Explorations into Urban Mobility Patterns
Using Aggregate Mobile Network Data

Paper# TSI-SOTUR-08-03

November 2008

Andres Sevtsuk
Massachusetts Institute of Technology
Explorations into Urban Mobility Patterns
Using Aggregate Mobile Network Data

MIT Portugal Program
Transportation Systems Focus Area

Research Domain:
Transportation Systems Integration

Research Project:
Strategic Options for Integrating Transportation Innovations and Urban Revitalization
(SOTUR)

Paper#:
TSI-SOTUR-08-03

November 2008

Andres Sevtsuk
PhD candidate
Department of Urban Studies and Planning
Massachusetts Institute of Technology
asevtsuk@mit.edu

This publication was made possible by the generous support of the Government of Portugal through the Portuguese Foundation for International Cooperation in Science, Technology and Higher Education and was undertaken in the MIT-Portugal Program.

Comments and edits to this publication were made by Charisma Choudhury and Chris Zegras of MIT and Francisco Camara Pereira of Universidade de Coimbra.
<table>
<thead>
<tr>
<th></th>
<th>Table of Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction....................................................................................................3</td>
</tr>
<tr>
<td>2</td>
<td>Using Erlang measures as indicators of population distribution......................4</td>
</tr>
<tr>
<td>3</td>
<td>The daily regularity in the population distribution of Rome..........................8</td>
</tr>
<tr>
<td>4</td>
<td>Understanding the activity patterns of places..............................................14</td>
</tr>
<tr>
<td>5</td>
<td>Conclusion ......................................................................................................19</td>
</tr>
<tr>
<td>6</td>
<td>References ......................................................................................................21</td>
</tr>
<tr>
<td>7</td>
<td>Acknowledgements ..........................................................................................22</td>
</tr>
</tbody>
</table>
1 Introduction

In this paper we explore the daily population distribution patterns in the city of Rome in Italy, using a unique database describing aggregate calling activity over time in a mobile phone network. We address the following questions: Does the population distribution in Rome follow routine hourly, daily and weekly patterns? And how do such patterns vary in different parts of the city? The first question is meant to explore whether the urban activity distribution follows a habitual trend. The aim of the second question is to investigate potential explanations to the observed trends in activity distribution between different parts of the city.

Using the cells of a mobile network as spatial units of analysis, we focus on the behavior of places rather than individuals. Previous research using travel surveys or tracking of individuals has demonstrated that individual daily travel patterns exhibit great regularity (Hanson & Huff 1988; Schlich & Axhausen 2003; Schafer 2000; Eagle & Pentland 2007; González et al. 2008) and that individual mobility patterns are strongly related with land-use patterns as well as the built environment of a city (Ewing & Cervero 2001; Hanson & Giuliano Ed. 2004; Maat et al. 2005; Vilhelmson 1999). Individual travel patterns provide an valuable description of personal navigation in the built environment, but lack the capacity to explain how a city as a system in its own right is used in daily operation. In this paper we explore whether regularities in personal travel are also reflected in the daily population distributions of specific areas, and the city as a whole. By classifying places according to their network activity signatures, we attempt to explain the observed daily population distribution trends through socio-economic characteristics of their respective areas. The findings are meant to inform transportation planners as well as spatial planners. Understanding mobility patterns from an aggregate city-wide perspective could allow researchers to better evaluate the usage loads that transportation infrastructure and the built fabric face in daily life, and lead to more informed decision making in their planning.

Obtaining detailed measurements on population movements has raised great difficulty in the past. Traditional origin destination (OD) matrixes that transportation planners commonly use, usually only estimate the total amount of trips between each OD cell per day. At best, they also detail the amount of trips at rush hours or weekends. They do not tell us how different locations around the city are frequented during different hours of a day, thus omitting a large proportion of actual trips\(^1\). In order to account for more precise patterns, researchers have started to use estimates from mobile phone networks. Tracking of individual mobile phones over long periods of time has allowed studies to approximate detailed travel trajectories of individuals (Ahas et al. 2005), estimate residential and employment locations through so called geographic ‘anchor points’ of individual trajectories where people tend to spend most time (Aasa et al. 2008; Eagle & Pentland 2007), and study the regularity of individual daily travel patterns (González et al. 2008).

