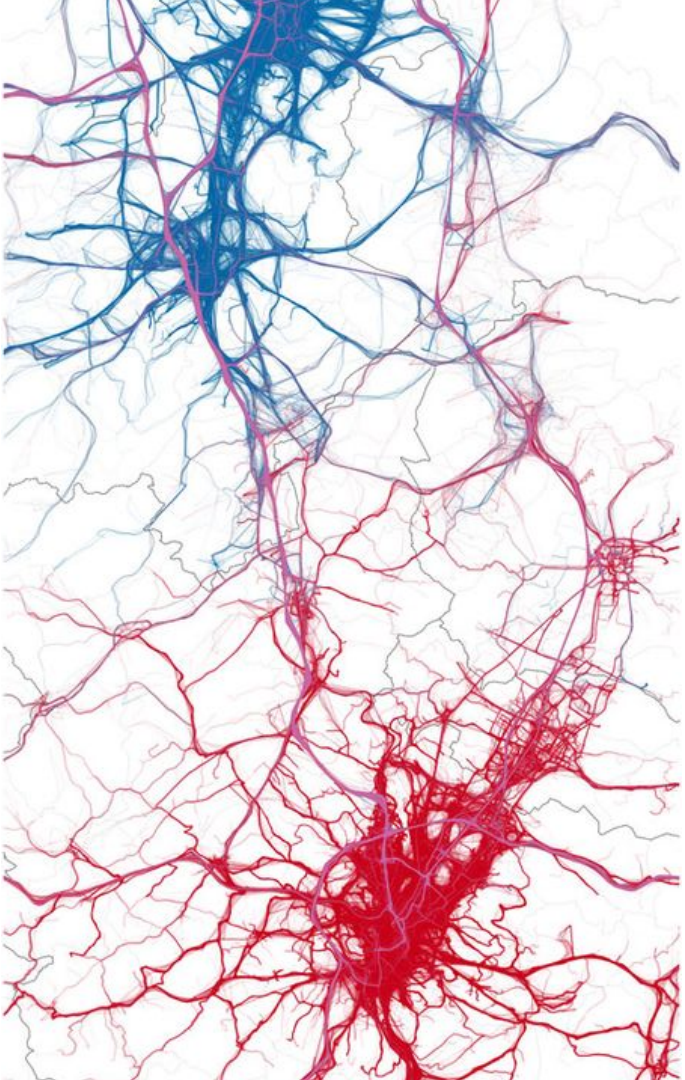


IN SUPREMÆ DIGNITATIS 1343

Consiglio Nazionale delle Ricerche

SCUOLA NORMALE SUPERIORE

Individual mobility laws

How far do individuals move between successive locations?

Do human displacements exhibit characteristic spatial scales?

Are mobility patterns dominated by recurrence or exploration?

How are visits distributed across locations?

How does visitation frequency depend on distance from home?

How many regularly visited places do individuals maintain?

How predictable are future movements from past trajectories?

Displacement lengths

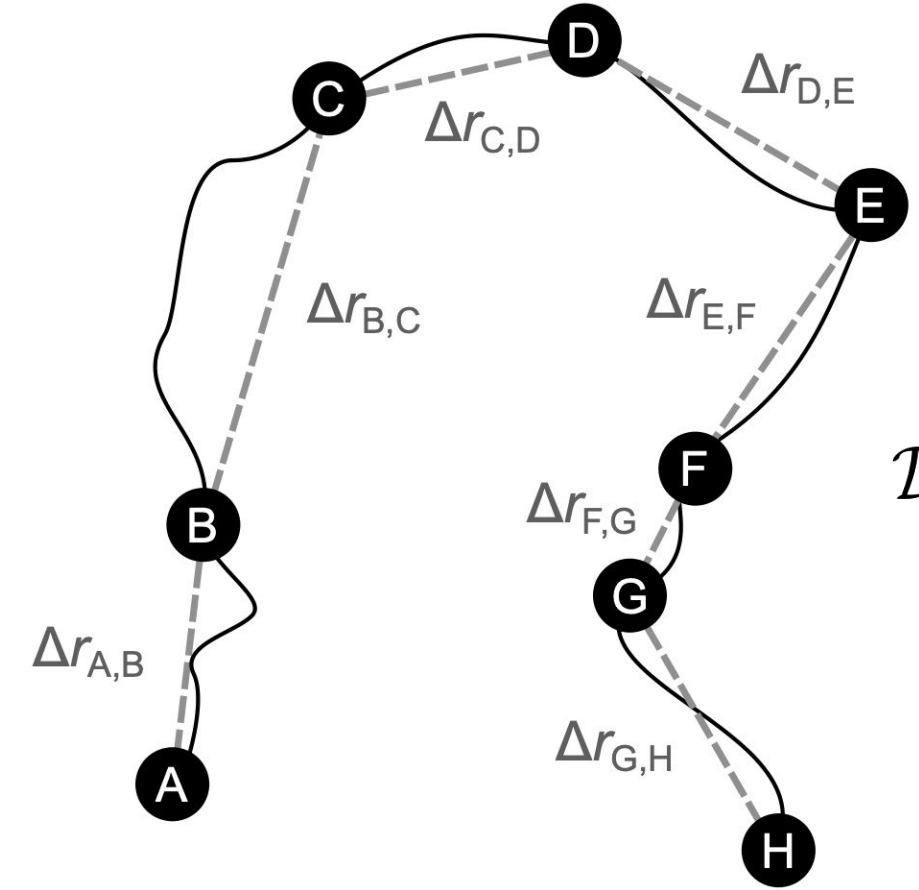
DEFINITION 6.1

Displacement lengths

Let $L_u = (\ell_1, \ldots, \ell_m)$ be the location sequence of an individual u. Let $\mathbf{r}$ be a function associating each location ℓ_i with a spatial position $\mathbf{r}(\ell_i) \in \mathbb{R}^2$.
The multiset of *displacement lengths* of u is defined as

$$\mathcal{D}_u = \left\{ \Delta r_{i,i+1} : i = 1, \ldots, m-1 \right\}. \tag{6.1}$$

where $\Delta r_{i,i+1} = d\big(\mathbf{r}(\ell_i), \mathbf{r}(\ell_{i+1})\big)$ and $d(\cdot, \cdot)$ is a distance function between two spatial positions.



$$\mathcal{D}_u = \{\Delta r_{A,B}, \ldots, \Delta r_{G,H}\}$$

What quantities can be derived from the multiset?

$$d_{\max}(u) = \max_{\Delta r \in \mathcal{D}_u} \Delta r$$

Max displacement travelled by an individual

$$d_{\text{tot}}(u) = \sum_{\Delta r \in \mathcal{D}_u} \Delta r$$

Total distance travelled by an individual

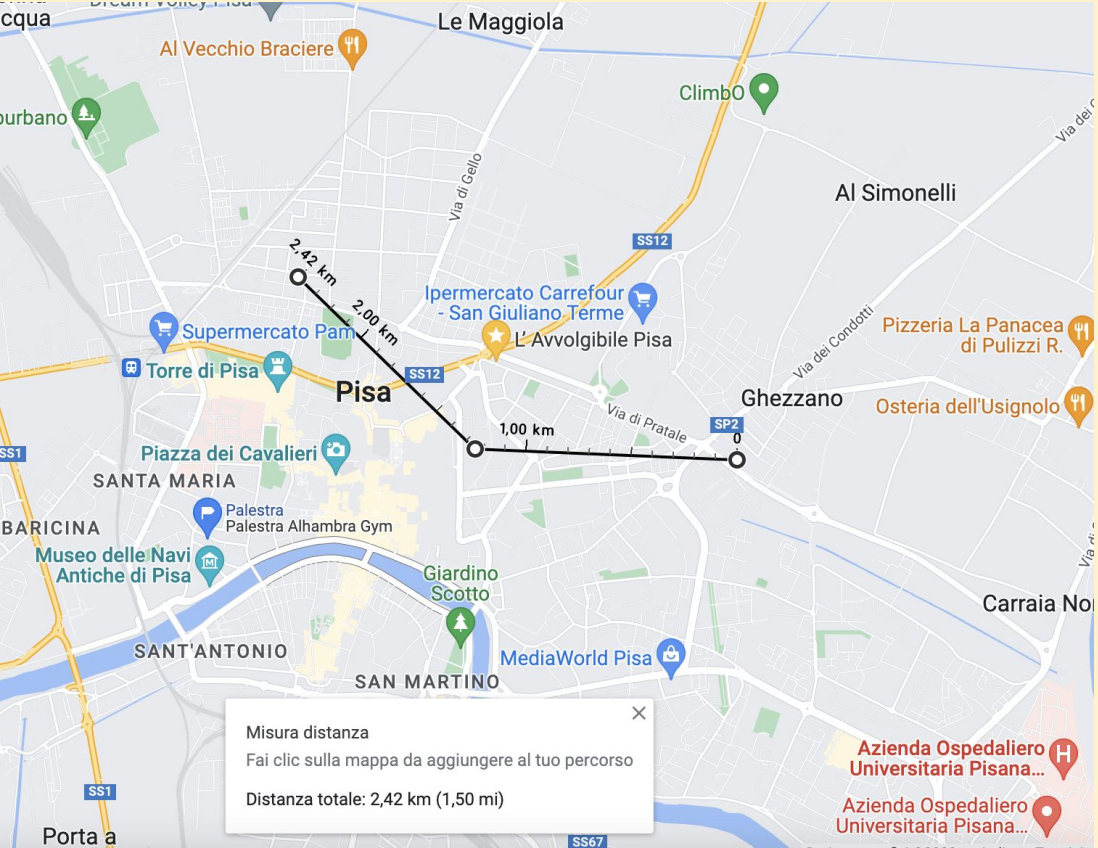
Pooled multiset of displacement lengths

Displacement lengths are commonly studied at the population level

$$\mathcal{D} = \biguplus_{u \in \mathcal{U}} \mathcal{D}_u = \{\Delta r_j^{(u)} : u \in \mathcal{U},\ j = 1, \ldots, m_u - 1\}$$

j-th displacement length
of individual u

How far do individuals move between successive locations?
Do human displacements exhibit characteristic spatial scales?



