

# Deriving Mobility Practices and Patterns from Mobile Phone Data

Fabio Manfredini, Paola Pucci, and Paolo Tagliolato

Dipartimento di Architettura e Studi Urbani, Politecnico di Milano, via Bonardi 3,  
20133 Milano, Italy  
{fabio.manfredini,paola.pucci,paolo.tagliolato}@polimi.it

**Abstract.** The paper addresses the issue of analyzing and mapping mobility practices by using different kinds of mobile phone network data, which provide geo-located information on mobile phone activity at a high spatial and temporal resolution. We will present and discuss major findings and drawbacks, based on an application carried out on the Milan urban region (Lombardy, Northern Italy).

**Keywords:** Mobile phone data, Density of space usage, Mapping urban domains, Milan urban region, mobility.

## 1 Introduction

Interpretative tools for the identification of mobility practices in the contemporary metropolis are needed, not only for the some known limitations of traditional data sources but also because new forms of mobility are emerging, describing new city dynamics and time-variations in the use of urban spaces by temporary populations. In Italy, the traditional data sources for urban and mobility investigations (ie surveys, census) have known limitations, including the high cost of surveys, the difficulty of data updating, the difficulty of describing city dynamics and time dependent variations in intensity of urban spaces usages by temporary populations at different scales.

New forms of mobility, close to daily mobility, are changing the way in which urban spaces are used. They are characterized both by being based on the use of transportation system, and by the efficient appropriation of information technologies (internet, mobile phone). They intensified the density of the moves with which we can read diversified uses of the city, that traditional sources of analysis are unable to return with continuity. In Lombardy region, the systematic mobility represents only the 29% of daily travels, which are attested on 2,65 travel/day in average, with a propensity to mobility that changes in relationship to the professional condition.

As underlined by some authors [1], [2], [3], [4], [5], [6], the changes in the management of mobility in contemporary cities are a useful key for understanding the transformations of times, places and modes of social life and work programs, structuring the metropolitan areas.

In this perspective, mobility may represent both a tool of knowledge and a project for urban planners, provided that a better understanding of different patterns of mobility in the form of "active biographies", which increase the range of "post-fordist living and labor styles" [7], is available.

Considering the role of mobility practices in social and spatial differentiation, it becomes important to formulate pertinent analytical approaches, aimed at describing the different densities of use of the city as a new challenge and a prerequisite for understanding the city and its dynamics.

Hence, from an analytical point of view, it becomes important to accompany the traditional quantitative approaches referred to a geographic displacement that tends to focus on movement in space and time, in an aggregate way and for limited periods, with data sources able to describe fine grain over-time variation in urban movements.

In this direction, an interesting contribution may come from mobile phone network data as a potential tool for the development of real-time monitoring, useful to describe urban dynamics, as it has been tested in several experimental studies [8], [9], [10].

The application researches were focused on two different products. Some studies dealt with aspects of representation of the data, emphasizing the most directly evocative aspects, to highlight how these data may represent the "Mobile landscapes" [9]. Other studies focused both on data-mining analysis and on the construction of instruments capable of deriving summary information and relevant data about the urban dynamics from cell-phone [8].

As opposed to the more traditional methods of urban surveys, the use of aggregated and anonymous mobile phone network data has shown promise for large-scale surveys with notably smaller efforts and costs [11].

If we consider the observed and aggregated telephone traffic as the result of individual behaviors and habits, we can treat mobile phone data as a functional source on the *real* use of the cities, capturing, for example, traces of temporary populations, which are difficult to intercept by traditional data source, but which, at the same time, increasingly affect urban practices both quantitatively and qualitatively.

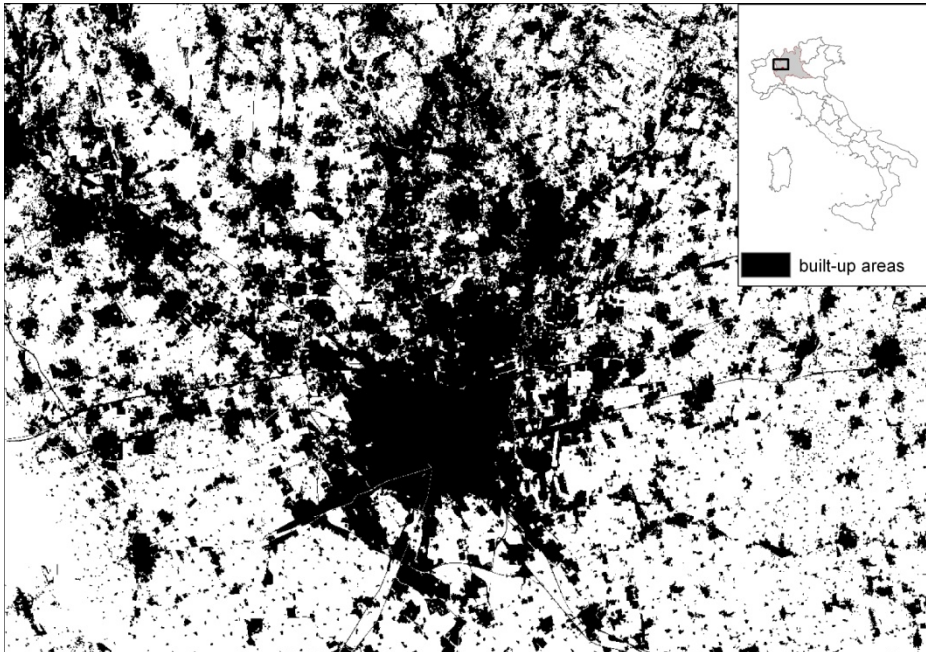
In this direction an increasing number of studies concerns the exploitation of mobile phone data in urban analysis and planning [12]. In particular an interesting issue regards the classification of urban spaces according to their users' practices and behaviors [11], [13]. In [14] the authors outline the fact that city areas are generally not characterized by just one specific use, and for this reason they introduce the use of c-means, a fuzzy unsupervised clustering technique for land use classification, which returns for each area a certain grade of membership to each class.

