Knowledge Discovery and Data Mining Laboratory - ISTI CNR

The origin of heterogeneity in human mobility

Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti

The patterns of mobility

In the last decade, the emergence of big mobility data allowed scientists from different disciplines to discover that our movements are not random, but follow their own laws:

- the mobility of an individual can be confined within a stable circle (defined by a *center of mass* and a *radius of gyration*);
- such circles are highly heterogeneous, since a power law was found in the distribution of the radius of gyration;
- despite the observed heterogeneity, by observing the past history the whereabouts of most individual can be predicted with very high accuracy.

Although these discoveries have doubtless shed light on interesting and fascinating aspects about human mobility, the origin of the observed patterns still remains unclear:

Why do we move so differently? What are the factors that shape our mobility? Which movements mainly determine the mobility of an individual?

Unveiling heterogeneity

In our data-driven study of human mobility, we exploit the access to two big mobility datasets:

- a GSM dataset of 67,000 users active in a big European country;
- a GPS dataset of 46,121 cars traveling in the region of Tuscany, Italy.

For each user in the two datasets, we computed several measures to characterize their individual mobility:

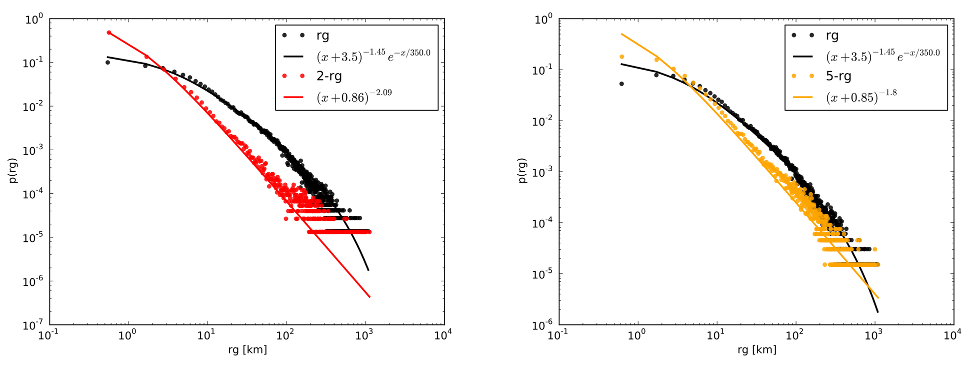
- **the radius of gyration** r_g, defined as

$$r_g = \sqrt{\frac{1}{N}\sum_{i \in L}(\vec{r}_i - \vec{r}_{cm})^2},$$

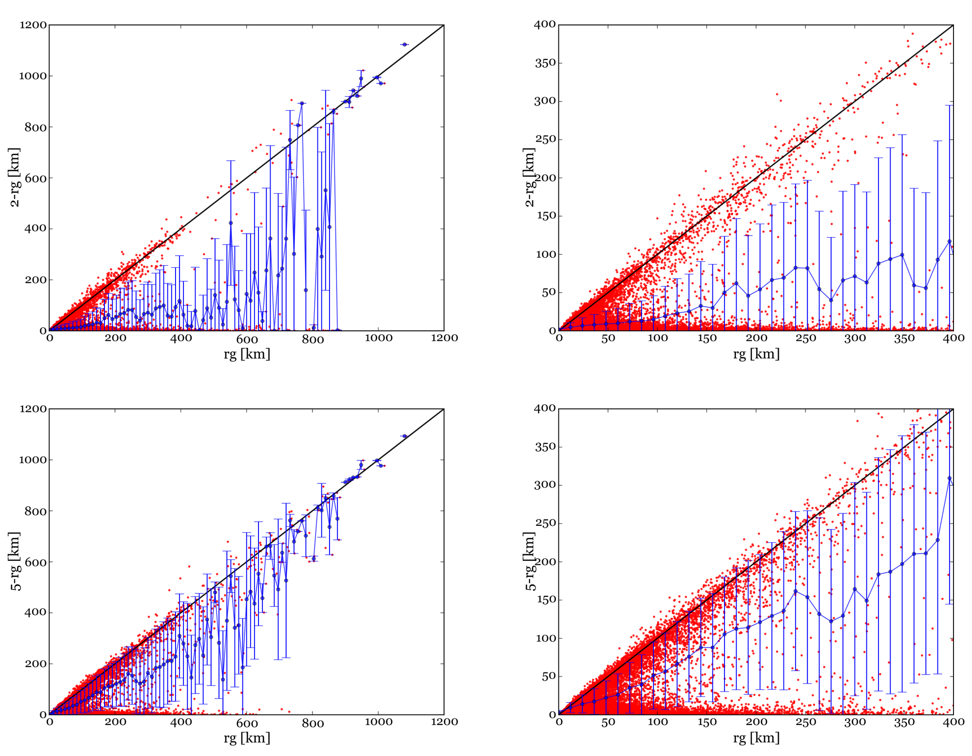
  where $\vec{r}_i$ represents the positions recorded for the user, and $\vec{r}_{cm}$ is the user's center of mass;
- the **k-radius of gyration** k-r_g, that is the radius of gyration computed on the k most frequent locations of a user;
- the **cluster radius of gyration** C-r_g, given a partition in m clusters of user's locations, it is the radius of gyration computed on the m most frequent locations of each cluster.