Collect the distances from your homes to Largo Bruno Pontecorvo 3, Pisa

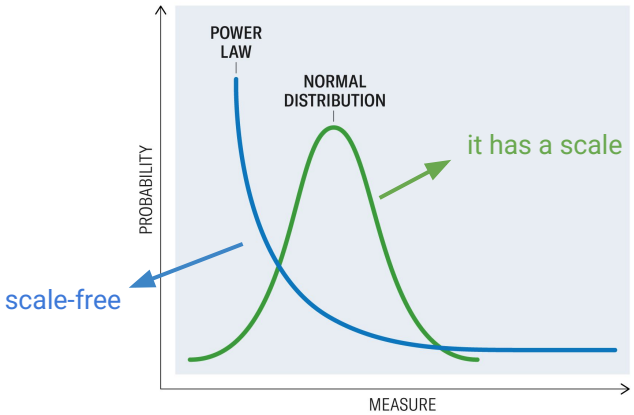
Displacement length probability

$$P(r) = \text{probability of finding a trip of length } r$$

What's the shape of the 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

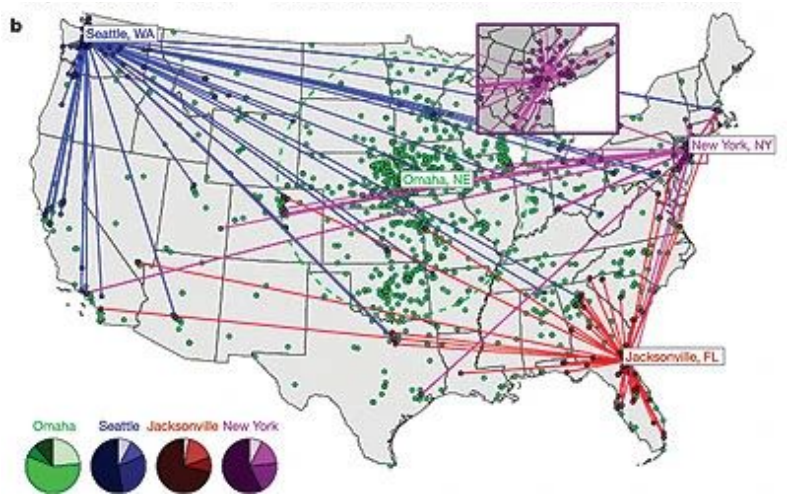




Tracking of dollar bills

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.

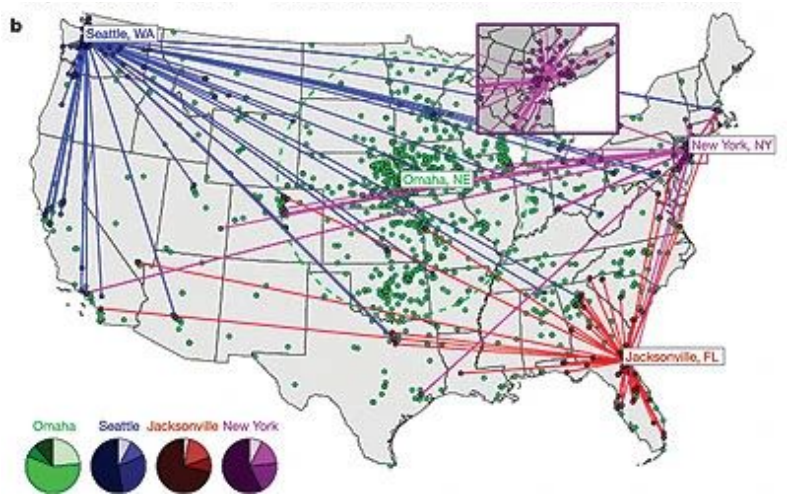
Tracking of dollar bills

Most banknotes within 10 km

- Seattle 53%
  NYC 58%
  Jacksonville 71%

Small but considerable fraction farer than 800 km

- Seattle 8%
  NYC 7%
  Jacksonville 3%



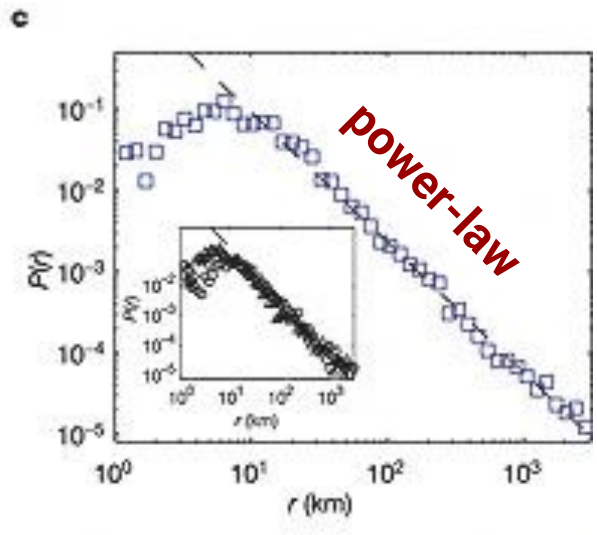
Trajectories of bank notes originating from four different places with travelling time $T < 14$ days.

Tracking of dollar bills

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

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

$$\beta = 1.59 \pm 0.02$$



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.

What limitations do dollar bill records have?

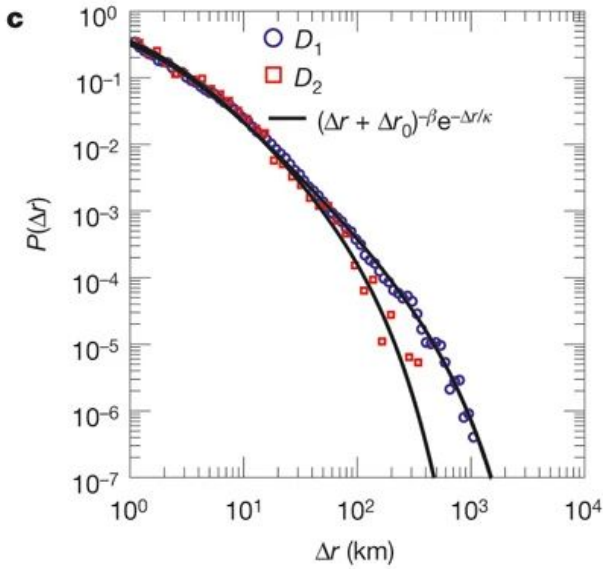
A banknote do not correspond to a single individual: it is exchanged among multiple people

Mobile Phone Records

González et al., 2008:

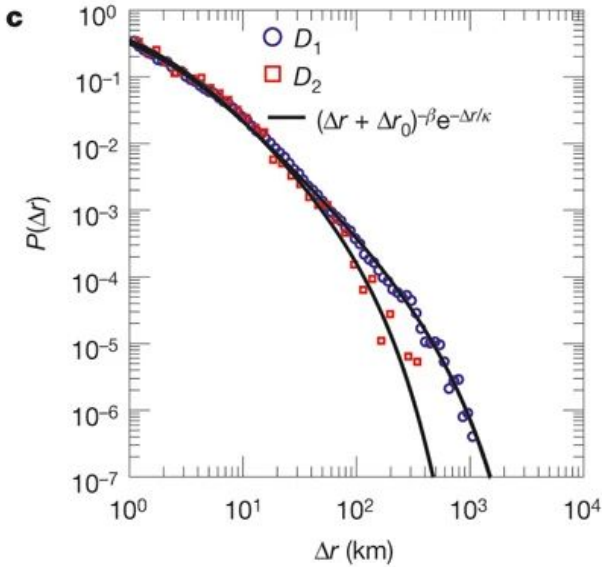
- Dataset D1 (CDRs):
  - Users: 100,000
  - Records: 16,264,308

- Dataset D2 (CPRs):
  - Users: 206
  - Records: 10,407



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

Mobile Phone Records



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

Truncated power-law

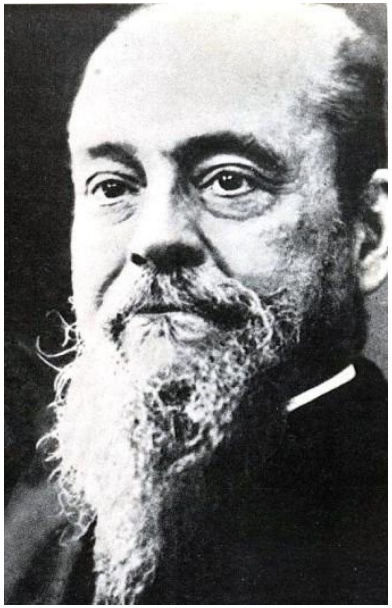
$$P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} \exp(-\Delta r/\kappa)$$

$$\beta = 1.75 \pm 0.15$$

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

$$\kappa_{\text{CDR}} = 400\text{km} \quad \kappa_{\text{CPR}} = 80\text{km}$$

Vilfredo Pareto's 80/20 rule



A few wealthy individuals earned most of the money; the majority of the population earned small amounts.

Incomes follow a power law.

80/20 rule: roughly 80 percent of money is earned by only 20 percent of the population.

