

# Human Mobility Characterization from Cellular Network Data

Richard A. Becker<sup>†</sup>   Ramón Cáceres<sup>†</sup>   Karrie J. Hanson<sup>†</sup>   Sibren Isaacman<sup>◇</sup>  
Ji Meng Loh<sup>†</sup>   Margaret Martonosi<sup>◇</sup>   James Rowland<sup>†</sup>   Simon Urbanek<sup>†</sup>  
Alexander Varshavsky<sup>†</sup>   Chris Volinsky<sup>†</sup>

<sup>†</sup> AT&T Labs – Research, Florham Park, NJ, USA

<sup>◇</sup> Princeton University, Princeton, NJ, USA

<sup>†</sup> {rab, ramon, karrie, loh, jrr, urbanek, varshavsky, volinsky}@research.att.com

<sup>◇</sup> {isaacman, mrm}@princeton.edu

## ABSTRACT

Characterizing human mobility patterns is critical to a deeper understanding of the effects of people’s travel on society and the environment. Location information from cellular telephone networks can shed light on human movements cheaply, frequently, and on a large scale. We have developed techniques for analyzing anonymous cellphone locations to explore various aspects of human mobility. In particular, we have analyzed billions of location samples for hundreds of thousands of people in each of the Los Angeles, San Francisco, and New York metropolitan areas. Our results include measures of how far people travel every day, estimates of carbon footprints due to home-to-work commutes, density maps of the residential areas that contribute workers to a city, and relative traffic volumes on commuting routes. We have validated our techniques through comparisons against ground truth provided by volunteers, and against independent sources such as the US Census Bureau. Throughout our work, we have taken measures to preserve the privacy of cellphone users. This article presents an overview of our methodologies and findings.

## 1. INTRODUCTION

An improved understanding of human mobility patterns would yield insights into a wide variety of important societal issues. For example, evaluating the impact of human travel on the environment depends on knowing how large populations move about in their daily lives. Similarly, understanding the spread of a disease hinges on a clear picture of the ways that humans themselves move and interact. Other examples abound in fields such as urban planning, where knowing how people come and go can help determine where to deploy infrastructure and how to reduce traffic congestion.

Human mobility researchers have traditionally relied on expensive data collection methods, such as surveys and direct observation, to get a glimpse into the way

that people move about. This high cost typically results in infrequent data collection or small sample sizes. For example, a national census produces a wealth of information on where millions of people live and work, but it is carried out only once every ten years.

In contrast, data from cellular telephone networks can help characterize human mobility cheaply, frequently, and on a large scale. Billions of people keep a phone near them most of the time. Since cellular networks need to know the rough location of all active phones to provide them with voice and data services, location information from these networks holds the potential to revolutionize the study of human mobility.

We have used anonymized Call Detail Records (CDRs) from a cellular network to shed light on the mobility patterns of large numbers of people. CDRs are routinely collected by wireless service providers for billing and to help operate their networks, for example to identify congested cells in need of more resources. Each CDR contains information such as the time a phone placed a voice call or received a text message, as well as the identity of the cellular antenna with which the phone was associated at that time. When joined with information about the locations and directions of those antennas, CDRs can serve as sporadic samples of the approximate locations of the phone’s owner.

CDRs are an attractive source of location information for three main reasons. One, they are collected for all active cellular phones, which number in the hundreds of millions in the United States and in the billions worldwide. Two, they are already being collected to help operate the networks, so that additional uses incur little marginal cost. Three, they are continuously collected as each voice call and text message completes, thus enabling timely analysis.

At the same time, CDRs have two significant limitations. One, they are sparse in time because they are generated only when a phone engages in a voice call or

text message exchange. Two, they are coarse in space because they record location only at the granularity of a cellular antenna. It is not obvious a priori whether CDRs provide enough information to characterize human mobility in any useful way.

Since 2009, we have pursued a research program aimed at developing sound analysis techniques for exploring various aspects of human mobility using CDRs. We have shown that CDRs can indeed be used to accurately characterize important aspects of human mobility. Our results to date include the following:

- We have determined how far anonymous populations of hundreds of thousands of people travel every day in each of the Los Angeles, San Francisco, and New York metropolitan areas.
- We have calculated the carbon emissions due to the home-to-work commutes of those same populations, accounting for differences in modes of travel as well as distance.
- We have identified which residential areas contribute what relative numbers of workers and of holiday parade attendees to a suburban city—Morristown, New Jersey (NJ).
- We have estimated relative traffic volumes on the main commuting routes into Morristown.

