A Tale of One City: Using Cellular Network Data for Urban Planning

Richard A. Becker Ramón Cáceres Karrie Hanson Ji Meng Loh
Simon Urbanek Alexander Varshavsky Chris Volinsky

1. INTRODUCTION

With the continuing urbanization of the world's population and the rapid growth of cities, urban planners are faced with many challenges, including heavily congested roads, overzealous development, and increasing pollution. To efficiently address these problems, urban planners need to develop a better understanding of the dynamics of modern cities. This includes studying the flow of people into and out of cities as well as the use of the commercial and residential parts of a given city. Understanding these issues is key for urban planners to build the cities of tomorrow while also satisfying increased resource and environmental constraints.

Cellular network data has the potential to revolutionize the study of city dynamics. Such study is a time-consuming and expensive process, entailing techniques such as surveys and vehicle counting. Large-scale commuting studies can take years to complete, and many municipalities learn about new trends only infrequently when a detailed census is released. In contrast, cellular networks must know the approximate locations of all active cellular phones in order to provide them with communication services. Given the ubiquity of these phones and their almost constant proximity to their owners, cellular networks can be used to opportunistically sense the locations of large populations of people. They thus provide a means to monitor city dynamics frequently, cheaply, and at an unprecedented scale.

In this paper, we explore the use of anonymized Call Detail Records (CDRs) to capture city dynamics. CDRs document the location of the wireless cell carrying every voice call and Short Messaging Service (SMS) transaction, as well as the time when the transaction occurs. Because CDRs are routinely collected by all service providers for operational, planning, and billing purposes, the incremental cost and resources required to analyze this data are minimal. However, CDRs have several limitations. First, they are sparse in time because they are generated only when a transaction occurs, rendering cellphone users invisible at all other times. Second, they are coarse in space because they record locations at the granularity of a cell tower sector, giving an uncertainty on the order of one square mile for each transaction. Nonetheless, the convenience and prevalence of CDR collection makes it worthwhile to investigate how much can be learned from these records.

In this paper, we present several ways in which CDRs can be used to provide important information about city dynamics to urban planners. We analyze 2 months of cellular traffic in and around Morristown, New Jersey (NJ), a suburban city in the United States (US) with approximately 20,000 residents. We demonstrate the feasibility of our approach through tabulation, statistical analysis, and visualization. Specifically, we make the following contributions:

- We demonstrate how CDRs can be used to determine which residential areas in and around a city contribute daytime workers to that city, i.e., the city's *laborshed*. Furthermore, we validate our methodology by comparing the laborshed derived from CDRs to that derived from the US Census, and show the strong similarity between the two. (Section 3)
- We show how CDRs can be used to determine which residential areas contribute late-night revelers to a city, i.e., the city's *partyshed*. (Section 4)
- We show how CDRs can be used to capture the lifebeat of a city through *lip plots*, a novel visualization of cellphone activity in the individual wireless cells covering that city. (Section 5)
- We demonstrate that it is possible to group city dwellers and visitors into categories based on cellphone usage patterns, through a novel application of an unsupervised clustering algorithm. We further show that combining this clustering analysis with contour maps reveals interesting aspects of city dynamics that could be of great value to urban planners. (Section 6)

2. DATASET

This section presents the methods we used to collect, anonymize, and analyze cellular network data, along with general characteristics of our dataset. It also describes the steps we took to preserve the privacy of individuals.

2.1 Methodology and Characteristics

We collected anonymized Call Detail Records from the cellular network of a large US communications service provider.

We captured transactions carried by the 35 cell towers located within 5 miles of the center of Morristown, NJ, a suburban city in the greater New York City metropolitan area. These 35 cell towers house approximately 300 antennas pointed in various directions and supporting various radio technologies and frequencies. Our goal was to capture cellular traffic in and around the town and choosing the 5-mile radius allowed us to cover both Morristown proper and its neighboring areas.

The data was collected and anonymized by a party not involved in the data analysis. In place of the phone number of the person involved in a transaction, each CDR contains an anonymous identifier consisting of the 5-digit billing zip code and a unique integer. Each CDR also contains the starting time of the voice or SMS event, the duration of the event, and the locations and azimuths of the antennas of cell tower antennas associated with the event.

We collected voice and SMS traffic for 60 days between November 29, 2009 and January 27, 2010. In total, we collected 15 million voice CDRs and 26 million SMS CDRs for 475,000 unique phones.

2.2 Privacy Measures

Given the sensitivity of the data, we took several steps to ensure the privacy of individuals. First, only anonymous records were used in this study. In particular, personally identifying characteristics were removed from our CDRs. CDRs for the same phone are linked using an anonymous unique identifier, rather than a telephone number. No demographic data is linked to any cellphone user or CDR.

Second, all our results are presented as aggregates. That is, no individual anonymous identifier was singled out for the study. By observing and reporting only on the aggregates, we protect the privacy of individuals.