Location importance

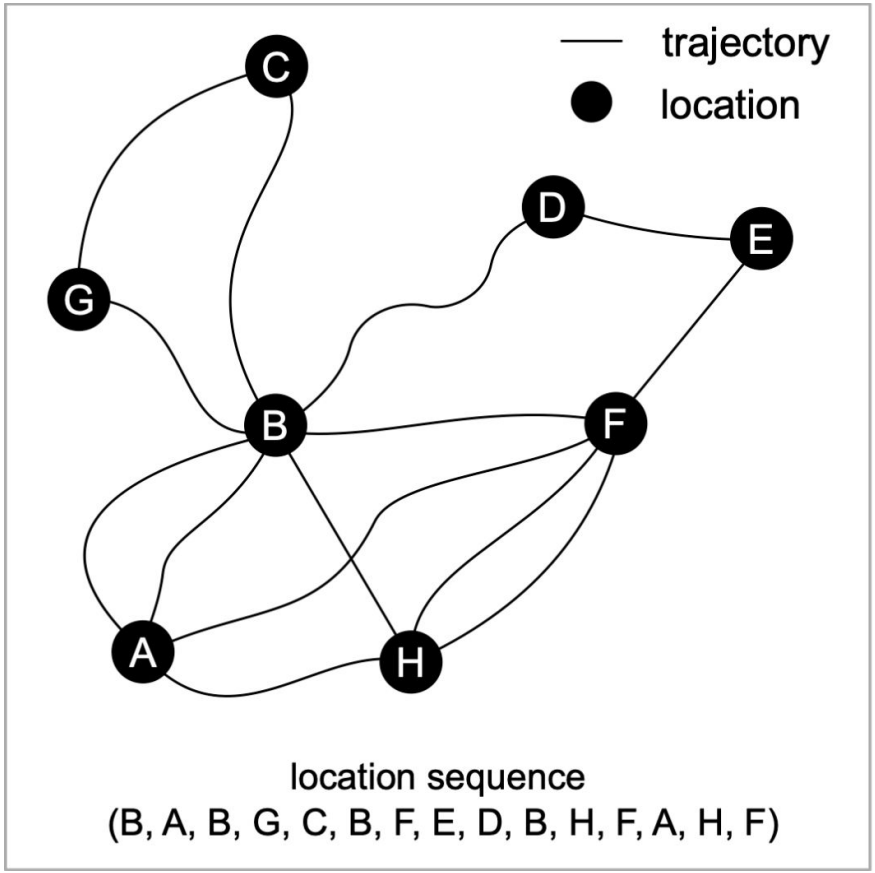
DEFINITION 6.2

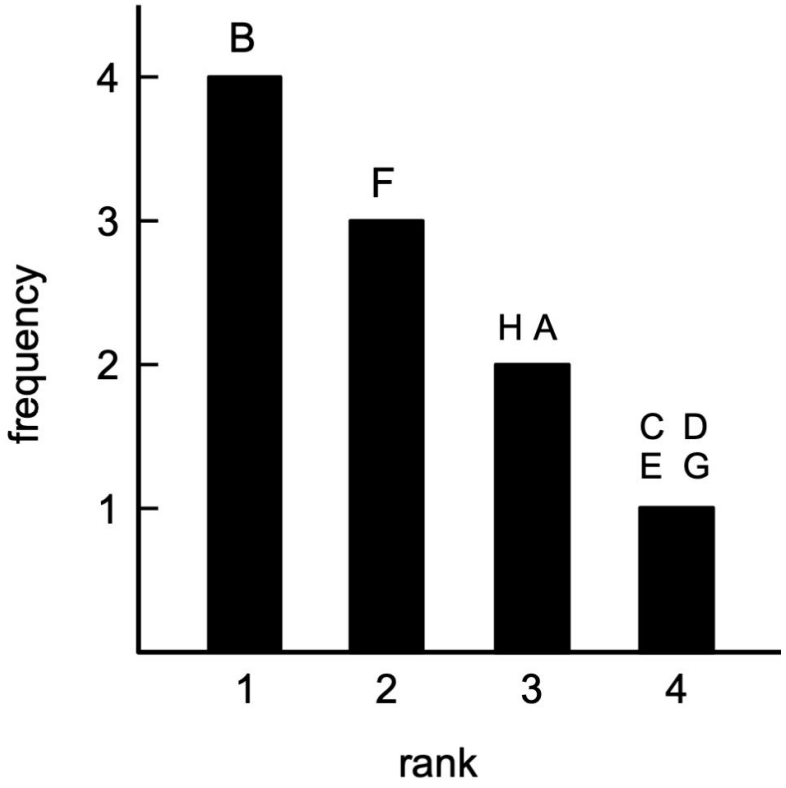
Location frequency

Let $L_u = (\ell_1, \ldots, \ell_m)$ be the location sequence of an individual u, and let $\mathcal{L}_u$ be the set of distinct locations in L_u.
The *location frequency* of $\ell \in \mathcal{L}_u$ is defined as

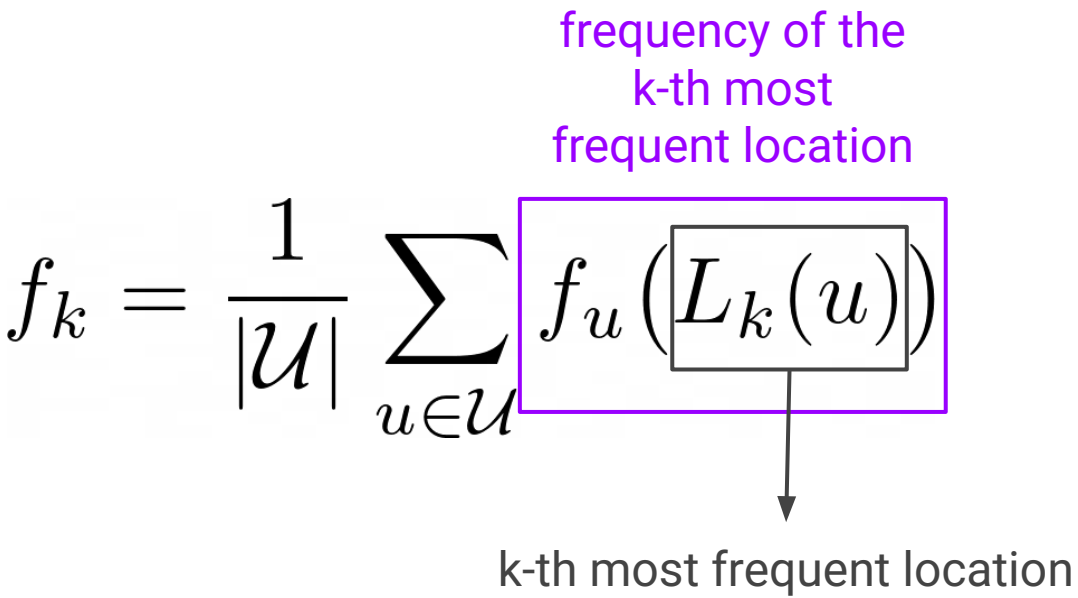
$$f_u(\ell) = \frac{n_\ell}{m} \tag{6.6}$$

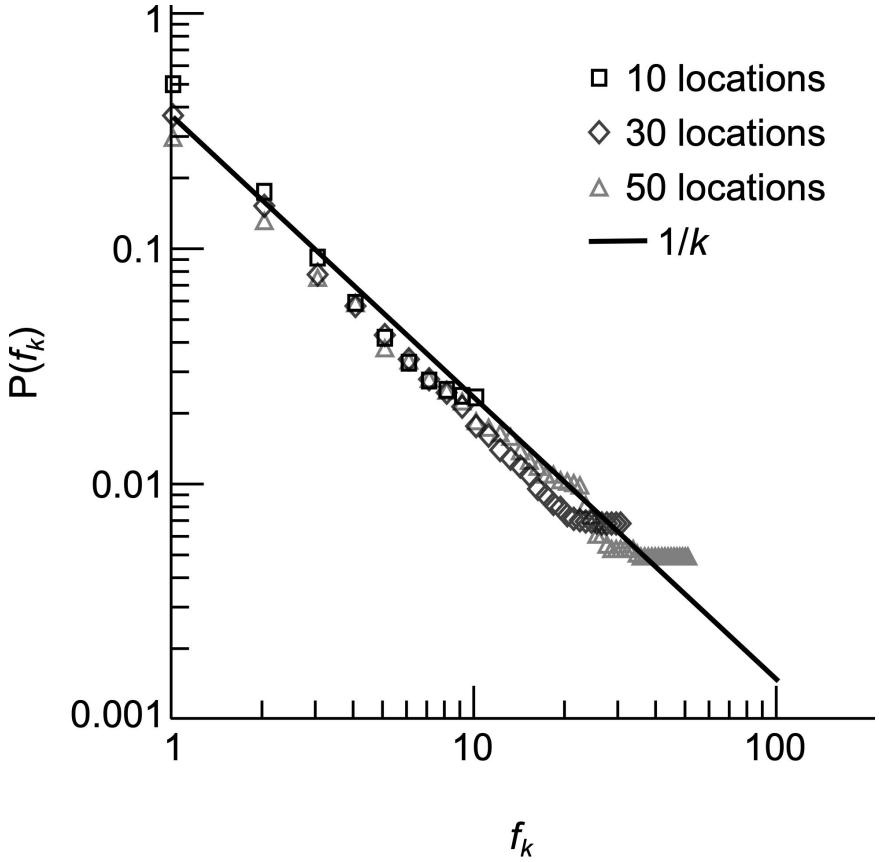
where n_ℓ denotes the number of visits u made to ℓ, i.e., how many times ℓ appears in $\mathcal{L}_u$, and m is the length of the location sequence.
Let $L_1(u), L_2(u), \ldots, L_{|\mathcal{L}_u|}(u)$ be an ordering of the locations in $\mathcal{L}_u$ such that $f_u\big(L_1(u)\big) \geq f_u\big(L_2(u)\big) \geq \cdots \geq f_u\big(L_{|\mathcal{L}_u|}(u)\big)$. $L_k(u)$ is called the k-th *most frequent location* of u.


trajectory
location
C
D
E
G
B
F
A
H
location sequence
(B, A, B, G, C, B, F, E, D, B, H, F, A, H, F)


