

contributed articles

DOI:10.1145/2398356.2398375

Anonymous location data from cellular phone networks sheds light on how people move around on a large scale.

BY RICHARD BECKER, RAMÓN CÁCERES, KARRIE HANSON, SIBREN ISAACMAN, JI MENG LOH, MARGARET MARTONOSI, JAMES ROWLAND, SIMON URBANEK, ALEXANDER VARSHAVSKY, AND CHRIS VOLINSKY

Human Mobility Characterization from Cellular Network Data

AN IMPROVED UNDERSTANDING of human-mobility patterns would yield insight into a variety of important societal issues. For example, evaluating the effect of human travel on the environment depends on knowing how large populations move about in their daily lives. Likewise, understanding the spread of a disease requires a clear picture of how humans move and interact. Other examples abound in such fields 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 traditionally rely on expensive data-collection methods (such as surveys and direct observation) to glimpse the way people move about. This cost typically results in infrequent data collection or small sample sizes; for example,

the U.S. national census produces a wealth of information on where hundreds of millions of people live and work but is carried out only once every 10 years.

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

We have analyzed billions of anonymized Call Detail Records (CDRs) from a cellular network to characterize the mobility patterns of hundreds of thousands of people. CDRs are routinely collected by wireless-service providers for billing and to help operate their networks by, say, identifying 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 the time). When joined with information about the locations and directions of these antennas, CDRs can serve as sporadic samples of the approximate locations of the associated phones' owners.