We have validated our results by comparing them against ground truth provided by volunteers, and against independent sources such as the US Census Bureau. Throughout our work, we have taken measures to preserve individual privacy. The rest of this article presents an overview of the methodologies and findings of our human mobility studies based on cellular network data.

## 2. PRIVACY & TERMINOLOGY

Although CDRs are a valuable source of data for mobility studies that could benefit society at large, cellular customers rightfully have the expectation that their individual privacy will be preserved. We take several active steps to protect privacy. First, all our CDRs are anonymized by someone not involved in the data analysis. Each cellular phone number is replaced with an identifier consisting of a unique integer.

Second, we use only the minimal information needed for our studies. Our simplified CDRs consist of the anonymous phone identifier, date, and time of a voice call or text message; the elapsed time of a call (zero for a text message); the cellular antennas involved in the event; and the phone’s billing ZIP code. Our data does not include demographic information for the subscriber or any information about the other party in the communication. In some of our studies we use the billing ZIP code as a rough estimate of the phone owner’s home

location. We excluded business subscribers from all our datasets because those billing ZIP codes generally do not correspond to home locations.

Finally, we present only aggregate results. We do not focus our analysis on any individual phones, aside from those of a group of volunteers who gave us permission to look at their records.

In addition to these active steps, it is in the nature of CDRs to give only temporally sparse and spatially coarse information about a phone. A CDR is generated only when the phone is used for a call or text message—at all other times the phone is invisible to us. We only know the location of the phone in an approximate way, based on the antennas that were involved with the call. Because an antenna often covers an area greater than one square mile, our spatial resolution is limited.

A brief note on terminology surrounding cellular network equipment will help in understanding the rest of this article. We refer to a cell *tower* as the location of equipment placed on a free-standing tower, atop a building, or on some other physical structure. In general, each tower hosts multiple *antennas*, each handling a particular radio technology and frequency (e.g., Universal Mobile Telecommunications System at 850 MHz) and pointing in a specific compass direction (e.g., north). All the antennas that point in the same direction from the same tower cover what we call a *sector*.

## 3. DAILY RANGE OF TRAVEL

How far do people travel every day? We can approximate this quantity by finding the maximum distance between any two cell towers that a phone contacts in one day, and call this distance the *daily range*. This section presents some of our findings regarding the daily range of people living in three major metropolitan areas in the United States: Los Angeles (LA), San Francisco (SF), and New York (NY).

We gathered anonymous location data for cellular phones whose owners live in the metropolitan regions of interest. First, we identified ZIP codes within a 50-mile radius of the LA, SF, and NY city centers. These ZIP codes correspond to the colored regions in Figure 2. Second, we obtained anonymized CDRs for a random sample of phones with billing addresses in those ZIP codes. Third, so as to exclude people who do not live near their billing address, we removed all CDRs for phones that appeared in their base ZIP code fewer than half the days they had voice or text activity.

Table 1 describes our newest dataset for each region. Each dataset contains hundreds of millions of location samples for hundreds of thousands of phones over 3 months of activity, with 12-18 median location samples per day for each phone.

We compared our sets of phones against US Census data [24] and confirmed that the number of sampled

	LA	SF	NY
Total unique phones	318K	241K	267K
Total unique CDRs	1395M	701M	1095M
Median CDRs/phone/day	18	12	18
Median calls/phone/day	6	5	7
Median texts/phone/day	6	3	5

**Table 1: Characteristics of our Call Detail Record datasets for the Los Angeles, San Francisco, and New York metropolitan areas. Each dataset spans 91 consecutive days from April 1 to June 30, 2011.**

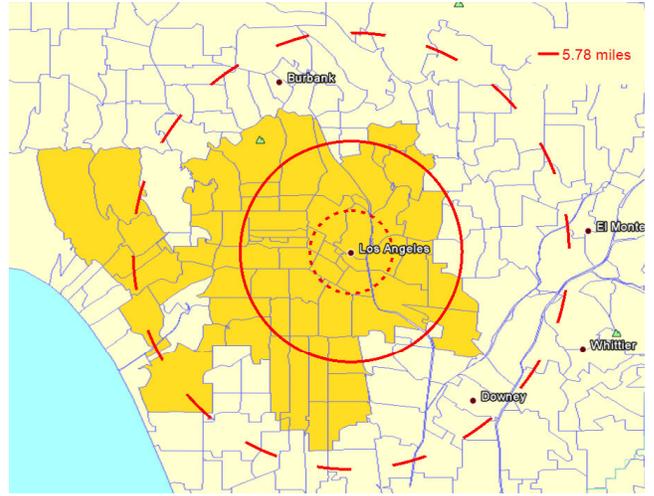
phones in each ZIP code is proportional to the population of that ZIP code. We therefore believe that our datasets are representative of the populations at large in the regions of interest.

