

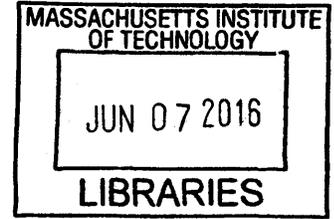
Navigating Congested Cities: Understanding Urban Mobility Using New Data Sources

by

Serdar Çolak

B.S., Boğaziçi University (2008)

M.Sc., Boğaziçi University (2010)



ARCHIVES

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

© Massachusetts Institute of Technology 2016. All rights reserved.

Signature redacted

Author

Department of Civil and Environmental Engineering
April 29, 2016

Signature redacted

Certified by

Marta C. González
Associate Professor
Thesis Supervisor

Signature redacted

Accepted by

Heidi Nepf

Donald and Martha Harleman Professor of Civil and Environmental
Engineering
Chair, Graduate Program Committee

Navigating Congested Cities: Understanding Urban Mobility Using New Data Sources

by

Serdar Çolak

Submitted to the Department of Civil and Environmental Engineering
on April 29, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Transportation

Abstract

Rapid urbanization and technological innovations sparked the generation of massive volumes of data that is continually improving in resolution. In particular, mobile phones, having reached penetration levels above 97% in Europe and Americas according to the World Bank, transformed into passive sensors of urban mobility by signaling movement at the individual level. The data generated by these devices has a wide range of applications concerning how people and cities interact through the infrastructure. This thesis presents new analysis tools that utilize large geolocated datasets to provide new insights towards human mobility, road networks, congestion, and energy. In the first part of this work, we analyze the emergence of vehicular congestion in an urban road network through the use of a simple traffic flow model. We show that spatial constraints and the topology of the road network are determinant factors that shape the nature of the city's phase transition to a congested state. In the second part, we outline a methodology that processes raw geolocated data to extract aggregate mobility information that is comparable to local surveys and existing origin-destination matrices for five different metropolitan areas. Next, we analyze how the unique congestion fingerprint of a city is produced through the combination of travel demand, population density, road supply and route choice. We evaluate the potential of implementing socially aware routing solutions for congestion alleviation, and assess the implications of such solutions. Finally, we couple urban travel demand with energy demand of electric vehicles, and present their relationship while exploring the potential benefits of optimized arrival hour and charging timeshifts.

Thesis Supervisor: Marta C. González
Title: Associate Professor

Acknowledgments

First and foremost, I'd like to express my gratitude to my advisor, Professor Marta C. González, for her continuous guidance and support over the last five years. Through her mentorship I am leaving MIT with a stronger mind, for which I will remain indebted.

I would like to thank my committee, Professor Barnhart and Professor Amin, for their insightful comments and help in strengthening my work. I also would like to thank my colleagues and labmates along the way: Pu Wang, Christian Schneider, Carlos Herrera, Jameson Toole, Yingxiang Yang, Suma Desu, Antonio Lima, Arda Halu, and Emre Can Kara, among many others. Thank you for making 1-151 a fun and happy place.

Above all, I would like to thank my parents, my brother and his family, for their continuous love and support. They made this dissertation possible.

Cease, cows, life is short.

Dedicated to my parents, Hamit Çolak and Berrin Çolak.

Contents

1	Introduction	23
1.1	Literature Review	23
1.1.1	Big Data	23
1.1.2	Human Mobility	25
1.1.3	Transportation Demand Modeling	31
1.1.4	Traffic Flow	38
1.1.5	Science of Cities	40
1.2	Dissertation Overview	43
2	On the role of spatial dynamics and topology on network flows	47
2.1	Introduction	47
2.2	Methods	49
2.3	Results	51
2.4	Discussion	58
3	Extracting Origin Destination Information from Mobile Phone Data	59
3.1	Introduction	60
3.1.1	Description of Data	62
3.2	Methods	64
3.2.1	System Architecture and Implementation	64
3.2.2	Estimating Origin-Destination Matrices	68
3.3	Results	84
3.3.1	Trip Tables and Survey Comparison	86

3.3.2	Road Network Analysis	88
3.3.3	Bipartite Road Usage Graph	88
3.3.4	Visualization	95
3.4	Discussion	96
4	Understanding congested travel in urban areas	99
4.1	Introduction	99
4.2	Methods	101
4.2.1	Mobile phone data	101
4.2.2	Census and travel survey data	103
4.2.3	Extraction of validated origin-destination information	104
4.2.4	Road networks	107
4.2.5	Traffic flow and travel time	108
4.2.6	Traffic assignment	110
4.3	Results	114
4.3.1	Approach	114
4.3.2	Travel times	116
4.3.3	Selfish Routing	123
4.3.4	Weight of Social Good	126
4.3.5	Travel time benefit distributions	126
4.4	Discussion	128
5	Coupling Electric Vehicle Charging with Urban Mobility	131
5.1	Introduction	132
5.2	Methods	136
5.2.1	Data	136
5.2.2	Electric Vehicle OD Estimation	137
5.2.3	Optimization Model	138
5.3	Results	140
5.3.1	Estimating Electric Vehicle Mobility	140
5.3.2	Electric Vehicle Charging Session Profiles	141

5.3.3	Energy and Travel Demand Relationship	144
5.3.4	Assessment of the Impact of Charging Timeshifts	147
5.4	Discussion	151
6	Conclusion	155
6.1	Summary	155
6.2	Future Work	157
6.2.1	Analytics Engines for Mobility	157
6.2.2	Data-Driven Policies	158
6.2.3	Mobility and Sociodemographics	159
6.2.4	Incentivization for Overall Benefit	160

List of Figures

1-1	An illustration of a mobile phone trace of a typical user, and the uniqueness of such traces versus the number of spatiotemporal points taken at random [56]. Five points are enough to distinguish one person from all others in the dataset almost completely.	25
1-2	Distribution of the radius of gyration, r_g , obtained for each person using their mobile phone logs by the use of Equation 1.1. $P(r_g)$ follows a truncated power-law [70].	27
1-3	(a) The hourly regularity $R(t)$ of users throughout a typical week [154]. $R(t)$ ranges between 0.55 and 0.9, implying that humans are highly regular in terms of locations they stay in. (b) The number of unique locations visited $S(t)$ versus $t(h)$, time in hours. $S(t)$ grows sublinearly with $t^{0.6}$ for r_g levels ranging from 32 to 256 kilometers. [153] (c) The visitation frequency F_k of the k th most visited location follows Zipf's law, across segments with S ranging from 20 to 60. [153]	28
1-4	Probabilities as percentages $p(ID)\%$ of finding each of the 17 enumerated mobility motifs across Paris and Chicago and mobile phone data, surveys, and models. The red nodes depict home location, the fractions below motifs denote the number of most observed configurations to those possible. [144]	29
1-5	A color coded depiction of K_{road} and a breakdown of the MDSs for two road segments. K_{road} values are typically higher for road segments that are more central. The degree distribution is approximated by a normal distribution. [171]	30

1-6	(a) An illustration of the gravity and radiation models for Davis County, UT and Madison County, AL and the T_{ij} estimates from these locations to Washington County, UT and Houston County, AL, respectively. (b-c-d) Comparison of measured and estimated T_{ij} for both models. Green boxes imply that the model estimate lies between the 9th and 91th percentiles of the data. [149]	32
1-7	Flowchart describing the 4-step modeling process, highlighting the four main steps: generation, distribution, modal split and assignment. These steps are carried out repeatedly, depending on the outcomes of the evaluation. [126]	33
1-8	An illustration of activity based models. HBW stands for home-based-work trips, HBO stands for home-based other trips. Trips are evaluated based on activities obtained from locations uniquely labeled for each traveler. [126]	37
1-9	The fundamental diagram resulting from the cell transmission model [51], where q_{max} represents maximum flow and is sustained by the road for densities within $[k_a, k_b]$. At k_j the road reaches its jam density and speed drops to zero.	38
1-10	Scaling relationships of (a) building sizes and (b) employment against population. (c) scaling of employment and population density with respect to distance from the city center. [17]	41
1-11	Scaling relationships of population and size to other city properties [26, 25]	42

2-1	(a) Number of particles W versus time t for the Internet model with different loading rates for the shown simple network with node and edge betweenness values mapped as node size and edge width, respectively. (b) A second order transition at the critical value $R_c = 7$ for this network. (c) A schematic representation of the SPQM for a link with $C = 2$, $\tau = 3$ and $V = 7$. At $t = 2$, the link reaches its volume capacity and no incoming particle is allowed entering the link.	50
2-2	Transitions of a non-periodic lattice as estimated by simulations and by analytical solutions for (a) the IM and (c) the PQM. Transitions of a completely rewired network with $\rho = 1$ for (b) the IM and (d) the PQM for varying tendencies α to have long distance links. Analytical solutions are consistent with simulations.	52
2-3	(a) Distributions of the timespans through which the congested element operates at its outflow capacity at R_c for the IM and PQM, following a power law. (b) The average number of timesteps normalized by T before a gridlock is observed, versus the volume capacity. (c) Change in the nature of the transition to congestion of a non-periodic lattice of size $N = 625$ with specified levels of volume capacities. (d) The critical loading rate R_c for varying V/τ values. R_c settles to the PQM-limit as V/τ increases.	56
2-4	A congestion map of the San Francisco road network for $R = 36$ vehicles per timestep. Colors represent the times within which the road segments become congested as a consequence of the spillover. The black circle denotes the origin of congestion. (inset) PQM and SPQM transitions for the San Francisco road network, where $R_c^{\text{PQM}} = 40$ and $R_c^{\text{SPQM}} = 30$	57
3-1	A flowchart of the system architecture.	65
3-2	An illustration of different stages of the stay point algorithm. Figure taken from [88].	72

3-3	Pseudocodes for initialization and selection of candidate stays for the stay point algorithm.	73
3-4	Pseudocodes for agglomerative clustering and final pass for the stay point algorithm.	74
3-5	Schematic example of phone records converted to daily trips for a typical mobile phone user. Activities are inferred in stay locations and daily trips are measured by time of day between these stays.	75
3-6	Maps depicting the expansion factors in Boston and in Rio de Janeiro in the specified resolutions, the distributions of the values of these factors, and the comparisons between CDR population and the scaled up census population.	78
3-7	Pseudocode for home and work expansion of the OD creation algorithm.	79
3-8	Pseudocode for trip counting in the OD creation algorithm.	80
3-9	A spatial illustration of flows comparing inter and intra town/subdistrict OD pairs.	81
3-10	Our efficient implementation of the incremental traffic assignment (ITA) model. A sample OD matrix is divided into two increments and then split into two independent batches each.	83
3-11	Pseudocode for the ITA algorithm.	85
3-12	Correlations between OD matrices produced by our system and those derived from travel surveys at the largest spatial aggregation of the two models. In Boston, this refers to towns, in San Francisco, MTC superdistricts, in Rio, census superdistricts, and in Lisbon, freguesias. The larger of these area units (e.g. towns in Boston), the better our correlations, while correlations at the smallest aggregates(e.g. freguesias in Portugal), correlations are lower. However, more work must be done to understand uncertainties in estimates provided by both models.	89
3-13	Distributions of travel volume assigned to a road and the volume-over-capacity (V/C) ratio for the five cities. The values presented in the legend refers to the fraction of road segments with $V/C > 1$	90

3-14	A graphical representation of the bipartite network of roads and sources (census tracts), with edge sizes mapping the number of users using the connected road in their individual routes.	91
3-15	Distributions of k_{road} and k_{source} for the five cities.	91
3-16	Maps depicting the proposed road classification, summarized in the legend, for the five subject cities.	93
3-17	Two screen images from the visualization platform. (a) The trip producing (red) and trip attracting (blue) census tracts using Cambridge St., crossing the Charles River in Boston. (b) Roads used by trips generated at the census tract including MIT.	94
4-1	A typical depiction of rows of CDR data in Boston. User 12345678 makes a call from location A (Davis square), then goes on to make two calls from location B (Boston City Hall), then makes one call location C (MIT) at noon and another later, and makes one final call again from the location A at 8pm.	102
4-2	The population, vehicle usage rate and the expansion factor distributions of the five subject cities.	105
4-3	Properties of the road networks of five subject cities. Distributions of road segment length L , free travel time t_f , hourly vehicle flow capacity C , and a measure of total supply S	109
4-4	BPR function for obtaining travel time from volume and capacity. . .	110
4-5	A depiction of connector modeling. Tract centroid T is connected to nearest four intersections.	110
4-6	Algorithms used in the computation of equilibrium. [57, 122]	113

4-7	Illustration of routing equilibrium. (a) In this small network, 100 drivers are going from A to D. The road labels represent the costs of travel as a function of vehicle flows. User equilibrium allocates the flows between paths as $f_{ABD} = f_{ACD} = 25$ and $f_{ABCD} = 50$, and the average travel time is 3.75 minutes for all drivers. Socially optimal flows decrease total travel time to 3.5 by $f_{ABD} = f_{ACD} = 50$ and $f_{ABCD} = 0$, with road <i>BC</i> remaining unused. (b) Achieved percentage of potential savings for increasing values of social good weight λ : 10% and 20% social good weight results in 40% and 60% of potential savings, respectively.	115
4-8	The maps of VOCs (volume over capacity) of the roads in the user equilibrium configuration. The depicted cities are (a) Boston, USA, (b) San Francisco Bay Area, USA, (c) Lisbon, Portugal, (d) Porto, Portugal, and (e) Rio de Janeiro, Brazil. Higher VOCs are generally observed in highways as they provide faster means of travel. (Boston is 2x the distance scale.)	119
4-9	Comparisons of cities and their congested travel. (a) Distributions of commuting trip distances, d , in the morning peak period with parameters of the fitted lognormal distribution depicted in the legend. (b) Distribution of trip free flow speeds, v_f , and in traffic conditions, v_t . (c) Commuting travel times versus route distances of commuters, d_r . (d) Estimates of overall mean % of time lost in congestion versus population density p for TomTom Traffic Index estimates and our analysis. (e) Relationship of overall mean % congestion to the demand to supply ratio, Γ , for the five subject cities, with errorbars specifying the standard deviation. (f) Average population density ρ as a function of distance from most dense area in the region, r	120
4-10	The probability distribution fits for the straight-line distances.	121

4-11	% congestion distributions for the five cities. (a) Overall % congestion levels obtained from user equilibrium compared to values from TomTom. (b) Distributions of % congestion for trips for the five cities.	123
4-12	Travel time comparisons and potential savings. (a) Comparison of travel times and their distributions between user equilibrium versus routes obtained from the online map provider. Origin-destination samples consist of 2000 origin destination pairs with the highest commuting flow magnitudes for each city. (b) The percentage of potential savings in average commuting times for the five cities for varying levels of social good weight of routing. (inset: the travel time savings represented in actual minutes)	125
4-13	Benefit and congestion decrease distributions for different weights of social good. (a) A depiction of three route alternatives with the corresponding travel times for a trip from Union Square to San Francisco Airport for $\lambda = 0$; $\lambda = 0.2$ and $\lambda = 1$, respectively. (b) Counts of vehicle trips and observed travel time benefits for $\lambda = 1$ and $\lambda = 0.1$. Negative benefits refer to increase in travel times for vehicles sacrificing for the social good. The spread of the distributions increase for higher λ . (c) The response of distributions of percentage decrease in time lost to congestion to increasing values of λ . The skewness towards positive values of congestion decrease indicate movement towards more optimal configurations.	127
5-1	Coupling PEV charging with urban mobility. (a) Mobile phone trajectories are used to estimate mobility patterns. (b) Charging sessions used to characterize session and electricity demand curves. (c) These findings are combined to analyze the relationship between commuting patterns and electricity demand. (d) Charging activity is shifted to relieve peaks in demand and generate savings.	135

5-2	OD estimation for PEV drivers in the Bay Area. The probability distributions of commuting distances, D , and commuting travel times, T , of all vehicle trips and PEV trips estimated through income information and trip distances.	142
5-3	PEV charging session profiles. (a) Distribution of N_{day} the number of sessions per day for each driver ID starting from the day of first record (right inset: number of new driver IDs added every month, left inset: distribution of N_{EVSE} , the number of unique EVSEs visited by every driver ID.) (b) The distributions of h_a and h_d , the arrival and departure hours to and from an EVSE. (inset: the distributions of Δh_a and Δh_d , the interarrival and interdeparture times for a driver ID visiting a specific EVSE.) (c) The distribution of E_S , the total energy withdrawn per session (inset: the distribution of δ_S , session durations.) (d) The power consumption as a function of the normalized session duration segmented by total duration groups.	145
5-4	Relationship between OD trips and energy consumption. (a) The probability distributions of session energy E_S obtained from charging sessions compared to those of distance, D , free commuting travel time T and commuting travel time in traffic, T_c , estimated from ODs. (inset: the cumulative distributions) (b) Linear regressions of the average distance of Nissan Leaf commuting trips to a zipcode, D^z , versus the average morning session energy for these PEVs for that zipcode, E_S^z and E_M^z . The bands represent the 95% confidence intervals, and point sizes are proportional to the number of sessions in that zipcode. (inset: a regression of the first differences of the energy and distance measurements)	148

5-5 Assessing the benefits of minimizing peak power. (a) An illustration of a sample of sessions in the flexible model. The charging event within a session is shifted such that the overall peak power is minimized. (b) The decrease in peak power measurements for each model and varying d . Peak shaving of up to 38% can be achieved. (c) The influence of the proposed changes on travel times. Majority of drivers are not influenced, the worst case is a few individuals suffering from an additional 20 minute delay for $d = 4$. (d) The demand charge portion of the monthly bill and distribution of savings for varying values of d . The flexible model is able to generate savings worth approximately 1400\$ for $d = 4$ 152

List of Tables

3.1	A comparison of the extent of the data involved in the analysis of the subject cities.	64
3.2	Trip tables estimates. Where possible, our results are compared to estimates made using travel surveys. For each city, we report the number of person trips in millions for a given purpose or time. Trip purposes include: home-based work (HBW), home-based other (HBO), and non-home-based (NHB). Trip periods include: 7am-10am (AM), 10am-4pm(MD), 4pm-7pm (PM), and the rest of the day (RD). We note that the exact boundaries of the surveys do not exactly coincide with those used in our estimation so direct comparisons are not exact. In general, trip magnitudes align closely, with the exception of Rio de Janeiro, where the survey results report far too few trips, illustrating the difficulty of obtaining sensible measurements via certain techniques. No comparisons could be found for Porto. *Note that the Lisbon Survey only contains estimates of vehicle trips in millions. . .	87
4.1	A comparison of the extent of the data involved in the analysis of the subject cities.	117
4.2	A comparison of general properties of the subject cities.	117
4.3	KS test statistics for lognormal, exponential and power-law distribution fits for the straight-line distances.	118

4.4	Comparison of cost findings in the subject cities for the morning peak hour. Colored rows indicate the loss of travel times from free travel times to socially optimal flows, and from socially optimal flows to user equilibrium flows for commuters, respectively (FTT: free travel time, SO: social optimum, UE: user equilibrium, % S: percentage of total congestion attributed to selfish routing, $S = 100 * Benefit/Loss$) . .	124
4.5	Regression statistics for travel time estimations.	124
5.1	Characteristics of PEV drivers. Distribution of (a) average daily miles driven and (b) annual income by PEV drivers in California, USA [62].	137

Chapter 1

Introduction

This chapter consists of two sections. The first paints a picture of the literature that sets the context for the subject of the work described in this thesis. More specifically, the literature review is divided into five distinct parts: the explosion of data, human mobility, transportation demand modeling, traffic flow and science of cities. The second and last section presents an overview of the remainder of the dissertation for the reader.

1.1 Literature Review

1.1.1 Big Data

In early 20th century, information was mostly stored in analog, namely in paper, vinyl, cassette players or floppy disks. Starting from late 1980s, with the introduction of personal computers, data started being digitized. A decade later, digital storage started to make its way into personal computers in the form of compact disks, portable hard disks, and flash drives. The new millennium introduced a sharp decline in hardware and storage costs, where at the same time, as predicted by Moore's Law, transistor density in integrated circuits continued to double every two years, fueling further this paradigm shift. Today, most new data is stored digitally- and older data is being digitized as well. Technological innovation continually replaces analog

devices with digital ones, consequently producing data in increasing volume, variety and velocity. This led to cloud storage and computing solutions, more specifically, a group of remote servers hosted over the Internet that process or store data. Companies like Amazon have built business practices around providing web services to anyone willing at low rates.

This trend of digital devices replacing old analog ones combined with low costs of computing and storage consequently started the age of what is labeled as the Internet of Things. What this phrase refers to is a network of devices, such as one's cellphone, thermostat, vehicle or any other electronic equipment that can communicate with other similar devices by collecting and transferring data. When considered in the context of urban living, the collection of devices act as a swarm of sensors signaling information about the elements of a city: where its people are, how congested the roads are, what the air pollution levels are, and so on. This implication results in a grand vision, a smart city, where all infrastructure can be maintained electronically through generated data and can be returned to the people as valuable information. For example, the information we generate through the use of cellphones feed the live traffic maps we use to check the traffic. The same information is also used by traffic control centers to act on problematic regions. Therefore it can be said that the massive amount of data generated can be utilized in assisting decision making for policy, as well as improving quality of life for citizens. This dissertation makes significant use of one such type of dataset, namely mobile phones, for the goal of generating information to assist both policy makers and citizens. These devices log the location of every call to a certain precision, which can be used for various purposes for the benefit of society. In our case that purpose is to better understand urban mobility and propose solutions to improve it.

A resulting and understandable concern is that of privacy and security. The connectedness of electronic devices and improved understanding and control of complex sociotechnical systems envisioned by the ideas of internet of things and smart cities come at the cost of loss of privacy and increased security risks. In their work [56], de Montjoye et al. demonstrate the privacy boundaries in mobile phone data. As

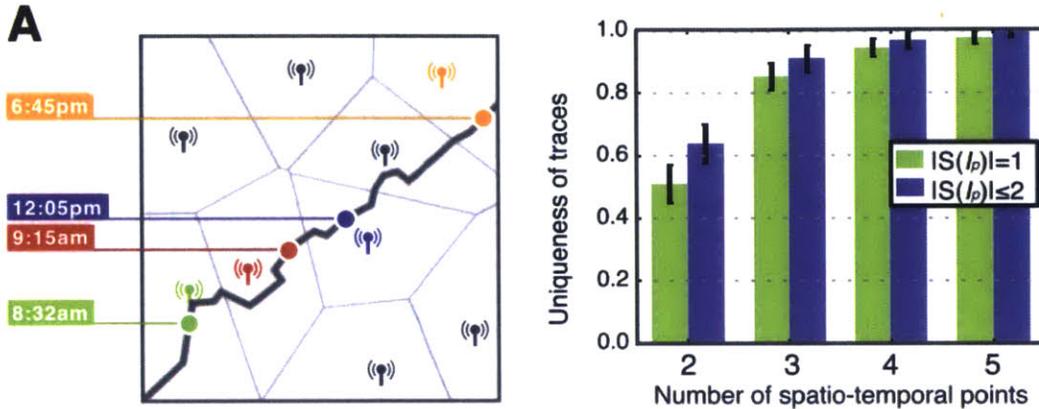


Figure 1-1: An illustration of a mobile phone trace of a typical user, and the uniqueness of such traces versus the number of spatiotemporal points taken at random [56]. Five points are enough to distinguish one person from all others in the dataset almost completely.

illustrated in Figure 1-1, the movements of a mobile phone user can typically be obtained at a certain spatial resolution from their digital traces. In this work, authors also measure the uniqueness of these traces for increasing number of spatiotemporal points to find that as few as four such points randomly picked is enough to distinguish the specific user out of millions. This finding specifically highlights the need of appropriate security and privacy precautions that need to be taken by data providers and users to keep the information safe.

1.1.2 Human Mobility

The subject of understanding how people move in a massive scale is aptly referred to as human mobility. This line of research is relatively recent, as its introduction is simultaneous with the permeation of large geolocated datasets into academia.

One of the first works in this area was carried out by Brockmann et al. [32], where authors presented an analysis of how bank notes circulated in the US. To gather this unique data, the authors creatively created a platform where the holders of specifically annotated banknotes reported their location. Considering the fact that bank notes can only be moved by the people who are in their possession, the set of all these

check-ins generated a valuable dataset on human mobility at the country scale. In this work, the authors demonstrate that Δr , the displacement between consequent records used to measure the movement characteristic, follows a scale-free random walk otherwise referred to as a Levy flight process. That is, the probability distribution $P(\Delta r)$ is heavy tailed, and humans infrequently make long trips.

