The origin of heterogeneity in human mobility ranges

Luca Pappalardo
Department of Computer Science
University of Pisa
Largo Bruno Pontecorvo 3, 56127 Pisa, Italy
lpappalardo@di.unipi.it

ABSTRACT

In the last decade, scientists from different disciplines discovered a great heterogeneity in human mobility ranges, since a power law characterizes the distribution of the characteristic distance traveled by individuals, the so-called radius of gyration. The origin of such heterogeneity, however, still remains unclear. In this paper, we analyze two mobility datasets and observe that an individual's locations tend to be grouped in dense clusters representing geographical mobility cores. We show that the heterogeneity in human mobility ranges is mainly due to trips between these mobility cores, while it is greatly reduced when individuals are constrained to move within a single mobility core.

CCS Concepts

•**Applied computing** → *Physics; Mathematics and statistics;*

Keywords

human mobility; mobility data mining; mobile phone data; GPS data; data science; Big Data

1. INTRODUCTION

In the last decade the availability of big mobility data, such as GPS tracks from vehicles and mobile phone data, offered a series of novel insights on the quantitative patterns characterizing human mobility. In particular, scientists from different disciplines discovered that human movements are not completely random but follow specific statistical laws. The mobility of an individual can be confined within a stable circle defined by a center of mass and a radius of gyration [7, 12]. Interestingly, such circles are found to be highly heterogeneous since a power law characterizes the distribution of the radius of gyration of individuals [7, 14]. Although these discoveries have doubtless shed light on interesting aspects about human mobility, the origin of the observed patterns still remains unclear: what is the origin of the heterogeneity in human mobility ranges? Answering this question is of great importance in contexts like urban planning and the design of smart cities, since it can be helpful for crucial problems such as movement prediction [3, 20] and activity recognition [11, 8, 15].

In this paper, we address this question by performing a data-driven study of human mobility. In our analysis we exploit the access to two mobility datasets, each storing the trajectories of about 50,000 individuals. We observe that the locations visited by the individuals tend to cluster in dense groups, representing meaningful geographical units or *mobility cores*. We then compute for every individual her inter-core characteristic traveled distance and her intra-core characteristic traveled distance, which are defined by the radius of gyration computed on the trips between mobility cores and the trips within mobility cores respectively. From the comparison of the total radius of gyration of an individual with her intra- and inter-core radius of gyration we observe two main results. First, a strong linear correlation emerges between the total radius of an individual and her inter-core radius, suggesting that the mobility range of an individual is mainly determined by trips between mobility cores. Second, the distribution of the characteristic intra-core radius of gyration has a peak suggesting that individuals show typical mobility ranges when constrained to move within mobility cores. Our results, which emerge on different types of mobility data and at different geographical and temporal scales, suggest that people perform two types of trips: intra-core trips and inter-core trips, the latter being the origin of the observed heterogeneity in mobility ranges.

The paper is organized as follows. Section 2 summarizes some works relevant to our topic. Section 3 introduces the two mobility datasets we analyze and Section 4 describes the measures of individual human mobility we use during the analysis. Section 5 shows the results of our work and finally Section 6 concludes the paper.

2. RELATED WORK

The availability of Big Data on human mobility allowed scientists from different disciplines to discover that traditional mobility models adapted from the observation of animals [5, 6] and dollar bills [2] are not suitable to describe people's movements. Indeed, at a global scale humans are characterized by a huge heterogeneity, since a power law emerges in the distribution of the radius of gyration, the characteristic distance traveled by individuals [7, 12]. Despite this heterogeneity, through the observation of past mobility history the whereabouts of most individuals can be predicted with an accuracy higher than 80% [4, 18]. Moreover, according to their recurrent and total mobility patterns individuals naturally split into two distinct mobility profiles, namely returners and explorers, which show communication preferences with individuals in the same mobility profile [14].

The patterns of individual human mobility have been observed in both GSM data and GPS data [7, 12], and have been used to build generative models of individual human mobility [10, 18, 14], generative models to describe human migration flows [17, 21, 9], methods to discover geographic borders according to recurrent trips of private vehicles [16], methods to predict the formation of social ties [3, 20], and classification models to predict the kind of activity associated to individuals' trips on the only basis of the observed displacements [11, 8, 15]. Bagrow et al. exploit network science techniques to split the mobility of individuals into mobility units, or mobility habitats [1]. They find a relationship between the total radius of gyration of an individual and the trips between the main mobility habitats. In this paper we investigate the existence of mobility groups at different geographical levels. We use data mining clustering techniques (instead of network techniques) to aggregate an individual's locations into clusters.

