Urban Analysis for the XXI Century: Using Pervasive Infrastructures for Modeling Urban Dynamics

Enrique Frías-Martínez

Telefónica Research
Ronda de la Comunicación s/n, 28050, Madrid, Spain
efm@tid.es

Abstract — **The recent adoption of ubiquitous computing technologies has enabled the capture of large amounts of spatio-temporal data about human motion. The digital footprints computed from these datasets provide complementary information for the study of social and human dynamics, with applications in urban planning and transportation. In this context, cell phones and cell phone networks, due to its pervasiveness, can be considered sensors of human behavior and as such, can be used as proxies to study urban environments. In this paper we describe three applications that present the potential of using cell phone networks infrastructures for modeling urban dynamics: (1) the identification of dense areas; (2) the identification and classification of land uses and (3) the identification of routes within an urban landscape. Our approach provides a complementary source of information for traditional urban analysis techniques without the limitations that these techniques have.**

I. Introduction

With the increasing capabilities of mobile devices, individuals leave behind footprints of their interaction with the urban environment. As a result, new research areas, such as urban computing and smart cities, focus on improving the quality of life in an urban environment by understanding the city dynamics through the data provided by ubiquitous technologies [1][2].

Traditionally, urban analysis and the study of urban environments have used data obtained from surveys to characterize specific geographical areas or the behavior of groups of individuals. However, new data sources (including GPS, Bluetooth, Wi-Fi hotspots, geo-tagged resources, etc.) are becoming more relevant as traditional techniques face important limitations, mainly: (1) the complexity and cost of capturing survey data; (2) the fact that typically the data available is aggregated; (3) the fact that the data is static and represents a snapshot of the situation in a specific moment in time and (4) the increasing unwillingness of individuals to provide (what they perceived to be) personal information. One of the new data sources relevant for the study of urban environments are cell phone records, as they contain a wide range of human dynamics information (ranging from mobility, to social context and social networks) that can be used to characterize individuals or geographical areas.

A city is an inherently self-organized human-driven organization where individuals and their behavior play an important role in defining the pulse and the dynamics of the city. This implies that, in order to efficiently model human mobility, individual information is necessary in order to reflect that location is, at least in part, each individual's decision [3]. The datasets captured by ubiquitous computing technologies inherently reflect individual information relating to mobility and social dynamics and also capture those dynamics at different moments in time, thus making it possible to study the evolution of different parameters over time. As a result the models obtained using this data sources overcome some of the main limitations of traditional urban analysis approaches based on questionnaires.

In this paper we present three applications that use the data provided by a cell phone network infrastructure to demonstrate how the information extracted can be used to model city dynamics and as an extra source of information for urban analysis. The first application is the detection of dense areas, i.e. the identification of areas within an urban environment where people tend to go at specific moments during each day. The second application consists on automatically identifying land uses (i.e., residential, industrial, commercial, etc.) from the information provided by the cellular antennas, and the third application identifies routes within an urban environment, i.e. the amount of people that travels from the coverage area of one tower to another at specific moments in time. The models obtained from these applications have a variety of applications for urban planning or transport planning that will be detailed in each section. Although all the applications we are going to present have used data originating from cell phone networks, the same techniques and the same models can be generated and complemented with other resources of pervasive geo-localized data such as twitter, Flickr, foursquare, or the logs generated by any location-based service.

Some authors have already used cell phone traces to implement urban analysis studies. Among others, Ratti et al. [4] used aggregated cell-phone data to analyze urban planning in Milan, Eagle et al. [5] identified behavioral patterns from the information captured by phones carrying logging software, and in [6] the authors use Bluetooth to characterize pedestrian flow data. Other examples are [7], which monitors the dynamics of Rome and obtains clusters of geographical areas measuring cell phone towers activity using Erlangs (being 1 Erlang 1 person using the phone for 1 hour); Horanot et al. [8], that analyze four different geographical spots at different times in Bangkok; or Reades et al. [9] that use eigendecomposition to study the time structure and find correlations between the number of Erlangs and the commercial activity of an area.

The rest of the paper is organized as follows: first we present the basic concepts of a cell phone network infrastructure that define how the information is captured. After that we present the three applications previously presented, focusing mainly on the modeling results. The paper finishes with the conclusions section.