Finally, each CDR only included location information for the cellular towers with which a phone was associated during a voice call or at the time of a text message. The phones were effectively invisible to us aside from these events. In addition, we could estimate the phone locations only to the granularity of the cell tower antenna coverage area. Although the effective radius of an antenna depends upon tower height, radio power and terrain, a given antenna on a cell tower has an uncertainty of about 1 square mile [10].

3. CALCULATING AND VALIDATING THE LABORSHED

Understanding the daily flow of people in and out of a city is important for urban planning. In particular, for a city with a commercial district, understanding where workers live can help manage vehicular traffic flow and plan public transportation services. For example, Morristown is a regional center of commerce and shopping, with a developed downtown area, many office complexes, a large hospital, and the county courthouse. It draws a large worker base from the nearby suburbs and even some from the much larger New York City. The geographical area representing where a city's workers live is known as its *laborshed*.

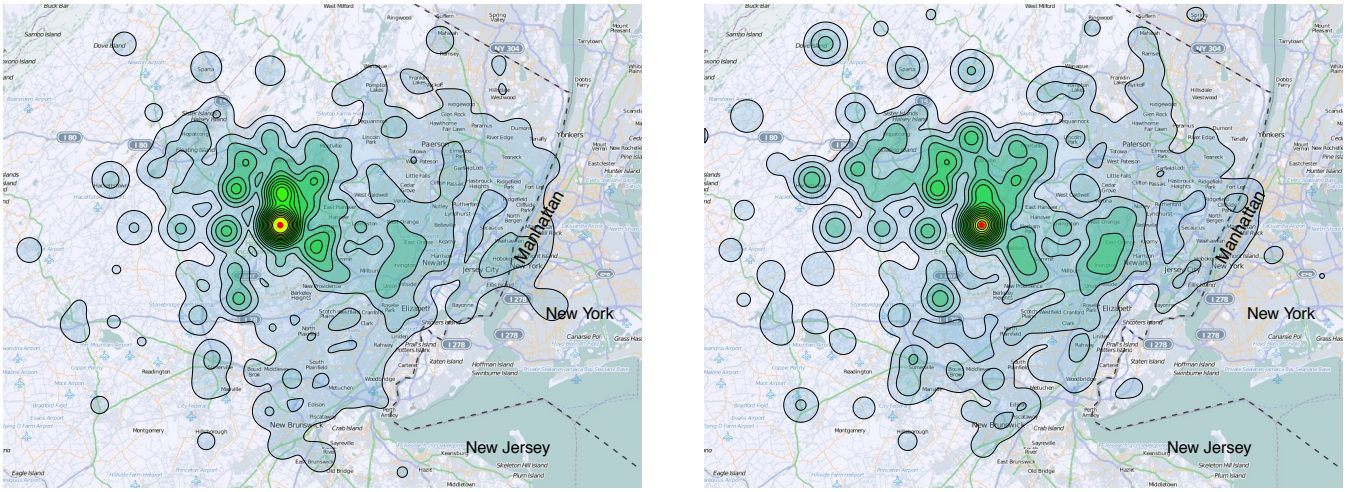
We used our CDR data to calculate the laborshed for Morristown. We classified a cellphone user as a worker if she is frequently observed in Morristown during business hours (9am to 5pm, Monday to Friday). More specifically, a worker needs to satisfy two conditions. First, she needs to engage in an average of at least 4 calls/messages per week during business hours, involving one of the Morristown cell towers. Second, she needs to make those calls/messages on an average of at least 2 unique weekdays per week. We derived these thresholds experimentally and observed that moderate changes to these values did not affect our results. We then used the billing ZIP codes of these users to obtain the geographical distribution of their residences.

We validate our CDR-based laborshed results by comparing them to publicly available US Census data. Specifically, we used the 2000 Census Transportation Planning Package (CTPP) [1], which contains information about people's places of residence and work. We mapped census tracts to ZIP codes by calculating the proportion of a census tract that fell within each ZIP code of interest.

In contrast to the census definition of a worker, ours may include some residents of Morristown who use their cellphones there during business hours but are not employed, such as students. We corrected this overcount in two steps. First, we estimated a correction factor by using four neighboring ZIP codes and calculating the ratio of workers in these towns, computed from the CDRs, to the number of workers given by the census data. We found this ratio to be 1.5. Second, we divided the CDR-based worker numbers for Morristown ZIP code by this factor.

Figure 1 shows contour maps of the Morristown laborshed as calculated from the corrected CDR data (a) and from the census data (b). We do not expect the two maps to be identical, because our CDR records only show those who are actively generating calls or SMS records, and reflect only the activity of the part of the population on our company's network. Additionally, the census data we had access to is a decade old, and commuting patterns might have changed significantly in that time. However, we expect the maps to be similar if our methodology is sound. Indeed, we see that the geographic distributions are quite similar, especially for the regions close to Morristown, but also for the area near Newark, roughly midway between Morristown and New York City, and for the region to the south and east of Morristown. The CDR numbers are lower than the census numbers for the more distant northwestern region, and somewhat higher for New York City.

