Understanding individual human mobility patterns

Marta C. Gonzalez, Cesar A. Hidalgo & Albert-Laszlo Barabasi

Nature Vol.453, 5 Jun 2008

Motivation

- Importance of human mobility patterns
  - Urban planning
  - Traffic forecasting
  - Spread of biological
  - Mobile viruses
- Lack of tools to monitor the individuals' locations

Random Walk Model

- Human trajectories are often approximated with various random walk or diffusion models.
  - Many unknown influence (job- and family-imposed restrictions)
- Animal trajectory is approximated by Levy flight.
  - A random walk for which step size Δr follows a power-law distribution $P(\Delta r) \sim \Delta r^{-(1+\beta)}$
  - This finding has been generalised to humans.

Bank Note Dispersal

- Bank note dispersal is a proxy for human movement.
  - Money is carried by individuals.
- Human trajectories are best modelled as a continuous-time random walk.
  - Most of the time we travel only over short distances.
  - Occasionally we take longer trips.

Bank Note Dispersal

- Each consecutive sighting of a bank note reflects the composite motion of two or more individuals who owned the bill.
- It is unclear whether the distribution reflects
  - 1. the motion of individual users
  - 2. some previously unknown convolution between population-based heterogeneities and individual human trajectories

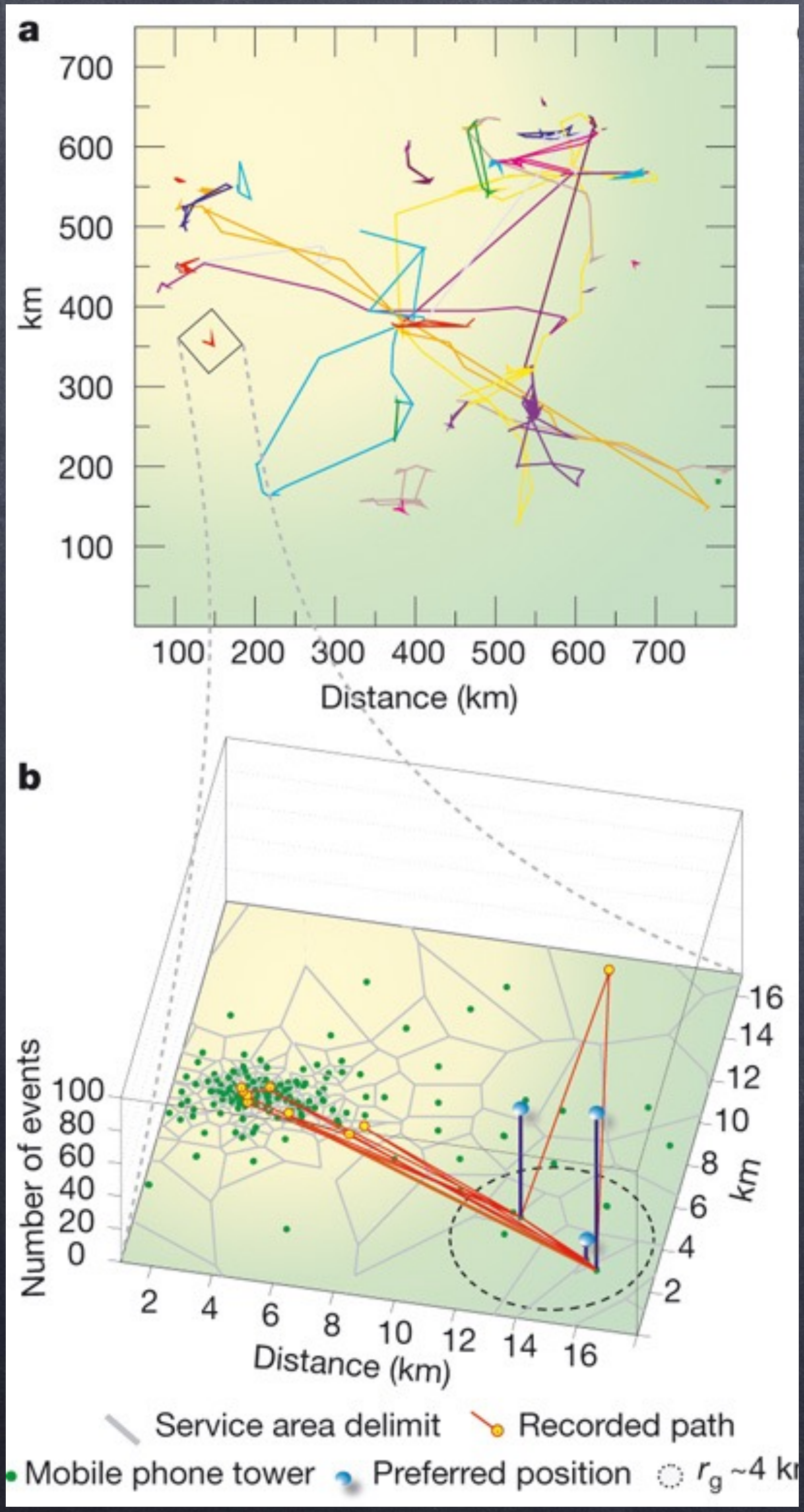
Mobile Phones

- Mobile phones are carried by the same individual during his/her daily routine
  - They offer the best proxy to capture individual human trajectories.

Data Sets

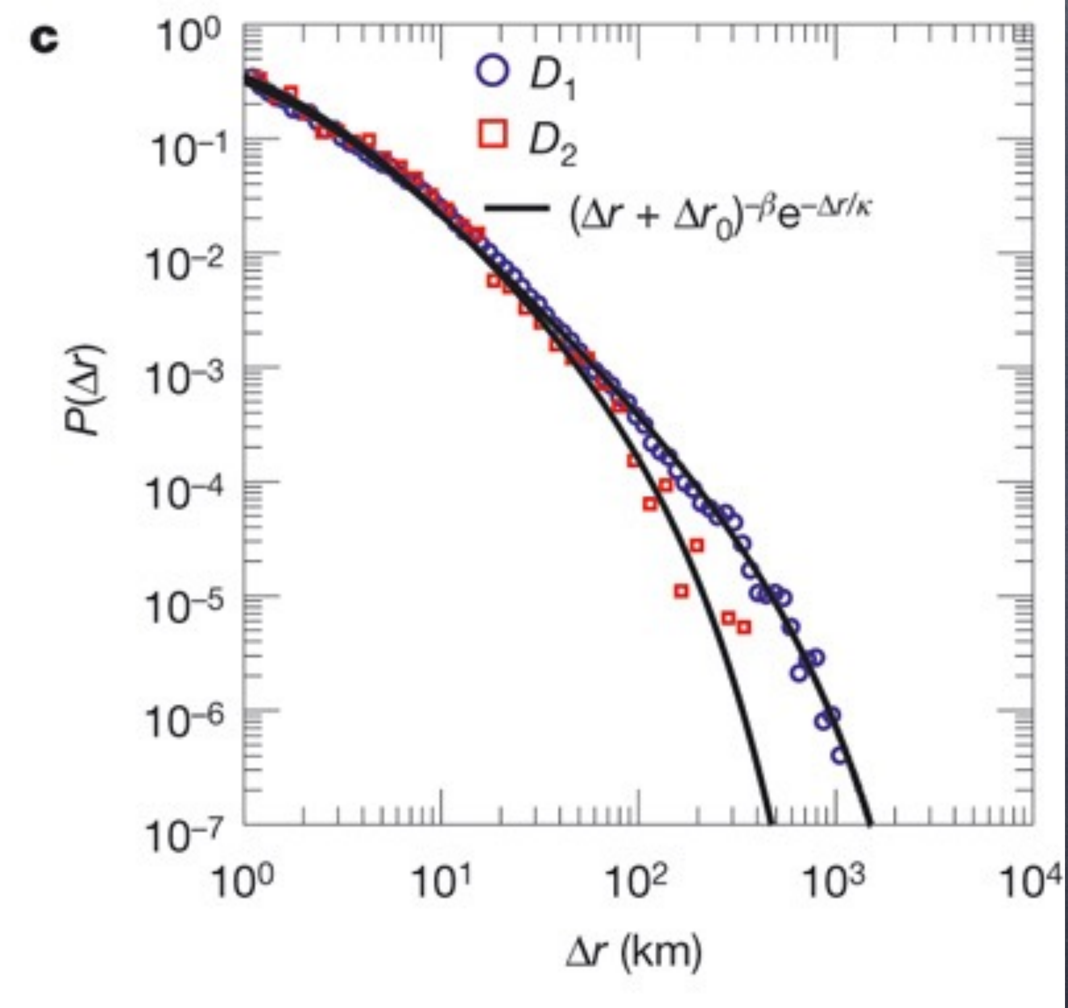
- D1: mobility patterns recorded over a 6-month period for 100,000 individuals
  - Individuals were selected randomly from a sample of more than 6 million anonymised mobile phone users.
  - The location of the tower was recorded each time a user initiated or received a call or a text message.
  - The time between consecutive call followed a bursty pattern.
- D2: the location of 206 mobile phone users
  - The location was recorded every two hours for an entire week.