II. Data Collection in a Cell Phone Network

In order to capture human dynamic information for implementing urban analysis, we present some basic concepts about how cell phone networks work. Cell phone networks are built using a set of Base Transceiver Stations (BTS) that are in charge of communicating mobile phone devices with the cell network. The area covered by a BTS is called a sector. A BTS has one or more directional antennas (typically two or three, covering 180 or 120 degrees, respectively) that define a cell and all the cells of the same BTS define the sector. At any given moment in time, a cell phone is covered by one or more antennas. Depending on the network traffic, the phone selects the BTS to connect to. The geographical area covered by a cell depends mainly on the power of individual antennas. Depending on the population density, the area covered by a cell ranges from less than 1 Km^2, in dense urban areas, to more than 3 Km^2, in rural areas. Each BTS has latitude/longitude attributes that indicate its location, a unique identifier BTS_{id}, and a polygon representing its sector. For simplicity, we assume that the sector of each BTS is a 2-dimensional non-overlapping region, and we use Voronoi diagrams to define the covering areas of the set of BTSs considered. Figure 1(left) presents a set of BTSs with the original coverage for each cell, and (right) the simulated coverage obtained using Voronoi diagrams. While simple, this approach gives us a good approximation of the coverage area of each BTS.

CDR databases are populated when a mobile phone connected to the network, makes or receives a phone call or uses a service in the network (e.g., SMS, MMS, etc.). In the process, the information regarding the time and the BTS where the user was located when the call was initiated is logged, which gives an *indication* of the user's geographical location at a given moment in time. Note that no information about the *exact* user's location inside a sector is known. Furthermore, for a given call it is possible to store not only the initial BTS during the period of a call, but also all BTSs used during it in case caller/callee move to other cells in the network (*hopping*). The set of fields typically contained in a CDR include: (1) originating encrypted phone number, (2) destination encrypted phone number, (3) identifier of the cell that handled the originating phone number (if it is a phone number of the carrier), (4) identifier of the sector that handled the destination phone number(if it is a phone number of the carrier), (5) date and time of the call and (6) duration of the call and (7) extra information regarding error codes detailing the cause for call termination, indicators of roaming, etc.

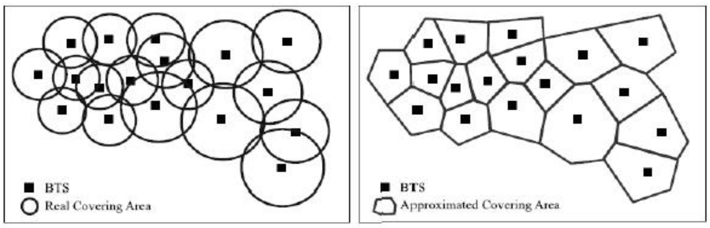

Figure 1. (Left) Example of a set of BTSs and their coverage and (right) approximated coverage obtained applying Voronoi tessellation.

III. Identification of Dense Areas

Some of the applications of smart cities and the study of social dynamics include traffic forecasting [1], modeling of the spread of biological viruses [2], urban and transportation design [1] and location-based services [6]. A challenging and interesting problem common to all this applications is the identification of areas of high density of individuals and their evolution over time. This information is of paramount importance for, among many others, urban and transport planners, emergency relief and public health officials, as it provides key insights on where and when there are areas of high density of individuals in an urban environment. Urban planners can use this information to improve the public transport system by identifying dense areas that are not well covered by the current infrastructure, and determine at which specific times the service is more needed. On the other hand, public health officials can use the information to identify the geographical areas in which epidemics can spread faster and, thus, prioritize preventive and relief plans accordingly.

The problem of dense area detection was initially presented in the data mining community as the identification of the set(s) of regions from spatio-temporal data that satisfy a minimum density value. This problem was initially solved for spatial and multidimensional domains [10], and later for spatio-temporal domain [11]. In the former proposals, no time dimension is considered, while in the later ones only moving objects, typically represented by GPS sensors that continuously report their locations, are considered. Common to all of the above methods is that a fixed-size non-overlapping grid or circle employed to aggregate the values over the spatial dimensions are considered. Therefore, these methods "constrain" the shape of the detected areas and, generally, identify dense areas that are a superset/subset of the desirable dense areas. Ideally, we seek a technique that is able to detect dense areas whose shape is as similar as possible to the underlying dense geographical areas.

We have developed a new technique, the Dense Area Discovery (DAD-MST) algorithm to automatically detect dense areas in cell phone networks. Our approach, unlike the previous approaches, is not based on fixed-size grids, but on the natural tessellation of the spatial domain, thus overcoming the limitations of all the previous approaches. The DAD-MST algorithm, given a set of BTSs, BTS={bts_1,…,bts_n}, that gives coverage to a geographical region R, discovers the optimal disjoint subsets of BTS that cover areas within R where the number of unique users reaches a maximum in a specific time period. An exhaustive exploration of all possible disjoint subsets of BTS becomes a daunting task as the number of BTS increases. Thus, the algorithm was based on the Maximum Spanning Tree (MST) algorithm that selects, at each step, the best subsets of BTS. The rest of the section focuses on the application of DAD-MST, its algorithmic details and the result evaluation can be found in [12].

