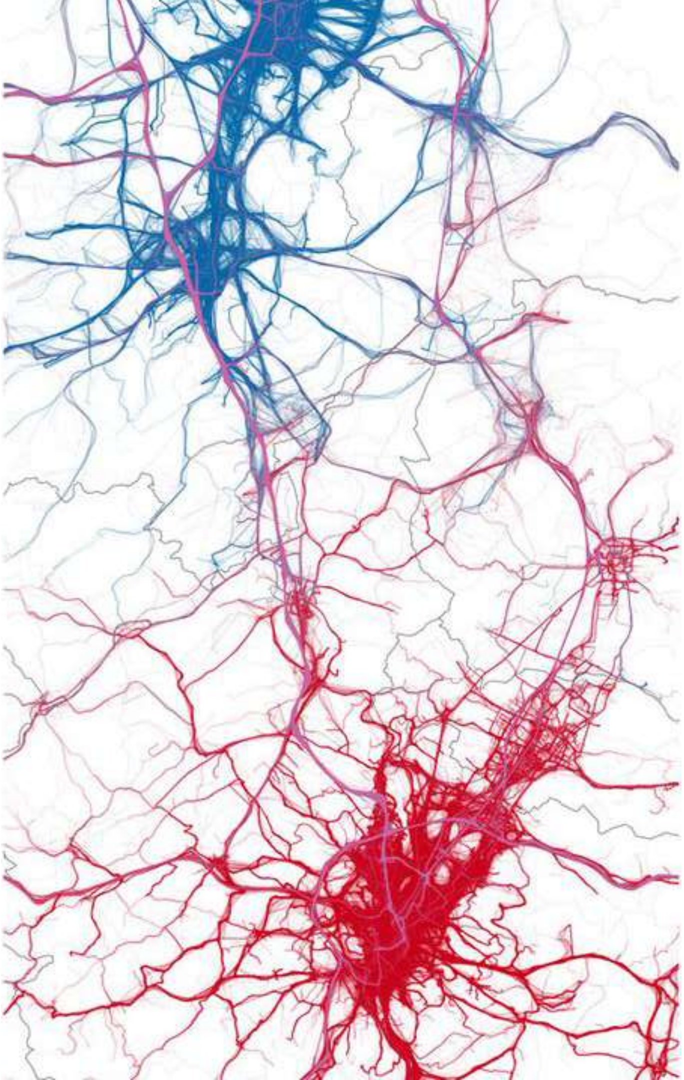


IN SUPREMÆ DIGNITATIS 1343

Consiglio Nazionale delle Ricerche

SCUOLA NORMALE SUPERIORE

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?

