D4.2 Typology of urban mobility patterns

Project acronym: EUNOIA
Project title: Evolutive User-centric Networks for Intraurban Accessibility
Grant Agreement number: 318367
Funding Scheme: Collaborative project
Project start date / Duration: 01 Oct 2012 / 26 months
Call Topic: FP7-ICT-2011-8
Project web-site: http://www.eunoia-project.eu/
Deliverable: D4.2 Typology of mobility patterns
Edition 1
Date 30/11/2014
Status Final
Dissemination Level: Public
Authoring and Approval

<table>
<thead>
<tr>
<th>Prepared by</th>
<th>Thomas Louail (CEA-IPhT - Researcher)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maxime Lenormand (IFISC – Researcher)</td>
</tr>
<tr>
<td>Reviewed by</td>
<td>Marc Barthélemmy (CEA-IPhT - Principal Researcher)</td>
</tr>
<tr>
<td></td>
<td>José Ramasco (Deputy Scientific Coordinator)</td>
</tr>
<tr>
<td></td>
<td>Ricardo Herranz (Deputy Management Coordinator)</td>
</tr>
<tr>
<td>Approved by</td>
<td>Maxi San Miguel (Project Coordinator)</td>
</tr>
</tbody>
</table>

Record of Revisions

<table>
<thead>
<tr>
<th>Edition</th>
<th>Date</th>
<th>Description/Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edition 1 Draft 1</td>
<td>24/11/2014</td>
<td>Initial version</td>
</tr>
<tr>
<td>Edition 1 Draft 2</td>
<td>28/11/2014</td>
<td>Second version</td>
</tr>
<tr>
<td>Edition 1</td>
<td>30/11/2014</td>
<td>Minor editorial corrections</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Approved for submission to the EC</td>
</tr>
</tbody>
</table>
# Table of Contents

1. **INTRODUCTION** .................................................................................................................................................. 1

2. **CROSSCHECKING DIFFERENT SOURCES OF MOBILITY INFORMATION** ................................................................. 3

3. **EXTRACTING A TYPOLOGY OF MOBILITY PATTERNS FROM LARGE INDIVIDUAL MOBILITY NETWORKS** .................. 6

4. **COMPARISON OF THE SPATIAL STRUCTURE OF COMMUTING IN SPANISH CITIES** ......................................................... 8
   4.1. **Scaling of the numbers of residential hotspots and employment hotspots with the population size of cities** ............... 8
   4.2. **Proportions of different types of commuting flows in Spanish cities** ..................................................................... 9
   4.3. **Spatial properties of the different types of commuting flows** .................................................................................. 10
   4.4. **Classification of cities based on their commuting structure** .................................................................................. 12

5. **CASE STUDY: EVOLUTION OF THE SPATIAL STRUCTURE OF MOBILITY IN LONDON DURING A WEEK** ................... 13

6. **CASE STUDY: INFLUENCE OF SOCIAL CHARACTERISTICS ON THE WAY PEOPLE MOVE AND SPEND THEIR MONEY IN MADRID AND BARCELONA** ................................................................. 16

7. **SUMMARY AND PERSPECTIVES** ............................................................................................................................ 20

**ANNEX I. REFERENCES** ............................................................................................................................................... 22

**ANNEX II. ABBREVIATIONS AND ACRONYMS** ........................................................................................................... 23
1. Introduction

EUNOIA Work Package 4 'Data Analysis' has been dedicated to the analysis of large sets of individual human mobility data in several European cities. The objective was to make good use of the masses of individual data generated by ICT devices — such as mobile phone records, tweets, credit card data, public transport cards data — but also from more traditional data resulting from mobility and transport surveys, provided by Local Authorities and Census Offices. As introduced in our position paper (EUNOIA, 2012), the data analysis pursued several goals, among them:

- Characterising mobility patterns in different types of cities, i.e. with different geographical, structural, political, institutional, socio-economic, and cultural characteristics;
- Uncovering the relations between the spatial organisation of activity centres and the structure of mobility in the city.