Results

We start our analysis by comparing the distributions of r_g and k-r_g. While the former is well fitted by a power law ($\beta = -1.45$) with an exponential cutoff, the latter losses the cutoff and shows a higher exponent for the power law ($\beta = -2.09$). The higher the k, the closer the k-r_g curve is to the r_g one, the exponent of the power law gradually decreases and the exponential cutoff starts to emerge.



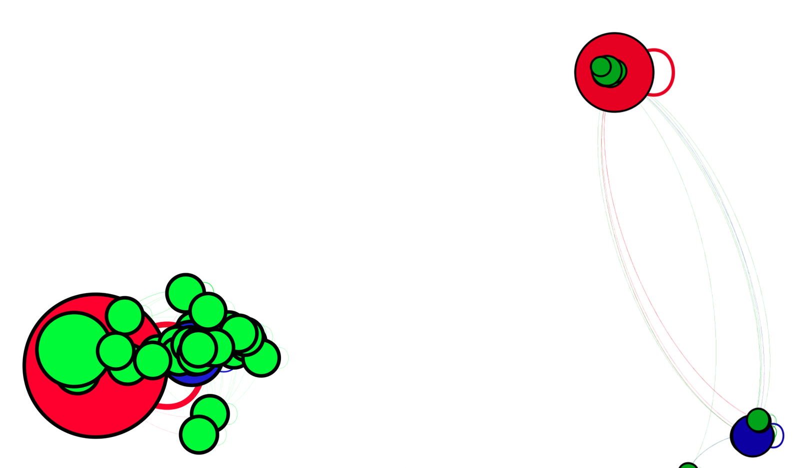
From the correlation between r_g and 2-r_g two different profiles clearly emerge: the radius of users on the "diagonal" is mainly determined by the two most important locations. Conversely, for the other category of users (the majority) two location are not sufficient to synthesize their characteristic traveled distance. Such tendency is independent from the scale and from the number k of locations considered into the computation of the k-radius.



What is the shape of the mobility of users in the two profiles?

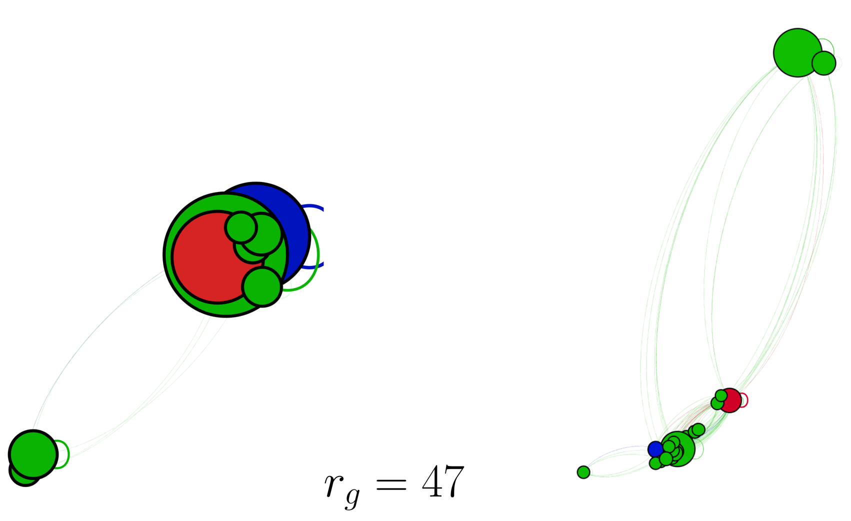
To address this question, we randomly chosen some individuals and reconstructed their individual mobility networks (a graph where nodes are locations and edges routes between locations).