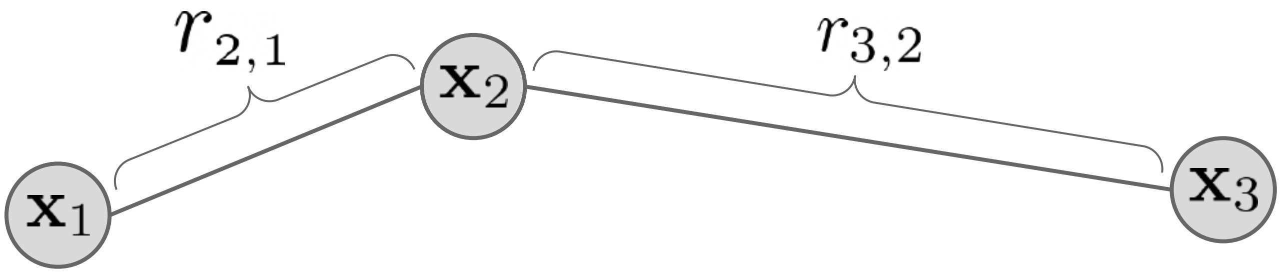
Distances

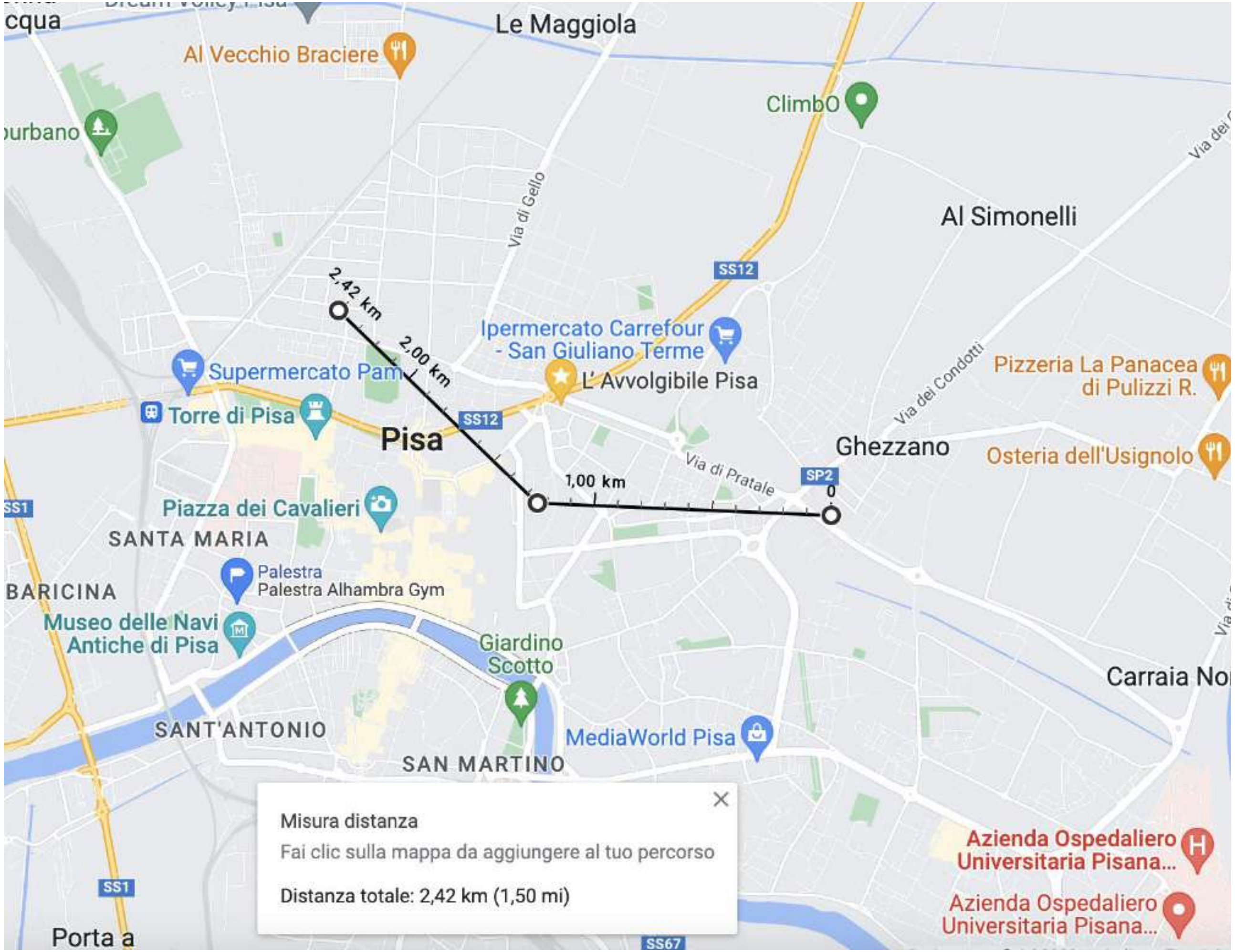
Travel distance (jump length)

Distance between
two consecutive locations
visited by a moving object

$$r = |\mathbf{x}_2 - \mathbf{x}_1|$$

Earth distance





Le Maggiola
Al Vecchio Braciere
ClimbO
Al Simonelli
SS12
2,42 km
2,00 km
1,00 km
0
Ipermercato Carrefour - San Giuliano Terme
L'Avvolgibile Pisa
Supermercato Pam
Torre di Pisa
Pisa
Via di Gello
Via di Pratale
Via dei Condotti
SP2
Ghezzano
Pizzeria La Panacea di Pulizzi R.
Osteria dell'Usignolo
Piazza dei Cavalieri
SANTA MARIA
BARICINA
Palestra
Palestra Alhambra Gym
Museo delle Navi Antiche di Pisa
Giardino Scotto
SANT'ANTONIO
SAN MARTINO
MediaWorld Pisa
SS1
SS67
Azienda Ospedaliero Universitaria Pisana...
Azienda Ospedaliero Universitaria Pisana...
Misura distanza
Fai clic sulla mappa da aggiungere al tuo percorso
Distanza totale: 2,42 km (1,50 mi)

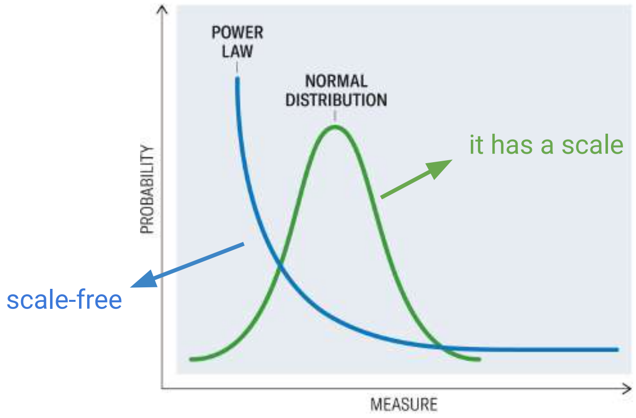
Travel distance probability

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

What's the shape of this distribution?

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.

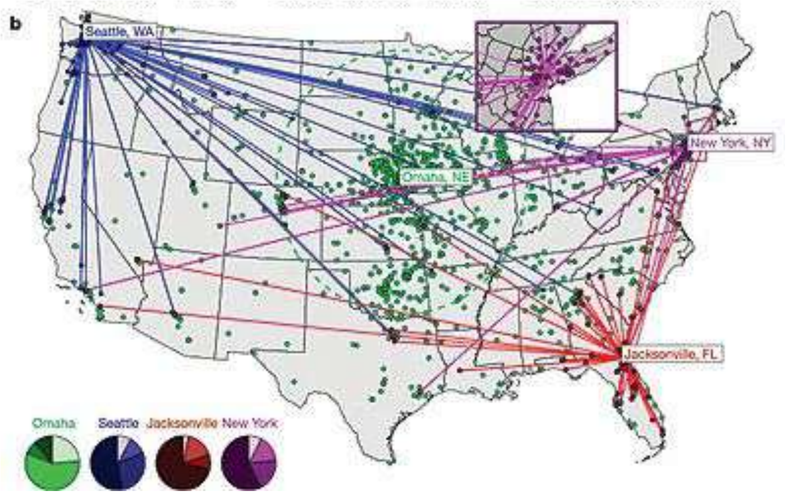


Tracking of dollar bills

Tracking of dollar bills

Brockmann et al., 2006:

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

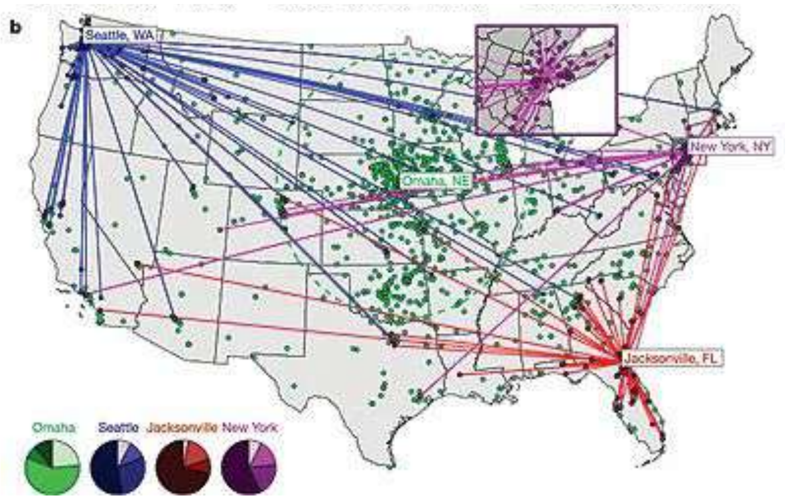


Trajectories of bank notes originating from four different places with travelling time $T < 14$ days.

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%



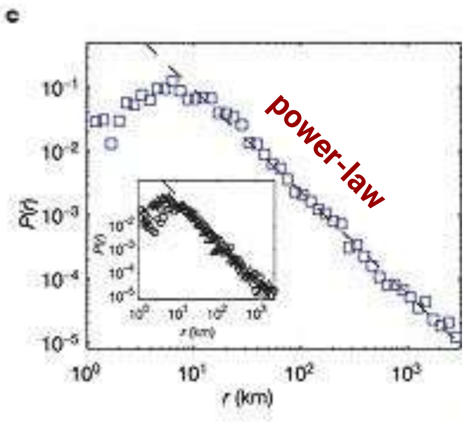
Trajectories of bank notes originating from four different places with travelling time $T < 14$ days.

Tracking of dollar bills

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

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

$$\beta = 0.59 \pm 0.02$$



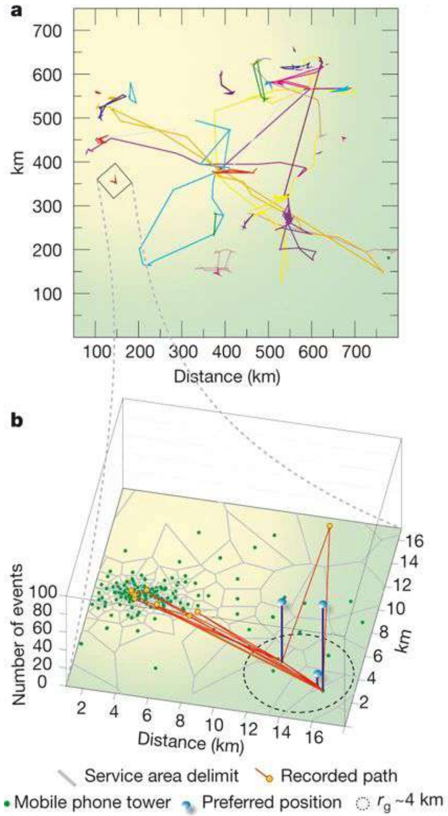
Measured P(r) of traversing a distance in less than T = 4 days. The inset shows P(r) for metropolitan areas, cities of intermediate size, small towns.

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

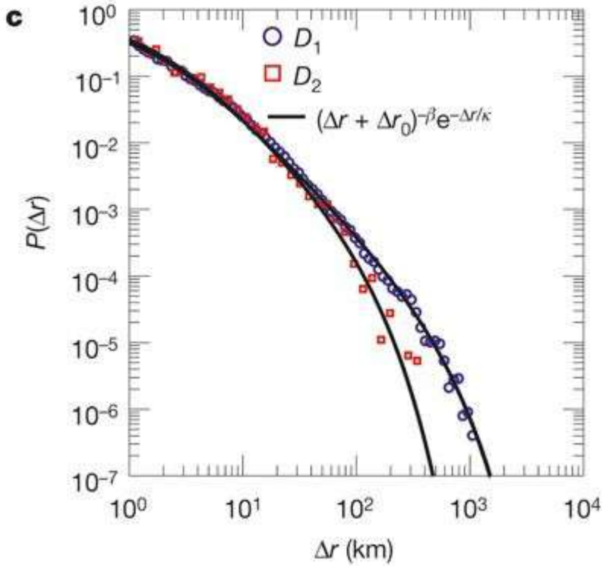


Week-long trajectory of 40 mobile phone users

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



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

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

$$\beta = 1.75 \pm 0.15$$
$$r_0 = 1.5km$$
$$\kappa_{D_1} = 400km$$
$$\kappa_{D_2} = 80km$$