We computed a phone’s daily range by calculating distances between all pairs of cell towers contacted by the phone on a given day, and selecting the maximum distance between any two such towers. To validate our methodology, we recruited volunteers who logged their actual locations for one month and gave us permission to inspect their CDRs for that same period. The median difference between daily ranges computed from CDRs and those derived from the ground-truth logs was less than 1.5 miles, giving us confidence in our range-of-travel results. See [13] for more details.

Numerous insights about human mobility arise from the study of daily ranges. For example, the median of a phone’s daily range values over the duration of a dataset is an approximation of the most common daily distance traveled by the phone’s owner. Similarly, the maximum daily range across a dataset corresponds to the longest trip taken during that time.

Figure 1 gives a visual representation of the median daily ranges for residents of central LA, SF, and NY. The darker yellow areas correspond to ZIP codes in the City of Los Angeles, the City of San Francisco, and the Borough of Manhattan. These darker areas do not include the surrounding communities also represented in our complete metropolitan-area datasets. The radii of the red circles are proportional to the median daily ranges for residents of the corresponding shaded areas. As shown, people who live in the City of Los Angeles travel longer distances on a typical day than people who live in the City of San Francisco, who in turn travel longer distances than people who live in Manhattan.

By analyzing similar datasets from different time periods, we have made additional spatial and temporal comparisons between the daily ranges of various populations. For example, people throughout the LA region travel farther on a typical day than people throughout the NY area. In contrast, the longest trips taken by residents of Manhattan are much longer than those



(a) Los Angeles



(b) San Francisco



(c) New York

**Figure 1: Median daily range of cellphone users who live in central LA, SF, and NY (darker yellow areas). The radii of the inner, middle, and outer circles represent the 25th, 50th, and 75th percentiles of these ranges across all users in that area. All maps are drawn to the same scale.**

taken by residents of central Los Angeles. Furthermore, people in both the LA and NY regions tend to travel shorter distances in the winter months than in the summer months, with the effect being more pronounced in NY. For a more complete description of our daily range results, please refer to [13] and [14].

#### 4. CARBON FOOTPRINTS

Evaluating the environmental impact of human travel is of urgent interest to society at large. A person’s commute between home and work can account for a significant portion of their overall carbon footprint. We can estimate the carbon emissions due to these commutes by combining our datasets of cellphone locations with a US Census dataset on mode of transport to work (e.g., automobile, bus, train) [24] and a table of carbon emissions by mode of transport [4].

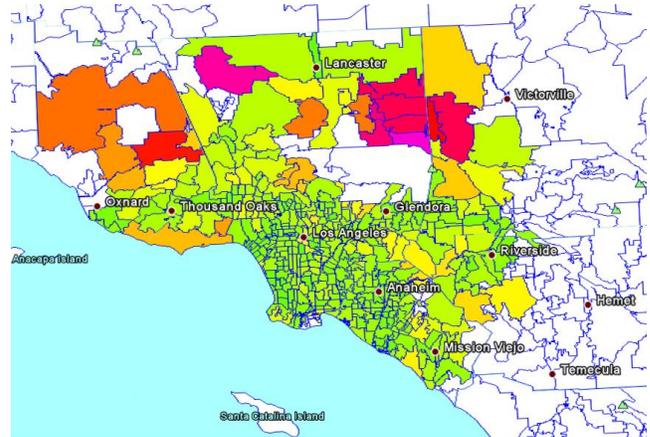
We first devised an algorithm that uses CDRs to identify important places in people’s lives, defined as places that a person visits frequently or where they spend a lot of time. We further identified the likely home and work locations from among those important places, then calculated the home-to-work commute distance. Our approach, described in more detail and validated in [12], uses a series of clustering and regression steps to accomplish these tasks. We found that our commute distance estimates were within one mile of the ground-truth distances provided by volunteers.