\(^1\) The U.S. Department of Transportation estimates that only 17.7% of all trips are now work related. Source: The 2001 National Household Travel Survey, daily trip file, U.S. Department of Transportation
Mobility patterns in these studies have been analyzed from the individual’s point of view, which due to the sensitivity of the data, has raised significant privacy concerns.\(^2\) Difficulty of obtaining, as well the limited sample size of individuals tracked, has previously hampered the possibility of exploring general city-wide mobility patterns, and studying the relationship between mobility and urban land use through mobile network data. In our case we use aggregate network activity data similar to Ratti et al. 2006, and Reades et al. 2007, which describes the general distribution of a population rather than individuals, known as Erlang measures in the industry.

2 Using Erlang measures as indicators of population distribution

Mobile phone usage is now ubiquitous in many countries. In several European countries there are more registered mobile phones than people. In Italy, the location of our case study, there are 1.24 mobile phones per person (CIA World Factbook, 2005). The calling patterns observed in the network thus offer great potential for estimating the distribution patterns of the population. However, the data also has its reservations: unlike OD matrixes, or individual cell-phone tracking, Erlang data does not indicate where a caller comes from or goes to, it simply estimates the amount of callers in a given network cell at a given time. Furthermore, the actual number of people in a cell is not explicitly shown in the Erlang data, but must be derived through statistical procedures involving error, as described below.

We obtained data for this study from Telecom Italia Mobile (TIM), the largest service provider in the Italy. Besides TIM, there are three other large service providers in the country: Omnitel Vodafone, Wind and Blue. TIM is currently the market leader in the city, supplying about 40.3% of the share. This constitutes approximately one million users in Rome, less than half of the city’s population. TIM’s data used in this study did not describe the activity of all the registered users in Rome, but only those who were actively engaged in phone calls during the measurement periods in Rome. We extracted longitudinal data on 398 network cells, which fall within the ring road that delimits the city, at 15 minute intervals over 24 weeks in 2006. Figure 1 shows the voronoi polygons of the cell areas that we analyzed.

---

Unfortunately the precise number of clients in each cell over every 15-minute period is not available in Erlang measurements. Instead the Erlang measurements indicate the aggregate call volume in particular cells. An Erlang measure is essentially a use multiplier per unit time. The use of one mobile phone for one hour in a particular cell constitutes one Erlang, whereas the use of two phones for half an hour each also constitutes one Erlang. Therefore Erlang values are affected by both the amount of calls done by each user, and each call’s duration, they do not tell us exactly how many users were connected to the network in a particular cell. The values are strongly affected by the likelihood of people to make phone calls: if the likelihood of making a call in a given cell increases, during a rush hour for instance, then the Erlang values also increase, regardless whether the actual population in the cell has changed or not. It is clear that there are periods when the entire population is more likely to make calls, daytime hours as opposed to night time hours for instance. In order to overcome this bias, we normalized all the Erlang values in individual cells by the current total Erlang value in the whole city. Given that the total urban population remains fairly stable during normal workdays, the fluctuations that we see in the city-wide sum of Erlang values show when people are more likely to make calls. Normalizing the raw Erlang values by the sums eliminates the city-wide calling behavior trends from the data and gives us a proportional measure ranging from zero to one, which indicates what percentage of all Erlang values originate from a particular cell, as shown in Equation 1. If we multiplied this proportion by the
total known population of the city, then we could obtain a rough estimate of how many people are located in each cell at every time period. If this estimate for the proportional distribution of people were further divided by the area of the network cell, we could obtain an approximation for the population density per unit of area.

**Equation 1 Estimating the city-wide proportion of call activity in each cell from Erlang data.**

\[ p_i = \left( \frac{e_i^t}{E} \right) \]  

where

- \( p \): proportion of city-wide calling activity in a cell
- \( i \): cell identifier
- \( e \): Erlang value of a cell
- \( E \): sum of Erlang values in all cells in the city
- \( t \): time period

When the proportional Erlang values are depicted across the city during different hours of the day we obtain a picture of the relative changes in the urban population movements. Figure 3 illustrates the change in the proportional distribution of city-wide activity before and after the morning rush hour on an average business day morning. At 8AM we notice activity spreading across cells in the central core of the city, reaching outward towards the periphery. By 11AM, the proportion of activity in the peripheral areas has decreased, and the largest share of activity originates from the historic center of the city. Depicting such changes of call volume distribution at 15 minute intervals can reveal a gradual description of the population movements throughout a day.

