Human Mobility

Name of First Author and Name of Second Author

Abstract Mobility is a core component of human behavior and has important impacts on urban and transportation planning, modeling social and biologic spreading, and understanding economic outcomes. Though a rich literature exists with methods to quantify, model, and predict the movements of individuals and populations, access to data has hindered progress. The rise of ubiquitous mobile computing promises to change this. Capable of passively recording large amounts of high resolution spatio-temporal data, there are tremendous new opportunities to study human mobility and its role in social phenomena. Here we present an overview of works utilizing these new data to understand and model human movement. We begin by discussing the historic roots of research in this area and the new data sources now available to it. Next, we present recent works that take inspiration from statistical physics, machine learning, and network theory and apply them to to modeling social phenomena related to human movement at the individual and aggregate levels.

1 Introduction

Mobility has been a steering force for much of human history. Whether it was tribal mixing and migration impacting population genetics or great explorers discovering new worlds and trade routes, the movement of peoples determines the dynamics of numerous social and biological processes. Humanity now finds itself at the center of another great migration, this time into cities and urban regions. In 2008 the United Nations reported that for the first time, more than 50

Name of First Author
Name, Address of Institute, e-mail: name@email.address

Name of Second Author
Name, Address of Institute e-mail: name@email.address

Urban and transportation planners have long been interested in the flow of vehicles, pedestrians, or goods from place to place. They use these insights to identify critical points in infrastructure, how urban form influences function, and how to optimize public services such as busses or subways. Epidemiologists rely heavily on models of human movement to predict and prevent disease outbreaks. As air travel makes it possible for viruses to quickly jump continents and dense urban spaces facilitate human-to-human contagion, these models must capture considerable complexity. Disease is not the only contagion affected by the spatial dynamics of people. Information, in the form of ideas or gossip, is spread from person to person in offices and cafes across the world. Social scientists have theorized that these interactions are at the heart of culture that impacts crime rates, social mobility, and economic growth.

Despite the rich set of questions and theory developed thus far, progress testing and verifying them has been slowed by a lack of reliable and accessible data sources. Traditional sources of data on human mobility are often prohibitively expensive surveys that are plagued by small sample sizes and bias. Given these impediments, researchers have developed numerous statistical methods to carefully treat this data and account for biases and allow for generalizability.

The past decade has given rise to ubiquitous mobile computing wherein billions of individuals are able access people, goods, and services through telecommunication devices such as mobile phones and applications that run on them. The penetration of these devices is astounding. The six billion mobile phones currently in use is nearly triple the estimated two billion internet users. These devices and applications passively record the actions, social connections, and movements of their users with high spatial and temporal resolution. While the collection, storage, and analysis of this data presents very real and important privacy concerns, it also offers an unprecedented opportunity for researchers to learn about human behavior. Various location a services leverage cellular antennas, wifi access points, and GPS receivers to measure the geographic position of users to within a few hundred meters or less. With billions of data points captured on millions of users each day within a city, new research into computational social science [23] has begun to augment and replace sparse, traditional data sources, answering old questions and raising new ones.

In this chapter, we present an overview of mobility research in this new, data rich environment. We begin by briefly describing a variety of new sources of mobility data then detail statistical models that describe them. We first focus on models of individual mobility that have consistently demonstrated three important characteristics of human mobility; we are slow to explore new places, predictable, and unique. We also discuss efforts to add context to observed movement patterns and motifs and infer a location's semantic meaning to an individual. Finally, we present models that describe aggregates of human mobility such as the the flow of people from place to place. Throughout this chapter, we point out specific applications of these models to real systems like congestion management, economic growth, and the spreading of both information and disease.

2 New Data Sources

Traditional data sources for human mobility range from census estimates of large scale migrations and commuting to travel diaries filled out by individuals. These surveys are generally expensive to administer and participate in as they require intensive manual data recording and transcription. They also require that individuals recall large amounts of information on when, where, and how they have traveled that may be prone to mistakes and biases. The results are surveys that cover only a day or week at a time and collected from small samples of the population (often less than 1

Wireless carriers, however, are far from the only entity making use of location information. Applications that run on these devices can also access location data, opening the door for sophisticated mapping and navigation services. A variety of sensors from GPS to wifi allow devices to pinpoint a user's location to within just a few meters. Protocols like bluetooth and NFC allow users to discover and connect to other devices within a few meter radius, creating ad hoc sensor and proximity networks. Other applications such as social networks, incorporate geographic information both implicitly and explicitly. Many networks allow individuals to list their home cities, but connect to users anywhere, revealing the relationship between social ties and geography. Other networks incorporate mobility more explicitly such as Twitter, which allows users to geotag tweets with precise locations, or FourSquare that invites users to 'check-in' at establishments and broadcast their location to friends.

