notified when your bills are found next.

[Show Me an Example of a Bill Tracking Report](#)

Portions U.S. Patent #8,682,917

309,182,332 Total Bills Entered    \$1,662,473,399 Total Dollar Value    1,117 Bills Entered Today    118 Hits Today

Please **Logon** or **Register**    Background...    Theme...    v

Hide Status    Copyright © 2022 WheresGeorge.com LLC. All rights reserved. Privacy & Terms    Site designed by [Royce Redfern](#)

**CURRENCY TRACKING PROJECT**

The top image shows a US one-dollar bill with a tracking overlay. The overlay includes the website URL 'www.WHERESGEORGE.com', the serial number 'L 2824 5A', and a red box containing the text 'Where am I?'. A green circular stamp with the text 'THIS IS A REGISTERED BILL' is also visible. The bottom image shows another US one-dollar bill with a tracking overlay. The overlay includes the text 'Track where I go next' and a blue circular stamp with the text 'SEE WHERE I HAVE BEEN'. Both images feature the 'UNITED STATES CURRENCY TRACKING PROJECT' logo.

FEDERAL RESERVE NOTE

THE UNITED STATES OF AMERICA

THIS NOTE IS LEGAL TENDER FOR ALL DEBTS, PUBLIC AND PRIVATE

www.WHERESGEORGE.com

L 2824 5A

WASHINGTON, D.C.

Where am I?

THIS IS A REGISTERED BILL

ONE DOLLAR

IN GOD WE TRUST

Track where I go next

SEE WHERE I HAVE BEEN

www.WHERESGEORGE.com

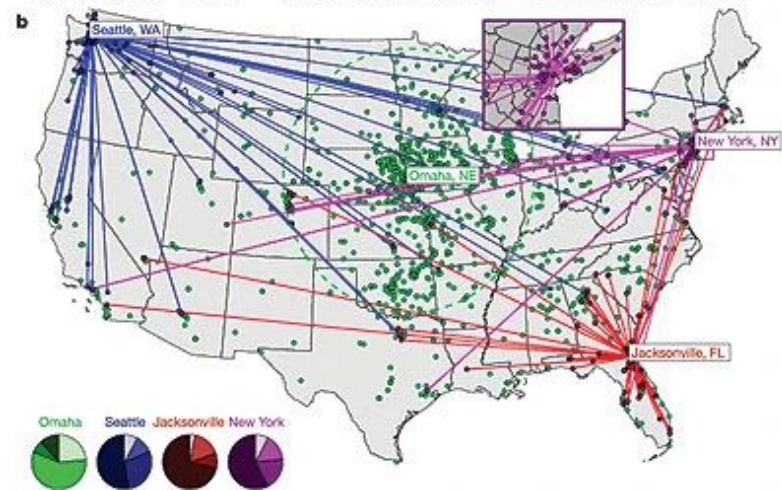
UNITED STATES CURRENCY TRACKING PROJECT



# Tracking of dollar bills

Brockmann et al., 2006:

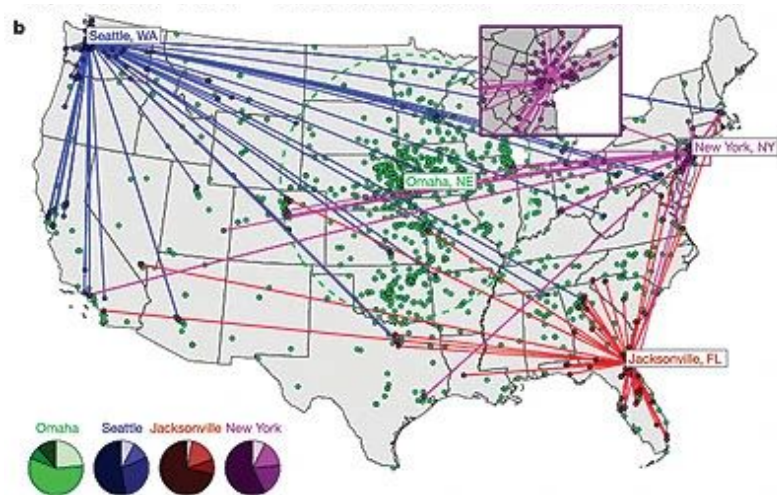
- Dollar bills: 464,670
- Records: 1,033,095
- Area: US  
(excluding Alaska and Hawaii)



Trajectories of bank notes originating from four different places with travelling time  $T < 14$  days.

# Tracking of dollar bills

- Most bank notes are reported close the initial entry,  $r \leq 10km$ 
  - Seattle 53%, NYC 58%, Jacksonville 71%
- A small but **considerable** fraction is reported at large distances,  $r > 800km$ 
  - Seattle 8%, NYC 7%, Jacksonville 3%



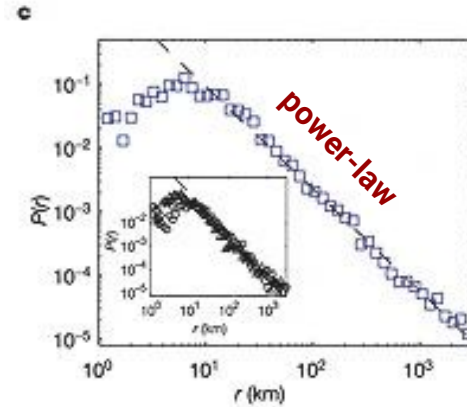
Trajectories of bank notes originating from four different places with travelling time  $T < 14$  days.

# Tracking of dollar bills

- Probability of traversing a distance in 1-4 days (20,540 bills)

$$P(r) \sim r^{-(1+\beta)}$$

$$\beta = 0.59 \pm 0.02$$

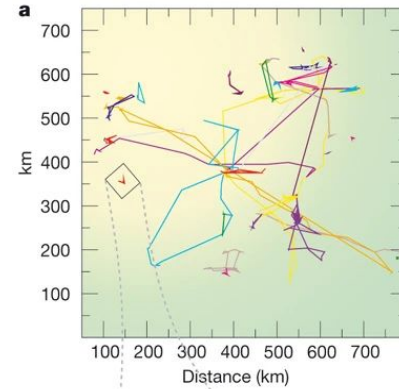


Measured  $P(r)$  of traversing a distance in less than  $T = 4$  days. The inset shows  $P(r)$  for metropolitan areas, cities of intermediate size, small towns.

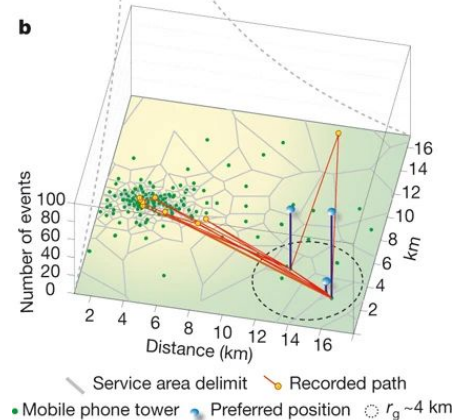
# Mobile Phone Records

González et al., 2008:

- Dataset D1 (CDRs):
  - Users: 100,000
  - Records: 16,264,308
- Dataset D2 (CPRs):
  - Users: 206
  - Records: 10,407



Week-long trajectory of 40 mobile phone users

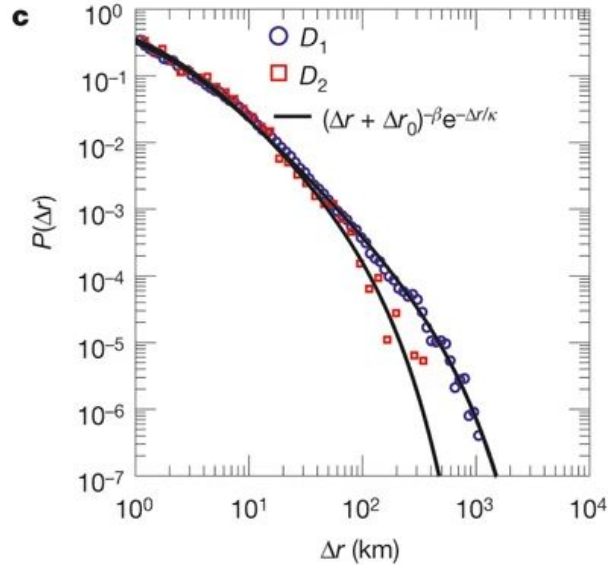


Detailed trajectory of a single user

- 186 two-hourly reports
- 12 locations.