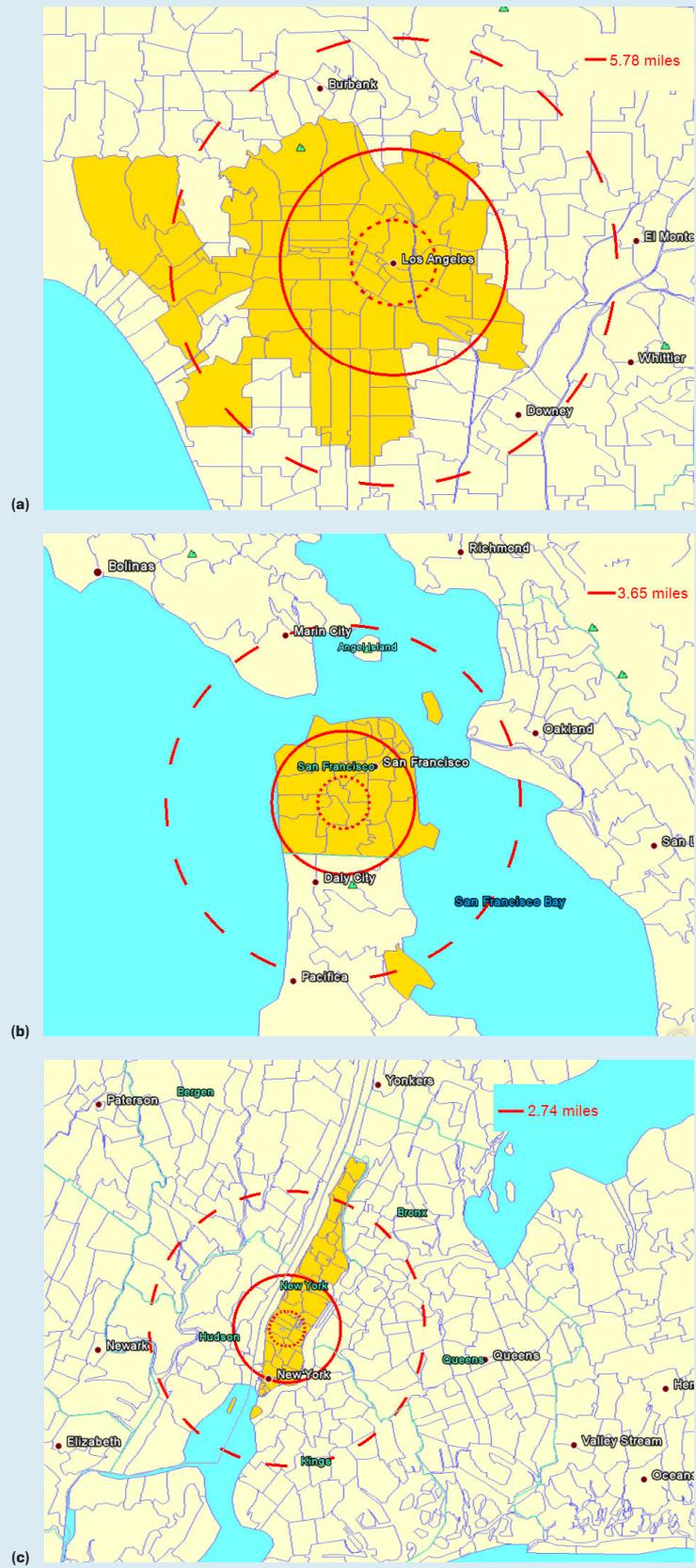
» key insights

- Cellular telephone networks enable the study of human mobility at low cost and on an unprecedented scale.
- Results from such studies have broad applicability in mobile computing, urban planning, ecology, and epidemiology.
- We have developed and validated techniques for analyzing billions of anonymous location samples to determine the daily range of travel, carbon footprint of home-to-work commutes, and other mobility characteristics of hundreds of thousands of people living in the Los Angeles, San Francisco, and New York metropolitan areas.



Figure 1. Median daily range of cellphone users living 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.



CDRs are an attractive source of location information for three main reasons: They are collected for all active cellular phones, numbering in the hundreds of millions in the U.S. and billions worldwide; they are already being collected to help operate the networks, so additional uses incur little marginal cost; and they are continuously collected as each voice call and text message completes, enabling timely analysis.

At the same time, CDRs have two significant limitations: They are sparse in time because they are generated only when a phone engages in a voice call or text-message exchange; and they are coarse in space because they record location only at the granularity of a cellular antenna. Not obvious *a priori* is whether CDRs provide enough information to characterize human mobility in a useful way.

Since 2009, we have pursued a research program aimed at developing sound analysis techniques for exploring aspects of human mobility using CDRs and shown that CDRs are indeed useful for accurately characterizing important aspects of human mobility. Our results to date include the following:

Daily travel. We have determined how far anonymous populations of hundreds of thousands of people travel every day in the Los Angeles, San Francisco, and New York metropolitan areas;

Carbon emissions. We have calculated the carbon emissions due to the home-to-work commutes of these populations, accounting for differences in distance and modes of travel;

Number of workers and event goers. We have identified which residential areas contribute what relative number of workers and holiday parade attendees at a suburban city—Morristown, NJ; and

Traffic volumes. We have estimated relative traffic volumes on the main commuting routes into Morristown.

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

Privacy and Terminology

Though 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:

Anonymization. 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;

Minimal information. 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; and

Aggregate results. We present only aggregate results and do not focus our analysis on 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, the nature of CDRs is 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; the phone is invisible to us at all other times. We know only the location of the phone in an approximate way, based on the antennas 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 the article. We refer to a cell tower as the location of equipment placed on a freestanding 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 (such as Universal Mobile Telecommunications System at 850 MHz) and pointing in a specific compass direction (such as north). All antennas pointing in the same direction from the same tower cover what we call a sector.

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 a phone contacts in one day, calling this distance the daily range. Here, we present some of our findings regarding the daily range of people living in three major metropolitan areas in the U.S.: 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. We identified ZIP codes within a 50-mile radius of the LA, SF, and NY city centers, corresponding to the colored regions in Figure 2. We obtained anonymized CDRs for a random sample of phones with billing addresses in those ZIP codes. And, so as to exclude people not living 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.

The table here describes our most recent dataset for each region, with each dataset containing hundreds of millions of location samples for hundreds of thousands of phones over three months of activity, with 12–18 median location samples per day for each phone.

We compared our sets of phones against U.S. Census data²⁴ and confirmed the number of sampled phones in each ZIP code is proportional to the population of that ZIP code. We there-

fore believe our datasets are representative of the populations at large in the regions of interest.

We computed each 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 the 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; for more, see Isaacman et al.¹³

The study of daily ranges yields numerous insights about human mobility. 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.