**Figure 2 Proportional distribution of city-wide activity. The intensity of the shade in cells denotes the share of city-wide activity. Left: 8AM, Right: 11AM on an average business day.**

However, the population distribution estimates obtained from raw Erlang measures rest on several assumptions. First, the raw data needs to be normalized as shown in Equation.
1 in order to make the measurements obtained at different hours of the day comparable.

Second, even after the normalization, the estimates still assume that the average call characteristics during any given measurement period are uniform across all cells in the city. If this assumption is violated, then the Erlang measure of a particular cell can be inflated without any population movements, by simply having longer or more numerous calls per user in a particular cell. There is little doubt that the likelihood of making a call, and the average duration of a call vary between cells (e.g. a person in a train station might be more likely to make a call than a person in a park). This bias is hard to account for since knowing the likelihood of making a phone call in different areas of the city at the exact same time would require an extensive observational survey, or more detailed information from data providers. We hope to address the issue in more detail in future work. If we could obtain the information describing the total number of calls and individual location updates (the number of registered phones in each cell, regardless of engaging in calls or not) from service providers, then we could determine the average call characteristics in each cell during different hours of a day, and correct for the errors.

The third set of biases can raise from the particularities of the given operator’s network. On the technical side, it is not always clear whether the call activity handled by a particular cell describes the actual call volume in its respective area. Some cells can occasionally reach the peak of their capacity, in which case they forward the call handling to other adjacent, less occupied cells. It is also not clear what fraction of calls might not be handled by the nearest cell due to coverage shortages caused by geographic obstacles in the environment of the cell. Our tests comparing the nearest voronoi serving areas with the actual estimated serving areas by the operator, suggested that the voronoi areas of cells provide a rather good estimate for locating calls.

From a sampling standpoint it is also questionable whether measuring the network usage of Telecom Italia’s customers’ would necessarily approximate all inhabitants of Rome. Despite the huge sample size (over one million customers in Rome), we do not know whether the customers are uniformly spread through all social, economic, gender and age groups of the population. Such information is unfortunately highly proprietary amongst service providers. In an attempt to find substantive evidence for possible segmentation amongst TIM’s customers, we compared the available services plans of the three competing operators in Rome, and interviewed a small sample of users in Rome. We concluded that TIM’s services are targeted to attract all socio-economic and age groups of customers.

Despite these shortcomings, the data does exhibit potential for broad and frequently sampled analysis of the urban population distribution. There is little doubt that more useful and more accurate data from cell phone networks will come available for planning analysis in the years to come, but at present, Erlang measures are readily available on most mobile phone networks and offer standard measurements on different hardware platforms. Unlike individual tracking data, Erlang measurements do not present a threat on individual privacy and are therefore generally easily obtainable. In the following we present a case study using Erlang estimates in the city of Rome, Italy.
3 The daily regularity in the population distribution of Rome

Using the estimates of the call volume distribution over a long period of time, we can use multiple tests to analyze whether there is regularity in the amount of activity that a cell contains at different times of the day or week. For example, does the cell around the Pantheon contain a predictable amount of activity at 10am, 11am, 12am and so on, on all Mondays within the sample of our data? Or are the observed standard deviations too large to exhibit any periodic qualities in the cell’s activity at these times? What other periodic cycles might the urban population distribution follow? In this section we address these questions using several different methods.

Whether regularity is present or not can be found by dividing the cell’s observed standard deviation by its observed average Erlang value for a given time period. The resulting ratio gives an idea of the hours when the activity in a cell follows a generally repetitive pattern. If the ratio is smaller than one, then the deviations are smaller than average Erlang values, and some regularity is found in the cell’s activity. Such a calculation is shown during a 24 hour cycle for seven days of a week in a randomly selected cell in Figure.