In this document we focus on mobility patterns and on the structure of mobility in cities, to propose typologies of trips, of mobility centres (hotspots) and a classification of cities according to their mobility structure. The most important and regular type/motivation of mobility inside urban areas is commuting (journey to work). To study the organisation of mobility at the city scale we then focused on commuting flows in cities. Commuting flows come as OD (Origin-Destination) matrices, a classic tool of urban transport planning. The OD matrix describes a network, directed and weighted, and in the general case this mobility network is time-dependent. These matrices are usually extremely difficult and costly to measure, and it is only recently, thanks to technological advances such as the GPS or the democratisation of ICT devices, that we can reconstruct OD matrices from the geolocated data produced by the individuals’ devices. Since the mid-2000's and the first research articles in the field, individual geolocated data generated by ICT devices, mobile phone in the first place, have been mainly used to study the statistical and spatial properties of individual human mobility. So far studies have generally relied upon a unique source of ICT data, in most cases mobile phone data. In contrast in EUNOIA WP4 we have tried to exploit several sources of data simultaneously available for some cities, to inspect urban dynamics, and to cross-check the information provided by various sources of data.

A first important analysis is then to ensure that different OD matrices constructed from different mobility data sources provide indeed the same information on the structure of mobility in the city, whether they are the result of surveys conducted by Local Authorities and Census Offices, or the result of computational methods filtering huge databases of passive data produced by individuals’ mobile phones, tweets, or transport cards (section 2).

Secondly, the OD matrix contains the complete information on the commuting flows in a city, and for that reason it fails to immediately provide clear, synthetic and useful information of the structure of mobility in the city. One of our goals was then to propose methods to extract synthetic and stylised facts that would help to better understand the collective mobility patterns observed at the city scale (section 3). We then applied the developed methods to thirty cities in Spain (section 3), and then in London (section 4), in order to compare and classify these cities according to their commuting structure. During the project’s second year, we also had access to an additional source of individual data, namely credit cards data. In addition to the time and location associated to each transaction of a given customer, these data also inform us on some demographic and social
characteristics of the card owner. With a statistical approach, they thus allow to assess the influence of demographic and social characteristics on mobility in the city, if any (section 6).
2. Crosschecking different sources of mobility information

The pervasive use of new mobile devices has allowed a better characterisation in space and time of human concentrations, and of individual mobility in general. While classical data sources, such as surveys or census, have a limited level of geographical resolution (e.g., districts, municipalities, counties are typically used) or are restricted to generic workdays or weekends, the data coming from mobile devices can be precisely located both in time and space. Most previous works in this field have used a single data source to study human mobility patterns. In this EUNOIA WP4 study, we have performed instead a cross-check analysis by comparing the spatial organisation of journeys extracted from data collected from three different sources: Twitter, census and cell phones. This work has been published in *PlosOne* in July 2014 (Lenormand et al., 2014).

Our analysis has focused on the urban areas of Barcelona and Madrid, for which data of the three types were available. We have assessed the correlation between the datasets on different aspects: the spatial distribution of people concentration, the temporal evolution of people density, and finally the commuting (journey to work) patterns of individuals.

The first aspect considered refers to the population concentration in different parts of the cities. This point is of great importance in the analysis and planning of urban environments, including the design of new services or of contingency plans in case of disasters. Our results show that both Twitter and cell phone data produce similar density patterns both in space and time, with a Pearson correlation close to 0.9 in the two cities analysed.

![Figure 1. Correlation between the spatial distribution of Twitter users and mobile phone users for the weekdays (aggregation from Monday to Thursday) and from noon to 1pm for the metropolitan area of Barcelona (cells of 4km²).](image)

(a) Scatter-plot composed by each pair (Number of Twitter users, Number of mobile phone users), the values have been normalized (dividing by the total number of users) in order to obtain values between 0 and 1. The red line represents the perfect linear fit with slope equal to 1 and intercept equal to 0. ((b)-(c)) Spatial distribution of Twitter users (b) and mobile phone users (c). In order to facilitate the comparison of both distributions on the map, the proportion of users in each cell is shown (always bounded in the interval [0,1]).

The second aspect considered has been the temporal distribution of individuals, which allow us to determine the type of activity that is most common in specific urban areas. We show that similar temporal distribution patterns can be extracted from both Twitter and cell phone datasets. Indeed, for both the Twitter data set and the mobile phone data we have identified three temporal distribution patterns associated with three type of land use: Residential, Business and Nightlife (Figure 2).
Figure 2. Temporal distribution patterns for the metropolitan area of Barcelona (cells of 4km²). (a), (c) and (e) Mobile phone activity; (b), (d) and (f) Twitter activity; (a) and (b) Business cluster; (c) and (d) Residential/leisure cluster; (e) and (f) Nightlife cluster. We have divided the week into four groups: from the left to the right: the weekdays (aggregation from Monday to Thursday), Friday, Saturday and Sunday).