This work was immediately followed by [70], where authors used the mobile phone traces of 6 million anonymized individuals from a European city to test a similar hypothesis. The work set out to investigate the underlying mechanism that produces the scale-free outcome of $P(\Delta r)$. In doing so, authors introduced a new metric r_g , the radius of gyration, to measure an individual’s movement:

$$r_g(t) = \sqrt{\frac{1}{N(t)} \sum_{i=1}^{N(t)} (\mathbf{r} - \mathbf{r}_{cm})^2} \quad (1.1)$$

where $N(t)$ is the number of observed locations and \mathbf{r}_{cm} is the center of mass for the user. The results, similar to that in [32], showed that $P(\Delta r)$ followed a truncated power-law, as shown in Figure 1-2. This seminal work tested three hypotheses of which the first was that made in [32]. Gonzalez et al. concluded that the statistics of individual trajectories $P(\Delta r_g | r_g)$ and the heterogeneity of $P(r_g)$ convolve to produce the resulting $P(\Delta r)$. Essentially, the Levy flight process proposed in [32] did in fact hold but only in each individual’s characteristic r_g . Moreover, once normalized for r_g and anisotropized, mobility patterns collapsed onto a single universal pattern. This finding pointed to a very significant implication: humans are unique yet very similar in the patterns with which they move. Following this work, new sources of geolocated movement data opened up previously unimaginable avenues in understanding movement of individuals.

A large body of work soon followed, mainly focused on providing an improved understanding of aggregate individual mobility. Among these works, [154] and [153] provide a good overview about the ways people visit locations, and some of the main contributions are shown in Figure 1-3. First, people demonstrate high regularity $R(t)$, the probability that in a specific hour of the week, a user will be in their most visited

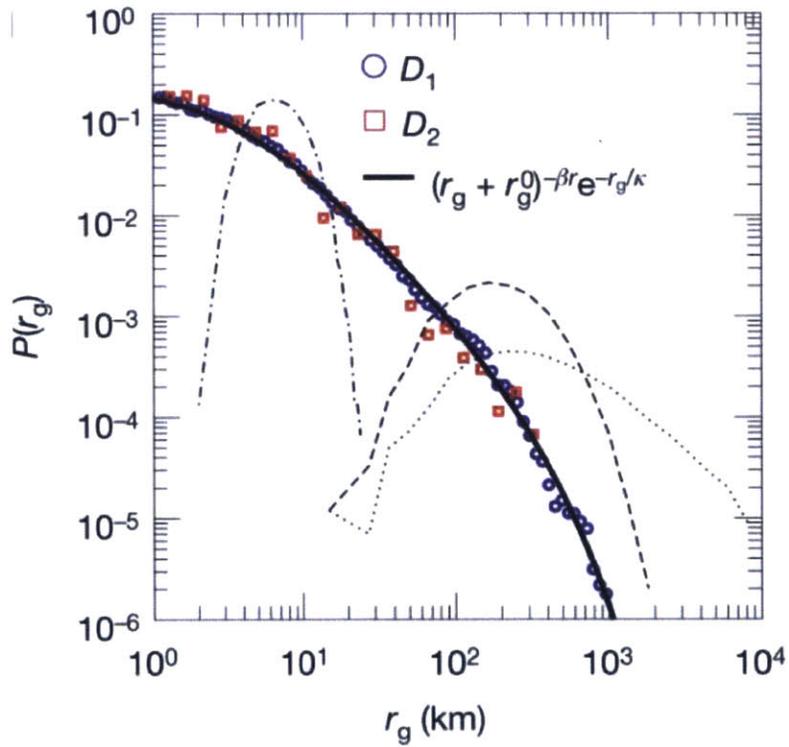


Figure 1-2: Distribution of the radius of gyration, r_g , obtained for each person using their mobile phone logs by the use of Equation 1.1. $P(r_g)$ follows a truncated power-law [70].

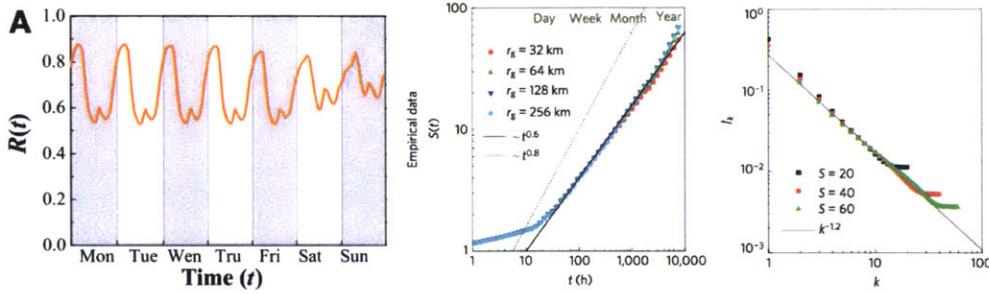


Figure 1-3: (a) The hourly regularity $R(t)$ of users throughout a typical week [154]. $R(t)$ ranges between 0.55 and 0.9, implying that humans are highly regular in terms of locations they stay in. (b) The number of unique locations visited $S(t)$ versus $t(h)$, time in hours. $S(t)$ grows sublinearly with $t^{0.6}$ for r_g levels ranging from 32 to 256 kilometers. [153] (c) The visitation frequency F_k of the k th most visited location follows Zipf’s law, across segments with S ranging from 20 to 60. [153]

location for that hour. In fact, an average person is to be found in their most visited location for that hour more than 80% of the time, a staggering number providing a lower bound to location prediction. Secondly, for a range of values of r_g , the number of unique locations visited $S(t)$ grows sublinearly with time. People are slow explorers, and are biased towards visiting locations they already have. Thirdly, f_k , defined as the visitation frequency of the k th most visited location for a user, follows Zipf’s law for varying levels of S . In other words, the frequency at which people visit a place is inversely proportional to its ranking based on the number of times that place is visited. In overall, these findings show that people sublinearly explore new locations, are highly regular in their visited locations and in fact stay relatively confined to a specific subset of these locations.

Schneider et al. further built on these findings in [144], where they analyzed the motifs with which people visited locations using a similar mobile phone dataset. As demonstrated in Figure 1-4, people demonstrate only about 19 distinct motifs formed by at most 6 unique locations. The work also presents an analytically tractable Markov chain framework that reproduces empirical findings from the mobile phone data as well as those observed in previously conducted surveys.

The evolution of this literature increasingly showed the potential of utilizing geolocated data to understand movement. Within several years, models that could

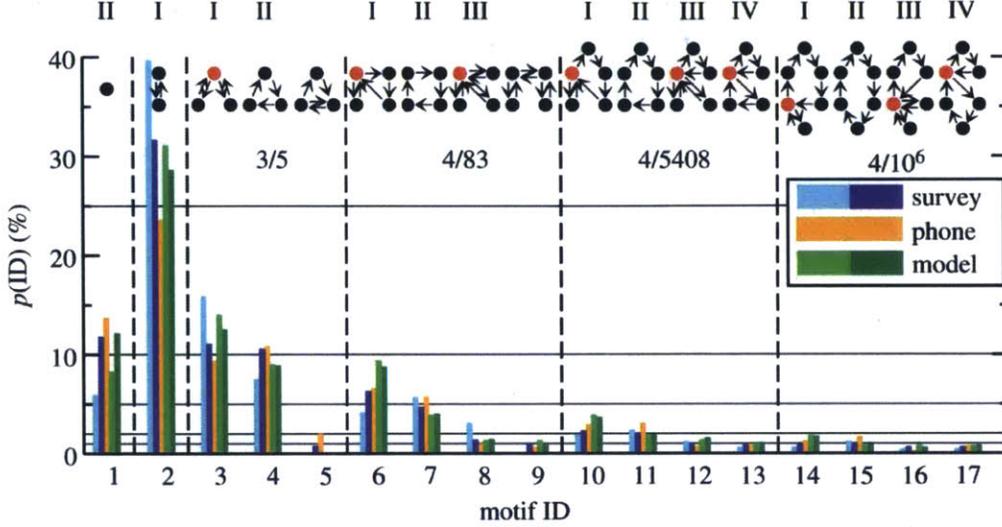


Figure 1-4: Probabilities as percentages $p(ID)\%$ of finding each of the 17 enumerated mobility motifs across Paris and Chicago and mobile phone data, surveys, and models. The red nodes depict home location, the fractions below motifs denote the number of most observed configurations to those possible. [144]

replicate surveys at the aggregate level were produced with promising results. The know-how required to transform these newly generated datasets into insights can feed our solutions to complex sociotechnical problems. The most obviously related complex system is transportation. The next step, and the big challenge, is bringing these new findings and approaches into the specifics of transportation and urban travel.

In this context, Wang et al. [171] made the first attempt to bring the demand information obtained from mobile phones into the transportation world. In this work, the authors aimed to analyze road usage patterns in two metropolitan areas in the US: Greater Boston and the San Francisco Bay. By estimating transient origin-destination patterns from user trajectories, authors estimate the vehicular demand for a typical weekday morning in both regions. They define MDS for each road segment, namely the major driver sources, from which they measure K_{road} as the count of how many distinct MDSs a road segment attracts travellers from. Authors combine this metric with the topological measure of betweenness centrality [63, 13] that measures the tendency of a segment in the road network to lie on shortest paths, and provide a novel classification of roads in a metropolitan area: connectors that have high

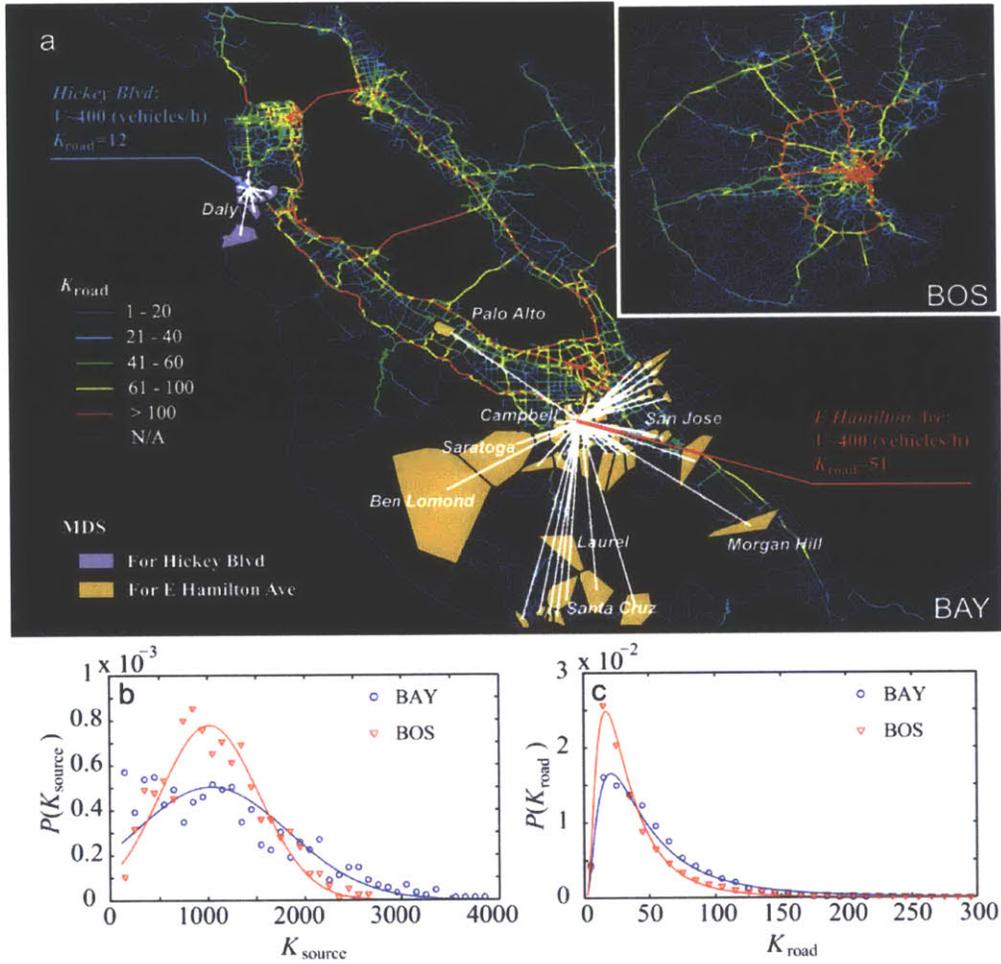


Figure 1-5: A color coded depiction of K_{road} and a breakdown of the MDSs for two road segments. K_{road} values are typically higher for road segments that are more central. The degree distribution is approximated by a normal distribution. [171]

betweenness and high K_{road} , peripheral connectors that have high betweenness but low K_{road} , attractors that have low betweenness but high K_{road} , and locals which have low betweenness and low K_{road} . Figure 1-5 illustrates some of these findings in this work.

On another front very closely related to transportation, Simini et al. [149] test commonly used gravitational trip distribution model in the transportation literature. Analogous with Newton's law of gravitation, the gravity model estimates T_{ij} (number of people that move between two locations i and j) as

$$T_{ij} = \frac{m_i^\alpha n_j^\beta}{f(r_{ij})}, \quad (1.2)$$

where m_i and n_j are the populations in the corresponding locations and r_{ij} is the distance between them. α , β , and $f(r_{ij})$ are used for calibration. The various weaknesses of the gravity model (including the lack of a rigorous derivation and the free form of $f(r_{ij})$ among others) led the authors to develop the radiation model:

$$\langle T_{ij} \rangle = \frac{m_i n_j}{(m_i + s_{ij})(m_j + s_{ij})}, \quad (1.3)$$

where s_{ij} represents the total population in the radius r_{ij} centered at i excluding m_i and n_j . Their findings show that the commuting flows agree substantially better with the radiation model, as shown in Figure 1-6. Since this work, a large body of work was produced to improve the traditional models and those newly proposed [181, 96, 107].

1.1.3 Transportation Demand Modeling

One of the biggest challenges in transportation is accurate modeling of urban mobility, as it is a very complex problem with high dimensionality. People, places, activities, roads, public transportation services all come together to form a very heterogeneous, almost organic system. When we consider a typical morning commute at the level of a single individual, many factors determine the nature and the characteristics of this trip. The most important factors are personal: where one lives, goes to work, walks his/her dog, where his/her choice of grocery store is, and so on. Some other these factors are the sociodemographic characteristics of the person, namely income, education level, and social network. Some other factors are external: modes of travel available, weather conditions, events such as concerts or sports games in vicinity for that specific time. Among these, the cost of travel for each mode, not only monetary cost but also that of time, further influences the traveler's choices. In overall, these factors and possibly many others are all significant in one's choices regarding whether or not to make a trip and if so, in what way one makes it.

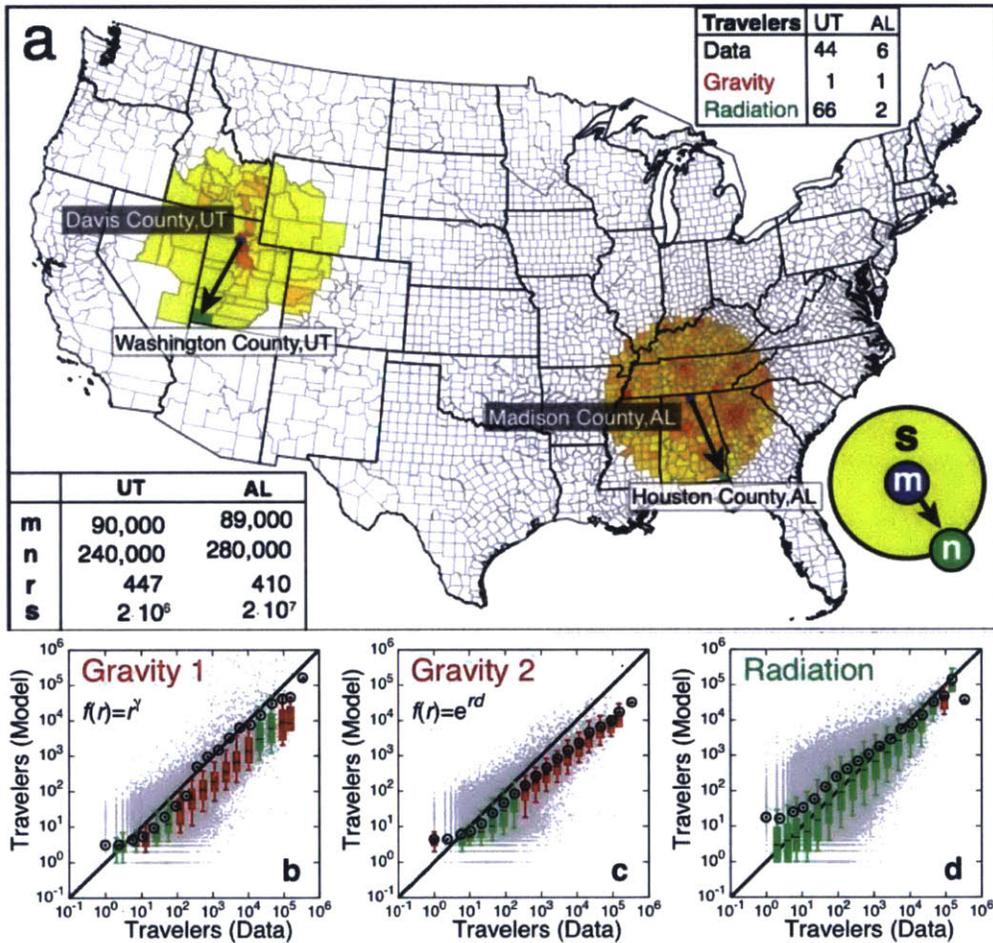


Figure 1-6: (a) An illustration of the gravity and radiation models for Davis County, UT and Madison County, AL and the T_{ij} estimates from these locations to Washington County, UT and Houston County, AL, respectively. (b-c-d) Comparison of measured and estimated T_{ij} for both models. Green boxes imply that the model estimate lies between the 9th and 91th percentiles of the data. [149]

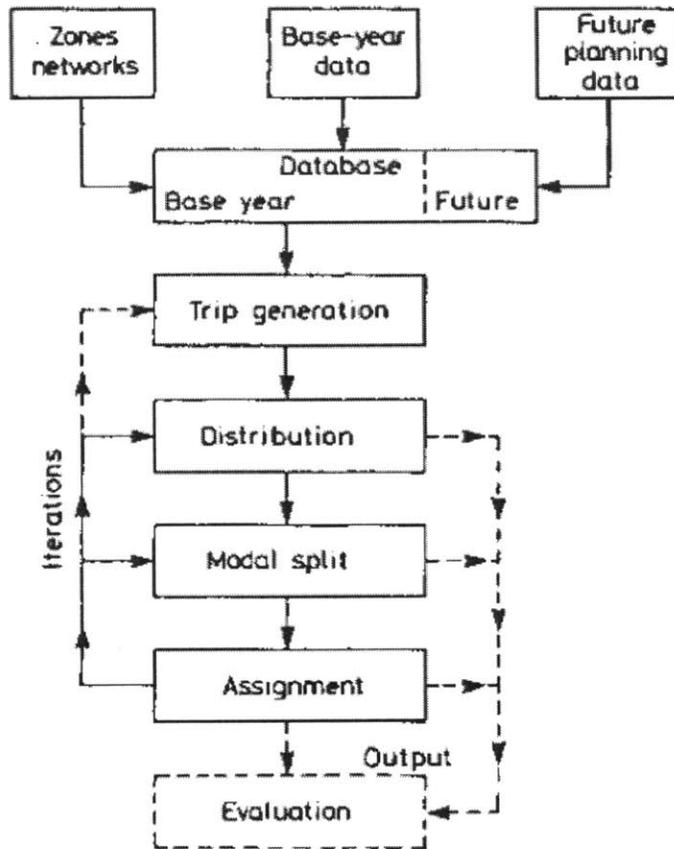


Figure 1-7: Flowchart describing the 4-step modeling process, highlighting the four main steps: generation, distribution, modal split and assignment. These steps are carried out repeatedly, depending on the outcomes of the evaluation. [126]

Understandably, accurately predicting the travel demand for a metropolitan area of millions of residents is a very difficult task. Transportation planners undertook this problem starting early 20th century and developed a series of methods to model most of these concepts over time. The general framework that encapsulates these methodologies is referred to as the *four step model*, as summarized in Figure 1-7. This modeling framework begins with gathering population data and forming a set of TAZs (Traffic Analysis Zones) over the region of analysis. Additional inputs such as economic activity, a breakdown of points of interest such as shops, schools, and parks are also obtained. First step, *trip generation*, consists of estimating the total number of trips generated by each zone. The next step, *distribution*, produces what's known as the origin-destination matrix: a breakdown of the number of total trips between any two zones by time and purpose. *Modal split* models the choice of mode of transport for each traveler taking into account the characteristics of various route options, modes as well as the traveler's sociodemographic characteristics. The final step is the *assignment* of these trips to the network of the corresponding mode.

Trip generation models aim to estimate the total number of trips generated and attracted by each location, broken down by time and purpose. For example, how many car trips are made from Harvard Square to Boston Common on a typical Monday morning? To incrementally estimate these numbers, one method is growth-factor models. These models intake previous findings for these numbers, and estimate the growth factor as a function of current and future parameters such as population, car ownership, income, and others. The more general solution in earlier 20th century was building statistical models (such as regressions) to estimate trip generation rates for parameters such as those aforementioned. However due to the inherent limitations regarding dependency, category analyses were proposed in 1960s [176]. These models stratified the population by each parameter, for example, one strata would consist of the group of households whose income range between 2000 and 3000 \$ per month and own 1 car. A regression would be carried out independently for each strata.

Trip distribution models are considerably more complex. The production and attraction values obtained for each zone do not provide the information regarding the

distribution of these trips, specifically in terms of pairs. Trip distribution aims to breakdown the trip number totals to a matrix, where each row and column corresponds to a zone, the sum of rows and columns correspond to the production and attraction of that zone. This matrix is called an Origin-Destination (OD) matrix. As mentioned in Section 1.1.2, synthetic models like the gravity model are among the first examples of trip distribution. For example, gravity model states that the number of trips between two locations are proportional to each of their population, and inversely proportional to the distance between them, as described in Equation 1.2. Various calibration and fitting techniques along with different forms of the distance relationships exist in literature. One other model that has gained traction is the intervening opportunities model, which states that the decline in trip magnitudes is not an outcome of distance, but rather opportunities [159, 145].

Modal split is an inherently different problem as it tries to model how individuals make a specific choice among various options. At the core of this problem lies understanding human behavior and choice. Therefore at this point in the four step modeling framework, a paradigm shift occurs towards a more disaggregate approach. In this context, discrete choice models are the most commonly accepted methodology [23]. The simplest of such models, is the multinomial logit model (MNL):

$$\pi_{ij} = \frac{\exp(u_{ij})}{\sum_{k=1}^J \exp(u_{ik})} \quad (1.4)$$

where π_{ij} stands for the probability of i making choice j , for utility values calculated for each choice u_{ik} . Probit models have also been widely used. These models have since been significantly improved upon and used in complex ways such as mixed, nested or combination models over the years.

The fourth and last step, assignment, has significantly evolved over the twentieth century. With the foundation of game theory [118], the idea of equilibrium permeated into transportation modeling. The typical equilibrium concept in economics is the point of balance between supply and demand where the marginal cost of producing a good is equal to its marginal revenue. In the context of transportation, this occurs

on the roads- people choose routes in a way that minimizes the travel time, i.e. incurred cost, until they no longer can. Traffic assignment models try to solve this problem, that is, how does the network look like in equilibrium, and conversely, out of equilibrium.

One seminal work on this topic is that of J.G.Wardrop [172], which focuses on how rational drivers choose routes and its outcomes in terms of travel times. The paper is widely known for Wardrop's first and second principles and the formalization of the equilibrium concept in the context of route choice. The first principle states that the journey times in all routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route. The second principle asserts that in equilibrium, the average journey time is minimum. These findings are another description of Nash equilibrium [118]. Beckmann et al. [19] mathematically formulated this concept in traffic networks, namely user equilibrium.

Typically, demand information in the form of an OD matrix, a road network representation with accurate capacity and speed profiles, and rules regarding route selection are the inputs of traffic assignment. The first is, as described earlier, obtained in trip generation and distribution steps. The road network, is generally at hand in most transportation consulting companies for their subject cities. Nowadays, open-source repositories such as Open Street Maps (OSM) can be used to obtain this information. The third input, namely the route choice model, can be rather simple or complex. Old methods of assignment include static incremental traffic assignment, where the demand would be assigned to the network in batches in an effort to approximate equilibrium. Starting late 90s, route and path based assignment procedures made use of computers in solving for equilibrium more efficiently and accurately. The forefront of traffic assignment research today is more concerned with implementations of dynamic traffic assignment, where OD information varies temporally and stochastic components are incorporated into route choice [93, 86].

