Chapter 2
Modeling and Understanding Intrinsic Characteristics of Human Mobility

Jameson L. Toole, Yves-Alexandre de Montjoye, Marta C. González, and Alex (Sandy) Pentland

Abstract Humans are intrinsically social creatures and our mobility is central to understanding how our societies grow and function. Movement allows us to congregate with our peers, access things we need, and exchange information. Human mobility has huge impacts on topics like urban and transportation planning, social and biologic spreading, and economic outcomes. So far, modeling these processes has been hindered by a lack of data. This is radically changing with the rise of ubiquitous devices. In this chapter, we discuss recent progress deriving insights from the massive, high resolution data sets collected from mobile phone and other devices. We begin with individual mobility, where empirical evidence and statistical models have shown important intrinsic and universal characteristics about our movement: we, as human, are fundamentally slow to explore new places, relatively predictable, and mostly unique. We then explore methods of modeling aggregate movement of people from place to place and discuss how these estimates can be used to understand and optimize transportation infrastructure. Finally, we highlight applications of these findings to the dynamics of disease spread, social networks, and economic outcomes.

J.L. Toole (✉)
Engineering Systems Division, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
e-mail: jamesontoole@gmail.com

Y.-A. de Montjoye • A. (Sandy) Pentland
Media Lab, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
e-mail: yva@mit.edu; pentland@mit.edu

M.C. Gonzáles
Department of Civil and Environmental Engineering, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
e-mail: bgoncalves@gmail.com

2.1 Introduction

Mobility has been a steering force for much of human history. The movement of peoples has determined the dynamics of numerous social and biological processes from tribal mixing and population genetics to the creation of nation-states and the very definition of our living areas and identities. Urban and transportation planners, for example, have long been interested in the flow of vehicles, pedestrians, or goods from place to place.

With more than half of the world's population is now living in urban areas,[1] understanding how these systems work and how we can improve the lives of people using them is more important than ever. Insights from models informed by novel data sources can identify critical points in road infrastructure, optimize public services such as busses or subways, or study how urban form influences its function. Epidemiologists are also relying heavily on models of human movement to predict and prevent disease outbreaks [1, 2] as global air travel makes it possible for viruses to quickly jump continents and dense urban spaces facilitate human-to-human contagion. This has made understanding human movement a crucial part of controlling recent disease outbreaks.[2] Finally, social scientists are increasingly interested in understanding how mobility impacts a number of social processes such as how information spreads from person to person in offices and cafes across the world. These interactions have been theorized to impact crime rates, social mobility, and economic growth [3, 4] and understanding their dynamics may improve how we live, work, and play.

The growing need to understand and model human mobility has driven a large body of research seeking to answer basic questions. However, the lack of reliable and accessible data sources of individual mobility has greatly slowed progress testing and verifying these theories and models. Data on human mobility has thus far been collected through pen and paper surveys that are prohibitively expensive to administer and are plagued by small and potentially biased sample sizes. Digital surveys, though more convenient, still require active participation and often rely on self-reporting [5]. Despite the development of statistical methods to carefully treat this data [6–8] new, cheaper, and larger data sources are needed to push our understanding of human mobility efforts further.

The evolution of technology over the past decade has given rise to ubiquitous mobile computing, a revolution that allows billions of individuals to access people, goods, and services through "smart" devices such as cellular phones. The penetration of these devices is astounding. The six billion mobile phones currently in use triples the number of internet users and boast penetration rates above 100 % in the developed word, e.g. 104 % in the United States and 128 % in Europe.[3] Even

---

[1]United Nations Department of Economic and Social Affairs—World Urbanization Prospects—2014 Update. http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf.

[2]http://www.worldpop.org.uk/ebola/.

[3]GSMA European Mobile Industry Observatory 2011 http://www.gsma.com/publicpolicy/wp-content/uploads/2012/04/emofullwebfinal.pdf.

in developing countries, penetration rates are of 89 %[4] and growing fast. These devices and the applications that run on them passively record the actions of their users including social behavior and information on location[5] with high spatial and temporal resolution. Cellular antennas, wifi access points, and GPS receivers are used to measure the geographic position of users to within a few hundred meters or less. While the collection, storage, and analysis of this data presents very real and important privacy concerns [9, 10], it also offers an unprecedented opportunity for researchers to quantify human behavior at large-scale. With billions of data points captured on millions of users each day, new research into computational social science [11] has begun to augment and sometimes replace sparse, traditional data sources, helping to answer old questions and raise new ones.

In this chapter, we present an overview of mobility research in the current data rich environment. We describe a variety of new data sources and detail the new models and analytic techniques they have inspired. We start by exploring research on individuals that emphasizes important intrinsic and universal characteristics about our movement: we are slow to explore, we are relatively predictable, and we are mostly unique. We then discuss efforts to add context and semantic meaning to these movements. Finally, we review research that models aggregates of human movements such as the flow of people from place to place. Throughout and at the end of this chapter, we point out applications of this research to areas such as congestion management, economic growth, or the spreading of both information and disease.

2.2 New Data Sources

Traditional data sources for human mobility range from census estimates of daily commutes to travel diaries filled out by individuals. These surveys are generally expensive to administer and participate in as they require intensive manual data encoding. To extract high-resolution data, individuals are often asked to recall large amounts of information on when, where, and how they have traveled making them prone to mistakes and biases. These challenges make it hard for surveys to cover more than a day or week at a time or to include more than a small portion of the population (typically less than 1 %).

Mobile phones, however, with their high penetration rates, represent a fantastic sensor for human behavior. A large fraction of location data from mobile phones are currently in the form of call detail records (CDRs) collected by carriers when users perform actions on their devices that make use of the telecommunications

---

[4]ITU. (2013) ICT Facts and Figures http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf.

[5]Lookout (2010) Introducing the App Genome Project https://blog.lookout.com/blog/2010/07/27/introducing-the-app-genome-project/.
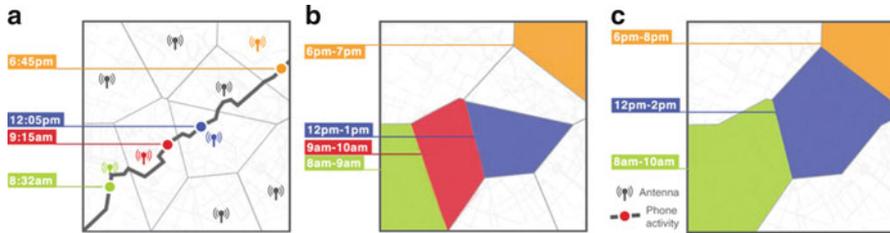
Fig. 2.1 Mobile phones are increasingly being used to collect high-resolution mobility data. This figure from de Montjoye et al. [9] depicts (**a**) a sequence of calling events made by a user at different locations. (**b**) These events are localized to the area served by the closest mobile phone tower to the use and (**c**) can be aggregated into individual specific neighborhoods where a user is likely to be found at different times of the day or week

network. The location of each device at the time a call, text, or data request is registered (Fig. 2.1) is recorded by carriers for billing, network performance, and legal purposes. Locations are inferred either by observing the tower through which the phone is connected or by triangulation with nearby towers. With the increasing use of mobile phones, each individual generates tens to hundreds of these digital breadcrumbs on a daily basis and this number is only increasing. Through specific agreements or through open-data challenges [12], location data on millions of users is readily available to researchers and has been used extensively to augment and sometimes replace traditional travel surveys. This data now forms the core of numerous new mobility studies and models some of which we describe below.