B
F
H A
C D
E G
4
3
2
1
frequency
1
2
3
4
rank

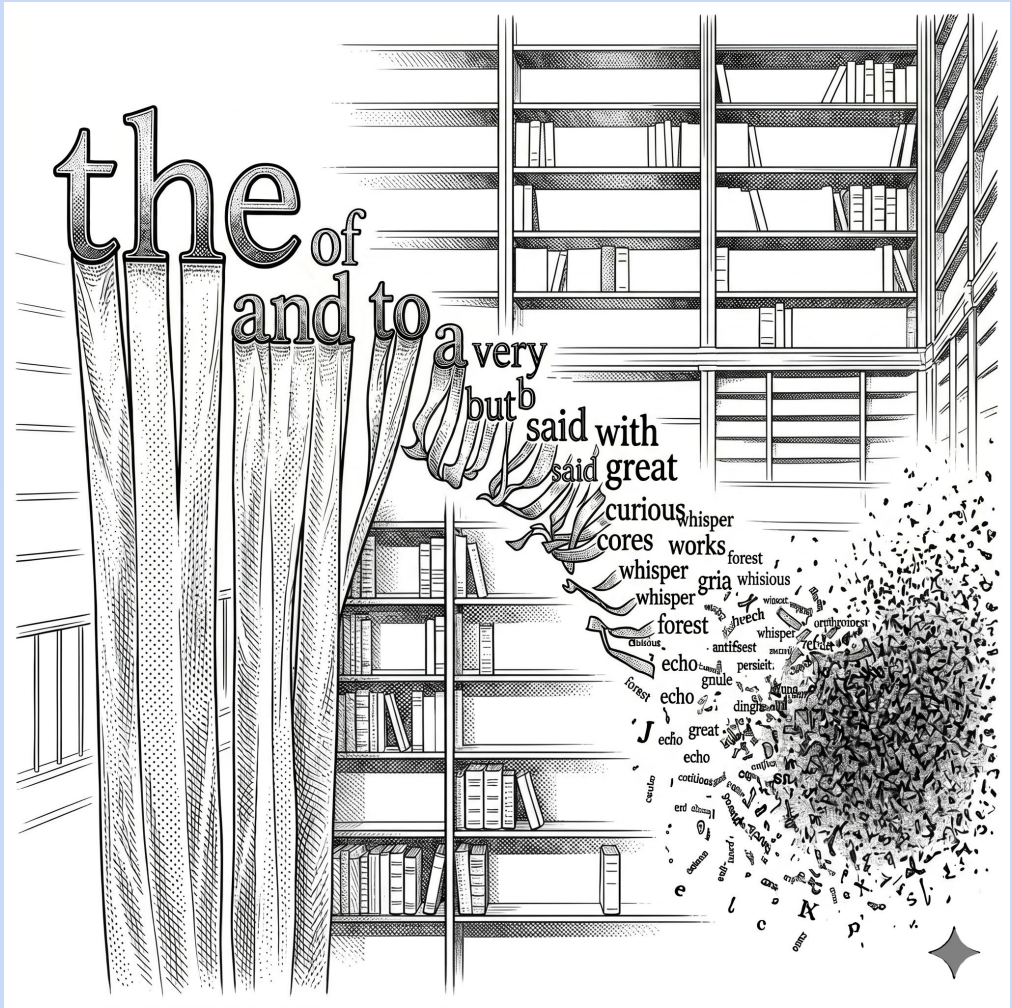
Rank-frequency distribution





González et al., 2008

$$f_k \sim \frac{1}{k}$$



DID YOU KNOW THAT?

Words in your language follow **Zipf's law**

Zipf's law

If you rank words in a language by how often they appear:

- the second most common word appears half as often as the first
- the third one-third
- the fourth one-fourth
- etc.

Example: In English, the word "the" accounts for 7% of all words. The next most frequent word, "of", accounts for 3.5%, following by "and", and so on.

Why?

Principle of least effort:
speakers and listeners balance efficiency and clarity

Zipf's law also appear in city sizes, webpage popularity, and more.

Compute the distribution of f_k for this text:

https://www.gutenberg.org/cache/epub/76871/pg76871-images.html

Radius of gyration

DEFINITION 6.5

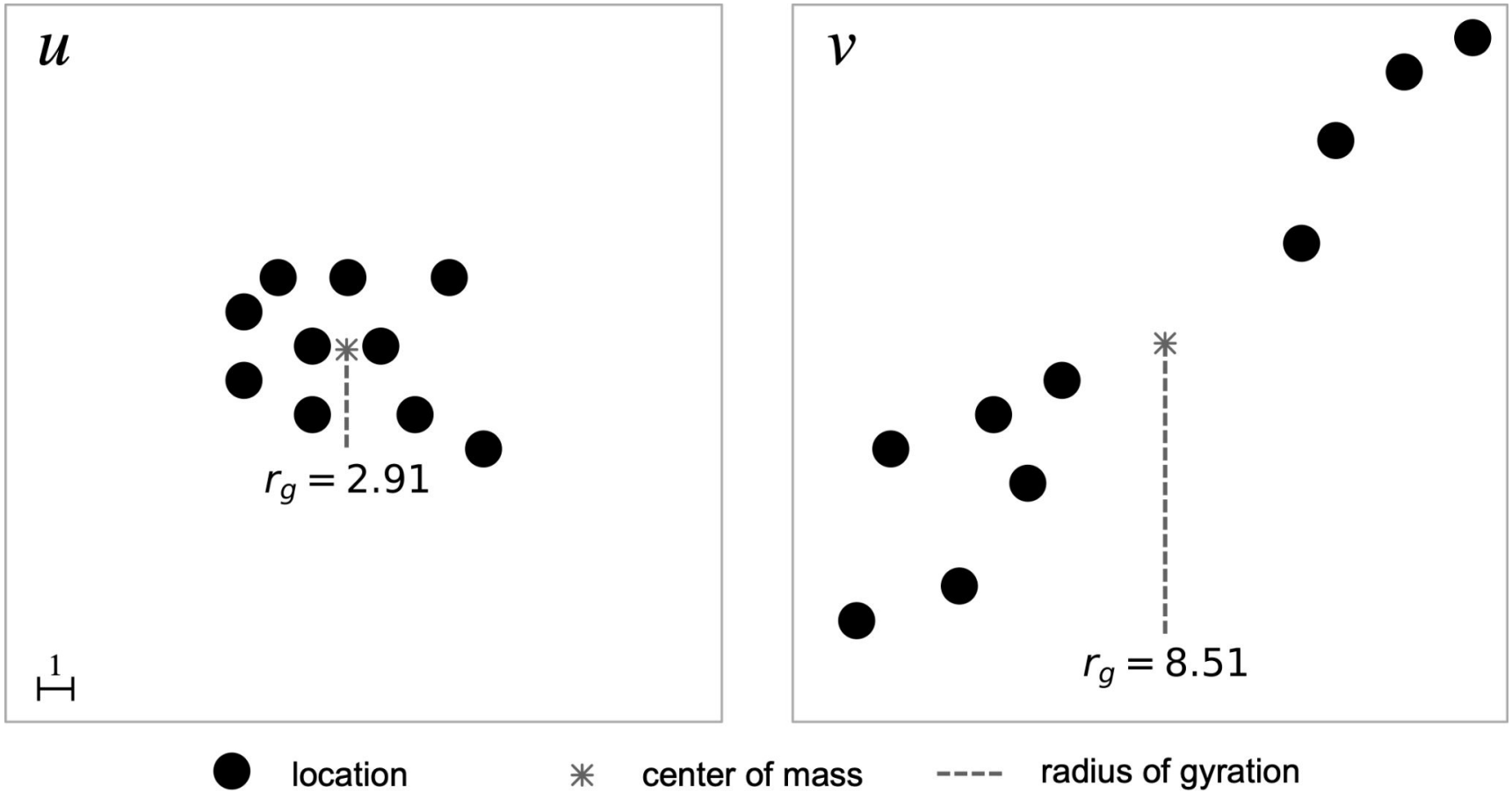
Radius of gyration

Let $T_u = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$ be the mobility trajectory of an individual u, where each $\mathbf{p}_i \in \mathbb{R}^2$ represents the spatial position (e.g., latitude and longitude) of u at time i.
The *radius of gyration* of u is defined as:

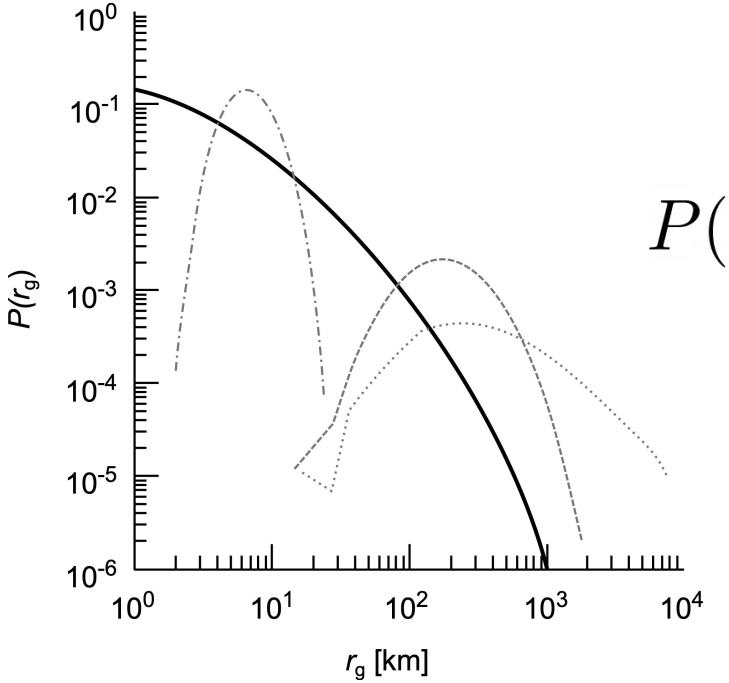
$$r_g(u) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d(\mathbf{p}_i, \mathbf{r}_{\text{cm}})^2}, \tag{6.12}$$

where $d(\cdot, \cdot)$ denotes a distance function between two spatial positions, and $\mathbf{r}_{\text{cm}} \in \mathbb{R}^2$ is the centre of mass of u's mobility trajectory, defined as:

$$\mathbf{r}_{\text{cm}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{p}_i\,. \tag{6.13}$$


u
$r_g = 2.91$
1
v
$r_g = 8.51$
location
center of mass
radius of gyration

Mobile Phone Records



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

Truncated power-law

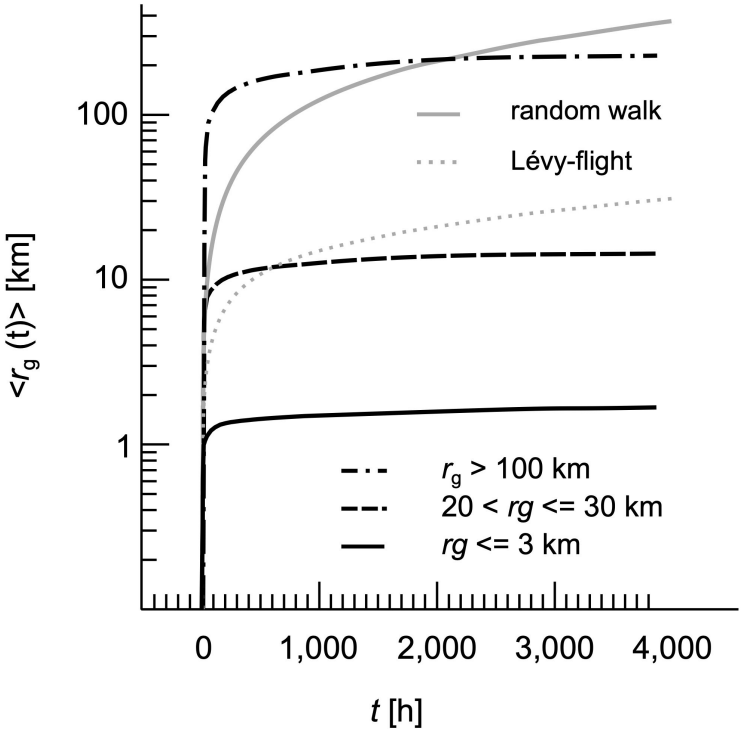
$$P(r_g) = (r_g - r_g^0)^{-\beta_{r_g}} \exp(-r_g/\kappa_{r_g})$$