3. MOBILITY DATA

GSM data. Our first data source consists of anonymized mobile phone data collected by a European mobile carrier for billing and operational purposes. The mobile phones carried by individuals in their daily routine offer a good proxy to study the structure and dynamics of human mobility: each time an individual makes a call the tower that communicates with her phone is recorded by the carrier, effectively tracking her current location. The datasets consists of Call Detail Records (CDR) describing the calls of 67,000 individuals during three months selected from 1 million users provided that they visited more than two locations during the observation period and that their average call frequency was $f \geq 0.5$ hour^{-1}. Each call is characterized by timestamp, caller and callee identifiers, duration of the call and the geographical coordinates of the tower serving the call. We reconstruct a user's movements based on the time-ordered list of phone towers from which a user made her calls [7].

GPS data. Our second data source is a GPS dataset storing information about the trips of 46,000 private vehicles traveling in Tuscany during one month. The GPS traces are provided by Octo Telematics[1], a company that provides a data collection service for insurance companies. The GPS device embedded into a vehicle's engine automatically turns on when the vehicle starts, and the sequence of GPS points that the device transmits every 30 seconds to the server via a GPRS connection forms the global trajectory of a vehicle. We exploit the stops of the vehicles to split the global trajectory into several sub-trajectories, corresponding to the trips performed by the vehicle. We set a stop duration threshold of at least 20 minutes to create the sub-trajectories, in order to avoid short stops like traffic lights: if the time interval between two consecutive observations of a vehicle is larger than 20 minutes, the first observation is considered as the end of a sub-trajectory and the second one is considered as the start of another sub-trajectory. We also performed the extraction of the sub-trajectories by using different stop duration thresholds (5, 10, 15, 20, 30 and 40 minutes) without finding significant differences in the sample of trips and in the statistical analysis we present in this paper. We assign each origin and destination point of the obtained sub-trajectories to the corresponding Italian census cell, using information provided by the Italian National Institute of Statistics (ISTAT). We describe the movements of a vehicle by the time-ordered list of census cells where the vehicle stopped [14].

[1]http://www.octotelematics.com/

GSM vs GPS. The GSM and the GPS datasets differ in several aspects [13, 12]. The GPS data refers to trips performed during one month (May 2011) in an area corresponding to a single Italian region, while the mobile phone data cover an entire European country and a period of observation of three months. The GPS data represents a 2% sample of the population of vehicles in Italy [12], while the mobile phone dataset covers users of a major European operator, about the 25% of the country's adult population [7, 14]. The trajectories described by mobile phone data include all possible means of transportation. In contrast, the GPS data refers to private vehicle displacements only. The fact that one dataset contains aspect missing in the other dataset makes the two types of data suitable for an independent validation of human mobility patterns.

4. MOBILITY MEASURES

The radius of gyration r_g is a standard measure to describe the characteristic distance traveled by an individual, defined as [7, 12]:

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

where L is the set of locations visited by the individual, $\mathbf{r}_i$ is a two-dimensional vector describing the geographical coordinates of location i; n_i is the visitation frequency of location i; $N = \sum_{i\in L} n_i$ is the total number of visits of the individual, and $\mathbf{r}_{cm}$ is the center of mass of the individual defined as the mean weighted point of the visited locations [7, 12]. The distribution of the radius of gyration is well fitted by a power-law with exponential cutoff, as measured on mobile phone data [7, 14] and GPS data [12, 14].

Given a partition of an individual's locations in m groups, or *mobility cores*, we define a *dominant location* D_i as the most visited location in group i, i.e. the preferred location of the individual when she visits locations in group i (see Figure 1). We define the *inter-core radius* r_g^{inter} of an individual as the radius of gyration computed on her m dominant locations ($m \geq 2$), and the *intra-core radius* r_g^{intra} as the radius of gyration computed on the locations of a given mobility core. Table 1 summarizes the mobility measures we use in our analysis and Figure 1 schematizes some of the concepts introduced above.

measure	symbol
radius of gyration	r_g
dominant location	D_i
intra-core radius of gyration	r_g^{intra}
inter-core radius of gyration	r_g^{inter}

Table 1: The mobility measures used in our study and the corresponding mathematical notation.

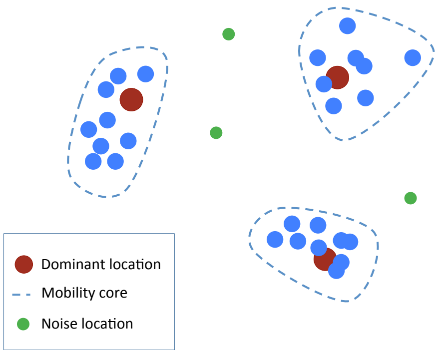


Figure 1: The image illustrates the locations visited by an individual. Blue circles are visited locations, groups of circles within blue dashed shapes are mobility cores, red circles are dominant locations. Green circles are noise locations that are not part of any mobility core. The radius of gyration is computed on all the circles, the inter-core radius on red circles, the intra-core radius on the circles within the same dashed shape.