We then applied this approach to our large CDR datasets for the LA, SF, and NY metropolitan areas (previously described in Section 3) and computed the distribution of commute distances across the population of each ZIP code in our regions of interest. We found that our estimates were within one mile of the average commute distances for these same regions as published by the US Bureau of Transportation Statistics [23].

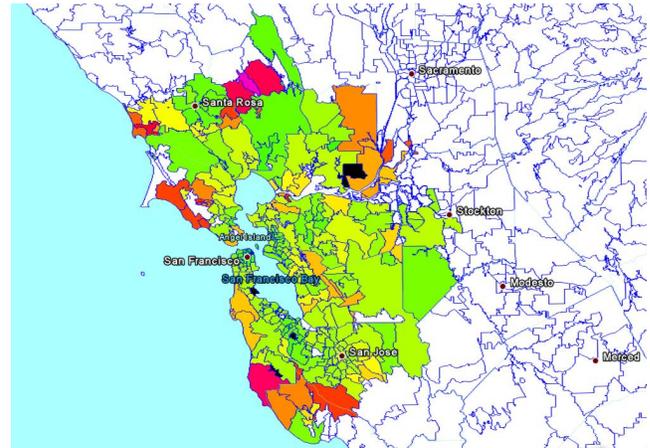
Finally, we joined our distributions of commute distances with the publicly available distributions of modes of transport per ZIP code and of carbon emissions per mode of transport per passenger. Figure 2 shows our results in the form of heat maps, where color corresponds to the median carbon emission per commute across the people in each ZIP code. Colors are ordered so that greener ZIP codes correspond to lower carbon emissions, with yellow, orange, red, and purple ZIP codes showing increasing emissions.

In the NY area, increasing distance from Manhattan correlates with increasing carbon footprint. In contrast, LA is more uniform throughout, except for parts of Antelope Valley (in the northeast portion of the map), which are separated from downtown LA by a mountain range that must be driven around. The results for SF are between those for NY and LA.

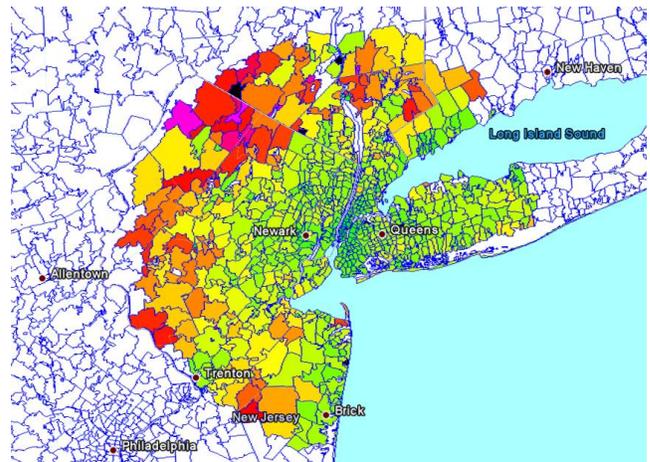
These patterns match well with generally understood movement patterns in each city. Popular knowledge in-



(a) Los Angeles



(b) San Francisco



(c) New York

**Figure 2: Median carbon emissions per home-to-work commute of cellphone users who live in the LA, SF, and NY metropolitan areas. Greener ZIP codes denote smaller carbon footprints, ranging through yellow, orange, red, and purple as footprints grow. All three maps use the same geographic and carbon scales, and the emissions are scaled linearly.**

icates that in NY, a great many people commute into Manhattan, while in LA, there is no single concentration of jobs. SF has at least two major job centers, one focused in the City of San Francisco proper, and the other in Silicon Valley. Thus, unlike NY, SF has more than one strong jobs focus, but unlike LA, it has some clear areas of jobs focus.

Beyond identifying patterns of carbon emissions, we can also compare raw carbon values. For instance, although they are hard to see in Figure 2, Manhattan ZIP codes have the lowest carbon footprints of all ZIP codes studied. These low carbon footprints are presumably due to the nearness to work of many people’s homes, and to effective public transportation.

### 5. LABORSHED & PARADESHED

City and transportation planners are interested in knowing the home locations of people who work in and visit their city. This information is useful, for instance, in forecasting road traffic volumes during the morning and evening rush hours. The set of residential areas that contribute workers to a city is known as the city’s *laborshed*.