The last question studied has been the extraction of mobility networks in the shape of Origin-Destination commuting matrices. We shown that at high spatial resolution, in grid cells with sides of 1 or 2 km, the networks obtained with cell phones and Twitter are comparable. Of course, the integration time needed for Twitter is higher in order to obtain similar results. Twitter data can run in serious problems too if instead of recurrent mobility the focus is on short-term mobility, but this point falls beyond the scope of this work. Finally, the comparison with census data is also acceptable: both Twitter and cell phone data reproduce the
commuting networks at the municipal scale from an overall perspective (Figure 3). Still and although good on average, the agreement between the three different datasets is broken in some particular connections that deviate from the diagonal in our scatterplots. This can be explained by the fact that the datasets come from different sources, were collected in different years and may have different biases and level of representativeness. For example, Twitter is supposed to be used more by younger people. The explanation of these deviations and whether they are just stochastic fluctuations or follow some rationale could be an interesting avenue for further research.

![Figure 3](image)

Figure 3. Comparison between the non-zero flows obtained with the three datasets for the Barcelona’s case study (the values have been normalized by the total number of commuters for both OD tables). Blue points are scatter plot for each pair of municipalities. The red line represents the $x=y$ line. (a) Twitter and mobile phone. (b) Census and mobile phone. (c) Census and Twitter.

To conclude, our results show that the three data sources are providing indeed comparable information. Even though the representativeness of Twitter geolocated data is lower than that of mobile phone and census data, the correlations between the population density profiles and mobility patterns detected by the three datasets are close to one in a grid with cells of 4km$^2$ and 1km$^2$. This level of correlation supports the feasibility of interchanging the three data sources - at the spatio-temporal scales considered – when one wishes to analyse the structure of mobility at the city scale. These results set a basis for the reliability of previous works basing their analysis on single ICT datasets only. And similarly, the door to extract conclusions from data coming from a single data source (due to convenience of facility of access) is open as long as the spatio-temporal scales tested here are respected. Hence these results have been an important basis for other studies done in WP4 that has focused on the analysis of the spatial structure of commuting by using OD matrices extracted from mobile phone data (see the corresponding section in D4.2 ‘Typology of mobility patterns’).
3. Extracting a typology of mobility patterns from large individual mobility networks

In the introduction we have mentioned that the increasing availability of urban data opens exciting possibilities of renewed quantitative approaches to cities and to many urban dynamics. Additionally, in the previous section we have shown that at the city scale, the commuting trips revealed by these new sources of data are in accordance with those revealed by travel surveys. Nevertheless, while exciting, these enormous amounts of data also bring some new problems to solve. A central one is the extraction of useful and synthetic information from these tremendous datasets. We are interested in extracting coarse-grained information and stylized facts that encode the essence of the human, urban, phenomenon under study - in our case individual mobility in cities - and that any reasonable model or urban mobility should reproduce. Indeed, such meso-scale information helps us understand the system, to compare different systems, and also to propose models that could explain the process under study.

An OD matrix encapsulates the complete information on individuals’ flows in a city, at a given spatial scale and for a specific purpose, generally commuting (journey to work). It is a large, weighted and directed network, and as such it does not provide a clear, synthetic and useful information on the structure of the mobility in the city. More generally, it is usually very difficult to extract high-level information from large networks and state-of-the-art methods such as community detection\(^1\) and stochastic block modelling fail to give us an expressive image of the organisation of mobility at the city scale. This is particularly true in the case of commuting networks in cities, where edges represent flows of individuals that travel daily from their residential area to their main activity area. Several types of trips can be distinguished in these commuting networks. Some flows constitute the backbone of the mobility in the city by connecting major residential neighbourhoods to employment centres, while other flows converge from smaller residential areas to important employment centres, or diverge from major residential neighbourhoods to smaller activity areas. In addition, the spatial properties of these commuting flows are fundamental in cities and a relevant method should be able to take this aspect into account.