Over the years and still today, the four step model has formed the solid foundation of travel demand modeling. In their quest to improve this methodology, modelers focused on the reasons for a trip, that is, for what purpose do people travel? This

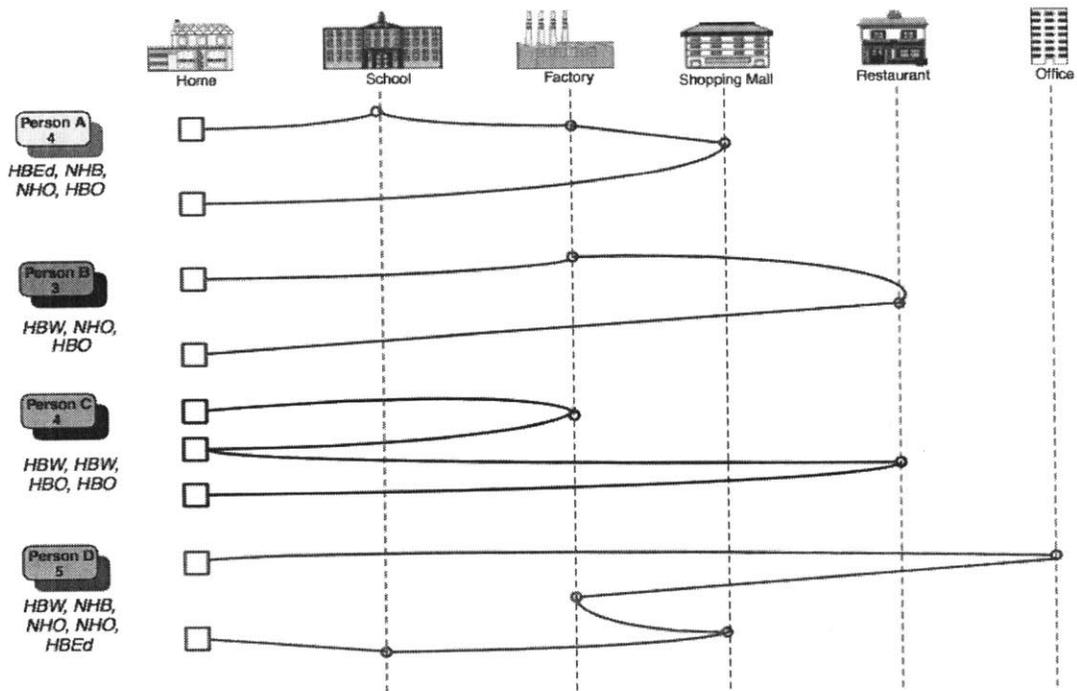


Figure 1-8: An illustration of activity based models. HBW stands for home-based-work trips, HBO stands for home-based other trips. Trips are evaluated based on activities obtained from locations uniquely labeled for each traveler. [126]

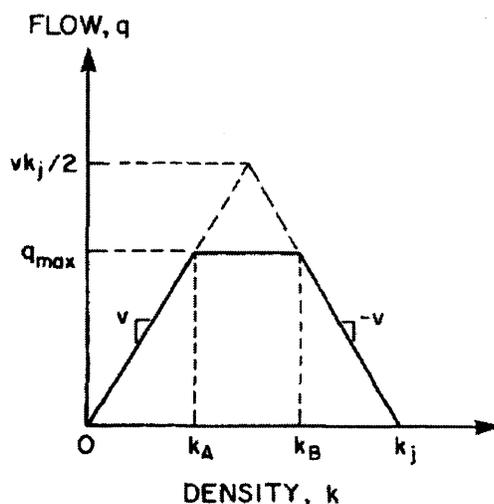


Figure 1-9: The fundamental diagram resulting from the cell transmission model [51], where q_{max} represents maximum flow and is sustained by the road for densities within $[k_a, k_b]$. At k_j the road reaches its jam density and speed drops to zero.

idea led to the interest in better understanding activities and how they influence trips. In other words, instead of looking at the total number of trips between two locations, modelers wanted to understand the travel diary of an individual, composed of a series of activities such as grocery shopping, commuting, picking up kids from school that generate trips. Figure 1-8 illustrates the general concept of activity based modeling, which has become the current state-of-the-art in practice.

1.1.4 Traffic Flow

The modeling of traffic flow on roads is a separate yet rich area of study. Besides the easily measurable speed, researchers quantified density and flow, as the number of vehicles per unit of roadway and per unit time passing through it, respectively. Based on these measurements, microscopic and macroscopic models were developed [116]. Several fundamental flow diagrams, a commonly used label for the relationship between the flow and the density in a road segment, were proposed and are still being revised and calibrated today [131]. One commonly used traffic flow model is the cell transmission model (CTM), developed by Daganzo et al. [51]. In this model, cars fill

up discrete cells that form the road segment. The number of cars that flow into and out of each cell is kept track of, and thus the number of cars in each cell. This model of flow continuity enables CTM to capture traffic jams and shockwaves. The basis of CTM can be mathematically described as,

$$\begin{aligned}
 n_{i+1}(t+1) &= n_i(t) \\
 n_i(t+1) &= n_i(t) + y_i(t) - y_{i+1}(t) \\
 y_i(t) &= \min(n_{i-1}(t), Q_i(t), N_i(t) - n_i(t))
 \end{aligned}
 \tag{1.5}$$

where $n_i(t)$ is the number of vehicles in cell i at time t , $y_i(t)$ is the inflow to i at t , $Q_i(t)$ is the flow capacity into i , and $N_i(t)$ is the maximum number of vehicles for cell i . These formulas describe flow between cells such that flow capacities and spatial constraints are conserved. The resulting fundamental diagram is shown in Figure 1-9.

Cellular automata approaches regarding modeling traffic flow has also attracted a lot of attention, mostly from the physics community. One such model is called the Nagel-Schreckenberg (NS) model, where the road is again divided into cells of a fixed duration Δt . Each cell is either occupied or unoccupied with vehicles. Each vehicle, can accelerate if not at maximum velocity and decelerate if there is a vehicle in the following cell. With added noise, NS model reproduced various properties of traffic flow as well as providing valuable insights regarding jams and shockwaves.

Today, complex micro and macrosimulators are able to combine traffic flow theory and general traffic demand models to provide good flow estimates and sensitivity analyses. MITSIMLAB, a microscopic traffic simulation laboratory developed at MIT is a prime example [12]. DYNAMIT, a real time model system for network management and emergence response is developed by the same group [22]. Similarly, resesarchers at ETH Zurich have built MATSIM, a highly capable multi-agent transport simulation tool [9].

1.1.5 Science of Cities

Starting from the beginning of history, humans increasingly clustered together to generate economies. Archeological findings date the earliest *cities*, as we understand them today, back to 5000 BC. Today, popularly referred UN statistics state that more than half of the world's population lives in cities, understandably inspiring scientists to try to understand how they form, evolve and operate. Scientists from various disciplines have aimed to understand how cities grow in population and in area, how the distribution of population density is evolves, how cities stay monocentric or transform into polycentricity, and how infrastructures, interactions, socioeconomic metrics, and rents are interrelated. The commonality in these attempts how difficult the problem is: cities are results of highly complex interactions between similarly complex elements. To clarify, one can picture an individual, simply at the microscopic scale. The individual is difficult to model, even when simplified to be rational, as exemplified by topics in Sections 1.1.2 and 1.1.3. These individuals live and move in the city: they go to restaurants, meet with friends, pay rent, consume electricity. The outcome of this behavior in aggregate determines how dense, expensive, safe, congested, clean the city is. At the mesoscopic scale, the problem is aimed towards understanding the intricacy of how a city expresses itself. This leads to the question of how universal these patterns of expression are, in other words, analyzing a system of cities. In addition to the spatial scale, the science of cities also vary by temporal resolution: understanding city growth requires analysis that span decades, whereas in the problem of disaster response, even seconds are important. Physicists, economists, urban planners, engineers among others have worked in various spatial and temporal scales of cities. This section focuses on the literature at the macroscopic level.

A typical methodology in understanding a system is to assess how it scales with respect to its various properties. In the context of cities, area and population are the two key properties. Gibrat's Law for cities states that the city area and population sizes can be approximated by lognormal distributions. In his interesting work extending Gibrat's claims [17], Batty examines city and building sizes, populations,

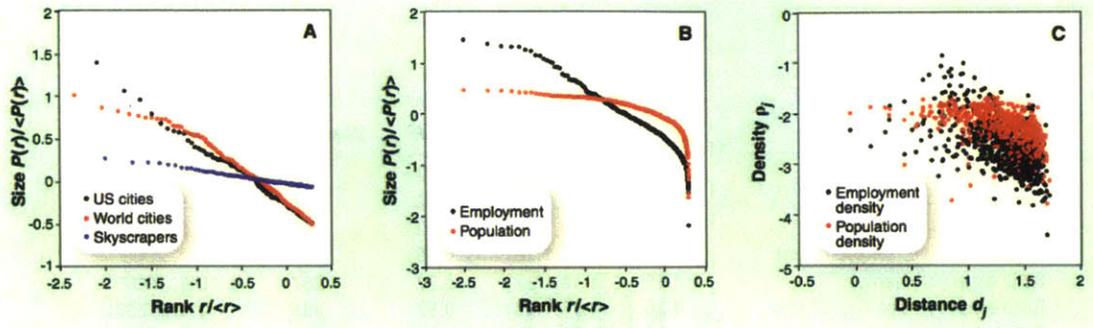


Figure 1-10: Scaling relationships of (a) building sizes and (b) employment against population. (c) scaling of employment and population density with respect to distance from the city center. [17]

population densities and employment. The author argues that growth rates are far from being random, nor are they favoring large cities. In fact, growth generally is shifted towards smaller cities, which is arguably the consequence of the higher benefits of economies of scale in smaller cities both in terms of income and expenses, and the increased stress imposed on the road and power infrastructures. Figure 1-10 shows the findings in this work, specifically that building sizes and employment scale with population. The scaling against distance from what can be accepted as the central business district, or the most densely populated area, or the city center can also be observed.

In [25, 26], Bettencourt et al. more thoroughly examine how various properties of cities come together and how they scale. These works summarize the observed exponent ranges for proposed models for various city properties. The authors scale population N to determine network volume, length, interactions per capita and rents to provide estimates on the ranges of the exponents with the acceptable assumptions of $D = 2$ for number of dimensions and $H = 1$ for the fractal dimension. The observed exponent ranges are in good agreement with the models, as shown in Figure 1-11. In [26], authors beautifully classify scaling exponents. Biological organizations have $\beta < 1$, are driven by efficiency, and grow sigmoidally with long-term population limits. Productivity, and generation of wealth drive sociological organization and grow with a series of collapses and booms, thus $\beta > 1$.

Y	β	95% CI	Adj- R^2	Observations	Country-year
New patents	1.27	[1.25,1.29]	0.72	331	U.S. 2001
Inventors	1.25	[1.22,1.27]	0.76	331	U.S. 2001
Private R&D employment	1.34	[1.29,1.39]	0.92	266	U.S. 2002
"Supercreative" employment	1.15	[1.11,1.18]	0.89	287	U.S. 2003
R&D establishments	1.19	[1.14,1.22]	0.77	287	U.S. 1997
R&D employment	1.26	[1.18,1.43]	0.93	295	China 2002
Total wages	1.12	[1.09,1.13]	0.96	361	U.S. 2002
Total bank deposits	1.08	[1.03,1.11]	0.91	267	U.S. 1996
GDP	1.15	[1.06,1.23]	0.96	295	China 2002
GDP	1.26	[1.09,1.46]	0.64	196	EU 1999–2003
GDP	1.13	[1.03,1.23]	0.94	37	Germany 2003
Total electrical consumption	1.07	[1.03,1.11]	0.88	392	Germany 2002
New AIDS cases	1.23	[1.18,1.29]	0.76	93	U.S. 2002–2003
Serious crimes	1.16	[1.11, 1.18]	0.89	287	U.S. 2003
Total housing	1.00	[0.99,1.01]	0.99	316	U.S. 1990
Total employment	1.01	[0.99,1.02]	0.98	331	U.S. 2001
Household electrical consumption	1.00	[0.94,1.06]	0.88	377	Germany 2002
Household electrical consumption	1.05	[0.89,1.22]	0.91	295	China 2002
Household water consumption	1.01	[0.89,1.11]	0.96	295	China 2002
Gasoline stations	0.77	[0.74,0.81]	0.93	318	U.S. 2001
Gasoline sales	0.79	[0.73,0.80]	0.94	318	U.S. 2001
Length of electrical cables	0.87	[0.82,0.92]	0.75	380	Germany 2002
Road surface	0.83	[0.74,0.92]	0.87	29	Germany 2002

Data sources are shown in *SI Text*. CI, confidence interval; Adj- R^2 , adjusted R^2 ; GDP, gross domestic product.

Urban scaling relations	Observed exponent range	Model ($D = 2, H = 1$)	Model D, H
Land area $A = aN^\alpha$	[0.56,1.04]	$\alpha = \frac{2}{3}$	$\alpha = \frac{D}{D+H}$
Network volume $A_n = A_0 N^\nu$	[0.74,0.92]	$\nu = \frac{5}{6}$	$\nu = 1 - \delta$
Network length $L_n = L_0 N^\lambda$	[0.55,0.78]	$\lambda = \frac{2}{3}$	$\lambda = \alpha$
Interactions per capita $\bar{l}_i = l_0 N^\delta$	[0.00,0.25]	$\delta = \frac{1}{6}$	$\delta = \frac{H}{D(D+H)}$
Socioeconomic rates $Y = Y_0 N^\beta$	[1.01,1.33]	$\beta = \frac{7}{6}$	$\beta = 1 + \delta$
Network power dissipation $W = W_0 N^\omega$	[1.05,1.17]	$\omega = \frac{7}{6}$	$\omega = 1 + \delta$
Average land rents $P_L = P_0 N^{\delta_L}$	[0.46,0.52]	$\delta_L = \frac{1}{2}$	$\delta_L = 1 - \alpha + \delta$

Figure 1-11: Scaling relationships of population and size to other city properties [26, 25]

These works have led to many others [15, 16, 11, 99, 100] analyzing how mobility and congestion shapes cities, and how polycentricity evolves, among others. More importantly, they provide a benchmark in understanding the underlying mechanism of growth and productivity in cities as well as uncovering the nontrivial relationships between various metrics. This new understanding of cities will eventually produce more informed and educated local policies.

1.2 Dissertation Overview

In the second section of this thesis, we discuss the problem of congestion from a physics perspective, with the goal of building on the expertise reviewed in Section 1.1.4. Particle flows in spatial networks are susceptible to congestion. Our goal is to analyze the phase transitions of road networks to a state of congested transport and the influence of both topology and spatial dynamics on the emergence of this congested state. Many previous studies in the physics community have focused on the flows of data packets in the internet and therefore have reasonably overlooked the role of space and time in particle flows in the networks. In the case of data packets, the travel time between nodes is negligible and queues accumulate in the nodes which have limited capacity to process the packets and congestion occurs. In contrast, in road networks queues are formed on the links which have limited spatial capacity, and in turn velocities depend on the density of travelers on each link. In this context, transportation research focuses on capturing traffic flow by making use of the fundamental diagrams that empirically relate flow, density and speed in road segments and utilizing them in macroscopic link models. Our goal in this section is to build on the simple model of the Internet by adding temporal and spatial dimensions of particle flow and hence construct a framework to analyze congestion in spatial networks from a network science perspective. Under simplified demand profiles, and simple flow dynamics, how can we describe the inherent physical phenomenon that is congestion from the perspective of space and topology? We systematically show that the value of the critical loading rate at which congestion emerges is affected

by the addition of spatial dynamics, changing the nature of this transition from a continuous to a discontinuous one. Our numerical results are confirmed by introducing an analytically solvable framework. As a case of study, we explore the implications of our findings in the San Francisco road network where we can locate the roads that originate the congested phase. These roads are the spatially constrained, and not necessarily those with high betweenness as predicted by models without spatial dynamics.

In the third section of this work, we focus on origin-destination matrix estimation using mobile phone data. A brief literature on travel demand estimation was presented in Section 1.1.3. Typically, the necessary information to initiate the four step model is obtained through surveys. These surveys require meticulous design to capture the mobility of the whole population by the smallest sample of households and as accurately as possible. The people in the selected households are all asked questions about the trips they make, the times they make them, the modes they use, where they work, and other detailed information. However despite the quality of the data obtained from these surveys, they have a very small sample size, they're expensive, time consuming, and they require a lot of design, supervision and quality control. On the other hand, every time we use our mobile phones, a crumb of data containing a timestamp and an approximate geolocation is generated and stored in the service providers' servers. This practice was initially adopted by these companies for billing purposes, as these call logs eventually were used to calculate the monthly bill for the consumer. As phones become more and more capable, consumers regularly use data services to not only make calls but to access applications, check emails, and even listen to music or read. In consequence, the size and potency of the data a consumer generates is constantly increasing. Moreover, along with the rapid progression of mobile technology over the last years the adoption has also drastically increased: As of January 2014, 90% of American adults own a cellphone, and 58% own a smartphone. The ubiquity of mobile phone usage is persistent in not only the United States but around the world: usage statistics obtained by the World Bank indicate an average of 0.93 mobile cellular subscriptions per person for the world. This value drops down

only to 0.63 for a less developed region like Sub-Saharan Africa. In this context, it can be argued that data generated from mobile phone use has a lot of implications for use in transportation modeling due to its richness and ubiquity. Our goal is to test the power of such data for various cities across the world against the results obtained from surveys obtained by local authorities. We show that at certain aggregation levels, CDRs have the potential to inform four step models quite strongly, and at least support the household travel survey in its role in transport modeling.

The fourth section of this work is focused on bringing together the two concepts targeted separately in the previous two sections, namely road networks and travel demand, while providing a comparison of cities to understand the aforementioned interplay to improve on the literature reviewed in Sections 1.1.2 and 1.1.5 from a transportation point of view. Currently, most urban areas are ridden with congestion. Increasing population along with agglomeration of goods and opportunities leads to increased population densities, which in turn leads to peaks in demand for infrastructures. Energy and mobility are prime examples, load curves for urban areas are highly peaked, and access to downtown areas in the morning peak traffic is often infuriatingly difficult. Specifically in developing countries, congestion levels are staggeringly high: TomTom, a leading GPS company has reported above 60% congestion in cities like Moscow, Istanbul, Rio de Janeiro and Mexico City. In efforts to alleviate congestion, cities rely on construction of bigger roads, introduction of new bus lines, carpool lanes, in more dire cases congestion pricing and most extremely road space rationing. Drivers, to escape traffic, have increasingly resorted to real time traffic tools which, when collectively used, arguably makes traffic worse. In this section, we begin by implementing a modern static traffic assignment model to obtain comparable estimates of travel times for five cities: Boston, San Francisco Bay Area, Rio de Janeiro, Lisbon, and Porto. We compare our findings to values obtained from an online map provider. We analyze how total vehicle kilometers travelled and the total network capacity, when combined with population density, can help explain a city's response to travel demand. We then analyze the potential travel time benefits of socially aware routing behavior. From another perspective, we calculate how much

time is lost solely due to the fact that drivers minimize their own travel times. We demonstrate distributions of benefits and losses at the individual level, that add up to generate overall benefit for the society.

In the fifth section, we move our focus towards the electrification of transportation, namely the electric vehicles (EVs). The electrification of transportation introduces a spatiotemporal tie between the traditionally independent power and transportation infrastructures through EV charging. Due to the mobile nature of energy storage in EVs and the dependency of the resulting energy demand to the trip characteristics, understanding this relationship is crucial. In this section, we use origin-destination information obtained from mobile phone location data along with EV charging session data to study the coupling between energy demand and commuting patterns in Bay Area, California. We first estimate the mobility patterns specific to EV drivers from origin destination information obtained from mobile phone data. We then explore charging behavior to characterize arrival and departure times, visitation patterns, session durations and flexibility. Next, we analyze the relationship between commuting behavior and energy consumption by coupling the mobility information with energy demand. Finally we develop a smart charging scheme that shaves peak power load by incorporating smart charging and arrival hour modification, and quantify and assess the potential benefits and the applicability of such solutions.

In the sixth and final section, we conclude this dissertation with an overarching discussion and potential avenues for future work.

Chapter 2

On the role of spatial dynamics and topology on network flows

Particle flows in spatial networks are susceptible to congestion. In this section, we analyze the phase transitions of these networks to a state of congested transport and the influence of both topology and spatial dynamics on its emergence. We systematically show that the value of the critical loading rate at which congestion emerges is affected by the addition of spatial dynamics, changing the nature of this transition from a continuous to a discontinuous one. Our numerical results are confirmed by introducing an analytical solvable framework. As a case of study, we explore the implications of our findings in the San Francisco road network where we can locate the roads that originate the congested phase. These roads are the spatially constrained, and not necessarily those with high betweenness as predicted by models without spatial dynamics.

2.1 Introduction

Flow networks are inherently liable to congestion. The ability of these networks to handle demand at reasonable levels is crucial as otherwise a congested phase of

This chapter is based on [48].

transport affects the performance across the entire network. Therefore, it is of prime interest to analyze how and where networks begin to undergo a transition to a congested state and the dynamics of its response. Dissecting flow patterns is essential to address this problem. In this context, flow of data packets in the Internet is well understood, as analyses of their traffic dynamics and phase transitions are abundant [125, 60, 157, 55, 8, 110]. The transition point to congestion is in this case well established through analytical solutions and simulations [6, 71, 72, 187, 178]. To interpret the role of a network backbone for managing flows, optimal paths and minimum spanning trees have been studied [163, 177, 40]. The most relevant metric to determine the vulnerability of internet flows is the betweenness centrality for shortest paths, because it determines the critical element generating congestion in the network [63, 13].

All these previous studies have overlooked the role of space and time in particle flows in the networks; a reasonable assumption in Internet applications. In this case the travel time between nodes is negligible and queues accumulate in the nodes which have limited capacity to process the packets and congestion occurs. In contrast, in transportation networks travel time of vehicles or individuals is crucial. Queues are formed on the links which have limited spatial capacity, and in turn their velocity depends on the density of travelers on each link. In this context, transportation research focuses on capturing traffic flow by making use of the fundamental diagrams that empirically relate flow, density and speed in road segments and utilizing them in macroscopic link models [67, 53]. The cell transmission model [51, 52] and the simple point queue models [92, 93, 120] are well established among such traffic flow models. Alternatively, cellular automata models for vehicular traffic [117, 43] have also been used to mimic traffic flow behavior along with many other discrete stochastic models [146, 116]. In none of these cases, the interaction of the spatial dynamics with the network topology have been addressed. Our goal here is to build on the simple model of the Internet by adding temporal and spatial dimensions of particle flow and hence construct a framework to analyze congestion in spatial networks from a network science perspective. The proposed framework can be extended to flows in other kinds of spatial networks [10, 14, 105], and more importantly could open new avenues of

the research on urban road networks that would go beyond modeling and topological analysis for the statistical physics community[139, 15, 183].

2.2 Methods

We begin by recalling the scheme in [6, 72, 8, 157, 178, 187, 71, 55, 125, 60] and refer to it as the internet model (IM): the network is loaded with R identical particles at each timestep t with randomly assigned origins and destinations. A fixed shortest path routing table guides particles towards their destination. Nodes can transmit as many particles per timestep as their *outflow capacity*, C , and travel between two nodes takes a unit timestep. Queues of particles form at the nodes, and they can grow infinitely large. Particles are exempt from joining the queue at their destination and are removed from the system upon arrival. The network response is measured by the order parameter H [72]:

$$H(R) = \lim_{t \rightarrow T} \frac{\langle \Delta W \rangle}{R \Delta t}, \quad (2.1)$$

where W denotes the number of particles in the system, $\langle \Delta W \rangle$ is the average change in the number of particles still in the system after a timestep, Δt is the unit timestep and T is the length of the simulation. Figure 2-1 (a) and (b) depict the IM. For low values of R , the network reaches a rate of particle arrival equal to the loading rate. W remains constant and consequently $H = 0$. Conversely, if R exceeds a certain threshold R_c , a linear increase in W with a slope of H is observed due to excessive queueing. This behavior maps a second-order phase transition to congestion.