A. City Dynamics from a Dense Area perspective

We collected cell phone data in the form of CDR for the city of Guadalajara (state of Jalisco, Mexico). We used a sample of the dataset containing the calls of over one million anonymized unique customers over a period of four months, with around 50 million CDR entries collected with 500 BTS towers.

Figure 2 shows some landmarks of the city under study, having as a reference the subway system. The two subway lines in the city (represented by dotted black lines), run East-West (L1) and North-South (L2) with one central station in common. For reference purposes the central station is denoted by C, with stops north of C denoted as N1 to N7, stops south of C denoted S1 to S11 and stops east of C denoted E1 to E9. The downtown area is geographically located around C, E1, E2, E3 and E4. Near C we find university buildings, government offices and parks. The vicinities of E1 and E2 form the commercial part of the city with markets, commercial streets and hotels. E3 and E4 have more university buildings and night life area. The rest of L1 services mainly residential areas. Regarding L2, around S3 to S11 there are mainly residential neighborhoods with light industrial areas. N2 to N7 serve residential areas with some commercial and entertainment places. The map also indicates other places such as a Stadium complex (S) and the city Zoo (Z), the main zoo of the country. For the areas not commented, as a general rule, there is a mixture of residential areas (with different densities) and light industrial areas, with the north and north-west having more affluent areas than the south.

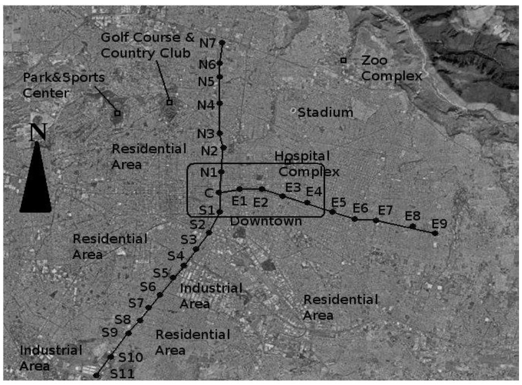

Figure 2. Main landmarks of the city of Guadalajara, Jalisco, Mexico.

Figures 3 and 4 depict the graphical representation of the dense areas detected for the city using the DAD-MST algorithm during mornings (7am to 11:59am), afternoons(12pm to 4:59pm), evenings (5pm to 9:59pm) and nights (10pm to 3am) in term of number of users during weekdays and weekends respectively. The area R was chosen to cover the city and its metropolitan area as to reflect the social dynamics between the metropolitan belt and the city. The area presented in the figures is a smaller rectangle of 20km*15km centered in the city in order to appreciate downtown dense areas. Note that the same color in different time frameworks does not imply the same density, just the same relative importance. The combination of Figures 3 and 4 shows the evolution of dense areas during weekdays and weekends and thus an indication of how people move in the city. It has to be noted that the dynamics reflected by the evolution of dense areas does not necessarily imply flocks of individuals moving from one area to another, just density of individuals changing over time. The top five dense areas presented in each map of Figures 3 and 4 have been numbered from 1 to 5 (being dense area #1 the one with the highest density) to facilitate the reading of the relevance. If one number is not present is because it is not located in the area of 20km*15km showed in the Figures 3 and 4 but is present somewhere in R, i.e. in those cases the dense areas are located outside the city somewhere in the metropolitan belt.

Following the temporal sequence of the evolution of dense areas during weekdays (see Figure 3) it can be observed that downtown is covered by a big dense area in the morning, focusing on the university, the commercial and the governmental district. Also in the morning, the second and fifth dense areas are located in residential zones and dense areas #3 and #4 are outside the city. This indicates that in those hours although downtown is the top dense area, the metropolitan belt of the city also has important density of individuals. In the afternoon, the top dense area that appeared in the morning disappears, and a new dense area, ranked #2, located around E1 to E4 and focusing on the commercial area appears. Also in the afternoon, 3 out of the top 5 dense areas identified are in the metropolitan belt. Note that because of the nature of the data, we are strictly representing the number of people that have used their cell phone in that period of time. It can be the case that once people get to the working place or to the university in downtown, cell phones are not used with the same frequency. This would motivate the disappearance of the dense area in the government and university districts. On weekday evenings and nights the activity is concentrated in downtown, with the top dense area around the commercial and business districts and aligning with the subway lines. Both in the evening and at night there is a shift, when compared to mornings and afternoons, in the localization of the top dense areas from the metropolitan belt to downtown. To represent that shift, in the evening 4 out of the 5 top dense areas are in the city, while at night the 3 top ones are in the city.