Data Sets



- a: Week-long trajectory of 40 mobile phone users
- b: The detailed trajectory of a single user
  - 186 two-hourly reports

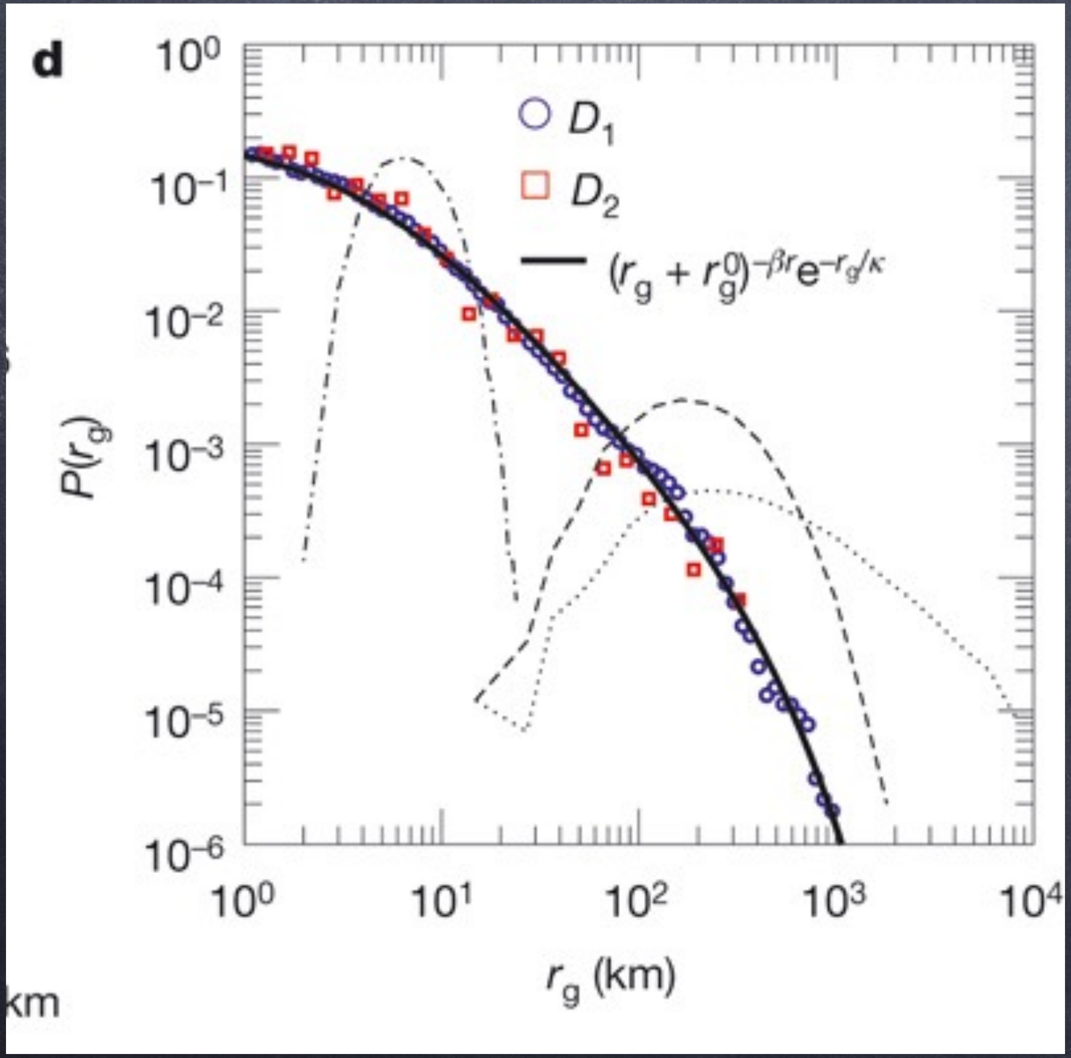
Jump Size Distribution



- Truncated power-law distribution
  - $P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} exp(-\Delta r/\kappa)$
- Three distinct hypotheses
- A. Each individual follows a Levy flight
- B. The distribution captures a population-base heterogeneity
- C. A convolution of hypotheses of A and B