Spatial networks, being embedded in two-dimensional space, give rise to three heterogeneities that need to be captured. First is non-uniform travel times. This differs from the Internet where data packets hop from one node to the next in a single timestep. Second, these networks carry flows along the links. A particle in a spatial networks has a specific position on the link it is traveling on. The third source of heterogeneity is a consequence of this: particles occupy physical space and gradually

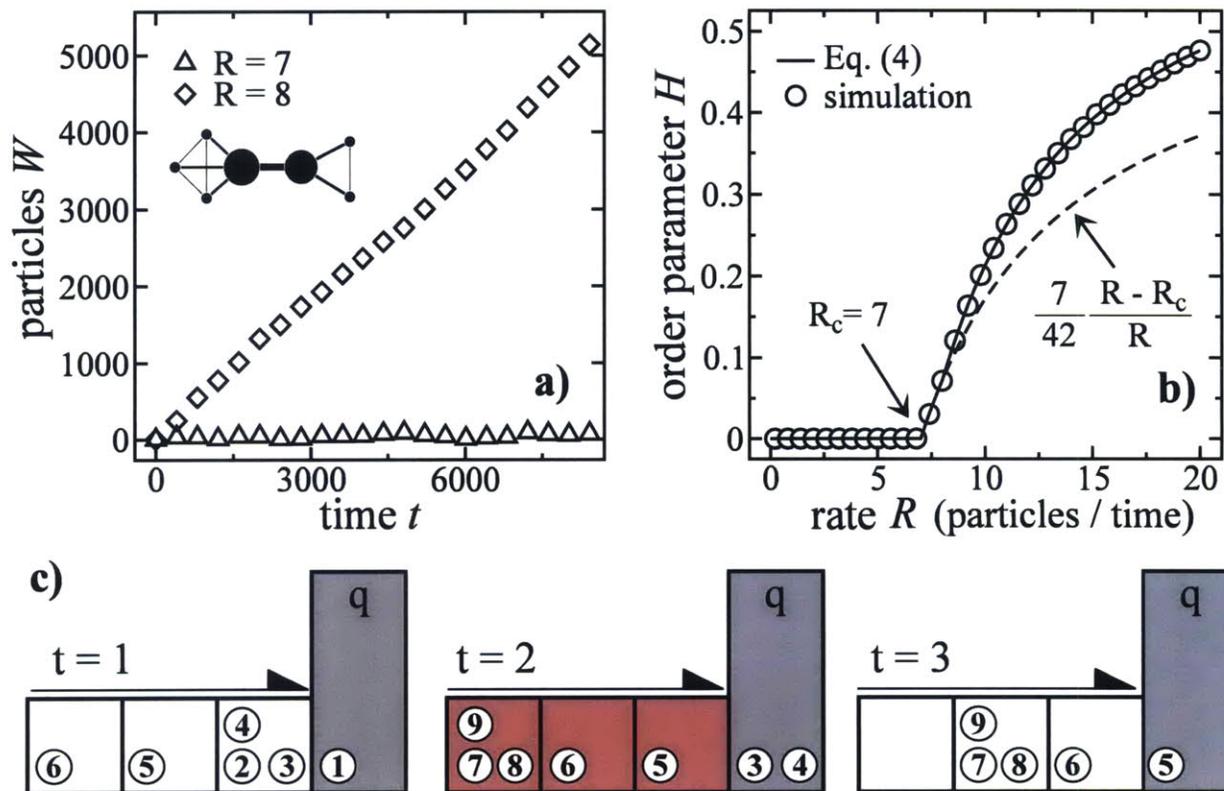


Figure 2-1: (a) Number of particles W versus time t for the Internet model with different loading rates for the shown simple network with node and edge betweenness values mapped as node size and edge width, respectively. (b) A second order transition at the critical value $R_c = 7$ for this network. (c) A schematic representation of the SPQM for a link with $C = 2$, $\tau = 3$ and $V = 7$. At $t = 2$, the link reaches its volume capacity and no incoming particle is allowed entering the link.

fill the segment. Among the various models of traffic flow in the transportation literature aimed to address these issues, the point-queue model (PQM) used in [120, 93, 92, 59] is an adaptation of the IM that shifts flow from nodes to links, incorporating the non-uniform travel time distribution and thereby making the flow analysis very similar to that of the internet. Particles traverse the link freely by hopping through τ slots of unit travel time to join a queue at the end of the link from which they will be discharged at the outflow capacity. The total travel time consists of the free travel time and the delay, namely, the timespan between the particle entering the queue and exiting it. The spatial point-queue model (SPQM) incorporates a single additional constraint to the PQM: every link has an upper limit for the number of particles it can hold at once. We will refer to this value as the *volume capacity* of the link, V . Links cannot accept any new particles when they reach their volume capacity, as illustrated in Figure 2-1 (c). This additional constraint has a crucial effect on the nature of the network response, as links at volume capacity clog upstream links and cause them to succumb to congestion as well. This spreading of congestion occurs at rates that depend on the loading rate and the network topology along with specific link properties. The rate of particles unable to travel determines the speed with which the congestion spills, which makes the spreading process non-binary unlike traditional spreading models in the literature. Although the SPQM share some aspects with several directed percolation models, the movement of non-identical particles along predetermined spatial shortest paths with non-binary spreading is uncommon and therefore relatively unstudied [75, 83].

2.3 Results

The critical loading rate R_c^{IM} has been shown [6] to be equal to

$$R_c^{\text{IM}} = N(N - 1)(C_{\text{max}}/B_{\text{max}}^N), \quad (2.2)$$

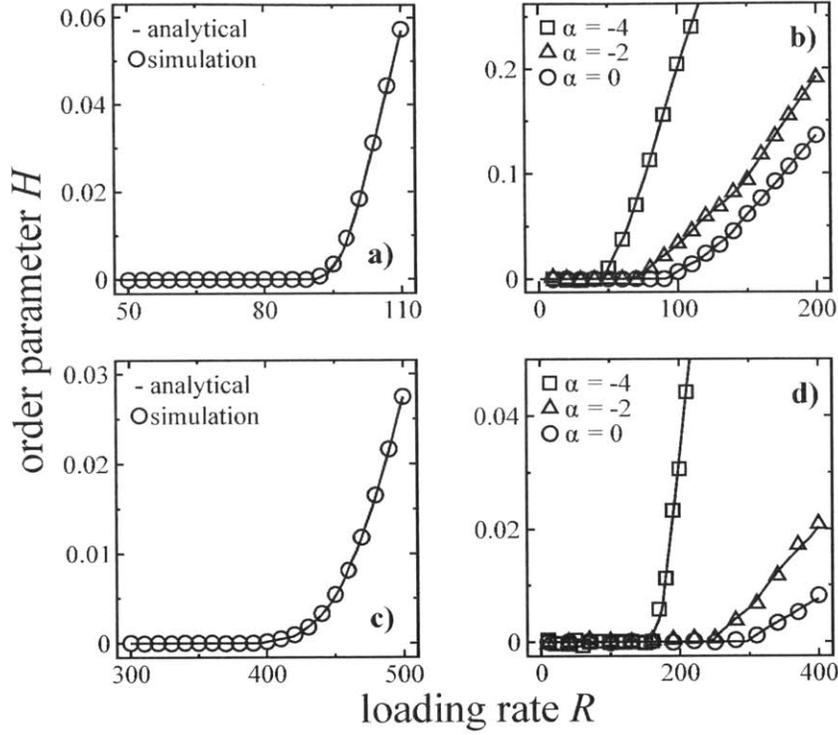


Figure 2-2: Transitions of a non-periodic lattice as estimated by simulations and by analytical solutions for (a) the IM and (c) the PQM. Transitions of a completely rewired network with $\rho = 1$ for (b) the IM and (d) the PQM for varying tendencies α to have long distance links. Analytical solutions are consistent with simulations.

where N is the network size measured by the number of nodes, B_{\max}^N is the maximum node betweenness and C_{\max} is the outflow capacity of this node. This relation arises from the fact that inflow to a node is proportional to its betweenness centrality [69]. At R_c , the inflow is equal to the outflow at the node with the minimum C_{\max}/B_{\max}^N value. For the PQM, we adjust this equation by replacing the node betweenness by an edge betweenness value,

$$R_c^{\text{PQM}} = N(N-1)(C_{\max}/B_{\max}^{\text{E}*}) \quad (2.3)$$

, where $B_{\max}^{\text{E}*}$ refers to the maximum *modified betweenness*. This modification is necessary since particles do not join the queues in the final links, which should be omitted from the edge betweenness calculations.

Next, we introduce a framework to analytically calculate the entire transition curve to congestion. For $R > R_c$, particle inflow at certain elements will be larger than the outflow. We define R_c^i as the critical loading rate specific to element i . For large R , the outflow of congested links are maximized to capacity, which in consequence affects the inflow to the links downstream. To account for this we define the *delay factor*, $D_i(R)$, referring to the fraction of paths through i that are not suffering from delay as,

$$D_i(R) = \mathcal{H}(R_c^i - R) + \frac{C_i}{I_i(R)} \mathcal{H}(R - R_c^i), \quad (2.4)$$

where $\mathcal{H}(x)$ is the Heaviside step function and C_i and $I_i(R)$ are the outflow capacity and the inflow of element i for loading rate R . $D_i(R) = 1$ suggests no congestion for element i , whereas lower values indicate levels of congestion. Using this definition, the inflow $I_i(R)$ at a specific loading rate R can be quantified as,

$$I_i(R) = \frac{\sum_{k \in \Gamma(i)} \prod_{j \in k} R D_j(R)}{N(N-1)} \quad (2.5)$$

$$H(R) = \sum_{i \in \mathcal{N}} (I_i(R) - C_i) \mathcal{H}(I_i(R) - C_i), \quad (2.6)$$

where $\Gamma(i)$ is the set of paths passing through element i . Eq.(2.6) accounts for all the delay factors of the elements upstream of element i by going through the shortest paths. Eq. (2.4) and (2.5) form a set of coupled equations that can be solved to obtain the inflows for every element. $H(R)$ is obtained by summing all positive values of $I_i(R) - C_i$. Figure 2-1 (b) also reveals the exact solution for the simple network around the critical point with the dashed curve $H = (7/42)(1 - R_c/R)$. In order to test different network topologies and examine the effect of space on criticality, we use a non-periodic lattice as a substrate and rewire each edge (i, j) with probability ρ to a new destination j^* chosen with probability proportional to $d(i, j^*)^\alpha$ where $d(i, j)$ denotes the Euclidean distance [174, 91]. Figure 2-2 reveals that the simulations and analytical results perfectly coincide for transitions in both a two dimensional non-periodic lattice of $N = 1225$ and $C = 4$, and its completely rewired instance. Critical rates decrease for smaller ρ as newly introduced shortcuts have higher betweenness

values. As ρ increases, the value of α and its effect on network topology becomes more pronounced: rewired links in networks with lower α values are more localized and therefore maximum betweenness values are higher in these networks. Consequently the increase in H is sharper for lower α . Figure 2-2 (b) and (d) show sharper transitions and lower R_c for decreasing α values for network instances with $\rho = 1$.

At R_c , the link that triggers congestion, also referred to as the critical element, is expected to fluctuate between free flow and congested phases. Figure 2-3 (a) exhibits the frequency distribution of the timespans at which this most critical element operates at its outflow capacity, as an indicator of the temporality of the phase transition. Results show that these timespans follow a power law with exponents of -0.58 ± 0.04 for the IM and -0.48 ± 0.04 for the PQM, independent of the network topology. SPQM exhibits a different behavior. For low volume capacities links tend to fill up, causing links upstream to fill as well. Fluctuations cause a *gridlock*, a condition where all elements of a cycle are completely filled and hence flows stop. In case of a gridlock, the order parameter increases very sharply. In Figure 2-3 (b), for varying volume capacities we measure the average number of timesteps it takes for a gridlock to occur, t^g , normalized by the length of the simulation T . It can be observed that for lower V , gridlocks are observed relatively quickly. For large volume capacities, PQM and SPQM have the same R_c , which we will refer to as the PQM-limit. In either model, at steady state, queues are not expected to be necessarily empty but rather steady in their size. If the volume capacity of a link is smaller than this steady state queue size, particles will be blocked in the upstream link which consequently may suffer from a decrease in its outflow due to this clogging effect. This suggests a lower critical point for a network with active volume capacity constraints, correspondingly $R_c^{\text{PQM}} \geq R_c^{\text{SPQM}}$. Therefore, to realize the congestion-free transport to the fullest, the PQM-limit should be aimed. Figure 2-3 (c) illustrates the response of a non-periodic lattice with $N = 625$ for decreasing levels of volume capacity and reveals that both the critical loading rate and the nature of the transition is affected by V . Figure 2-3 (d) shows the effect of the volume levels on R_c . In point queue models, the number of particles on a link is the sum of those in the queue, which was shown to be pro-

portional to the modified betweenness, and those that are traveling. The expected inflow to a single slot of travel on a link is proportional to the actual betweenness of that link, hence the expected number of particles traveling is τB^E . Therefore for small V and strictly deterministic inflows, a link reaches its maximum volume when $V = R_c (\tau B^E + B^{E*})_{\max} / N(N - 1)$. The inflow to the link is proportional to the actual routing betweenness so the critical point can be expressed as,

$$R_c = N(N - 1) \min \left(\frac{C}{B_{\max}^{E*}}, \frac{V}{(\tau B^E + B^{E*})_{\max}} \right), \quad (2.7)$$

where B^E is the edge betweenness. Consequently a linear increase in R_c is observed up to the point where $V = C(\tau B^E + B^{E*})_{\max} / B_{\max}^{E*}$. However, the stochastic nature of the model, along with the first-order nature of its transition, causes fluctuations that force R_c/C below this analytical bound as it converges to the PQ-limit.

Figure 2-4 depicts the transitions for the PQM and the SPQM for the San Francisco road network with $N = 1152$ and an average degree of 3.2. The network is discretized by unit travel times of 10 seconds. Outflow capacity of a road segment is obtained by using the speed limit and the number of lanes. Volume capacities are estimated for every road segment assuming that the volume capacity is reached when speed drops to half of the speed limit. Under these assumptions, the PQM-limit is not reached as $R_c^{\text{PQM}} = 40$ (14400 vehicles/h) and $R_c^{\text{SPQM}} = 30$ (10800 vehicles/h). To capture network response in the SPQM, segment volumes are recorded at different time steps of the simulation $t = 720$ (2h), $t = 1440$ (4h) and $t = 2160$ (6h) for $R = 36$, slightly above the critical load. Figure 2-4 illustrates the network response by mapping road segments that have reached 80 percent of their volume capacities in the given time periods by color. Congestion originates from an artery leading to the downtown area and anisotropically spreads to other regions. After 6 hours, most network elements are suffering from congestion.

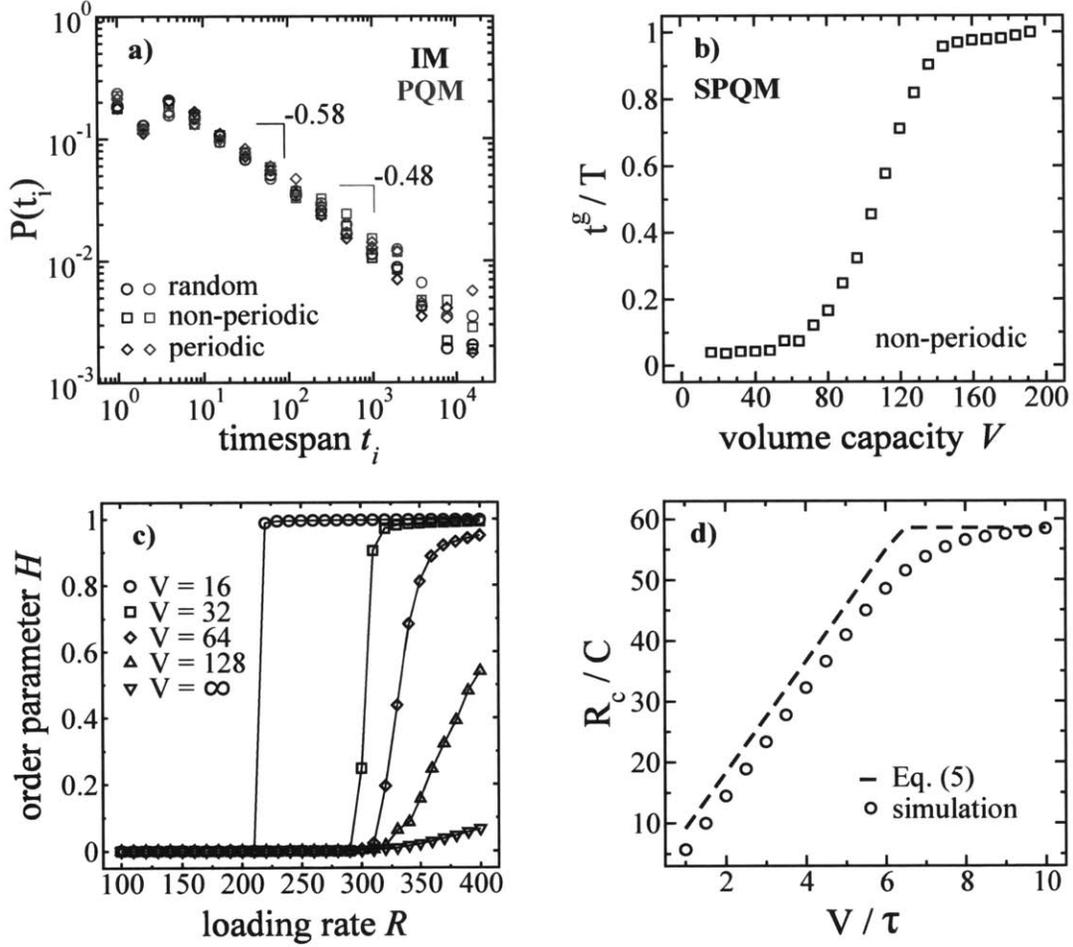


Figure 2-3: (a) Distributions of the timespans through which the congested element operates at its outflow capacity at R_c for the IM and PQM, following a power law. (b) The average number of timesteps normalized by T before a gridlock is observed, versus the volume capacity. (c) Change in the nature of the transition to congestion of a non-periodic lattice of size $N = 625$ with specified levels of volume capacities. (d) The critical loading rate R_c for varying V/τ values. R_c settles to the PQM-limit as V/τ increases.

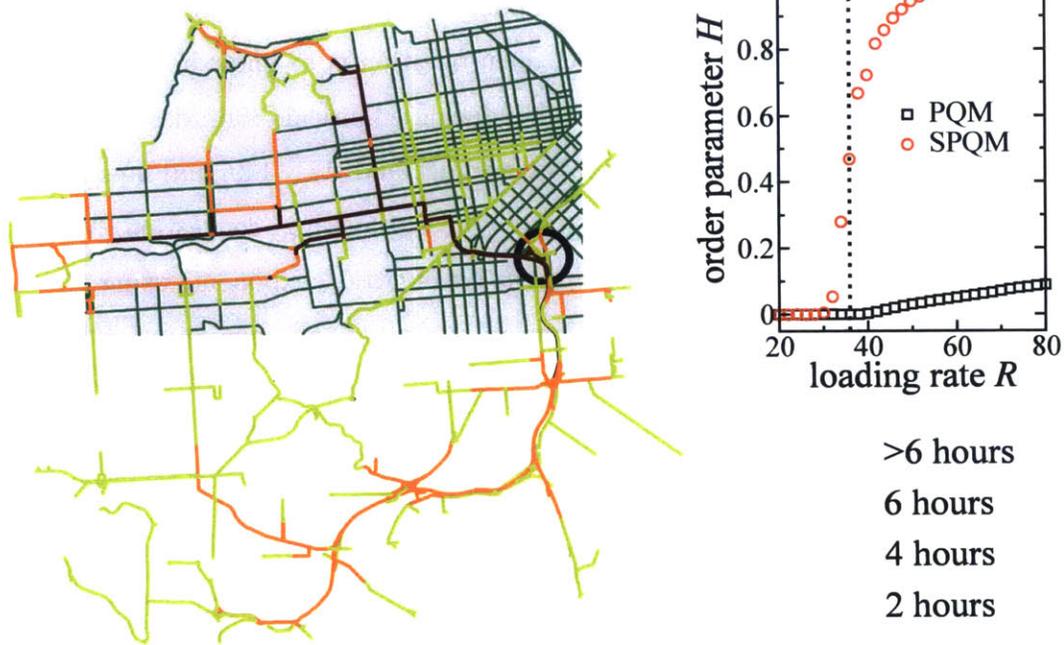


Figure 2-4: A congestion map of the San Francisco road network for $R = 36$ vehicles per timestep. Colors represent the times within which the road segments become congested as a consequence of the spillover. The black circle denotes the origin of congestion. (inset) PQM and SPQM transitions for the San Francisco road network, where $R_c^{\text{PQM}} = 40$ and $R_c^{\text{SPQM}} = 30$.

2.4 Discussion

Our findings suggest that in the SPQM, the critical road segment cannot accept incoming vehicles as it saturates to its volume capacity before it reaches its outflow capacity. This outcome can be traced back to the inherence of congestion in downtown areas: cities with high population densities have concentrated spatial demand distributions, which result in the inadequacy of urban space to accommodate such concentrated flows. This work is a step further on a systems analysis applied to congestion in roads. In further studies, population and facility distributions can be modeled. In the applications domain an open question is to know how the studied transition to congestion is influenced by introducing empirical origin-destination matrices that represent the real population's travel demand [171, 149] and how network topology contributes towards reducing- or maybe aggravating- congestion. While in the domain of phase transitions, analyzing the set of critical exponents using directed percolation as a benchmark case remains as an intriguing and open question [75, 83].

Chapter 3

Extracting Origin Destination Information from Mobile Phone Data

Rapid urbanization is placing increasing stress on already burdened transportation infrastructure. Ubiquitous mobile computing and the massive data it generates presents new opportunities to measure the demand for this infrastructure, diagnose problems, and plan for the future. However, before these benefits can be realized, methods and models must be updated to integrate these new data sources into existing urban and transportation planning frameworks for estimating travel demand and infrastructure usage. While recent work has made great progress extracting valid and useful measurements from new data resources, few present end-to-end solutions that transform and integrate raw, massive data into estimates of travel demand and infrastructure performance. Here we present a flexible, modular, and computationally efficient software system to fill this gap. Our system estimates multiple aspects of travel demand using call detail records (CDRs) from mobile phones in conjunction with open- and crowdsourced geospatial data, census records, and surveys. We bring together numerous existing and new algorithms to generate representative origin-destination matrices, route trips through road networks constructed using open and crowd-sourced data repositories, and perform analytics on the system's output. We

This chapter is based on [45] and [164].

also present an online, interactive visualization platform to communicate these results to researchers, policy makers, and the public. We demonstrate the flexibility of this system by performing analyses on multiple cities around the globe. We hope this work will serve as unified and comprehensive guide to integrating new big data resources into customary transportation demand modeling.

3.1 Introduction

The accelerating growth of cities has made the estimation of travel demand and the performance of transportation infrastructure a critical task for transportation and urban planners. To meet these challenges in the past, methods such as the widely used four-step model and more recent activity based models were developed to make use of available data computational resources. These models combine meticulous methods of statistical sampling in local [50, 150] and national household travel surveys [158, 138] to process and infer trip information between areas of a city. The estimates they produce are critically important for understanding the use of transportation infrastructure and planning for its future [169, 155, 104, 98, 78, 77, 76, 101, 41, 21].

While the surveys that provide the empirical foundation for these models offer a combination of highly detailed travel logs for carefully selected representative population samples, they are expensive to administer and participate in. As a result, the time between surveys range from 5 to 10 years in even the most developed cities. The rise of ubiquitous mobile computing has lead to a dramatic increase in new, *big data* resources that capture the movement of vehicles and people in near real time and promise solutions to some of these deficiencies. With these new opportunities, however, come new challenges of estimation, integration, and validation with existing models. While these data are available nearly instantaneously and provide large, long running, samples at low cost, they often lack important contextual demographic information due to privacy reasons, lack resolution to infer choices of mode, and have their own noise and biases that must be accounted for. Despite these issues, their use for urban and transportation planning has the potential to radically decrease the

time in-between updated surveys, increase survey coverage, and reduce data acquisition costs. In order to realize these benefits, a number of challenges must be overcome to integrate new data sources into traditional modeling and estimation tools.

Analyzed on its own, data generated by the pervasive use of cellular phones has offered insights into abstract characteristics of human mobility patterns. Recent work has found that individuals are predictable, unique, and slow to explore new places [70, 32, 56, 154, 153, 39, 37]. The availability of similar data nearly anywhere in the world has facilitated comparative studies that show many of these properties hold across the globe despite differences in culture, socioeconomic variables, and geography. The benefits of this data have been realized in various contexts such as daily mobility motifs [144, 147], disease spreading [20, 175] and population movement [102]. While these works have laid an important foundation, there still is a need to integrate these data into transportation planning frameworks.

To make these new data useful for urban planning, we must clarify their biases and build on the progress made by transportation demand modeling even in the face of limited data resources. We must combine this domain knowledge with new algorithms and metrics to better understand travel behaviors and the performance of city infrastructure and we must update technologies to accommodate the computational requirements of processing massive geospatial data sets. Individual survey tracking and stay extraction [7], OD-estimation and validation [36, 121, 171, 84], traffic speed estimation [11, 186], and activity modeling [129, 135] have all been explored using new massive, passively collected data. However, these studies generally present alternatives for only a few steps in traditional four-step or activity based models for estimating travel demand or fail to compare outputs to travel demand estimates from other sources. Moreover, many methods offered to date lack portability from one city to many with minimal additional data collection or calibration required.

Here we fill this gap with a modular, efficient computational system that performs many aspects of travel demand estimation billions of geo-tagged data points as an input. We review and integrate new and existing algorithms to produce validated origin-destination matrices and road usage patterns. We begin by outlining the system

architecture in section 3.2.1. In section 3.2.1 we explain our methods of extracting, cleaning, and storing road network information from a variety of sources. We discuss recent advances in OD creation from mobile phone data in section 3.2.2 and implement a simple, parallel incremental traffic assignment algorithm for these trips in section 3.2.2. We present comparisons of these results to estimates from traditional survey methods in section 3.3.1. Finally, in sections 3.3.2, 3.3.3, 3.3.4 we present a variety of measurements that can be made with the proposed system as well as an online, interactive visualization for conveying these results to researchers, policy makers, and the public. To demonstrate the flexibility of the system, we perform these analyses for five metro regions spanning countries and cultures: Boston and San Francisco, USA, Lisbon and Porto, Portugal, and Rio de Janeiro, Brazil.