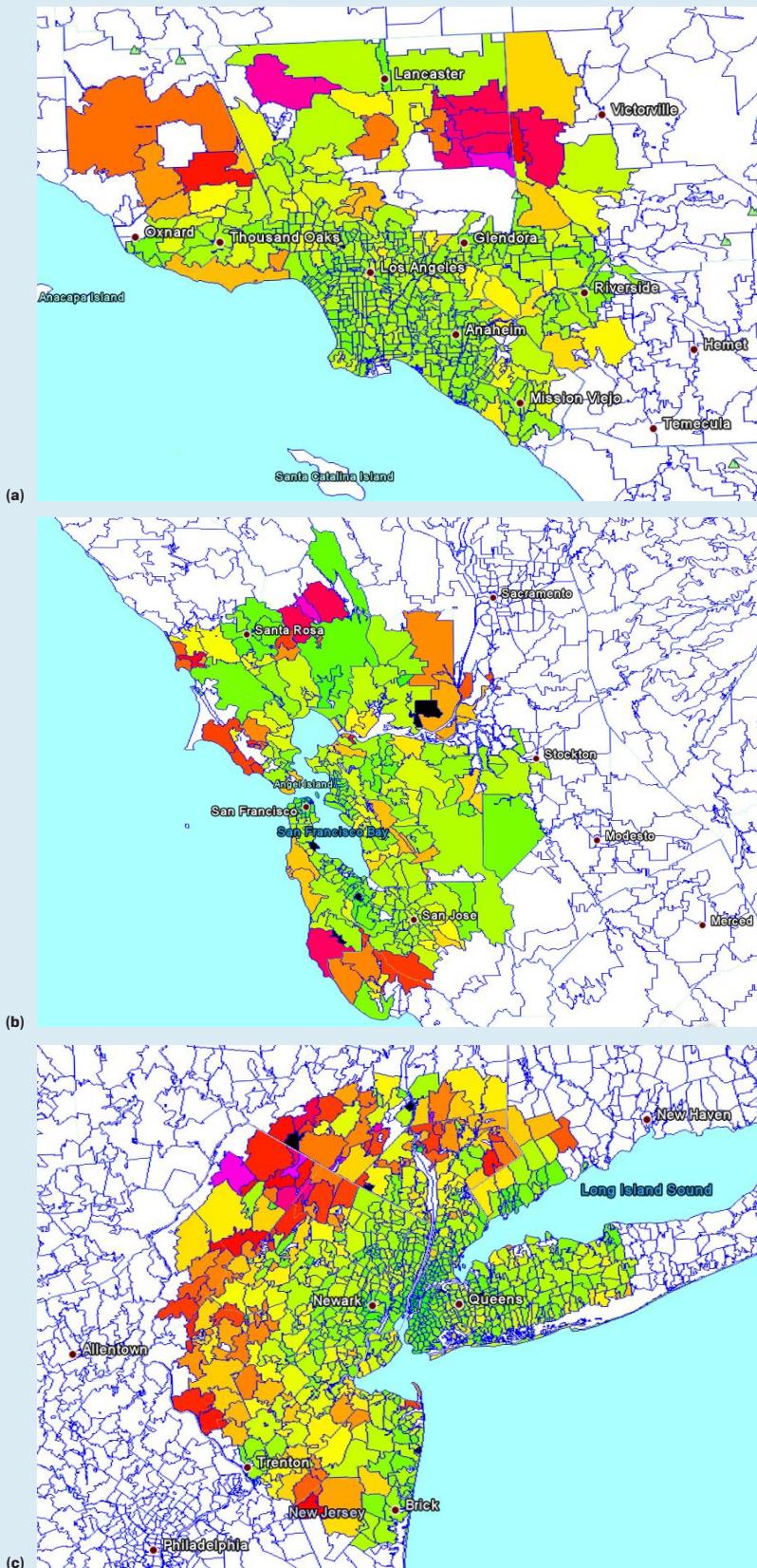
Figure 1 is a visual representation of the median daily ranges for residents of central LA, SF, and NY; the darker yellow areas correspond to ZIP codes in the City of Los Angeles, the City of San Francisco, and the Borough of Manhattan. These areas do not include the surrounding communities also represented 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 living in the city of Los Angeles travel longer distances on a typical day than people living in the city of San Francisco, who in turn travel longer distances than people living in Manhattan.