Radius of Gyration

- $r_g(t)$: characteristic distance travelled by user a during time t.



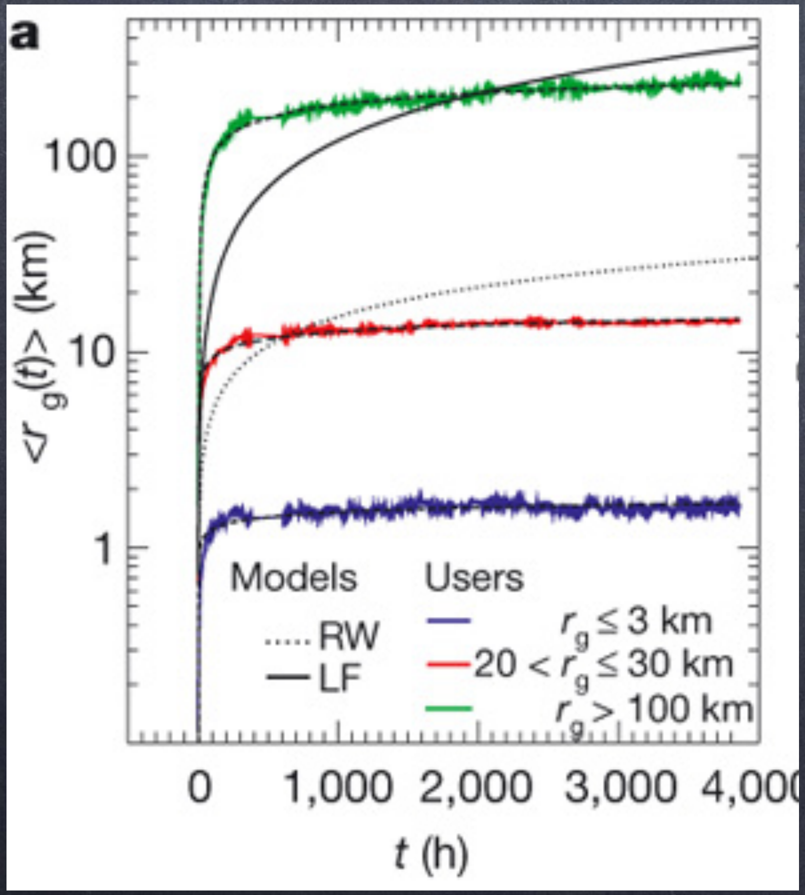
- Truncated power-law
  - $P(r_g) = (r_g + r_g^0)^{-\beta} exp(-r_g/\kappa)$
- An ensemble of agents following
  - Dotted: Random Walk(RW)
  - Dashed: Levy flight(LF)
  - Dot-dashed: Truncated Levy flight(TLF)

Radius of Gyration

- All these were not sufficient to explain the truncated power-law distribution $P(r_g)$
- The difference in the range of typical mobility patterns of individuals r_g has a strong impact on the truncated Levy behaviour of individuals.
- Ruling out hypotheses A.