Radius of gyration

Characteristic distance of an individual

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

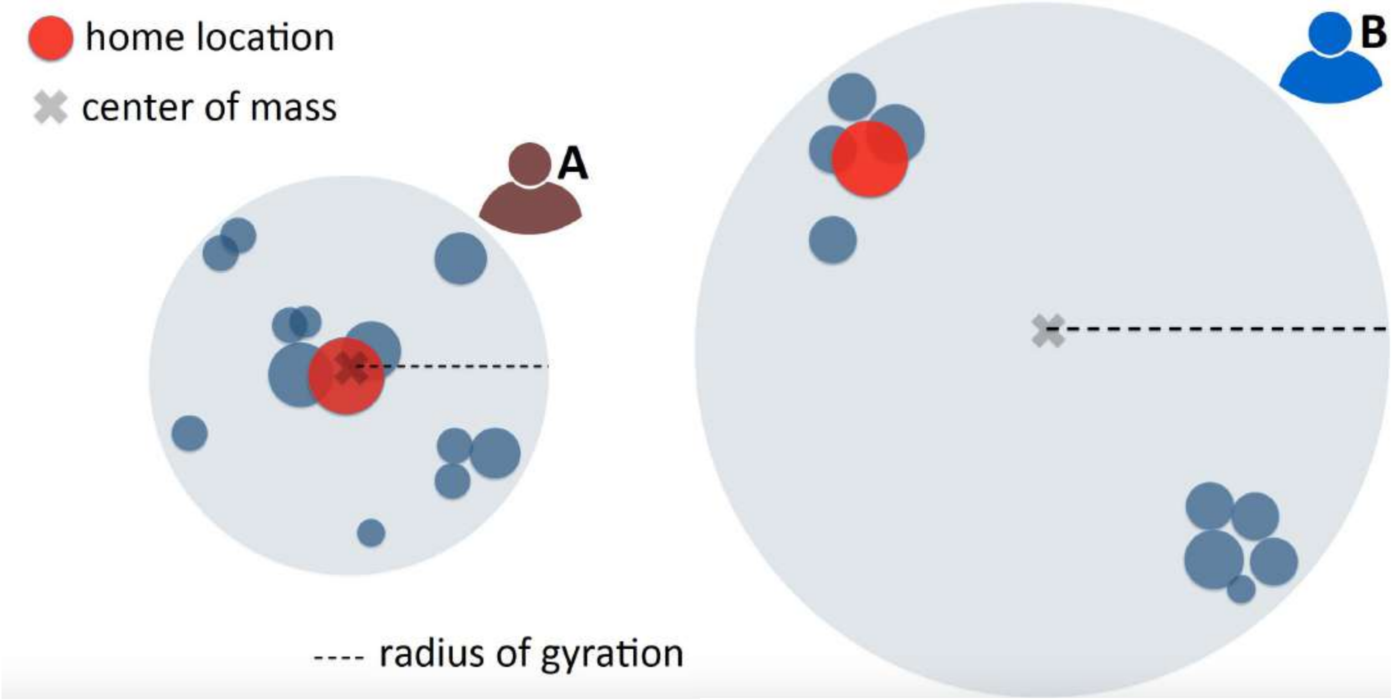
Center of mass

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

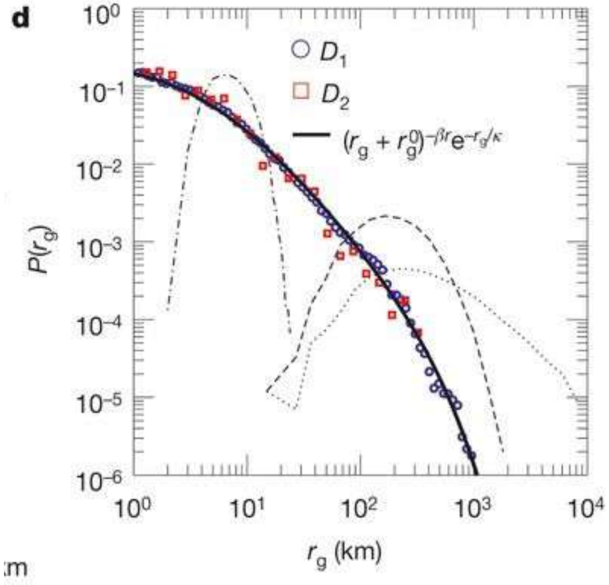
n_u number of records

$\mathbf{r}_i$ position

Radius of gyration



Radius of gyration



Measured $P(r_g)$ on datasets D1 and D2. The dotted, dashed and dot-dashed curves show P(rg) obtained from null models

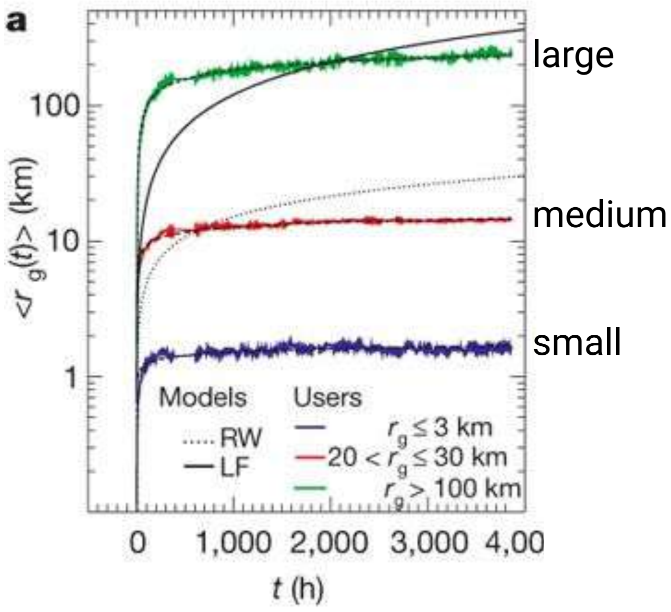
$$P(r_g) = (r_g + r_g^0)^{-\beta_r} \exp(-r_g/\kappa)$$