3.1.1 Description of Data

Travel surveys are typically administered by state or regional planning organizations and are integrated with public data such as census tracts and the demographic characteristics of their residents, made available by city, state, and federal agencies. New data sources, however, come from new providers. Large telecommunications companies, private applications, and network providers collect and store enormous quantities of data on users of their products and services, presenting computational challenges for storing and analyzing them. Billions of phone calls must be processed, data from open- and crowd- sourced repositories must be parsed, and results must be made more accessible to individuals that generated them. At the same time, it is critical that measurements from these new sources are statistically representative and corrected for biases inherent in new data. This process requires integration of new pervasive data with reliable (though less extensive) traditional data sources such as the census or travel surveys. We combine the following data sets to illustrate the capabilities of the system architecture here proposed:

1. *Call Detail Records (CDRs)*: At least three weeks of call detail records from mobile phone use across each subject city. The data includes the timestamp

and the location for every phone call (and in some cases SMS) made by all users of a particular carrier. The spatial granularity of the data varies between cell tower level where calls are mapped to towers and triangulated geographical coordinate pairs where each call has a unique pair of coordinates accurate to within a few hundred meters. Market shares associated with the carriers that provide the data also vary. Personal information is anonymized through the use of hashed identification strings. For reference, 6 weeks of CDR data from the Boston area containing roughly 1 billion calls made by 1.6 million unique users consumes roughly 70 gigabytes of disk space in its raw format. In cities with longer observation periods, data size quickly becomes a performance issue.

2. *Census Data:* At the census tract (or equivalent) scale, we obtain the population and vehicle usage rate of residents in that area. For US cities, the American Community Survey provides this data on the level of census tracts (each containing roughly 5000 people). Census data is obtained for Brazil through IBGE (Instituto Brasileiro de Geografia e Estatística) and for Portugal through the Instituto de Nacional de Estatística. All cities analyzed in this work have varying spatial resolutions of the census information.
3. *Road Networks:* For many cities in the US, detailed road networks are made available by local or state transportation authorities. These GIS shapefiles generally contain road characteristics such as speed limits, road capacities, number of lanes, and classifications. Often, however, these properties are incomplete or missing entirely. Moreover, as such road inventories are expensive to compile and maintain, they simply do not exist for many cities in the world. In this case, we turn to OpenStreetMaps (OSM), an open source community dedicated to mapping the world through community contributions. For cities where a detailed road network cannot be obtained, we parse OSM files and infer required road characteristics to build realistic and routable networks. At this time, the entirety of the OSM database contains roughly 4 terabytes of geographic features related to roads, buildings, points of interest, and more.

Table 3.1: A comparison of the extent of the data involved in the analysis of the subject cities.

	City				
	Boston	SF Bay	Rio	Lisbon	Porto
Population (mil.)	4.5	7.15	12.6	2.8	1.7
Area (1000km ²)	4.6	18.1	4.5	2.9	2.0
# of Users (mil.)	1.65	0.43	2.19	0.56	0.47
# of Calls (mil.)	905	429	1,045	50	33
# of cell towers	N/A	892	1421	743	335
# of Edges (ths.)	21.8	24.3	22.7	28.1	15.1
# of Nodes (ths.)	9.6	11.3	22.1	16.1	8.6
# of Tracts	732	1139	729	295	272

4. *Survey and Model Comparisons:* Wherever possible, we obtain the most recent travel demand model or survey from a particular city and compare the results to those output by our methods. In Boston, we use the 2011 Massachusetts Household Travel Survey (MHTS) and upscale trips according to standard procedures, in San Francisco, the 2000 Bay Area Transportation Survey (BATS), in Rio de Janeiro, a recent transportation model output provided by the local government, and in Lisbon, the most recent estimates from the MIT-Portugal UrbanSim LUT model that uses the 1994 Lisbon transportation survey as input[61]. We found no recent travel survey or model for Porto.

Table 3.1 compiles descriptive statistics for these data sources for each city we explore in the latter sections of this section.

3.2 Methods

3.2.1 System Architecture and Implementation

Architecture

The system architecture to integrate the data sources above must be flexible enough to handle different regions of the globe which may have different data availability and quality and efficient enough to analyze massive amounts of data in a reasonable

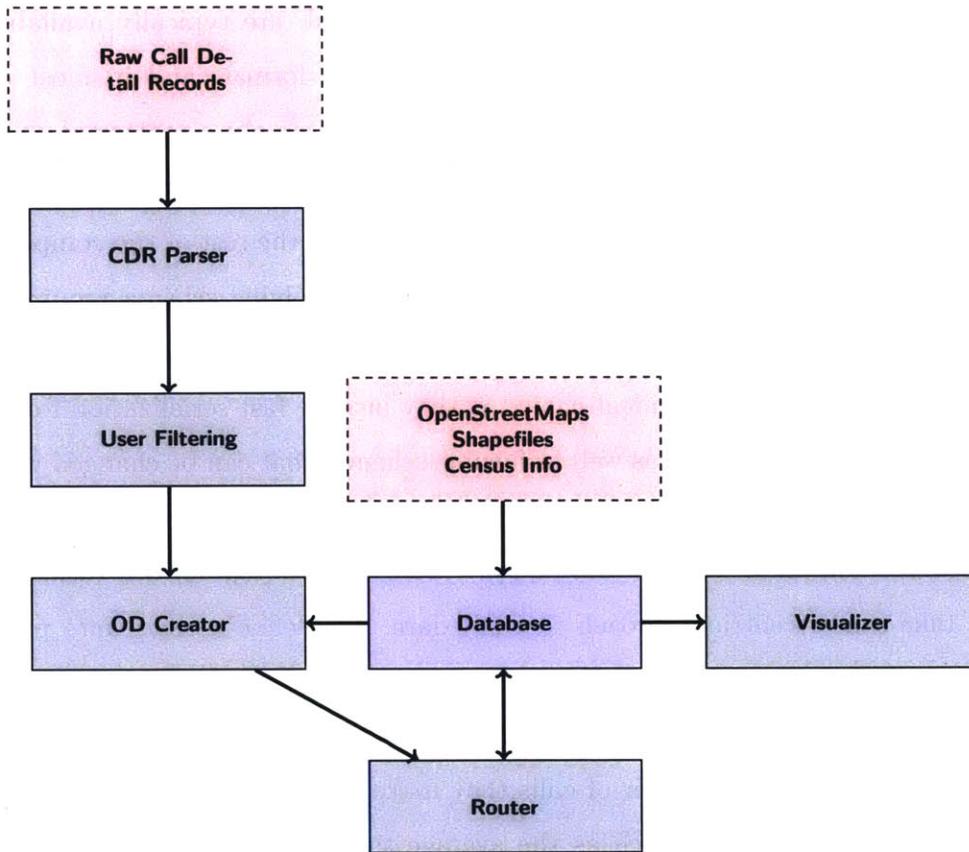


Figure 3-1: A flowchart of the system architecture.

amount of time. The proposed system must also be modular, so that components can be updated easily as new technologies and algorithms become available. To meet these requirements, we choose an object-oriented approach with loose schema requirements. A final object is to make results accessible to a range of end users via online, interactive visualization. To satisfy these constraints, we propose the system architecture depicted in Figure 3-1.

Parsing, Standardizing, and Filtering User Data

One of the biggest challenges in parsing and analyzing travel survey data is the incredible variety in data schema, collection, and reporting practices. Each planning organization typically constructs its own set of data codes and definitions and provides

data in unique formats. This makes it very difficult to compare surveys done in different cities. Call detail records, on the other hand, are typically available for many cities from the same provider and in the same format, and in most cases, translating between the formats of different carriers is simply a matter of shuffling columns. The first component of our system is a simple architecture to convert all CDR data to a standard format that can be expected by the rest of the components.

Given the size of these data sets and the rapidly evolving schema requirements of new models, choosing the proper data structure is critical. Google's open source Protocol Buffer Library is an ideal choice as they provide fast serialization for speed and space efficient file storage as well as flexible schemas that can be changed without compromising backwards compatibility. These structures were designed to serve some of the largest databases in the world and are more than enough for our task.

We take a user centric approach to CDR data. We define a *user_data* protocol buffer message that will form the core data structure for our custom User class in an object-oriented programming model. Each User object can be assigned a number of attributes such as the number of calls they make, their home and work locations, and mobility characteristics such as the average time between calls or the average distance traveled on each trip. More sophisticated methods can compute the number and distribution of their trips and even expand them based on census information. We define similar structures and classes for OD matrices, trips, and census data. The serialization routines built into the protocol buffer library ensures that storage of raw data is efficient. To analyze a new city, the user only needs to write two simple routines, one to parse a single line of the CDR file and populate relevant user attributes and one to populate census data objects. Standardizing the CDR data format in this way makes it very easy to compare the output of our estimation models across different cities.

Creating and storing geographic data

A relational database is used to store road network and census information for every city in a standard format. Given the current cost of computing resources, these

systems provide adequate performance for storing static GIS and census data and have convenient, mature interfaces for easy access. We also use this database to store aggregated results from our estimates so that they can be made available to interactive web APIs and visualization platforms. We use a Postgres and the open source spatial extension PostGIS to store and manipulate census and road network data.

While census tract or TAZ (Traffic Analysis Zone) polygons and demographic information are stored in this database, it is computationally inefficient to perform point-in-polygon calculations for each user or call record in our CDR dataset. To dramatically speed these computations, we rasterize polygons into a small pixel grid, where pixel values is a unique identifier for the census tract covering that pixel. This raster is then used as a look-up table to convert the latitude and longitude of calls into census tract IDs. The rasterization introduces some error along the borders of tracts, but these errors are minimized by making pixel sizes much smaller than the size of the raster and resolution of the location estimates of calls (between 10m and 100m).

While the platform supports road networks supplied by local municipalities in the form of shapefiles, we have implemented a parser to construct routable road networks from OpenStreetMap (OSM) data due to its global availability. Transportation networks in OSM are defined by *node* and *way* elements. Nodes represent points in space that can refer to anything from a shop to a road intersection, while ways contain a list of references to nodes that are chained together to form a line. In our context, relevant ways are those used by cars and relevant nodes are intersections within the road network. Ways and nodes may also contain a number of tags to denote attributes such as "number of lanes" or "speed limit". Many roads, however, do not include the whole set of attributes necessary for accurate routing. For example, city roads often lack speed limit information required to estimate the time cost, which in turn is used to find shortest paths based on total travel time. To infer this missing data, our system supports the creation of user-defined mappings between highway types and road properties. For example, ways tagged as "motorways" are generally major

highways and have a speed-limit of 55 mph in the Boston area. They tend to have 3 lanes in each direction. "Residential" roads, on the other hand, have a speed-limit of 25mph and 1 lane in each direction. Each road segment is also given a capacity based on formulas suggested by the US Federal Highway Administration. Using these mappings, we parse the OSM xml data to create a routable, directed road graph with all properties required to estimate realistic costs driving down any given road.

We implement two additional cleaning steps to improve efficiency. The first filters out irrelevant residential roads. These small local roads are filtered from our network, as they are not central to the congestion problem, yet tend to increase computation time significantly. Finally, in OSM data, a node object can refer to many things, for example an actual intersection or simply a vertex on a curve used to draw a turn. The latter case results in a network node with only one incoming and one outgoing edge (assuming U-turns are not allowed). These nodes are superficial and increase network size and routing algorithm run times needlessly. We simplify networks by removing these nodes from the network and only connecting true intersections, keeping the geographic coordinates of the nodes so that link costs still reflect actual geographic length of roads rather than straight line distances between start and end points. The parsed and cleaned edges are then loaded into the Postgres database, preserving attributes and geometry. Pseudo-code of the algorithm to parse and simplify OSM networks can be found in Algorithm 1 in the supplementary materials.

3.2.2 Estimating Origin-Destination Matrices

The following sections review algorithms for transforming billions of geo-tagged data points into validated origin destination matrices and assigning these flows to transportation infrastructure. Some of these algorithms are important for their deviation from traditional approaches and some are important for their computational efficiency, a requirement when faced with such massive data sets.

Measuring Flow

Current methods to estimate the flow of people or vehicles from place to place in a city generally fall into two categories: four-step or activity based approaches. The former class of models breaks the process into a sequence of four steps from which it earns its name. The first three steps in a four-step model – trip generation, distribution, and mode choice – are designed to estimate origin-destination matrices containing the number of trips from place to place within a city. Traditional modeling approaches use data from travel surveys possibly combined with land use and point of interest information to generate estimates of trip production and attraction for locations. These trips are then distributed from their point of origin to destinations across the city using gravity or radiation models. Modes of transit are assigned using models estimated from survey data and information on the transit infrastructure. More recent activity based models approach travel demand from an individual level. Assuming that travel demand is created by the need to fulfill activities, these models use similar survey data to estimate utility curves for travels and predict behaviors using probit or logit models based on these preferences.

While new data sources such as CDRs do not provide the same detailed demographic and contextual information about individuals or trips, they do provide an opportunity to measure travel more directly. With billions of data points, high spatiotemporal resolution, and long observation periods, passive data collected by mobile devices provide unparalleled scale of observation. New methods to estimate travel demand must balance trade offs between small, but complete data for a short period of time and large, but incomplete data over a longer period of time. In both cases noise and biases must be carefully dealt with to produce valid measurements. In this section we adapt and integrate previous works that have tackled parts of this problem into a full implementation of travel demand estimation for cities.

Mobile phones offer good, but imperfect measurements of geographic position. The coordinates of a mobile phone event are either recorded as the location of a nearby tower through which the event was routed or as a triangulation based signal

strength from multiple towers. This creates uncertainties of a few hundred meters in estimates of a user’s location. Moreover, observations are only recorded when an individual uses his or her device, resulting in heterogeneous sampling frequencies between users and at different times for a given user. While sampling rates and data density are increasing rapidly with rising penetration rates and usage, these issues present statistical challenges.

Initial methods by Wang et al. construct *transient origin-destination matrices* by simply counting a trip for pair of consecutive calls made within the same hour from two different towers. However, this method lead to an abundance of short trips and provided a very biased view of movement. Instead, mobile phone trajectories must be de-noised to remove spurious points or calls made in the the middle of routes rather than origins or destinations. To extract meaningful locations, termed as *stays*, algorithms have been developed to smooth out this noise and control for these biases. Jiang et al. provide a thorough review of these techniques in [88] and we adapt the stay point algorithm originally described by Zheng et al. in [188].

Given a user’s trajectory of spatiotemporal points $P = \{p_1(x_i, y_i, t_i) \mid i \in [0, n]\}$, the goal is to discover meaningful locations at which a user repeatedly stays for a significant amount of time. The algorithm begins by considering each call in a time ordered sequence. Two consecutive (p_i, p_{i+1}) points are considered to form the start of a *candidate set* of points at the same semantic location if the distance between them is less than a threshold $\Delta r_{i,i+1} < \delta$. Subsequent points are added to this candidate set if they also meet this criteria, e.g. p_{i+2} is added if $\Delta r_{i+1,i+2} < \delta$. The result is a candidate set $S = \{p_s(x_s, y_s, t_s), \dots, p_t(x_t, y_t, t_t)\}$ containing a number of consecutive calls. A candidate set is considered to represent a single *candidate stay* if time between the first and the last observation in the subsequence S are separated by a time greater than a threshold $\Delta t_{m,n} > \tau$. The geographic location of a candidate stay is set to be at the centroid of points in S . Due to noise in locations and daily call frequencies, multiple candidate stays that are actually the same place may be estimated at a slightly different geographic coordinate on different observation days. To account for this, a final agglomerative clustering algorithm is used to consolidate candidate

stays to a single semantic location regardless of the temporal sequence of individual calls. Though many agglomerative clustering algorithms exist, we implement a simple, efficient grid based approach by assigning each filtered location to a grid cell and then defining a final *stay point* as the centroid of all filtered locations in each cell. A final pass through the original calls assigns any call within a distance δ from a stay point to that stay point regardless of whether or a not a consecutive call was recorded from that location. This algorithm removes noisy or spurious outliers from the data set while preserving as much information on visits as possible. It may also be run on both triangulated and tower-based CDR data, in the latter case it removes noise associated with calls from the same location being routed through different nearby towers due to environmental factors. Pseudo-code can be found below.

With de-noised trajectories of stay points, the next step is to infer contextual information about each location. Alexander et al. and Colak et al. [3, 45] improve on methods by Wang et al. and Iqbal et al. [171, 84] by using visit frequencies and temporal data to infer contextual information such as a location's function or trip purpose. A user's *home* location is defined as the stay point they are observed at most frequently between the hours of 8pm and 7am on weeknights. Their *work* location is defined as the stay point other than home that a users visits the most between the hours of 7am and 8pm on weekdays. Because many individuals do not work, we leave the work location blank if the candidate location is not visited more than once per week or if the location is less than 500m from their home location. All remaining non- home or work stay points are designated as *other*. Figure 3-5 illustrates the aforementioned procedure.

Daily trips are estimated from filtered users by analyzing consecutive observations at different stay points during a given time window. They begin by defining an *effective day* as a period between 3am one morning and 3am on the next consecutive morning. This definition is used to minimize the number of trips that are prematurely ended due to the assumption that users start and end each day at home. A home-based work (HBW) trip is counted if a user is observed to travel between home and work, a non-home based (NHB) trip is counted if a user moves between two non-

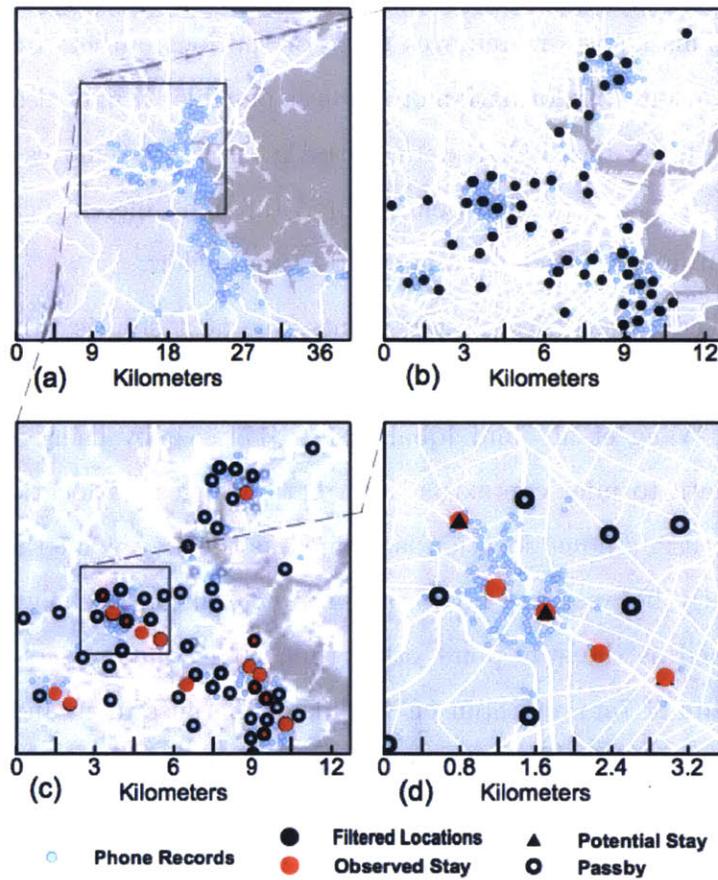


Figure 3-2: An illustration of different stages of the stay point algorithm. Figure taken from [88].

ALGORITHM 2: Stay Point Algorithm - Step 1 - Initialize

```
1: {Each user object has a number of attributes}
2: call = a call object with an associated latitude, longitude, stay index
3: calls = vector of a user's calls ordered by timestamp
4: candidateSet = empty set of consecutive calls that meet criteria for a stay
5: candidateStays = a vector of centroids from candidate sets
6:  $\delta$  = distance threshold between consecutive calls (in meters)
7:  $\tau$  = time threshold between entry into and exit from the stay (in seconds)
8: ds = a grid size for the agglomerative clustering algorithm (in meters)
9: stayCalls = an empty vector of calls from stay points
10: {Notes}
11: *Centroid(callSet) returns an object whose latitude and longitude are the
    centroid of all points in the input
12: *DistanceBetweenCalls(call1, call2) returns the geographic distance between
    calls in meters
13: *TimeBetweenCalls(call1, call2) returns the time between call in seconds
```

ALGORITHM 3: Stay Point Algorithm - Step 2 - Candidate Stays

```
1: {For each user, loop through all calls and find candidate stays}
2: candidateIndex = 0
3: candidateSet = {}
4: for i = 0 to i = calls.size() - 2 do
5:   if DistanceBetweenCalls(calls[i], calls[i + 1]) <  $\delta$  then
6:     candidateSet.append(calls[i + 1])
7:   else
8:     if TimeBetweenCalls(candidateSet[0], candidateSet[end]) >  $\tau$ 
       then
9:       for call in candidateSet do
10:        call.stayIndex = candidateIndex
11:        candidateStay = Centroid(candidateSet)
12:        candidateStays.append(candidateStay)
13:        candidateSet = {calls[i]}
14:        candidateIndex = candidateIndex + 1
```

Figure 3-3: Pseudocodes for initialization and selection of candidate stays for the stay point algorithm.

ALGORITHM 4: Stay Point Algorithm - Step 3 - Agglomerative Clustering

```
1: grid = construct a uniform grid that covers all of a user's calls with cell
   dimensions ds × ds
2: stayIndex = 0
3: for grid cells containing a candidateStay do
4:   candidateStays = {listofcandidateStayinCell}
5:   stay = Centroid(candidateStays)
6:   for call made from a candidateStay in this cell do
7:     call.longitude = stay.longitude
8:     call.latitude = stay.latitude
9:     call.stayIndex = stayIndex
10:    stayCalls.append(call)
11:   stayIndex = stayIndex + 1
```

ALGORITHM 5: Stay Point Algorithm - Step 4 - Final Pass

```
1: {Final pass to add any remaining calls to the stay}
2: for i = 0 to i = calls.size() do
3:   if call not part of a stay and DistanceBetweenCalls(call, stay) >  $\delta$  for
   any stay then
4:     call.longitude = stay.longitude
5:     call.latitude = stay.latitude
6:     call.stayIndex = stayIndex
7:     stayCalls.append(call)
8: Sort stayCalls by timestamp
9: _____
```

Figure 3-4: Pseudocodes for agglomerative clustering and final pass for the stay point algorithm.

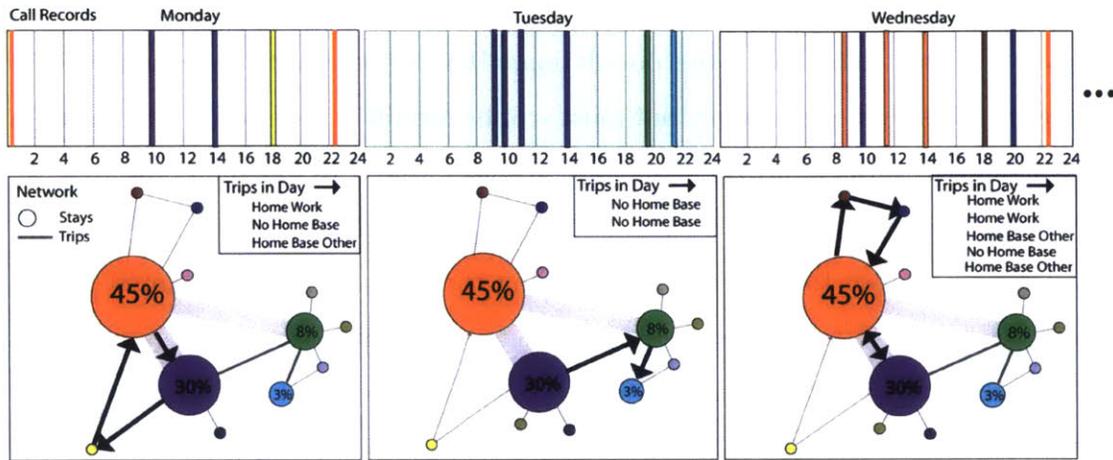


Figure 3-5: Schematic example of phone records converted to daily trips for a typical mobile phone user. Activities are inferred in stay locations and daily trips are measured by time of day between these stays.

home stay points, and a home-based other (HBO) trip is counted if a user is observed moving between their home location and a location labeled as other.

Though a user must have traveled between two different observed stay points at some in time, we do not know the precise departure time. We assign a random departure time based on the conditional probability that user departed during an hour between the time they were last observed at the origin and the time they were first observed at the destination. This conditional probability function for departure time can be derived from surveys such as the National Household Travel Survey or estimated empirically using observed call frequencies of all users over the course of the day. Alexander et al. show that this method produces CDR trip departure time distributions in line with multiple surveys for the Boston region. Having assigned departure times and purposes to each trip, we can construct trips made by a given user. Generally, we are interested in trips between geographic areas such as towns or census tracts so here we convert origin and destination points to IDs of the tract of zone they are in. The result is a vector of trips between locations in the city for each user in our data set.

While a trip represents an observation of movement of at least one person between two locations, we expand these trip counts to represent all individuals in a city.

Expansion is a critical step in models relying on survey data where the sample sizes are typically less than 1% of the population. Here we generally have hundreds of thousands of users in our sample, but must still be careful to control for differences in market share and usage rates across a city. We first scale trips based on how often an individual uses their phone. For each user, we calculate the average number of trips made during a given time window by dividing the number of trips counted by the number of days that user was observed making a call. This step effectively measures the average number of trips a user makes between two locations on a day given that they are observed in our data set.