At the same time that users are sharing their movements on through applications, there has also been an explosion of other geographic data layers thanks to the digitization of maps. The advancement of GIS software and falling data storage prices make it possible for small and large cities to offer their public mapping data to citizens in an online format. Efforts like the U.S. Census Bureau's TIGER-line repository, San Francisco's OpenSF, and New York Citiy's PLUTO databases offer quantities of publicly accessible geographic data on everything from building footprints and the location of individual trees to the exact path of highways and administrative boundaries. Private efforts such as Google Maps are giving individuals and planners access to high resolution satellite imagery, route planning, and point of interest information through free or low cost APIs while open and crowd sourced initiatives like OpenStreetMap allow anyone in the world to contribute to and download digital maps of roads, buildings, subways, and more. Put together, these resources provide a digital map of the world that acts as a rich backdrop on which to study human mobility and the infrastructure built to facilitate it on a planetary scale.

Infrastructure is getting smarter as well. Toll booths can count and track cars while mapping and navigation applications synthesize data from thousands of travelers to build accurate, real-time traffic estimates and provide routing information accordingly. Subways, busses, taxis, and bike shares are increasingly moving to electronic fare systems that record when millions of users enter and exit transportation systems. These automatic vehicle location (AVL) and systems often post live feeds of where vehicles are and predictions of system performance to help users better plan trips. This data also helps the operators of these systems better plan for and respond to demand, reducing costs and increasing system performance. For example, Uber, an on-demand car service, uses historical usage data to minimize the time a user has to wait for a car to arrive and the time that cars spend empty without passengers, improving system efficiency for both riders and drivers. Applications such as NextBus, that aggregate realtime vehicle location feeds from hundreds of public transit entities to display and predict when the next bus will arrive make unreliable transit systems easy to use.

Put together, these new data assets are enabling a wealth of new opportunities to study human mobility and offer enormous potential to provide the knowledge and insights back to their creators. Not only do they collect data at a scale and richness that is prohibitively expensive to match via traditional methods, this data is also robust to manipulation by conscious or unconscious biases. The scope of surveys is limited by the ability of individuals to remember details of when and where they were weeks or months ago and it requires too much effort to manually enter data for long periods of time. Digital devices, however, passively record this information without any effort required from the user and provide a signal that is difficult to fake. While immense and rich, this data is not without pitfalls. Old biases and hidden variables do not disappear simply because the data is large and new ones are often introduced. The challenge and ultimate goal is to use this data to expand upon and augment our current understanding of human mobility in a way that makes use of these new tools while still providing confidence in their conclusions.

3 Individual Mobility Models

Understanding mobility at an individual level entails collecting and analyzing sets of times, places, and semantic attributes of how users travel between them and what they are doing when they get there. For example, on a typical morning I may wake up at home, walk to a local cafe to pick up coffee on the way to the subway that I take to work. After work I may go to the grocery store or meet a friend for dinner before returning home only to repeat the process the next day. Our goal is to understand underlying patterns in this mobility and the mechanisms that give rise to them using the new, high resolution data now available.

Early work has drawn a great amount of inspiration from statistical physics, modeling mobility as random walk or diffusion processes. For instance, a crowdsourcing project known as 'Where's George' asked volunteers to enter the geographic location and serial number of dollars bills in order to build up a travel history of various banknotes. Because bills are primarily carried by people as they travel from store to store, a note's movement can serve as a proxy for human movement. By applying mathematics for describing continuous time random walks to the trajectories of the bills, researchers found that their movement appears to follow a Levy flight process, a random walk characterized by a significant (fat-tailed) probability of making long jumps. These findings are aligned with common sense observations that humans tend to make many short trips in a familiar area, but also take longer journey's from time to time.

However, the movement of these bills does not tell the whole story. Using a dataset containing call detail records (CDRs) for more 100,000 users collected over a 6 month period in a European country, Gonzalez et al showed that human mobility patterns are more subtle than previous studies had suggested. The distribution of displacements for the entire population was well 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 cut-toff distances between 80km and 400km. To understand the mechanism that gives rise to this distribution, the authors borrowed a quantity, radius of gyration r_g, from polymer physics:

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

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 over an observation period t. Examining the distribution of r_g in the population, Gonzalez et al again found it 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, but that it could not be explained by the assumption that all individuals moved according to the same Levy flight process. Instead, the authors found that this distribution is driven by the convolution of a heterogeneous population of Levy flight distributions, 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 Levy flight process up to trips of some individual specific characteristic distance r_g, after which the probability of long trips drops far faster than would be expected from a purely random process.