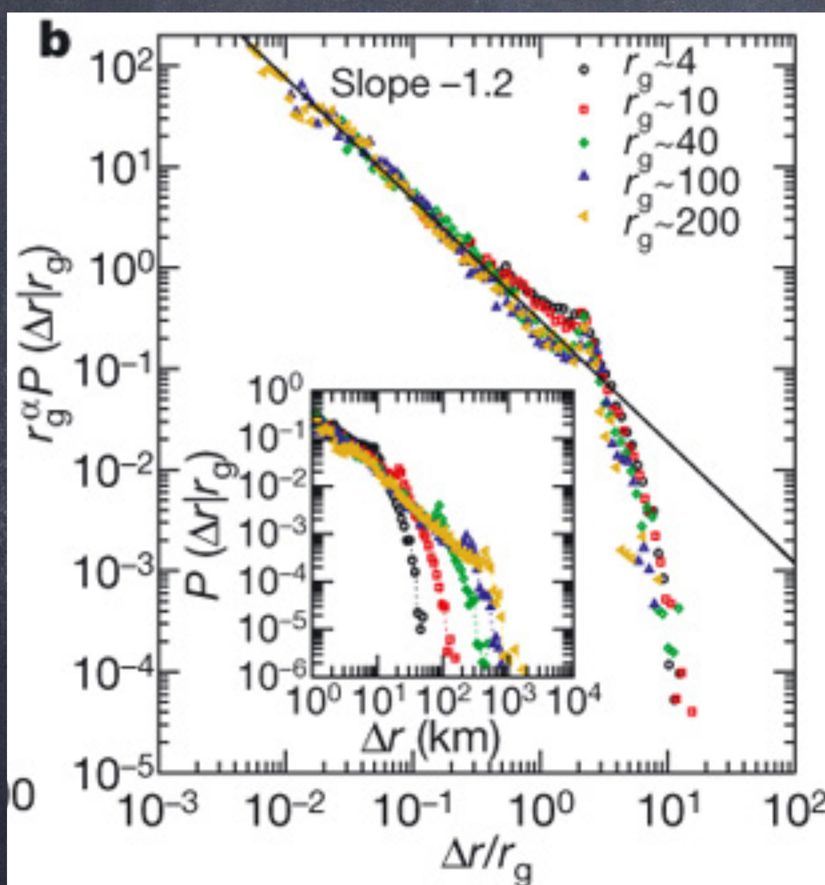
Time Dependence of r_g

- For RW, LF and TLF, r_g should increase with time
  - The longer we observe a user, the higher the chance that he/she travel farther.



- Logarithmic increase
  - A manifestly slower dependence
  - A saturation process

Jump Size Distribution $P(\Delta r|r_g)$



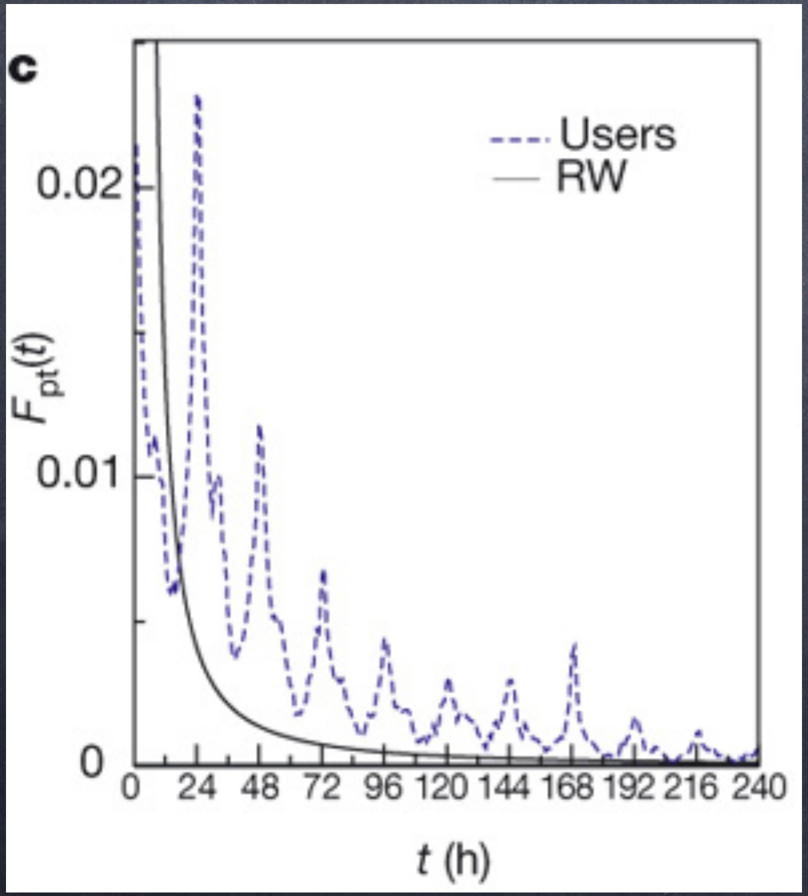
- Inset: small/large r_g - s/l distance
- Rescale: collapse into a curve
- $P(\Delta r|r_g) \sim r_g^{-\alpha} F(\Delta r/r_g)$
- $F(x)$: r_g-independent
  - $F(x) \sim x^{-\alpha}$ for x < 1
  - $F(x)$: rapidly decrease for x >> 1
- Individuals' travel patterns may be approximated by Levy flight up to a distance characterised by r_g