Due to differences in daily usage of mobile phones among the population, not every user makes enough calls on a typical day to infer their movement patterns. For this reason, we must filter out users that do not make enough calls. This step requires trade-offs between sample size and amount of data we have on each selected user. Because we will eventually be routing these trips through the transportation network, it is important to correctly estimate the total number of trips taken as well as the distribution of trips across the city. In practice, we find that filtering out users who we measure to make fewer than 2.5 trips per day leaves a large sample size of active users and results in valid estimates of trip tables and OD matrices as shown in subsequent sections. Those implementing these methods may find that different filtering criteria produce samples suited for different tasks.

We then expand the average trip counts of filtered users to account for market penetration rates. As with survey participants, the ratio of cell phone users to the population is not uniform within the region. Each user is assigned a home census tract and expansion factors are computed for each tract by measuring the ratio of the number of users assigned there and the reported population. In cities such as Boston, these expansion factors tend to be less than 10, but can be higher in places with lower market share. They are generally much lower than surveys which may only choose two or three individuals to represent hundreds or thousands in an area. Each user's typical daily trip volumes are then multiplied by the expansion factor corresponding to their home tract and the now represent the movements of some fraction of the

tracts population. The spatial and magnitude distributions of the expansion factors obtained for Boston and Rio de Janeiro are shown in Figure 3-6.

Finally, we may wish to consider only trips via a certain mode, e.g. vehicle trips. Though CDR data does not provide resolution required to measure mode choice, vehicle trips can be approximated by weighting person trips by vehicle usage rates in the home census tract of users. In this way, full OD matrices for vehicle or person trips are computed by summing the expanded trip volume computed for all users between all pairs of census tracts. We also construct partial OD matrices containing only trips of a certain purpose during a certain time window. Due to the relative consistency of CDR data around the world, we can adopt this same OD creation procedure in all cities. Pseudo-code to generate OD matrices has been adapted from [3, 45] and can be found at the end of this section. The results from this method are compared to the output of traditional models where applicable. A mapping of raw origin destination flows for Boston and Rio de Janeiro are depicted in 3-9. Trip tables and correlations plots can be found below in section 3.3.1.

Trip Assignment

Having estimated OD flows, our next task is to efficiently assign these trips to transportation infrastructure, in this case a road network. The first step takes tract to tract OD matrices and distributes trips among nodes, or intersections. A trip originating in a census tract is assigned uniformly at random to an intersection in that tract and to an intersection within its destination tract. This distributes flows such as not to create artificial congestion points and reflects general uncertainty in the exact origin of trips. Other approaches, however, may consist of using abstract centroid nodes unique to each tract and connect to a number of other intersections within that tract using what's referred to as centroid connectors. With intersection to intersection flows, the next task is to assign traffic to routes.

Traffic assignment is another mature domain that has been studied extensively by urban and transportation planners. Static non-equilibrium models approaches consist of treating all users as homogenous agents who make route choices prior to

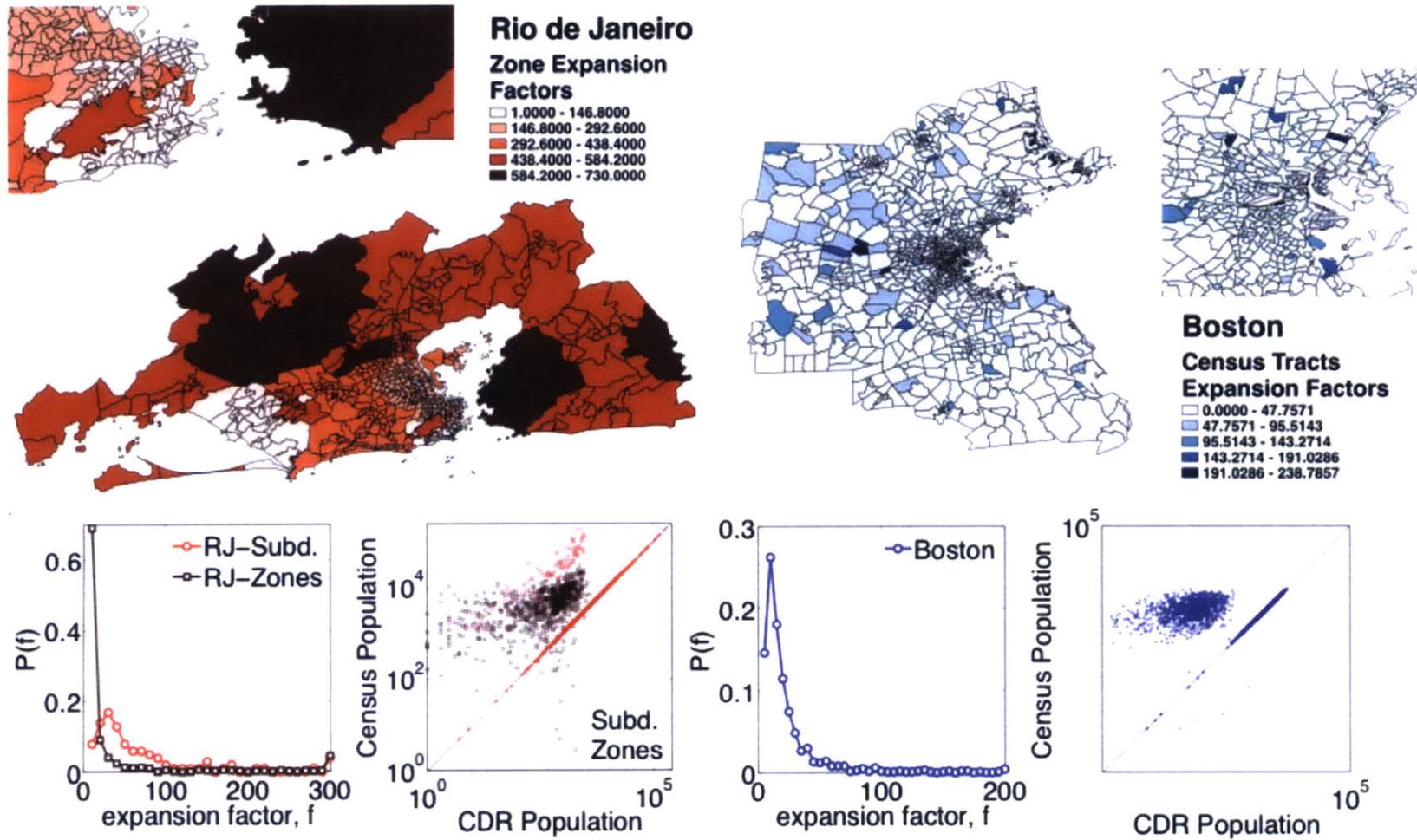


Figure 3-6: Maps depicting the expansion factors in Boston and in Rio de Janeiro in the specified resolutions, the distributions of the values of these factors, and the comparisons between CDR population and the scaled up census population.

ALGORITHM 6: OD Creation Algorithm - Step 1 - Home / Work Expansion

```
1: {Data objects}
2: tracts = census tract data objects containing demographic variables
3:  $OD(o, d, p, t) = 0$  for origin o, destination d, purpose p, and period t
4: _____
5: {Detect home and work for all users and compute expansion factors}
6: for user in users do
7:   user.stays = vector of calls at stay points sorted by time
8:   user.home = index of stay point visited the most between 8pm and 7am
   on weekdays
9:   user.work = index of non-home stay point visited the most between 7am
   and 8pm on weekdays
10:  if user visits work less than once per week then
11:    user.work = null
12:  for stay in user.stays do
13:    stay.label assigned as home, work, or other
14:  user.weekdays = number of weekdays a user records a stay
15:  user.workdays = number of weekdays a user records a stay at work
16:  tract[user.home].numUsers = tract[user.home].numUsers + 1
17: for tract in tracts do
18:  tract.expansionFactor = tract.population/tract.numUsers
```

Figure 3-7: Pseudocode for home and work expansion of the OD creation algorithm.

ALGORITHM 7: OD Creation Algorithm - Step 2 - Trip Counting

```
1: {Count and expand trips}
2: for user in users do
3:   trips = empty vector to store trips taken by a user
4:   for i = 1 to i = user.stays.size() do
5:     s0 = user.stays[i - 1]
6:     s1 = user.stays[i]
7:     if s0 == s1 then
8:       continue
9:     if s0 and s1 are on the same effective day then
10:      trip = new trip from s0 to s1
11:      trip.purpose = PurposeFromLabels(s0, s1)
12:      trip.workday = true if workday for user, false otherwise
13:      trip.departure = GetConditionalDepartureTime(s0, s1)
14:      trips.append(trip)
15:    else s0 and s1 are not on the same effective day
16:      morning = create trip from home to first recorded stay
17:      night = create trip from last recorded stay to home
18:      trips.append(morning)
19:      trips.append(night)
20:   for trip in trips do
21:     o = trip.origin
22:     d = trip.destination
23:     p = trip.purpose
24:     t = trip.departure
25:     if trip.workday == true then
26:       flow = tract[user.home].expansionFactor / user.workdays
27:     else
28:       flow = tract[user.home].expansionFactor / user.weekdays
29:      $OD(o, d, p, t) = OD(o, d, p, t) + flow$ 
30:
31: {Notes}
32: *PurposeFromLabels(s0, s1) returns a trip purpose (HBW, NHB, HBO)
   based on the label of origin and destination stays
33: *GetConditionalDepartureTime(s0, s1) returns a departure time based on
   the observation times at origin and destination
34: *an effective day is defined as a period between 3am today until 3am on the
   next consecutive morning
```

Figure 3-8: Pseudocode for trip counting in the OD creation algorithm.

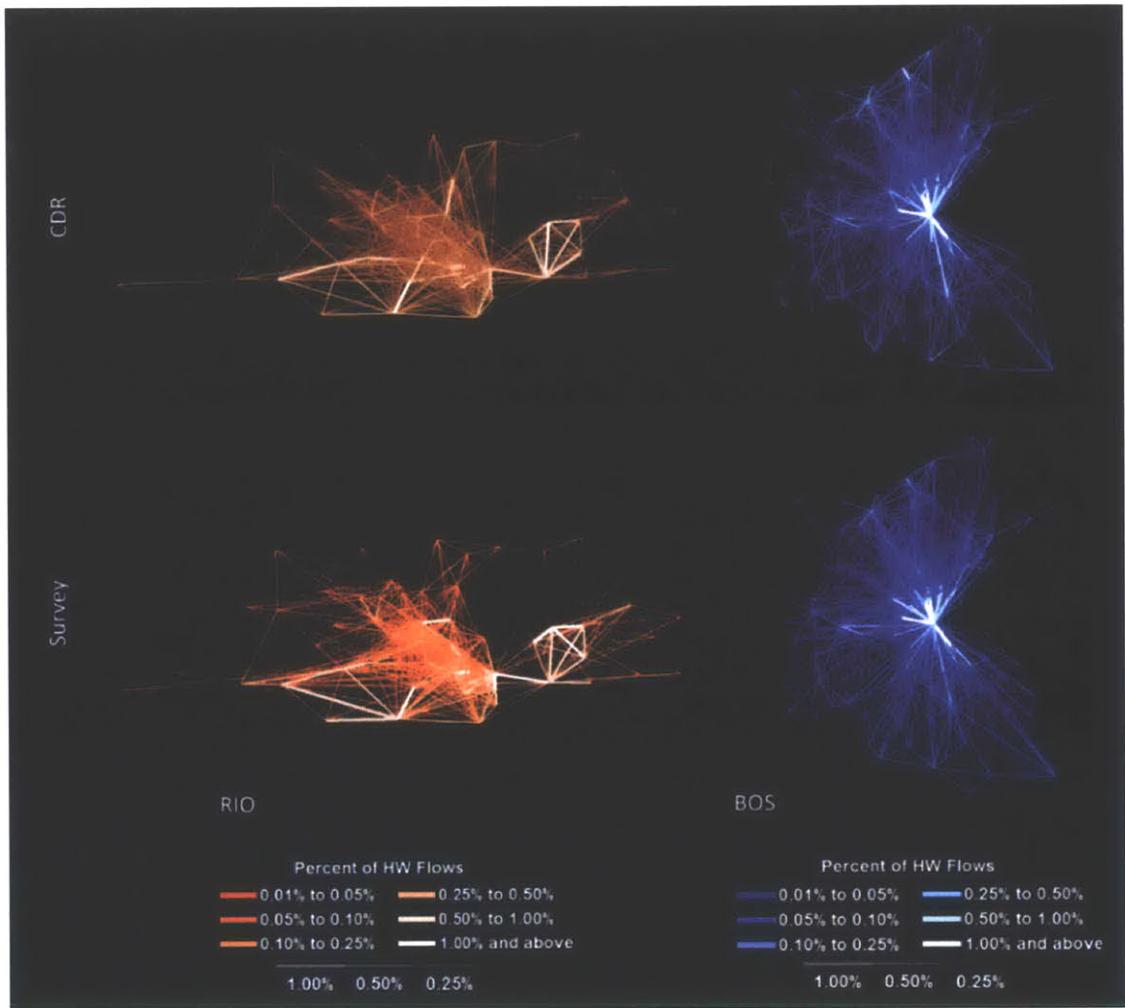


Figure 3-9: A spatial illustration of flows comparing inter and intra town/subdistrict OD pairs.

departure based on some heuristic related to current traffic conditions (e.g. the path that minimizes travel time). Incremental Traffic Assignment (ITA) is a variant of these static non-equilibrium assignment models that assigns batches of trips serially and updates costs between increments, as an improvement over the simplest all-or-nothing assignment methods. However, it is known that dynamic equilibrium models are more realistic in assigning trips as outcomes are closer to the Wardrop principles [172], or Nash Equilibria, where drivers seek paths that minimize their travel time and in the final traffic conditions, no driver has an incentive to change their route. To take a step further from static models, Dynamic Traffic Assignment (DTA) [111] models take an iterative and temporally more coherent approach. The addition of these complexities help model traffic flow at finer granularity, enabling road segments to have different conditions within themselves and consequently the representation of phenomena like congestion spill-back, FIFO principle, and others [48].

Our system is modular so that it may implement any number of traffic assignment algorithms. Here, however, we take a simple ITA approach, as it is computationally efficient for many trip pairs in detailed road networks and allows us to keep track of each vehicle as it is routed through the network. We develop a set of tools to perform large scale routing and traffic assignment using parallelization for speedups. First, the parsed and optimized road network is loaded into a graph object. In our implementation, we use the Boost Graph Library for its flexibility and efficiency. We can then compute shortest paths based on a user defined cost (in this case travel time on road segments). We choose the A* algorithm among the wide range of shortest path algorithms, as it's widely used in routing on geographic networks for its flexibility and efficiency. The A* algorithm implements a *best-first-search* using a specified heuristic function to explore more promising paths first. The euclidian distance between nodes provides an intuitive heuristic that ensures optimal solutions are found. While this algorithm provides the same results as Dijkstra's algorithm, we find that it becomes more efficient to compute paths one by one for sparse OD matrices.

On most city roads, free-flow speeds are rarely achieved due to congestion. As a

Full OD		Increment 1 width=0.7		Increment 2 width=0.3	
(o,d)	flow	(o,d)	flow	(o,d)	flow
(1,2)	1000	Batch 1 (1,2)	700	Batch 1 (1,2)	300
(1,3)	100	Batch 1 (1,3)	70	Batch 1 (1,3)	30
(2,3)	250	Batch 1 (2,3)	175	Batch 1 (2,3)	75
(3,2)	100	Batch 2 (3,2)	70	Batch 2 (3,2)	30
(4,3)	1000	Batch 2 (4,3)	700	Batch 2 (4,3)	300
(5,4)	500	Batch 2 (5,4)	350	Batch 2 (5,4)	150

Figure 3-10: Our efficient implementation of the incremental traffic assignment (ITA) model. A sample OD matrix is divided into two increments and then split into two independent batches each.

result, traffic patterns may significantly change the time costs associated with using a particular route. To address this, we implement an Incremental Traffic Assignment (ITA) algorithm [126]. A simplified schematic explaining the procedure can be seen in Figure 3-10. This algorithm assigns trips in a series of increments and updates the costs of edges in the network based on the number of vehicles that were previously assigned to that road between increments. For example, the first increment assigns 40% of trips for each pair assuming each driver experiences free-flow speeds. The travel time cost associated with every road segment is then adjusted based on how many drivers were assigned to that road and the total number of cars a road can accommodate in unit time. The next 30% of drivers are then routed in the updated conditions. This process is repeated until all users have been assigned a route. The shortcoming of this method is that once a driver has been assigned a route it does not change, and consequently the approach does not converge to Wardrop’s equilibrium even for very small increment sizes. Yet we use it here due ease of implementation and the fact that it is still insightful for the purposes of demonstrating the implementation of a modular data-driven travel demand model. Future work will explore the use of newer methods.

Relating travel performance to traffic conditions has been a long standing problem in transportation. Many different characterizations exist, ranging from conical volume-delay functions to more complex approaches [31, 156, 2]. One of the most simplistic and common metrics used in determining the travel time associated with a

specific flow level is the ratio between the number of cars actually using a road (volume) and its maximum flow capacity (volume-over-capacity or V/C). At low V/C , drivers enjoy large spaces between cars and can safely travel at free-flow speeds. As roads become congested and V/C increases, drivers are forced to slow down to insure they have adequate time to react. Based on the volume-over-capacity (V/C) for each road, costs are updated according to Eq. 3.1, where $\alpha = 0.15$, $\beta = 4$ are used per guidelines set by the Bureau of Public Roads¹.

$$t_{current} = t_{freeflow} \cdot (1 + \alpha(V/C)^\beta) \quad (3.1)$$

Though increments must be routed in serial, all routes discovered within an increment are independent. To speed up the routing process, we divide all trips in an increment into batches and send these batches to different threads for parallel computation. Because the road network remains fixed in each increment, we only need to store a single graph object shared by all threads. When a shortest path is found, we walk that path and increment counts of the number of vehicles that were assigned to each road and sum the counts from all batches after the increment has finished. We also keep track of the origin and destination census tracts of the assigned vehicles in a bipartite graph for later analysis. After all trips have been routed, we compute final V/C ratios and other metrics of each segment and update these values in the database so they can be used for other applications or visualization. Pseudo code for this ITA procedure can be found in Algorithm 3-11.

3.3 Results

In the following sections we demonstrate the range of outputs provided by our system. We first report trip tables and compare origin-destination matrices produced by our system to available estimates made using travel surveys. We then report road network performance as well as characteristics of road usage patterns enabled by the

¹Travel Demand Modeling with TransCAD 5.0, User's Guide (Caliper., 2008).

ALGORITHM 8: Incremental Traffic Assignment

```
graph = road network
OD(p, t) = origin-destination matrix for purpose p and time window t
B = a bipartite network containing roads and census tracts
incrSize = vector of increment sizes, e.g. [0.4, 0.3, 0.2, 0.1]
nBatches = number of threads to use
for i = 0 to i < incrSize.size() do
  for b = 0 to b < nBatches do
    create new thread
    batch = GetBatch(OD, b)
    for all o, d pairs in batch do
      flow = OD[o, d].flow · incrSize[i]
      route = A*(o, d, graph)
      for all segment s in route do
        s.flow = s.flow + flow
         $B_{e \rightarrow o} = B_{s \rightarrow o} + flow$ 
    wait for all threads to finish
  for segment s in graph do
     $s.cost \leftarrow s.freeFlowTime \cdot (1 + \alpha(\frac{s.flow}{s.capacity})^\beta)$ 
```

* *GetBatch*(*OD*, *B*) returns only the subset of *OD* pairs pertaining to a batch

* *A**(*o*, *d*, *graph*) returns the shortest path between *o* and *d* if a path exists

Figure 3-11: Pseudocode for the ITA algorithm.

construction of a bipartite road usage network.

3.3.1 Trip Tables and Survey Comparison

In order to understand when and where these new data will be effective and how the results differ from traditional approaches, we compare the output of our system to previous travel surveys wherever possible. In four of the cities studied, we find estimates of travel demand from surveys: the 2011 Massachusetts Household Travel Survey (MHTS) in Boston, the 2000 Bay Area Travel Survey (BATS) in San Francisco, a 2013 transportation plan in Rio de Janeiro, and estimates from a 2012 LUT model in Lisbon[61]. While these surveys do not always produce all estimates we are able to generate with our system, we make comparisons wherever possible.

Trip tables report the total number of trips of a given purpose or during a given time of day for a city and represent the total load placed on transportation infrastructure. In Table 3.2, we report trip tables for each city in this study. We find close agreement with trip tables estimated using CDR data and surveys in Boston and the San Francisco Bay Area and less agreement in Rio de Janeiro. We note, however, that the 3.74 million person trips estimated for Rio is far too low given the population of the region and highlights the difficulty in finding reliable planning resources in many areas. Finally, we note that in Lisbon, the survey results represent vehicle trips only, while we report person trips. When adjusting for mode car ownership rates in Portugal, our numbers align more closely. We were unable to find a survey or model for comparison in Porto.

In addition to trip tables, it is also necessary to compare the distribution of trips from place to place around the city. In order to make this comparison, the area unit of analysis for the survey and our model must be aligned. Given the resolution of mobile phone data, our system is designed to create ODs at the census tract (or equivalent) level while many surveys aggregate to larger traffic analysis zones or super districts. For comparison, we aggregate the OD matrices from CDRs to the coarser grained resolution provided by the survey and compare results. Figure 3-12 show correlation histograms comparing OD matrices at the largest spatial aggregation available pro-

Table 3.2: Trip tables estimates. Where possible, our results are compared to estimates made using travel surveys. For each city, we report the number of person trips in millions for a given purpose or time. Trip purposes include: home-based work (HBW), home-based other (HBO), and non-home-based (NHB). Trip periods include: 7am-10am (AM), 10am-4pm(MD), 4pm-7pm (PM), and the rest of the day (RD). We note that the exact boundaries of the surveys do not exactly coincide with those used in our estimation so direct comparisons are not exact. In general, trip magnitudes align closely, with the exception of Rio de Janeiro, where the survey results report far too few trips, illustrating the difficulty of obtaining sensible measurements via certain techniques. No comparisons could be found for Porto. *Note that the Lisbon Survey only contains estimates of vehicle trips in millions.

City	HBW	HBO	NHB	AM	MD	PM	RD	Total
Boston	5.76	8.99	6.72	3.71	7.68	5.75	4.33	21.47
MHTS	3.22	12.83	9.49	5.32	8.87	8.20	3.15	25.54
SF Bay	4.07	10.05	7.04	4.47	7.81	5.35	3.53	21.16
BATS	4.60	11.54	4.66	4.18	6.90	4.22	3.00	20.80
Rio	9.92	17.17	11.46	7.71	14.09	10.47	6.29	38.55
Survey	2.06	-	-	1.31	1.19	1.24	-	3.74
Lisbon	1.08	2.01	1.21	0.79	1.67	1.26	0.58	4.30
Survey*	0.61	-	-	-	-	-	-	-
Porto	0.49	0.87	0.46	0.32	0.70	0.54	0.27	1.83
Survey	-	-	-	-	-	-	-	-

duced by our methods and those produced by traditional methods. In general we find very high correlations in Boston, San Francisco, and Rio, with lower correlations in Lisbon. Lisbon, however, has the smallest units of aggregation and these results demonstrate the limitations of these comparisons at very high spatial resolutions. We hope future work explores how these correlations relate to the modifiable area unit problem. Finally, there is significant uncertainty in all models and we hope future works will explore this uncertainty further.

3.3.2 Road Network Analysis

The first output of this procedure is volume, congestion (volume-over-capacity), and travel times for all road segments. Using the outcomes of our analyses, we calculated the distributions of volumes on roads, along with V/C s in Figure 3-13. Interestingly, the results suggest qualitatively similarly distributed volumes and V/C s for our five subject cities. Moreover, our findings are consistent with general congestion studies that identify Rio de Janeiro as one of the most congested cities in the world and the San Francisco Bay Area not far behind. Smaller cities such as Boston and Porto have fewer problems with congestion.