The circle represents the radius of gyration centred in the user's centre of mass.

# Mobile Phone Records



$$P(r) = (r + r_0)^{-\beta} \exp(-r/\kappa)$$

$$\beta = 1.75 \pm 0.15$$

$$r_0 = 1.5 \text{ km}$$

$$\kappa_{D_1} = 400 \text{ km}$$

$$\kappa_{D_2} = 80 \text{ km}$$

Measured  $P(r)$  of travel distances obtained for D1 and D2. The solid line indicates a truncated power law.

# Radius of gyration

Characteristic distance  
of an individual

$$r_g(u) = \sqrt{\frac{1}{n_u} \sum_{i=1}^{n_u} (\mathbf{r}_i - \mathbf{r}_{cm})^2}$$

Center of mass

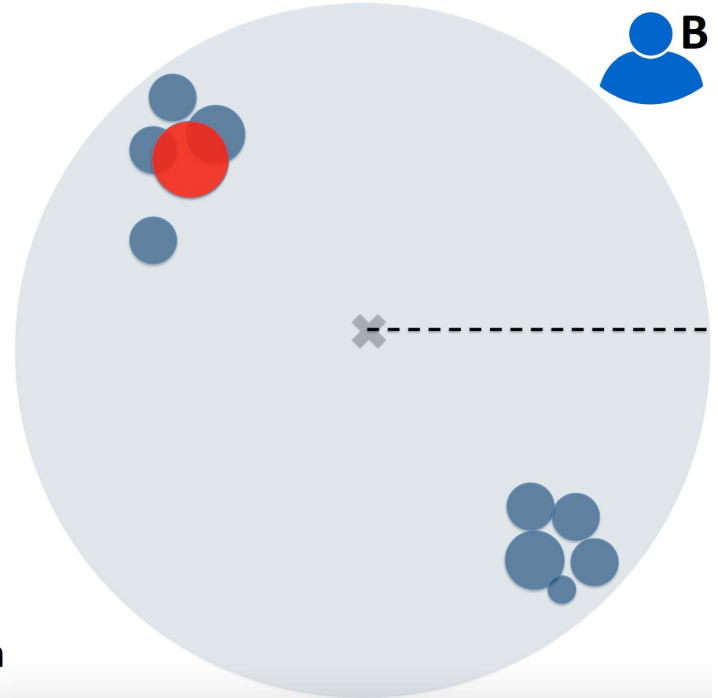
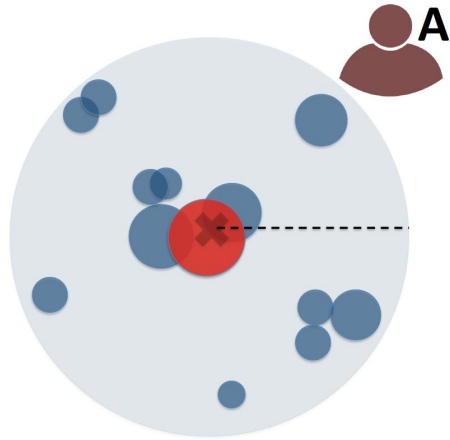
$$\mathbf{r}_{cm} = \frac{1}{n_u} \sum_{i=1}^{n_u} \mathbf{r}_i$$

$n_u$	number of records
$\mathbf{r}_i$	position

# Radius of gyration

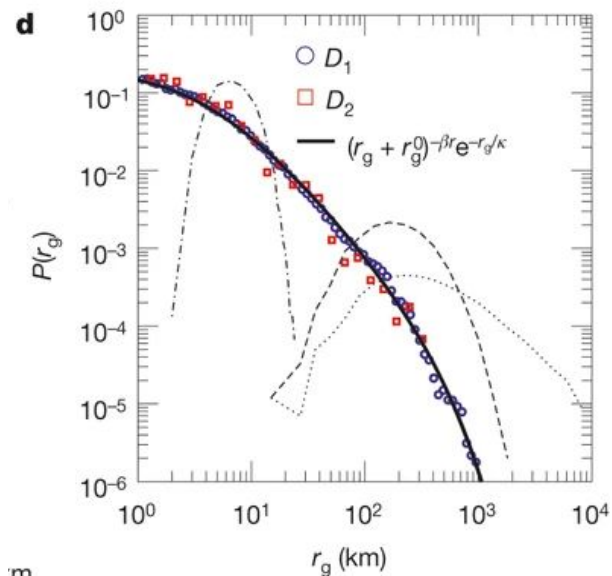
● home location

✕ center of mass



---- radius of gyration

# Radius of gyration



$$P(r_g) = (r_g + r_g^0)^{-\beta_r} \exp(-r_g/\kappa)$$

$$r_g^0 = 5.8 \text{ km}$$

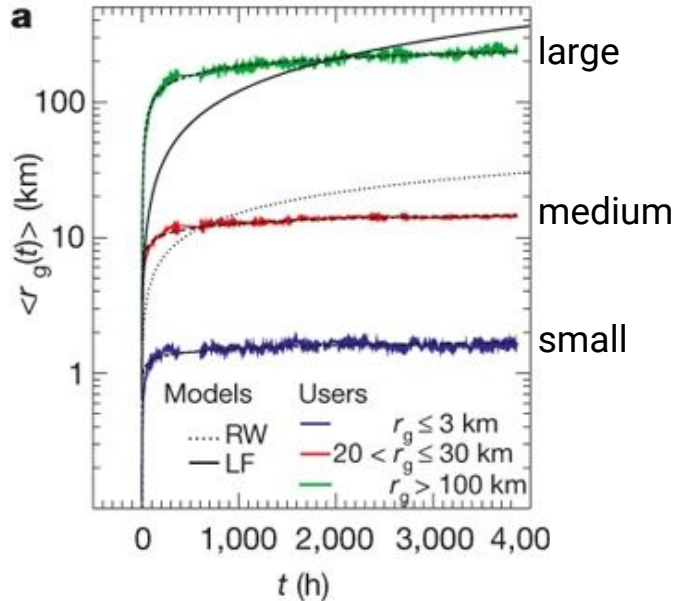
$$\beta_r = 1.65 \pm 0.15$$

$$\kappa = 350 \text{ km}$$

Measured  $P(r_g)$  on datasets  $D_1$  and  $D_2$ . The dotted, dashed and dot-dashed curves show  $P(r_g)$  obtained from null models



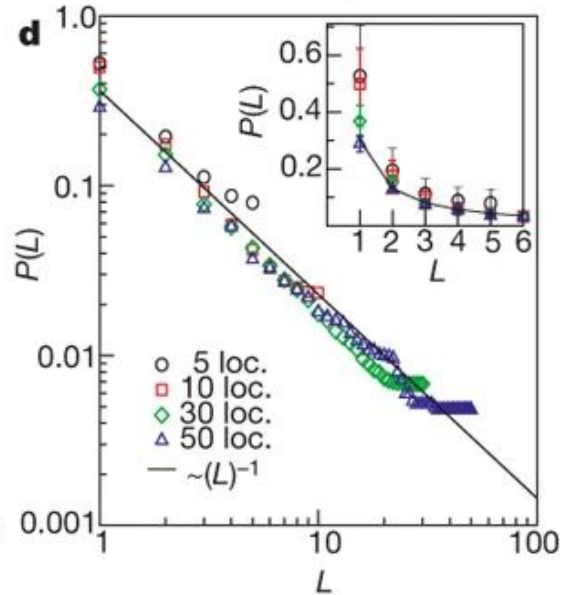
# Radius of gyration - time evolution



- Radius increases logarithmically with time
- Indicating a saturation process

Radius of gyration versus time for users of three groups. The black curves correspond to the analytical predictions for the random walk models. The dashed curves corresponding to a logarithmic fit.