Further investigation revealed the source of this behavior. Human movement, unlike random processes, feature frequent returns to previously visited locations such as home or work. The nature of these returns was found to follow a very particular pattern. An individual returns to a previously visited location with a probability proportional to that location's rank (relative to that user's other visits) $P(L)\ 1/L$. These non-random, predictable return visits are unaccounted for in random walk and Levy flight models and are at the heart of deviations of observed behavior from random processes. Finally, Gonzalez et al examined the probability density functions of individuals' positions in space, $\Theta(x,y)$. Diagonalizing this distribution by each individuals inertial tensor, the distributions of persons in different cities can be compared despite differences in local geography or orientation. Interestingly, these rectified distributions display directional dependence or anisotropy the ratio of which $S = \sigma_x/\sigma_y$ depends on the radius of gyration of an individual. Removing this anisotropy by dividing each spatial direction by its standard deviation causes the spatial probability density function of individuals $\Theta(x/\sigma_x, y/\sigma_y)$, with any radius of gyration to collapse to the same shape collapse onto the same shape.

Additional studies have measured similar quantities using data from CDRs as well as location based check in services such as Foursquare or Twitter, finding similar results [5] These approaches show that human mobility can be studied en masse via data collected from mobile phones and that it can be modeled as the convolution of heterogenous Levy flight distributions, that the probability an individual returns to a previously visited location is well approximated by a power-law, and that proper scaling of spatial distribution function reveals similar underlying behavior. Taken together, they demonstrate the effectiveness of applying models inspired by statistical physics to human movement.

Subsequent work has built off of this tradition. Song et al [38] demonstrate a three important of scaling laws of individual human mobility patterns (again using mobile phone data). First, the number of unique locations visited by mobile phone $S(t)$ users scales sublinearly with time $S(t)\ t^{\mu}$ where $\mu = 0.6$. Second, the probability a user returns to a previously visited locations (again) scales with the inverse of the rank of that location $P(L)\ L^{-\zeta}$ where $\zeta = 1.2$. And third, the mean displacement of an individual from a given starting point shows slower than logarithmic growth, demonstrating the extremely slow diffusion of humans.

Song et al [38] then propose a simple model of human mobility that is capable of capturing these three empirical regularities. Starting at time t, an individual will make a trip at some future time Δt drawn from a fat-tailed probability distribution. With probability $\rho S^{-\gamma}$, the individual travels to a new, never-before visited location some distance Δr away, where Δr is again drawn from a fat-tailed distribution characterized in previous models. With probability $1 - \rho S^{-\gamma}$ an individual returns to a previously visited location that is selected based on the inverse rank equation identified previously. Though this simple model is capable of reproducing many statistical and scaling properties of human mobility, it does 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).

In addition to these analytic and generative models of human mobility, there is another deep literature that uses machine and statistical learning to model and predict mobility in the future. Using the same mobile phone dataset, Song et al [39] applied principals from information theory to show that the upper bound on predictability of a typical individual's movements is a surprisingly high 93% and found that no user displayed a potential predictability of less than 80%. Moreover, this work introduced a new metric, regularity $R(t)$, defined as the probability a user if found at their most visited location during a given hour t, and the quantified the number of unique locations visited during a typical hour of the week $N(t)$. These quantities have routinely been measured in different data sets in different cities and countries and are consistent among them [5].

While the work of Song et al places theoretic bounds on the maximum predictability of human mobility, the authors do not speculate on which algorithms achieve this. To that end, a number of statistical learning techniques developed for and applied to mobility data at various spatial and temporal scales. Some of the first studies in this area, predating even analytic computations, use Markov and hidden Markov models and information on underlying transportation networks to predict transition between mobile phone towers within cities to improve quality of service of wireless networks through proper resource allocation [21, 26, 41, 22]. At smaller scales, researchers have used various trajectory estimation and Kalman filtering algorithms to predict movements in smaller spaces such as college campuses [29, 24]. Cho et al [7] exploit the temporal periodicity of human mobility in their Periodic Mobility Model (PMM) as well as the propensity of mobility to involve social behavior in a Period Social and Mobility Model (PSMM). At their core, these models amount to mixture models in two-dimensional space that attempt to learn probability distributions that a user will be at any given location at a given time from previous location data and, in the case of the PSMM, the location history of social contacts. Multivariate nonlinear time series forecasting has produced similar results [12, 35] predicting where an individual will be in the next few hours or at a given time of a typical day. Finally, principal component analysis has been applied to extract typical spatiotemporal patterns of users that can be used to predict future behavior. In one of the first studies to mine the behavior of college students using mobile phones, Eagle et al found that an individual's behavior could be represented as a linear combination of just a few "eigenbehaviors", which are represented as temporal vectors whose components represent activities such as being at home or work. Not only do these typical behavioral profiles change based on the type of student being examined, they can also be used to predict future behaviors, label social interactions, and perform long range (on the order of months and year) forecasts of mobility [13, 14, 32].