$$\beta_{r_g} = 1.65 \pm 0.15$$

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

$$\kappa_{r_g} = 350\text{km}$$

Mobile Phone Records



Radius increases logarithmically with time, indicating a saturation process

DEFINITION 6.6

k-radius of gyration

Let $L_u = (\ell_1, \ldots, \ell_m)$ be the location sequence of an individual u, and let $\mathcal{L}_u$ be the set of distinct locations in L_u. Let $\mathbf{r} : \mathcal{L}_u \to \mathbb{R}^2$ be a function associating each location with a spatial position. Let $L_1(u), L_2(u), \ldots, L_{|\mathcal{L}_u|}(u)$ be the locations ordered by decreasing frequency. For a given integer k, define the set of the k most frequent locations as

$$\mathcal{L}_u^{(k)} = \{L_1(u), \ldots, L_k(u)\}.$$

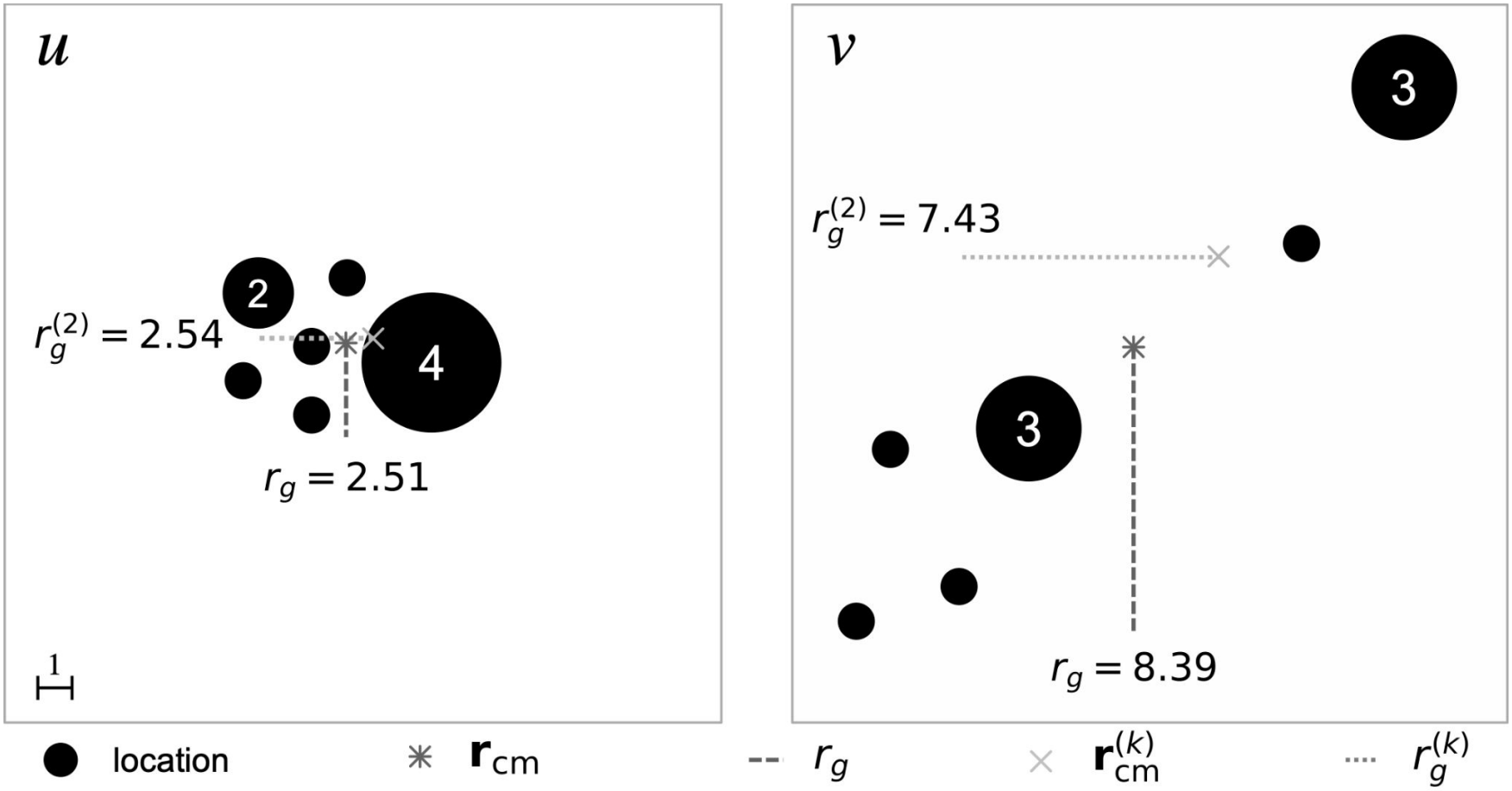
The *k-center of mass* of u is defined as

$$\mathbf{r}_{\mathrm{cm}}^{(k)}(u) = \frac{1}{\sum_{\ell \in \mathcal{L}_u^{(k)}} f_u(\ell)} \sum_{\ell \in \mathcal{L}_u^{(k)}} f_u(\ell)\, \mathbf{r}(\ell), \tag{6.14}$$

where $f_u(\ell)$ denotes the location frequency of ℓ and $\sum_{\ell \in \mathcal{L}_u^{(k)}} f_u(\ell)$ represents the fraction of visits occurring in the top-k locations.

The *k-radius of gyration* of u is defined as

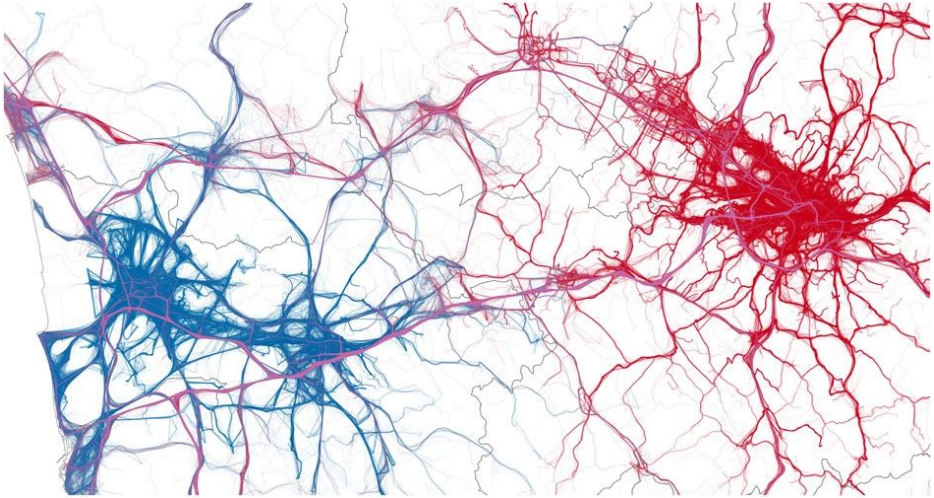
$$r_g^{(k)}(u) = \sqrt{\frac{1}{\sum_{\ell \in \mathcal{S}_u^{(k)}} f_u(\ell)} \sum_{\ell \in \mathcal{S}_u^{(k)}} f_u(\ell)\, d\left(\mathbf{r}(\ell), \mathbf{r}_{\mathrm{cm}}^{(k)}(u)\right)^2}. \tag{6.15}$$


u
2
4
$r_g^{(2)} = 2.54$
$r_g = 2.51$
1
v
3
3
$r_g^{(2)} = 7.43$
$r_g = 8.39$
location
$\mathbf{r}_{\text{cm}}$
r_g
$\mathbf{r}_{\text{cm}}^{(k)}$
$r_g^{(k)}$

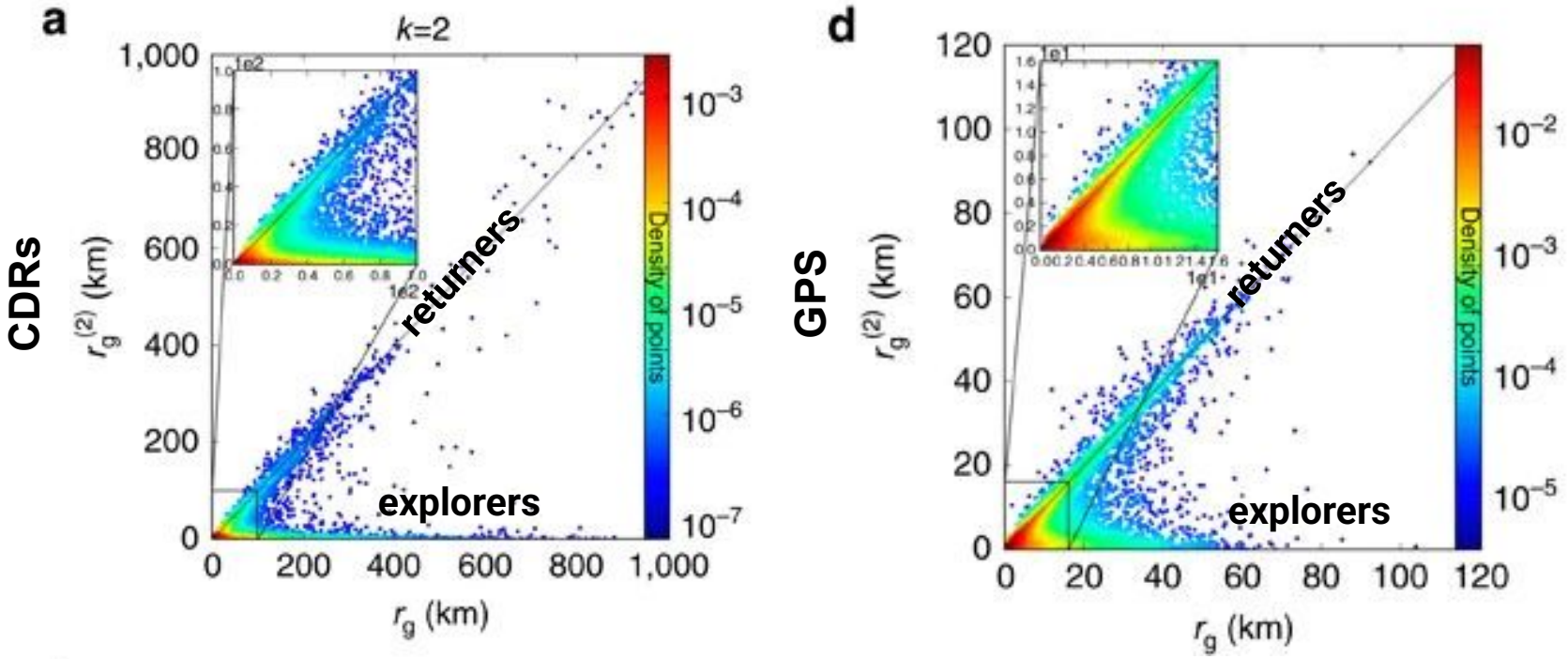
CDRs and GPS data

Pappalardo et al., 2015:

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

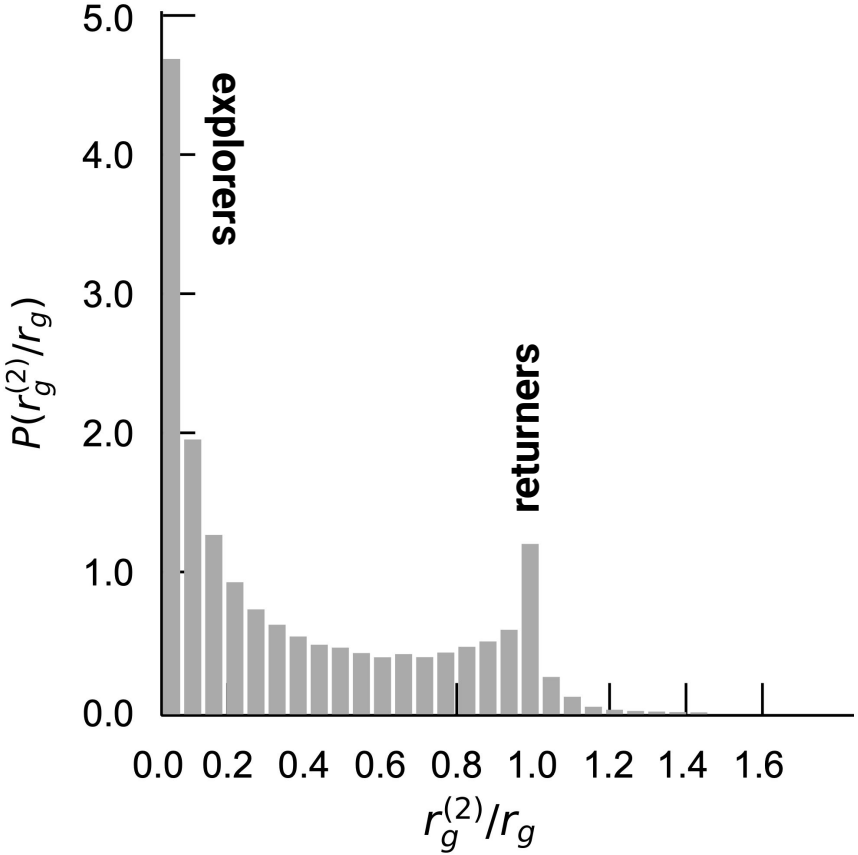


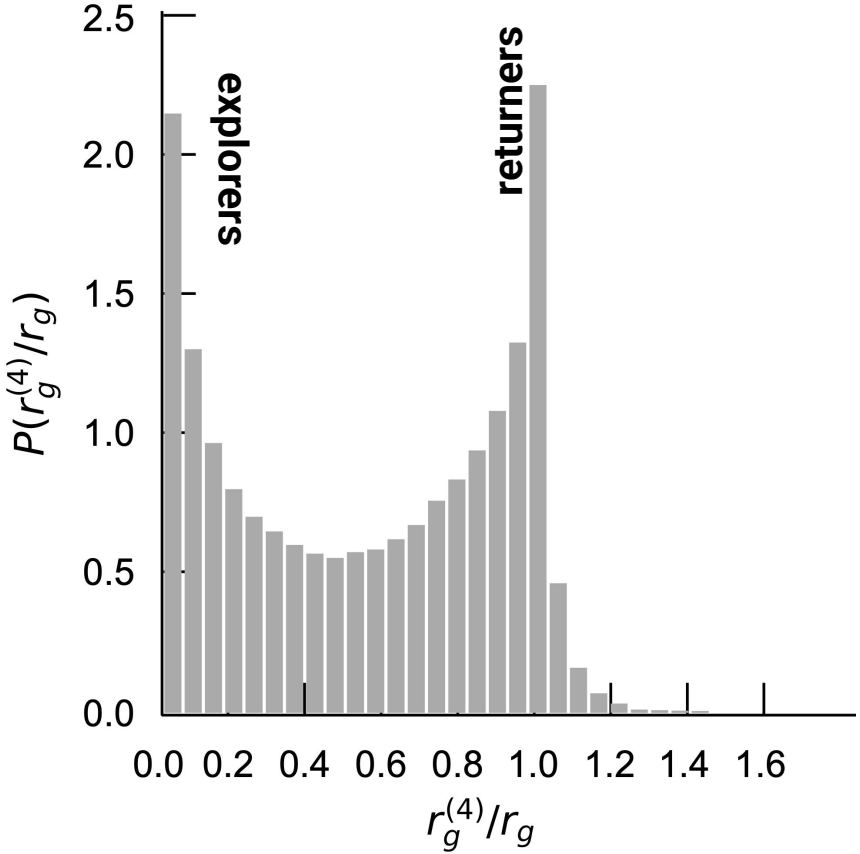
Returners and Explorers



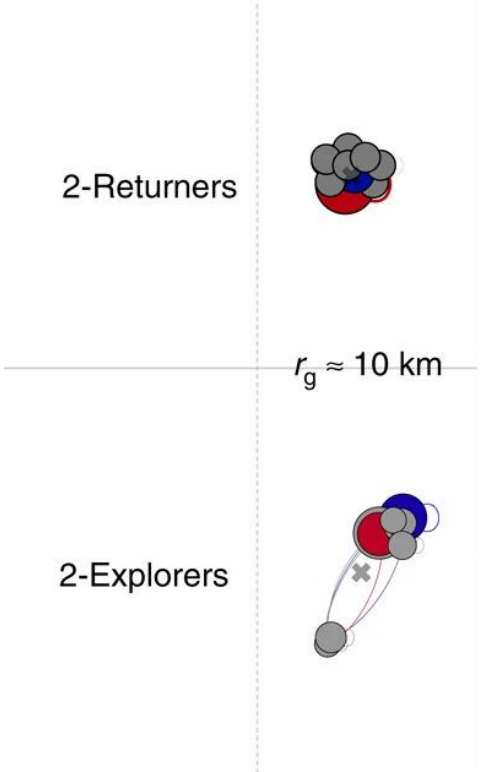
Correlation between total rg and $rg^{(k)}$ for k=2, 4, 8 for CDRs and GPS traces. Each point is coloured from blue to red, indicating the density of points in the corresponding region.