# Location frequency



- Rank each location based on how many times an individual is recorded there
  - E.g.,  $L=3$  is the third-most-visited location for an individual

$$P(L) \sim 1/L$$

Frequency of visiting locations for users observed to visit 5, 10, 30 and 50 locations.  $L$  is the rank of the location listed in the order of the visit frequency. 40% of the time individuals are found at their first two preferred locations

People devote most of their time to a few locations, spending their time to places with diminished regularity

# k-radius of gyration

Recurrent  
characteristic distance  
of an individual

$$r_g^{(k)} = \sqrt{\frac{1}{N_k} \sum_{i=1}^k w_i (\mathbf{r}_i - \mathbf{r}_{cm}^{(k)})^2}$$

k-center of mass

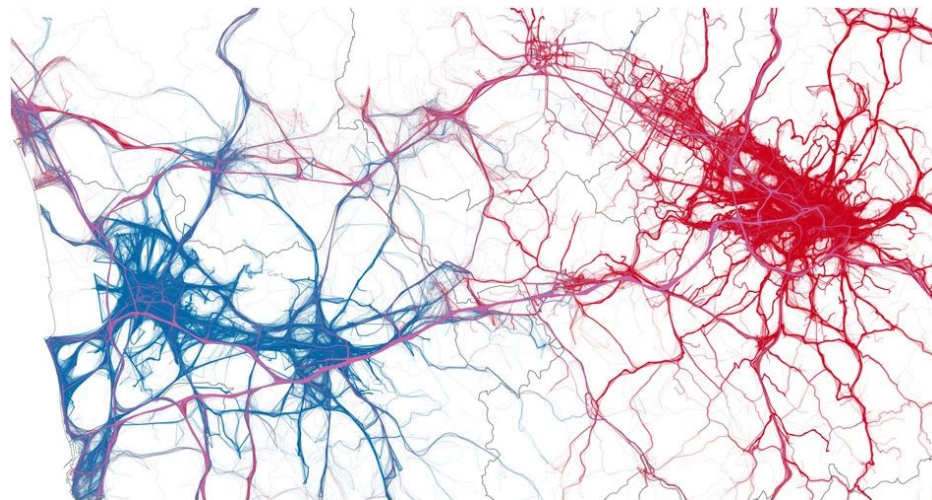
$$\mathbf{r}_{cm}^{(k)} = \frac{1}{N_k} \sum_{i=1}^k w_i \mathbf{r}_i$$

$N_k$  number of records  
in location  $k$

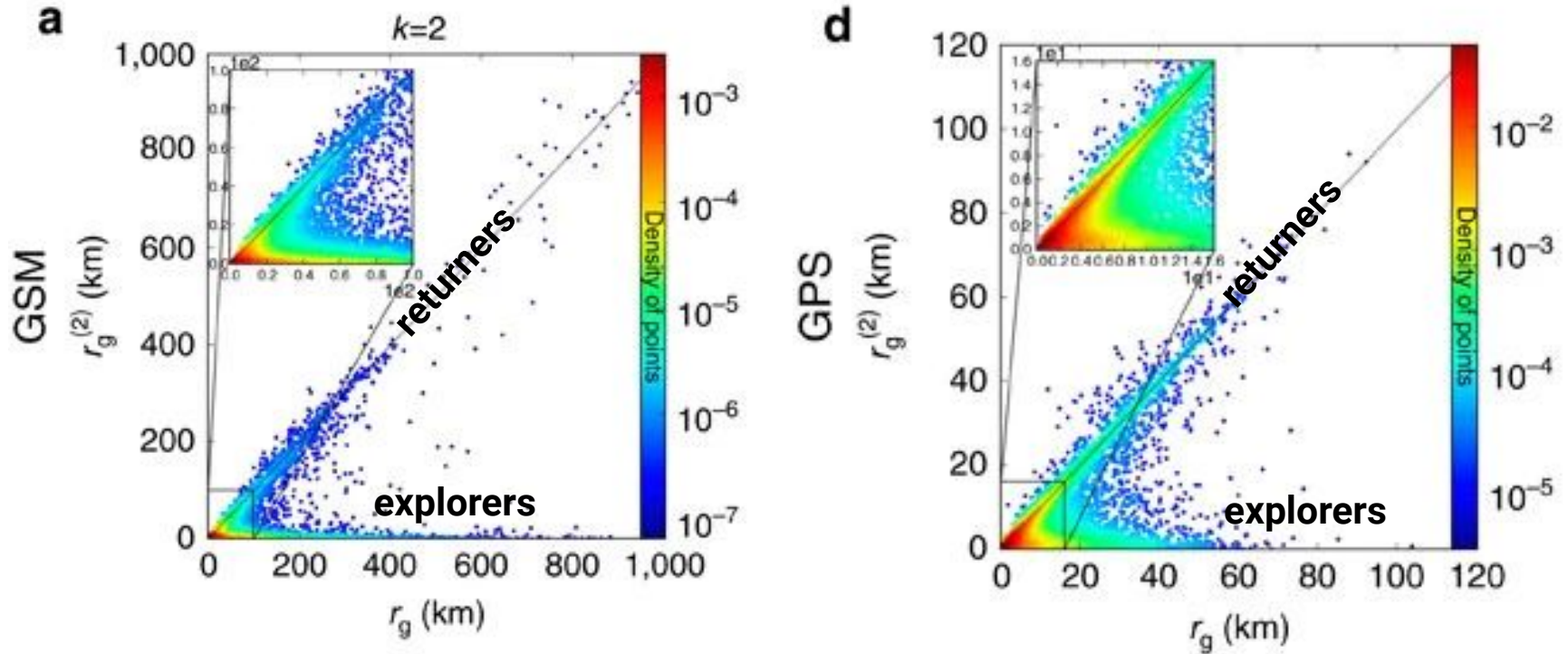
# Mobile Phone Records

Pappalardo et al., 2015:

- CDRs:
  - Users: 67,000
- GPS traces:
  - Users: 46,000



# Returns and Explorers



Correlation between total  $r_g$  and  $r_g^{(k)}$  for  $k=2, 4, 8$  for CDRs and GPS traces. Each point is coloured from blue to red, indicating the density of points in the corresponding region.