5. RESULTS

For every individual in the two datasets, we partition her locations in mobility cores by using the DBSCAN clustering algorithm [19], which extracts dense groups of points according to two input parameters: *eps*, the maximum search radius; and *minPts*, the minimum number of points (locations) to form a cluster. Every location have two features, the latitude and the longitude of the location's position on the space. The DBSCAN algorithm uses the latitude and longitude of locations to group them in clusters according to the input parameters *minPts* and *eps*. We set $minPts = 2$ and $eps = 5, 10, 50, 100$km in our experiments and eliminate the noise clusters produced by the algorithm, i.e. locations that do not belong to any dense cluster of locations according to the input parameters (see Figure 1).

We compute the distribution of the number of obtained (non-noise) clusters per individual, at different values of *eps* parameter (see Figure 2). We observe a peaked distribution where the majority of individuals have few mobility cores, e.g. two mobility cores when $eps = 5$km and one mobility core when $eps = 100$km, and individuals having more than ten mobility cores are extremely rare (Figure 2). The fact that the algorithm produces non-noise clusters indicates that that the locations of an individual are not randomly distributed but tend to aggregated in dense groups of locations, representing geographical units of individual mobility. Our distribution of cores per person is in contrast with previous works which build mobility groups using network science techniques [1], where most users possess 5-20 mobility groups and only ≈7% of users have a single mobility group.

We also compare an individual's radius of gyration r_g with her inter-core radius r_g^{inter}, observing a strong linear correlation (see Figure 3). Since the inter-core radius is computed on the dominant locations of the individual's mobility cores, this result suggests that the radius of gyration is mainly determined by the tendency of an individual to partition her mobility in different geographical units. If we compute the distribution of individuals' intra-core radius r_g^{intra}, indeed, we do not obtain a power law anymore (Figure 4): a peak emerges from the distribution of r_g^{intra} for low *eps* suggesting that, when restricted to move within mobility cores, individuals show typical radii of gyration. In summary, our analysis suggests that: (i) individuals tend to split their mobility in dense groups of locations (mobility cores); (ii) the distance between the dominant locations in mobility cores generates the observed heterogeneity in human mobility ranges; (iii) the heterogeneity is indeed greatly reduced when individuals are constrained to move *within* mobility cores.

Interestingly, we observe that similar results emerge from both the mobile phone dataset, which captures displacements by any transportation means in an entire European country during three months, and the GPS dataset, which only captures movements by private vehicles occurred in Tuscany during one month.

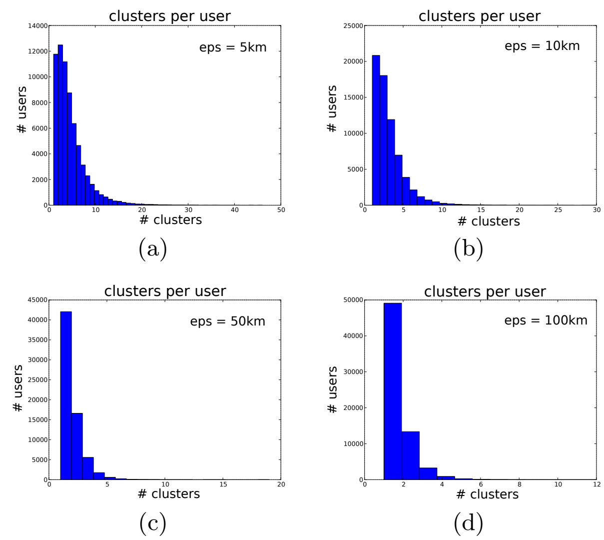


Figure 2: Distribution of the number of clusters per individual on the GSM dataset for $eps = 5, 10, 50, 100$km (the GPS dataset produces similar results). The plots highlight a clear tendency of locations to cluster in dense groups. We observe that: (i) the majority of individuals have few mobility cores (2 or 3), (ii) as *eps* increases the mode of the distribution approaches to one.