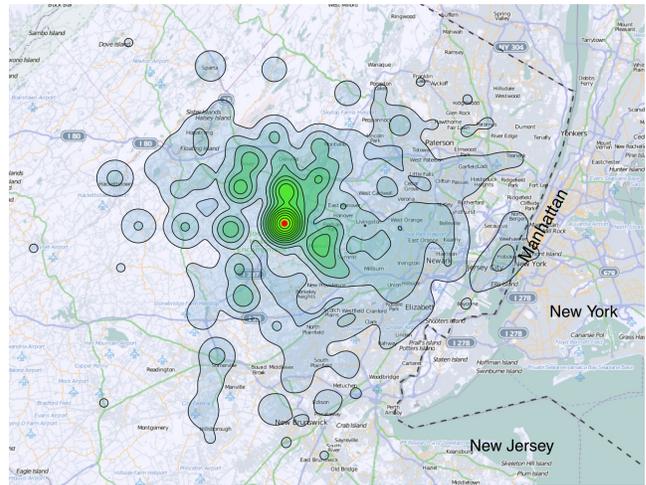
To study an example laborshed, we captured all transactions carried by the 35 cell towers located within 5 miles of the center of Morristown, NJ, a suburban city with approximately 20,000 residents. These 35 cell towers house approximately 300 antennas pointed in various directions and supporting various radio technologies and frequencies. Our goal was to capture cellular traffic in and around the town. Choosing the 5-mile radius allowed us to cover both Morristown proper and its neighboring areas.

We thus obtained anonymized CDRs for 60 consecutive days between March 1 and April 29, 2011. In total, we collected over 17 million voice CDRs and 39 million text CDRs for more than 472,000 unique phones.

We identified Morristown’s laborshed from the CDRs as follows. We first classified as Morristown workers those cellphone users with significant activity inside Morristown during business hours (9am to 5pm, Monday to Friday). We then used billing ZIP codes to identify their places of residence. This method produced counts of Morristown workers by residential ZIP code.

We validated our results by comparing them with data from the US Census. We confirmed that the number of workers we attributed to each ZIP code was strongly correlated with the number of workers in the same ZIP code as published in the “Journey to Work” tables of the 2000 US Census Transportation Planning Package [24]. Our analysis and validation methodology are described in more detail in [2].

Figure 3 shows a geographic representation of Morristown’s laborshed, with darker colors indicating the home areas of larger numbers of Morristown workers.

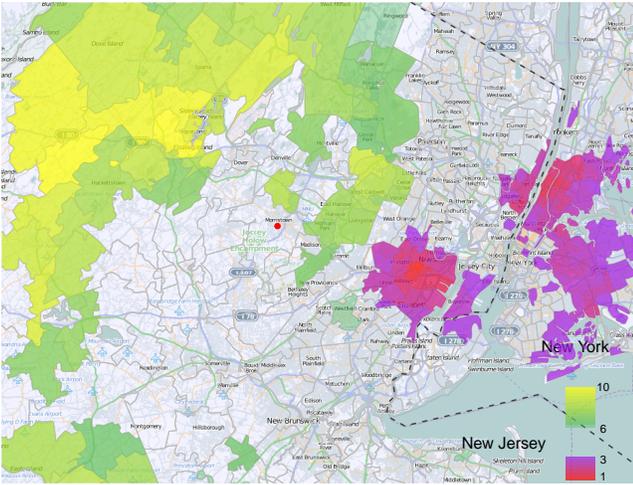


**Figure 3: Laborshed of Morristown.** The red dot is at the city center. Contour lines divide regions of different concentrations of workers’ homes. Workers are identified as those who use their cellphones in Morristown during weekday business hours. Most workers come from nearby areas, but some come from as far as Manhattan.

Interestingly, there seem to be many more workers coming from the area immediately north of Morristown than from the south. These two areas have similar population densities, so this difference may be related to geography, demographics, or transportation infrastructure. Furthermore, although population density increases dramatically to the east (as one gets closer to Manhattan), we see almost as many workers coming in from the west, perhaps because Morristown is a regional center of commerce. However, there do seem to be workers who make long “reverse commutes” from areas of New Jersey close to Manhattan. All of these facts could be useful to policy makers deciding on future municipal and regional mass transit initiatives.