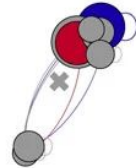
# Returners and Explorers

2-Returners

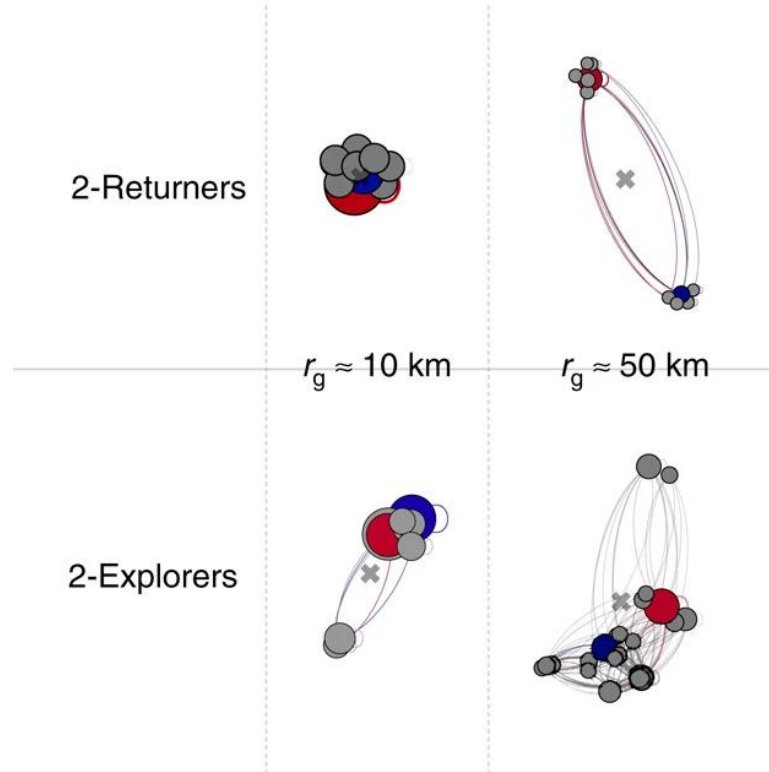


$r_g \approx 10$  km

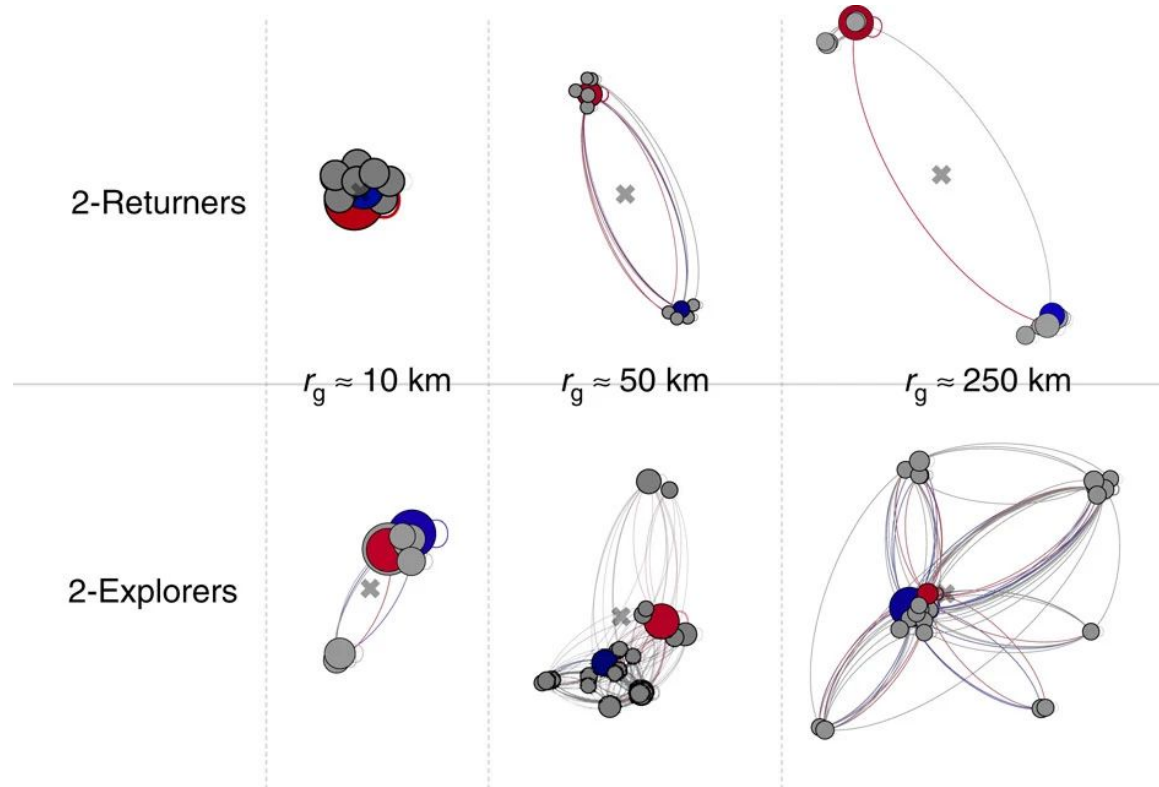
2-Explorers



# Returners and Explorers



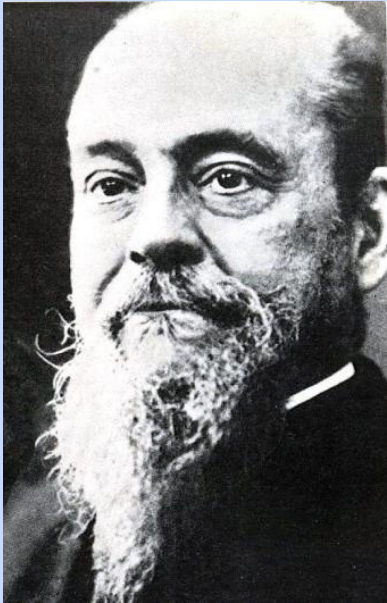
# Returns and Explorers





## INTERVALLO

# Vilfredo Pareto and the 80/20 rule



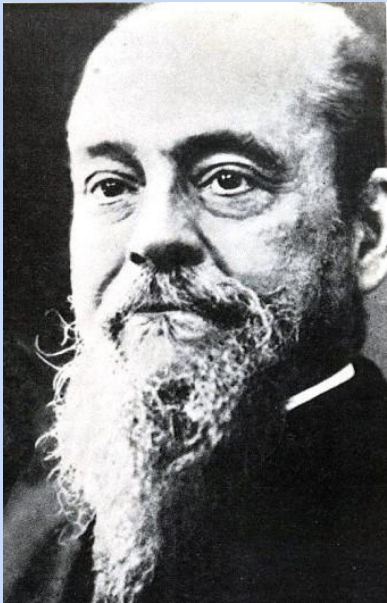
He noticed that in Italy a few wealthy individuals earned most of the money, while the majority of the population earned rather small amounts.

He connected this disparity to the observation that **incomes follow a power law**, representing the first known report of a power-law distribution.

**The 80/20 rule:** Roughly 80 percent of money is earned by only 20 percent of the population.

# INTERVALLO

## Vilfredo Pareto and the 80/20 rule

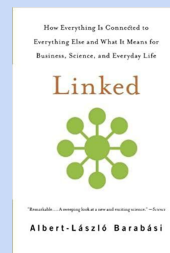


The 80/20 rule emerges in many areas:

- 80% of profits are produced by 20% of the employees
- 80% of decisions are made during 20% of meeting time
- 80% of links on the Web point to only 15% of webpages
- 80% of citations go to only 38% of scientist
- 80% of links in Hollywood connected to 30% of actors
- **The 1% phenomena:**
  - In the US, 1% of the population earns 15% of the total income
  - signature of income disparity, it is a consequence of the power-law nature of the income distribution

# References

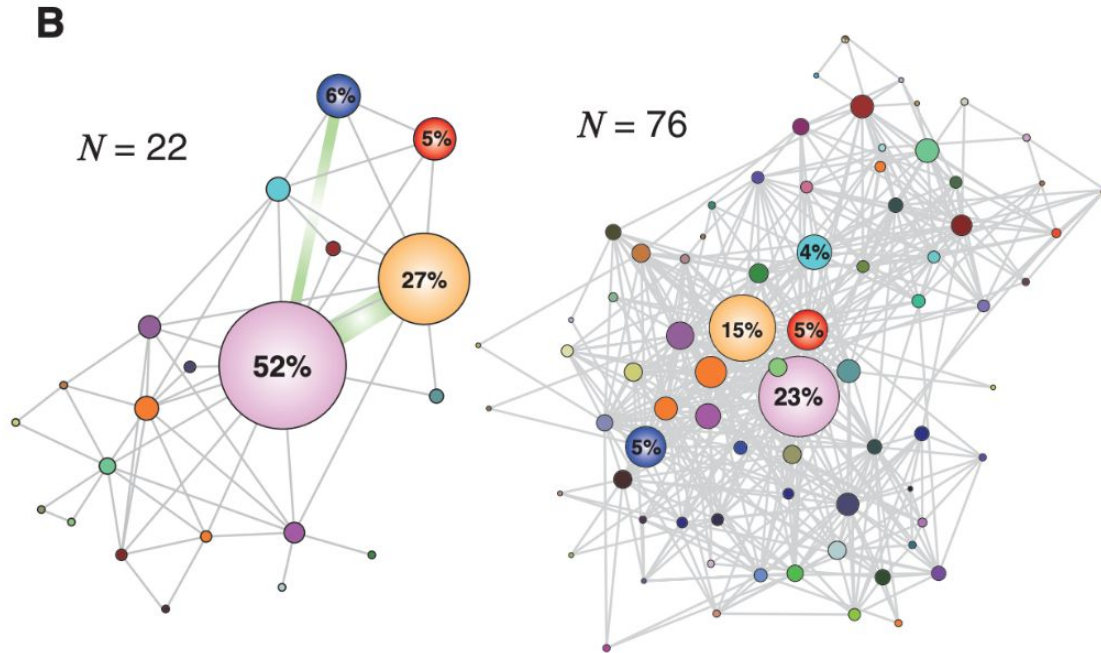
- [article] [We Need to Let Go of the Bell Curve](#), Harvard Business Review, 2022
- [article] [Visualizing power-law distributions](#), Capital as Power, 2019
- [book] [Linked: the New Science of Networks](#), A.-L. Barabasi
- [book] [Chi troppo chi niente](#), E. Ferragina