Studying the contour maps themselves reveals a high-density region centered directly over Morristown, indicating a large concentration of people who both live and work in Morristown. We also see that many more workers come from areas north of Morristown than south, and that there seems to be a



(a) Call Detail Records

(b) 2000 US Census

Figure 1: Morristown laborshed maps calculated from (a) CDR data and (b) US Census data. The two maps show similar patterns, indicating that CDR data provides a plausible estimate of where Morristown workers live. The red dots represent the center of Morristown.

cluster of workers to the east, in the more heavily populated areas close to New York City. Additionally, there are some pockets of workers who come from towns west of Morristown. This information could be useful to urban planners, for example to reduce traffic congestion by organizing new transit and bike routes or park-and-ride programs.

Another way to validate our laborshed results is to graph in a scatterplot the number of workers who live in each ZIP code as calculated from our CDR data against the corresponding numbers from the census data. Figure 2 shows the result, with the two axes on a logarithmic scale. With perfect agreement (up to a multiplicative factor), the points would fall on a straight line. For comparison, we show the $y = x$ equality line as a dotted line, and the best linear fit, $y = .387x$ as a solid line. The correlation coefficient is 0.81. Again, we do not expect perfect agreement, but there is a clear correspondence between the CDR- and census-based numbers, and if we want to roughly estimate numbers of people, we might multiply the CDR numbers by $1/.387$.

To summarize, we argued that understanding where workers live is important for urban planning and showed that the laborshed generated from CDRs matches closely that obtained from census data. These results give confidence in the validity of our approach. Our approach has significant advantages over the census because its low cost makes it practical to generate laborshed results much more frequently, for example every few months instead of every ten years.

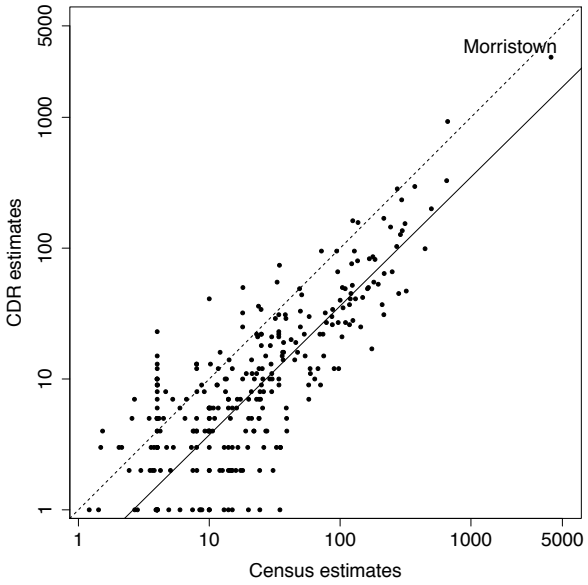


Figure 2: Scatterplot showing agreement between Morristown laborshed numbers from CDR data and US Census data. Each point represents one ZIP code. The solid line shows the best linear fit, where the CDR count equals 0.387 of the Census count. If the CDR estimates exactly matched the Census numbers, the points would fall on the dotted line.

4. CALCULATING THE PARTYSHED

We can apply techniques similar to those described in the previous section to other groups of people besides daytime workers. As one example, in this section we focus on people who are active late at night. Like many cities, Morristown

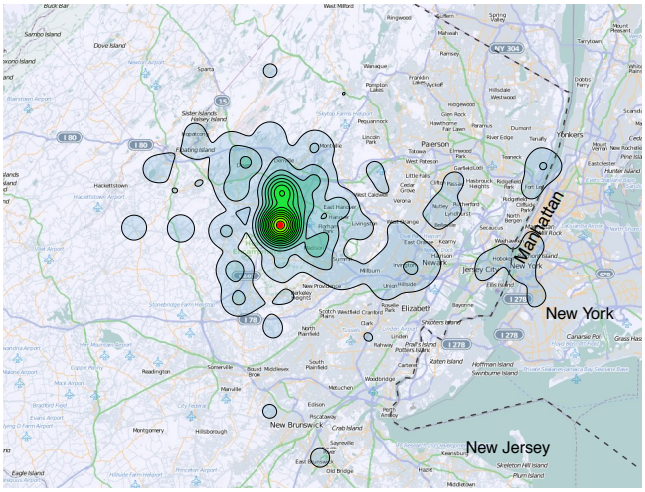


Figure 3: Morristown partyshed map showing the home locations of people who used their cellphones during weekend late nights in downtown Morristown. Comparing to the earlier laborshed maps, partiers' homes are concentrated in areas closer to Morristown than workers' homes.

has a lively bar and restaurant scene that attracts people from other communities as well as locals. In analogy with our earlier laborshed calculations, we refer to the geographical distribution of where this group lives as the *partyshed*.