$$r_g^0 = 5.8km$$

$$\beta_r = 1.65 \pm 0.15$$

$$\kappa = 350km$$

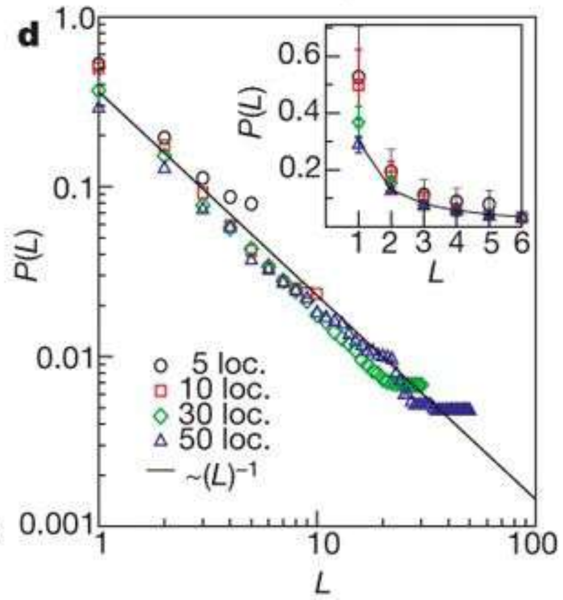
Radius of gyration - time evolution



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.

- Radius increases logarithmically with time
    - Indicating a saturation process

Location frequency



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

- 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

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

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

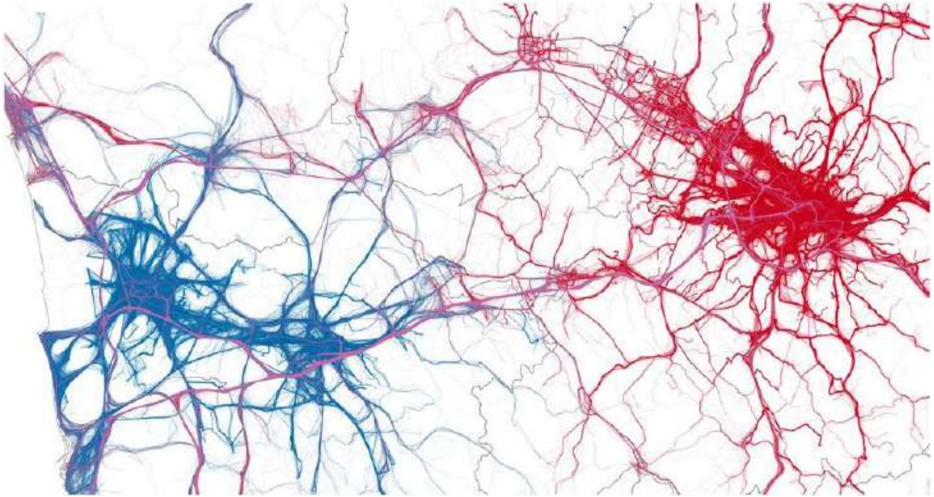
$$\mathbf{r}_{cm}^{(k)} = \frac{1}{N_k} \sum_{i=1}^{k} w_i \mathbf{r}_i$$

N_k number of records in location *k*

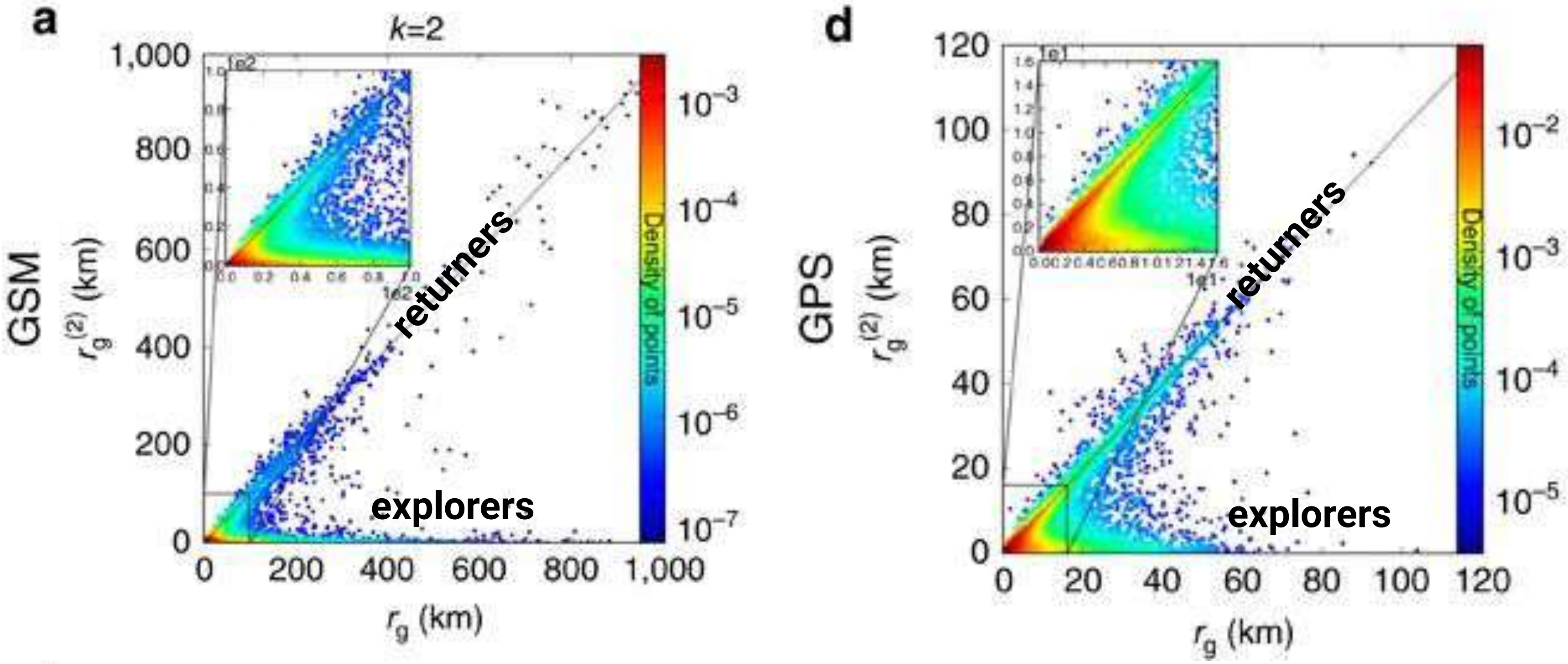
Mobile Phone Records

Pappalardo et al., 2015:

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

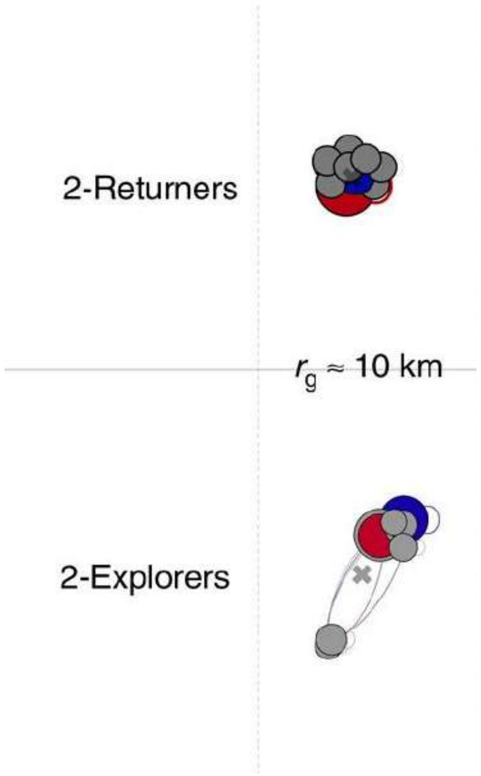


Returners and Explorers

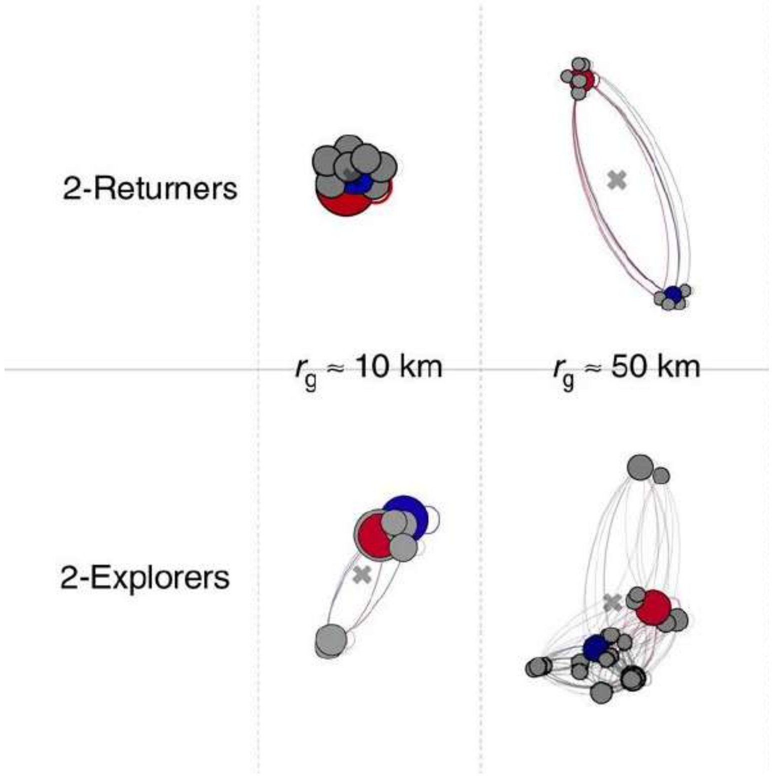


Correlation between total rg and rg$^{(k)}$ for k=2, 4, 8 for CDRs and GPS traces. Each point is coloured from blue to red, indicating the density of points in the corresponding region.