Following the temporal sequence of the evolution of dense areas during weekends (see Figure 4) it can be observed that in the morning and in the afternoon the top dense area (#1) is outside the city. In both cases there are dense areas in the commercial and business districts in downtown, although they are not in the top 3. Two dense areas appear north of downtown, which include the stadium complex (marked with S*)* and the zoo (marked with Z) respectively. Both dense areas are present in the four range hours during weekends (note that both complexes cover a small geographical area of the dense areas identified). During the evenings and at night the top dense area is in downtown around the subway stops E1 to E4, i.e. in the commercial and nigh life area. Both evening and nights during weekends have the same dense areas, indicating that there is no change in the dynamics of the city. As it happened during weekdays, in the evening and at night there is a shift in the top dense areas from the metropolitan belt to the city. Both during weekdays and weekends residential areas are identified south of downtown.

The identification of dense areas in residential neighborhoods is tightly related to the density of housing, where lower income neighborhoods tend to have a higher density than more affluent neighborhoods. This is probably one of the reasons why residential dense areas are identified mainly in the south and none in the north. A direct application of this knowledge regarding the social dynamics of the city is to help in the decision process of the design of the public transport infrastructure. In general, as mentioned in the previous section, there is an alignment between the subway lines (especially N1-N11 and E1 to E7) and the dense areas, indicating that the main dense areas are covered. The dense areas of the south-west are not directly covered by S1-S11, although they are in the vicinity. Note that our algorithm has also identified dense areas in the southeast corner of the city, where there is no subway service right now. From an urban planning perspective, this information could be used to propose line extensions to public transport officials. Also it is relevant that dense areas are very relevant in the metropolitan belt, and typically in the top 5 most important ones, so enough means of communication (buses, trains, etc.) have to communicate those areas with the city, specially between afternoon and evening when dense areas shift from the metropolitan belt to the city.

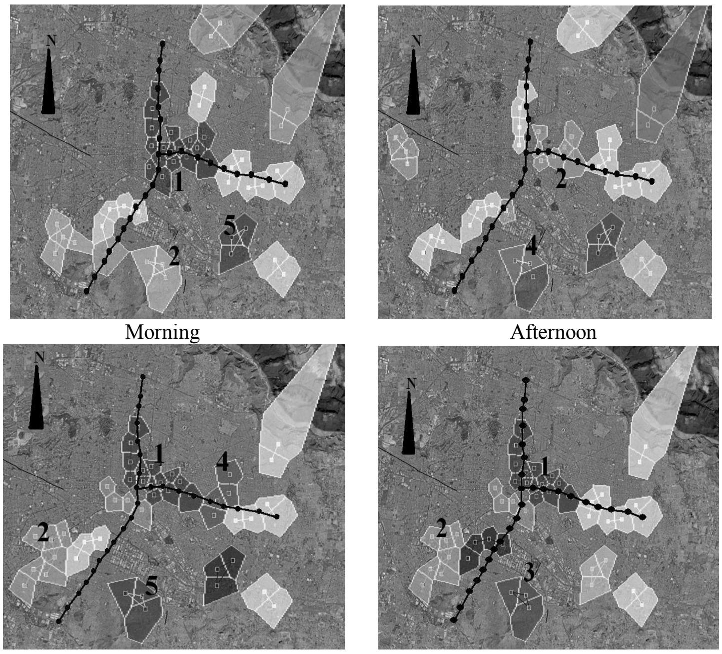

Figure 3. Area of 20km*15km of the city under study and the dense areas detected by DAD-MST in the morning, afternoon, evening and night during weekdays.