In this general context, we used mobile phone data provided by Telecom Italia, the main Italian operator, in order to test the potentialities of this information in describing the density of use of urban spaces at different temporal and spatial scales as a precondition:

- to identify temporary populations and different forms of mobility that structure the relationships in the contemporary city;
- to propose diversified management policies and mobility services that city users require, increasing the efficiency of the supply of public services.

## 2 Mobile Phone Data

Milan is placed in an urban region which goes far beyond its administrative boundaries (see fig.1). The core city and the whole urban area have been affected in the last 20 years by changes in their spatial structures and have generated new relationships between the centre and suburbs. At the moment, the urban region of Milan is a densely populated, integrated area where 4.000.000 inhabitants live, where there are 370.000 firms and large flows of people moving daily in this wide area [15].



**Fig. 1.** Map of built-up areas in the Milan urban region (2007). Source: DiAP elaboration of DUSAF 2.1 data.

In order to analyze the complex temporal and spatial patterns emerging from mobile phone data, we used two different types of data provided by Telecom Italia.

### 2.1 Telephone Traffic Intensity – Erlang

The first data type concerns the mobile phone traffic registered by the network over the whole Milan urban region (Northern Italy).

Data are expressed in Erlang, namely the average number of concurrent contacts in a time unit. In the present case, the data represent the telephone traffic intensity every 15 minutes and was supplied by Telecom Italia in a spatialized form. From the telephone traffic recorded by each cell of the network, the telephone provider

distributed the measurements, by means of weighted interpolations, throughout a tessellation of the territory in 250 meters x 250 meters squared areas (pixels)<sup>1</sup>.

We performed time series analysis on this data along a period of 14 days in March 2009 (March 18<sup>th</sup> till March 31<sup>st</sup>), in order to evaluate specific characteristics of population behaviors at an hourly and daily base. We then applied a novel geo-statistical unsupervised learning technique aimed at identifying useful information on hidden patterns of mobile phone use. We will show that these hidden patterns regard different usages of the city in time and in space and that they are related to the mobility of individuals. The results return new maps of the region, each describing the intensity of one of the identified mobility pattern on the territory. This highlights, in our opinion, the potentials of this data for urban planning and transport research studies.

## 2.2 Aggregated Tracks of Users – Hourly Origin Destination Matrices

The second typology of data consists in localized and aggregated tracks of anonymized mobile phone users. It is an origin-destination datum derived from the Call Detail Record database. Italian privacy policies severely constrain the use of these data, even for research purposes. In the framework of a collaboration with Telecom Italia (T-Lab), we arrived at the definition of a datum which was free from privacy constraints, consisting in an aggregation of users' displacements based on CDR records. Telecom engineers set up a system for the automatic and blind extraction of data of this kind.

The system is fed with the CDR and a tessellation of a certain geographical region (in the present case the Lombardy Region). The output consists in time series of Origin-Destination matrices (where origin and destination zones are the tessellation's tiles), equivalent to a function  $F(o,d,t) \rightarrow n$  which, at time  $t$  ( $t$  varying in within the 24 hours of a given day), assigns to origin  $o$  and destination  $d$  the number  $n$  of distinct users that performed some mobile phone activity<sup>2</sup> within  $o$  at time  $t-1$  and a subsequent activity within  $d$  at time  $t$ .

The CDR's raw informations are available at the level of the antennas which handled the activity. The distribution of antennas in space depends on the amount of mobile phone traffic that needs to be managed. In dense urban areas we therefore observe a high density of antennas while in the suburbs the density of antennas may be very low. For the positioning of a user within a certain tile of a tessellation to be reliable, a technical constraint imposed that the tiles contained at least 13 antennas.