Though generally less common than CDRs, applications running on smartphones may access even more precise estimates of a user's position. A variety of these sensors, from GPS to wifi, can pinpoint the location of a device to within just a few meters and can record data every few minutes [13]. Similarly, protocols such as bluetooth and NFC (near field communication) allow devices to discover and connect to one another within a few meter radius, creating ad hoc sensor and social proximity networks [14]. Some of these applications and underlying social-networks explicitly add crucial context to mobility data. Foursquare invites users to "check-in" at specific places and establishments, Twitter will automatically geotag tweets with precise coordinates from where they were sent, and the Future Mobility survey app passively maintains an activity diary [5] requiring little input from users.

Infrastructure and public services have also become much smarter and now collect data on their usage to improve and help plan operations. Toll booths automatically count and track cars and this data has helped create accurate and real-time traffic estimates used by mapping and navigation services to provide better routing information. Subways, streetcars, and busses use electronic fare systems that record when millions of users enter and exit transportation systems to help better predict demand. In addition to smarter public infrastructure, the ecosystem created by digital devices has given birth to entirely new transportation services such as Hubway, the Boston bike rental service, that collects data of every bike ride

and has even released some publicly[6] or Uber, an on-demand car service, that uses historical usage data to balance the time a user has to wait for a car to arrive and the time drivers spend without clients. Finally, on-board devices and real-time data feeds from automatic vehicle location (AVL) systems power applications such as NextBus to track the location of thousands of busses and subways across the world to display and predict when the next bus will arrive. While smart infrastructure comes with its own privacy challenges [15],[7] vehicle and public transport data offer additional information to urban planners and mobility modelers to better understand these systems.

Finally, most practical mobility models need to properly account for geography such as mountains and rivers, transportation infrastructure such as bridges and highways, differences in density between urban and rural areas, and numerous other factors. Thankfully, the digitization of maps has led to an explosion of geographic data layers. Geographic information systems (GIS) have improved dramatically while falling data storage prices have made it possible for small and large cities to offer their public mapping data to citizens in an online, machine readable format. The U.S. Census Bureau's TIGERline program, San Francisco's OpenSF, and New York City's PLUTO data warehouse are just a few sources that offer huge repositories of publicly accessible geographic data on everything from building footprints and the location of individual trees in a city. Open- and crowd-sourced initiatives like OpenStreetMap allow anyone in the world to contribute and download high-resolution digital maps of roads, buildings, subways, and more, even in developing areas that may not have institutional resources to create them. Private efforts such as Google Maps and MapBox offer high-resolution satellite imagery, route planning, or point of interest information through free or low cost APIs. Put together, these resources provide a digital map of the world that serves as a rich backdrop on which to study human mobility and the infrastructure built to facilitate it.

Put together, new sources from CDRs to public transport data, from mobile phone applications to AVLs generate a data with size and richness prohibitively expensive to match via traditional methods. Collected passively and without any effort from the user, this data is often more robust to manipulation by conscious or unconscious biases and provides a signal that is difficult to fake. While we are convinced of the potential of this data, it is always important to remember that it is not without pitfalls. It would be illusory to think that all of the old biases or hidden variables would simply disappear because the data is large. In some cases, data is only recorded when an individual interacts with a device which may bias when samples are taken [16]. Similarly it is important to keep in mind that even if it covers a significant fraction of the population this data might not be representative. Finally, these data generally come stripped of context. We do not know why an individual

---

[6]Hubway Data Visualization Challenge (2012) http://hubwaydatachallenge.org/.

[7]New York taxi details can be extracted from anonymized data, researchers say (2014) http://www.theguardian.com/technology/2014/jun/27/new-york-taxi-details-anonymised-data-researchers-warn.

has chosen to move or what they will be doing there. For these reasons, sampling and robust statistical methods are still–maybe more than ever[8]—needed to use this data to augment our current understanding of human mobility while still providing robust conclusions. We now discuss a number of studies that aim to do just this.

2.3   Individual Mobility Models

Understanding mobility at an individual level entails collecting and analyzing sets of times, places, and semantic attributes about how and why users travel between them. For example, on a typical morning one may wake up at home, walk to a local coffee shop on the way to the bus that takes them to work. After work they may go to the grocery store or meet a friend for dinner before returning home only to repeat the process the next day. The goal of modeling this mobility is to understand the underlying patterns of individuals using new high resolution data. Models can be used to plan infrastructure or public transport. Furthermore, models provide insights into the underlying nature of human behavior helping us understanding how we are slow to explore, relatively predictable, and mostly unique.

Early modeling work draws a great amount of inspiration from statistical physics, with numerous efforts making parallels with human mobility and random walk or diffusion processes. One of the used data from the crowd-sourced "Where's George" project. Named after George Washington, whose head appears on the $1 bill, the project stamped bills asking volunteers to enter the geographic location and serial number of the bills in order to build a travel history of various banknotes. As bills are primarily carried by people when traveling from store to store, a note's movement serves as a proxy for human movement. Modeling the bills trajectories as continuous random walks, Brockmann et al. found that their movement appears to follow a Lévy flight process [20]. This process is characterized by subsequent steps whose angular direction is uniformly distributed, but whose step-lengths follow a fat-tailed distribution. While small jumps are most probable, bills have a significant probability of making long jumps from time to time. These findings are aligned with observations that humans tend to make many short trips in a familiar area, but also take longer journey's now and then.