It is tempting to assume that such predictability is the result of high levels of similarity between all individuals in a region. Perhaps the pull of monocentric downtowns or the structure of transportation systems funnel users to the same places and route choices. de Montjoye et al [10] explored this hypothesis and found that, while predictable, an indiviuals movement patterns are also unique. Using CDR data on 1.5 million users, the authors measure how unique mobility traces of an individual are within a population. They begin by defining a trace T as a set of spatio temporal points, each containing a location and a timestamp. Sets of randomly chosen spatio temporal points I_p are said to be compatible to a trace if $I_p \subseteq T$. The traces of each user are then compared to all possible sets of randomly chosen points and the unicity of a trace is measured by the number of other users who are also compatible with that set. Applying this measure to the entire dataset, the authors find that just four spatio temporal points is enough to uniquely identify 95% of all users. These results raise many questions about the privacy of massive, passively collected datasets, but also highlight an incredibly interesting nuance of human mobility: though individuals are predictable, they are also unique.

Thus far, much of the work discussed has focused on concrete measures of mobility and has treated two different geographic locations as semantically different places. If two individuals are not at the same place at the same time, their movement patterns are different. Recent work, however, as began to abstract from these constraints and explore how we move between places that form similar semantic functions such as home or work. Schneider et al. [36] introduced the concept of mobility motifs by examining abstract trip chains over the course of a day. A daily mobility motif is defined by some number of locations to visit and a particular order that a person visits them over the course of a day. More formally, these motifs constitute directed networks where nodes are locations and edges represent 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). On average, they find individuals visit 3 different places in a given day. They next construct all possible mobility motifs for a given number of locations n (this set consists of all directed networks of size n where all nodes are reachable from at least one starting node) and compute the frequencies that those motifs appear in human mobility data from call detail records. Shockingly, despite over 1 million ways to travel between 6 or fewer locations, 90

Like recurring daily trips, recurring daily social encounters were also the subject work by Sun et al [40] that used data from bus passengers in Singapore to track repeated co-locations between passengers. The authors found that an average individual encounters roughly 50 people per trip and that these trips are highly periodic, occurring at intervals associated with working hours as well as daily and weekly trips. Temporal encounter networks were constructed by connecting individuals if they rode the same bus at the same time. A pair of individuals who have encountered each other can expect to meet 2.5 times over the course of a week. Again, the probability of these encounters was periodic, 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. These results shed light on the surprising familiarity that we may have with strangers during our daily movements.

In summary, new data sources are shedding new light on the movement patterns of individuals over numbers spatial and temporal scales. At the city scale, over weeks and months, human movement is characterized by slow exploration, preferential return to previous visited places, and predictable uniqueness. These qualities has been exploited by 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. More abstract explorations of daily mobility motifs and encounters with strangers show even more predictability in how we travel and who we travel with, strangers or not. Despite the nearly endless number of places to visit and order to visit them in, we choose only a few and repeat these choices each day. Individual mobility patterns, however, are not the only level of granularity of interest to researchers. In the next section we will discuss works with the goal goal of describing aggregate movement and flows of many individuals from place to place.

4 Aggregate Mobility

For the purposes of planning urban spaces, optimizing transportation networks, studying the spread of ideas or disease, it is often necessary to aggregate the movement of individuals to compute flows of people from place to place. For example, origin-destination (OD) matrices serve to store the number of people traveling from any location to any other and are critical to transportation planning. Sometimes, constructing these aggregate mobility statistics is simply a matter of summing the movements of each individual, but, as is often the case for complex systems, collective behavior emerges that can described by its own set of models. One widely used approach to modeling aggregate mobility behaviors is the ”four step model”, made of the following stages: 1) trip generation, 2) trip distribution, 3) mode choice, and 4) route assignment [27, 31].

In the first stage, the number of individuals traveling from each zone is typically estimated using information on socioeconomic demographics and land use then these trips are connected to destination points using some distribution algorithm. The specific mode shares (e.g. foot, bus, or car) are computed in the mode choice stage and finally routes are assigned by applying a path finding algorithm to the transportation network specific to a chosen mode. Like individual mobility models, aggregate models at each step are often inspired physical processes.

Some of the earliest techniques for estimating OD matrices (the first two stages of the four step model) are gravity models that borrow directly from Newton's law of gravitation. The number of trips T_{ij} taken from place i to place j is treated as a function of the population of each place m_i and m_j and some function of the the distance between them $f(r_{ij})$. In the fully parameterized version of this model, $T_{ij} = a\frac{m_i^{\alpha} m_j^{\beta}}{f(r_{ij})}$. One recovers the classical gravity model from physics by setting $\alpha = \beta = 1$, and $f(r_{ij}) = r_{ij}^2$. Typically, though, these parameters are generally calibrated for each specific application using survey data. The intuition behind such models is that the population of a place, or it's mass, is responsible to generating and attracting trips and thus the total flux between the two places should be proportional to some product of the two masses while the distance between them mitigates the strength of this connection.

