Ranges of Human Mobility in Los Angeles and New York

Sibren Isaacman*, Richard Becker†, Ramón Cáceres†,
Stephen Kobourov‡, Margaret Martonosi*, James Rowland† , Alexander Varshavsky†

*Dept. of Electrical Engineering, Princeton University, Princeton, NJ, USA
{isaacman,mrm}@princeton.edu
†AT&T Labs – Research, Florham Park, NJ, USA
{rab,ramon,jrr,varshavsky}@research.att.com
‡Dept. of Computer Science, University of Arizona, Tucson, AZ, USA
kobourov@cs.arizona.edu

***Abstract*—The advent of ubiquitous, mobile, personal devices creates an unprecedented opportunity to improve our understanding of human movement patterns. In this work, we study human mobility in Los Angeles and New York by analyzing anonymous records of approximate locations of cell phones belonging to residents of those cities. We examine two data sets gathered six months apart, each representing hundreds of thousands of people, containing hundreds of millions of location events, and spanning two months of activity. We present, compare, and validate the daily range of travel for people in these populations. Our findings include that human mobility changes with the seasons: both Angelenos and New Yorkers travel less in the winter, with New Yorkers showing a more pronounced decrease in mobility during the cold months. We also show that text messaging activity does not by itself accurately characterize daily range, whereas voice calling alone suffices. Finally, we demonstrate that our methodology is accurate by comparing our results to ground truth obtained from volunteers.**

I. Introduction

An improved understanding of human mobility would yield insights into a variety of societal issues. There are long-standing examples in urban planning, where knowing how people move can help determine where to deploy infrastructure such as public transit stations [10]. Similarly, understanding the ways in which disease spreads hinges on a clear picture of the ways that humans themselves spread [1]. Of more recent concern, determining how far people travel in the course of their daily lives sheds light into their environmental impact.

Location information from cellular telephone networks has the potential to revolutionize the study of human mobility. Researchers have traditionally relied on surveys and observations of relatively small numbers of people to get a glimpse into the way that humans move about. In contrast, cellular networks must know the rough location of the millions of active cell phones in their coverage areas in order to provide the phones with voice and data services. Given the almost constant physical proximity of cell phones to their owners, location data from these networks has great promise as a tool for the large-scale characterization of human mobility.

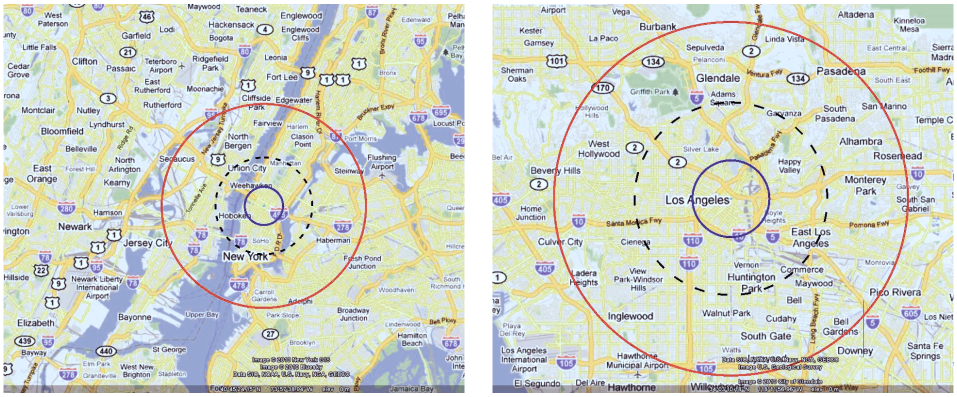
In this work, we use location information from a cellular network to characterize human mobility in two large metropolitan areas in the United States (US): Los Angeles (LA) and New York (NY). Specifically, we analyze anonymous records of approximate cell phone locations at discrete times when the phones are in active use. These records are in two datasets gathered roughly six months apart. Each dataset spans two months of activity for hundreds of thousands of phones and has hundreds of millions of location events.

We focus on aggregate statistics of daily travel. We define *daily range* as the maximum distance that a phone, and by assumption its owner, has been seen to travel in one day. We proceed to make spatial and temporal comparisons of these ranges. For example, as shown in Figure 1, cell phone users in downtown LA have median daily ranges that are nearly double those of their Manhattan counterparts.