In (Louail et al., 2014b) we have proposed a simple and versatile method that we designed to compare the structure of large, weighted and directed mobility networks. The guiding idea is that a simple and clear picture of the mobility structure in the city is provided by the distribution of flows between different types of nodes. The method extracts a very simple footprint of the commuting structure, under the form of a 2*2 matrix that separates the flows in four categories, and then quantifies the proportion of each of those four types of commuting flows in the city:

- The **integrated** flows (I) of individuals that live in major residential neighbourhoods (**residential hotspots**) and that work in the city’s main employment areas (**employment hotspots**);
- The **divergent** flows (D) of individuals that live in **residential hotspots** and do not work in employment hotspots (but in smaller activity areas);
- The **convergent** flows (C) of individuals that do not live in **residential hotspots** but work in employment hotspots;
- Finally the more **random** flows (R) of individuals who neither live nor work in hotspots.

\(^1\) See for example the review of S. Fortunato (Fortunato, 2010, Physics Reports).
For each of the four categories we can sum the number of commuters concerned and divide it by the total number of commuters in the OD matrix, which gives the proportion of individuals in each of the four categories of flows. In other words, for a given city, we reduce the original OD matrix to a 2*2 matrix, where each of the four terms I (integrated), C (convergent), D (divergent) and R (random) is comprised between 0 and 1, and \( I+C+D+R=1 \).

**Figure 4.** The ICDR method decomposes the commuting flows in the city in four categories: the Integrated flows (I) from hotspot to hotspot, the Convergent flows (C) to hotspots, the divergent flows (D) originating at hotspots and finally the random flows (R), which are neither starting nor ending at hotspots. For each city with its origin-destination matrix, we can compute the importance of each commuting flow category and get a simple picture of the mobility structure in the city.
4. Comparison of the spatial structure of commuting in Spanish cities

We then applied the ICDR method on Origin-Destination (OD) matrices extracted from mobile phone data that were available for a five weeks period in thirty-one Spanish urban areas. (NB: we detail our method to extract OD matrices from individual mobile phone records in Annex 3 'Extraction of OD matrices from individual mobile phone data').

4.1. Scaling of the numbers of residential hotspots and employment hotspots with the population size of cities

The first step of the ICDR method consists in determining residential and employment hotspots. To that end we used the ‘LouBar’ method that we developed during the project’s first year, and that is documented in EUNOIA document D4.1, and also in (Louail et al., 2014). A first interesting measure is to count the number of residential hotspots and employment hotspots in cities, and see how these numbers scale with the population size of cities. As one can see on Figure 5, these numbers for residential (blue dots, left panel) and employment hotspots (red dots, right panel) scale sublinearly with the population size of the city. The number of work hotspots grows significantly slower than the number of residential hotspots, showing that in Spanish urban areas residential areas are statistically more decentralised, when compared to activity centres -- that we intuitively expect to be less numerous.

![Graph A](image1.png)

**Figure 5. Numbers of hotspots versus population size for the 31 Spanish cities studied.** (a) Number of residential hotspots vs. population size of the city (b) Number of work hotspots vs. population size of the city. Both scaling relations are sublinear, indicating an "economy of scale" in the spatial organisation of cities. The exponent is remarkably smaller in the case of the work/main daily activity hotspots, which means that in Spanish cities the number of important places that concentrate many jobs and daily amenities, grows slower than the number of important residential places.
4.2. Proportions of different types of commuting flows in Spanish cities

We calculated the ICDR values for each city's commuting OD matrix. For the 31 Spanish urban areas under study we obtain the ICDR values shown in Figure 6. In Figure 6(a), the ICDR values are plotted versus the population size of the cities. We see that globally the proportion $I$ of individuals that commute from residential hotspots to employment hotspots decreases as the population size $P$ increases, while the proportion $R$ of 'random' flows increases and the proportions $C$ and $D$ of convergent and divergent flows seem surprisingly to stay constant whatever the population size of the city. In Figure 6(b) we plot the same ICDR values, but this time sorted by decreasing values of $I$. We see more clearly that the $I$ and $R$ values alone are the relevant parameters to distinguish one city from another.

The decay of the proportion of integrated flows in favour of more 'random' flows when the population size increases traduces that an effect of growth among Spanish cities is the decentralization of both activity places and residences. As cities get bigger, their numbers of residential and employment hotspots grow (sublinearly), but these hotspots catch a smaller part of the commuting flows.