We defined two distinct tessellations which could gave us the possibility to map and to interpret main spatial patterns of mobile phone users' mobility: the first is more

---

<sup>1</sup> More formally we can define the Erlang  $E_{xj}$  relevant to the pixel  $x$  and to the  $j$ -th quarter of an hour as:

$$E_{xj} = \frac{1}{15} \int_{15(j-1)}^{15j} N_x(t) dt$$

where  $N_x(x)$  is the number of mobile phone using the network within pixel  $x$  at time  $t$ , hence  $E_{xj}$  is the temporal mean over the  $j$ th quarter of an hour of the number of mobile phones using the network within pixel  $x$ .

<sup>2</sup> With mobile phone activity we intend each interaction of the device with the mobile phone network (i.e. calls received or made, SMSs sent or received, internet connections, etc..).

fine-grained, and it was obtained by a data-driven process taking into account the spatial distribution of antennas; the second, even if more coarse, was more directly related to administrative boundaries, consisting in the aggregation of adjacent municipalities' polygons.

We obtained the two tessellations as follows:

- Automatic clustering of antennas (526 polygons). Each zone is an aggregation of Voronoi cells obtained from the points of location of the antennas. More precisely we proceeded by clustering the positions of the antennas by means of an agglomerative hierarchical clustering algorithm (complete linkage, euclidean distance): we firstly cut the hierarchical tree in order to obtain 100 clusters; we then selected the groups with more than 500 antennas and we cut the corresponding subtrees according to an *inconsistency coefficient* less than a given threshold (see e.g. [18]), obtaining a final sufficiently balanced partition (i.e. with an homogeneous number of antennas per cluster). Finally, for each cluster, we calculated the polygon corresponding to the union of the voronoi cells of its antennas. Hence we obtained the tessellation having these polygons as tiles;
- Automatic aggregation of municipalities (313 polygons): Each zone is an aggregation of municipalities. Each zone contains not less than 13 antennas. An automatic procedure has been created in order to build new zones in an iterative manner.

On these tessellations it was possible to map the direction and the intensity of mobile phone users' movements at an hourly basis. The data set was collected in different working days: five Wednesday respectively in July, August, September, October and November 2011. Using this data, we performed an analysis aimed at evaluating the overall mobility of cell phone users in the Lombardy region.

## 3 Analysis

### 3.1 Treelet Decomposition of Erlang

Erlang measures can give insights on different aspects of the urban area to which they refer, and their analysis can be developed with various scopes: the segmentation of the area into districts characterized by homogeneous telephonic patterns; the identification of a set of reference signals able to describe the different patterns of utilization of the mobile phone network in time.

Treelet decomposition is an effective dimension reduction technique for Erlang profiles and, more generally, for data with peculiar functional features, like spikes, periodicity, outliers.

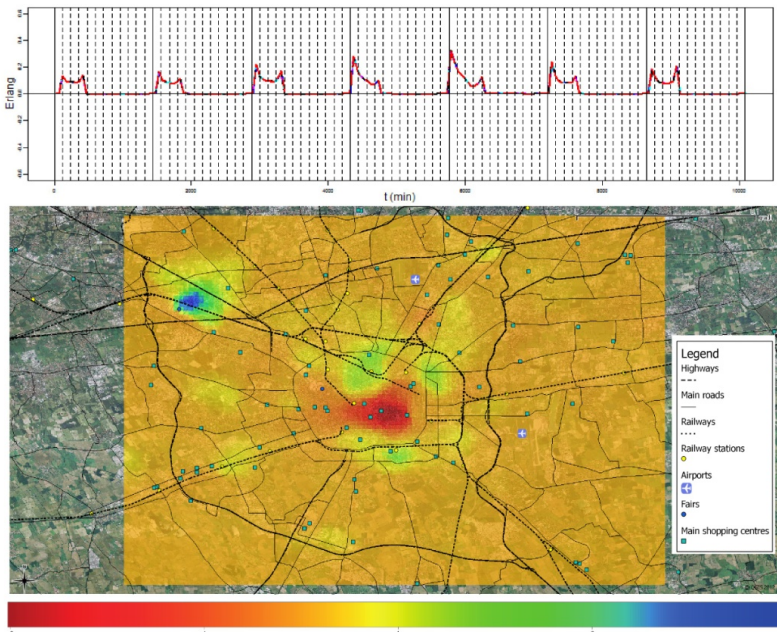
The methodology of Treelet decomposition [16], [19] allowed us to obtain: a reference basis reporting the specific effect of some activities on Erlang data; a set of maps showing the contribution of each activity to the local Erlang signal. The idea behind our approach is that different basic profiles (each being one element of the treelet decomposition basis) of city usages can concur in the same place and that the overall observed usage of a certain place is the superimposition of layers of these profiles.

We selected some results as significant for explaining specific patterns both of mobility and of city usages (commuting, nightly activities, distribution of residences, non systematic mobility). We tested their significance and their interpretation from an urban analysis and planning perspective at the Milan urban region scale.