**Predictability**

# Individual Mobility Network

A network where nodes are an individual's visited locations and edges movements between locations



## **The role of randomness**

1. What is the role of randomness in human mobility?
2. To what degree are our movements predictable?

# Entropy

Random entropy

$$S^{rand} = \log_2 N$$

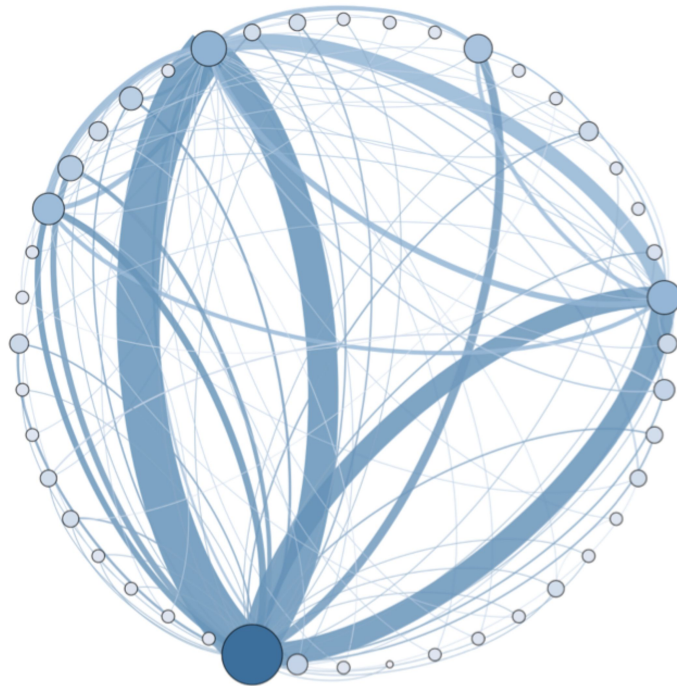
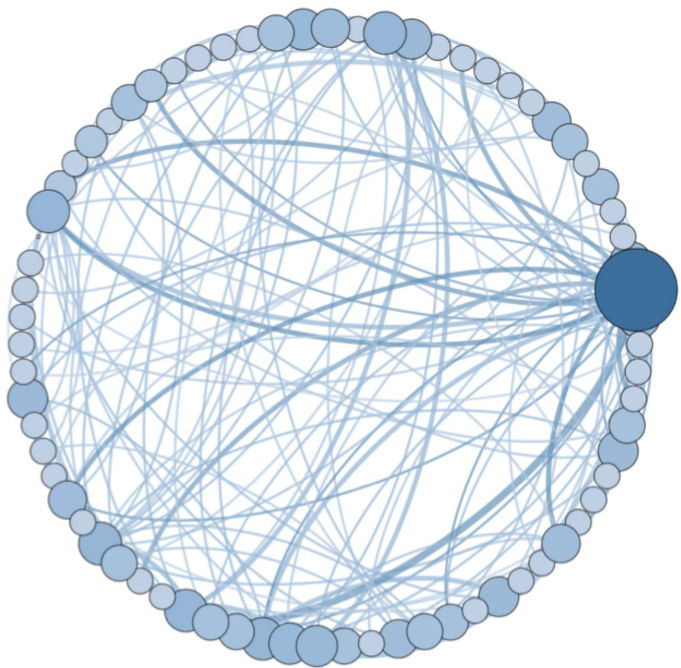
Uncorrelated entropy

$$S^{unc} = - \sum_{i=1}^n p_i \log_2 p_i$$

Real entropy

$$S = - \sum_{T'_i \subset T_i} p_{T'_i} \log_2 p_{T'_i}$$

# Who's the most predictable?



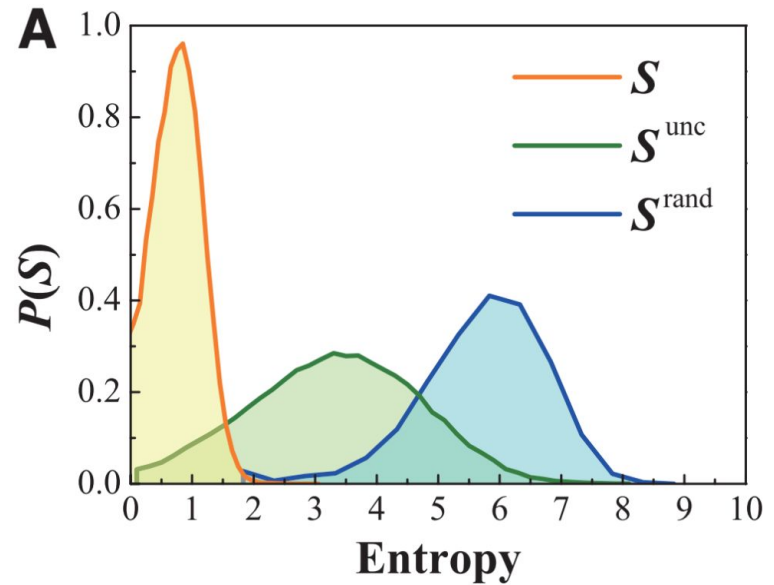


# Entropy

Song et al., 2010:

- 50,000 users (CDRs)
- $S$  peaks at 0.8

$$2^{0.8} = 1.74$$

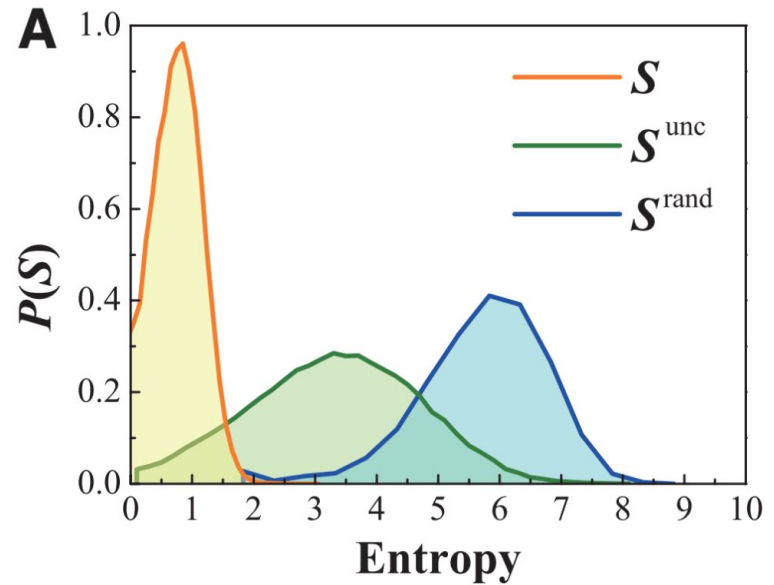


# Entropy

Song et al., 2010:

- 50,000 users (CDRs)
- $S$  peaks at 0.8

$$2^{0.8} = 1.74$$

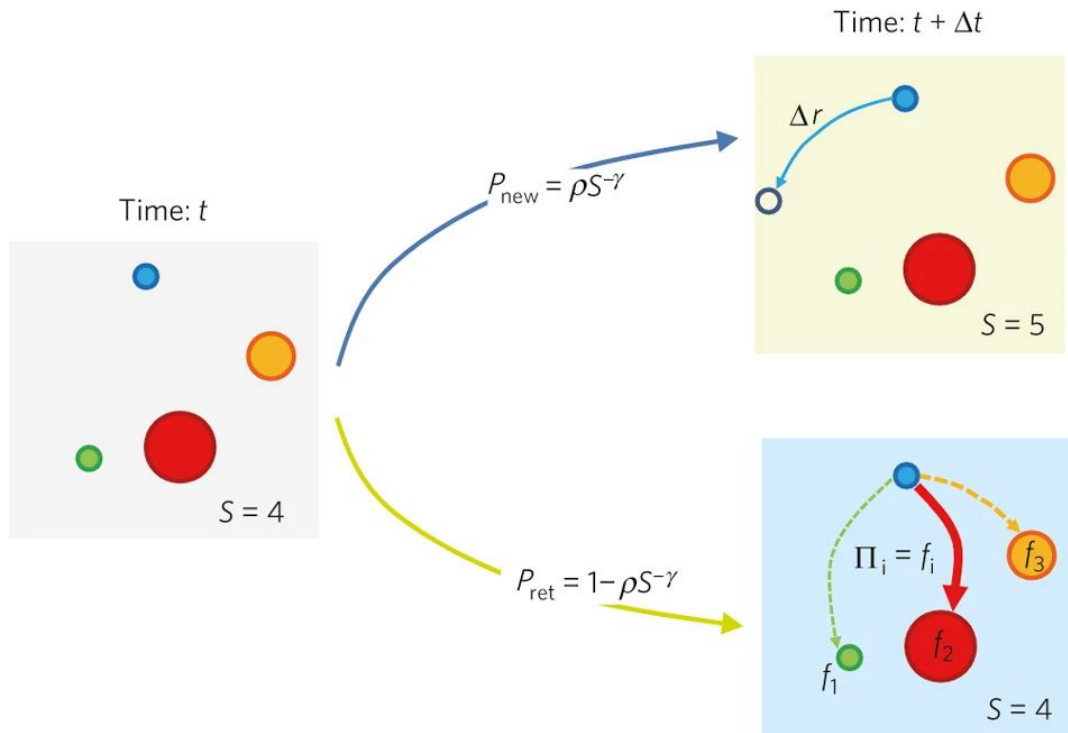


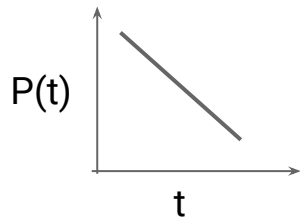
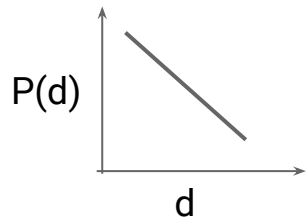
# References

- [paper] [Human Mobility: Models and Applications](#), Barbosa et al., Physics Report, 2018, Section 3.1.1
- [paper] [The scaling laws of human travel](#), Brockmann et al., Nature, 2006
- [paper] [Understanding individual human mobility patterns](#), Gonzalez et al., Nature, 2008
- [paper] [Returners and Explorers Dichotomy in Human Mobility](#), Pappalardo et al., Nature Communications, 2015
- [paper] [Limits of Predictability in Human Mobility](#), Song et al., Science 2010

# **Modelling Individual Human Mobility**

# Exploration and Preferential Return Model (EPR)

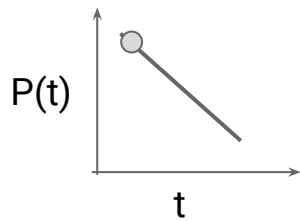
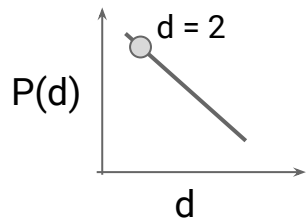




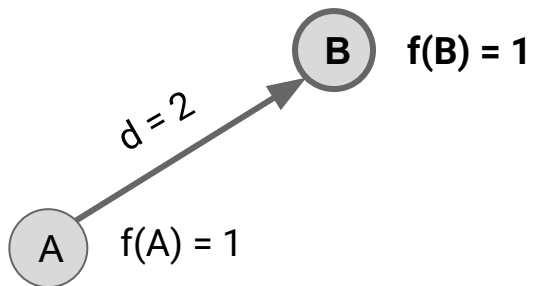
Location	Time	Event
A	0	

**A**  $f(A) = 1$

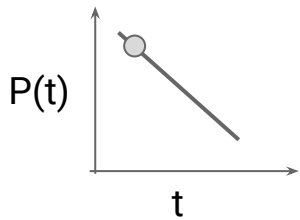
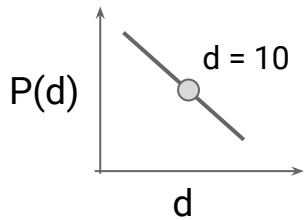
$S = 1$



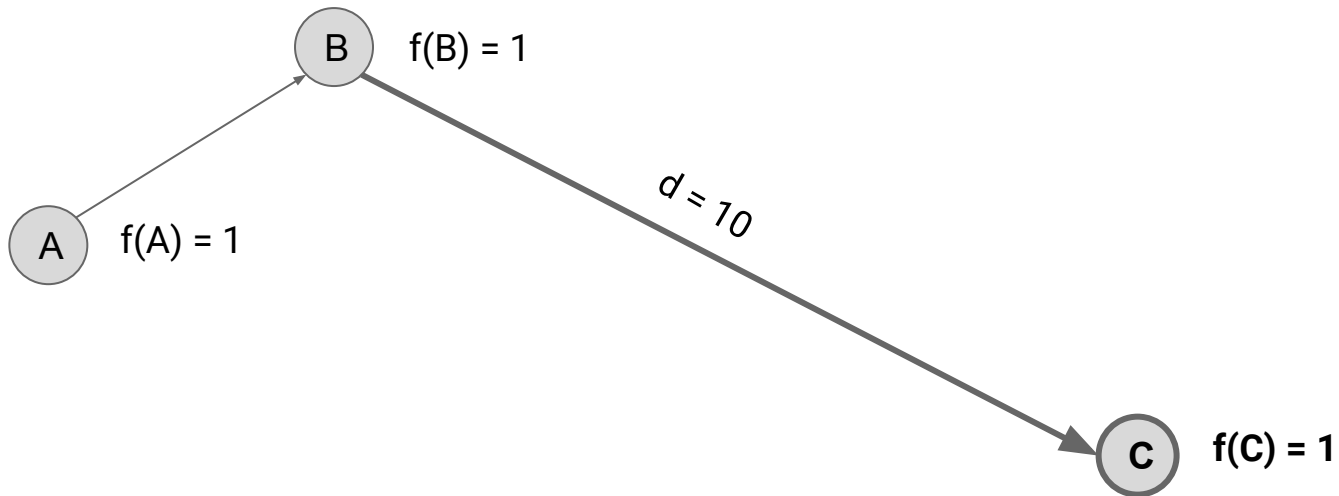
Location	Time	Event
A	0	
B	2	Explore



$S = 2$

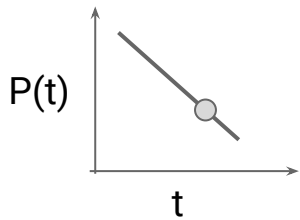
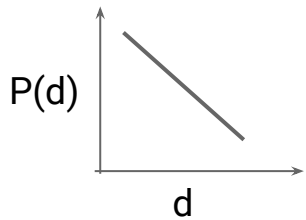