![Image of Figure 6](image_url)

**Figure 6. ICDR values of the 31 Spanish cities under study.** (a) $I$ (integrated), $C$ (convergent), $D$ (divergent) and $R$ (random) values versus population size. Globally $I$ decreases while $R$ increases when the population size increases (b) Same ICDR values but this time ranked by decreasing order of $I$. By definition, we have for each city $I+C+D+R=1$. It is remarkable that $I$ and $R$ dominate and seem almost sufficient to distinguish cities, while $C$ and $D$ are almost constant whatever the city size.

In order to evaluate to what extent the ICDR signatures of cities are characteristic of their commuting structure, we compared these values to the ones returned by a null model of random commuting flows. The principle is that for each city, we generate random OD matrices of the same size than the original, empirical OD matrix of the city, with random flows of individuals, but still that preserve the number of residents and the number of workers in each part of the city (one could read (Louail et al, 2014 b) for details). The comparison of the empirical values with those obtained with a null model demonstrate that for Spanish cities, the larger the city and the further its organization from a random one. This result is somewhat counter-intuitive, since we expect that the larger the city, the more ‘disordered’ is the mobility of individuals. This is not the case, and in the largest cities there exists a commuting backbone that in fact couldn’t result from purely random movements of individuals. To the contrary it is a footprint of the city's structure and of its history.
4.3. Spatial properties of the different types of commuting flows

Another important aspect of the citizens commuting flows is the average distance of each of the four different types of flows exhibited by the method, and also the relation between the city size and the distances travelled by individuals that live in that city. For each city we have calculated the average distance travelled by individuals per type of commuting flows (Integrated, Convergent, Divergent and Random). The resulting average distances are plotted in red on Figure 7. On this figure we can see that for all categories of flows the average distance travelled by individuals increases with population size, which is an expected effect as the area of Spanish cities also grows when the population size increases. We also observe that the average distance of convergent flows (C) increases faster than the average distance of the three other types of flows (we note that the average distance associated to convergent flows increases more than the distance associated to divergent flows (D), showing that these flows are not ‘symmetric’ as one could naively expect). This result means that for this set of Spanish cities, individuals that commute from small residential areas to important activity centres travel on average a longer distance than all other residents. This observation could be an indication that for our set of cities, residential areas have spatially expanded in the last decades while activity centres remained at their location, leading to longer commuting distances for residents of these Spanish urban areas.

Interesting information is also provided by the comparison of distances measured in the data with average distances measured in random OD matrices generated by a null model of random commuting flows. The average distances associated to the null model are plotted in blue in Figure 7(a). We see that for all types of flows the distances measured in the data are in fact shorter than those generated by the null model. This is another clear indication of the spatial organization of individual flows in cities. It also highlights the importance of the travel time budget in the choice of a residential location. Remarkably enough, the distances of convergent flows (C) are both the largest and the ones that increase the fastest with population size, indicating a low degree of efficiency on this aspect in the biggest cities.
Figure 7. Spatial structure of commuting in Spanish cities revealed by the application of the ICDR method. (a) Average commuting distance vs. population size, per type of commuting flows. (b) ratio \( \frac{D(\text{Null model})}{D(\text{data})} \), per type of flows. The ‘LouBar’ criteria is used to define residential and employment hotspots.

On Figure 7(b) we have plotted the ratio between the average distance returned by the reference null model, and the corresponding distance measured in data, for the four types of flows. Values lower than 1 indicate that the average commuting distance generated by the null model is indeed shorter than the distance observed in the city. Surprisingly, we observe that some small cities display a value less than one, which indicates the lesser importance of space, and the lesser importance of rational/reasonable spatial organisation at shorter spatial scales. We also see that this ratio increases faster for random flows (R) than for the three other types of flows (I, C, D), suggesting a remarkable spatial structure of these random flows.
4.4. Classification of cities based on their commuting structure

Finally the ICDR signature of the cities’ OD matrices allows clustering cities with respect to the structure of their commuting patterns. We have measured the Euclidian distance between the cities' ICDR signatures and we have then performed a hierarchical cluster analysis. Figure 8 shows the dendrogram resulting from the classification. Four well-separated clusters of cities are identified on this dendrogram. Remarkably the classification shows that largest cities are grouped together and are characterized by a larger proportion of ‘random’ flows (R) of individuals both living and working in parts of the city that are neither dominant residential and activity centres. This increasing proportion of random flows can be interpreted as an increased facility, in bigger urban areas, to commute from any part of the city to any other part (Louail et al., 2014 b).