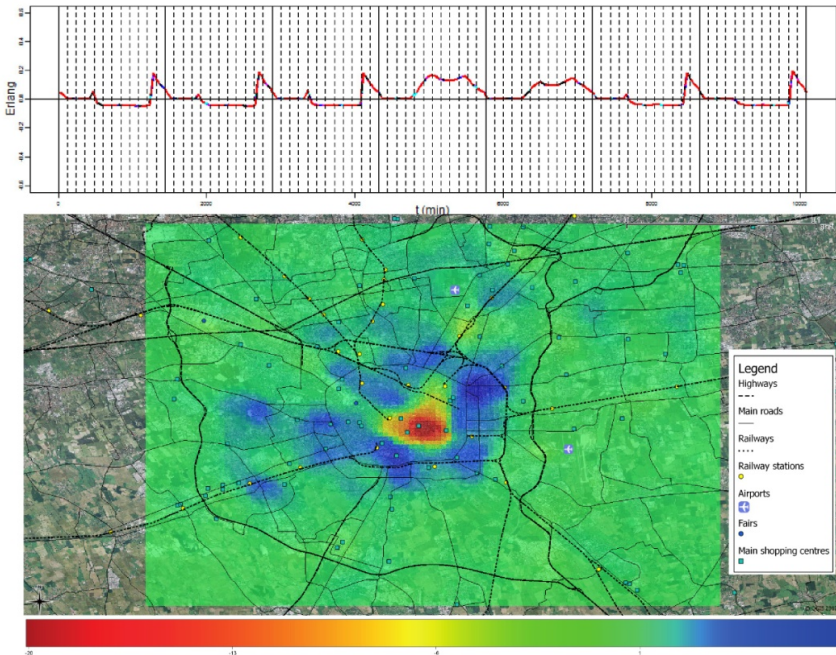
Each of the following figures represents one of the extracted city usage profile and is organized as follows:

- Top panel depicts the considered basic profile: x-axis represents time, spanning 7 days from Wednesday to Tuesday at an hourly rate. The dotted lines correspond to 2 hours while the continuous lines separate the different days of the week; y-axis: Erlang values;
- Bottom panel depicts a map of intensity values: colors show how much the upper profile concur to explain one place’s overall (telephone) use pattern.

Figure 2 is about the density of mobile phone activity late at night (in particular from midnight until 8 am). We can observe here some interesting hot spots where values are very high. For example, the exhibition district in the Northern Western side of the map. In the considered period an important Fair (the 2009 International Milan Design Week) was held and the peak fits well with the nightly activities necessary for the mounting and the organization of the site. Another point of interest is the Fruit and Vegetable Wholesale market in the South Eastern part of the region where consistent night work happens for delivering and distributing products that come from whole Italy and abroad. The city centre is characterized by a relative low value, according to the absence of relevant nightly activity inside it.



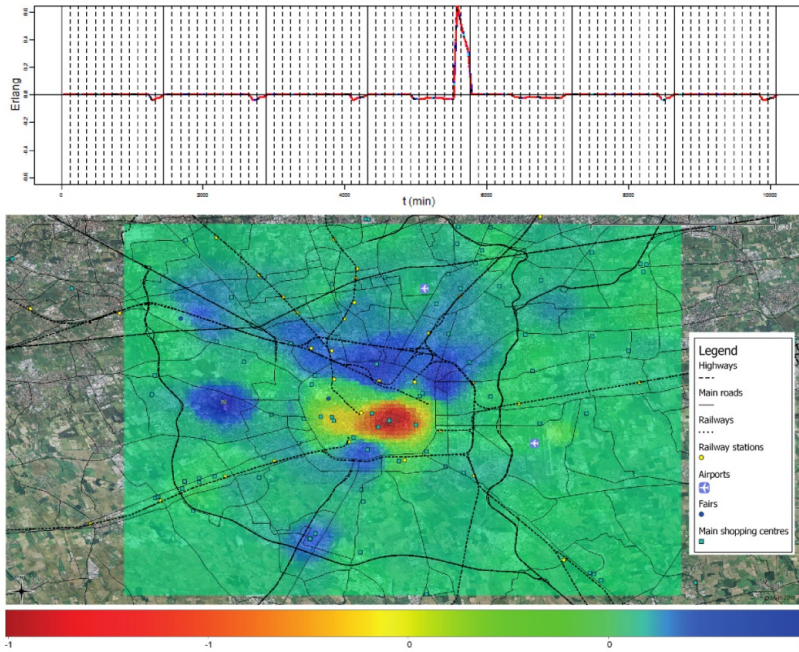
**Fig. 2.** Nightly activity. Hot spots highlight the presence of night work.