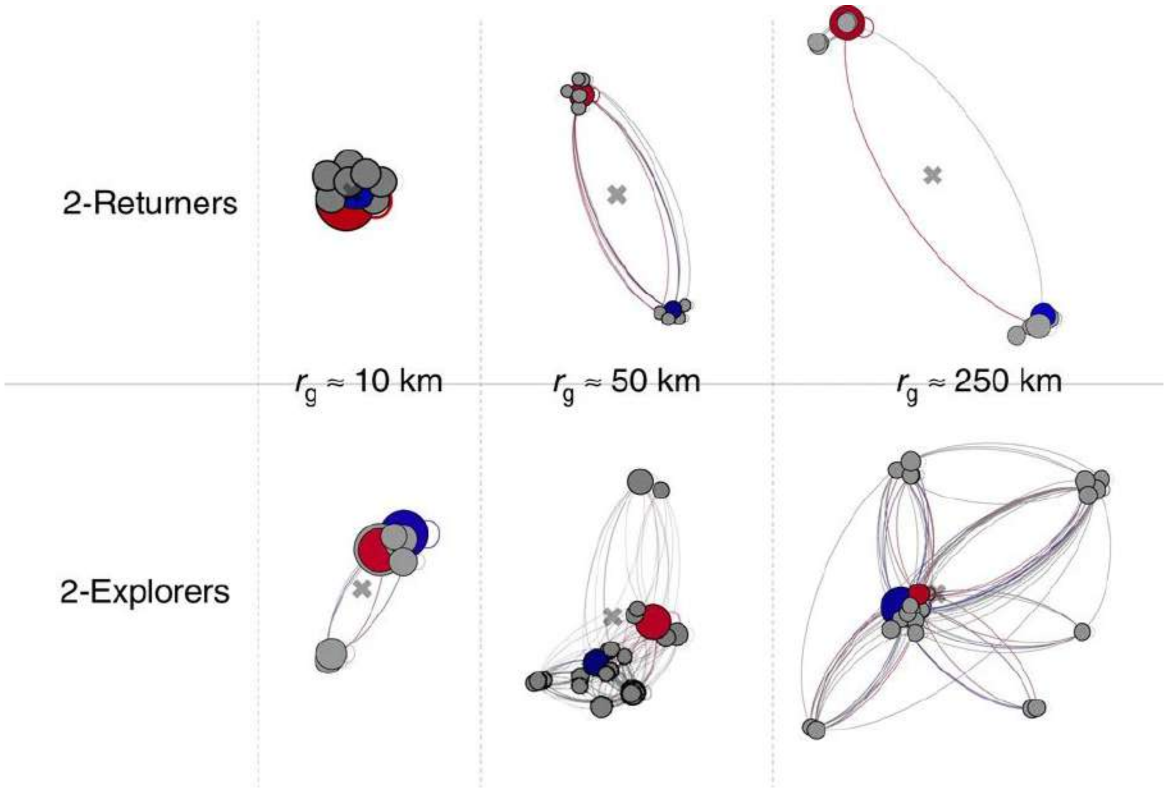
Returners and Explorers



Returners and Explorers



Returners 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.

Vilfredo Pareto and the 80/20 rule

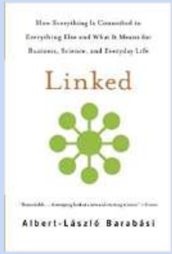
The 80/20 rule emerges in many areas:

- 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

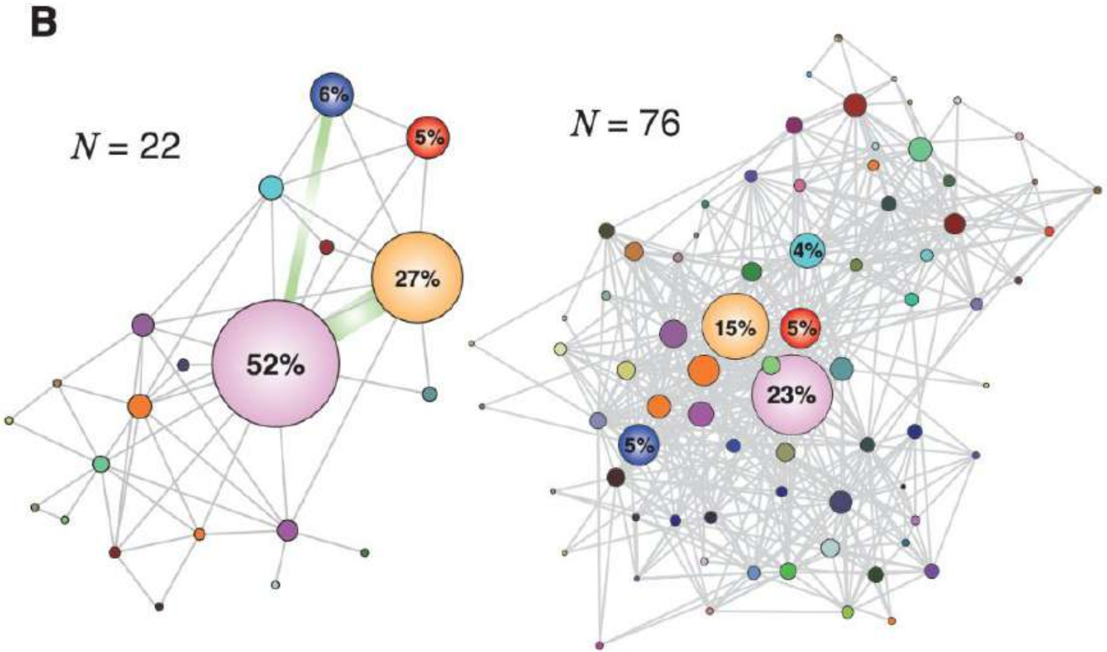


- [book] Chi troppo chi niente, E. Ferragina

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

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

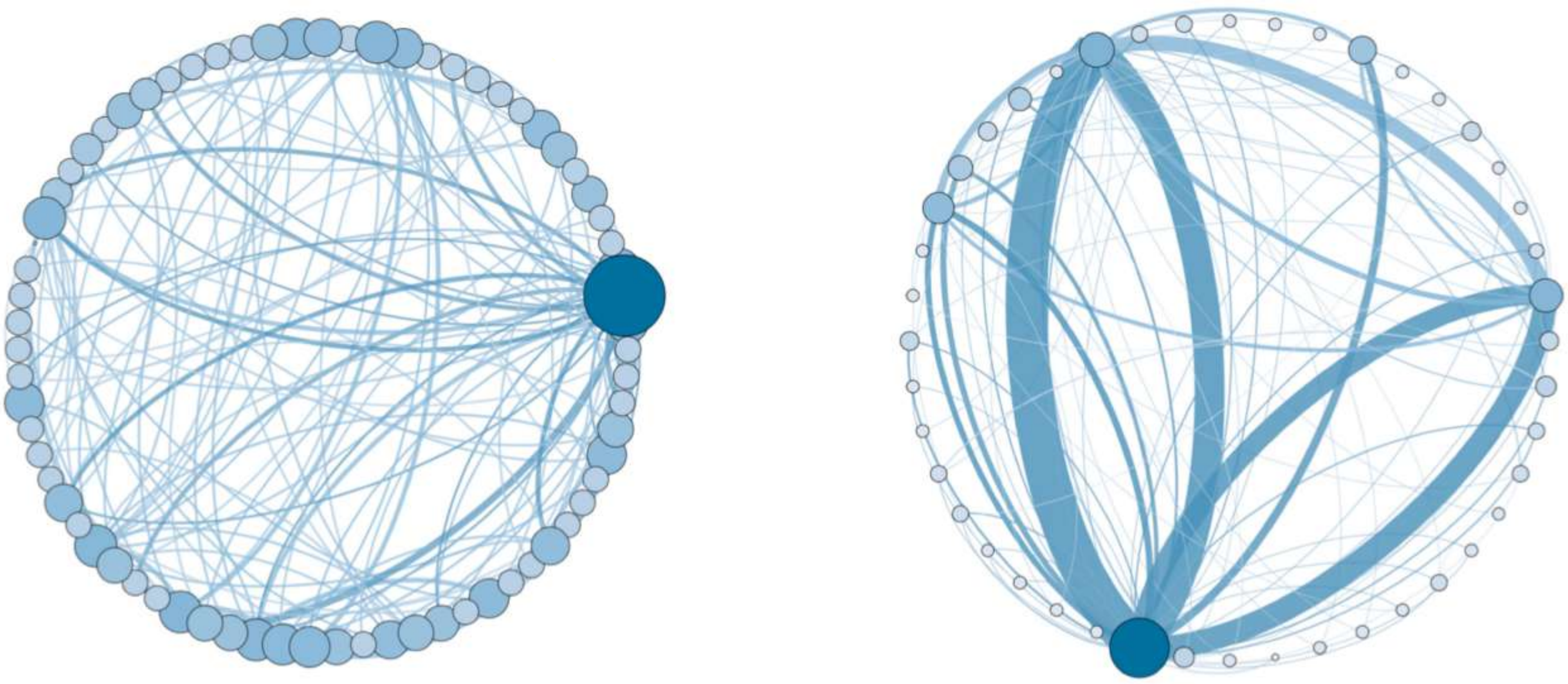
Uncorrelated entropy

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

Real entropy

$$S = -\sum_{T_i' \subset T_i} p_{T_i'} \log_2 p_{T_i'}$$

Who's the most predictable?

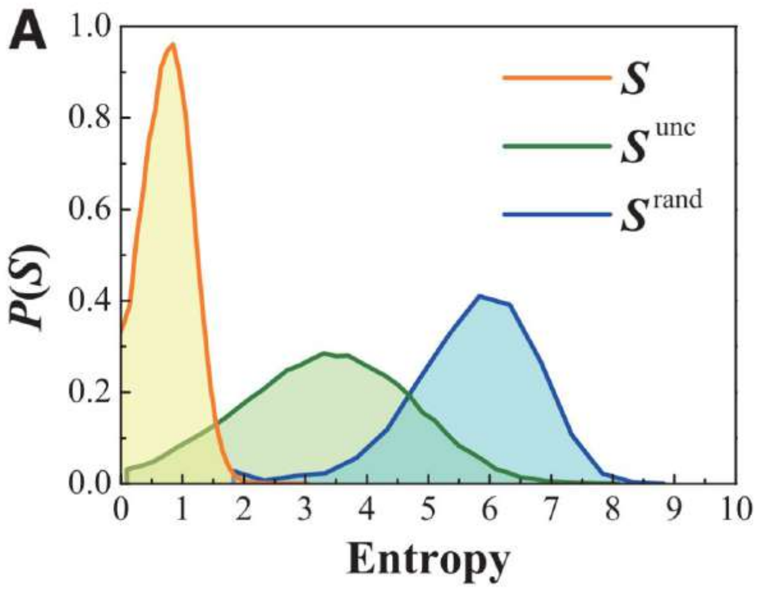


Entropy

Song et al., 2010:

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

$$2^{0.8} = 1.74$$

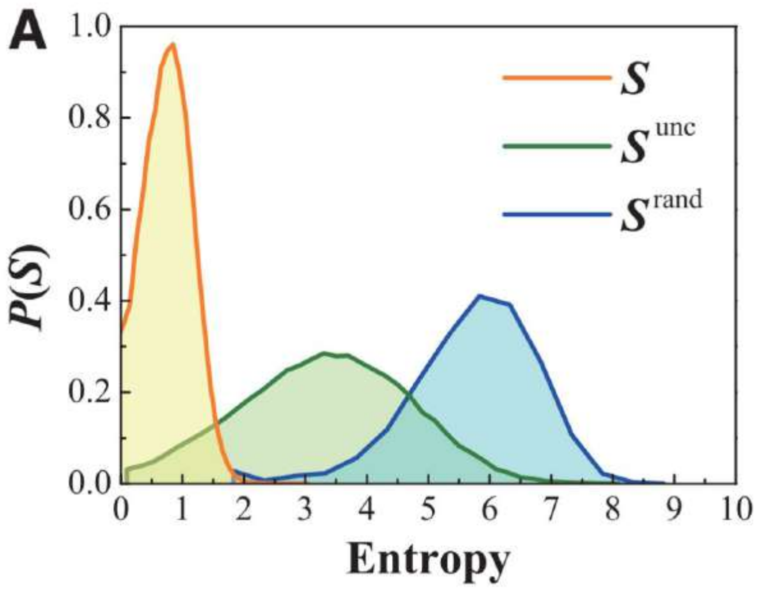


Entropy

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

Homework

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

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 is 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

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

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