In 2008, Gonzalez et al. [17] showed that the movement of these bills does not tell the whole story. Using a CDRs dataset of more than $100,000$ users over a 6-month period in a European country (Fig. 2.2a), they showed that the step-length distribution for the entire population was better approximated by a truncated power-law $P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} \exp(-\Delta r/\kappa)$ with exponent $\beta = 1.79$ and cutoff distances between 80 km and 400 km. This suggests that Lévy flights are only a good approximation of individual's mobility for short distances. To understand the
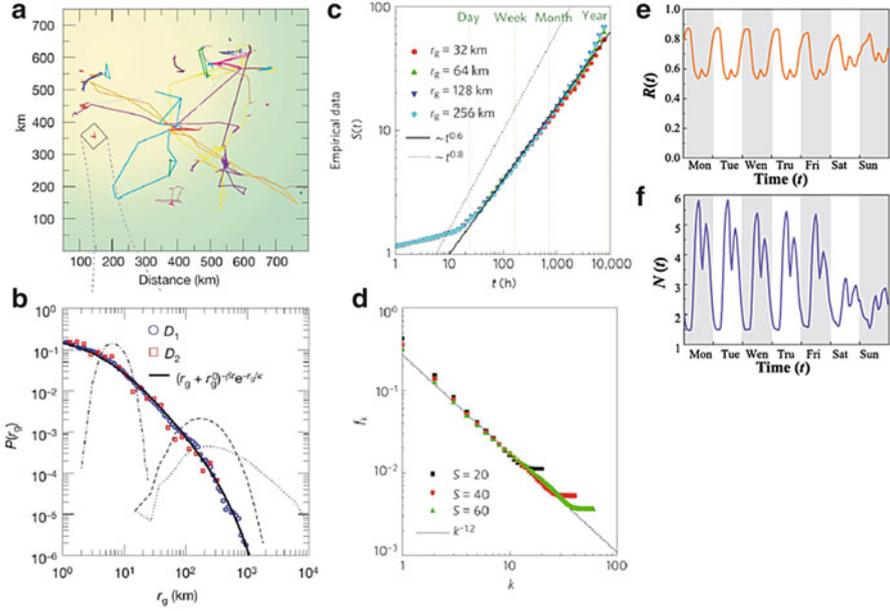
---

[8]Flowing data—Where People Run in Major Cities http://flowingdata.com/2014/02/05/where-people-run/.

Fig. 2.2 (**a**) Individual mobility trajectories are passively collected from mobile devices [17]. (**b**) Measuring the distribution of radius of gyrations, r_g within a population of 100,000 users in a European country reveals considerable heterogeneity in typical travel distance of individuals. Moreover, this distribution cannot be explained by modeling each individual's movement as realizations of a single Lévy flight process [17]. (**c**) and (**d**) Show the slower than linear growth in new locations visited over time $S(t)$ and that the probability a location is visited next is inversely proportional to the frequency it has been visited in the past [18]. (**e**) This preferential return contributes to strikingly high predictability $R(t)$ over time while (**f**) the number of unique locations visited in any given hour is highly periodic and corresponds to the sleep-wake cycles of individuals [19]

mechanism that gives rise to this distribution, the authors borrowed a quantity from polymer physics known as the "radius of gyration" r_g:

$$r_g(t) = \sqrt{\frac{1}{N(t)} \sum_{i=1}^{N(t)} (\mathbf{r} - \mathbf{r_{cm}})^2}, \qquad (2.1)$$

where N(t) are the number of observed locations and r_{cm} is the mean location of the user during the observation period. In essence, the radius of gyration is a measurement of the characteristic distance an individual travels during an observation period t. The authors then showed that the distribution of r_g in the population is itself well approximated by a truncated power-law with $r_g^0 = 5.8$ km, $\beta_{r_g} = 1.65$, and a cutoff of $\kappa = 350$ km (Fig. 2.2b). Simulations suggest that the step-length distribution of the entire population is produced by the convolution

of heterogeneous Lévy flight processes, each with a different characteristic jump size determined by an individual's radius of gyration. Put differently, each person's mobility can be approximated by a Lévy flight process up to trips of some individual characteristic distance r_g. After this distance, however, the probability of long trips drops far faster than would be expected from a traditional Lévy flight.

Further investigation by the authors revealed the source of this behavior: the idiosyncrasy of human movements. Unlike random processes, humans are creature of habits and tend to return to previously visited locations such as home or work. The nature of these returns was also found to follow a very particular pattern. An individual returns to a previously visited location with a probability proportional to that location's rank $P(L) \sim 1/L$ amongst all the places he or she visits. These non-random, predictable return visits are unaccounted for in random walk and Lévy flight models and have been shown to be at the heart of deviations of observed behavior from random processes. Additional studies [21] have found similar patterns in both other CDRs datasets and Foursquare or Twitter check-ins.

Subsequent work by Song et al. [18] further studied how individual-specific locations need to be taken into account in mobility models. Using a similar CDR dataset, the authors showed three important characteristics of human behavior. First, the number of unique locations visited by individuals $S(t)$ scales sub-linearly with time $S(t) \sim t^\mu$ where $\mu = 0.6$ (Fig. 2.2c). Second, the probability an individual returning to a previously visited locations scales with the inverse of the rank of that location $P(L) \sim L^{-\zeta}$ where $\zeta = 1.2$ (Fig. 2.2d), a phenomena labeled as "preferential return." Third, the mean displacement, Δr, of an individual from a given starting point shows slower than logarithmic growth, demonstrating the extremely slow diffusion of humans in space. In essence, these findings pinpoint the dampening of explorative human movement overtime. Long jumps are observed so infrequently that they do not affect the average displacement of individuals. The authors then propose a new model of human mobility to capture these three characteristics. The model is as follows: starting at time t, an individual will make a trip at some future time Δt drawn from a fat-tailed probability distribution measured from CDRs. With probability $\rho S^{-\gamma}$, the individual travels to a new, never-before visited location some distance Δr away, where Δr is drawn from the fat-tailed distribution characterized in the previous model. With probability $1 - \rho S^{-\gamma}$ an individual returns to a previously visited location according to the inverse rank equation.

These early models do not attempt to recover periodic aspects of movement (e.g., daily commuting) or semantic meaning of visits (e.g., to visit a friend or go shopping), or attempt to do so. They do, however, emphasize important statistical and scaling properties of human mobility and often successfully reproduce them. Taken together, these models show that human explore slowly space, returning more often than not to known places and with less long steps than predicted by a power-law distribution.

Approaching the problem from the perspective of machine and statistical learning, another set of models has uncovered and explored another facet of human mobility: how predictable we are. In [19], Song et al. used information theory

metrics on CDRs to show the theoretical upper-bound on predictability using three entropy measures: the entropy S, the random entropy S^{rand}, and the uncorrelated entropy S^{unc}. They then use their empirical distributions to derive an upper bound on a user's predictability ($\prod^{\text{max}}$, $\prod^{\text{rand}}$, and $\prod^{\text{unc}}$). On average, the potential predictability of an individual's movement is an astounding 93 % and no user displayed a potential predictability of less than 80 %. To further quantify predictability, the author introduced two new metrics. They defined regularity $R(t)$ as the probability a user is found at their most visited location during a given hour t, along with the number of unique locations visited during a typical hour of the week $N(t)$ (Fig. 2.2e, f). Both show strong periodicity and regularity. These quantities have since been measured in different data sets in different cities and countries and have been shown to be consistent among them [21].

While the previous study provided a theoretic upper bound on the predictability of an individual, a number of statistical learning techniques have been developed to make predictions in the traditional sense. Early work in the area, predating even analytic computations, used Markov models and information on underlying transportation networks to predict transitions between mobile phone towers within cities. These models have been used to improve quality of service of wireless networks through proper resource allocation [22–25]. Later work incorporated various trajectory estimation and Kalman filtering algorithms to predict movements in small spaces such as college campuses [26, 27].