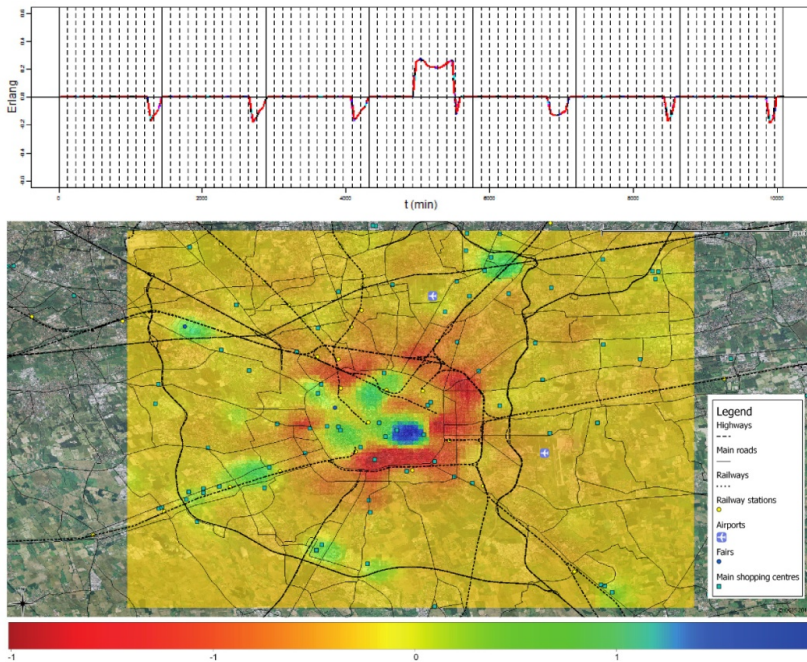
**Fig. 3.** Concentration of activities during evenings of working days and during daytime (from 8 am until 8 pm) of the week end: residential districts of the Milan urban region

Figure 3 puts in evidence some locations with high concentration of mobile phone activity during the evening of the working days and during daytime (from 8 am until 8 pm) of the week end. It shows a significant correspondence with main residential districts of the Milan urban region. It highlights a relevant concentration of homes along the second circular ring of the city, where the density of resident population reaches the highest value of Milan, but also in some municipalities with a residential profile and social housing in the south, south-west and in the north of the metropolitan area (Corsico, Rozzano, Sesto S.G). The city centre of Milan appears as a void, a fact that is consistent with the changes that occurred in the last decades, namely a gradual replacement of the residents with activities mainly related to the services sector and the commercial sector.

Figure 4 shows places with high density of activity during Saturday evening, from 8 pm until midnight. Focusing on the core city area, we notice several interesting patterns: a high activity in some places where there are many pubs and restaurants near the Milan Central Station, in the Navigli District, in the Isola Quarter and in other ambits characterized by the presence of leisure spaces (Filaforum Assago in the south of Milan), but also of activities in a continuous cycle as the hospitals. This treelet has proven to be effective in describing the temporal profile of the city lived by night populations during Saturday.



**Fig. 4.** Density of activity during Saturday evening (8pm-midnight). Saturday night population: leisure and hospitals



**Fig. 5.** Mobility practices. Saturday (10am- 8pm), shopping and leisure activity.



Figure 5 highlights another relevant mobility pattern, which is difficult to intercept through database traditionally used in urban studies: the shopping activity and, in general, the leisure activity. The map represents the density of mobile phone use during Saturday, from 10 am to 8 pm. Shopping and leisure are two of the main reasons of mobility in contemporary cities: they belong to the category of unsystematic mobility, and they significantly contribute to the even more complex mobility patterns that can be observed in the Milan urban region due to the distribution of commercial centers, commercial streets and, in general, of activities (museums, touristic sites, cinemas, just to cite some) inside and outside the city. These places attract, especially in certain days of the week, a huge amount of population coming from a vast territory that goes far beyond the administrative boundaries of the city. The map is the result of this spatial pattern and shows an important concentration of mobile phone traffic in the city centre and in other several places outside the city (most of them corresponding to the presence of commercial centers). The mainly residential areas, recognized in the previous Figure 3, are consequently characterized by the lowest value.

### 3.2 Aggregated Tracks of Mobile Phone Users

The analysis of the activity of mobile phone users permitted to put in evidence the main hourly distribution of origin destination movements of a huge sample of people (more than one million per day).

We started from the hourly origin destination matrices (October, 19<sup>th</sup> 2011) of mobile phone users among the 526 zones (tiles) of the more fine grained tessellation. For each zone it was available a set of directed connections towards the other and for each connection it was available the number of traced users.

Our goal was to find a synthetic visualization. Our proposal is to consider “prevalent fluxes” of mobility at different hours of a typical working day defined as the sum vector of all the fluxes moving from each zone.