Figure 3 Std. Deviation Erlang / Avg. Erlang in a randomly selected cell during each day of the week.

We see that in this particular cell standard deviation appears to be lower than average Erlang values between 8.00am - 11.00pm on business days, and 9.00am – 10.00pm on weekends. Certain regularity in this cell’s activity is thus observable during these hours of the day. We found it typical to encounter higher deviations during time periods with low call traffic, and lower deviations during time periods with high call traffic. Standard deviations are highest, up to twice the average Erlang values, at around 3.00-4.00am.
However, even during daytime hours, the ratio in this particular cell remains between 0.85 and 1, meaning that the deviations constitute $\frac{6}{7}$th the size of average Erlang values. Regularity in the activity of this cell is thus weak even during daytime hours. The low regularity can be partially explained by the fact that our sample data ranged from August, the month of almost complete vacation in Rome, through September and October to November, which are more typical working months of the year.

Figure 4 illustrates the same analysis on all of the 389 cells that we studied. It shows that the values falling below the number one line (indicated as dotted) occur between 8am and 11pm on all days. It is during this time period that the Erlang values in cells are most predictable. However, the general regularity of daily population distribution differs across cells. Figure 5 illustrates the standard errors between raw values and average values of activity share in cells. Larger standard errors, denoted by darker tones, represent cells whose activity share is generally less predictable and more variables across a sample of 7 weeks. Since the activity shows regularity during daytime hours, it is important to ask how we might explain the observed distributions that repeat from day to day. We shall describe and attempt to explain some of these patterns in the second half of the paper. Explaining why the regularity in activity distribution varies in different parts of the city is a challenge we have to address in another paper.

**Figure 4 Weekly plot of std. deviation / average Erlang across 389 cells in Rome**

**Figure 5 Standard errors between the raw values and average values of activity share in cells. Between 8am and 11pm. Larger standard errors, denoted by darker tones, represent cells whose activity share is generally less predictable and more variables across a sample of 7 weeks.**
An important question from city planning and transportation planning point of view is *How does the population distribution in a city vary from day to day and week to week?* Hanson & Huff 1988 researched the question using individual travel surveys and concluded that behavior does not follow a weekly cycle closely enough for a one-week travel record to assess the level of day-to-day variation present in the individual's record. Schlich & Axhausen 2003 studied different travel-diary based methods to measure similarity of travel behavior and found that there is more consistency in individual time budgets for travel, than trip characteristics. They also concluded that travel behavior is more stable on work days. Mobile network activity data provides a unique opportunity to compare these results with the overall distribution of activity in the city as a whole on different days.

We selected a 7 week period in fall 2006, where our data had least gaps, and explored how much consistency there is in the activity distribution in each cell. We only looked at the distribution patterns between 8am and 11pm due to the larger deviation at other times, as explained above. We first derived the overall average share of activity in each cell at every hour of the day through every day of the week, and then compared how much the distribution patterns on specific days (49 days in our sample) differed from the average. Weekends were separated from weekdays.

**Figure 6** Observed correlations between activity shares in cells at different hours of a day on specific days (left) and average days (right). The two top row comparisons illustrate weekdays, and the bottom row weekends (N= 65,535)
The findings in Figure 6 suggest that there is great regularity in the activity distribution of the city on both weekdays and weekends. The raw data of individual business days indicates that 89.97% of the activity distribution from 8am to 11pm is explained by the distribution pattern of an average business day (t=1356.75, p<0.001). On weekends, 90.43% of the variation in activity patterns of individual days is explained by the distribution pattern on an average weekend day (t=705.05, p<0.001). This suggests that despite possible variations in daily trips of individuals, the urban areas covered by network cells contain a highly consistent share of people at different hours on all business days, and a different but equally consistent share during weekend days. The difference in the share of activity that an average cell contains on an average business day and an average weekend day is relatively minor. The correlation between the two is 83.20% (t=170.9, p<0.001), which attests to the dense and mixed land-use neighborhoods of Rome.