Temporal periodicity was used by Cho et al. [28] in their Periodic Mobility Model and social behavior incorporated in the Period Social and Mobility Model. These approaches derive the probability distribution of a user to be at any given location at a given time from previous location data. The latter also account for the location history of social contacts. The authors used these models to estimate that as much as 30 % of our trips may be taken for social purposes. Multivariate nonlinear time series forecasting produced similar results [29, 30] predicting where an individual will be either in the next few hours or at a given time of a typical day. These models, however, are all focused on predicting the geographic position of individuals at different times and do not attempt to understand what individuals may be doing there or any other semantics of place.

Though acquiring semantic information about mobility is more difficult than simply measuring geographic coordinates, it provides a much richer abstraction to study behavior. In one of the first studies to mine the behavior of college students using mobile phones, Eagle et al. [31] gave a few hundred students smart phones that recorded not only locations, but also asked users to label each place with its function such as home or work. Applying principal component analysis to these abstract movements from semantic place to semantic place (as opposed to geographic movements alone), the authors found that an individual's behavior could be represented as a linear combination of just a few "eigenbehaviors." These eigenbehaviors are temporal vectors whose components represent activities such as being at home or being at work. They can be used to predict future behaviors, perform long range forecasts of mobility, and label social interactions [14, 32]. The price paid for such detailed predictions, however, is the need for semantic

information about locations. Geographic positions need to be tagged with attributes such as home or work in order for them to be grouped and compared across individuals.

Another approach to studying more abstract measurements of individual location information comes from recent work by Schneider et al. [33]. The authors introduced "mobility motifs" by examining abstract trip chains over the course of a day. A daily mobility motif is defined a set of locations and a particular of visits. More formally, these motifs constitute directed networks where nodes are locations and edges are trips from one location to another. For example, the motif of an individual whose only trips in a day are to and from work will consist of two nodes with a two directed edges (one in both directions). Counting motifs in mobility data from both CDRs and traditional travel surveys, they found that, on average, individuals visit three different places in a given day. They then construct all possible daily motifs for a given number of locations n and compute their frequencies. Shockingly, while there exists over 1 million ways for a user to travel between 6 or fewer locations, 90 % of people use one of just 17 motifs and nearly a quarter follow the simple two location commute motif introduced earlier (Fig. 2.3a). The authors found similar results in travel survey data and introduced a simple Markov model for daily mobility patterns which reproduces empirical results.

It is tempting to hypothesize that high theoretical and practical predictability results from high levels of similarity between individuals in a region. Perhaps the pace of life, full of mono-centric downtowns, or the structure of transportation systems funnel users to the same places and route choices. de Montjoye et al. [9] explored this hypothesis and found that, while predictable, an individual's movement patterns are highly unique. The authors introduced "unicity" $\mathcal{E}_p$ as the fraction of traces uniquely defined by a random set of p spatiotemporal points where a trace T is a set of spatiotemporal points, each containing a location and
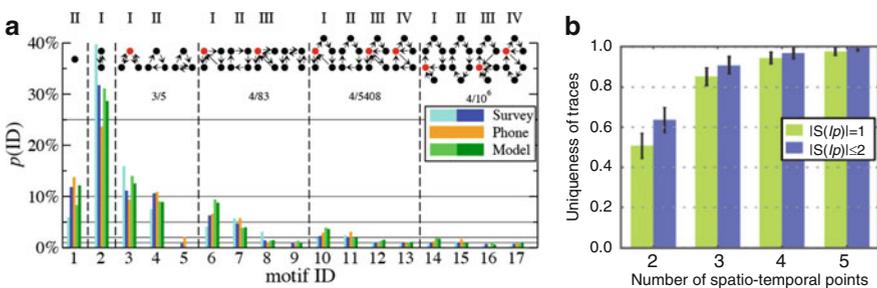

Fig. 2.3 (**a**) Removing geographic coordinates from locations and only focusing on a set of unique places and the directed travel between them, mobility motifs reveal that the daily routines of people are remarkably similar. Despite over 1 million unique ways to travel between 6 or fewer points, just 17 motifs are used by 90 % of the population. Moreover, the frequency of their appearance in CDR data matches very closely with more traditional survey methods [33]. (**b**) Despite this similarity and predictability, our movement displays a high degree of unicity. Just four spatiotemporal points is enough to differentiate a user from 95 % of all others individuals [9]

a timestamp. A trace is said to be uniquely defined by a set of points I_p if it is the only trace that matches I_p in the entire dataset. Applying this measure to a CDR dataset on 1.5 million users, the authors found that just four spatiotemporal points is enough to uniquely identify 95 % of all users (Fig. 2.3b). The authors further study unicity when the data is coarsened spatially or temporally. They found $\mathcal{E} \sim (v * h)^{\beta}$ unicity decrease as a power function with the spatial (v) and temporal resolution of the data (h) and that $\beta \sim -p/100$. Taken together, these equations show that unicity decreases slowly with the spatial and temporal resolution of the data and that this decrease is easily compensated by the number of points p. High uniqueness in human mobility traces exists across many spatiotemporal scales. These results not only raise many questions about the privacy of massive, passively collected metadata datasets, but also highlight an incredibly interesting nuance of human mobility: though individuals are predictable, they are also unique.

Merging concepts of predictability and unicity, work by Sun et al. [34] used temporal encounter networks to study repeated co-locations between passengers using data from bus passengers in Singapore. Temporal encounter networks were constructed by connecting individuals if they rode the same bus at the same time. An average individual encountered roughly 50 people per trip and these trips were highly periodic, occurring at intervals associated with working hours as well as daily and weekly trips. A pair of individuals who encountered each other tended to meet an average of 2.5 times over the course of a week. The distribution of time between encounters reveals strong periodicity, with passengers riding the same bus to work in the morning riding the same home, or riding the same bus at the same time each morning. This finding illustrates the idiosyncrasies of human mobility. Not only we visit a few places very during the day, but we also do so at the same times and by the same routes. Amazingly, though both of these results suggest that our unicity should be low, the previous work shows us that this is not the case.

In summary, new data sources have allowed researchers to show that, over weeks and months, human movement is characterized by slow exploration, preferential return to previous visited places, exploration of daily motifs, and predictable uniqueness. These regularities have been used to develop algorithms capable of predicting movement with high degrees of accuracy and have been shown to mediate other important processes such as social behavior and disease spread. Individual mobility patterns, however, are not the only level of granularity of interest to researchers, city planners, or epidemiologist. Aggregate movement can be either derived from individual level model or modeled as an emergent, personified phenomena. In the next section, we discuss works and models which aim at describing and modeling aggregate movement and flows of many individuals from place to place.

2.4   Aggregate Mobility

Aggregated mobility is used for planning urban spaces, optimizing transportation networks, studying the spread of ideas or disease, and much more. Perhaps the largest component in these models are origin-destination matrices that store the

number of people traveling from any location to any other at different times or by different means. Like many complex systems, aggregate behavior is often more than the sum of individual parts and can be modeled separately. Additional layers of complexity are also needed to account for and sometimes explain individual choice of mode of transportation or route as described by the "four step model" [8, 37].