Two insights emerge from the observation of the networks of diagonal users: i) the higher the radius of gyration, the higher the distance between L_1 and L_2; ii) less frequent locations tend to thicken around L_1 and L_2, like satellites gravitating around major planets.



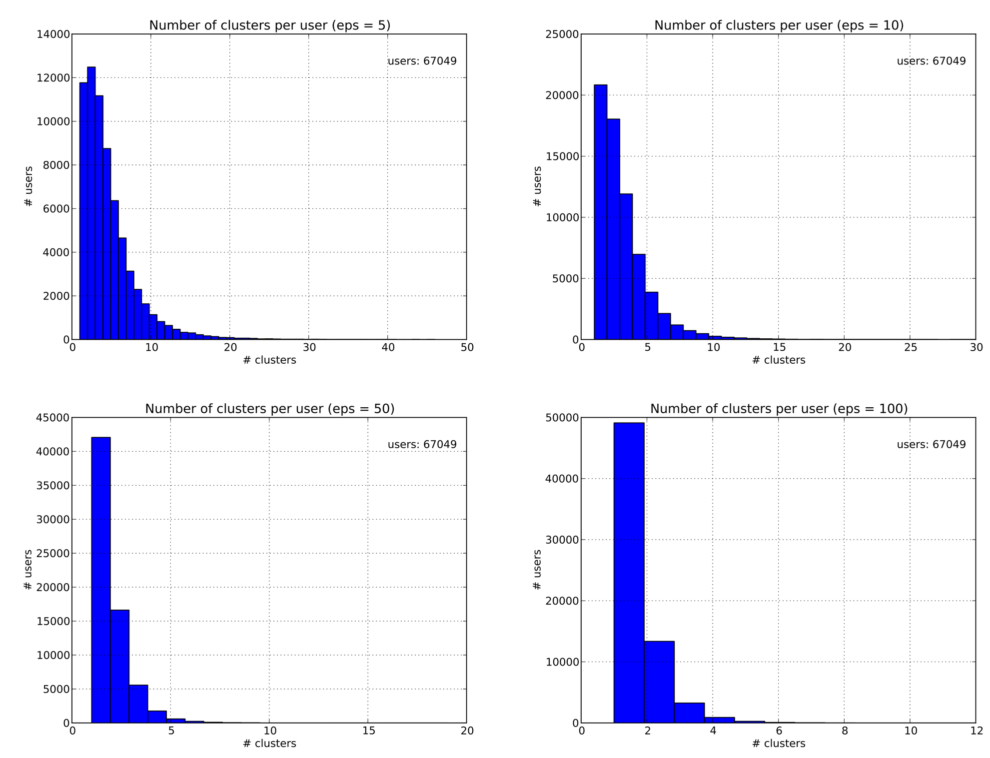
$r_g = 8$ $r_g = 48$

In the other set of users, L_1 and L_2 tend to be close independently of the value of the radius, while other less frequent locations appear very far apart from the L_1-L_2 group.