Characteristics of CDR datasets for the LA, SF, and NY metropolitan areas, with each dataset spanning 91 consecutive days, April 1 to June 30, 2011.

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

Figure 2. Median carbon emissions per home-to-work commute of cellphone users living 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 these maps use the same geographic and carbon scales; emissions are scaled linearly.



By analyzing similar datasets from different time periods, we 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 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, see Isaacman et al.¹³ and Isaacman et al.¹⁴

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 his or her overall carbon footprint. We can estimate the carbon emissions due to these commutes by combining our datasets of cellphone locations with a U.S. Census dataset on mode of transport (such as automobile, bus, and train)²⁴ and a table of carbon emissions by mode of transport.⁴

We devised an algorithm that uses CDRs to identify important places in people's lives, defined as places a person visits frequently or spends a lot of time. We further identified the likely home and work locations from among these important places, then calculated the home-to-work commute distance. Our approach, described in more detail and validated in Isaacman et al.,¹² uses a series of clustering and regression steps to accomplish these tasks. We found 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 described earlier 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 U.S. 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 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 an increasing carbon footprint; in contrast, LA is more uniform throughout, except for parts of Antelope Valley (northeast portion of the map) separated from downtown LA by a mountain range drivers must go 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 indicates 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, another in Silicon Valley approximately 40 miles to the south. 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 also compared raw carbon values. For instance, though difficult to see in Figure 2, Manhattan ZIP codes have the smallest carbon footprints of all ZIP codes studied, presumably due to the nearness to work of many people's homes, as well as to an extensive public transportation infrastructure.

Laborshed and Paradeshed

City and transportation planners are interested in knowing the home locations of people who work in and visit their city. The information is useful in, say, forecasting road-traffic volume during 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 five miles

of the center of Morristown, NJ, a suburban city with approximately 20,000 residents. These 35 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 five-mile radius allowed us to cover both Morristown proper and its neighboring areas. We obtained anonymized CDRs for 60 consecutive days, March 1 to April 29, 2011, thus collecting more than 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 classified as Morristown workers those cellphone users with significant activity inside Morristown during business hours (9 A.M. to 5 P.M., 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 2000 U.S. Census, confirming 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 U.S. Census Transportation Planning Package.²⁴ Our analysis and validation methodology are described in more detail in Becker et al.²

Figure 3 is a geographic representation of Morristown's laborshed, with darker colors indicating the home areas of larger numbers of Morristown workers. 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 the difference may be related to geography, demographics, or transportation infrastructure. Furthermore, though population density increases dramatically to the east (as one gets closer to Manhattan), we see almost as many workers coming from the west, perhaps because Morristown is a regional center of commerce. However, there do seem to be workers making long "reverse commutes" from areas of New Jersey close to Manhattan. All these facts could be useful to policymakers deciding on future municipal and regional mass-transit investments.

Our methodology allows us to estimate the flow of people in and out

Figure 3. Laborshed of Morristown, NJ; the red dot denotes the city center.

Contour lines divide regions of different concentrations of workers' homes, with workers identified as those who use their cellphones in Morristown during weekday business hours. Most workers are from nearby areas, but some are from as far as 40 miles away in Manhattan.