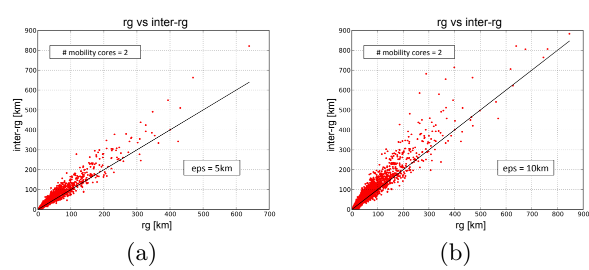
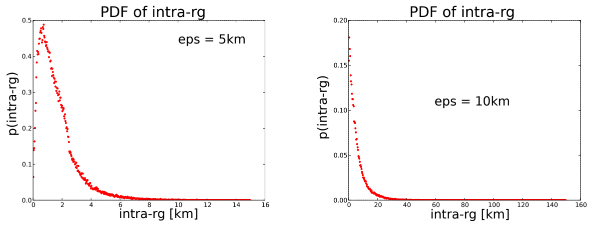


Figure 3: Radius of gyration (on x axis) versus inter-core radius (y axis) of individuals having two mobility cores, for $eps = 5$km (a) and $eps = 10$km (b). Plots refer to the GSM dataset (the GPS dataset produces similar results).



(a) (b)

Figure 4: Distribution of intra-core radius r_g^{intra} across individuals in the GSM dataset (the GPS dataset produces similar results), for $eps = 5$km (a) and $eps = 50$km (b). We observe that, for $eps = 5$km, the distribution is not a power law anymore but a peak emerges denoting a characteristic radius of gyration (a). For $eps = 50$km the distribution starts approaching a power law.

6. CONCLUSIONS

In this paper we showed that the locations visited by individuals tend to cluster in a small number of mobility cores. The radius of gyration computed on the dominant locations of each mobility cores highly correlates with the standard radius of gyration, meaning that the characteristic distance traveled by individuals is mainly determined by their dominant locations. Moreover, individuals show homogenous radii of gyration when constrained to travel within mobility cores. Our results showed that individual human mobility is composed by two types of trips: intra-core trips, which represent movement within a given geographical unit, and inter-core trips, which define trips between locations belonging to different mobility cores and generate the heterogeneity observed in human mobility ranges. As future work, we plan to investigate deeply the structure of intra- and inter-trips and quantify the contribution of every single intra- or inter-trip in shaping the characteristic traveled distance of an individual.

7. ACKNOWLEDGMENTS

This work has been partially funded by the EU under the FP7-ICT Program by project Petra n. 609042, under H2020 Program by projects SoBigData grant n. 654024 and Cimplex grant n. 641191.

8. REFERENCES

[1] J. Bagrow and Y.-R. Lin. Mesoscopic structure and social aspects of human mobility. *PLoS ONE*, 7(5), 2012.

[2] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[3] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'11, pages 1082–1090. ACM, 2011.

[4] N. Eagle and A. Pentland. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63:1057–1066, 2009.

[5] G. M. V. et al. Lévy flight search patterns of wandering albatrosses. *Nature*, 381:413–415, 1996.

[6] G. R.-F. et al. Lévy walk patterns in the foraging movements of spider monkeys. *Behavioral Ecology and Sociobiology*, 55(25), 2003.

[7] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.

[8] S. Jiang, J. F. Jr, and M. González. Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, 25:478–510, 2012.

[9] W. S. Jung, F. Wang, and H. E. Stanley. Gravity model in the korean highway. *EPL (Europhysics Letters)*, 81:48005, 2008.

[10] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella. Human mobility models for opportunistic networks. *IEEE Communications Magazine*, 49(12):157–165, 2011.

[11] L. Liao, D. J. Patterson, D. Fox, and H. Kautz. Learning and inferring transportation routines. *Artif. Intell.*, 171(5-6):311–331, Apr. 2007.

[12] L. Pappalardo, S. Rinzivillo, Z. Qu, D. Pedreschi, and F. Giannotti. Understanding the patterns of car travel. *The European Physical Journal Special Topics*, 215(1):61–73, 2013.

[13] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, and F. Giannotti. Comparing general mobility and mobility by car. In *Proceedings of the 1st BRICS Countries Congress (BRICS-CCI) and 11th Brazilian Congress (CBIC) on Computational Intelligence*, 2013.

[14] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, and A.-L. Barabasi. Returners and explorers dichotomy in human mobility. *Nature Communications*, 6, 09 2015.

[15] S. Rinzivillo, L. Gabrielli, M. Nanni, L. Pappalardo, D. Pedreschi, and F. Giannotti. The purpose of motion: Learning activities from individual mobility networks. In *Proceedings of International Conference on Data Science and Advanced Analytics*, DSAA'14, 2014.

[16] S. Rinzivillo, S. Mainardi, F. Pezzoni, M. Coscia, D. Pedreschi, and F. Giannotti. Discovering the geographical borders of human mobility. *Künstliche Intelligenz*, 26(3):253–260, 2012.

[17] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.

[18] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327:1018–1021, 2010.

[19] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison Wesley, 2006.

[20] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabási. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 1100–1108, New York, NY, USA, 2011. ACM.

[21] G. K. Zipf. The p1p2/d hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677–686, 1946.