Though used heavily for decades, such models have their limitations. Simini et al [37] identify a number of shortcomings with the gravity model, the main two being the large number of parameters needed for calibration (often leading to overfitting) and the failure of the model to account for density. For example, the flow of people between two locations will be the same whether or not there is an entire city between the two or an empty desert. Intuitively, one would expect that trips between places would be affected by what is between the two. To capture this, the authors again borrow from a physical process, this time radiation and absorption. Termed the ”radiation model”, we imagine individuals being emitted from a place at a rate proportional to it's population and being absorbed by other locations at a rate proportional to the population there. The probability that an emitted person arrives at any particular place is then related to the probability they are not absorbed before getting there. Mathematically, this process results in the following function for computing the number of trips between two locations: $T_{ij} = T_i \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}$, 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 first striking fact about this formula is that is does not directly depend on the distance between the two places. Instead, it is the opportunities between the two places that determines the flux of trips between them. Second, the formula is non-parametric and does not require calibration for different regions. Simini et al go on to show that the radiation model provides superior estimates of OD flows than the gravity model in a variety of contexts related to transportation, commuting, and information.

While the radiation model is an improvement over the gravity model in many respects, it has proven most appropriate for areas the size of counties or larger. Within cities, its accuracy decreases. Because the model implicitly accounts for differences in population density, it breaks down in situations where density is more uniform and distances are small. For example, 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. However, the introduction of a single scaling parameter α in the function describing the conditional probability an individual is absorbed at a location was shown to be enough to correct for these distortions and provide a model that works on any length scale [48, 20].

Activity based models can serve as alternatives to the four step model or as a tool during one of the stages [4]. In these models, it is assumed that all trips are made in order to fulfil certain needs or desires of an individual. The goal, then is to identify what those needs are for different segments of the population and how they are typically fulfilled. For example, one might construct a logit model of mode choice using data from travel surveys and diaries that predicts which type of transportation will be used for a trip given the demographics of an individual and environmental factors. That model can then be applied to other areas of the city where demographic and environmental data is provided. These models are closely related to agent based models that attempt to simulate the behaviors of city residents on the micro-scale and aggregate them to measure macro performance of transportation systems.

The final step of the four step model is route assignment. Given that a number of individuals are estimated to travel between two points using a particular mode, 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 their performance. For example, given the huge number of possible routes that a commuter could drive in the morning, which do they take? These estimates are often made assuming that individuals following shortest paths that minimize some cost function, typically related to travel time and distance. In the case of a road network, speed limits can be combined with segment lengths to estimate free flow travel times. Shortest path algorithms such as Dijkstra's algorithm, which finds shortest paths between all pairs of nodes in a network, or the popular A-star algorithm, which is a "best-first" search algorithm that relies on local information and heuristic functions, are then used to find a drive a route that minimizes travel times.

In most cases, however, drivers rarely encounter completely empty freeways. To model the impact of congestion on network costs, iterative traffic assignment (ITA) algorithms are used. First, trips are split into segments containing only a fraction of total flow between two points (typically 40%, 30%, 20%, and 10%). Trips in each segment are then routed along shortest paths independently of all other trips in that segment while counts are kept of how many trips were assigned to use each road segment. At the end of each iteration, the travel times are adjusted according to a BPR function that takes into account the current congestion on a road segment given by the ratio between the volume of traffic assigned to a road thus far and the rated capacity of the road. As roads because more congested, the travel times quickly increase causing drivers in later iterations to be assigned to different, less congested routes. At the end of assignment, one has both total volume on each road segment, as well as congestion and updated travel times. These values can then be validated against traffic counters, speed sensors, or data from vehicle fleets like taxis and busses.

While a variety of data sources were traditionally used at each step of the four step model, the current explosion of new data resources has dramatically reduced the cost of increased the efficiency of validation. The Mobile Millenium project, for example, first began combining real time GPS data from smart phones with physical sensors from road networks and data feeds from vehicle fleets such as taxis and busses to understand and predict traffic flows [33, 19, 18]. Subsequently, Wang et al [44] showed it is possible to generate valid road usage estimates using historic CDR data. Their procedures first scale mobile phone activity in areas to correct for differences in market share, then measure trips by counting consecutive phone calls of individuals as they move through the city. These trips are then routed using an ITA algorithm like the one described above, giving road usage patterns for a metro-region. Given the availability of CDR data and high quality road network information (thanks to crowdsourced initiatives like OpenStreetMap), this analysis can be replicated in many cities, allowing us to compare road usage patterns across cities.

Using this approach, Wang et al show that the distribution of traffic volume and congestion (measured as volume over capacity or VOC) following distributions that are well fit by an exponential mixture model that depend on the number major and minor roadways in a cities network. Beyond just looking at traffic and congestion distributions, the authors describe usage patterns of drivers by constructing a bipartite usage graph that connects locations in the city to roads used by those travelers. Universal patterns are found in the degree distributions of the projected networks of each node type. Comparing the typological properties of roads as measured by their betweenness centrality to their behavioral importance given by the number of unique locations that contribute traffic to a road, Wang et al are able to classify the function of a road. For example, a bridge may be topologically important because it is the only way to cross a river, but a mainstreet 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 a small fraction (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. Similar work shown that this analysis can be used to predict traffic jams [43, 45].