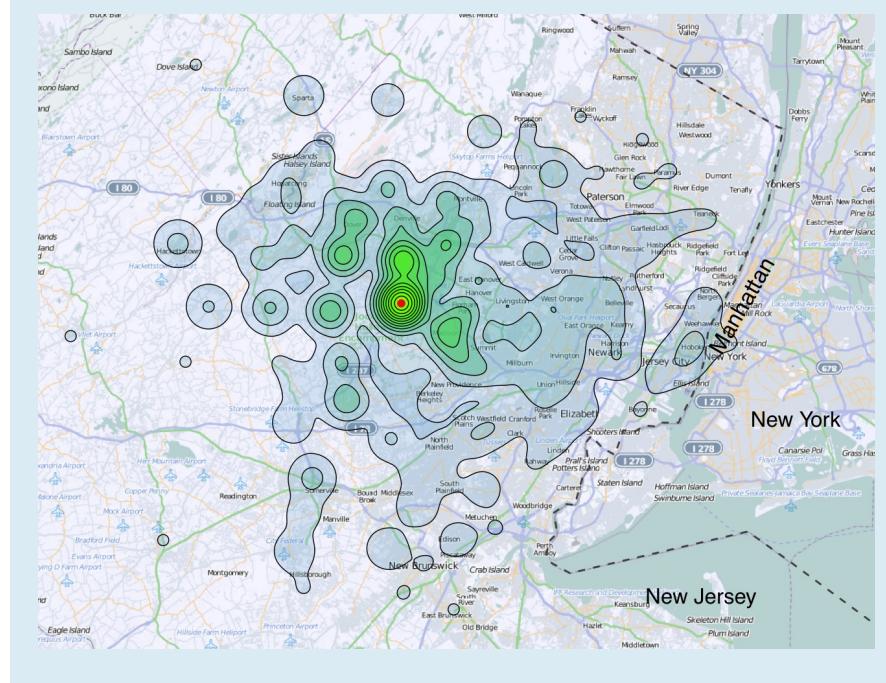
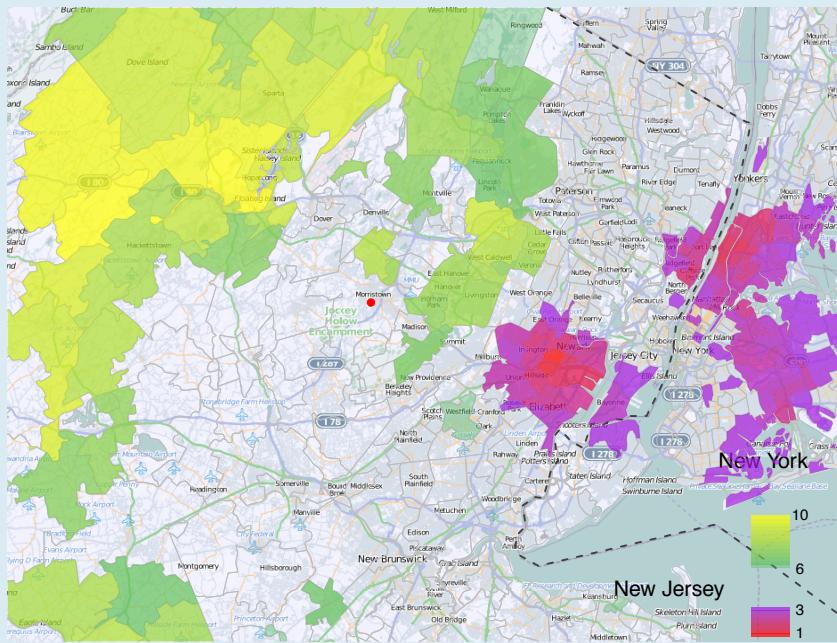


Figure 4. Paradeshed of Morristown, NJ; the red dot denotes the city center.

Five times as many people were in Morristown for the St. Patrick's 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 and purple-red areas less than that average. Communities contributing near the average are not colored to highlight the outliers.



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, and regional events). Knowing where people come from can help them in advertising for the event and easing traffic congestion.

One such occasion in Morristown was the St. Patrick's Day Parade on Saturday March 12, 2011, from 11 A.M. to 3 P.M. We repeated our analysis for obtaining the laborshed but on cell-phone transactions handled during the time of the parade by the antennas pointing along the parade route. Figure 4 is the resulting paradeshed, with people coming for the parade, compared with data for the same antennas and time interval on a typical Saturday without special events. The parade is a county affair, so we would have expected 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. Prior to the

advent of cellular networks, it was notably difficult for local officials to obtain this information except through expensive surveys.

Traffic Volume

The quality of life in any urban area is directly influenced by the frustration, pollution, time lost, and noise of traffic congestion. Efforts by planners to improve traffic flow while not sacrificing street life need 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 volume on major commuting routes into Morristown, we used the same data-collection procedure we used to calculate the laborshed, as described earlier. However, in this case we recorded activity in and around Morristown from December 2009 to January 2010. We used two filters to obtain an appropriate subset of CDRs for the study: First, to retain data about moving vehicles, we used only voice CDRs including antennas on at least five 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 by identifying 15 common commuting routes (13 driving routes and two train routes) radiating from the town center. We obtained ground-truth data for them by driving/riding each one four times (two in each direction), using at least two phones calling each other on each drive/ride. We obtained the CDRs for these calls to both train and test our algorithms. From our training data, we determined a reference pattern of cellular sectors used by calls on each of the 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 a challenge.