Jump Size Distribution $P(\Delta r|r_g)$

- $P(\Delta r) = \int_0^\infty P(\Delta r|r_g)P(r_g)dr_g$
- The jump size distribution $P(\Delta r)$ is the convolution between the statistics of individual trajectories $P(\Delta r|r_g)$ and the population heterogeneity $P(r_g)$
- Consistent with hypothesis C.

Mechanism Stabilising r_g

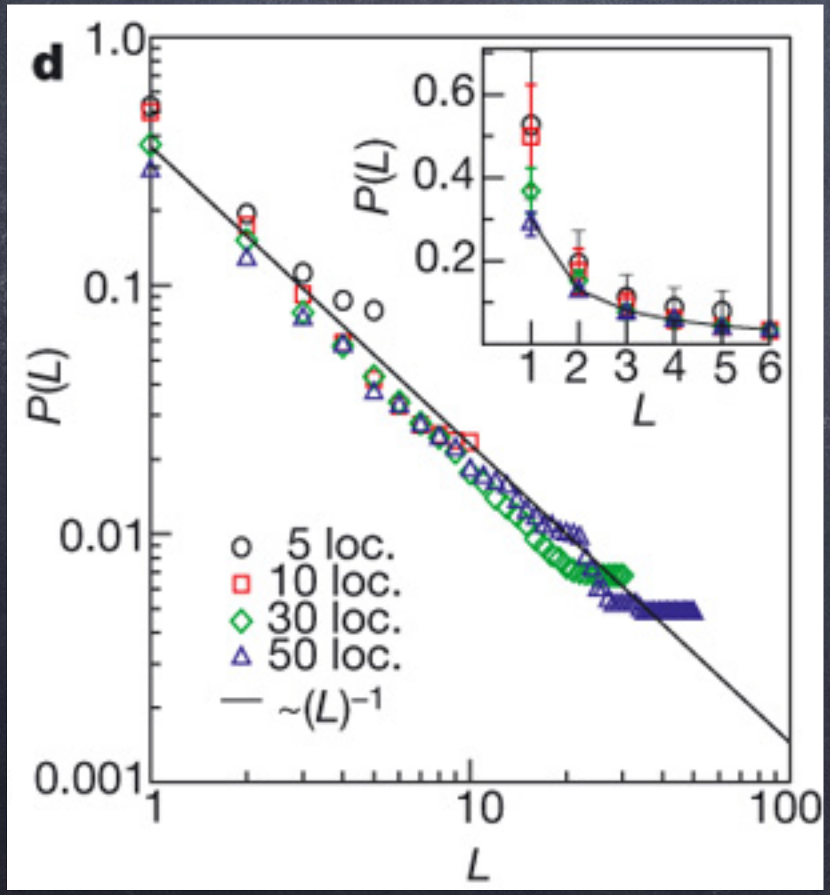
- The return probability $F_{pt}(t)$: a user returns to the position where he/she was first observed after t hours.



- The return probability is characterised by several peaks at 24h, 48h and 72h.