The correlation between the city-wide activity distribution on specific weekdays over different weeks (i.e. activity distribution from 8am to 11pm on Mondays, Tuesdays etc.) is even stronger. 91.13% of the variation in a specific business day (i.e. Monday) is explained with the distribution patterns of the average Monday (t=1461.54, p<0.001). Specific Saturdays and Sundays in the data have a 92.44% correlation with the average Saturday and Sunday respectively (t=801.68.05, p<0.001).

It is interesting to note that the regularity of distribution is equally strong on both business days and weekends. Schlich & Axhausen 2003 found that, based on individual commuting patterns, trips on weekends are less regular than on business days. Our data suggests that this difference does not manifest itself in the city-wide distribution of activity. However, the overall consistency of the distributions across the city in a seven week sample does not suggest that the same level of consistency is observed in individual cells. The scatter above and below the hypothetical regression line in the plots of Figure 6, shows that during some measurements, individual cells can deviate considerably from the average. Schlich & Axhausen also noted that measured similarity between day-to-day patterns declines if the method captures more of the complexity. The measure we have been using- the city-wide consistency of population distribution in cells- is obviously a very general one, and captures very little complexity of individual’s travel patterns. It is thus expected that using measures as general as ours, a higher consistency would appear. However, the general high consistency of daily activity distribution across the city implies that the collective movements that rise from a multitude of individual behaviors are more consistent than those of individuals.

Before proceeding to the description of the daily activity patterns we can also explore the nature of the periodic cycles themselves. If there is regularity in the activity patterns of network cells, what are the periods during which cells’ activity cycles typically repeat? We saw above that the Erlang values are more stable during daylight hour. In studying a longer continuous dataset, we can also see other characteristic cycles in the data.
In order to analyze all possible periodic qualities in the sample data we used Fast Fourier Transformations (FFT) on the seven-week data. FFT breaks the average weekly signal of each cell into constituent sinusoidal functions with specific frequencies and amplitudes, and distinguishes the magnitudes of multiple independent periods from the same signal as show in Figure 7. The dataset we used for FFT was proportional as above, where the individual cell’s Erlang values were divided by the city-wide Erlang sum at the time of each measurement resulting in a \( p \) value for each cell (Equation 1). This was necessary in order to distinguish the population distribution cycles from the calling behavior cycles that characterize the city as a whole. The analysis thus distinguishes the periods when the proportion of city-wide activity in cells’ repeats and eliminates the city-wide cycles that characterize people’s likelihood to make calls.

**Figure 7 FFT analysis. Magnitudes and frequencies of proportional activity \( p \) in a typical week across 389 cells in Rome.**

The FFT analysis revealed that the largest magnitude periodic quality in a week is a twenty four hour cycle, which indicates that the proportional activity in an average cell reiterates from day to day. The importance of this cycle suggests that the distribution of activities in a city does indeed follow a daily routine. There are great differences in the magnitudes of the 24 hour cycle between different cells, depending on how much the daytime and nighttime occupancy of an area differ. The cells with the highest amplitudes in a 24 hour cycle represent areas with highest differences between the maximum and minimum proportions of city-wide activity during measurement periods. In other words, these cells accommodate a relatively large share of all call activity in a city at some points over 24 hours, and a relatively low proportion at others.

The next highest amplitude after the 24 hour cycle signifies the three and a half day period, distinguishing weekends from weekdays. The activity distribution clearly differs between business days and weekends, which intuitively illustrates the lack of home-work trips and the presence of alternative travel patterns during weekends. This broadly confirms earlier findings in regular differences between work day and weekend travel patterns (Vilhelmson, 1999).

The 12-hour and 8-hour cycles, which FFT also clearly distinguished in Figure 7, intuitively resemble the night/day, and working/non-working hours of the city when the
population distribution reorganizes into predictable patterns at home, work or other daily activities. Compared to earlier OD survey based studies, these cycles present a more detailed view of mobility patterns within a single day. The 8-hour cycles signify that not only is the population distribution of Rome similar from day to day, but there is also a resemblance between the population distribution during work hours, and before or after work hours. The overall importance of the 8-hour cycles attests to the mixed land-use pattern Rome, where jobs and residences often occupy the same neighborhoods.