Like their individual-focused counterparts, many of these aggregate models are inspired by physical processes. Some of the earliest techniques for estimating origin-destination matrices are gravity models which have been used to model flows on multiple scales, from intra-city to international [8, 38]. Borrowed directly from Newton's law of gravitation, the number of trips T_{ij} taken from place i to place j is modeled as a function of the population of each place m_i and m_j and some function of the distance between them $f(r_{ij})$. The intuition is that the population of a place, its mass, is responsible for generating and attracting trips and thus the total flux between the two places should be proportional to the product of the two masses while the distance between them mitigates the strength of this connection. In the fully parameterized version of this model, an exponent is applied to the population at the origin and destination $T_{ij} = a \frac{m_i^\alpha m_j^\beta}{f(r_{ij})}$ to account for hidden variables that may be specific to local regions or populations. While the classical gravity model from physics is recovered by setting $\alpha = \beta = 1$, and $f(r_{ij}) = r_{ij}^2$, these parameters are generally calibrated for specific application using survey data.

Gravity models, however, are not without limitations. First, they rely on a large number of parameters to be estimated from sparse survey data which often leads to overfitting and, second, they fail to account for opportunities that exist between the two masses of people. The latter fault results in the same flow of people being estimated between two locations whether there is an entire city or an empty desert between them. Intuitively, one would expect that trips between places would be affected by the intervening opportunities to complete a journey. These shortcomings lead Simini et al. to develop the radiation model [35]. Again borrowing from physics (this time radiation and absorption), they imagined individuals being emitted from a place at a rate proportional to its population and absorbed by other locations at a rate proportional to the population there. In this model, the probability that an emitted person arrives at any particular place is a function of their probability of not being absorbed before getting there. The model is as follows: $T_{ij} = T_i \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}$, where T_i is total number of trips originating from location i and s_{ij} is the population within a disc centered on location i with a radius equal to the distance between i and j. The radiation model does not directly depend on the distance between the two places, taking instead into account the opportunities in-between them (Fig. 2.4a). Unlike the gravity model, the radiation model is parameterless and requires only data on populations to estimate flow. The authors showed that despite its lack of parameters, the radiation model provides better estimates of origin-destination flows than the gravity model for areas the size of counties or larger.

Yang et al. adapted Simini's radiation model to correct for distortions caused at different scales [39]. They showed the original radiation model's lower accuracy in urban environment is due to the relatively uniform density and small distances
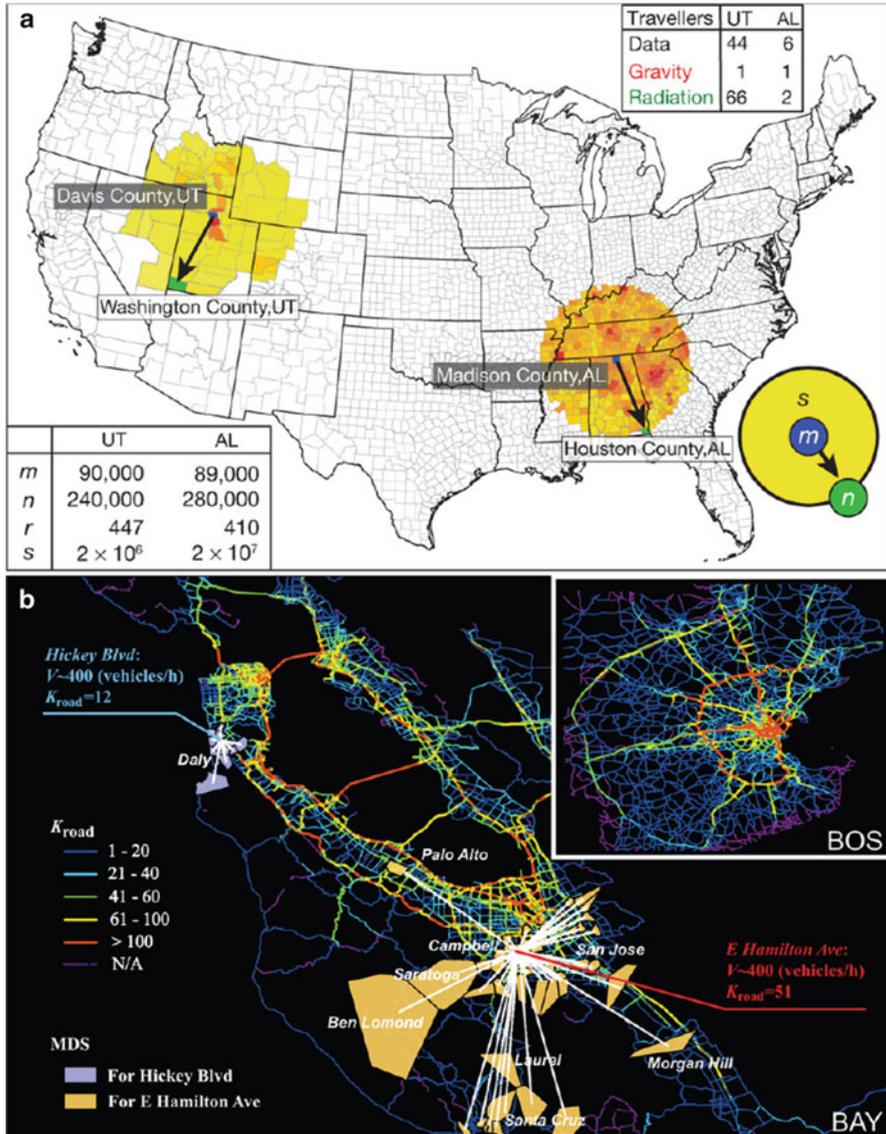
Fig. 2.4 (**a**) The radiation model accounts for intervening opportunities, producing more accurate estimates of flows between two places than more traditional gravity models [35]. (**b**) Routing millions of trips measured from CDR data to real road networks makes it possible to measure the importance of a road based on how many different locations contribute traffic to it, K_{road}. Understanding how transportation systems perform under different loads presents new opportunities to solve problems related to congestion and make infrastructure more efficient [36]

that characterize cities. In dense urban areas, distances are all relatively short and an individual may choose to visit a particular location due to hedonic attributes regardless of whether it is convenient to get to or not. Yang et al. subsequently introduced a scaling parameter α in the function describing the conditional probability an individual is absorbed at a location. This single parameter was enough to correct for these distortions and to provide a model that works on any length scale. Moreover, the authors suggested that for urban areas, the density of points of interest (POIs) such as restaurants and businesses is a better predictor of the absorption of a place than its population. Iqbal et al. [40] have demonstrated an improved way to extract valid, empirical OD matrices from CDR data to validate the model.