In a previous workshop paper, we presented preliminary results from our study of human mobility [9]. Our contributions in this paper go beyond our earlier ones in a number of ways. One, we work with a second set of cell phone activity records gathered at a different time of year. This added dataset allows us to make comparisons across both time and space, for example range comparisons between spring and winter. Two, our new dataset includes text messaging activity in addition to voice calling activity. This added data enables a comparison of mobility estimates based on different types of activity: voice, texting, and both types combined. Three, we compare our range results to ground truth provided by opt-in volunteers to verify whether cell phone activity in fact serves as a good proxy for user locations.

Our main findings include the following: First, people travel less during the winter than during the summer, with the effect being more pronounced in NY than in LA. Second, text messaging activity alone yields a vastly underestimated view of daily travel, while voice calling activity is representative of the full range. Third, daily ranges derived from cell phone activity match the ground-truth ranges provided by volunteers, thus validating our methodology. In addition to these new findings, we observed many of the same human mobility characteristics when using our winter dataset as we did with our earlier spring dataset.

To summarize, this paper demonstrates that location information from cellular networks can bring to light significant aspects of human mobility. We have identified aspects that vary across space and time, and other aspects that remain fairly constant across populations. The rest of this paper describes in more detail our datasets, our analysis methodology, our validation of this methodology, and our range results.



(a) Manhattan

(b) Downtown Los Angeles

Fig. 1. Maps giving a visual representation of the median daily ranges of cell phone users in Manhattan and downtown Los Angeles. The radii of the inner, middle, and outer circles represent the 25^{th}, 50^{th}, and 75^{th} percentiles, respectively, of these ranges across all users in a city. Ranges for all users in a city are made to originate in a common point for clarity of display. Maps are drawn to the same scale.

II. Datasets

A. Data Collection Methodology

Our datasets are comprised of anonymized Call Detail Records (CDRs) from a random set of cellular phones whose billing addresses lie within specific geographic regions.

Defining Geographic Regions of Interest

We first developed a target set of 891 ZIP codes located in the Los Angeles and New York areas. In the LA area, the ZIP codes cover the counties of Los Angeles, Orange and Ventura. In the NY area, ZIP codes cover the five New York City boroughs (Manhattan, Brooklyn, Bronx, Queens, and Staten Island) and ten New Jersey counties that are close to New York City (Essex, Union, Morris, Hudson, Bergen, Somerset, Passaic, Middlesex, Sussex, and Warren). Our selected ZIP codes cover similarly sized areas in LA and NY.

Anonymized CDR Contents

We then obtained anonymized CDRs for a random sample of phones in each ZIP code. The CDRs contain information about two types of events involving these phones: voice calls and text messages. In place of the phone number, 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 text event, the duration of the event, the locations of the starting and ending cell towers associated with the event, and an indicator of whether the phone was registered to an individual or a business.

Excluded Categories of Phones

Our goal is to understand aggregate mobility of people in particular regions of the country, and to compare them analytically where possible. As such, our study omits from consideration two sets of phones from the original CDRs.

First, we omitted phones registered to businesses, retaining only phones registered to individuals. This step avoids, for example, the situation where a cellular service reseller based in a ZIP code of interest would cause us to study large numbers of phones that are not representative of that ZIP code.

Second, we removed from our sample those phones that appeared in their base ZIP code fewer than half the days they had voice or text activity. We assumed that the owners of such phones now live in other parts of the country but have retained their old billing addresses (e.g., they are college students). Therefore, their daily travel patterns may not be representative of the geographical areas we are interested in.

Metric	LA Spring	LA Winter	NY Spring	NY Winter
Total Unique Phones	106K	97K	78K	71K
Total Unique CDRs	74M	247M	62M	161M
Median Calls Per Day	9	8	10	9
Median Texts per Day	N/A	4	N/A	3

TABLE I
General data characteristics of the Spring and Winter datasets. Texting records were not collected in the Spring.

After these two filtering steps, our CDRs are a useful representation of mobility and telephone usage in the regions of interest. While there will always be some people using personal phones for business (and vice versa), we have compared our filtered CDRs against US Census Data for the regions of interest [18] and find a strong correlation between the expected and actual number of users in each ZIP code.