Returners and Explorers

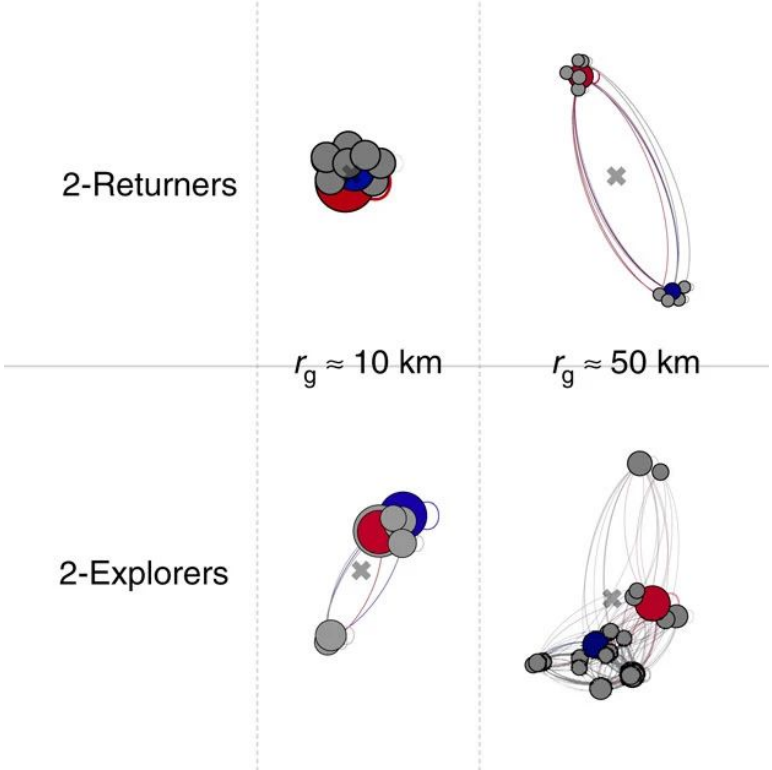




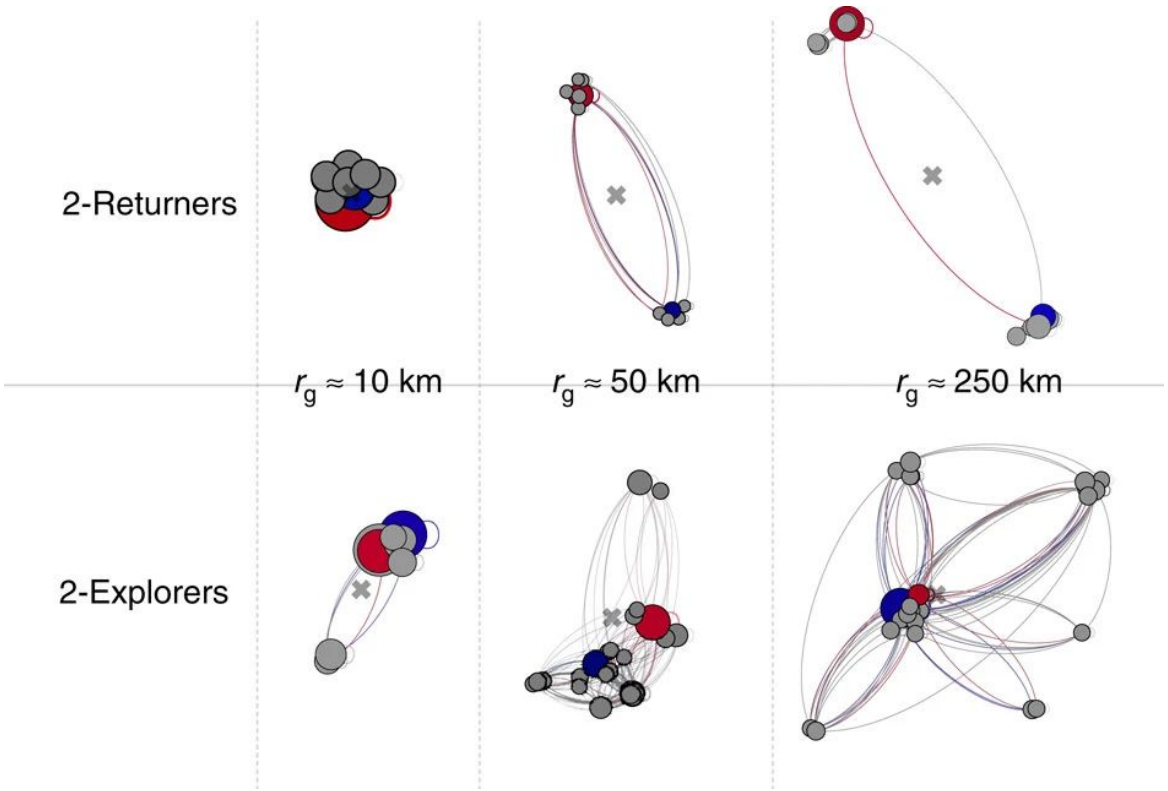
Returners and Explorers



Returners and Explorers



Returners and Explorers



Entropy and Predictability

DEFINITION 6.7

Temporal-uncorrelated entropy

Let $L_u = (\ell_1, \ldots, \ell_m)$ be the location sequence of an individual u, and let $\mathcal{L}_u$ denote the set of distinct locations appearing in L_u.
The *temporal-uncorrelated entropy* of u is given by

$$S_{\text{unc}}(u) = -\sum_{\ell \in \mathcal{L}_u} f_u(\ell)\, \log_2 f_u(\ell). \tag{6.16}$$

where $f_u(\ell)$ is the location frequency of location ℓ.

Temporal-uncorrelated entropy

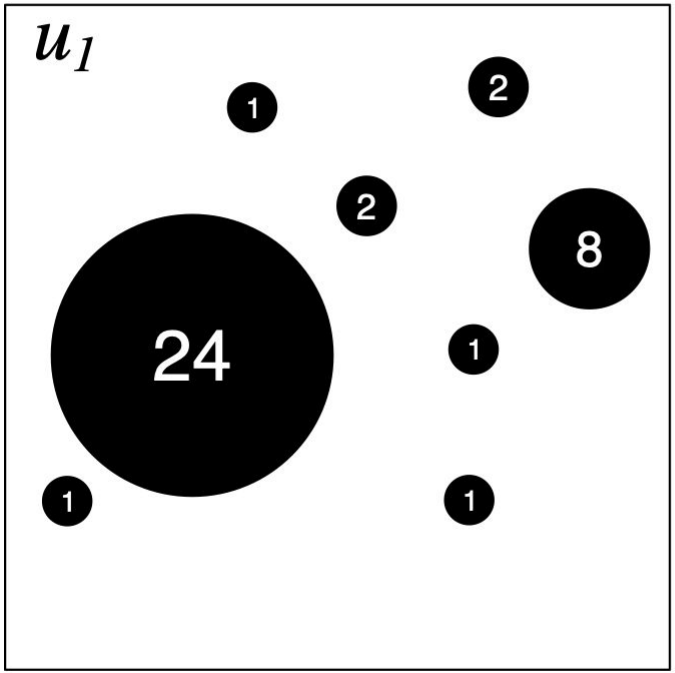
It is expressed in bits as it is defined using logarithms in base 2. For example:

$$S_{\mathrm{unc}}(u) = 3\mathrm{bits}$$

means that the uncertainty is equivalent to that of choosing uniformly at random among:

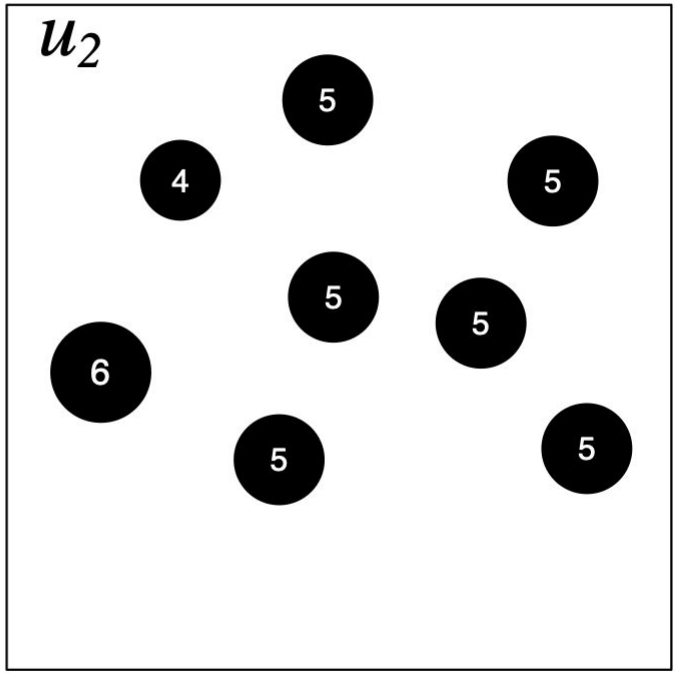
$$2^3 = 8$$ locations

Temporal-uncorrelated entropy



visits = 40

locations = 8



Visit distribution p_i	S_{unc} (bits)	$2^{S_{\text{unc}}}$	Interpretation
$(1, 0, 0, 0)$	0.00	1.00	Always in one place (zero uncertainty)
$(0.9, 0.1, 0, 0)$	0.47	1.38	Almost always in one place
$(0.7, 0.2, 0.1, 0)$	1.16	2.23	Mostly two relevant locations
$(0.5, 0.25, 0.125, 0.125)$	1.75	3.36	About 3–4 effective locations
$(0.4, 0.3, 0.2, 0.1)$	1.85	3.60	Fairly spread across locations
$(0.25, 0.25, 0.25, 0.25)$	2.00	4.00	Uniform distribution (max uncertainty)

DEFINITION 6.8

Real entropy

Let $L_u = (\ell_1, \ldots, \ell_m)$ be the location sequence of an individual u.
The *real entropy* of u is defined as:

$$S_{\text{real}}(u) = -\sum_{L'_u} P(L'_u) \log_2 P(L'_u) \tag{6.17}$$

where the sum runs over all time-ordered subsequences L'_u of L_u, and $P(L'_u)$ denotes the probability of observing the subsequence L' in the sequence.

Real entropy

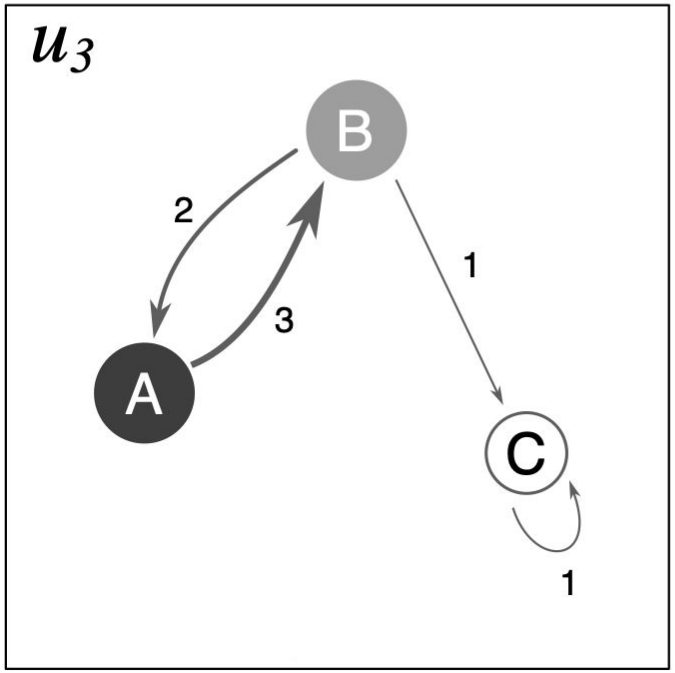
It is the effective number of possible next locations once temporal correlations are taken into account. For example:

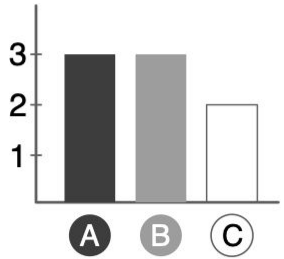
$$S_{\text{real}}(u) = 0.8 \text{ bits per step}$$

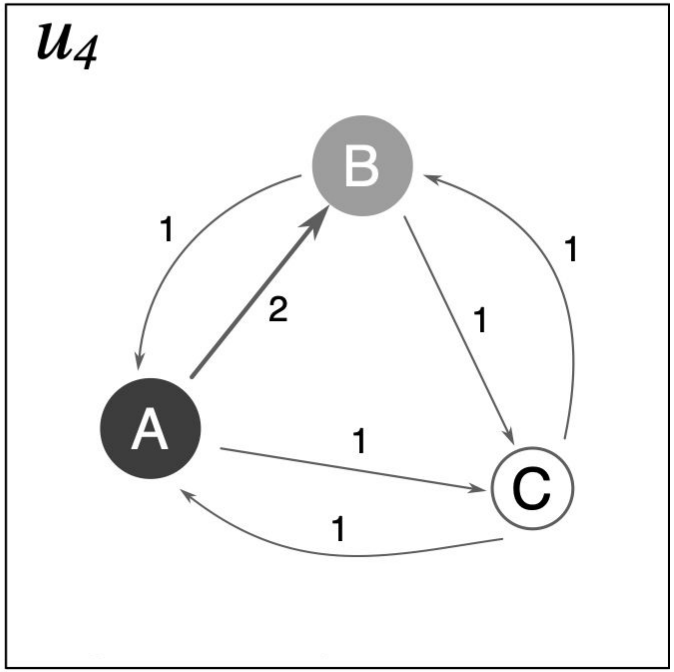
at each step the individual behaves as choosing among:

$$2^{0.8} \approx 1.74 \text{ locations}$$

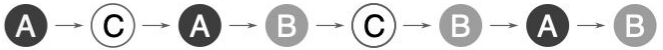
Real entropy











Temporal-uncorrelated vs Real entropy

✅ $S_{\mathrm{real}}(u) \leq S_{\mathrm{unc}}(u)$

or

$S_{\mathrm{real}}(u) \geq S_{\mathrm{unc}}(u)$

?