Finally, activity-based approaches [6] model user intent more explicitly. They hypothesize that all trips are made to fulfill certain needs or desires of an individual. Travel and survey diaries are used to identify those needs for different segments of the population and how they are typically fulfilled. This knowledge can then be used by the model given the demographics of individuals and environmental factors. These models are closely related to agent-based models simulating the behavior of city residents and rely heavily on the idea of economic utility.

From a practical perspective, city planners need to know not only how many people will go from point A to point B at a certain time of the day but also the mode of transportation and route choice of these individuals. For example, we would like to predict which route they will take so that we can properly estimate the stress placed on transportation systems and potentially optimize performance. Models of route choice typically assume that individual rationally chose the path from A to B that minimize some cost function such as total travel time or distance. Paths can be computed on a road network using shortest path algorithms such as the traditional Dijkstra algorithm or A-Star, an extension that enjoy better performance thanks to heuristics. Other informations such as speed limits can also be taken into account to estimate free flow travel times.

More advanced models are needed to account for the impact of congestion as drivers rarely encounter completely empty freeways. Iterative traffic assignment algorithms model congestion endogenously [41]. Trips are first split into segments containing only a fraction of total flow between two points. Trips in each segment are then routed along shortest paths independently of all other trips in that segment keeping counts of how many trips were assigned to each road. The travel times are then adjusted according to a volume delay function that accounts for the current congestion on a road where congestion is computed as the ratio between the volume of traffic assigned to the segment and the capacity of the road (referred to as volume-over-capacity). Trips in the next segment are then routed using updated costs until all flow has been accounted for. In this way, as roads become more congested and the travel time increases, drivers in later iterations are assigned to different, less congested routes. Values of total volume on each road, congestion, and travel times can then be validated against traffic counters, speed sensors, or data from vehicle fleets like taxis and busses but also smartphones such as in the Mobile Millennium project [42–45].

Wang et al. [36] further explored the use of CDRs as input for these iterative algorithms to estimate traffic volume and congestion. After correcting for differences in market share and vehicle usage rates, they measure trips by counting consecutive phone calls of individuals as they move through the city to generate flow estimates that were then routed. Using this approach, Wang et al. show the distribution of traffic volume and congestion to be well approximated by an exponential mixture model. This model depends on the number of major and minor roadways in a cities network. Using the same approach, the authors describe the usage patterns of drivers by a bipartite usage graph connecting locations in the city to roads used by those travelers (Fig. 2.4b). Roads can be defined by the number of locations that contribute traffic to them while place can be described by the roads used to visit them. The "function" of a road can then be classified by comparing its typological to its behavioral importance. For example, a bridge may be topologically important because it is the only way to cross a river, but a main street may be behaviorally important because it attracts motorists from many different neighborhoods. Using these measures, researchers were able to devise congestion reduction strategies that target the 2 % of neighborhoods where trip reduction will have the largest network wide effect. They found this smart reduction strategy is three to six times as effective as a random trip reduction strategy. Further work used this analysis to predict traffic jams [46, 47].

Private cars, however, are not the only mode of transportation studied. Using smartphones and AVL data, researchers have been mapping the routes followed by public transport and even privately owned mini-buses in the developing countries [48–50]. Similarly, data on air travel has been increasingly available to study aggregated mobilities between cities for applications in epidemiology (see below).

2.5 Human Behavior and Mobility

While of obvious interest to travelers, urban planners and transportation engineers, people's movement strongly impacts other areas. Though by no means an exhaustive list, we highlight three areas here: social behavior, disease and information spread, and economic outcomes. Many of these dynamics are discussed in greater detail in further sections of this volume.

2.5.1 Mobility and Disease Spread

Human movement via cars, trains, or planes has always been a major vector in the propagation of diseases. Consequently, the human mobility data and models discussed so far have increasingly been used to study the propagation of diseases. For example, CDR data has been used to map mobility patterns in Kenya helping researchers in their fight against Malaria [1, 54]. More recently, CDR and other data
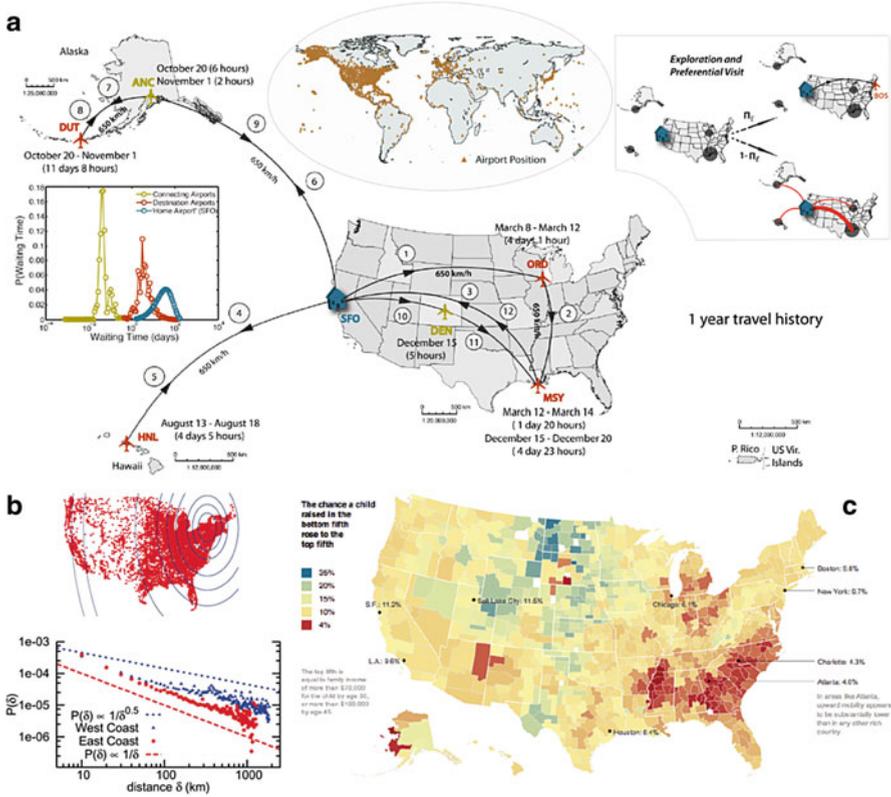
Fig. 2.5 (**a**) Global air travel has dramatically increased the speed at which diseases can spread from city to city and continent to continent [51]. (**b**) Mobility also impacts social behavior as we are far more likely to be friends with someone who lives nearby than far away [52]. (**c**) Mobility and the access it provides has strong correlations with economic outcomes. Children have dramatically different chances at upward economic mobility in certain places of the United States than others [53]

from West-Africa has been used to model regional transportation patterns to help control the spread of Ebola.[9] Finally, air travel data has become central to the study of global epidemics when planes allow an individual to travel between nearly any two points on the globe in a matter of hours. The global airline network therefore often determines how potent an epidemic could be and its likely path across the globe [2, 51, 55, 56] (Fig. 2.5a).