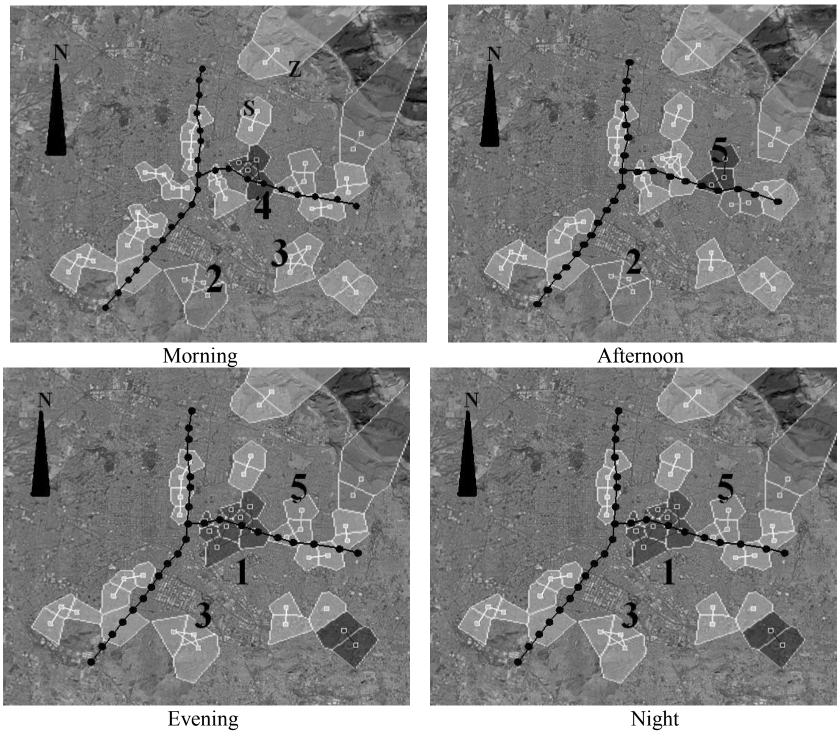

Figure 4. Area of 20km*15km of the city under study and the dense areas detected by DAD-MST in the morning, afternoon, evening and night during weekends.

IV. Land Use Identification in Urban Landscapes

Using the information contained in a CDR database we can characterize the uses given to specific urban areas. The geographical areas in which the city is going to be divided will be defined by the Voronoi tessellation of the set of BTSs, and each area will be characterized with the corresponding BTS activity (the signature of the BTS tower). The identification of land uses can be automatically done by clustering the set of signatures using clustering techniques (such as k-means).

We define the activity of a BTS, and by extension of its area of coverage, as the number of calls that are managed by that BTS over a given period of time. In our study we have measured the activity every five minutes. Human dynamics are well differentiated between week days and weekend days [13], and those differences will translate into different BTS levels of activity. In order to preserve that information we opt to build each BTS signature as the concatenation of the averaged activity of the BTS during weekdays (Monday to Friday) and weekends (Saturday and Sunday) (see Figure 5 for some examples).

Our study has been done using CDR data collected from Madrid during a period of 1 moth, from October 1st 2009 to October 31st 2009. The area covered by the city is of 400Km2, with more than 3 million inhabitants, and is served by 1100 towers that collected over 100 million interactions. Once each area of coverage has been characterized with its BTS signature, k-means was applied to identify land uses. The optimum number of clusters was identified to be 5. More details about the definition of the signatures and the clustering process can be found in [14][15].

Figure 5 presents the signatures of the class representatives for the five land uses found. Obviously, these behaviors are heavily influenced by cultural characteristics, and as such, although similar land uses could be identified for other countries, the signatures would be shifted according to cultural and social routines. Also, Madrid is a traditional European city in the sense that the old part of the city concentrates commercial, office and residential areas, which should difficult the identification of land uses. The clusters are characterized by:

- Cluster 1: This cluster is characterized by the fact that the activity takes place mainly during weekdays, especially in working hours, and weekend activity is almost nonexistent. During weekdays the activity is heavily focused between 10AM and 14PM and another peak of activity between 16:00 and 19:00 hours. This cluster shows a clear work related activity and the hypothesis is that the BTS coverage areas included in this cluster are used as industrial parks and/or office areas.

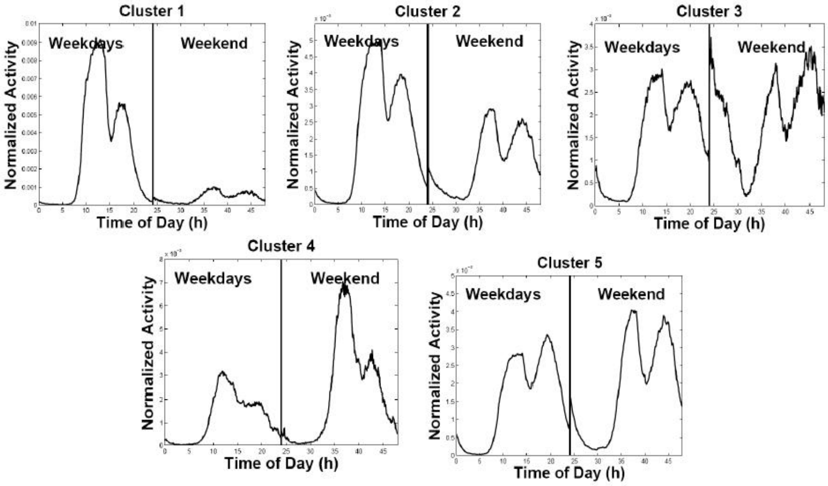

Figure 5. From left to right and top to bottom, class representatives of cluster 1 through cluster 5 of the land uses identified by k-means when using the weekday-weekend aggregation and Euclidean distance.