We identify the partyshed cohort using CDRs by setting different criteria and thresholds than the ones we used for the laborshed. We look for cellphone users who have voice call or text messaging activity on late weekend nights (10pm to 3am, Fridays to Sundays). Again, we apply the constant correction of 1.5 described in Section 3. The resulting partyshed is shown in Figure 3. It appears that the distribution of partiers is considerably more concentrated in and near Morristown than the distribution of workers. Nonetheless, there is still some representation of people who live far away, even as far as New York City. Knowing where groups of revelers come from and where they return at the end of the night could allow towns to tailor services such as late-night shuttle buses intended to keep inebriated drivers off the road.

5. CAPTURING THE LIFEBEAT OF A CITY

In this section, we demonstrate that it is possible to identify patterns of human activity in different parts of a city by observing cell phone usage in different cell tower antenna coverage regions. Studying these patterns may allow city officials to model the typical flow of people between different parts of the city over time. Monitoring these patterns may in turn allow the timely detection of anomalies such as dangerous overcrowding surrounding a popular music concert. We refer to these patterns as the *lifebeat* of a city.

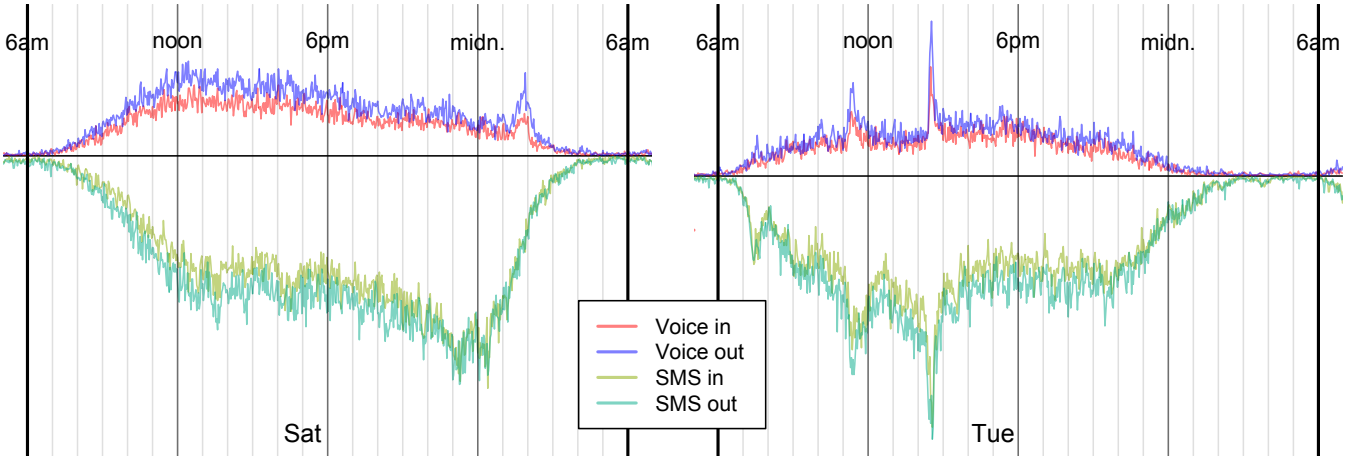
To analyze data from multiple cell tower antennas simultaneously, we developed a novel visual display capable of representing the multivariate nature of the data. We first aggregate the underlying data in two-minute intervals and then aggregate by the day of the week, which allows the study of day-specific patterns. In total, we end up with 720 two-minute bins for each day of the week for both voice calls and SMS messages. We then use the principle of small multiples [11] to display the data for all combinations of the partitioned variables.

Before illustrating how viewing the behavior of multiple antennas simultaneously can provide powerful insights, we first present plots for two specific antennas in Morristown in detail. Figure 4 shows usage plots captured on two different days of the week for two antennas located on the same cell tower but pointing in different directions. The x-axis represents time, starting and ending at 6am. The height of the plot shows the amount of traffic: height above the axis represents voice call volume, while height below the axis represents SMS volume. By using these opposite directions of the axes we avoid overplotting and retain the ability to recognize shapes at a glance. In addition, our visual cognitive system is good at evaluating symmetry quickly, thus we can quickly asses whether voice usage strongly deviates from SMS usage. For both types of traffic, color is used to distinguish inbound vs. outbound traffic. The resulting shape of the plot when used for within-day activity resembles lips and hence we call this type of visualization *lip plots*.

The patterns of these two plots are strikingly different. Figure 4(a) shows data from a Saturday in one part of town. The SMS traffic dominates the voice traffic, and the volumes keep rising throughout the day with maximal usage between 11 PM and 1 AM. The voice traffic, despite being dominated by the SMS traffic, has a noticable spike at 2AM. This is the cell tower antenna that points to the downtown area including the majority of the restaurants and bars in town. The spikes might represent late night revelers, and in particular the 2AM voice spike might reflect the fact that the bars close at 2AM, and patrons are looking for a ride home.