In addition to road networks, other new forms of transportation such as the rise of global air travel are changing the way we think about human mobility. Individuals can now travel between nearly any two cities on the globe in a matter of hours. Using network theoretic approaches, the structure of travel between cities has been studied to reveal a hub and spoke logistics network that seeks to minimize the diameter (maximum number of hops between any two points) of the networks while also minimizing the cost associated with operating flights [15]. The resulting network has topological properties similar to the scale-free networks seen in social systems, but with a community structure that arises from spatial and geopolitical constraints [17].

New data sources are generating opportunities to map aspects of human mobility that have gone previously unstudied. For example, in rapidly developing cities like Dhaka, public transit systems are often inadequate, unfinished, or underfunded. Extensive informal transit networks, often privately owned mini-buses operating on crowdsourced routes shuttle thousands of people around the city each day. Using smart devices, researchers are now able to map these routes, creating the first ever schedules and route diagrams of transit services in the world's largest cities [11, 6, 34]. Even in cities like New York and San Francisco, mobility is rapidly changing. Data from automatic vehicle location (AVL) and smart fare cards on public transit are empowering a new generation of applications that make it easier to use public transportation services. Car sharing services such as Uber and ZipCar are all but eliminating the need to own a car. Powered by data from mobile applications, sophisticated demand modeling and optimization algorithms ensure that residents can have mobility on-demand.

The ubiquity of mobile devices collecting data on movement extends across cultures and continents. Analysis that uses mobile phone data to measure road network performance can be repeated in nearly any city from San Francisco to Rio de Janeiro to Beijing. While these comparisons highlight differences in some mobility characteristics, they also point to a number of similarities that suggest some underlying mechanism is at work. For example, levels of traffic, congestion, and the ways in which places contribute traffic to roads can all be described by simple statistical distributions [44].

While the study of human mobility is of immediate interest to travelers and the transportation engineers who build and maintain the systems that facilitate them, the movement of people has enormous direct and indirect roles in a number of other human behaviors. While not nearly an exhaustive list, we highlight three such areas here: social behavior, disease and information spread, and economic outcomes. Many of these dynamics will be discussed in greater detail in further sections of this volume.

4.1 Mobility and Social Behavior

While many trips are made to performed alone to complete individual tasks, a huge fraction of trips surround people. Inspite of new communications technologies that make 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, movement is a means to a social end. The link between social contacts and mobility has becoming increasingly prominent in research as it is often through location based social networks or communication devices that mobility data is collected. Liben-Nowell first measured the empirical fact that the probability of being friends with another individual on an online social network decreases with the distance between those two individuals at a rate inversely proportional to the distance between them [25]. Toole et al [42] showed that understanding and incorporating the geographic structure of social networks is critical to predicting how information will spread through them.

Subsequent work verified these findings in other social networks [1, 16] and found that social contacts were very useful in predicting where an individual would travel next [7, 12]. Cho et al [7] finde that 50%-70% of mobility can be explained as periodic behavior, with another 10%-30% related to social interactions. Analytic models have also been proposed to incorporate these dynamics. For example, Grabowicz et al [16] incorporate these features by having individuals travel in a continuous 2D space in a model similar to the model described above and in [38], but here an individual travel's is determined by the location of their contacts, instead of their own previous visits. With probability p_v (a free parameter in the model), an individual moves to the location of a friend and with probability $1-p_v$, they choose a random point to visited some distance Δr away (again choosing the jump distance from a fat-tailed distribution). Upon arriving at a new location, the individual can choose to form social ties with other individuals within a radius with probability p (fixed at $p = 0.1$ empirical measurements) or random individuals anywhere in the space with probability p_c (another free parameter). Again, a simple model is able to reproduce many empirical relationships found in social and mobility data.

4.2 Mobility and Disease Spread

Just as cars, trains, and planes transport humans around the globe, people serve as vectors for diseases. The properties of human travel through air networks, for example, has become the central focus of those studying global epidemics. Whereas historic pandemics such as the bubonic plague that swept through Europe in the 14th century traveled roughly at the speed of horse drawn carriages and migration by foot [8], modern flight allows an individual to travel between nearly any two points on the globe in a matter of hours, sometimes bringing disease with them. It is the global airline network that determines, at a high level, how potent an epidemic will be and its likely path across the globe [30, 2, 3, 28, 9]. Mobile phones are extending these studies to smaller and smaller scales at with higher resolution spatial data. For example, by using CDR data to map mobility patterns in Africa, researchers have been able to improve vaccine distribution strategies and better allocate scarce resources to fight Malaria and other diseases [47, 46]. Richer data available at lower cost, will continue to boost efforts to contain these outbreaks and mitigate their effects.