Location	Time	Event
A	0	
B	2	Explore
<b>C</b>	<b>5</b>	<b>Explore</b>

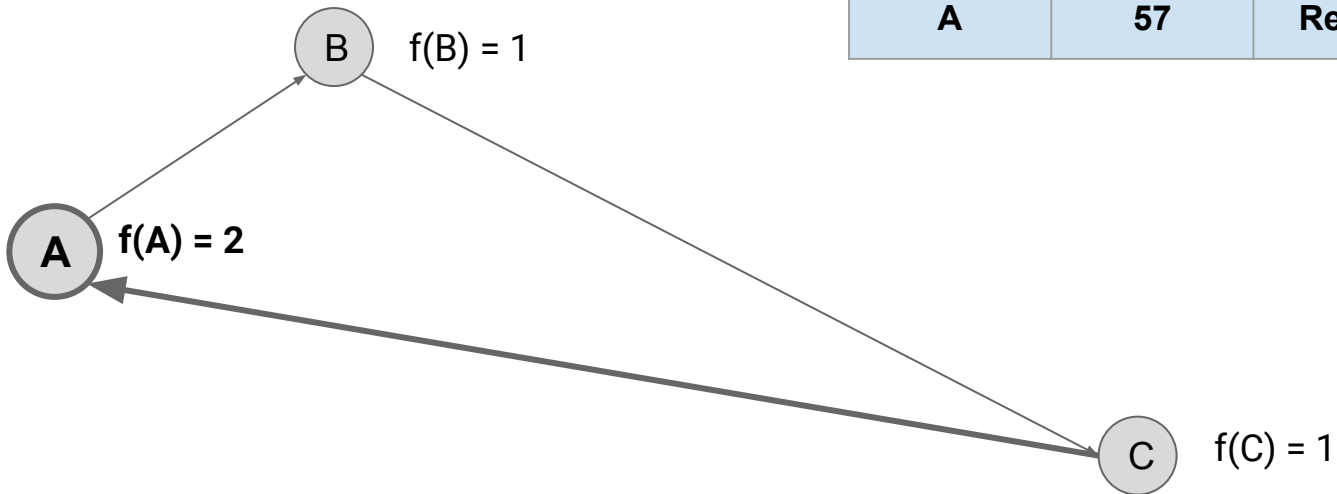


$S = 3$

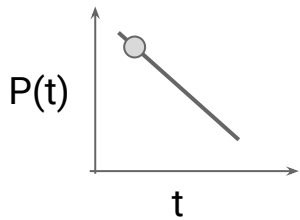
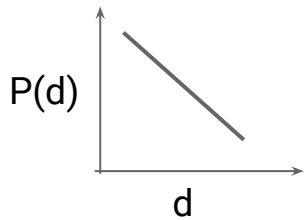




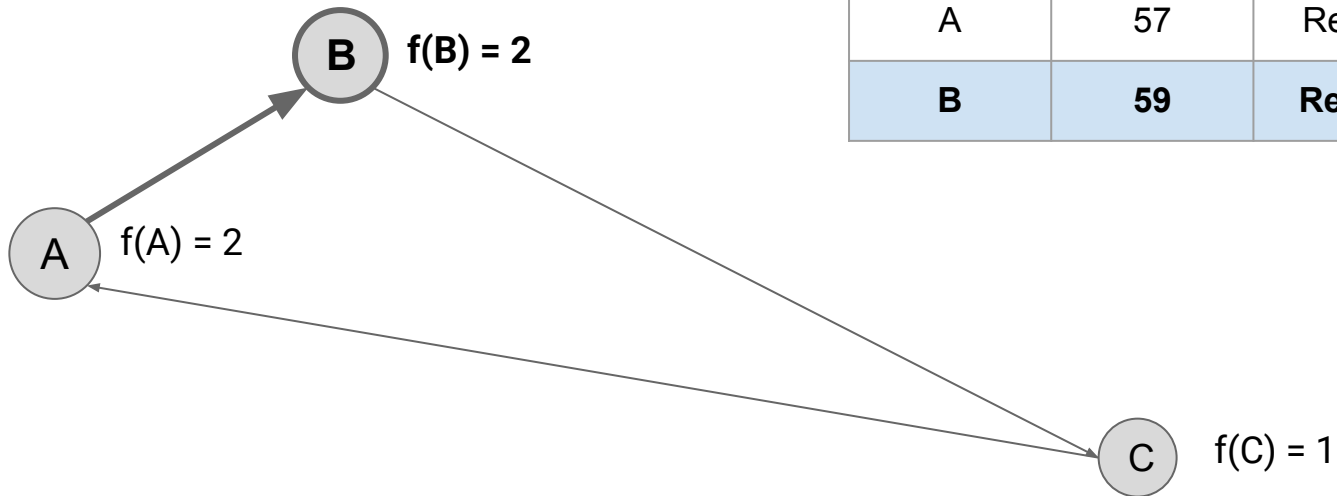
Location	Time	Event
A	0	
B	2	Explore
C	5	Explore
<b>A</b>	<b>57</b>	<b>Return</b>



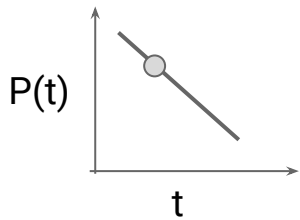
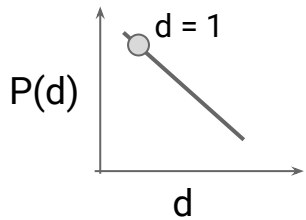
$S = 3$



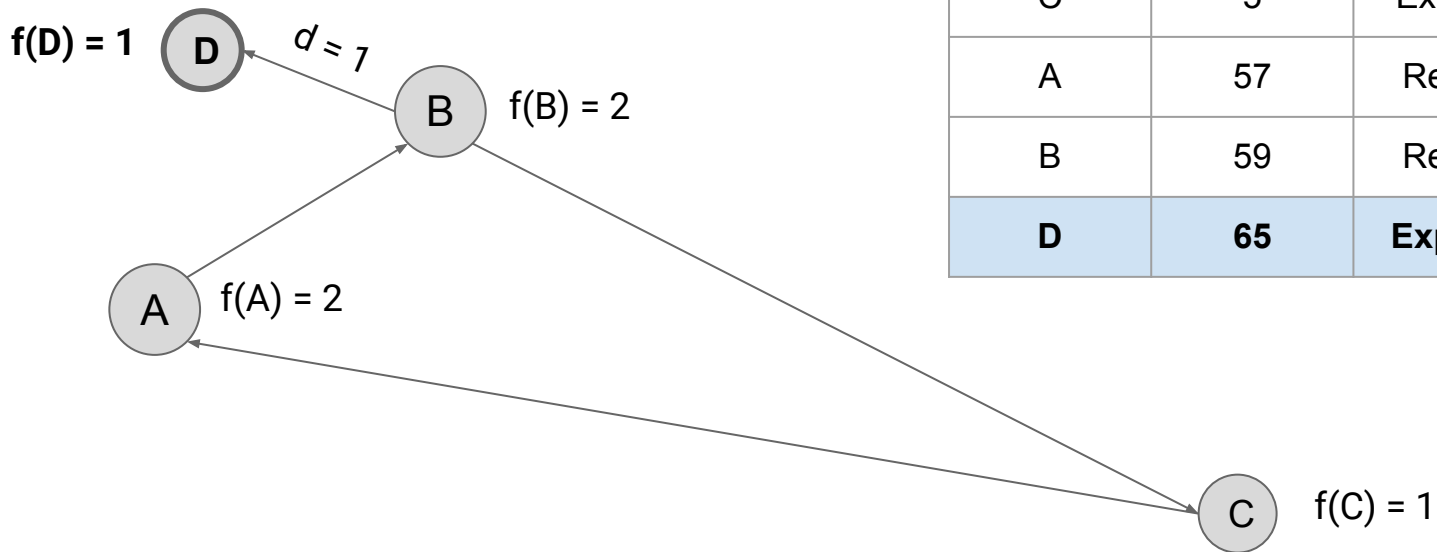
Location	Time	Event
A	0	
B	2	Explore
C	5	Explore
A	57	Return
<b>B</b>	<b>59</b>	<b>Return</b>



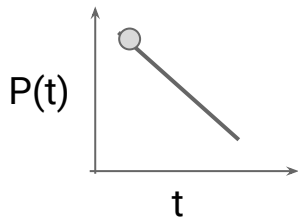
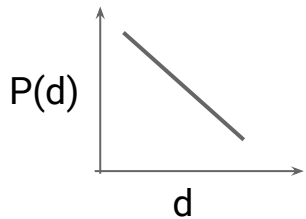
$S = 3$