The plot in Figure 4(b) has quite a different pattern. This plot is from a weekday and also shows a majority of data from SMS traffic as opposed to voice. But here the majority of traffic is in the morning and in particular, we see spikes in SMS usage at 7AM, 11AM and 2PM. Only at 2PM do we see a similar spike in the voice traffic. This cell antenna points towards the town's high school, and could reflect the communication patterns of the students there, texting before and after school and during lunch. The larger 2PM spike at the end of the school day might reflect calls between students and parents, where voice channels would be more likely.

But how can one get a sense of all of the traffic in a given area in a quick and visually appealing manner? We devised a visualization that shows all of the data for all of the days and antennas by treating the cell tower, segment, and technology as partitioning variables in a large display. Figure 5 is a subset of such a display showing three antennas of the main cell tower downtown, but the complete display



(a) Downtown Antenna

(b) High School Antenna

Figure 4: *Lip plots* of voice call and SMS volumes show unusual spikes highlighting local patterns or events in Morristown. Call volume (plotted upwards; inbound: red; outbound: blue) and SMS volume (plotted downwards; inbound: light green; outbound: dark green) on two antennas are shown. The antenna in (a) points towards the commercial and restaurant district and the antenna in (b) points towards the high school. A voice peak occurs Saturday at 2AM when the bars close. Both voice and SMS peaks occur Tuesday when the school lets out.

showing all antennas of this tower is best viewed by printing it out on large-scale paper and hanging it on a wall (a high-resolution, zoomable version of this complete display is available at `http://bit.ly/BigLipPlot`). In these composite displays, the individual lip plots are laid out in a grid. Each row represents a tower antenna with a pictogram on the left annotating the particular combination: the size of the circle represents the frequency, the direction of the line represents the direction of the segment and its color represents the technology (red for 3G and blue for 2.5G).

We are interested in comparing relative behavior across segments, therefore each row of the composite plot is scaled individually to remove the influence of the volume per segment. However, it is important to recognize segments with a small volume as patterns for such segments will be inherently more noisy, therefore we have included a bar chart of the volume for each activity type on the right side of each row.

These composite plots show intricate patterns of communication and reward careful scrutiny. Here are some of the things we learn from looking at the larger plot:

- There is a heterogeneity in the patterns. Each row reflects a single direction so any directional patterns can be quickly seen. There is a lot of variability across directions, reflecting differing usage patterns in different parts of the city.
- There is a wide variance in the volume covered by the different directions, as shown in the bar plots on the right. Certain directions simply have more traffic than others.
- The relationship between SMS and voice changes by direction. The middle row appears to have much more SMS traffic relative to voice than the other rows. This is particularly apparent on the weekends.
- Small volume antennas have high variance, and often result in giving the lips a 'fuzzier' look.

6. IDENTIFYING USAGE PATTERNS

In this section, we show how cell phone activity can be used to group city dwellers into categories useful to urban planners. For example, do people who commute to Morristown stay after work to eat dinner and hang out, or do they head home as quickly as possible? Do those who live there head downtown after work to grab dinner? Are the users who shop on a weekend day the same as those that hang out in the bars on the weekend nights? Some of these questions can be addressed by clustering users into groups based on their cell phone usage profiles and studying these groups in more detail. We use an unsupervised clustering approach that has no prior assumptions on what these user profiles might look like. Our clustering algorithm identifies clusters of behavior and returns a typical member of that cluster represented by the mean behavior in that group.

In order to capture usage patterns, we aggregate voice and SMS usage separately into bins. Each bin represents a particular hour of the day and day of the week, giving us a total of 168 bins (24×7). We do not differentiate between incoming and outgoing events, as our analysis shows that there is a strong correlation between the two. A voice call contributes to a bin an amount proportional to the duration of the call falling into that bin (e.g. the call contributes a full unit to the bin if it spans the entire hour). For SMS, a bin contains