• Cluster 2: This cluster represents a hybrid land use. During weekdays there is activity during commercial hours, with two peaks of similar normalized intensity (0.005 and 0.004) at around 12AM and 19PM. There is a relevant weekend activity although of less intensity (both peaks at around 0.003). This behavior could be considered commercial. Nevertheless, due to the mixed activities that take place on the city, the signature could also represent office and/or residential areas.

• Cluster 3: The third cluster has two elements that characterize its land use: (1) activity during weekdays and weekends has the same relevance and (2) during weekends there is a strong activity between 0AM and 5AM, indicating nightlife environments. The land use derived from this behavior is of nightlife areas: restaurants, bars, etc.

• Cluster 4: This class representative shows that activity during weekends is more than twice that of weekdays, and that this activity is focused between 12PM and 5PM. This land use probably characterizes weekend leisure activities.

• Cluster 5: This signature shows that activity during weekdays and weekends is of the same magnitude. During weekdays, the second peak of activity is of higher magnitude than the morning peak, while during weekends both peaks have the same magnitude. These characteristics imply residential areas, where individuals come back home after work.

A. Validation

Ideally, any validation should be done using a ground truth. Urban planning departments usually have some information available on urban land use. Nevertheless, in our experience, the information available has the following characteristics: (1) it is regarding how land use is planned not on the actual use of the land, which is not necessarily the same; and (2) urban planning focuses mainly on defining residential and industrial areas, not the variety of uses we have discovered. Considering the previous factors the validation has been done using our expert knowledge of the city.

Figure 6 shows the geographical representation of clusters 1 through 4 in a map of Madrid. Generally speaking, the city is contained inside two concentric ring roads (M-30 and M-40). The area inside the M-30 (the smaller ring) contains the city center as well as the main business, tourist and commercial areas, all of them mixed with residential areas. The area comprised between the M-30 and M-40 contains mainly residential districts and industrial parks. Due to space constraints cluster 5 has not been presented, although it is geographically formed by the areas not included in the previous clusters. The validation of each cluster consists on checking to which extent the interpretation of the signatures given in the previous section actually correlates with the geographical location of the clusters. Our main findings are:

• Cluster 1 - Industrial Parks & Office Areas (Fig.6(a)): The geographical areas included are mainly around Castellana, which concentrates the main business areas of the city. Also included are industrial parks and office areas situated in the north and east of the city (including both the Telefónica campus, known as Distrito C, and Telefónica Research buildings), hospital complexes and some public buildings (such as the Congress). The assumption made for this cluster seems to hold true, although not only for offices and industrial areas but also for hospital complexes and public buildings.

• Cluster 2 - Commercial Areas (Fig.6(b)): The areas included in this cluster focus mainly around the Salamanca (marked with S), Chamartin (Ch) and Moncloa (M) districts. These districts have a strong commercial activity, but are also densely populated residential areas. It also contains some Industrial Parks (IP) with a strong commercial activity. Our hypothesis is partially validated as the areas included are mainly commercial-residential.

• Cluster 3 - Nightlife areas (Fig.6(c)): The correspondence in this case with nightlife areas is straightforward, as the main nightlife areas of the city are included: Callao (C), Bilbao (B), Moncloa (M), Alonso Martinez (AM), Huertas (H). Also there is activity in Mercamadrid, the wholesale food market of the city, whose activity takes place early in the morning, and as such is clustered as a nightlife area.

• Cluster 4 - Leisure areas (Fig.6(d)): As in the previous case the behavior of this cluster is highly correlated with the geographical location of the cluster. Good examples are parks (Casa de Campo), golf clubs (GC), the Horse Racing Track (H), Plaza de Oriente, El Prado Museum and the flea market (which happens Sunday mornings).

• Cluster 5 - Residential areas: The area covered by this cluster accounts for approximately 60% of the geographical area. It is highly correlated with residential areas, mainly in the south and west of the city. We can conclude that the land use assumptions made in the previous section with the cluster representatives are validated because there is a strong correspondence between those uses and the infrastructures included in the geographical representation of each cluster.

We can conclude that the land use assumptions made in the previous section with the cluster representatives are validated because there is a strong correspondence between those uses and the infrastructures included in the geographical representation of each cluster. The knowledge provided by the class representatives can be used to automatically classify the land use of other cities. In [15] we present a classifier for the land uses of Barcelona using the land use signatures identified.