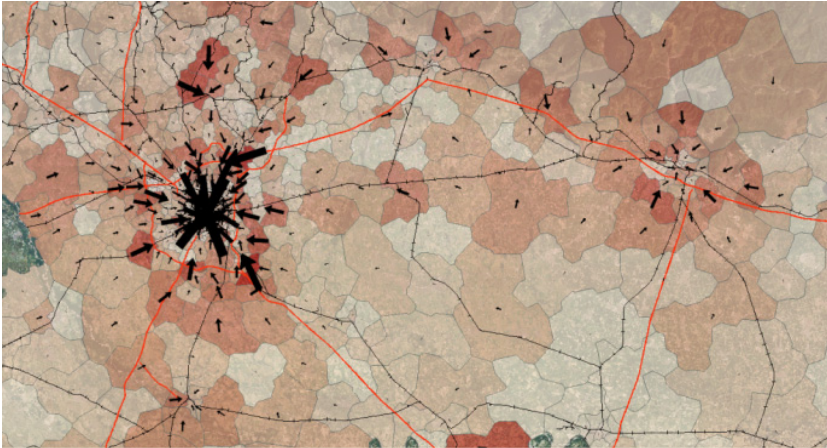
More specifically we associated to each origin destination flux, the vector applied to the centroid of the origin tile, directed to the centroid of the destination tile and having the magnitude given by the value in the origin destination matrix corresponding to the selected origin and destination. We then calculated, for each tile, the sum vector of all the vectors applied to their centroid. It is characterized by two dimensions: the magnitude, which is function of the magnitudes of the original vectors, and the angle which expresses the prevalent direction of the flux.

A set of maps of the sum vector moving from each zone at different hours has been produced in order to highlight the main patterns of mobility during a typical working day (October 19<sup>th</sup>, 2011). The interested reader can visit our interactive web version of the map showing prevalent fluxes of mobility in a working day at the following URL: <http://www.ladec.polimi.it/maps/od/fluxes.html> .

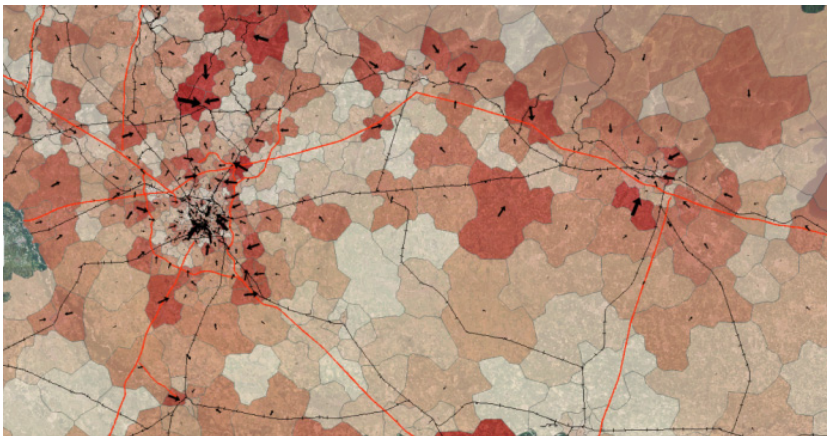
The maps cover a wide area that goes from Milan in the West to Brescia in the East and comprises many populated city regions (Pavia in the South, Monza in the North and Bergamo between Milan and Brescia).

The length and the thickness of the arrow are proportional to the magnitude of the sum vector and they are related to the prevalent fluxes of mobile phone users at specific hours. The convergence of travels toward the main centers during the

morning, the more complex direction of movements during the afternoon, are some interesting phenomena emerging from our analysis. The maps represent also the main infrastructures (railways and highways) and the zones of the tessellation, colored according to the intensity of the outgoing fluxes (light red color corresponds to lower intensity while dark red corresponds to higher intensity).



**Fig. 6.** Aggregated flows of mobile phone users: 9 am – 2011-10-19



**Fig. 7.** Aggregated flows of mobile phone users: 5 pm – 2011-10-19

The maps show *the variability* of phenomena that conventional data sources, such as census data, cannot give for a typical day. A broad use of the territory and an articulation of daily moves are visible every hour. The maps can be used as meaningful tools for monitoring the use of the infrastructural networks and of the urban spaces. On the one hand, the morning map (9pm; fig 6) confirms a polarization of movements towards the main centers offering job opportunities and highlights also the most commonly used infrastructures. On the other hand, the aggregated flows of