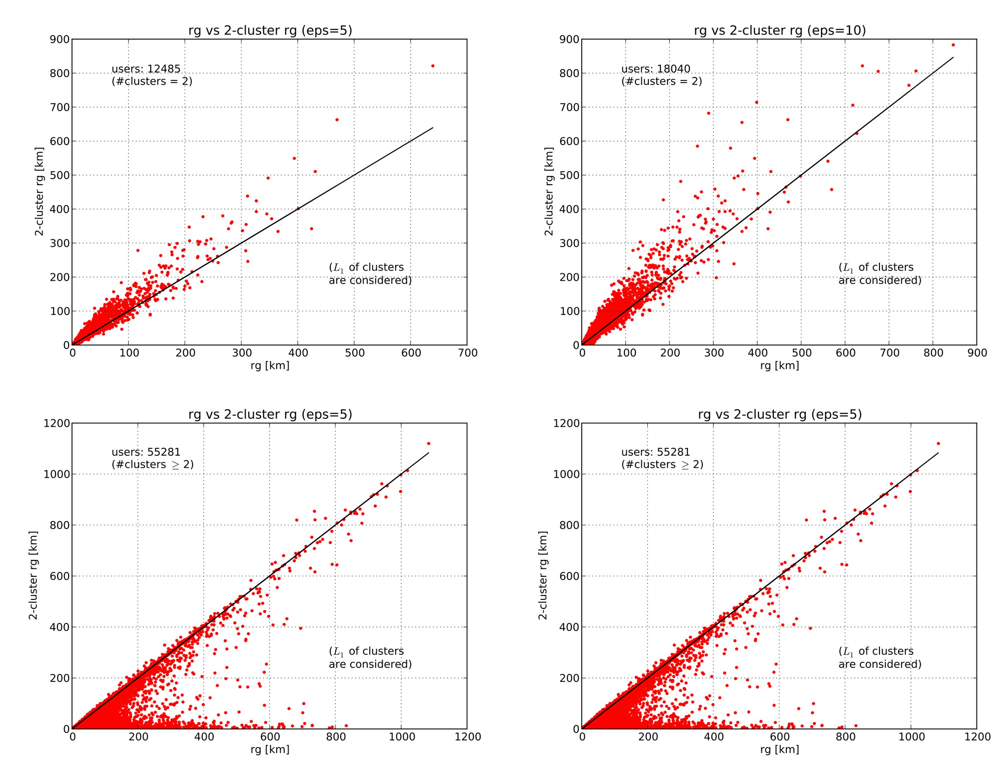


$r_g = 14$ $r_g = 47$

This suggests that ***the mobility of some individuals could be modeled in group of locations***, representing different urban areas, or different **mobility hearts**. To test this hypothesis, for each user we computed clusters of her locations through the DB-SCAN algorithm, using eps = 5, 10, 50, 100 km and minPts = 1 as parameters.

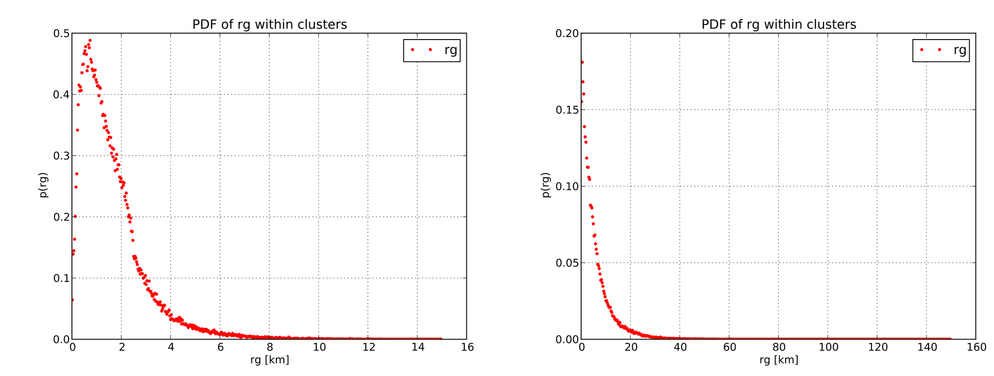


A strong linear correlation emerges from a comparison between rg and C-r_g, meaning that **our characteristic traveled distance is mainly determined by the dominant locations in our mobility hearts**. Indeed, if we discard some clusters, the splitting of the population into two profiles appears again in the r_g vs C-r_g correlation.



The distribution of the r_g within the mobility hearts is not a power law anymore: a peak now emerges from the distribution, suggesting that, **inside each mobility heart, users show a typical value in the characteristic traveled distance**.

In general, it is the distance between different mobility hearts that generates the observed heterogeneity, which is greatly reduced within the mobility hearts.



Contacts

Luca Pappalardo: lpappalardo@di.unipi.it
Filippo Simini: flppsmn@gmail.com
Salvatore Rinzivillo: rinzivillo@isti.cnr.it

European Conference on Complex Systems (ECCS'13) `http://www.eccs13.eu/`