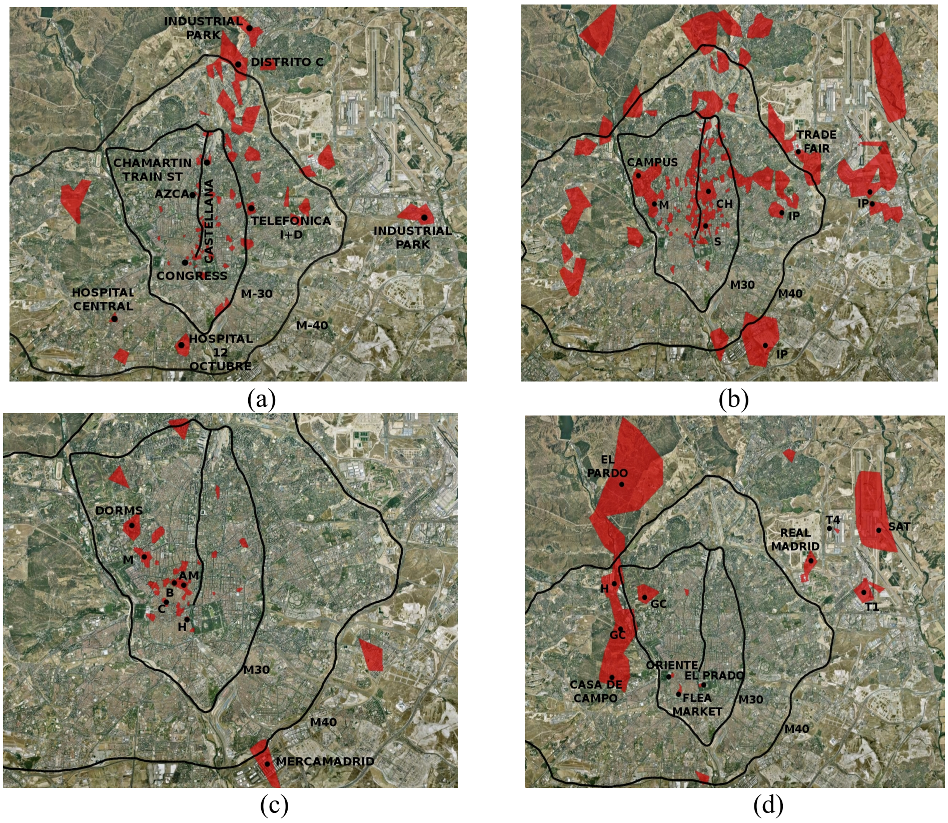

Figure 6. Geographical representation in a map of Madrid of: (a) Cluster 1 (Industrial Park – Office Areas), (b) Cluster 2 (Commercial-Residential), (c) Cluster 3 (Nightlife) and (d) Cluster 4 (Leisure).

V. Route Detection

The spatio-temporal information contained in CDRs can be aggregated to present the main commuting patterns and routes within a city. Although the routes can be presented at a BTS level, it makes more sense to group BTSs according to some geographical distribution (like neighborhoods or towns) in order to better understand the results.

Formally the problem can be presented as the identification of frequent changes between two geographical areas at specific moments in time. The literature already contains techniques to implement such approaches, mainly the Apriori algorithm [16] for learning association rules. In our approach, given a time slot (morning, afternoon, evening or night), we count the number of times there is a movement between all pairs of the geographical areas defined. After that, the support and confidence of each geographical commute can be obtained and the values used to filter the most relevant routes identified. As an example, in Figure 7 we present the main routes identified for the city of Madrid from the Moncloa neighborhood, during weekdays in the morning, indicating that the main commutes from that neighborhood are done towards Centro, Chamberi, Tetuan and Pozuelo (with their respective confidence values). Such information is key for the design of transportation routes or to correlate the current transportation routes with the actual needs of the city in order to define new public transport routes.

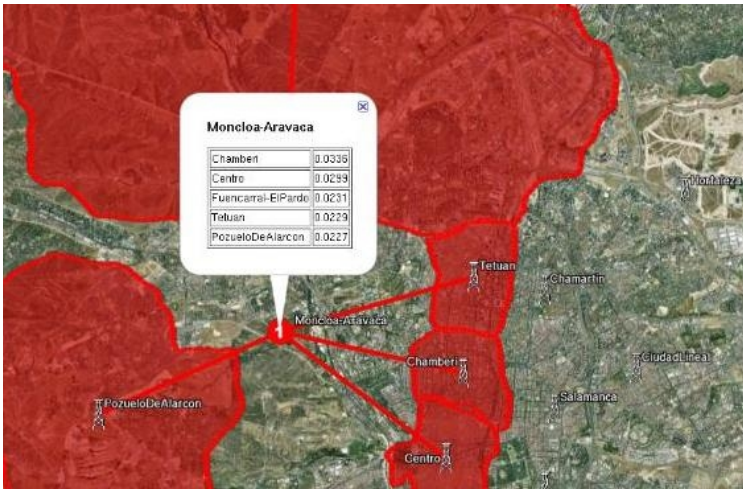

Figure 7. Example of the main commuting routes originating in Moncloa-Aravaca in a weekday morning (7am to 10am).