The activity cycles reflected in the FFT analysis can also be used to estimate the time periods when the activity distribution is experiencing most change. Calculating the average percentage change in activity across all cells in Rome can in this way approximate the rush hours of the city. Figure 8 illustrates the percent change of activity on weekdays and weekends. On weekdays the largest change in the city wide calling activity distribution occurs between 8am-10am and 5pm-8pm, which intuitively well match the time periods for traffic rush hours. On weekends the trend is less steep in the morning, but features a spike around 11am, which could represent the church going and other leisurely activities of Saturday and Sunday mornings. Unlike weekdays, when the distribution change falls to a steady rate around 10am, mobility remains in greater flux on weekend mornings, slowing down by approximately 2pm. It is interesting to note that the minimum average percentage change in the activity of a cell is around 6% per hour, which in the case of an average cell with a population of 8100 residents and an area of 0.93km², would denote a change of approximately 523 people per square kilometer in an hour. At the maximum city-wide average rate of change (16%) the estimate would suggest a difference of 1296 people square kilometer in an hour, suggesting that during rush hours we should experience up to 2.5 times the normal activity on streets.

Figure 8 Percent change in cell activity between 8am and 8pm on business days and weekends.

![Figure 8 Percent change in cell activity between 8am and 8pm on business days and weekends.](image)

We mentioned in section two that the estimates obtained from the network activity measures are still subject to the biases of non-uniform calling likelihood amongst different cells. An extreme case is presented by the cell containing the Parliament of Italy and several other federal offices on Piazza del Parlamento. Accommodating a very high concentration of politicians and administrative workers, the network activity originating
from this cell constituted 6.06% of the city’s total at 2 pm on Wednesdays. It is clear that this estimate does not represent a realistic projection for the population in that area: if multiplied by the city’s population of 2.7 million, it would mean that the cell should contain roughly 160 000 people. However, as the cell area is approximately 0.03 km², then the population density of the cell should be 4.98 million people per square kilometer. Such a bias is probably caused by the extremely frequent calling behavior amongst the politicians and employees who constitute the daytime population of the area.

4 Understanding the activity patterns of places

In the previous section we saw that some cells in Rome exhibit high magnitude cycles of activity, while others accommodate a more stable activity pattern throughout a typical day. In this section we ask how might we explain the differences or similarities between the activity patterns of cells?

Some of the observed patterns seem intuitive. Figure 9 illustrates the range between the minimum and maximum share of city-wide activity that cells accommodate in a typical daily cycle. The cells with the highest range contain a very high proportion of the urban population at some hours of the day, and a relatively low proportion at others. Figure 8 shows that the highest activity range is typically located in the heart of the historic city, which fills with tourists, and white collar workers at daytime, and houses few at night. Also cells that contained large office or industrial complexes, a major transit station, or uniquely high density housing, which fill at night and empty at daytime, or vice versa, stood out with high ranges in activity cycles. For instance, the single high range cell in the south of Rome is located in the center of the EUR office district. The lowest magnitude cycles, on the other hand, represent cells with relatively steady activity shares throughout a day. Cells with the lowest-range cycles were typically located in parks, unoccupied land, or major thoroughfares. These intuitive explanations need yet to be verified statistically.

Figure 9 The daily range of proportional activity in each cell between 8am and 11pm. The darkest tones represent cells that accommodate the highest range between the minimum and maximum proportion of city wide activity.
In order to explore a substantive explanation to the varying activity patterns between cells, we tested whether areas that share similar activity patterns also share similar socio-economic properties that might explain their activity trends.

We obtained the average weekly call activity log for each of the 389 cells from a 24-week dataset. In order to avoid the activities of any particular week dominating, we determined a ‘typical’ pattern for each day of the week (average signal for Monday, Tuesday etc.) The resulting data thus showed the average Erlang values in each cell in a seven day period at 15 minute intervals (672 counts per each cell).