3.3.3 Bipartite Road Usage Graph

In addition to measuring physical network properties of roads, the system architecture enables detailed analysis of individual road segments and neighborhoods within a city. Though the transient OD matrices constructed by Wang et al. [171] correlate poorly with OD matrices developed by the methods above and traditional surveys, their work highlights new metrics of road usage patterns that can be measured via these new data sources. To this end we create a bi-partite usage graph. Every time a route between two location is assigned, we traverse the path and keep a record of how many trips from each driver source (census tract) used each road. This record is then used to construct a bipartite graph containing two types of nodes: road segments and driver sources, as shown in Figure 3-14. Roads are connected to driver sources that

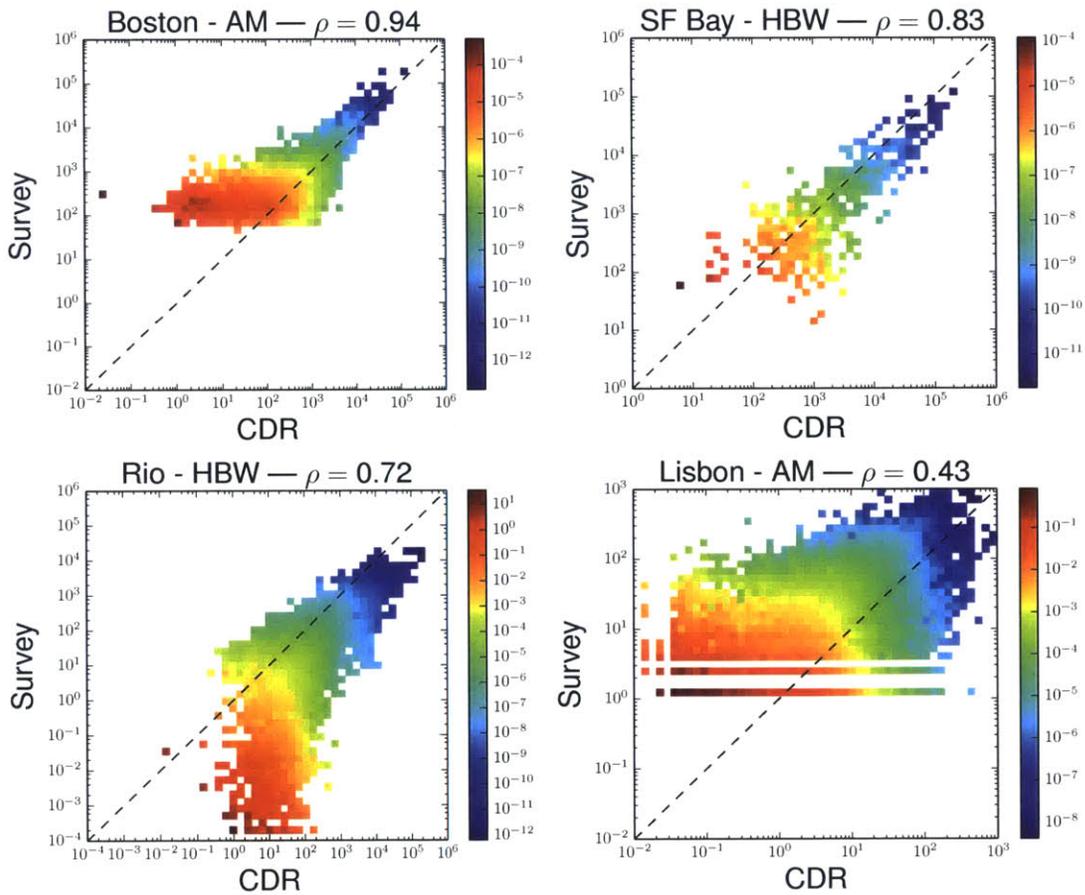


Figure 3-12: Correlations between OD matrices produced by our system and those derived from travel surveys at the largest spatial aggregation of the two models. In Boston, this refers to towns, in San Francisco, MTC superdistricts, in Rio, census superdistricts, and in Lisbon, freguesias. The larger of these area units (e.g. towns in Boston), the better our correlations, while correlations at the smallest aggregates (e.g. freguesias in Portugal), correlations are lower. However, more work must be done to understand uncertainties in estimates provided by both models.

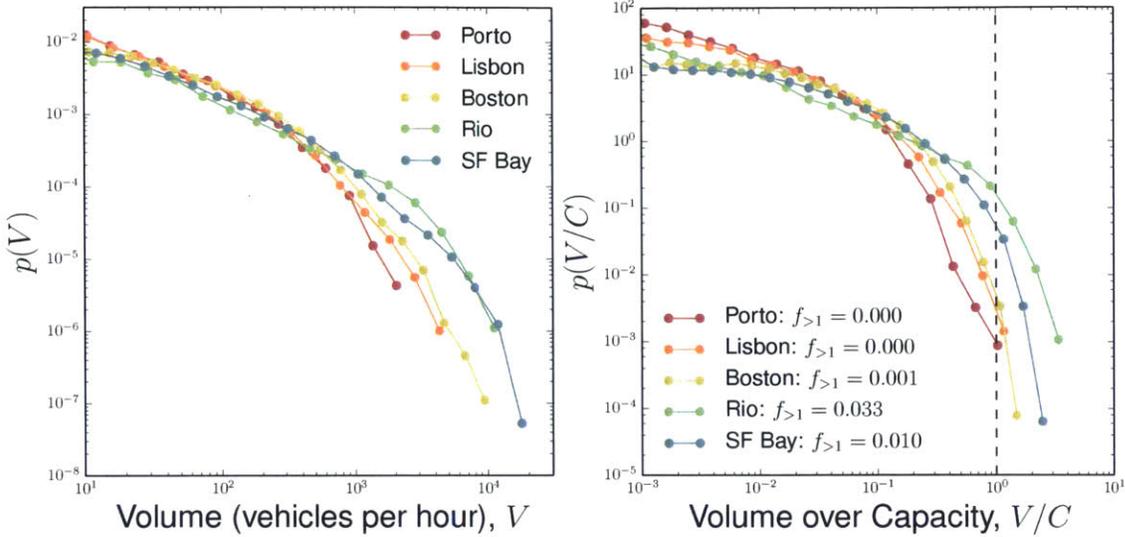


Figure 3-13: Distributions of travel volume assigned to a road and the volume-over-capacity (V/C) ratio for the five cities. The values presented in the legend refers to the fraction of road segments with $V/C > 1$.

contribute traffic to that segment and census tracts are connected to roads that are used by people who live here.

$$k_s^{road} = \sum_o A_{o \rightarrow s}, \quad k_o^{source} = \sum_s A_{o \rightarrow s} \quad (3.2)$$

$$A_{o \rightarrow s} = \begin{cases} 1, & \text{if vehicles from tract } o \text{ use road } s \\ 0, & \text{otherwise.} \end{cases}$$

We then examine the degree distributions of roads and census tracts using Eq.3.2 in this bipartite graph to reveal patterns of road usage in Figure 3-15. The number of roads used by residents of a given location is much more consistent between different cities and appears less affected by the size of the road network. On the other hand, the number of driver sources contributing traffic to a given road segment is broadly distributed, suggesting that most roads are *local* in that they serve only a few locations, while a few roads in the tail of the distribution are used for large fractions of the population. While this result is intuitive given that highways are designed for just this purpose, we hope future work explores the relationship between this bipartite usage graph and road network topology further.

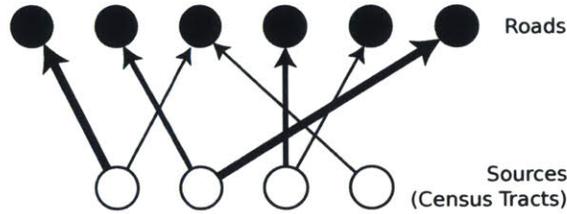


Figure 3-14: A graphical representation of the bipartite network of roads and sources (census tracts), with edge sizes mapping the number of users using the connected road in their individual routes.

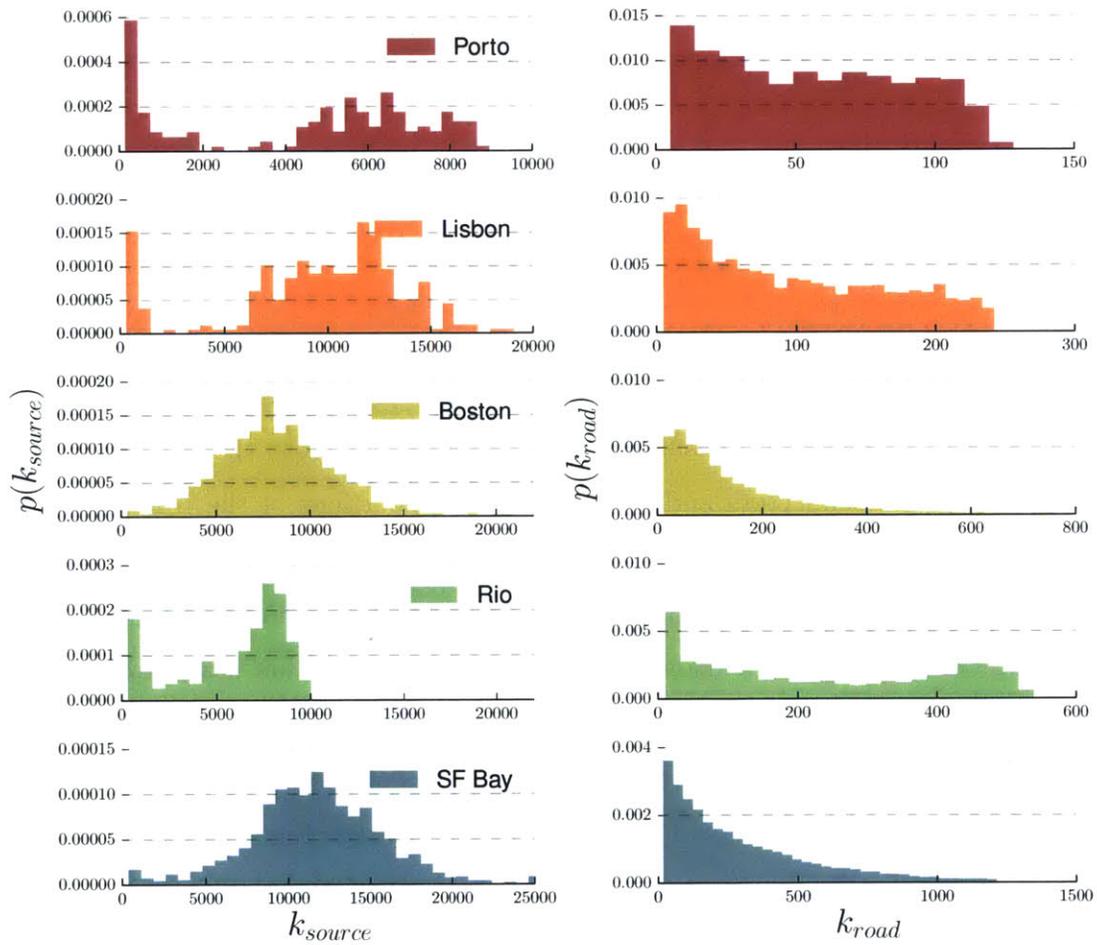


Figure 3-15: Distributions of k_{road} and k_{source} for the five cities.

An example of such an application was proposed by Wang et al. to classify road segments based on the relationship between topological and demand based metrics. Comparing the topological properties of roads in the physical network to the bipartite usage graph provides insights into their role in the transportation system. Edge betweenness centrality [119] captures the importance of a road by counting how many shortest paths between any two locations σ_{OD} must pass through that edge $\sigma_{OD}(e)$ (Eq. 3.3). While this measure captures some aspects of importance, it treats all potential paths as equally likely and tends to be biased towards geographically central links. The degree of a road in the bipartite usage graph reflects the number of locations in the city that actually rely on that road because trips were assigned there from actual travel demand. With these two metrics, betweenness centrality and a roads degree in the usage network, we can classify the role of a road in the cities transportation network.

$$bc_s = \sum_{o,d} \frac{\sigma_{OD}(s)}{\sigma_{OD}} \quad (3.3)$$

A simple classifier divides the betweenness usage degree space into four quadrants surrounding the point representing the 75th percentile for betweenness centrality and usage degree. Roads with betweenness and usage degree above the 75th percentile are both physical connectors and are used by large portions of the region. These roads tend to be bridges or urban rings. Roads with low betweenness, but high usage degree are attractors, receiving a higher proportion of trips than would be expected assuming uniform demand. Roads with high betweenness and low usage are physical connectors and serve an important purpose geographically, but may not be utilized by actual demand. Other roads, with low betweenness and low usage are local roads and primarily serve populations living and working nearby. Figure 3-16 shows each road according to this classification using data from the ODs calculated via mobile phones.

Finally, this bipartite framework of analysis allows us to augment visualizations of congestion maps in two ways. The first focuses on a single road segment. For

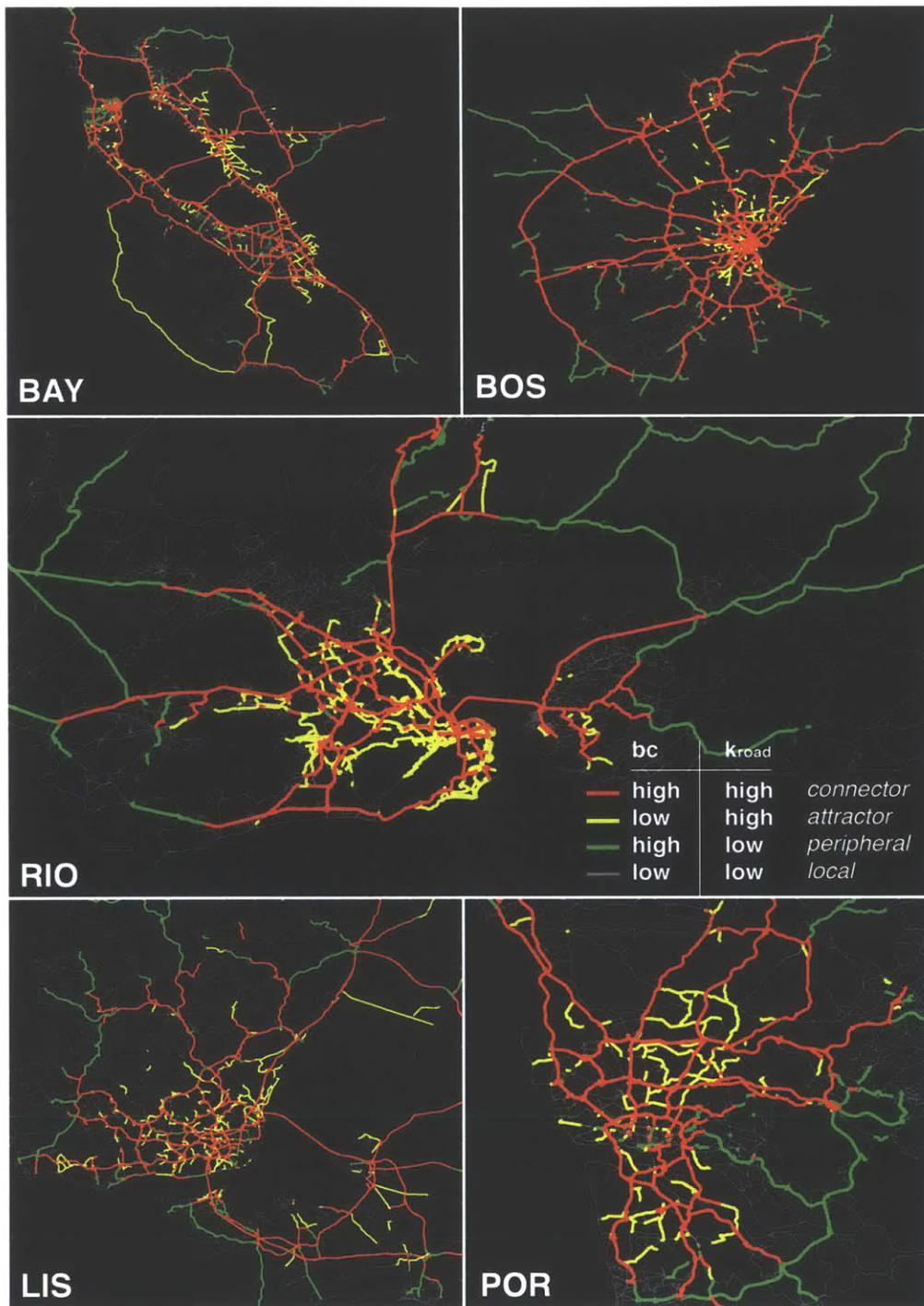


Figure 3-16: Maps depicting the proposed road classification, summarized in the legend, for the five subject cities.

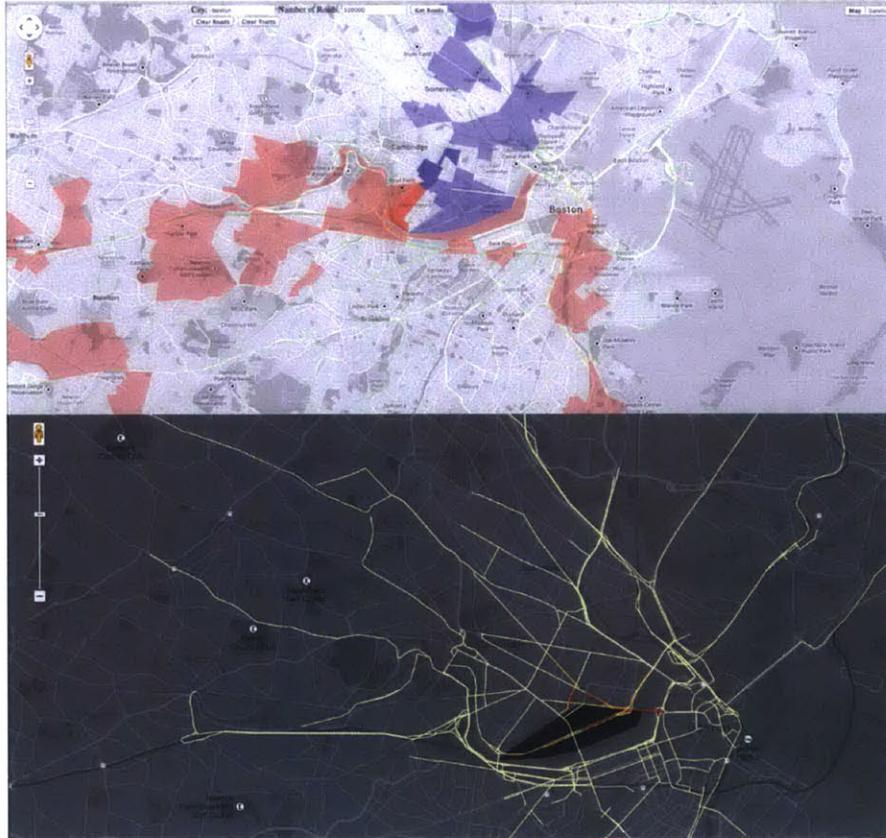


Figure 3-17: Two screen images from the visualization platform. (a) The trip producing (red) and trip attracting (blue) census tracts using Cambridge St., crossing the Charles River in Boston. (b) Roads used by trips generated at the census tract including MIT.

example, when we identify a segment of a highway that becomes highly congested with traffic jams each day, we can easily query the bipartite graph to obtain a list of census tracts where drivers sitting in that traffic jam are coming from and where they are going to. The census tract nodes can also be given attributes from containing any demographic data a user wishes. With this information, it is possible to identify leverage points where policy makers can offer alternatives to these individuals or even power applications such as car sharing, by notifying drivers that others sharing the same road may be going to and from the same places. Moreover, businesses considering products or services based on who may be driving by or near different locations may find value in these detailed breakdowns.

Rather than selecting a road segment node, we may also select a single census tract, and check its neighbors to construct a list of all roads used by individuals moving to or from that location. For example, for a given neighborhood in a city we can identify all major arteries that serve that local population. This information provides a detailed look at a central location based on how much road usage it induces. Moreover, geographic accessibility, critical to many socio-economic outcomes, can now be measured in locations that were previously understudied.

3.3.4 Visualization

To help make these results accessible to consumers and policymakers, we build an interactive web visualization to explore road usage patterns in each city. Most GIS platforms can connect directly PostGIS databases to visualize and analyze road networks with our estimated usage characteristics. While these platforms are preferred by advanced users familiar with GIS data, they are opaque to many consumers who may benefit from more detailed information on road usage. A simple API is implemented to query the database and generate standard GeoJSON objects containing geographic information on roads as well as computed metrics such as level of service. We also implement queries to answer questions such as "What are all the census tracts used by drivers on a particular road?" or "What are all roads used by a given location in the city?". These data are then parsed and displayed on interactive maps using any

of the available online mapping APIs and D3.js allowing users, with functionality that enables one to select individual roads and areas. Two screen images of this system is shown in Figure 3-17.

3.4 Discussion

This section has presented a full implantation of a travel demand model that uses new, big data resources as input. We have presented a system that combines and improved upon many disparate advanced in recent years to produce fast, accurate, and inexpensive travel demand estimates. We began by outlining methods to extract meaningful locations from noisy call detail records and estimate origin-destination matrices by counting trips between these places. Normalized and scaled trips counts are compared to estimates made using survey data in both trip tables and at the OD pair level. These flows are then assigned to road networks constructed from OpenStreetMap data using an incremental traffic assignment algorithm. As routes are assigned, a number of metrics on road usage are measured and stored.

While these results show great progress in making big data useful for transportation engineering, there are still limitations inherent in this data and our model. Specifically, we highlight three areas that are ripe for further study.

1. We have shown the the level of aggregation applied to OD matrices can affect the correlation observed between model outputs. This is a standard manifestation of the modifiable area unit problem and a more detailed exploration may indicate which levels of analyses were better suited for different data sources. Moreover, a more detailed analysis of uncertainty in model estimates may make it easier to assess their correlation and validity.
2. Our traffic assignment algorithm is efficient, but simple. In the future, a stochastic dynamic user equilibrium assignment methods should be explored and compared. Moreover, route choice modeling may be significantly improved by the availability of high resolution GPS trajectories of drivers. We believe our sys-

tem's modular design makes it easy to incorporate these new models.

3. Our mode choice model remains simple and will likely require more sophistication for modeling trips not taken in private vehicles. This, combined with improvements in route choice, may make it possible to estimate multi-modal trip demand, as public transportation, bike lanes, and even water transportation networks are included in OpenStreetMap data.

Transportation engineers and urban planners have a rich history estimating flows of people within cities and mapping this flow onto transportation infrastructure. However, these efforts are often constrained by limited data resources. The rise of ubiquitous mobile sensors has generated a wealth of new data on human mobility, but new tools must be developed to integrate these data and insights into traditional transportation modeling approaches. To this end, we have demonstrated a full implementation of a travel demand model utilizing mobile phone data as an input. We presented algorithms to generate routable road networks from open source data repositories, generate validated OD matrices and trip tables from CDR data, and route these trips through road networks using a paralleled ITA algorithm. We have demonstrated a number of possible analyses that can be performed on the output of this system including network performance and classification measurements and an online, interactive visualization platform.

As more data becomes available in the form of calls, gps traces, or real time traffic monitoring systems, we are excited at the prospect of updating and improving these systems further.

Chapter 4

Understanding congested travel in urban areas

4.1 Introduction

Cities have a long-standing history cultivating technological innovations which allow citizens to efficiently access goods and opportunities. However the ease of access has been increasingly difficult to maintain under rapid urbanization [68, 17, 26, 5, 25, 79, 85]. As growing population densities create excessive demand for cities' infrastructure, the increasing penetration and advancement of technology generates massive amounts of multidimensional data that can be utilized to study and mitigate this demand. Specifically, the availability of mobile phone data has led researchers to quantify fundamental spatiotemporal patterns to better understand human mobility in urban areas [70, 32, 153, 154, 56]. With the continuous increase in the volume and accuracy of new data sources, new methods that process and distill mobile phone data are consistently refined, and traditional models of mobility like the gravity, radiation, or activity based models are being updated in tandem [159, 136, 149, 179, 144, 171]. In the context of travel demand estimation, previous efforts focused on developing models that combine household travel surveys with census and land use information

This chapter is based on [47].

[126, 9]. Despite the robust methodology and meticulous implementation of these models, the high costs associated with obtaining the infrequent and small data has proven to be the bottleneck. To supplement these approaches, traffic simulations and demand estimation models have begun incorporating big data sources into their forecasts, building portable data pipelines to create data-driven decision making tools for policy makers [164, 3, 45].

Understanding of the complex interplay of road infrastructure and travel patterns to model travel times and congestion in not a single city but many at once has been a particular challenge in this line of research [100, 123, 99]. Road networks, the circulatory system sustaining a city's accessibility and cultivating its economic prosperity [94, 139, 14] are seized with congestion in most large metropolitan areas. In their 2013 report, TomTom, a leading GPS company, states that in cities like Moscow, Istanbul, Rio de Janeiro, Mexico City and Beijing, people on average spend more than 75% extra time traveling due to traffic. The resulting loss of time, money, and energy are borne by the city's citizens and travelers. Municipalities continually invest in road infrastructure construction and maintenance to increase supply, although controversies on whether more roads alleviate congestion persist [30]. Other efforts to reduce congestion aim to decrease driving demand by promoting alternative travel modes, high occupancy driving lanes, carpooling, congestion pricing and in extreme cases, road space rationing. Even with all these measures congestion remains inherent and drivers are increasingly leveraging real time information through GPS devices and online routing tools to move faster. With everyone having easy access to traffic information, drivers make decisions without coordination based on near-perfect information, resulting in suboptimal system configuration. This general trend of using raw real time information in decision making has significant implications, as it might be also used as a tool to guide drivers to make choices for the benefit of the city, thus creating a more optimal traffic configuration. The extent of the global inefficiency has been of great interest [168, 141, 140, 142] in many contexts ranging from wireless networks to transportation [170, 172, 29, 183, 148, 49]. Theoretical approaches to bring the system to optimality generally converge to marginal cost taxation, which

essentially forms the basis of congestion pricing schemes today [130, 151]. Despite the abundance of research on optimal flow configurations and their implications in the transportation, urban planning and economics literature, there is a shortage of works that utilize big data sources to understand the role of travel demand and actual travel times in metropolitan regions when comparing cities. This highlights a need to build a framework that can be replicated to systematically generate meaningful travel times to not only understand cities better but also test solutions to urban problems such as congestion or pollution.