---

[9]Cell-Phone Data Might Help Predict Ebola's Spread (2014) http://www.technologyreview.com/news/530296/cell-phone-data-might-help-predict-ebolas-spread/

2.5.2 Mobility and Social Behavior

Intent is a crucial element of human mobility and movement is often a means to a social end. Despite new communications technologies making it easier than ever to connect across vast distances, face-to-face interactions still play an important role in social behavior whether it is the employees of a company commuting to a central workplaces or friends meeting at a restaurant on a weekend. The link between social contacts and mobility has becoming increasingly prominent in research as mobility data is often collected through mobile phones or location-based social networks.

Using data from an online social-network, Liben-Nowell showed the probability of being friends with another individual to decrease at a rate inversely proportional to the distance between them suggesting a gravity model of the form discussed above [52] (Fig. 2.5b). Subsequent work verified Liben-Nowell findings in other social networks [57, 58] while Toole et al. [59] showed the importance of taking into account geography when studying social-networks and how information spreads through them. Moreover, geographic characteristics can be used to predict the social fluxes between places [60]. Conversely, social contacts are very useful in predicting where an individual would travel next [28, 29, 61] and Cho et al. find that while 50–70 % of mobility can be explained as periodic behavior, another 10–30 % are related to social interactions.

Models such as the one proposed by Grabowicz et al. [58] have subsequently been developed to incorporate this dynamic and evolve both social networks and mobility simultaneously. The authors incorporate social interactions by having individuals travel in a continuous 2D space where an individual travel's is determined by the location of their contacts and use location as a determinant of new social tie creation. The model is as follows: with probability p_v, an individual moves to the location of a friend, and, with probability $1-p_v$, they choose a random point to visit some distance Δr away. But, while social ties impact mobility, mobility can also impact social ties. Upon arriving at a new location, the individual can thus choose to form social ties with other individuals within a radius with probability p or random individuals anywhere in the space with probability p_c, a free parameter. Although simple, this model is able to reproduce many empirical relationships found in social and mobility data.

2.5.3 Mobility and Economic Outcomes

Mobility not only provides people with social opportunities, it also provides economic ones. Economists and other social scientists have developed numerous theories on the role of face-to-face interactions in socio-economic outcomes and economic growth. In-person meetings are thought to unlock human capital, making us productive [62, 63]. For example, jobs in dense cities tend to pay higher wages than the same jobs in more rural areas even after controlling for factors such as age

and education [64] in part due to productivity and creativity gains made possible by the rich face-to-face interactions that close spatial proximity facilitates. Universal urban scaling laws have been repeatedly found showing that societal attributes from the number of patents to average walking speed scales with population and theoretic models have been proposed that suggest density is at the heart of these relationships [3, 4, 65]. While density is one way of propagating these benefits, increased mobility is another. Poorer residents of cities have been, for example, shown to have better job prospects and higher chances of retaining jobs when given a personal car instead of being constrained by public transit [66]. Finally, Chetty et al. [53] found strong correlations between intergenerational economic mobility and variables related to the commuting times and spatial segregation of people (Fig. 2.5c). While we are only beginning to explore these relationships, early returns suggest that mobility is a critical component of many economic systems.

2.6   Conclusion

In this chapter, we reviewed a number of ways new data sources are expanding our understanding of human mobility. Applying methods from statistical physics, machine learning, and traditional transportation modeling, reproducible characteristics of human movement become visible. We explore slowly [17, 18], we are highly predictable [19, 29], and we are mostly unique [9]. Models of aggregate flows of people from place to place have also found success with analogies to statistical physics validated by new data sources [35]. More accurate measurements of city-wide traffic have made it easier than ever to assess the performances of transportation systems and devise strategies to improve them [36]. Valuable in their own rights, these insights have informed our understanding of other social phenomena as well, leading to more accurate models of disease spread, social interactions, and economic outcomes. As cities become home to millions of people each year, the insights gained from these new data are critical in making them more sustainable, safer, and better places to live.

References

 1. Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., et al. (2012, October). Quantifying the impact of human mobility on malaria. *Science, 338*(6104), 267–270.
 2. Colizza, V., Barrat, A., Barthélemy, M., & Vespignani, A. (2006, February). The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America, 103*(7), 2015–2020.
 3. Bettencourt, L. M. A. (2013). The origins of scaling in cities. *Science, 340*, 1438–1441.
 4. Pan, W., Ghoshal, G., Krumme, C., Cebrian, M., & Pentland, A. (2013) Urban characteristics attributable to density-driven tie formation. *Nature Communications, 4*, 1961.

5. Cottrill, C. D. A., Pereira, F. C. A., Zhao, F. A., Dias, I. F. B., Lim, H. B. C., Ben-Akiva, M. E. D., et al. (2013) Future mobility survey. *Transportation Research Record, 2354*, 59–67.

6. Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. Cambridge: MIT Press.

7. Hall, R. W. (Ed.) (1999). *Handbook of transportation science*. International series in operations research & management science (Vol. 23). Boston: Springer.

8. de Dios Ortúzar, J., & Willumsen, L. G. (2011). *Modelling transport*. Chichester: Wiley.

9. de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Nature Scientific Reports, 3*, 1376.

10. de Montjoye, Y.-A., Shmueli, E., Wang, S. S., & Pentland, A. S. (2014). OpenPDS: Protecting the privacy of metadata through SafeAnswers. *PLoS ONE, 9*, e98790.

11. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., et al. (2009). Computational social science. *Science, 323*(5915), 721–723.

12. de Montjoye, Y. A., Smoreda, Z., Trinquart, R., Ziemlicki, C., & Blondel, V. D. (2014, July). D4D-Senegal: The second mobile phone data for development challenge.

13. Aharony, N., Pan, W., Ip, C., Khayal, I., & Pentland, A. (2011). Social fMRI: Investigating and shaping social mechanisms in the real world. In *Pervasive and mobile computing* (Vol. 7, pp. 643–659).

14. Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology, 63*, 1057–1066.

15. Kosta, E., Graux, H., & Dumortier, J. (2014). Collection and storage of personal data: A critical view on current practices in the transportation sector. In *Privacy technologies and policy SE - 10* (Vol. 8319, pp. 157–176).

16. Ranjan, G., Zang, H., Zhang, Z.-L., & Bolot, J. (2012). Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mobile Computing and Communications Review, 16*(3), 33–44. http://dl.acm.org/citation.cfm?id=2412101.

17. González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782.

18. Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010, September). Modelling the scaling properties of human mobility. *Nature Physics, 6*(10), 818–823

19. Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010) Limits of predictability in human mobility. *Science, 327*(5968), 1018–1021.

20. Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature, 439*, 462–465.

21. Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011). Exploring millions of footprints in location sharing services. In *ICWSM* (pp. 81–88).

22. Kim, H. S. (2003, January). QoS provisioning in cellular networks based on mobility prediction techniques. *IEEE communications magazine, 41*(1), 86–92.

23. Liu, T., Bahl, P., Chlamtac, I. (1998). Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks. *IEEE Journal on Selected Areas in Communications, 16*, 922–935.