B. Dataset Characteristics

Our overall methodology resulted in location data for hundreds of thousands of phones split roughly evenly between LA and NY, with the number of phones in each ZIP code proportional to the population in that ZIP code. In addition to collecting data for two geographic regions, we also collected data for two time periods. The first period represents 62 consecutive days from March 15, 2009, to May 15, 2009. The second period covers 78 days from November 15, 2009, to January 31, 2010. We refer to these datasets as the Spring and Winter datasets, respectively. Table I offers some general characteristics of these datasets. As shown, each contains hundreds of millions of location events, with on the order of 10 events per phone per day.

C. 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 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 at the beginning and end of 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 coverage radius. Although the effective radius depends upon tower height, radio power, antenna angle, and terrain, these radii average about a mile, giving an uncertainty of about 3 square miles for any event [17].

III. Range of Travel

A. Analysis Methodology

In this study, we use the locations of cellular towers with which a phone is associated as approximations of that phone's locations. We compute a phone's *daily range* by calculating distances between all pairs of locations visited by the phone on a given day; the maximum pairwise distance between any two towers encountered on the same day is the daily range. Our daily range is a lower bound because we calculate distances "as the crow flies" and we do not necessarily see the phone at its extremes of travel. Daily range can only be computed on days in which at least one call or text message is observed.

By calculating the median and maximum values of these daily ranges over the duration of a dataset, we arrive at a phone's *median daily range* and *maximum daily range*, respectively. While the median daily range is an approximation of the "common" daily distance traveled, the maximum daily range corresponds to the longest trip taken across the dataset.

We categorize these ranges by whether they occurred on weekends or weekdays. Our reasoning is that for many people, a weekday range is more closely related to work-related travel (e.g., commuting, business trips), while weekend travel is more often done for pleasure.

We summarize our results with the help of boxplots and histograms. The boxplots depict five-number summaries of the complete empirical distributions of interest. The "box" represents the 25^{th}, 50^{th}, and 75^{th} percentiles, while the "whiskers" indicate the 2^{nd} and 98^{th} percentiles. The horizontal axes show miles on a logarithmic scale.

Nearly any difference between our medians is statistically significant due to our large sample sizes. We could have shown the variability in our data using "notched box plots" [13], where the size of a notch around the median represents the variation of the median. Boxplots whose notches do not overlap would be considered to have come from distributions with significantly different medians. However, because of the large sizes of our datasets, our notches would be imperceptibly small, about the same size as our median lines.

Finally, unless explicitly mentioned, all the figures in this section are based on the Winter dataset. For space reasons, we omit results involving the Spring dataset unless they are substantially different from the Winter results.

B. Spatial Comparisons

In [9], we found a number of differences in the daily ranges of people in the Los Angeles and New York areas based on our Spring dataset. Here, we also have access to our Winter dataset of locations for the same populations. As one would hope, our earlier results also hold for this new dataset.

For example, we again find that Angelenos generally travel farther than New Yorkers. Specifically, the median for weekday daily range is 5.5 miles in LA and 4.0 miles in NY. One likely explanation for this difference is that LA is generally more spread out than NY, so that people in LA travel farther while pursuing day-to-day activities. At the same time, we again see that Manhattanites exhibit maximum trip lengths that far exceed those of the most mobile Angelenos. In particular, the 75^{th} percentile of the maximum daily range is nearly 4 times larger for Manhattan residents than for their downtown Los Angeles counterparts. A possible explanation is the high concentration of business travelers in Manhattan.

Other results that hold across our datasets are the striking variations in mobility between subareas of the same city. Variations in median daily ranges between LA subareas span from 1.05 times at the 98^{th} percentile to 2.05 times at the 25^{th} percentile. Differences between NY subareas are even greater, spanning from 1.01 to 3.03 times at the 98^{th} and 25^{th} percentiles, respectively. One subarea from each city stands out. In NY, Manhattan tends to have significantly lower median daily ranges. This result is consistent with the idea that many people commute into the city center and those that already live there do not need to travel far. Conversely, in LA the outstanding region is Antelope Valley, which is on the outskirts. In this case, it appears that those who live far away from the city center tend to commute much further.