Mechanism Stabilising r_g

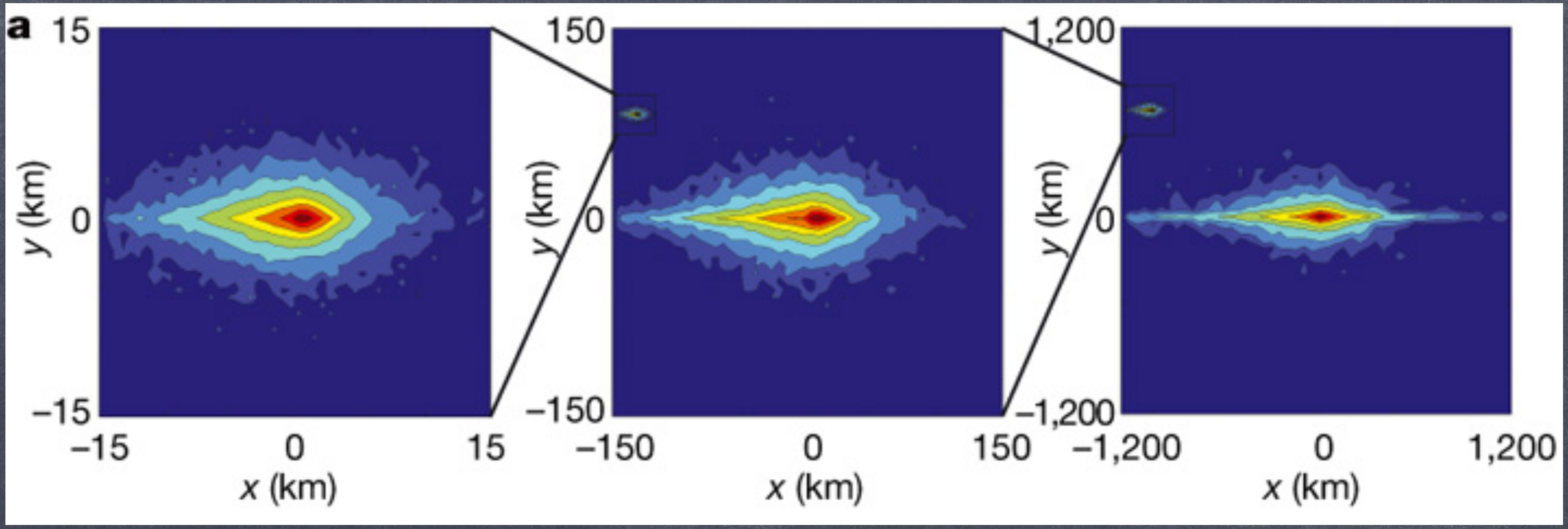
- We ranked each location on the basis of the number of times an individual was recorded there.



- Rank L: $P(L) \sim 1/L$
- The observed logarithmic saturation of $r_g(t)$ is rooted in the high degree of regularity.

Probability Density Function

- $\Phi_a(x, y)$: finding an individual a in a position (x,y)