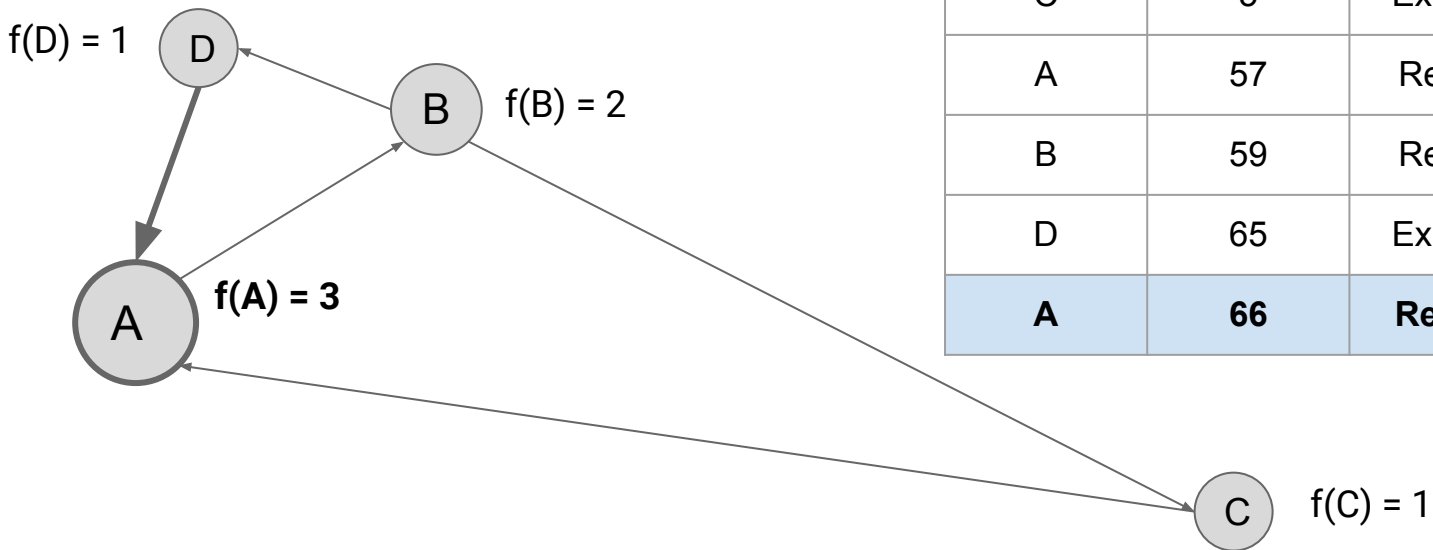
Location	Time	Event
A	0	
B	2	Explore
C	5	Explore
A	57	Return
B	59	Return
<b>D</b>	<b>65</b>	<b>Explore</b>



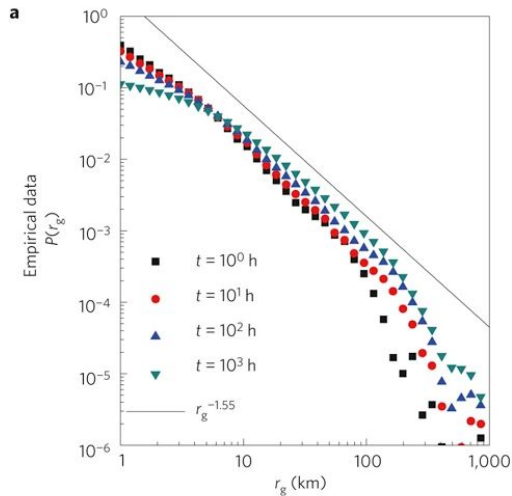
$S = 4$



Location	Time	Event
A	0	
B	2	Explore
C	5	Explore
A	57	Return
B	59	Return
D	65	Explore
<b>A</b>	<b>66</b>	<b>Return</b>

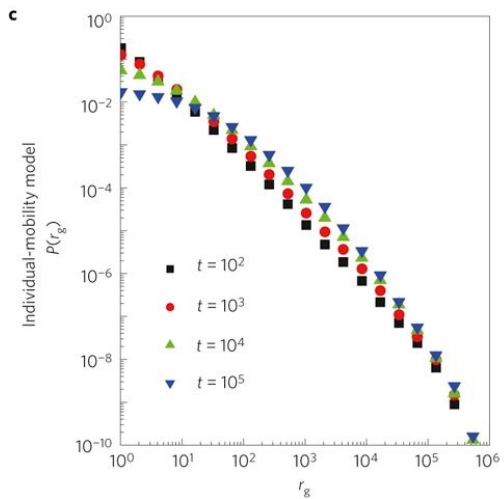
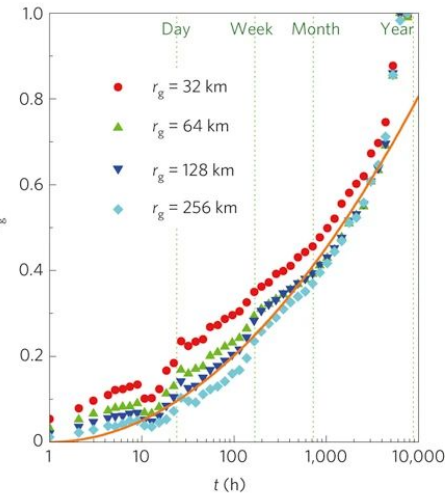


$S = 4$



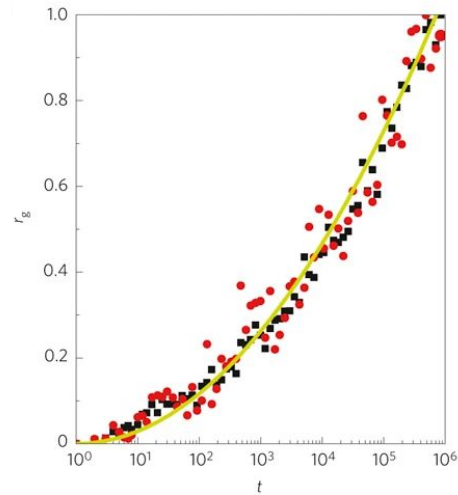
$P(r_g)$  for the mobile-phone users at different moments of time

The probability to return to a previously visited location as a function of the previous frequency of visitation  $f$



$P(r_g)$  for the EPR model using  $\alpha=0.75$ ,  $\beta=0.6$ ,  $\gamma=0.2$  and  $\rho=0.1$ , the values found to be of direct relevance to human mobility

Results for the EPR model



# References

- [paper] [Modelling the scaling properties of human mobility](#), Song et al., Nature Physics, 2010
- [paper] [Human Mobility: Models and Applications](#), Barbosa et al., Physics Report, 2018, Section 4.1

# Homeworks

to be delivered by Friday, October 21st, 2022



## Homework 5.1

Use `skmob` to compute the radius of gyration of all users in the [Brightkite dataset](#). Make a plot that shows the distribution of the radius of gyration over the population of users.

- Visualize in `folium` the  $r_{cm}$  and  $r_g$  for the top-10 users with the highest  $r_g$
- Compute the home location (HL) of each of these users, compute the distance between each user's HL and  $r_{cm}$
- Redefine  $r_g$  so that it is based on HL instead of  $r_{cm}$ , call it  $r_{g,h}$
- Visualize in `folium` HL and  $r_{g,HL}$  for each of the top-10 users with the highest  $r_g$
- Do the shapes of  $r_g$  and  $r_{g,HL}$  overlap? Compute the overlapping area using `shapely/geopandas`
- Submit a well-commented notebook



## Homework 5.2

Use the [Gowalla dataset](#) to estimate the overall popularity (i.e., number of visits) of each location in the dataset

- Plot the distribution of the locations' popularity. What's the shape of the distribution? Comment on it.
- Compute (using `skmob`) the uncorrelated location entropy of each location, plot its distribution.
- Show if there is a correlation between popularity and location entropy: Are more popular locations also the most "entropic" ones? Provide your interpretation of the result you get
- Repeat for the [Brightkite dataset](#)
- Submit a well-commented notebook

## Homework 5.3

Use the [Gowalla dataset](#) to compute the individual mobility networks (IMNs) of each user in the dataset

- Visualize the IMNs of the 1) top-10 individuals and 2) bottom-10 individuals based on their uncorrelated entropy.
- Extract proper network measures from the IMN of each user in the dataset (e.g., average clustering coefficient, average degree, number of nodes, etc.)
- Group individuals by this set of features, using the clustering algorithm you think is the most appropriate
- How many clusters do you find? Characterise and visualize the cluster medoids
- Repeat for the [Brightkite dataset](#)
- Submit a well-commented notebook

## Homework 5.4

Download your positions from [Google Maps](#). Plot the corresponding GPS trajectory. Plot the distribution of jump length.

- What's the shape of your jump length distribution? Comment.
- Compute your rg, and plot it in folium together with the center of mass
- What the distance between your home location and your center of mass?
- Repeat the steps above selecting only points in 2020. What's the difference between your overall rg and that during 2020?
- Compare your 2-rg and your overall rg. Are you a returner o an explorer? Comment on it
- Submit a well-commented notebook