C. Temporal Comparisons

Figure 2 plots the comparison of the median daily ranges between the Spring and the Winter datasets based on CDRs from voice calls only. Since we did not have CDRs for text messages in the Spring dataset, here we removed text CDRs from the Winter dataset for a meaningful comparison.

The figure shows that, on average, people's daily movement tends to be less during the winter months than during the spring. In NY, the 25^{th} percentile fell by 50% and the 75^{th} percentile fell by 23%, whereas in LA, the 25^{th} percentile fell by 37% and the 75^{th} percentile fell by 22%.

Although both LA and NY see drops in the daily ranges, it is suggestive that NY's drop is larger than LA's in all quartiles. Perhaps there is something intrinsic to the NY winter (e.g., inclement weather) that causes people there to move less than during the spring, more so than in LA.

We also studied how the travel patterns changed between the two datasets. To do so, we calculated the ratio between each individual's median daily range in the Spring dataset and the Winter dataset. Figure 3 breaks the results by weekday and weekend for NY and LA areas.

The results show an interesting trend. Although the 25^{th} and the 50^{th} percentiles stayed roughly the same for both LA and NY, the 75^{th} and the 98^{th} percentiles increased significantly. Thus, although the quartile of the population that falls between the 25^{th} and the 50^{th} percentiles did not change their travel patterns, the half falling above the 50^{th} percentile travelled farther during Spring. Also, the increase during the

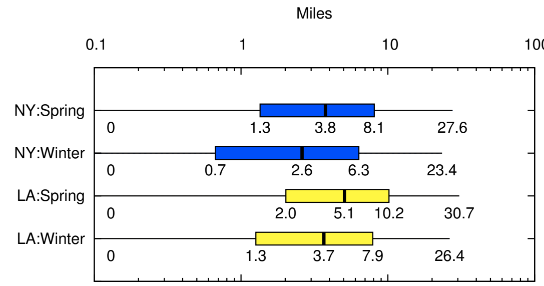


Fig. 2. Median daily range in Spring and Winter, using voice calls.

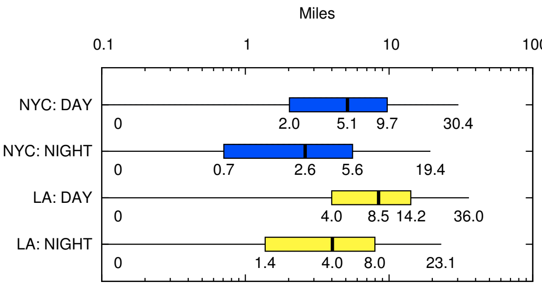


Fig. 5. Median day and night ranges. People travel much shorter distances during the evening hours.

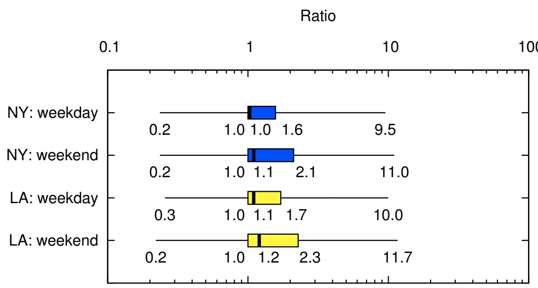


Fig. 3. Ratio between median daily ranges in Spring and Winter.

weekends is higher than during the weekdays. The ratio for the 75^{th} percentile during the weekends is higher than during the weekdays by 34% in NY and 33% in LA.

Similarly, we can take the ratio of an individual user's median daily range on weekends to the same user's median daily range on weekdays. Figure 4 shows the result of taking such a ratio. Over 50% of people move less during weekends than weekdays, regardless of location. However, the decrease is more pronounced in NY than in LA. In NY, the ratio of the 75^{th} percentile is 24% higher than in LA. Subregions exhibit the same ratios as the region they are in, within about 10%.

Our analysis points toward people generally being more mobile on weekdays than weekends. This is consistent with users whose primary travel is to and from work. On days off, a user may be likely to remain closer to home. This may also explain the drop during the Winter dataset. With many people on vacation around the end-of-year holidays, daily distances traveled by commuters could drop sharply. We note, however, that since we can only measure mobility when people use their phones, our results could also be skewed by people using their phones less on weekdays vs. weekends, or using their cell phones less where they have access to landlines.