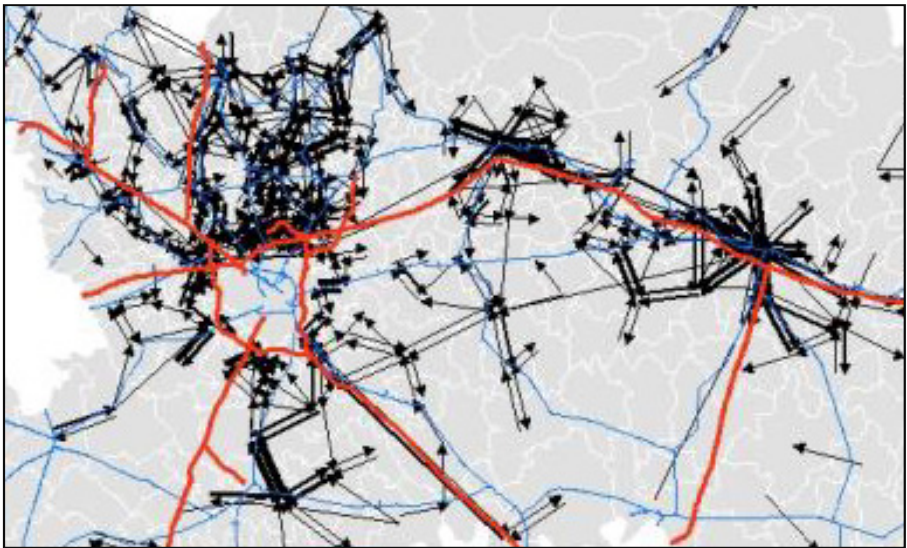
mobile phone users in the afternoon (at 5 pm; fig 7) allow to recognize significant places for shopping and leisure, that are attended after work. This type of information is difficult, if not impossible, to monitor through conventional data at a comparable spatial and temporal resolution.

The automatic aggregation of municipalities (313 polygons) has been used for producing maps of mobile phone users' fluxes directed from each zone towards the others. We used this tessellation in order to have a better visualization of the complex and more relevant outgoing and incoming connections between the zones.

Only flows of more than 100 users are shown by oriented lines that connect the centroid of each origin zone with the centroid of the destination zones. In order to better visualize the overall patterns of mobility in the region, in and out flows from/to Milan city were excluded.

The map (fig. 8) evidences relevant relations and fluxes towards the main cities of the Lombardy Region, but also some interesting patterns in the Northern Milan area where a high density of huge connections emerges.

It is also evident a linear element of interconnected centers along some important infrastructural corridors (i.e. the Sempione road in the northern western side of the region and along the highway in the western side of the map between Bergamo and Brescia). These maps show the complexity of daily mobility patterns that modify the hierarchical structure of the cities where traditionally the physical relationship between jobs and homes was the main reason of mobility. The density of fluxes at 5 pm (fig. 9) describes not only the return home, but also the unsystematic mobility related to individual habits, as an effect of the diversified uses of the Milan urban region.



**Fig. 8.** Origin destination fluxes of mobile phone users: 8 am, 2011–10-19

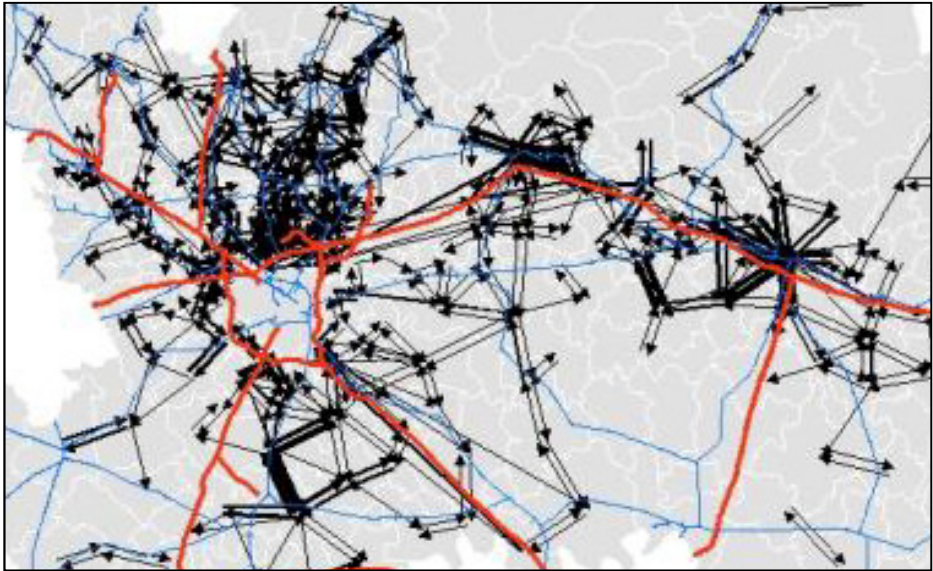


Fig. 9. Origin destination fluxes of mobile phone users: 5 pm, 2011-10-19

## 4 Implications for Policies

The research allowed us to test the potential of mobile phone data in explaining relevant urban usage and mobility patterns at the Milan urban region scale and in understanding the dynamic of temporary populations, two important topics that can be hardly intercepted through traditional data sources. This opens new implications for the urban research community which needs to elaborate new strategies to integrate traditional data with user generated data, such as mobile phone activity, in order to achieve a better comprehension of urban usages, in time and in space.

The presented data and methodology let the recognition of effective mobile populations in the urban environment. This knowledge can be exploited by decision makers for the definition of specific policies directed to temporary populations, which are more and more important in contemporary cities, otherwise ignored.