4.3 Mobility and Economic Outcomes

Mobility is often a means to an end. It provides access to things and people that we need to be happy, healthy, and productive. With this in mind, economists and other social scientists have hypothesized the role of human mobility in socio-economic outcomes and economic growth. In these theories, mobility is a means to unlock human capital that makes us productive and successful. For example, jobs in dense cities tend to pay higher wages than the same jobs in more rural areas even when controlling for factors such as age and education. Economists theorize this is due to productivity and creativity gains made possible by the rich face to face interactions that close spatial proximity facilitates. Meetings with strangers in coffee shops or with clients across town offer more opportunities to share of mutually beneficial information. While density is one way of propagate these benefits, increased mobility is another. Poorer residents of cities have better job prospects and higher chances of retaining jobs when given a personal car instead of being constrained geographically (and temporally) by public transit.

Access to space provides opportunities not only to greater numbers of jobs, but also improves the chances a person will find a job suited best for their particular skills, allowing them and their employer to achieve more. Moreover, these opportunities can have large and lasting effects on individuals and even their children. For example, a recent study used extremely rich data from tax records to observe millions of children for a number of decades with the goal of understanding intergenerational economic mobility. They found that, despite popular opinion to the contrary, income inequality has not increased significantly in the past 40 years. Instead, it has become more geographically segregated, with some regions of the country providing far better opportunities to children than other. While the authors do not claim causal relationships, they find some of the strongest correlations between intergenerational economic mobility and variables related to the commuting times and spatial segregation of people. Though many theories exists that may explain these correlations, testing them has been prohibitively expensive due to the need for reliable data on micro interactions between people and their surroundings. Until recently, data on these encounters has been exceedingly difficult to collect and analyze. However, the mobile devices we carry and use daily are designed, explicitly in some cases, to passively capture exactly these behaviors. Using this data to study human mobility and, more broadly, human behavior presents an enormous opportunity for researchers.

5 Conclusion

In this chapter we have reviewed a number of ways that new data sources are expanding our understanding of human mobility. The patterns of our movement can be are well modeled by drawing analogies from statistical physics and are characterized by periodic and preferential return to previously visited locations, high predictability, and uniqueness. Abstracting space and time with concepts such as daily mobility motifs reveal that though individuals may visit vastly different places over the course of the day, the number, order, and semantic meaning of these movements is shockingly consistent and stable within populations. Aggregating the movement of individuals into origin-destination matrices is critical for building, maintaining, and optimizing the infrastructure that enables it. With new data sources offering near realtime data streams containing information on the state and performance of transportation systems, engineers are better equipped than ever to improve them. At the same time, this data is giving a new lens through which to view social phenomena such as information and disease spread and how these things affect socio-economic outcomes. As the cost of smart devices and sensors falls and they become ever more ubiquitous, the opportunity to improve the lives of people across the globe will only grow larger.

References

1. Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*, pages 61–70, 2010.
2. Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. 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):21484–9, December 2009.
3. Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. 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):21484–9, December 2009.
4. Moshe E. Ben-Akiva and Steven R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, 1985.
5. Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Z. Sui. Exploring Millions of Footprints in Location Sharing Services. In *ICWSM*, pages 81–88, 2011.
6. Albert M. L. (Albert Man Loon) Ching. A user-flocksourced bus intelligence system for Dhaka, 2012.
7. Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining KDD 11*, KDD '11, page 1082. ACM Press, 2011.
8. Samuel K Cohn. Epidemiology of the Black Death and successive waves of plague. *Medical history. Supplement*, (27):74–100, January 2008.
9. Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. 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):2015–20, February 2006.