Computational complexity

For a sequence of length n, the number of contiguous subsequences is

$$\frac{n(n+1)}{2}$$

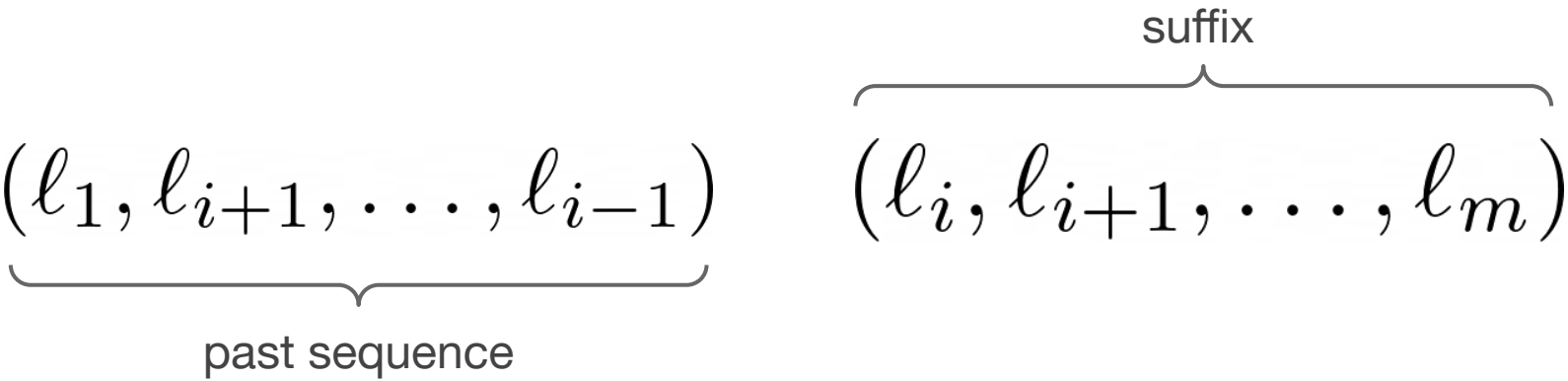
which scales quadratically with n.

$$n = 1000 \rightarrow (1000 \cdot 1001)/2 = 500,500$$

Lempel-Ziv estimation

Intuition:

- if the sequence is regular, long subsequences repeat
- if it is irregular, new patterns appear frequently



Lempel-Ziv estimation

Intuition:

- if the sequence is regular, long subsequences repeat
- if it is irregular, new patterns appear frequently

suffix

$$\underbrace{(\ell_1, \ell_{i+1}, \ldots, \ell_{i-1})}_{\text{past sequence}} \qquad (\ell_i, \ell_{i+1}, \ldots, \ell_m)$$

Never appeared

$$\Lambda_i = 1$$

Lempel-Ziv estimation

Intuition:

- if the sequence is regular, long subsequences repeat
- if it is irregular, new patterns appear frequently

$$(\ell_1, \ell_{i+1}, \ldots, \ell_{i-1}) \qquad (\ell_i, \ell_{i+1}, \ldots, \ell_m)$$

Never appeared

$$\Lambda_i = 2$$

Lempel-Ziv estimation

Intuition:

- if the sequence is regular, long subsequences repeat
- if it is irregular, new patterns appear frequently

$$\hat{S}_{\text{real}} \approx \frac{m \log_2 m}{\sum_{i=1}^{m} \Lambda_i}$$

length of the sequence

An example

$$L_{u_3} = (\underset{i}{A}, B, A, B, A, B, C, C)$$

$i = 1$ $\Lambda_1 = 1$

because the past sequence is empty

An example

$$L_{u_3} = (\underset{i}{A}, B, A, B, A, B, C, C)$$

$$i = 2 \qquad \Lambda_2 = 1$$

An example

$$L_{u_3} = (\boxed{A, B}, \underset{i}{\boxed{A}}, B, A, B, C, C)$$

$$i = 3$$

An example

$$L_{u_3} = (\underbrace{A, B}_{i}, \underbrace{A, B}_{i}, A, B, C, C)$$

$$i = 3$$

An example

$$L_{u_3} = (\underbrace{A, B}_{i}, A, B, A, B, C, C)$$

$$i = 3 \qquad \Lambda_3 = 3$$

An example

$$L_{u_3} = (\underbrace{A, B, A}, \underbrace{B}_{i}, A, B, C, C)$$

$$i = 4$$

An example

$$L_{u_3} = (\underbrace{A, B, A}_{i}, \underbrace{B, A}_{i}, B, C, C)$$

$$i = 4$$

An example

$$L_{u_3} = (\underbrace{A, B, A}_{i}, B, A, B, C, C)$$

$$i = 4 \qquad \Lambda_4 = 3$$

An example

$$L_{u_3} = (\boxed{A, B, A, B}, \underset{i}{\boxed{A}}, B, C, C)$$

$$i = 5$$

An example

$$L_{u_3} = (\underbrace{A, B, A, B}_{i}, \underbrace{A, B}_{i}, C, C)$$

$$i = 5$$

An example

$$L_{u_3} = (\underbrace{A, B, A, B}_{i}, A, B, C, C)$$

$$i = 5 \qquad \Lambda_5 = 3$$

An example

$$L_{u_3} = (A, B, A, B, A, \underset{i}{B}, C, C)$$

$$i = 6$$

An example

$$L_{u_3} = (\underbrace{A, B, A, B, A}, \underbrace{B, C}_{i}, C)$$

$$i = 6 \qquad \Lambda_6 = 2$$

An example

$$L_{u_3} = (\underbrace{A, B, A, B, A, B}_{i}, C, C)$$

$$i = 7 \qquad \Lambda_7 = 1$$

An example

$$L_{u_3} = (\underbrace{A, B, A, B, A, B, C}, \underbrace{C}_{i})$$

$$i = 8 \qquad \Lambda_8 = 1$$

Because the end of the sequence has been reached

An example

$$(\Lambda_1, \Lambda_2, \ldots, \Lambda_8) = (1, 1, 3, 3, 3, 2, 1, 1)$$

$$m = 8$$

$$\sum_{i=1}^{8} \Lambda_i = 15$$

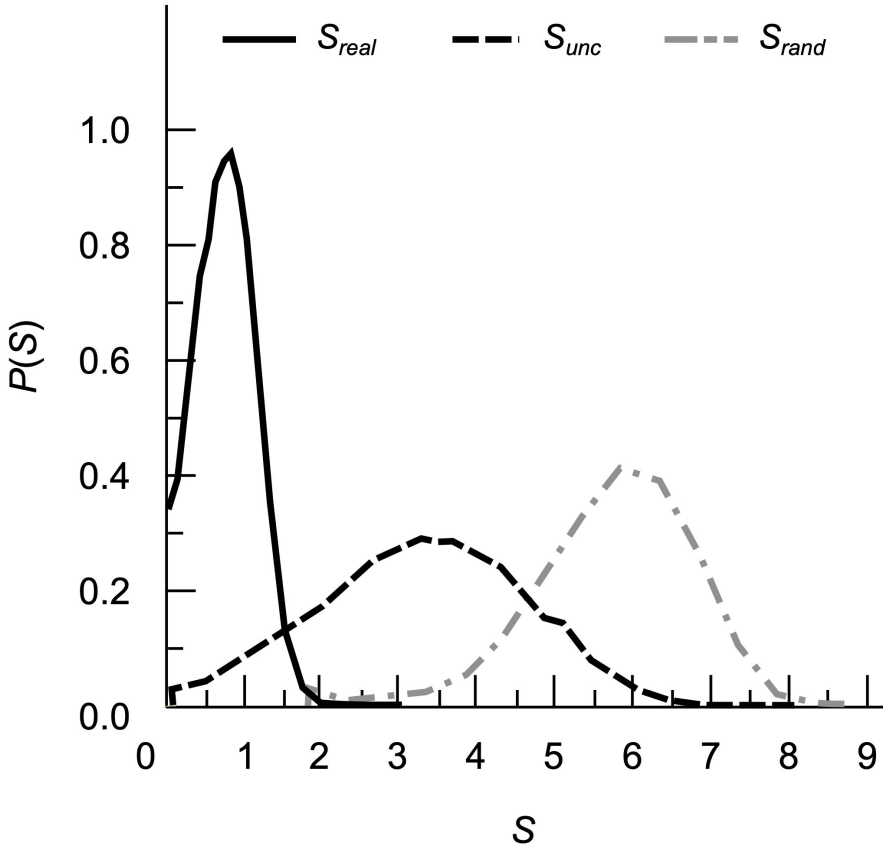
$$\hat{S}_{\text{real}}(u_3) \approx \frac{8 \log_2 8}{15} \approx 1.6$$

bits per step

Song et al., 2010:

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

$$2^{0.8} = 1.74$$



QUIZ

What does a power-law tail in the displacement-length distribution indicate?

A) All displacements are approximately the same length
B) Long displacements never occur
✅ C) Distribution has no characteristic distance scale in its tail
D) Displacement length is independent of distance

QUIZ

Under Zipf’s law, how does the frequency of a location depend on its rank k?

A) It increases in proportion to k.

✅ B) It decreases approximately in proportion to 1/k.

C) It is the same for every rank.

D) It decreases exponentially with k.

QUIZ

What does the radius of gyration measure?

A) How far a person travels in a typical day

✅ B) The spatial spread of visited locations

C) How often a person returns to their home

QUIZ

The radius of gyration increases logarithmically with time and then saturates. What does this suggest?

✅ A) The explored area expands, but its growth slows over time.

B) The person’s movements become progressively faster.

C) The person stops visiting previously known locations.

D) The explored area grows at a constant rate.

QUIZ

What best describes the distinction between returners and explorers?

✅ A) Returners repeatedly visit familiar locations; explorers tend to expand their spatial range.

B) Returners visit many locations once; explorers revisit only one location.

C) Returners move preferably within cities; explorers move between countries.

QUIZ

What does temporal-uncorrelated entropy use to describe an individual’s mobility?

A) The order of visits, without considering their frequency

✅ B) The frequency of visits to locations, without considering their order

C) The distance between every pair of locations

D) The time spent at each location, but not the locations themselves

QUIZ

Two people visit the same locations with the same frequencies, but one follows a much more repetitive routine. Which statement is correct?

A) Their temporal-uncorrelated entropies must differ.

✅ B) Their temporal-uncorrelated entropies are the same, but their real entropies may differ.

C) Their real entropies must be the same.

D) Neither type of entropy can be calculated.

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