Using principal component analysis on the 672 discrete measurements, we found that the dominant weekly trends in calling activity could be captured by the first three eigenvectors (a detailed explanation of eigen-decomposition can be found in Eagle & Pentland 2007, and Reades, Calabrese & Ratti 2007). Using three eigenvectors, we performed an average distance cluster analysis on all 389 cells to determine which areas had similar activity patterns over a week. The clustering fit test suggested that the most coherent structure was formed using 16 clusters. Many of the clusters in this solution contained only one or a few antennas, which was expected since cluster analysis is known for efficiently distinguishing outliers. We selected the three largest clusters that covered most of the cells for comparison with the socio-economic data in each cell.

In analyzing the locations of these first three Erlang clusters, a distinctively concentric geographic regularity emerged. Cells belonging to the first cluster (most numerous) were predominantly peripheral, cells in the second cluster surrounded the historic center, and cells in the third cluster were mostly located in the central core of the city. Figure 10 illustrates the standardized proportional activity shares in the three largest clusters from 8am to 11pm during business days and weekends. We can see that activity distributions
in the three clusters differ remarkably. The distinct behavior of each of the three clusters suggests a presence of particular activity types in the cells.

**Figure 10** Erlang signatures for clusters 1, 2 and 3 on an average business day and weekend (standardized values).

In an attempt to explain this grouping, we compared the distribution of demographic indicators and business establishments in each of the cells. The business distribution was obtained from the Yellow Pages data in Rome (it therefore does not include all businesses in the city, but only those that were listed). The residential distribution reflected the year 2000 census. The spatial spread of the city’s residential population and business locations are shown in Figure 11. Both categories were further broken down into specific age groups and business types, as shown in Table 1.

**Figure 11** Left: Residential population density. Right: Distribution of firms listed on Yellow Pages
We then performed a second cluster analysis of the same cells based on these demographic and business indicators in order to test whether the Erlang cluster solution would match with the clustering of demographic and business indicators. We conducted a principal component analysis and used the first three eigenvectors to cluster cells with similar demographic and business distributions together. The fit test showed the most coherent solution was found with 6 clusters, from which only three contained more than 5 cells each. We thus used these three largest clusters as a reference for the comparison with the three largest Erlang clusters described above.

A comparison between the Erlang and business/demographic clusters is compelling. Clusters with a similar demographic and business distribution correspond quite well with clusters with similar activity patterns. 63.14% of the cells in the three Erlang clusters match the same order demographic and business clusters, as shown in Figure 12. Comparing demographics and businesses separately with Erlang clusters, we found that Erlang clusters match better with the business distribution (60.05%) than with the population distribution (51.29%).

Table 1 describes the typical census and business characteristics of the three business/demographic clusters. The average number of residents in clusters 1, 2 and 3 is 3632.81, 9760.74 and 2969.4 respectively, and the number of businesses, 76.20; 227.61 and 292.55 respectively.

Cells in cluster 1 have the least amount of businesses but more residents than cluster 3. These seem to be the relatively lower density residential areas of Rome, which well matches the activity pattern of the first Erlang cluster as illustrated in Figure 10. These
cells contain a large proportion of night time activity on business days and low day time activity. During weekends, the day time activity remains above the mean till noon, which would be expected for residential areas.

Cells in cluster 2, have more residents than the other two clusters, less overall businesses than cluster 3, but more transportation, food, health, beauty and daily retail establishments. Cluster 2 thus seems to represent dense residential areas that contain local commerce. These indicators also corroborate the activity trend that we found in the second Erlang cluster in Figure 10. The cells accommodate a relatively high share of activity during weekday nights, and a very low share during business hours. During the weekends, however, these cells retain a large share of activity throughout the whole day, which might be explained by the larger presence of service and retail establishments.

Cells in cluster 3 have the highest amount of businesses, most notably hotels, restaurants, business services, government services and specialized retail, but the least amount of residents. This suggests that cluster 3 characterizes dense CBD neighborhoods where few people live. The corresponding Erlang signature in Figure 10 suggests that cells in cluster three have an almost opposite activity patterns than the first two clusters on weekdays: they contain very low activity at night, and a maximum during working hours before noon. Their activity drains out again between three and seven in the evening, and returns above the mean during late night hours, which seems to attest to their portion of business, entertainment and hotel establishments. During weekends, cluster three cells retain a high amount of activity throughout the night, dropping very low after four AM, which also seems to attest to their density of entertainment and accommodation facilities. A lower weekend peak occurs again between 8am and 11am, possibly indicating church or specialized retail visitors, after which the activity disappears again till the evening.