VI. Conclusions

Traditional urban analysis techniques are based on questionnaires which imply time and cost limitations. By extracting the knowledge generated by pervasive infrastructures, such as cell phone networks, we can generate models that explain urban dynamics and that complement traditional urban analysis approaches while overcoming the previous limitations. We have presented just three applications of how CDR data can be used to explain city dynamics at an aggregated level and how that information can be considered to define transport infrastructures or identify land uses. The examples are just intended as demonstrators of the capabilities of such approach, and can be used as building blocks to develop more specific applications. As future work we plan to develop similar applications but personalizing the study for different variables such as age groups, tourists or socio-economic levels.

References

[1] L. Liao and D.J. Paterson and D. Fox and H. Kautz, "Learning an inferring transportation routines," *Artificial Intelligence*, vol. 171, 2007.

[2] D. Brockmann, "Human mobility and spatial disease dynamics", *Review of Nonlinear Dynamics and Complexity, Wiley*, 2009.

[3] I. Beneson, "Modeling population dynamics in the city: from a regional to a multi-agent approach", *Discrete Dynamics in Nat. and Soc.*, 1999.

[4] C. Ratti, R. M. Pulselli, S. Williams and D. Frenchman, "Mobile Landscapes: using location data from cell phones for urban analysis", *Environment and Planning B: Planning and Design*, vol 33, no. 5, pp. 727-748, 2006.

[5] N. Eagle and A. Petland, "Reality Mining: Sensing Complex Social Systems", *Personal and Ubiquitous Computing*, vol. 10, no. 4, 2006.

[6] R. Ahas and U. Mark, "Location Based Services -- New Challenges for Planning and Public Administration", *Futures*, vol. 37, num. 6, 2005.

[7] J. Reades, F. Calabrese , A. Sevtsuk and C. Ratti, "Cellular Census: Explorations in Urban Data Collection", *IEEE Pervasive Computing*, vol. 6, num. 3, pp. 30-38, 2008.

[8] T. Horanont and R. Shibasaki, "Evolution of Urban Activities and Land Use Classification through Mobile Phone and GIS Analysis", *CUPUM*, 2009.

[9] J. Reades, F. Calabrese and C. Ratti, "Eigenplaces: analysing cities using the space-time structure of the mobile phone network", *Environment and Planning B: Planning and Design*, vol. 36, num. 5, pp. 824-836, 2009.

[10] R. Agrawal, J. Gehrke, D. Gunopulos and P. Raghavan, "Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications", *1998 SIGMOD Conference*, pp. 94-105, 1998.

[11] M. Ester, H.-P. Kriegel, J. Sander and X. Xu,"A local-density based spatial clustering algorithm with noise", *ACM SIGDD Conf. on Knowledge Discovery and Data Mining*, pp. 226-231, 1996.

[12] M. Vieira, V. Frias-Martinez, N. Oliver, E. Frias-Martinez, "Characterizing Dense Urban Áreas from Móbile Phone-Call Data: Discovery and Social Dynamics", *2nd. Int. Conf. on Social Computing (SocialCom 2010)*, 2010.

[13] J. Candia, M. Gonzalez, P. Wans, T. Schoenharl and A.-L. Barabasi, "Uncovering individual and collective human dynamics from mobile phone records", *J. Phys. A: Math. Theor.*, vol. 41, 2008

[14] V. Soto and E. Frias-Martinez, "Robust Land Use Characterization of Urban Lanscapes using Cell Phone Data", *1st Workshop on Pervasive Urban Applications, in conjunction with 9th Int. Conf. Pervasive Computing*, San Francisco, 2011.

[15] V. Soto and E. Frias-Martinez, "Automated Land Use Identification using Cell-Phone Records", *ACM Hotplanet 2011, 3rd Int. Workshop on Hot Topics in Planet-Scale Measurement, co-located with ACM Mobisys 2011*, Washington D.C.

[16] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases", *20th Int. Conf. on Very Large Data Bases VLDB*, pp. 487-499, 1994.