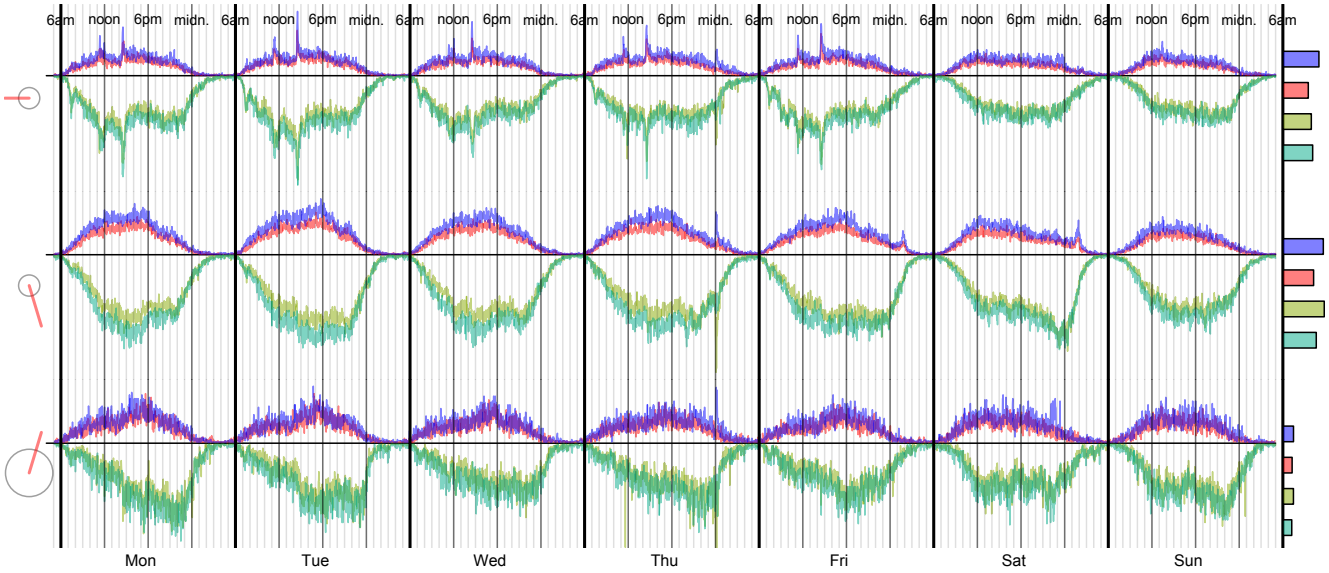


Figure 5: Composite lip plots of three different segments from a single cell tower for each day of the week. The lines with the circles at the left show compass directions of the antennas. The bars at the right show relative traffic volumes of in/out voice/SMS. Each antenna (row) has unique characteristics, e.g., the High School antenna (top row) shows the characteristic SMS spikes of school activity only on the weekdays.

the number of SMS events during a corresponding hour and day of the week. The result of the aggregation is two matrices, $Q = (q_{i,j})$ for voice and $P = (p_{i,j})$ for SMS, with seven columns corresponding to days of the week and 24 rows corresponding to hours of the day.

Finding stable clusters is a challenge in data sets like this due to the high variance of individual behavior. For low volume users, patterns might not present themselves clearly. If there are too many low volume users the clustering algorithm will wander too much, trying to fit signal to the noise. To combat this, we fit the clustering algorithm on a reduced, thresholded set of users, and then use the clusters found from this reduced set to assign a cluster label to every user. We use a threshold of 10 hours of voice traffic over the 60 day period (where an SMS counts as a call of 1.5 minutes), leaving us with approximately 26,000 accounts. Once the clusters are identified, each account is assigned a cluster based on the shortest Euclidean distance from the cluster means.

In order to make voice and SMS usage profiles comparable for the purpose of clustering we divide the bin counts by the global mean of each activity group. We use a k-means clustering algorithm with normalized usage vectors v_i consisting of the entries of both matrices P and Q: $v_i = (u_{i,1} \dots u_{i,336}) =$ $(p_{1,1} \dots p_{24,1}, p_{1,2} \dots p_{24,7}, q_{1,1} \dots q_{24,1}, q_{1,2} \dots q_{24,7})$ with $\forall i : \sum_j u_{i,j} = 1$ and a Euclidian distance measure. Despite the simplicity of the distance measure, which does not take into account the temporal relationship of the entries in the usage matrix, the resulting clusters are remarkably consistent and informative. The best result with respect to cluster size distribution is achieved for $k = 7$.

Figure 6 shows an image plot of the seven cluster means, each one representing a specific usage profile. Each profile consists of the voice usage matrix (left of the grey strip) and the SMS usage matrix (right of the grey strip). Each row of a matrix represents an hour, each column a day of week, starting with Monday, ending with Sunday. The color scale ranges from no usage (cream color) to most heavy usage (red). The number of the cluster represented by the profile is denoted in top left corner adjacent to a bar that shows the relative size of the cluster.

A few key usage types are evident, for example: cluster 2 consists of heavy voice users who have heaviest usage during business hours, are a bit less active on the weekend, and have little to no SMS usage; cluster 4 is an after-dinner voice caller; cluster 7 is a business-hour texter. Very few individual's profiles will look like these idealized cluster means, but we can calculate a distance from any single usage profile to these cluster means to determine that user's most likely cluster. In aggregate, this allows us to calculate the 'cluster mix' or the proportion of different types of user profiles that are present in the city at any given point of time. Additionally, we could track the clusters through time to see how the mix changes seasonally.

By combining our clustering results with contour maps, we can visualize and compare the geographical footprint of users belonging to different clusters. If the different clusters are really representing different segments of society, we might expect them to have a different geographical footprint. Urban planners might be interested in which of these clusters refer to groups of people that live in Morristown, as opposed to those that are coming to visit or are passing through.