Figure 12 Comparison of business & demographic data clustering (left) VS Erlang data clustering (right). Correlation: 63.14%
The comparison between cells with similar proportional activity patterns and their corresponding demographic and business indicators shows that the activity distribution of different urban areas, approximated through cell phone network usage, is indeed related to the socio-economic structure of the city. The characteristic business types present in each of the clusters match the intuitive activity trends observed in Erlang analysis. The clustering analysis illustrates that 63% of the cells, which share similar activity patterns, also share a similar residential demographic and business mix. Figure 10 distinguished the characteristic activity distribution patterns of three general types of urban areas with a different land uses. In future research, it would be interesting to analyze in greater detail what kind of activity patterns result from a particular mix of businesses and land uses. Such analysis could prove highly valuable for transportation researchers, traffic engineers, infrastructure planners, as well as mobility service providers.

5 Conclusion

Using call volume data from a wireless phone network in Rome, Italy, we explored two interrelated questions about the daily population distribution of a city: Does the distribution of the urban population, estimated through call volume, of Rome follow routine patterns? And if so, how could we explain the variation of such patterns in different parts of the city?

Unlike most studies on mobility patterns, which analyze mobility patterns via surveying individuals, we focused on the behavior of places rather than individuals. We used network cells as the spatial units of comparison of activity patterns in different parts of the city. Our findings were the following:

- There is very high regularity in the city-wide population distribution between 8am and 11pm on all days of a week. Since our estimates rely upon mobile phone call volumes, which are very low during night time, we were not able to confirm the regularity of activity distributions at night. The analysis of a large sample of aggregate mobile network data (approximately 1 million users) thus confirms earlier research findings from individual surveys, suggesting that the regularity of individual travels is also manifested in the regularity of the population distribution in particular areas, as well as the city as a whole.

- Unlike previous studies that have examined individual travel behavior and found large differences between weekday and weekend travels, our findings suggested that from an aggregate city-wide perspective, population distribution patterns in the city are approximately equally regular on weekdays and weekends. Our findings suggest that the collective movements that rise from a multitude of individual travel patterns in a city are more consistent than those of individuals.
- Whereas previous studies of individuals have generally found more consistency in personal travel-time budgets than spatial trip characteristics, our findings emphasize the regularity in spatial distribution rather than temporal distribution.

- Using signal processing and Fast Fourier Transformations (FFT) on a seven week dataset, we found that there are more important periodic cycles than only the daily and weekday/weekend cycles in the urban population distribution. The importance of the 8-hour and 12-hour cycles in our data pointed to similarities between the population distribution during work hours, and before or after work hours. The significance of the 8-hour cycles might be specific to our case-study: the city of Rome with a mixed land-use pattern, where jobs and residences often occupy the same neighborhoods.

- Using principal component and cluster analysis we found a strong correlation between areas that resemble in network usage and areas that resemble in demographic and business composition (63%). The specific mix of business establishments was generally a better predictor of activity in a cell than the residential demographics of the cell. This result confirms the existence of a strong relationship between daily mobility pattern of an area, and its land use.

The general consistency of our findings with previous scholarship suggests that, despite significant limitations and biases, aggregate mobile network usage data contains plentiful potential for analyzing urban mobility patterns. Future research is needed to explore the discrepancy with earlier individual based studies that have found that travel patterns are less regular on weekends. An additional inclusion of employment distribution, other socio-economic data, as well as measures of urban form in the land use indicators would also allow future research to better determine which characteristics of cell locations are most important for shaping the Erlang signatures of particular areas.
6 References


7 Acknowledgements

The ideas presented in this paper are a result of a collaborative effort by several researchers at the SENSEable City Lab at MIT. In particular, the work by Francesco Calabrese, Jon Reades and Mathew Jull has contributed toward the development of the concepts and methods presented above. We are greatly thankful for their intellectual support. We are also particularly grateful to the MIT Portugal program, under whose support and encouragement the research was effectuated.