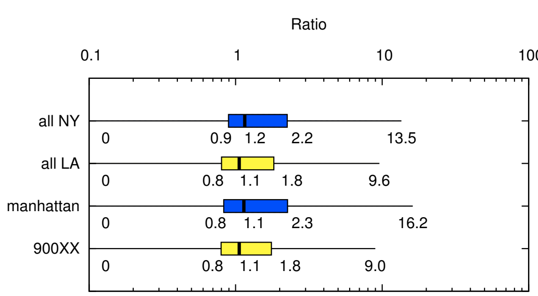


Fig. 4. Ratio between median daily ranges during the weekend and weekday (weekday/weekend). The plot uses only the Winter dataset.

Nevertheless, our data readily shows the effects of holidays on human mobility. For example, Thanksgiving and Christmas clearly stand out as days when many people exhibit their maximum daily ranges within the Winter dataset.

Finally, Figure 5 examines how people move during the "daylight hours" (defined as 7am to 7pm) as opposed to "night time hours". As expected, people move 1.56-2.86x farther during the day than the night, across quartiles. The result, in and of itself, is not a surprise, but the fact that we are able to detect this movement pattern lends legitimacy to our methods. In addition we once again observe the general trend of Angelenos traveling more than New Yorkers. This trend, then, seems to hold regardless of time of day.

D. Voice Calls vs. Text Messages

Recall that the Winter dataset includes CDRs from both voice calls and text messages. We studied the effect of each type of event on the median daily ranges. Figure 6 shows the median daily range computed using CDRs from voice calls only, text messages only, or both voice and text.

The results show that by using CDRs from voice calls only, we would underestimate the median daily ranges by 41% at the 25^{th} percentile, 20% for the 50^{th} percentile, and 11% at the 75^{th} percentile. Also, the results show that CDRs for text events only are unsuitable for estimating median daily ranges, though they serve well as a supplement.

Although it is not clear what causes the low apparent mobility when considering only text messages, it is possible that some of the effect is due to demographics of the users. Because text messages are used primarily by younger generations [12], the mobility estimates would capture a generation that travels mostly to and from school, perhaps a shorter average trip than that of their working parents.

E. Ground Truth for Daily Range

To validate our methodology, we set out to verify whether travel ranges derived from Call Detail Records correspond to actual ranges covered by people. To this end, we recruited five volunteers who agreed to log their actual locations and let us extract location information from their CDRs.

More specifically, the logs captured how long the volunteers spent at which locations throughout the day. Volunteers recorded their logs in whatever form was most convenient to

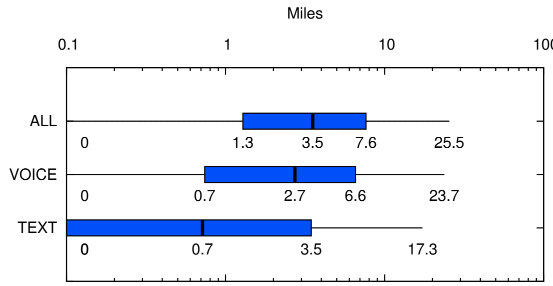


Fig. 6. Comparison of the median daily ranges when considering only one event type. Voice calls have larger ranges than text. For clarity of display, we show only NY area on the weekdays in the winter. Other regions/times exhibit similar behavior.

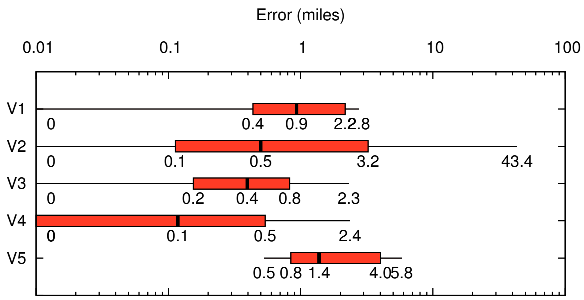


Fig. 7. Differences between daily ranges from 28 days of logs and Call Detail Records for 5 opt-in volunteers. All figures are in miles.

them. For example, some used calendar applications on their mobile phones. We later provided them with a web-based tool in which they could enter a street address or select a point on a map, and the tool would return the latitude and longitude for the address or point. They used this tool to convert the place names in their logs to lat-long pairs. We obtained 28 days of such logs for each of the 5 volunteers.