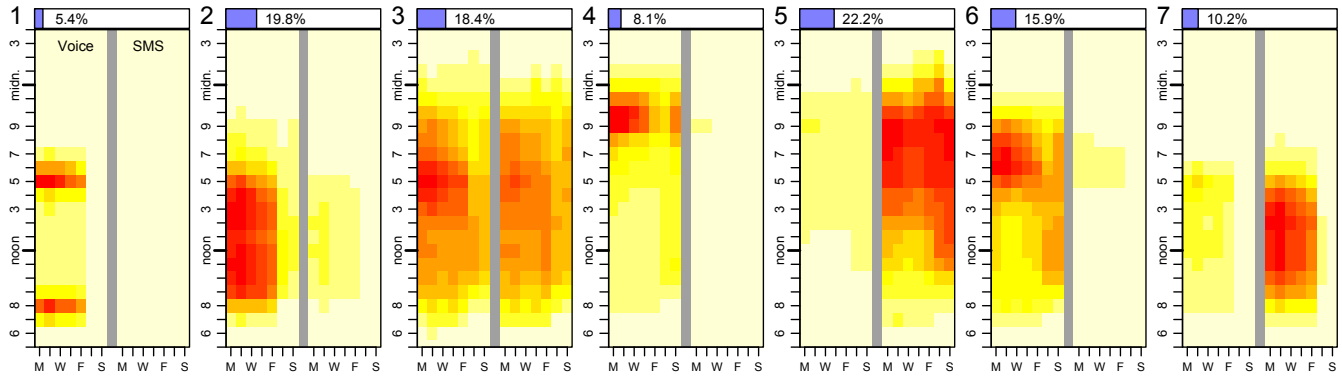


Figure 6: Seven cellphone usage patterns identified via clustering. Patterns emerge based on voice call and SMS volumes on different days of the week and hours of the day. Voice usage is shown on the left of the gray vertical bars, SMS usage on the right. Darker colors indicate higher volumes. The bar at the top shows the relative size of the cluster, the cluster number is on the top left. For example, cluster 1 shows only voice calls, just before and just after business hours. In contrast, cluster 7 shows primarily SMS usage during business hours.

Compare, for example, the usage profiles of clusters 5 and 7 in Figure 6. At first glance, these two profiles look fairly similar: both show heavy SMS usage with very little voice activity. The major difference is that cluster 5 shows usage on the weekends and at night, whereas cluster 7 is predominantly active during business hours. Figure 7 plots the geographical footprint of users belonging to these two clusters. Despite their similarity in usage, the two clusters have quite different geographical footprints. Cluster 7 is a geographically diverse group that by and large lives outside of Morristown, with significant clusters both to the east and west of town. This helps explain why we do not see much weekend activity for this profile - people who work in Morristown but live further away might not want to return to the town for their free weekend hours. Cluster 5 is very different - concentrated in and around Morristown, perhaps indicating students in the town. This example illustrates that combing clustering analysis with contour maps may reveal interesting aspects about city visitors.