Our methodology allows for estimating the flow of people in and out of a geographic area during arbitrary time periods. Of particular interest to city officials is how the mix of inhabitants changes during special occasions such as extreme weather, construction projects, or regional events. Knowing where people come from can help officials in advertising for the event and in easing traffic congestion.

One such occasion in Morristown was the St. Patrick’s Day Parade that took place between 11am and 3pm on Saturday, March 12, 2011. We repeated the analysis described above for obtaining the laborshed, but on cellphone transactions handled during the time of the parade by the antennas pointing along the parade route. Figure 4 portrays the resulting *paradeshed*, where people come from for the parade, compared with data for the same antennas and time interval but on typi-



**Figure 4: Five times as many people were in Morristown for the St. Patricks Day Parade as on a normal Saturday. To show the geographical distribution of parade attendees’ homes, we mapped the number of people coming from each surrounding ZIP code. Green-yellow areas contributed more than the parade-day average (i.e., more than 5 times the normal Saturday) and purple-red areas less than that average. Communities contributing near the average are not colored to highlight the outliers.**

cal Saturdays without special events. The parade is a county affair, so we expect the event to draw widely from other parts of the county (north and west of Morristown). Indeed, we see the areas north and west of Morristown showing large increases, while other areas south and east show smaller increases. Previously, it has been difficult for local officials to obtain this information except through expensive surveys.

## 6. TRAFFIC VOLUMES

The quality of life in an urban area is directly impacted by the frustration, pollution, time lost, and noise of traffic congestion. Efforts by planners to improve traffic flow while not sacrificing street life needs to be underpinned by a thorough understanding of existing traffic conditions. Since traditional methods of obtaining traffic data are expensive, we set out to determine whether we could estimate traffic volumes from CDRs.

To explore traffic volumes on major commuting routes into Morristown, NJ, we used the same data collection procedure we used to calculate the laborshed, as described in Section 5. However, in this case we recorded activity from December 2009 to January 2010.

We used two filters to obtain an appropriate subset of CDRs for this study. First, to retain only data about moving vehicles, we used only voice CDRs that included

antennas residing on at least 5 towers, as indicated by our own experiments to determine how motion was reflected in CDRs. We ignored text CDRs because text messages involve only a single location. Second, since we were interested in routes to and from the center of town, we used only CDRs with antenna sequences that began or ended at the tower handling calls for the core downtown area. After filtering, we were still left with tens of thousands of CDRs.

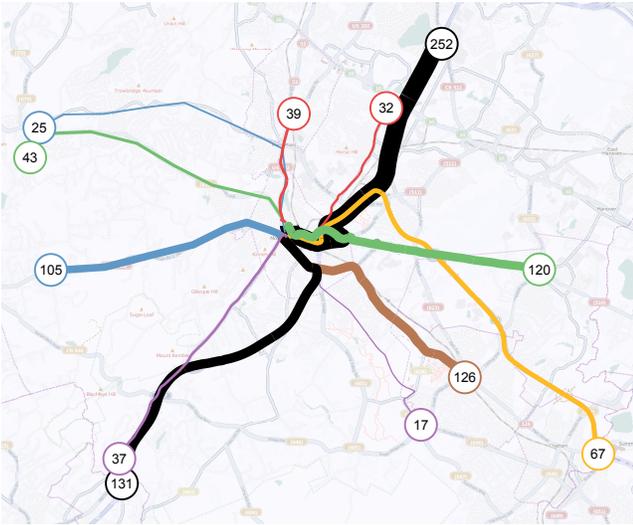
We began our study by identifying 15 common commuting routes (13 driving routes and 2 train routes) that radiate from the town center. We obtained ground-truth data for these routes by driving/riding each of them four times (two in each direction), using at least two phones calling each other on every drive/ride. We obtained the CDRs for these calls both to train and test our algorithms. From our training data, we determined a reference pattern of cellular sectors used by calls on each of our routes. We intentionally included some routes very close to one another and others that partially overlap, as routes do in real life. Some of our reference patterns were thus quite similar, making disambiguation challenging.

We then developed two methods for assigning CDRs to routes. The first uses a distance metric to assign a test CDR to the route with the closest reference pattern. We used a variant of Earth Mover’s Distance (EMD) as a metric that takes into account common subsets of sectors, the particular sequence of sectors, how long the call is associated with each sector, and tower locations.