Figure 7 shows the differences between the median daily ranges calculated from the logs and from the CDRs. The volunteers reported median daily ranges from 0.98 to 25.81 miles while the CDRs had daily ranges from 0.63 to 24.48 miles. In general, median errors per day were always less then 1.5 miles with the majority of errors being much lower. Only one user had egregious errors caused by not making calls at the end points of long journeys. These errors account for roughly 1% of all errors. Overall, our agreement with ground truth gives us confidence in our range of travel results.

F. Daily Distance vs. Daily Range

We define the *daily distance* as the sum of all the distances between consecutive CDRs for a user in one day. Thus, the daily distance is always at least as large as the daily range, and is a tighter lower bound on the total distance traveled by an individual. However, this metric provides less sense of the area over which the user travels, since it can be inflated by, say, making a large number of very short trips.

Table II compares the median daily distances to the median daily ranges in the NY and LA areas, broken down by weekend and weekdays. Although daily distances may differ from daily ranges by up to 10x, the median values remain about 2x. We

City	Days	Median Daily Distance	Median Daily Range	Ratio
NY	Weekdays	5.93	3.19	1.86
	Weekends	4.18	2.39	1.75
LA	Weekdays	9.18	4.50	2.04
	Weekends	6.03	3.50	1.80

TABLE II
COMPARISON OF THE MEDIAN DAILY DISTANCE TO THE MEDIAN DAILY RANGE IN THE WINTER DATASET.

can use these ratios to roughly convert between daily ranges and daily distances as appropriate. One would expect the daily distance to be at least twice the daily range since it should capture travel to and from the farthest destination. Some ratios in Table II are less than 2 because not all people make phone calls or send texts at the end points of their journeys.

The ratio of distance to range is roughly the same in the two cities and across the time periods. Thus, conclusions that we were able to draw from daily ranges are still valid when daily distances are discussed, although the numbers themselves are higher. For example, Angelenos continue to have median travel on weekends that is about 45% higher than New Yorkers (46% by daily range, 44% by daily distance).

IV. RELATED WORK

The usefulness of cellular network operational records has not gone unnoticed in the research community. González, Hidalgo and Barabási [7] used such records from an unnamed European country to form statistical models of how individuals move. From the 6 million phone users in their data set, their study focused on 100,000 users who made at least one call per day over a 6-month period. Whereas they were interested in modeling an individual, we are interested in differences in behavior between large populations.

Song et al. [17] study similar cellular network data from the point of view of predicting an individual's movements. Specifically, the authors consider the towers associated with phone users and show that given sufficient past history, one could guess the current location of a given user with high accuracy. These results are based on a more dense data set of 50,000 users over 3 months, selected from a larger set of 10 million, with the selection focusing on users who make at least one call every two hours. In the same paper they argue that age, gender, population density, and income have little impact on the accuracy of the prediction. Finally, they argue that different days of the week also have little impact, with weekends being slightly easier to predict. Once again, this work focuses on modeling the movements of individuals. However, our goal is to understand the differences at a more macroscopic level. Ultimately, we aim to attribute differences between cities to the geography, demographics, and development patterns of those cities, something that Song et al. explicitly remove from consideration. Thus, the work presented by Song et al. is in some ways orthogonal to our own. Nevertheless, it would be interesting to test their model on our datasets.

Other attempts at performing studies of user mobility also tend to focus on finer-grained movement patterns of individual users. Sohn et al. use GSM data to determine mobility modes, such as walking or driving, of three individuals [16]. Similarly, Mun et al. have developed PEIR [14] to track the environmental impact of individual users of the system. In addition to studying a much larger data set, we look at a completely different metric. Our goal is to look on a more macro scale at the ways in which whole populations behave.

In a step away from studying the patterns of individual users, Pulselli et al. and Ratti examine how call volume can be used as a proxy for population density in Milan [15]. Although call volumes allow them to infer general trends of motion through the city, we feel that our discrete location events provide a more direct picture of human mobility.