We then developed two methods for assigning CDRs to routes: One 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), a measure of the difference between two arbitrary probability distributions, 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. The other method uses as reference data the radio-frequency scans routinely performed by cellular network operators to measure network coverage. The scanner data contains signal-strength measurements stamped with global-positioning system (GPS) locations 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 being able to reuse data that is already available, without requiring additional data collection on every target route. It could easily be ex-

tended to larger-scale studies in other urban areas.

Both classification algorithms achieved approximately 90% accuracy on our test data, outperforming several other algorithms based purely on common subsets of towers, sectors, or antennas. Our route-classification algorithms and their accuracy are described in more detail in Becker et al.¹

Figure 5 shows the result of our route assignment to moving phones in the Morristown area, using the EMD-based algorithm applied to CDRs; the signal-strength-based method yields similar results. The widths of the lines superimposed on each route are proportional to the estimated traffic volumes on each route. The two wide black lines running roughly north and south correspond to the interstate highway that passes through Morristown. The counts shown at the beginning of each route are normalized to 1,000 moving phones. We compared our relative traffic volumes to traffic counts published by the New Jersey Department of Transportation¹⁷ and found a correlation coefficient of 0.77, giving us added confidence in the accuracy of our approach.

Related Work

The research community increasingly uses cellular network data to study human mobility, applying its findings to various domains, including urban planning,¹⁹ mobility modeling,¹⁰ social-relation inference,¹¹ and health care.³ Here, we survey a subset of that work most similar to our own.

Several efforts have 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 it is possible to characterize the intensity and spatio-temporal evolution of urban activities using call volume at cell towers. Reades et al.²⁰ studied call-volume activity in six distinct locations in Rome, Italy, showing 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 exam-

ine 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, finding 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. Though we have also studied how cellular network data can be used for urban planning, we pursued different research goals (such as calculating daily ranges, deriving and validating laborsheds, and estimating traffic volume).

In the domain of mobility modeling, Gonzalez et al.¹⁰ used cellular network data from an unnamed European country to form statistical models of how individuals move, finding human trajectories reflect 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, finding a high degree of predictability across a large user base largely independent of travel

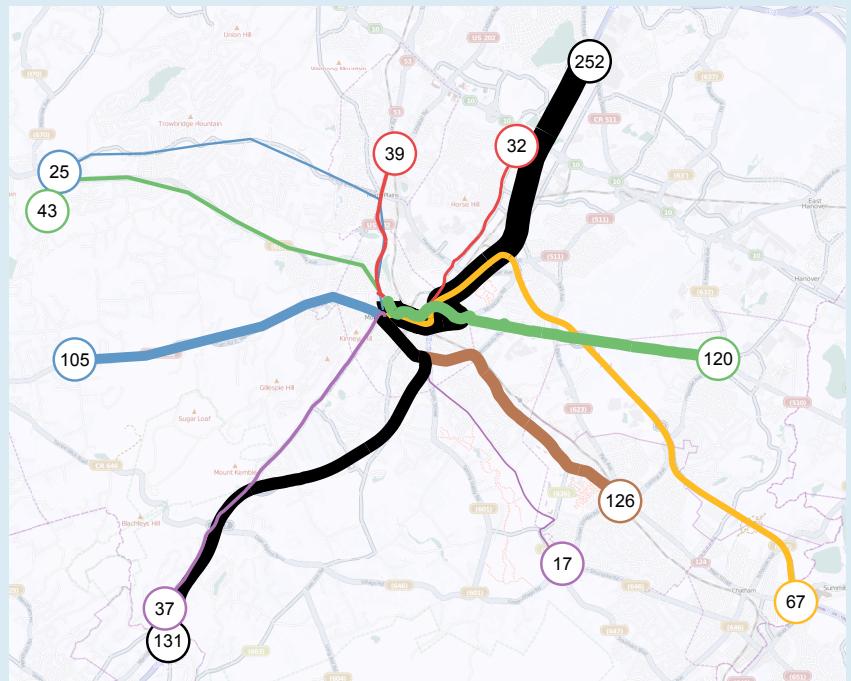
distances and other factors. Whereas these efforts modeled individuals, we focused on mobility differences between large populations in distinct geographic regions.