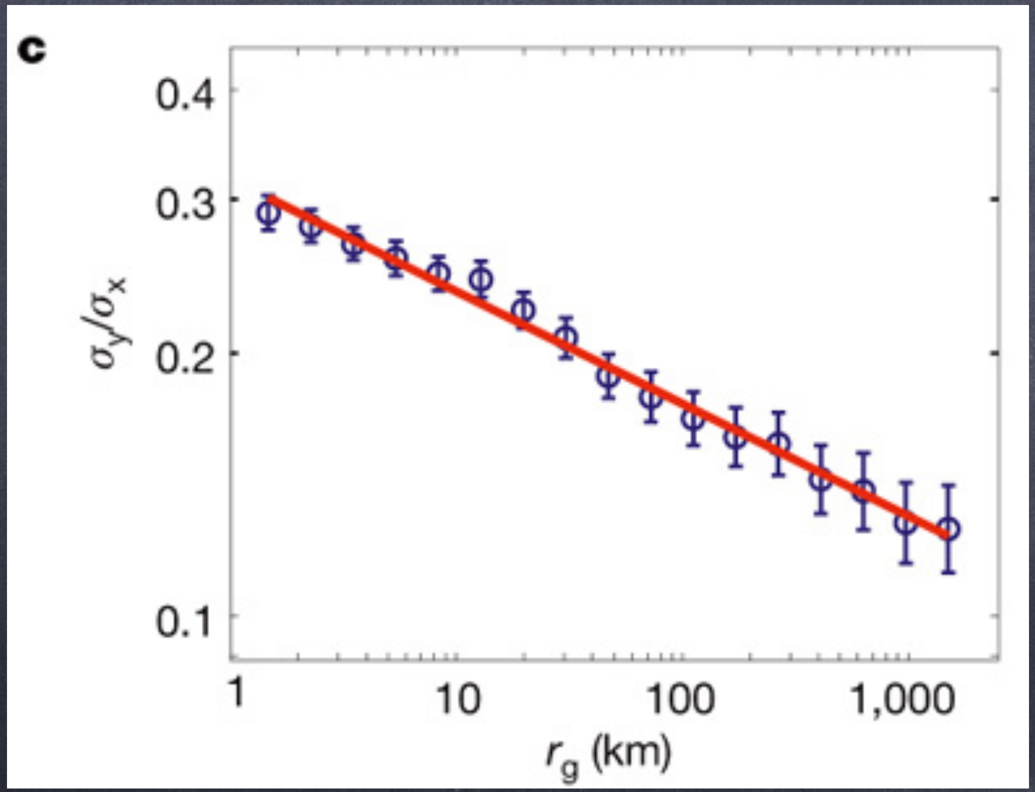


- The larger an individuals' r_g is, the more pronounced is this anisotropy.

Probability Density Function

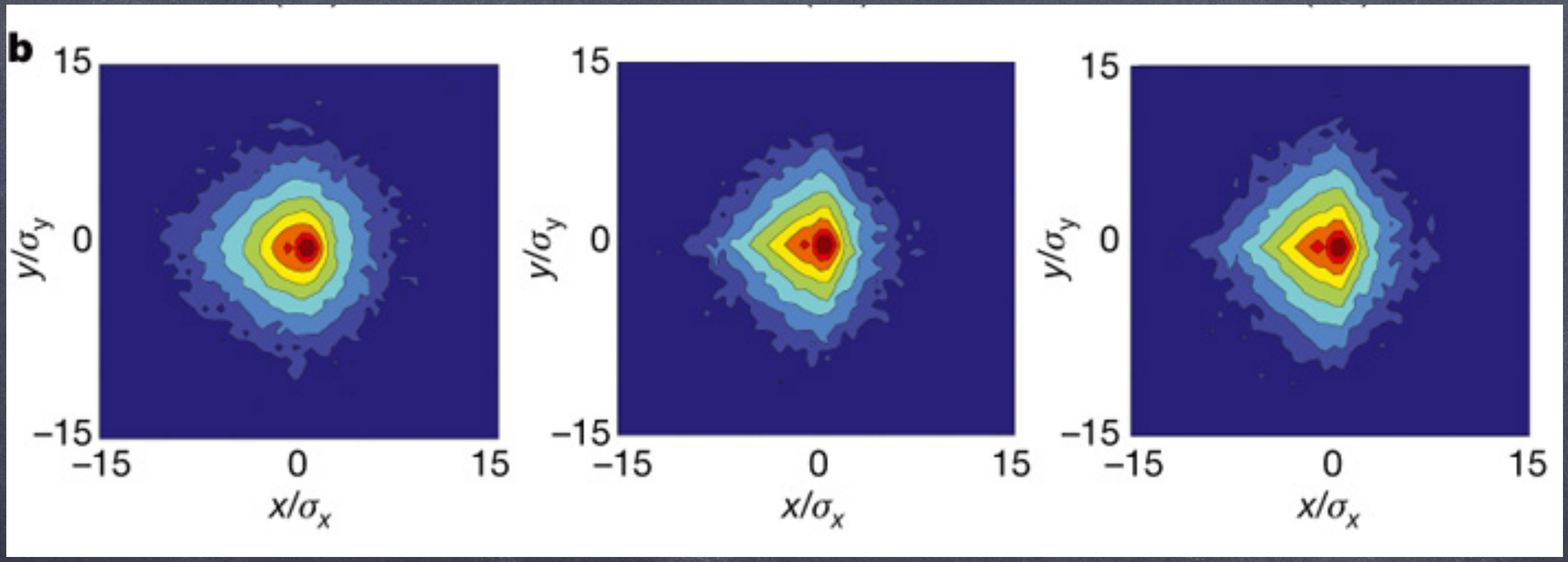
- The anisotropy ratio $S = \sigma_y/\sigma_x$

  $S \sim r_g^{\eta}$



Probability Density Function

- $\tilde{\Phi_a}(x/\sigma_x, y/\sigma_y)$: remove the anisotropies



- Similar for all groups of users

Conclusion

- Step size distribution capture a convolution of the population heterogeneity and the motion of individual users.
- Individuals display significant regularity.
- Key statistical characteristics of individual trajectories are largely indistinguishable after rescaling.
  - We can obtain the likelihood of finding a user in any location.