Our second method for assigning CDRs to routes uses as reference data the radio-frequency scans routinely performed by network operators to measure cellular network coverage. The scanner data contains GPS-stamped signal strength measurements from all observable antennas along major driving routes. Our classification algorithm estimates the likelihood of a given sequence of antennas being seen on a particular route, and selects the most likely route. This approach has the advantage of reusing already available data, thus not requiring additional data collection on every target route. It could easily be extended to larger-scale studies in other urban areas.

Both classification algorithms reached about 90 percent accuracy on our test data. They outperformed several other algorithms that are based purely on common subsets of towers, sectors, or antennas. Our route classification algorithms and their accuracy are described in more detail in [1].

Figure 5 shows the results of our route assignments to moving phones in the Morristown area, using the EMD-based algorithm applied to CDRs. (The signal strength-based method gives similar results.) The relative traffic volumes are normalized to a count per 1000 vehicles and are represented by the widths of the lines



**Figure 5: Relative traffic volumes on twelve commuting routes into the center of Morris-town as assigned by our route classification algorithms. Line widths are proportional to the estimated volumes. Counts shown at the beginning of each route are normalized to 1,000 moving cellphones.**

on each road. The two wide black lines that run roughly north and south correspond to the interstate highway that passes through town. We compared our relative traffic volumes to traffic counts published by the New Jersey Department of Transportation [17]. We found a correlation coefficient of 0.77, giving us added confidence in the accuracy of our approach.

## 7. RELATED WORK

The research community has been increasingly using cellular network data to study human mobility, and applying their findings to various domains, including urban planning [19], mobility modeling [10], social relation inference [11], and healthcare [3]. Below, we survey a subset of that work that is most similar to ours.

Several recent efforts explored how cellular network data can be used for urban planning. In studies of Milan, Italy, Ratti et al. [19] and later Pulselli et al. [18] demonstrated that it is possible to characterize the intensity and spatiotemporal evolution of urban activities using call volume at cellular towers. Reades et al. [20] studied call volume activity in six distinct locations in Rome, Italy, and showed that volume varied drastically between the studied locations and between weekdays and weekends. Girardin et al. [8] used tagged photographs from Flickr in combination with call volume data to determine the whereabouts of locals and tourists in Rome. They later repeated the study with only call volume data to examine differences in behavior between

tourists and locals in New York City [9]. Calabrese et al. [6] studied where people came from to attend special events in Boston, US. 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 pursued a different set of research goals, such as calculating daily ranges, deriving and validating laborsheds, and estimating traffic volumes.

In the domain of mobility modeling, González et al. [10] used cellular network data from an unnamed European country to form statistical models of how individuals move. They found that human trajectories show a high degree of spatial and temporal regularity, with each individual having a time-independent characteristic travel distance and returning often to a few characteristic locations. Song et al. [21] analyzed similar data to study the predictability of an individual’s movements. They found a high degree of predictability across a large user base, largely independent of travel distances and other factors. Whereas these efforts modeled individuals, we have focused on mobility differences between large populations in distinct geographic regions.

A complementary approach to collecting human mobility data from cellular networks is to collect it from the cellular phones themselves. For example, similarly to our route classification work, CTrack [22] maps a phone’s route by matching the cellular signal-strength fingerprints seen by a phone against a database of such fingerprints. More generally, there is a growing body of work in *participatory sensing*, which uses cellphones as sensors of location and other contexts [5, 7, 16]. Cellphone-based efforts have some attractive properties, most notably that they often have access to more and finer-grained sources of location information, such as GPS readings and WiFi fingerprints, than the cellular antenna identities found in Call Detail Records.

However, our network-based approach maintains important advantages. In particular, the cellphone-based approach typically requires that special software be installed and run on phones, which consumes power on those devices and in general tends to inhibit truly large-scale data collection. In contrast, we use information already collected by the network for all phones, which does not require additional software or consume extra power on mobile devices. As a result, our work has involved orders of magnitude more subjects than participatory sensing efforts to date.

## 8. CONCLUSION

Overall, our goal has been to make a case for the value of cellular network data in supporting a range of research and policy goals related to human mobility. Through several studies and publications over the past

two years or more, our group has demonstrated how Call Detail Records—despite their temporal sparseness and spatial coarseness—offer important insights into the movement patterns of individuals and communities.