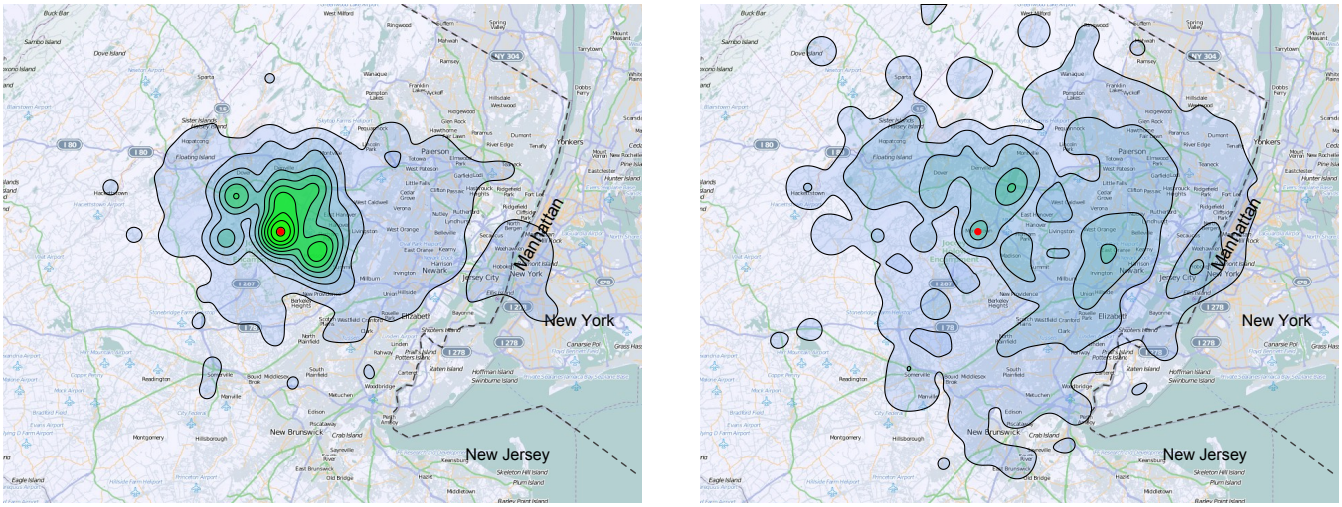
7. RELATED WORK

Several recent papers studied how cellular network data can be used for urban planning. In a case study in Milan, Italy, Ratti et al. [8] and later Pulselli et al. [7] demonstrated that it is possible to graphically represent the intensity of urban activities and their evolutions through space and time using call volume at cellular towers. Reades et al. [9] also looked at how call volume at a cell tower correlates with urban activities in the geographic vicinity of the tower. They studied call volume activity in six distinct locations in Rome, Italy and showed that the call volume varies drastically between the studied locations and between weekdays and weekends. They also proposed an algorithm for clustering together geographic locations that exhibit similarity in call volumes. Girardin et al. [3] used tagged photographs from Flickr in combination with the call volume data to determine the whereabouts of locals and tourists in Rome, Italy. They later repeated the study with only call volume data to examine the differences in behavior between tourists and locals in New York City [4]. Calabrese et al. [2] studied where people are coming from to attend special events in the Boston, MA area. They found that people who live close to an event are more likely to attend it and that events of the same type attract people from roughly the same home locations. Although we also study how cellular network data can be used for urban planning, we looked at a different set of research questions, such as deriving and validating labor shed, calculating party shed, capturing lifebeat of a city and clustering users based on their calling patterns.

González et al. [5] used cellular records from an unnamed European country to form statistical models of how individuals move. Isaacman et al. [6] looked at the difference in human mobility between New York and Los Angeles populations. Song et al. [10] studied the predictability of individual's movements and showed that given sufficient past history, one could guess the current location of a given user with high accuracy. Instead of determining how far people move or predicting an individual's movements, this paper studies how cellular network data can be used for urban planning.

8. CONCLUSIONS

The rapid growth of modern cities leaves urban planners faced with numerous challenges, such as high congestion and pollution levels. Effectively solving these challenges requires a deep understanding of existing city dynamics. In this paper, we describe methodology to study and monitor these dynamics by using Call Detail Records (CDRs), routinely collected by wireless service providers as part of running their networks. Our methodology scales to an entire population, has little additional cost, and can be continu-



(a) Cluster 5

(b) Cluster 7

Figure 7: The cellphone usage clusters shown in Figure 6 have very different geographic footprints. For example, cluster 5 (primarily SMS usage outside of school hours) has a Morristown-centric footprint while cluster 7 (SMS usage during business hours) draws people who live in a much wider area.

ally updated. This provides an unprecedented opportunity to study and monitor cities in a way that current practices are not able to do.

Specifically, we studied cellular network traffic around Morristown, NJ, a medium sized, suburban city over a 2 month period and presented three ways in which this data may be useful to urban planners. First, we demonstrated that cellular network data can be used to determine the geographical distributions of home locations of workers and partiers in a city and validated our methodology by comparing our results to the 2000 Census. Second, we presented a novel visualization technique and showed how it can be used to capture the 'life beat' of a city. Finally, we showed that it is possible to cluster city residents and visitors based on their cell phone usage using a novel application of an unsupervised clustering algorithm.

In the future, we plan to investigate how cellular network data can be used to identify anomalous events, such as parades, holidays, or disruptions due to traffic or weather incidents. We also plan to investigate the accuracy of estimating the geographical distribution of work locations of residents of a city. Finally, we plan to study the usefulness of capturing the temporal effects in our data to urban planners.

9. REFERENCES

[1] Bureau of Transportation Statistics. Downloaded from http://www.transtats.bts.gov.

[2] F. Calabrese, F. Pereira, G. DiLorenzo, L. Liu, and C. Ratti. The geography of taste: analyzing cell-phone mobility and social events. In *Proc. of International Conference on Pervasive Computing*, 2010.

[3] F. Girardin, F. Calabrese, F. Dal Fiorre, A. Biderman, C. Ratti, and J. Blat. Uncovering the presence and movements of tourists from user-generated content. In *Intn'l Forum on Tourism Statistics*, 2008.

[4] F. Girardin, A. Vaccari, A. Gerber, A. Biderman, and C. Ratti. Towards estimating the presence of visitors from the aggragate mobile phone network activity they generate. In *International Conference on Computers in Urban Planning and Urban Management*, 2009.

[5] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453, June 2008.

[6] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, J. Rowland, and A. Vasharsvky. A tale of two cities. In *Proc. of 11th ACM Workshop on Mobile Computing Systems and Applications (HotMobile)*, 2010.

[7] R. Pulselli, P. Ramono, C. Ratti, and E. Tiezzi. Computing urban mobile landscapes through monitoring population density based on cellphone chatting. *Int. J. of Design and Nature and Ecodynamics*, 3, 2008.

[8] 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*, 33, 2006.

[9] J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, 6, 2007.

[10] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327, February 2010.

[11] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.