Figure 8. Dendrogram resulting from the hierarchical clustering on cities based on their ICDR matrix. In front of each city name we indicate its rank in the hierarchy of population sizes. The biggest cities are clustered together. As cities get bigger, the ‘random’ component of their commuting flows increases, which traduces the fact that it is easier to commute from any place to any other in big cities.
5. Case study: evolution of the spatial structure of mobility in London during a week

In the context of the EUNOIA WP5, we have developed an activity-based travel demand model for London using MATSim. One objective was to make use of unconventional datasets to feed this simulation model, and also make use of these data for evaluation/validation of the outputs of the model. To that purpose we have analysed records from smart transport cards (Oyster cards) made available by Transport for London (TfL). We started with a sample of data for an entire week in July 2012. These data contain ~ 23.000.000 tap-in and tap-out events of individuals in 597 stations of the London rail network. For each 15-minutes time slice we have an OD matrix, meaning that the data inform us on the number of individuals that have travel from one station to another, and this for each couple of the 597 train stations.

The goal of this part of our work is twofold: (i) apply the ICDR method presented in the previous section 4 to highlight at an aggregated level, the evolution of the structure of mobility in a large city during an entire week; and (ii) extract stylized facts from these transport data that would be useful to calibrate the MATSim activity-based transport model developed in WP5. The large volumes of recorded tap-in and tap-out data cannot be used as such to feed or calibrate the simulation model at an individual scale. We first need to extract coarse-grained information that captures some key features of the mobility in London. Indeed a crucial feature of our MATSim model will be its capacity to reproduce some high-level stylised facts of the behaviour of the complex system under study.

This has been the guiding idea of this case study. We inspected the time dynamics of origin and destination hotspots in the tube, and the trips they polarize. How do these hotspots evolve during a week? How does the macro-structure of flows between hotspots evolve? What are the spatial properties of the dynamical spatial network determined by flow data?

In order to apply our ICDR method and quantify the importance of each of the 4 types of flows for each 15-minutes slice, we first determined origin and destination hotspots. In this case the hotspots are train/tube stations, and the problem is then to determine for each 15-minutes period, the London's train and metro stations whose number of entering/exiting commuters is much larger than for the other stations. There are two types of hotspots to be considered, origin hotspots and destination hotspots, and their number is not necessary the same. To determine the hotspots we used the ‘LouBar’ criteria of the general method we develop in the frame of EUNOIA WP4 (Louail et al. 2014) (see the corresponding annex in document D4.1). Figure 9 represents the number of stations that are origin hotspots and destination hotspots versus time.
Figure 9. Evolution of the numbers of origin and destination hotspots among the 597 stations of the London rail network, during a week. One can see a typical daily pattern during weekdays: the number of origin hotspots is maximal in the morning peak, when individuals commute to their workplace. It then decreases during the day, and reaches a minimum at night time. It is the opposite for destination hotspots, which is at its lowest level in the morning, and increases during the day. One can also notice than we observe the same pattern during the week-end.

We then applied our ICDR method and calculated the four values (see the two previous sections) at each point in time. Figure 10 represents the ICDR values of London vs. time, for the individuals travels made by train. We first notice I and R values change little during the course of the week, and stay in a small interval of values, [0.35-0.4] for R, and [0.15-0.2] for I. While a bigger city than Barcelona and Madrid, it is remarkable to note that the values for London confirm our observations when comparing mobility structure in Spanish cities through ICDR values: the main flows are the ‘random’ flows (R) between stations that are neither origin nor destination hotspots. One can also notice that in the course of the day, the proportion of convergent flows decreases at the benefit of divergent flows. The proportion of convergent flows (from small and medium-sizes origin stations to main destination stations) is maximal in the morning peak (7-9am), and then decreases all along the day, while the dynamics are opposite for divergent flows, whose weight is maximal in the evening peak (5-7pm). It is also remarkable that in the end of the day, we observe an increase of the proportion of integrated flows, from main origin stations to main destination stations, indicating that the journeys in the last hours of the day are more concentrated between a fewer number of stations. It is also remarkable that we observe the same pattern during the weekend.
Figure 10. Evolution of London’s ICDR values during the course of a week.
6. Case study: Influence of social characteristics on the way people move and spend their money in Madrid and Barcelona