24. Thiagarajan, A., Ravindranath, L., LaCurts, K., Madden, S., Balakrishnan, H., & Toledo, S. (2009). VTrack: accurate, energy-aware road traffic delay estimation using mobile phones. In *Proceedings of the 7th ACM conference on embedded networked sensor systems - SenSys '09* (pp. 85–98).

25. Krumm, J., Horvitz, E., Dourish, P., & Friday, A. (2006). Predestination: Inferring destinations from partial trajectories. *UbiComp 2006: Ubiquitous Computing, 4206*, 243–260.

26. Minkyong, K., Kotz, D., & Songkuk, K. (2006). Extracting a mobility model from real user traces. In *Proceedings - IEEE INFOCOM*.

27. Lee, K., Hong, S., Kim, S. J., Rhee, I., & Chong, S. (2009). SLAW: a new mobility model for human walks. In *IEEE INFOCOM 2009*.

28. Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility. In *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, KDD 11* (p. 1082). New York: ACM Press.

29. De Domenico, M. (2012). Interdependence and predictability of human mobility and social interactions. *Journal Pervasive and Mobile Computing, 9*, 798–807.

30. Scellato, S., Musolesi, M., Mascolo, C., Latora, V., & Campbell, A. T. (2011). NextPlace: A spatio-temporal prediction framework for pervasive systems. In *Pervasive computing,* Lecture notes in computer science (Vol. 6696, pp. 152–169). Heidelberg: Springer.

31. Eagle, N., & Pentland, A. (2006). Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing, 10*, 255–268.

32. Sadilek, A., & Krumm, J. (2012). Far out: Predicting long-term human mobility. In *AAAI* (pp. 814–820).

33. Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of the Royal Society, Interface the Royal Society, 10*(84), 20130246.

34. Sun, L., Axhausen, K. W., Lee, D.-H., & Huang, X. (2013, August). Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences, 110*(34), 13774–13779.

35. Simini, F., González, M. C., Maritan, A., & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature, 484*(7392), 8–12.

36. Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., & González, M. C. (2012, January). Understanding road usage patterns in urban areas. *Scientific Reports, 2*, 1001.

37. McNally, M. G. (2008, November). *The four step model*. Irvine: Center for Activity Systems Analysis.

38. Hansen, W. G. (1959, May). How accessibility shapes land use. *Journal of the American Institute of Planners, 25*(2), 73–76.

39. Yang, Y., Herrera, C., Eagle, N., & González, M. C. (2014, January). Limits of predictability in commuting flows in the absence of data for calibration. *Scientific Reports, 4*, 5662.

40. Iqbal, Md. S., Choudhury, C. F., Wang, P., & González, M. C. (2014, March). Development of origin-destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies, 40*, 63–74.

41. Spiess, H. (1990, May). Technical note-conical volume-delay functions. *Transportation Science, 24*(2), 153–158.

42. Samaranayake, S., Blandin, S., & Bayen, A. (2011). Learning the dependency structure of highway networks for traffic forecast. In *Proceedings of the IEEE conference on decision and control* (pp. 5983–5988).

43. Herring, R., Nasr, T. A., Khalek, A. A., & Bayen, A. (2010). Using mobile phones to forecast arterial traffic through statistical learning. *Electrical Engineering, 59*, 1–22.

44. Herrera, J. C., Work, D. B., Herring, R., Ban, X., Jacobson, Q., & Bayen, A. M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies, 18*, 568–583.

45. Jariyasunant, J. (2012). *Improving traveler information and collecting behavior data with smartphones*. PhD thesis.

46. Wang, J., Mao, Y., Li, J., Xiong, Z., & Wang, W.-X. (2015). Predictability of road traffic and congestion in urban areas. *PLoS One, 10*(4), e0121825. doi:10.1371/journal.pone.0121825. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121825.

47. Wang, P., Liu, L., Li, X., Li, G., & González, M. C. (2014, January). Empirical study of long-range connections in a road network offers new ingredient for navigation optimization models. *New Journal of Physics, 16*(1), 013012.

48. Lorenzo, G. D., Sbodio, M. L., Calabrese, F., Berlingerio, M., Nair, R., & Pinelli, F. (2014, January). AllAboard. In *Proceedings of the 19th international conference on Intelligent User Interfaces - IUI '14* (pp. 335–340). New York: ACM Press.

49. Ching, A. M. L. (2012). *A user-flocksourced bus intelligence system for Dhaka*. Diss. Massachusetts Institute of Technology.

50. Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S. H., & Ratti, C. (2014). Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences 111*(37), 13290–13294. http://www.pnas.org/content/111/37/13290.short.

51. Nicolaides, C., Cueto-Felgueroso, L., González, M. C., & Juanes, R. (2012). A metric of influential spreading during contagion dynamics through the air transportation network. *PLoS ONE,* 7, e40961.

52. Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., & Tomkins, A. (2005). Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America, 102*(33), 11623–11628.

53. Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). *Where is the land of opportunity? The geography of intergenerational mobility in the United States.* No. w19843. National Bureau of Economic Research. http://qje.oxfordjournals.org/content/early/2014/10/16/qje.qju022.full#cited-by.

54. Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., & Buckee, C. O. (2013, April). The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society, Interface / the Royal Society, 10*(81), 20120986.

55. Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. L., & Vespignani, A. (2009, December). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America, 106*(51), 21484–21489.

56. Meloni, S., Perra, N., Arenas, A., Gómez, S., Moreno, Y., & Vespignani, A. (2011, January). Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific Reports, 1*, 62.

57. Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: Improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web* (pp. 61–70).

58. Grabowicz, P. A., Ramasco, J. J., Gonçalves, B., & Eguiluz, V. M. (2013). Entangling mobility and interactions in social media. *PLoS One, 9*(3), e92196. http://dx.plos.org/10.1371/journal.pone.0092196.

59. Toole, J. L., Cha, M., & González, M. C. (2012). Modeling the adoption of innovations in the presence of geographic and media influences. *PLoS ONE,* 7(1), e29528.

60. Herrera-Yagüe, C., Schneider, C. M., Smoreda, Z., Couronné, T., Zufiria, P. J., & González, M. C. (2014). The elliptic model for communication fluxes. *Journal of Statistical Mechanics: Theory and Experiment, 2014*(4), P04022.

61. van den Berg, P., Arentze, T. A., & Timmermans, H. J. P. (2010). Size and composition of ego-centered social networks and their effect on geographic distance and contact frequency. *Transportation Research Record, 2135*, 1–9.

62. Kim, S. (1989). Labor specialization and the extent of the market. *Journal of Political Economy, 97*, 692–705.

63. Freedman, M. L. (2008). Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. *Journal of Urban Economics, 64*, 590–600.

64. Yankow, J. J. (2006). Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. *Journal of Urban Economics, 60*, 139–161.

65. Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007, April). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America, 104*(17), 7301–7306.

66. Gurley, T., & Bruce, D. (2005). The effects of car access on employment outcomes for welfare recipients. *Journal of Urban Economics, 58*, 250–272.