A complementary approach to collecting human-mobility data from cellular networks is to collect it directly from cellular phones themselves. For example, as in our route-classification work, CTrack²² 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 that 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 varied and finer-grain sources of location information (such as GPS readings and Wi-Fi fingerprints) than the cellular antenna identities in our CDRs.

However, our network-based approach maintains important advantages: In particular, the cellphone-based approach typically requires the installation and running of spe-

Figure 5. Relative traffic volume on 12 commuting routes to the center of Morristown, NJ, 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.



cial software on phones, consuming power on the devices and generally inhibiting truly large-scale data collection. In contrast, we use information already collected by the network for all phones and 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.

Conclusion

Our goal with this work has been to make a case for the value of cellular network data to support a range of research and policy goals related to human mobility. Through several studies since 2009, we have demonstrated how CDRs—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 them with other data (such as U.S. Census data on transportation use) we are able to generate estimates of home-to-work carbon footprints in a manner that can be updated much more frequently than typical census surveys, which are expensive and therefore infrequent. We also showed the use of CDR-based analysis to map laborshed statistics, helping predict how special events (such as a holiday parade) might influence commute and travel patterns.

These studies point to the great value of cellular network data for future urban-planning applications (such as traffic-congestion mitigation and mass-transit planning). Unlike 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.

Motivating all this work is the desire to glean useful statistics and models from the data without compromising the privacy of individual cellular telephone users. We employed various anonymization techniques to ensure privacy preservation. More broadly, we showed that a range of useful conclusions can be drawn about regional

mobility patterns based solely on anonymized, sampled, 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 will further improve the ability of scientists and planners to perform accurate, low-cost, privacy-preserving studies of human mobility.

Acknowledgments

Part of this work was performed while Sibren Isaacman was a doctoral student at Princeton University; it was supported by the National Science Foundation under Grant Nos. CNS-0614949, CNS-0627650, and CNS-0916246. Isaacman also acknowledges support from a Wallace Memorial Fellowship in Engineering from Princeton University and research internships from AT&T Labs.