Everyday, billions of individuals generate a large volume of geolocated data by using their mobile phone, GPS, public transport card or credit card. This vast amount of data coming from different data sources is opening new opportunities for the research in socio-technical systems. Indeed, geolocated data allow us to identify when and where people interact with ICT tools, for example, to give a phone call or to make a credit card payment, giving us some useful insights on human mobility. In addition, some datasets offer the opportunity to gather information about the content of the interaction or the operation with ICT tools. For example, we know where and when an individual use his or her mobile device to make a phone call but sometimes information about the content is also provided, such as the ID of the callee and the call duration. This information enables researchers to move further on the study of human behavior by analyzing how people interact with each other but also with their environment. Nevertheless, in most studies based on geolocated data a crucial piece of information is always missing: demographics characteristics of individuals using the ICT tools. However, quantitative studies of social networks dynamics have shown that people behave differently according to the gender and age.

In this running study, we analyze a credit card database containing over 40 million of bank card transactions in order to explore consumption habits and mobility patterns of bank customers in the two largest provinces of Spain (Figure 11) according to three demographic characteristics: gender, age and occupation.

![Maps of the transactions](image)

**Figure 11. Maps of the transactions.** The red dots represent the locations of the transactions on a map of the province of Madrid (a) and Barcelona (b). The small areas correspond to postcodes.

A first interesting result of this analysis has been to highlight the importance of demographic characteristics in the way people spend their money. We have plotted in Figure 12 the distribution of money spent according to demographics and business category. First, we observe that the proportion of money spent according to the business category is very different for men and women. Indeed, women spend more money than men in
Fashion, Food/Hypermarkets, Health and Wellness/Beauty whereas men spend more money than women in Automotive Industry, Bar/Restaurants, Technology and Transport. Next, we note that the proportion of money spent in Fashion, Food/Hypermarkets, Sports/Toys, Technology and Transport globally decreases with age. Inversely Automotive Industry, Health, Leisure and Travel Agencies are business categories for which the amount of money spent increases with age. Finally, we remark that differences also exist between people having different occupation. For instance, students spend more money in Bar/Restaurant, Fashion, Sports/Toys, Technology and Transport than others types of occupation.

Figure 12. Proportion of money spent according to customer's demographics and business' category.

To study the properties of customer mobility patterns, two indicators have been considered, the distance travelled by a customer between two consecutive transactions and the radius of gyration as defined in Gonzalez et al. (2008). These two indicators have been used to study the customer mobility patterns according to their demographic characteristics (gender, age and occupation). The results obtained are plotted in Figure 13 and 14. Based on these results, one can remark that, depending on his/her demographic characteristics, a customer can travel short or long distances and stays more or less closer to his/her centre of
mass. Three main differences are observed. First, women travel shorter distance than men and their trajectory stay closer to the centre of mass. Second, the average distance travelled between two consecutive positions and the radius of gyration decrease with age. Finally, an opposition between active and inactive is highlighted. Indeed, unemployed, retired and homemaker seem to travel shorter distances and stay closer to the centre of mass than other customers.

Figure 13. Distribution of the distance travelled by a customer between two consecutive transactions according to customer's demographic characteristics in Barcelona. (a) Probability density function of the distance traveled by a customer between two consecutive transactions. (b) – (d) Probability density function of the distance traveled by a customer between two consecutive transactions according to the gender (b), the age (c) and the occupation (d). The insets show the boxplot versions of the plots, the black point represents the average.
Figure 14. Distribution of the radius of gyration according to customer’s demographic characteristics in Barcelona. (a) Probability density function of the radius of gyration. (b) – (d) Probability density function of radius of gyration according to the gender (b), the age (c) and the occupation (d). The insets show the boxplot versions of the plots, the black point represents the average.
7. Summary and perspectives

We have performed several analyses on various large-scale mobility datasets that have allowed us to shed new light on several aspects of urban mobility. First we have summarised the results of a cross-check of three sources of mobility information available in the two largest Spanish cities: mobile phone records, tweets, and transport surveys, in Barcelona and Madrid. We have shown that these three data sources provide very well correlated information on commuting trips in the city (Lenormand et al., PlosOne, 2014).