Work by Girardin et al. used cell phone usage within cities to find locations of users in Rome [4] and New York City [5]. They were able to find where people clustered in these cities and major paths people take through the cities. They were also able to find differences between the behavior of locals and tourists. In addition to cell phone records, they relied on tagged photos uploaded to popular photo sharing websites. Girardin et al. look at short-term travel patterns to find detailed routes through a single city. We use a much broader dataset that expands our area of interest well beyond the borders of a single city, capturing calls made anywhere in the United States. We have also developed metrics that are general enough to allow comparisons between cities. Further, we have performed longitudinal comparisons into the stability of the datasets that Girardin's team have not be able to perform. Finally, we have a fundamentally different goal than Girardin et al. They aim to find dense regions of a city, while we aim to examine the differences in distances people travel.

Finally, there is a large body of work that examines social relations in the context of mobile communication networks. González, Lind and Herrmann [6] propose a model of social networks based on a system of mobile agents. Hidalgo and Rodriguez-Sickert [8] study the correlations between the structure of a mobile phone network and the persistence of its social links. Eagle and Pentland [2] introduce a system for sensing social systems with data collected from mobile phones over the course of several months. In a followup paper, Eagle and Pentland [3] show that they can identify the structure inherent in daily behavior with models of individuals and communities within the social networks of a population. Lambiotte et al. analyze statistical properties of social networks constructed from the records of a mobile phone company [11]. Our work so far has not explored the relationship between social links and human mobility, but it would be an interesting avenue of future research. We instead aim at developing metrics that can be used to quantify the differences in the ways that people from a particular region move.

V. Conclusions

Cellular phone networks can help solve important problems outside the communications domain because they can provide rich insights into the way people move. Scientists, practitioners, and policy makers in many fields can use human mobility data to explore existing problems and anticipate future ones. By analyzing anonymized records of cell phone locations, we have been able to draw novel conclusions regarding how people move in and around two major cities in the United States: Los Angeles and New York.

Using the concept of a daily range of travel, we have shown concrete differences between the mobility of Angelenos and New Yorkers. By comparing statistics drawn from different time frames, we have also found general truths about human movement that seem not to be tied to the metropolitan region in which people live. We have validated our methodology by comparing daily ranges drawn from cellular phone activity to ground-truth ranges provided by volunteers.

This paper demonstrates our approach to characterizing human mobility patterns on a large scale and without violating individual privacy. Our methodology has wide-ranging applications that can be used, for example, to examine correlations between human movements and world events such as disease outbreaks. We have already demonstrated that events such as national holidays are immediately evident. A better understanding of how such events affect human movement can inform a range of pursuits, from urban planning to disaster response. We plan to collaborate with researchers in these disciplines on applying our data to their problems.

References

[1] D. Brockmann, V. David, and A. M. Gallardo. Human mobility and spatial disease dynamics. *Workshop on Social Computing with Mobile Phones and Sensors: Modeling, Sensing and Sharing*, Aug. 2009.

[2] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4), 2006.

[3] N. Eagle and A. Pentland. Eigenbehaviors: Identifying structure in routine. *Behav. Ecol. Sociobiol.*, 63, 2009.

[4] 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.

[5] 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.

[6] M. González, P. Lind, and H. Herrmann. System of mobile agents to model social networks. *Phys. Rev. Lett.*, 96(8), Mar 2006.

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

[8] C. A. Hidalgo and C. Rodriguez-Sickert. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387(12), 2008.

[9] 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.

[10] The journey to work: Relation between employment and residence. Technical Report 26, American Soc. of Planning Officials, May 1951.

[11] R. Lambiotte, V. D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. V. Dooren. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21), 2008.

[12] R. Ling. The socio-linguistics of SMS: An analysis of SMS use by a random sample of Norwegians. *Mobile Communications: Re-Negotiation of the Social Sphere*, 2005.

[13] R. McGill, J. W. Tukey, and W. A. Larson. Variations of box plots. *The American Statistician*, 32, Feb 1978.

[14] 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. *International Conference on Mobile Systems, Applications and Services*, June 2009.

[15] 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.

[16] T. Sohn, A. Varshavsky, A. LaMarca, M. Y. Chen, T. Choudhury, I. Smith, S. Consolvo, J. Hightower, W. G. Griswold, , and E. de Lara. Mobility detection using everyday GSM traces. *International Conference on Ubiquitous Computing*, Sept. 2006.

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

[18] US census data. Downloaded from http://www.census.gov.