10. Yves-Alexandre de Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the Crowd: The privacy bounds of human mobility. *Nature Scientific Reports*, 3:1376, 2013.
11. Giusy Di lorenzo, Marco Luca Sbodio, Francesco Calabrese, Michele Berlingerio, Rahul Nair, and Fabio Pinelli. AllAboard. In *Proceedings of the 19th international conference on Intelligent User Interfaces - IUI '14*, pages 335–340, New York, New York, USA, February 2014. ACM Press.
12. Manlio De Domenico. Interdependence and Predictability of Human Mobility and Social Interactions. *csbhamacuk*, 2012, 2012.
13. Nathan Eagle and Alex Pentland. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, 10:255–268, 2006.
14. Nathan Eagle and Alex Sandy Pentland. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63:1057–1066, 2009.
15. Michael T Gastner and M. E.J. Newman. The spatial structure of networks. *The European Physical Journal B*, 49(2):247–252, 2006.
16. Przemyslaw A. Grabowicz, Jose J. Ramasco, Bruno Goncalves, and Victor M. Eguiluz. Entangling mobility and interactions in social media. page 16, July 2013.
17. R Guimerà, S Mossa, A Turtschi, and L A N Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22):7794–9, May 2005.
18. Juan C. Herrera, Daniel B. Work, Ryan Herring, X. Ban, Quinn Jacobson, and Alexandre M. Bayen. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, 18:568–583, 2010.
19. Ryan Herring, Tania Abou Nasr, Amin Abdel Khalek, and Alexandre Bayen. Using Mobile Phones to Forecast Arterial Traffic through Statistical Learning. *Electrical Engineering*, 59:1–22, 2010.
20. Md. Shahadat Iqbal, Charisma F. Choudhury, Pu Wang, and Marta C. González. Development of origindestination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74, March 2014.
21. H.S. Kim. QoS provisioning in cellular networks based on mobility prediction techniques. *IEEE Communications Magazine*, 41(1):86–92, January 2003.
22. John Krumm, Eric Horvitz, Paul Dourish, and Adrian Friday. Predestination: Inferring Destinations from Partial Trajectories. *UbiComp 2006: Ubiquitous Computing*, 4206:243–260, 2006.
23. David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. Computational Social Science. *Science*, 323(5915):721–723, 2009.
24. Kyunghan Lee Kyunghan Lee, Seongik Hong Seongik Hong, Seong Joon Kim Seong Joon Kim, Injong Rhee Injong Rhee, and Song Chong Song Chong. SLAW: A New Mobility Model for Human Walks. *IEEE INFOCOM 2009*, 2009.
25. David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33):11623–11628, 2005.
26. Tong Liu, Paramvir Bahl, and Imrich Chlamtac. Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks. *IEEE Journal on Selected Areas in Communications*, 16:922–935, 1998.
27. Michael G. McNally. The Four Step Model. *Center for Activity Systems Analysis*, November 2008.
28. Sandro Meloni, Nicola Perra, Alex Arenas, Sergio Gómez, Yamir Moreno, and Alessandro Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports*, 1:62, January 2011.
29. Kim Minkyong, David Kotz, and Kim Songkuk. Extracting a mobility model from real user traces. In *Proceedings - IEEE INFOCOM*, 2006.

30. Christos Nicolaides, Luis Cueto-Felgueroso, Marta C. González, and Ruben Juanes. A metric of influential spreading during contagion dynamics through the air transportation network. *PLoS ONE*, 7, 2012.
31. Juan de Dios Ortúzar and Luis G. Willumsen. *Modelling Transport*. 2011.
32. Adam Sadilek and John Krumm. Far Out: Predicting Long-Term Human Mobility. *AAAI*, pages 814–820, 2012.
33. Samitha Samaranayake, Sebastien Blandin, and Alexandre Bayen. Learning the dependency structure of highway networks for traffic forecast. In *Proceedings of the IEEE Conference on Decision and Control*, pages 5983–5988, 2011.
34. Paolo Santi, Giovanni Resta, Michael Szell, Stanislav Sobolevsky, Steven Strogatz, and Carlo Ratti. Taxi pooling in New York City: a network-based approach to social sharing problems. page 12, October 2013.
35. Salvatore Scellato, Mirco Musolesi, Cecilia Mascolo, Vito Latora, and Andrew T. Campbell. NextPlace: A spatio-temporal prediction framework for pervasive systems. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 6696 LNCS, pages 152–169, 2011.
36. Christian M Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C González. Unravelling daily human mobility motifs. *Journal of the Royal Society, Interface / the Royal Society*, 10(84):20130246, 2013.
37. Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):8–12, 2012.
38. Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, September 2010.
39. Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
40. L. Sun, K. W. Axhausen, D.-H. Lee, and X. Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34):13774–13779, August 2013.
41. Arvind Thiagarajan, Lenin Ravindranath, Katrina LaCurts, Samuel Madden, Hari Balakrishnan, Sivan Toledo, and Jakob Eriksson. 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*, pages 85–98, 2009.
42. Jameson L Toole, Meeyoung Cha, and Marta C González. Modeling the adoption of innovations in the presence of geographic and media influences. *PLoS ONE*, 7(1):e29528, 2012.
43. Jingyuan Wang, Yu Mao, Jing Li, Chao Li, Zhang Xiong, and Wen-Xu Wang. Predictability of road traffic and congestion in urban areas. July 2014.
44. Pu Wang, Timothy Hunter, Alexandre M Bayen, Katja Schechtner, and Marta C González. Understanding road usage patterns in urban areas. *Scientific reports*, 2:1001, January 2012.
45. Pu Wang, Like Liu, Xiamiao Li, Guanliang Li, and Marta C González. Empirical study of long-range connections in a road network offers new ingredient for navigation optimization models. *New Journal of Physics*, 16(1):013012, January 2014.
46. Amy Wesolowski, Nathan Eagle, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee. 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, April 2013.
47. Amy Wesolowski, Nathan Eagle, Andrew J Tatem, David L Smith, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee. Quantifying the impact of human mobility on malaria. *Science (New York, N.Y.)*, 338(6104):267–70, October 2012.
48. Yingxiang Yang, Carlos Herrera, Nathan Eagle, and Marta C González. Limits of predictability in commuting flows in the absence of data for calibration. *Scientific reports*, 4:5662, January 2014.