References

- Becker, R., Cáceres, R., Hanson, K., Loh, J.M., Urbanek, S., Varshavsky, A., and Volinsky, C. Route classification using cellular handoff patterns. In *Proceedings of the 13th International Conference on Ubiquitous Computing* (Beijing, 2011).
- Becker, R., Cáceres, R., Hanson, K., Loh, J.M., Urbanek, S., Varshavsky, A., and Volinsky, C. A tale of one city: Using cellular network data for urban planning. *IEEE Pervasive Computing* 10, 4 (Oct.–Dec. 2011), 18–26.
- Bengtsson, L., Lu, X., Thorson, A., Garfield, R., and von Schreeb, J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Medicine* 8, 8 (Aug. 2011).
- M.J. Bradley and Associates. *Comparison of Energy Use & CO₂ Emissions from Different Transportation Modes*. Report to American Bus Association, Washington, D.C., May 2007; <http://www.buses.org/files/ComparativeEnergy.pdf>
- Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., and Srivastava, M.B. Participatory sensing. In *Proceedings of the Workshop on World-Sensor-Web: Mobile Device-Centric Sensor Networks and Applications* (Boulder, CO, Oct. 2006).
- Calabrese, F., Pereira, F., DiLorenzo, G., Liu, L., and Ratti, C. The geography of taste: Analyzing cell-phone mobility and social events. In *Proceedings of the Eighth International Conference on Pervasive Computing* (Helsinki, May 2010).
- Cuff, D., Hansen, M., and Kang, J. Urban sensing: Out of the woods. *Commun. ACM* 51, 3 (Mar. 2008), 24–33.
- Girardin, F., Calabrese, F., Dal Fiorre, F., Ratti, C., and Blat, J. Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing* 7, 4 (Oct.–Dec. 2008), 36–43.
- Girardin, F., Vaccari, A., Gerber, A., Biderman, A., and Ratti, C. Towards estimating the presence of visitors from the aggregate mobile phone network activity they generate. In *Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management* (Hong Kong, June 2009).
- González, M.C., Hidalgo, C.A., and Barabási, A.-L. Understanding individual human mobility patterns. *Nature* 453, 5 (June 2008), 779–782.
- Hidalgo, C.A. and Rodriguez-Sickert, C. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications* 387, 12 (May 2008), 3017–3024.
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., and Varshavsky, A. Identifying important places in people's lives from cellular network data. In *Proceedings of the Ninth International Conference on Pervasive Computing* (San Francisco, June 2011).
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., and Varshavsky, A. Ranges of human mobility in Los Angeles and New York. In *Proceedings of the Eighth International Workshop on Managing Ubiquitous Communications and Services* (Seattle, Mar. 2011).
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., and Varshavsky, A. A tale of two cities. In *Proceedings of the 11th ACM Workshop on Mobile Computing Systems and Applications* (Annapolis, MD, Feb. 2010).
- Isaacman, S., Becker, R., Cáceres, R., Martonosi, M., Rowland, J., Varshavsky, A., and Willinger, W. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th ACM Conference on Mobile Systems, Applications, and Services* (Lake District, U.K., June 2012).
- Mun, M., Reddy, S., Shilton, K., Yau, N., Burke, J., Estrin, D., Hansen, M., Howard, E., West, R., and Boda, P. PEIR, the personal environmental impact report as a platform for participatory sensing systems research. In *Proceedings of the Seventh ACM Conference on Mobile Systems, Applications, and Services* (Krakow, Poland, June 2009).
- N.J. Department of Transportation; <http://www.state.nj.us/transportation/>
- Pulselli, R., Romano, P., Ratti, C., and Tiezzi, E. Computing urban mobile landscapes through monitoring population density based on cellphone chatting. *International Journal of Design and Nature and Ecodynamics* 3, 2 (2008).
- Ratti, C., Pulselli, R.M., Williams, S., and Frenchman, D. Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design* 33, 5 (2006), 727–748.
- Reades, J., Calabrese, F., Sevtuk, A., and Ratti, C. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing* 6, 3 (July–Aug. 2007), 30–38.
- Song, C., Qu, Z., Blumm, N., and Barabási, A.-L. Limits of predictability in human mobility. *Science* 327, 5968 (Feb. 2010), 1018–1021.
- Thiagarajan, A., Ravindranath, L.S., Balakrishnan, H., Madden, S., and Girot, L. Accurate, low-energy trajectory mapping for mobile devices. In *Proceedings of the Eighth USENIX Symposium on Networked Systems Design and Implementation* (Boston, Mar. 2011).
- U.S. Bureau of Transportation Statistics. Washington, D.C.; <http://www.transportstats.bts.gov>
- U.S. Census Bureau. Washington, D.C.; <http://www.census.gov>

Richard Becker (rab@research.att.com) is a Member of Technical Staff at AT&T Labs - Research, Florham Park, NJ.

Ramón Cáceres (ramon@research.att.com) is a Lead Member of Technical Staff at AT&T Labs - Research, Florham Park, NJ.

Karrie Hanson (karrie@research.att.com) is a Lead Member of Technical Staff at AT&T Labs - Research, Florham Park, NJ.

Sibren Isaacman (isaacman@cs.loyola.edu) is an assistant professor of computer science at Loyola University Maryland, Baltimore.

Ji Meng Loh (loh@njit.edu) is an associate professor at the New Jersey Institute of Technology, Newark, NJ.

Margaret Martonosi (mrm@princeton.edu) is the Hugh Trumbull Adams '35 Professor of Computer Science at Princeton University, Princeton, NJ.

James Rowland (jrr@research.att.com) is Director of Applied Data Mining Research at AT&T Labs - Research, Florham Park, NJ.

Simon Urbanek (urbanek@research.att.com) is a Principal Member of Technical Staff at AT&T Labs - Research, Florham Park, NJ.

Alexander Varshavsky (varshavsky@research.att.com) is a Senior Member of Technical Staff at AT&T Labs - Research, Florham Park, NJ.

Chris Volinsky (volinsky@research.att.com) is Executive Director of Statistics Research at AT&T Labs - Research, Florham Park, NJ.