We have then presented a new method that we developed to extract from OD matrices some useful and synthetic information on the structure of mobility in a city. The method relies on the identification of residential and employment hotspots, and aggregates commuting flows in four different types (integrated, convergent, divergent, random), depending whether they start and end from/to a hotspot or not. We have applied our method to all Spanish urban areas with more than 200,000 inhabitants. The method has allowed us to highlight several remarkable patterns in the commuting data extracted for these cities:

- On average and for all cities considered here, individuals that live in residential main hubs and that work in employment main hubs (integrated flows) travel shorter distances than the others (convergent, divergent and random flows);
- When the city size increases, the largest impact is on convergent flows of individuals living in smaller residential areas (typically in the suburbs) and commuting to important employment centres;
- Our classification of cities based on their ICDR values leads to groups with consistent population size, highlighting a clear relationship between the population size of cities and their commuting structure.

In addition, the comparison with a null model led us to interesting conclusions: flows in cities display a high level of spatial organisation and as the population size of the city grows, the increase of the Z-scores of all types of flows shows that the structure of the mobility is more and more specific and far from a random organization. We note that an interesting direction for future research would be to find some analytical arguments using simple models of city organisation for estimating these flows and how they vary with population size.

To illustrate the versatility of the proposed method, we applied it on temporal mobility data in London that give us complete information on the flows in the London train network during an entire week. We have shown that the proportion of different types of flows in London vary little during the week, and hence constitute a ‘signature’ of the city.

Our coarse-graining method provides a large-scale picture of individual flows in the city. In this respect it could be particularly useful for validating synthetic results of urban mobility models, and for comparing different models. An accurate modelling of mobility is indeed crucial in a large number of applications, including the important case of epidemic spreading, which needs to be better understood, in particular at the intra-urban level.

It would also be interesting to apply the proposed method to other mobility datasets, with different time and spatial scales, in order to test the robustness of these commuting patterns. Further studies focused on other cities and countries are needed at this stage in order to discuss the relevance of the proposed classification of cities based on their commuting structure, that groups cities according to their population size.
Another direction for future studies could be to inspect the time evolution of the ICDR values for datasets describing travel to work journeys over several decades (such as Census national travel surveys). The method could also be applied at larger spatial scale, for example to capture dominant effects in international migration flows. More generally, we believe that an important feature of the proposed method is its versatility as it could be applied to any type of data that is naturally represented by a weighted and directed network.

Finally we have made use of credit card data that contain sociodemographic information on the owner of the card, to see it such variables impact on the way people move and spend their money in the city space. We investigated the properties of people mobility patterns using three variables: the time elapsed between two consecutive transactions, the distance travelled by an individual between two consecutive transactions and the radius of gyration. Three main differences between groups of people were identified: differences between men and women, young and old people and active and inactive individuals. In the three cases, people of the first group (men, young people and active people) travel shorter distances and their trajectory stays closer to their centre of mass than individuals of the second groups (women, old individual and inactive people).

Among all the differences emphasised in this paper the one between men and women is the most difficult to explain. In all the comparisons we have carefully checked that this difference was not related to other sociodemographic variables and it was not the case. It could be interesting to verify whether this difference is related to other social characteristics such as the number of children for example. Indeed, the fact that the difference in terms of mobility patterns between men and women is less pronounced for old people and students may reflect that women with children move differently than women without children.
Annex I. References

Lenormand et al. (2014) Cross-Checking Different Sources of Mobility Information. PloS ONE 9(8) :e105184.


Annex II. Abbreviations and Acronyms

AUDES: Áreas Urbanas De ESpaña. Harmonized delimitation of Spanish metropolitan areas based upon functional criteria. Aggregation of municipalities is decided on the basis of the proportion of residents that work and commute to the metropolitan area;

BTS (tower): Base Transceiver Station. Fixed antenna that facilitates the linkage of the mobile phones to the mobile communication network. In CDR, the geolocation of a user is in fact the geolocation of the BTS tower to which its mobile phone is currently linked.

CBD: Central Business District.

CDR: Call Detail Record, also known as call data record. Data record produced by a telephone exchange or other telecommunications equipment that contains attributes that are specific to a single instance of a phone call or other communication transaction that was handled by that facility or device.

TfL: Transport For London.

UTM: Universal Transverse Mercator, a standard and universal system of cartographic projection.