Describing the trends of use of urban spaces, the maps of mobile phone data give important information for mobility policies: the lack of coincidence between the mobility practices in the peak hours in the morning and in the afternoon when the chains of displacements are very articulate and complex, allows to recognize not only the variability in mobility practices, but also the places where these practices are occurring.

The commuters between 8 am and 9 am, become city users between 5 pm and 7 pm. This phenomenon strictly affects land use and can pose new questions and indications for transport policy.

Indeed, if we overlaid the boundary of the institutional management of local public transport in the Milan area with the areas of mobility practices, taken from the mobile phone data, we could observe the "deep structural effects of the mobility of people on urban policies" and the obvious disconnection between fixed jurisdictions and "mobile factors" [17].

The variability in the space-time of use of urban spaces resulting by mobile phone data is also revealed by:

- the spaces of night leisure that define a geography of places densely crowded at Saturday night, that is quite different from the territories of night work during the week (Monday to Friday night);
- the shopping and leisure spaces during the weekend (between 10 am and 8 pm) show the inner city center of Milan, but also some commercial malls along the ring roads;
- the space of the residence, where most significant call-densities are concentrated in the evenings and on weekends and represent the "negative copy" of the work places;
- the spaces of temporary events (International Design Week) that attract a significant portion of tourists and city users who visit several places in the city that are not identified by traditional sources and are not limited to traditional exhibition spaces.

The same data helps us to question some interpretations in the literature on the erratic behaviors of metropolitan populations, on the nomadism that characterizes the contemporary practices, that surveys on mobile phone data have already undertaken [10]. Some research about a significant sample of mobile phone data have, in fact, contested interpretations of nomadism of contemporary populations.

If they confirm the high density of commuting, they also show the strong recursion of the paths. In other words we move more during the day, but according to the known and usual paths.

## References

1. Ehrenberg, A.: *L'individu incertain*, Paris, Calmann-Lévy (1995)
2. Urry, J.: *Sociology Beyond Societies*. Routledge, London (2000)
3. Kaufmann, V.: *Re-thinking mobility*. Ashgate, Aldershot (2000)
4. Ascher, F.: *Les sens du mouvement: modernités et mobilités*. In: Allemand, S., Ascher, F., Lévy, J. (eds.) *Le Sens du Mouvement*, Belin Bourdin, pp. 21–34 (2004)
5. Bourdin, A.: *Les mobilités et le programme de la sociologie*. *Cahiers Internationaux de Sociologie* 1(118), 5–21 (2005)
6. Scheller, M., Urry, J.: *The new mobilities paradigm*. *Environment and Planning A* 38, 207–226 (2006)
7. Nuvolati, G.: *Resident and Non-resident Populations: Quality of Life, Mobility and Time Policies*. *The Journal of Regional Analysis and Policy* 33(2), 67–83 (2003)
8. Ahas, R., Mark, Ü.: *Location based services—new challenges for planning and public administration?* *Futures* 37(6), 547–561 (2005)

9. Ratti, C., Pulselli, R.M., Williams, S., Frenchman, D.: Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design* 33(5), 727–748 (2006)
10. Gonzalez, M.C., Hidalgo, C.A., Barabási, A.-L.: Understanding individual human mobility patterns. *Nature* 453(7196), 779–782 (2008)
11. Reades, J., Calabrese, F., Sevtsuk, A., Ratti, C.: Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing* 6(3), 30–38 (2007)
12. Becker, R.A., Caceres, R., Hanson, K., Loh, J.M., Urbanek, S., Varshavsky, A., Volinsky, C.: A Tale of One City: Using Cellular Network Data for Urban Planning. *IEEE - Pervasive Computing* 10, 18–26 (2011)
13. Soto, Frias-Martinez: Automated land use identification using cell-phone records. In: *Proceedings of the 3rd ACM International Workshop on MobiArch*, pp. 17–22. ACM (2011)
14. Soto, Frias-Martinez: Robust land use characterization of urban landscapes using cell phone data. In: *The First Workshop on Pervasive Urban Applications (PURBA)* (2011b)
15. Balducci, A., Fedeli, V., Pasqui, G.: *Strategic planning for contemporary urban regions: city of cities: a project for Milan*. Ashgate, Burlington (2010)
16. Manfredini, F., Pucci, P., Secchi, P., Tagliolato, P., Vantini, S., Vitelli, V.: Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region. *MOX Report 25/2012* (2012)
17. Estebe, P.: *Gouverner la ville mobile*. PUF, Paris (2008)
18. Jain, A., Dubes, R.: *Algorithms for Clustering Data*. Prentice-Hall, Upper Saddle River (1988)
19. Vantini, S., Vitelli, V., Zanini, P.: Treelet Analysis and Independent Component Analysis of Milan Mobile-Network Data: Investigating Population Mobility and Behavior. In: *Analysis and Modeling of Complex Data in Behavioural and Social Sciences - Joint Meeting of of the Italian and the Japanese Statistical Societies* (2012)