To demonstrate the broad utility of CDR data, our work comprises several types of analyses. In one case, we demonstrated techniques for identifying important places in people’s lives from CDR traces. Coupling these with other data, such as US Census data on transportation usage, we can generate estimates of carbon footprints in a manner that can be updated much more frequently than typical census surveys, which are expensive and therefore infrequent. We have also shown the use of CDR-based analysis to map laborshed statistics, and to help predict how special events (e.g., a holiday parade) might influence commute and travel patterns.

These studies point to the immense value of cellular network data for use in future urban planning scenarios, such as traffic congestion mitigation and mass transit planning. In contrast to expensive and infrequent census approaches, the fact that CDR-based mobility data can be collected in unobtrusive ways has the potential to make broad use both cheaper and easier.

Underpinning all this work is the desire that useful statistics and models be gleaned from the data without impinging on the personal privacy of individual cellular telephone users. We employed a variety of anonymization techniques to ensure privacy preservation. More broadly, we have shown that a range of useful conclusions can be drawn about regional mobility patterns based solely on anonymized, sampled, and highly aggregated versions of the source mobility data.

Our most recent work seeks to provide fully synthetic models that mimic the individual and regional mobility patterns seen in the measured CDRs [15]. Such models, we believe, will further broaden the ability of scientists and planners to perform accurate, low-cost, and privacy-preserving human mobility studies.

## 9. REFERENCES

- [1] R. Becker, R. Cáceres, K. Hanson, J. M. Loh, S. Urbanek, A. Vasharsvsky, and C. Volinsky. Route classification using cellular handoff patterns. In *13th International Conference on Ubiquitous Computing (Ubicomp)*, 2011.
- [2] R. Becker, R. Cáceres, K. Hanson, J. M. Loh, S. Urbanek, A. Vasharsvsky, and C. Volinsky. A tale of one city: Using cellular network data for urban planning. *IEEE Pervasive Computing*, 10(4), 2011.
- [3] L. Bengtsson, X. Lu, A. Thorson, R. Garfield, and J. von Schreeb. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in haiti. *PLoS Med*, 8(8), 2011.
- [4] M. J. Bradley and Associates. Comparison of energy use & CO<sub>2</sub> emissions from different transportation modes. Report to American Bus Association, 2007.
- [5] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. In *Workshop on World-Sensor-Web (WSW06): Mobile Device Centric Sensor Networks and Applications*, 2006.
- [6] 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.
- [7] D. Cuff, M. Hansen, and J. Kang. Urban sensing: out of the woods. *Commun. ACM*, 51(3), 2008.
- [8] 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 *Int’l Forum on Tourism Statistics*, 2008.
- [9] F. Girardin, A. Vaccari, A. Gerber, A. Biderman, and C. Ratti. Towards estimating the presence of visitors from the aggregate mobile phone network activity they generate. In *International Conference on Computers in Urban Planning and Urban Management*, 2009.
- [10] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453, June 2008.
- [11] C. A. Hidalgo and C. Rodriguez-Sickert. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387(12), 2008.
- [12] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Vasharsvsky. Identifying important places in people’s lives from cellular network data. In *9th International Conference on Pervasive Computing (Pervasive)*, 2011.
- [13] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Vasharsvsky. Ranges of human mobility in Los Angeles and New York. In *8th International Workshop on Managing Ubiquitous Communications and Services (MUCS)*, 2011.
- [14] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, J. Rowland, and A. Vasharsvsky. A tale of two cities. In *11th ACM Workshop on Mobile Computing Systems and Applications (HotMobile)*, 2010.
- [15] S. Isaacman, R. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Vasharsvsky, and W. Willinger. Human mobility modeling at metropolitan scales. In *10th ACM Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2012.
- [16] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda. PEIR, the personal environmental impact report as a platform for participatory sensing systems research. *7th ACM Conference on Mobile Systems, Applications and Services (MobiSys)*, 2009.
- [17] NJ Department of Transportation. <http://www.state.nj.us/transportation/>.
- [18] 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.
- [19] 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.
- [20] J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, 6, 2007.
- [21] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327, February 2010.
- [22] A. Thiagarajan, L. S. Ravindranath, H. Balakrishnan, S. Madden, and L. Girod. Accurate, Low-Energy Trajectory Mapping for Mobile Devices. In *8th USENIX Symp. on Networked Systems Design and Implementation (NSDI)*, Boston, MA, March 2011.
- [23] US Bureau of Transportation Statistics. <http://www.transtats.bts.gov>.
- [24] US Census Bureau. <http://www.census.gov>.