In this work, we address this issue by coupling travel demand profiles and travel time estimates to analyze how efficiently people move across cities. We begin by modeling the supply by parsing publicly available OpenStreetMap data to obtain road networks. To model travel demand, we mine massive mobile phone datasets, also referred to as call detail records (CDRs) [28]. This procedure requires home and work location detection for millions of users, mining of their location shifts, and the proper sampling procedures to represent accurately the trip tables for the whole city.

4.2 Methods

4.2.1 Mobile phone data

Mobile phone datasets, also referred to as CDRs (Call Detail Records), used in this study consist of at least three weeks of records of all mobile phone users of a particular carrier across each subject city. Each individual call detail record consists of a hashed user identification string, a timestamp, and the location of the activity. The spatial granularity of the data varies between cell tower level, where calls are mapped to tower locations and distributed uniformly within the Voronoi cell that it forms, and triangulated geographical coordinate pairs, where each call has a unique pair of coordinates accurate to within a few hundred meters. Market shares associated with the carriers that provide the data also vary.

The nature of the activity varies: for all cities calls made by the user is included.

ID	Time	Lat	Lon	Location
12345678	24/12/2012 09:00:00	42.397	-71.121	A
12345678	24/12/2012 10:00:00	42.360	-71.057	B
12345678	24/12/2012 10:05:00	42.360	-71.057	B
12345678	24/12/2012 12:00:00	42.360	-71.094	C
12345678	24/12/2012 16:00:00	42.360	-71.094	C
12345678	24/12/2012 20:00:00	42.397	-71.121	A
...



Figure 4-1: A typical depiction of rows of CDR data in Boston. User 12345678 makes a call from location A (Davis square), then goes on to make two calls from location B (Boston City Hall), then makes one call location C (MIT) at noon and another later, and makes one final call again from the location A at 8pm.

Received calls, SMS activity, and various location signals may also be included. A minimum of three weeks of phone call records are available, although for some cities the period of the data is significantly longer. The granularity of the spatial component of the data in Rio de Janeiro is at the cell tower level: where calls are mapped to the Voronoi cells formed to model the coverage area of each tower. For other cities, the spatial information comes in triangulated latitude-longitude pairs, where each call has a unique pair of coordinates with an accuracy of roughly few hundred meters. Market shares associated with the carriers that provide the data also vary. Table 4.1 compiles descriptive statistics for these data sources for each city we explore in this section.

Each individual call detail record consists of a hash string identifying the mobile phone user, a timestamp marking the time of the activity, and the described spatial information regarding the activity. Figure 4-1 depicts an example daily log of a user living in Boston, where the location field is inferred as unique locations visited.

CDR data inherently contains noise, as expected in any similar dataset. One reason for noise is the set of algorithms mobile phone carriers use for tower-to-tower call balancing to improve service. This operation creates discontinuities in the data that do not represent actual movement. To remove this noise and correct for other similar discrepancies, we apply a procedure generally used for GPS traces, referred to as a stay-point algorithm. Jiang et al. provide a thorough review of these techniques in [88] and we adapt the stay point algorithm originally described by Zheng

et al. in [188]. In summary, stay-point algorithm simplifies a sequence of calls within a specified spatiotemporal area. In other words, calls within a certain radius and timeframe are bundled together. The *pass-by points* are removed, and *stays* remain. This mapping is made such that the representative point is the medoid of all such calls. For all cities here, except Boston where the data is triangulated, this algorithm is applied in a modified way. A tower-based CDR dataset only roughly describes the region from which the call was made, that is, the estimate of a user’s position is only known up to the Voronoi cell for that tower. For this reason, the simplification of the series of calls is applied by serializing the calls made from towers within a certain distance. For the temporal dimension, these calls are labeled as stays only if the user is known to be in that location for at least 10 minutes.

One key point worth noting is that CDRs are of passive nature: except for a very tiny portion of the data, a mobile phone user’s location information is only visible in the data when he/she interacts with his/her phone. Therefore it is certainly possible for a user to be in the location the data point classified as a pass-by, or alternatively be visiting other locations that cannot be distinguished due to lack of phone interaction. This issue and other similar shortcomings resulting from the nature of the data are discussed in detail in previous work [164, 3, 45].

4.2.2 Census and travel survey data

At the census tract (or equivalent) scale, we obtain the population, vehicle usage rate, and median income of residents in that area. For US cities, the American Community Survey provides this data on the level of census tracts (each containing roughly 5000 people). Census data is obtained for Brazil through IBGE (Instituto Brasileiro de Geografia e Estatística) and for Portugal through the Instituto de Nacional de Estatística. All cities analyzed in this work have varying spatial resolutions of the census information. Wherever possible, we obtain the most recent travel demand model or survey from the subject city and compare the results to those output by our methods. We use the 2011 Massachusetts Household Travel Survey (MHTS) for Boston, 2000 Bay Area Transportation Survey (BATS) for the Bay Area, a recent

transportation model output provided by the local government for Rio de Janeiro. For Lisbon, the most recent estimates from the MIT-Portugal UrbanSim LUT model that uses the 1994 Lisbon transportation survey as input are used. We found no recent travel survey or model for Porto.

At the census tract (or equivalent) scale, we obtain the population and the vehicle usage rate of residents in that area. For US cities, the American Community Survey provides this data on the level of census tracts (each containing roughly 5000 people). Census data is obtained for Brazil through IBGE (Instituto Brasileiro de Geografia e Estatística) and for Portugal through the Instituto de Nacional de Estatística. All cities analyzed in this work have varying spatial resolutions of the census information.

Figure 4-2 exhibits properties of the administrative boundaries used. Boston and Bay Area, regions in the United States, exhibit uniformity in their distributions of population per zone, as the populations are generally around 5000. Lisbon and Porto demonstrate higher deviations for a similar median, whereas the magnitude of the spread in Rio de Janeiro is higher than the other cities. To get an estimate of the vehicle usage rates, we use the following relationship:

$$VUR(i) = P_{\text{drive alone}}(i) + P_{\text{carpool}}(i)/S,$$

where $P_{\text{drive alone}}(i)$ and P_{carpool} are probabilities that residents in zone i drive alone or share a car, respectively. $S = 2.5$ is estimated to be the average carpool size [171].

Conversely, Boston and Bay have the highest vehicle usage rates whereas in Rio de Janeiro people are less car-oriented. To assess how similar our five cities are in terms of CDR data sampling we compare their expansion factors, defined as the ratio of the number of people living in a tract to the number of people assigned that tract as a home location. All cities have a mean below 100, although outliers exist.

4.2.3 Extraction of validated origin-destination information

Traditional modeling approaches to origin-destination (OD) information utilize data obtained from travel surveys, possibly combined with land use and point of interest

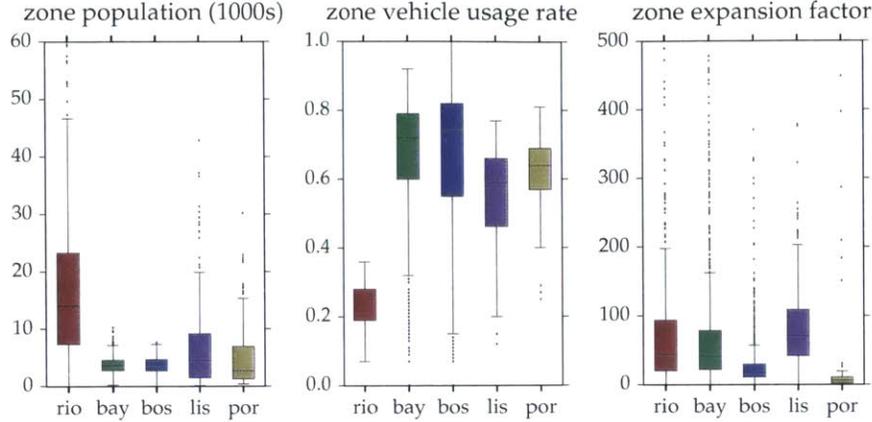


Figure 4-2: The population, vehicle usage rate and the expansion factor distributions of the five subject cities.

information to generate estimates of trip production and attraction for locations. While new data sources such as CDRs do not provide the same detailed demographic and contextual information about individuals or trips, they do provide many high resolution data points over a far longer observation period. Mobile phones offer good, but imperfect measurements of geographic position due to the uncertainty of the location estimates and the nonuniform sampling frequency.

Origin-destination (OD) information is traditionally modeled with data obtained from travel surveys, land use information and census data. First, estimates of trip production and attraction for zones are produced. These trips are then distributed among possible destinations across the city using calibrated gravity or radiation or similar models. Information from the survey are combined with mode choice models to split trips among travel alternatives. CDRs do not provide as detailed demographic and contextual information about travel patterns and behavior as household travel surveys do. Mobile phones offer good, but imperfect measurements of geographic position due to the uncertainty of the location estimates and the nonuniform sampling frequency. However millions of high resolution data points over a far longer observation period make CDRs a high potential data source. Methods developed to incorporate CDRs therefore aim to find a balance between a small and complete dataset that is household travel surveys, and a large but incomplete dataset, namely

CDRs.

In incorporating CDRs into such methods, Alexander et al. and Colak et al. [3, 45], outline a general framework. Location frequencies are found to estimate each location’s function for a user, and classify it as *home*, *work* or *other*. Consequently the trips between these locations are assigned a trip purpose: *home-based-work* (commuting, *home-based-other* or *non-home-based* are inferred. Morning peak commuting and total trips are estimated from filtered users by analyzing consecutive observations at different stay points during the morning peak period (6am-10am). These trips are then normalized to accurately represent actual daily number of trips by measuring how often a user uses their phone, their average number of trips, and the number of days that they were observed. Finally, the number of trips are expanded by the ratio of the population of the source tract to the number cell phone users in that tract. To consider trips made only by vehicles, we weigh obtained person trips by vehicle usage rates in the home census tract of users. To estimate the peak hour traffic volume, the morning period of 6am-10am was weighted in accordance to trip departure time distributions obtained in [45]. Peak hour demand occurs between 7:30am and 8:30am, and the average morning hour demand is multiplied by 1.5 to reflect the peak as per the departure time distributions. Another issue relating to the accuracy of findings is the choice of the administrative boundaries, that is, due to the spatial precision of the data, certain aggregation levels work better than others. This problem is analyzed in detail in previous work, where pseudocode to generate OD matrices and the comparisons to the outputs of traditional models can also be found [3, 45, 164].

Using this information of the trip distribution within the city, we estimate morning peak vehicular volumes from origins to destinations and compare the inferred travel times based on demand with the estimates of an online map provider in the respective routes and hour of the day. We then explore the relationship between travel distance and travel time across many cities. We show that the time lost due to congestion in each city can be accounted by a dimensionless parameter Γ that measures the ratio between the vehicular travel demand and the road infrastructure supply for the

city. To a lesser extent, the differences in congestion levels depend on the population density and the spatial distribution of population. Next, we calculate the detrimental effects of selfish routing by comparing obtained travel times to those that would be observed if the routes were selected to attain the social optimum. We then explore the bounds of the benefits of leveraging information technologies to influence route choices in ways that would help create a more optimal system configuration for vehicular travel. To do so, we implement a generalized selfish routing model that generates expected travel times for varying levels of consideration of overall social good, or λ . We analyze the system gains of socially aware driver behavior, as well as exploring the distributions of benefits and losses at the individual level. We present our findings for five major cities around the world: Boston and San Francisco Bay Area in USA, Rio de Janeiro in Brazil, and Lisbon and Porto in Portugal.

4.2.4 Road networks

For many cities in the US, detailed road network data are made available by local or state transportation authorities. These datasets generally are well maintained, however, many properties are often incomplete or missing entirely. For this purpose we infer required road characteristics to build realistic and routable networks using OpenStreetMap (OSM), an open source crowd sourced mapping tool.

While road networks supplied by local municipalities in the form of shapefiles can often be useful, we have implemented a parser to construct routable road networks from OpenStreetMap (OSM) data due to its global availability. *Nodes* in OSM data represent points representing points of interest or tags or an intersection, and *ways* contain references to nodes that are grouped. They may also contain attributes such as *number of lanes* or *speed limit*, although many roads have this information missing. What all roads have in common though is the road classification, varying between motorway, trunk, primary, secondary, tertiary, residential and trunk roads, as well as a some other irrelevant categories. For our purposes, we filter out roads with irrelevant categories, and residential roads as they are not central to the congestion problem, yet tend to increase computation time significantly. For easing computation, we also

simplify the network by collapsing roads with only one incoming and one outgoing road, if they're in the same road classification. To infer the missing data, we map assign every road a speed limit, number of lanes and a corresponding capacity based on its category and information in [173]. *Motorways* are generally major highways and have a speed-limit of 60 mph with 3 lanes in a direction, whereas primary roads are 40mph with 2 lanes. We assume the free travel time on a segment i is $t_{f,i} = 1.3 * L_i / v_i$, with L_i the road segment length and v_i the speed limit. To estimate the capacity C (vehicles per hour) of a road segment, we utilize the following relationship [173] using the speed v (mph) and the number of lanes n_l :

$$C = \begin{cases} 950 \times n_l, & \text{if } v < 40, \\ (1500 + 30 \cdot v) \times n_l, & \text{if } 40 \leq v < 60, \\ (1700 + 10 \cdot v) \times n_l, & \text{if } v \geq 60. \end{cases}$$

More information about the road networks can be found in Table 4.1.

Road network modeling is a lot more complex then the simple extraction of the topology. Realistic estimation of road capacities, lengths and travel times is essential. We demonstrate our findings in Figure 4-3. The road length and free travel times seem to follow a power-law, free travel times can range from ten seconds to as much as 20 minutes, and similarly for road lengths. Capacities are a direct result of road classes in OSM data: highways, trunks, primary, secondary and tertiary roads are all modeled to have different capacities and number of lanes. To assess overall supply more accurately, we also look at the product of the capacity and the length of the road networks. Our findings suggest that Bay Area, also in accordance with its size, has comparably larger supply.

4.2.5 Traffic flow and travel time

Relating travel performance to traffic conditions has been a long standing problem in transportation. Many different characterizations exist, ranging from conical volume-delay functions to more complex approaches.

A long-standing problem in highway engineering has been the characterization of

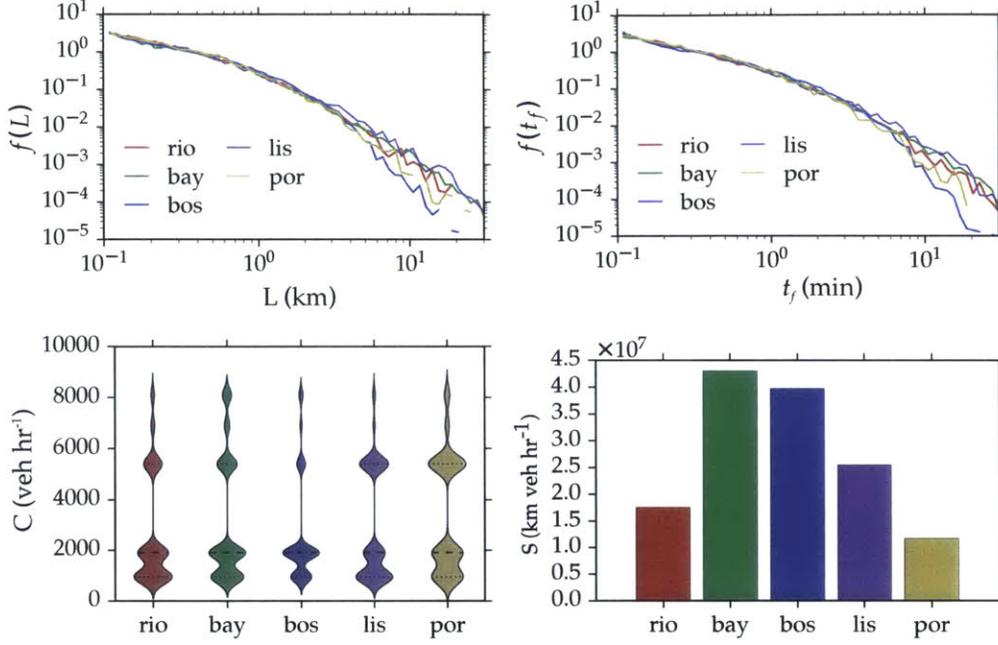


Figure 4-3: Properties of the road networks of five subject cities. Distributions of road segment length L , free travel time t_f , hourly vehicle flow capacity C , and a measure of total supply S .

the relationship between number of vehicles on a road segment, i.e. its *volume*, with the observed travel time on that road segment. Throughout the years a number of different characterizations have been developed ranging from conical volume-delay functions to more complex approaches [31, 156, 2]. One of the most simple and common metrics used in determining the travel time associated with a specific flow level is the ratio volume of vehicles on the road and its maximum flow capacity, also referred to as volume-over-capacity or VoC . At low $VoCs$, drivers enjoy large spaces between cars and can safely travel at free-flow speeds. As roads become congested and VoC increases, drivers are forced to slow down. Based on the guidelines set by the Bureau of Public Roads [106], the VoC of each road segment is used to estimate the travel time according to Eq. 4.1:

$$f_{BPR}(VoC, f_p) = t_f * \left(1 + \alpha (VoC)^\beta\right) * f_p, \quad (4.1)$$

where t_f refers to the travel time under free flow conditions. $\alpha = 0.6$ and $\beta = 4$ are

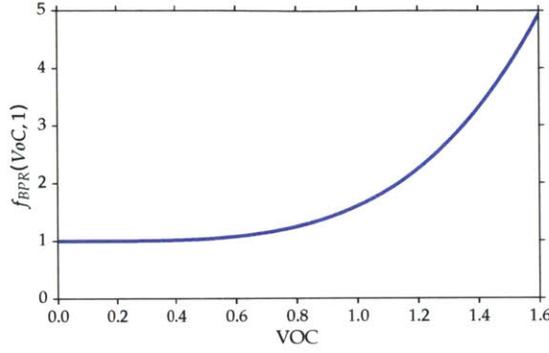


Figure 4-4: BPR function for obtaining travel time from volume and capacity.

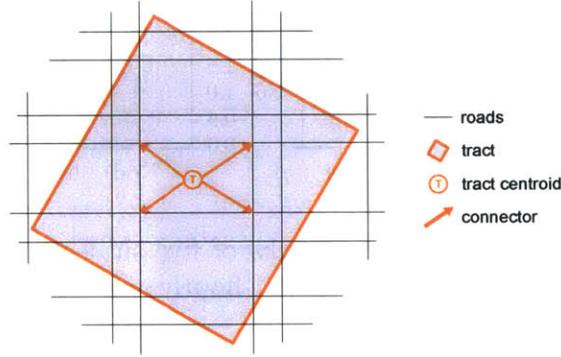


Figure 4-5: A depiction of connector modeling. Tract centroid T is connected to nearest four intersections.

calibration parameters. The relationship is depicted in Figure 4-4. f_p is a city-specific correction factor: $f_p^{bos} = 1.4$, $f_p^{rio} = f_p^{bay} = 1.3$, and $f_p^{lis} = f_p^{por} = 1.0$.

As a second calibration step, once the path-level travel times are obtained, we adjust the travel times by

$$t_c = t + k_{city} * t_{free}, \quad (4.2)$$

where $k_{bos} = k_{rio} = k_{bay} = -0.1$, $k_{lis} = 0$ and $k_{por} = 0.1$.

4.2.6 Traffic assignment

Traffic assignment is a mature domain that aims to bring together travel demand with road infrastructure to better understand traffic and has been studied extensively by urban and transportation planners. In this work, we follow an efficient, static, origin-

based assignment algorithm that focuses on the equilibration of a directed acyclic graph structure emanating from every origin node.

Traffic assignment is a very mature domain that has been studied extensively by urban and transportation planners. Static non-equilibrium models approaches consist of treating all users as homogeneous agents who make route choices prior to departure based on some heuristic related to current traffic conditions (e.g. the path that minimizes travel time). Incremental Traffic Assignment (ITA) is a variant of these static non-equilibrium assignment models that assigns batches of trips serially and updates costs between increments, as an improvement over the simplest all-or-nothing assignment methods. However, these methods results in solutions far from the Wardrop principles [172], where in the resulting system no driver should have an incentive to deviate from their route choice. Many methods to compute the equilibrium have been proposed in the literature [128], the easiest being from Frank-Wolfe (FW) solutions. FW based algorithms are quick to implement but slow to converge to the optimal solution. However they provide no information about which OD-pairs provide what amount of flow to which road segments. Path based algorithms take a step towards path enumeration, but in large networks with a high number of origin-destination pairs and alternative paths, the memory and computational requirement grow very quickly [87, 57, 86]. The more efficient approach is through the use of origin based algorithms, which are computationally feasible, have a fast convergence rate and do store path flows [95, 64]. More complex assignment models aim to take into account the variability in travel times by adding stochasticity to link travel times [54]. The process with which people choose routes is also of great interest to researchers, under the umbrella of route choice models. Prato (2009) presents a good overview of the wide literature on this subject, ranging from logit models to path set generation algorithms [132]. For the scope and the aggregate nature of our work, we opt to implement a static assignment model.

In this work, we will follow Algorithm B, proposed in [57] along with modifications and improvements outlined in [122], an origin based algorithm that focuses on the equilibration of a graph structure referred to as a bush, a directed acyclic graph

(DAG) emanating from every origin node introduced to the graph as the centroid of the origin tract. These structures are used with the reasonable assumption that in the equilibrium flows, no directed cycles should exist as no driver has an incentive to increase his/her travel time. The computational efficiency of this algorithm stems from the fact that DAGs can be traversed in linear time. The algorithm used in this work is outlined in Figure 4-6.

In these algorithms, the objective is to minimize the the distance between the current solution and the optimal solution. In this work, relative gap is used as the measure of convergence.

$$r_g = 1 - \frac{\sum_{o,d} t_{od} d_{od}}{\sum_{e \in E} t_e v_e}, \quad (4.3)$$

where t_{od} and d_{od} represent the demand and the travel time between an origin and a destination, and t_e and v_e represent the travel time and the volume on a road segment e . The numerator and the denominator essentially measure the same thing: the total travel time in the system. Theoretically, r_g is supposed to be equal to zero. This ensures that all drivers in the system are in fact taking the shortest possible routes, and the optimization problem is fully solved. Traffic assignment algorithms aim to bring r_g as close to zero as possible.

A critical design element of the implementation of origin based algorithms is the modeling of tract centroids, representing an aggregation of all the actual origins and destinations within the area, and the connectors, the hypothetical segments representing driver movement within the tract before joining the modeled road network [133]. Figure 4-5 depicts the implementation of connectors in this work, where tract centroids are connected to the four nearest intersections.

ALGORITHM B(N)

Initialize B as the shortest path tree rooted at the origin.
Assign all flows to links to B .
while $r_g > 0.001$
 do $\left\{ \begin{array}{l} \text{for all origins } o \\ \quad \text{do} \left\{ \begin{array}{l} \text{Add to } B_o \text{ edges } e \text{ with negative reduced costs.} \\ \text{Solve the Restricted Master Problem for } B_o. \\ \text{Simplify } B_o \text{ by removing } \{e | x_e = 0\}. \end{array} \right. \end{array} \right.$

RESTRICTED MASTER PROBLEM(Bush B , ϵ)

Update costs on all links on B .
Calculate the longest route tree with paths P_i and costs U_i .
Calculate the shortest route tree with paths p_i and costs u_i .
if $\max\{U_i - u_i, \forall i\} \leq \epsilon$, **stop**.
 else continue.
for all j
 do $\left\{ \begin{array}{l} \text{set of links in } p_i \text{ not in } P_i : S_j = p_i \setminus P_i \\ \text{set of links in } P_i \text{ not in } p_i : L_j = P_i \setminus p_i \\ \text{difference in costs to } j : g = (u_j - u_i) - (U_j - U_i) \\ \text{total marginal cost of sets } S_j \text{ and } L_j : h = \sum_{e \in S_j \cup L_j} c'_e \\ \text{flow to be shifted : } dx = \min\{g/h, \min\{x_e | e \in L_j\}\} \\ \text{add flow to shorter path : } x_e = x_e + dx, e \in S_j \\ \text{remove flow to shorter path : } x_e = x_e - dx, e \in L_j \\ \text{update travel times : } t_e, e \in S_j \cup L_j \end{array} \right.$

Figure 4-6: Algorithms used in the computation of equilibrium. [57, 122]

4.3 Results

4.3.1 Approach

We formalize the traffic problem by modeling route choice as follows: Every driver i makes a choice of the route p to their destination. This choice depends on a personal utility $u_i = \sum_{e \in p} c_e(x_e)$, expressed as the sum of the costs c of every road segment e along the chosen route. For simplicity, we assume that the cost of a road segment for driver i is equal to the travel time, $c_e(x_e) = t_e(x_e)$, where $t_e(x_e)$ represents the travel time t observed on road e for vehicle flow x_e . We can then define the total cost incurred by all users as $C = \sum_{e \in E} x_e t_e(x_e)$. The flow configuration that results in the optimal cost is referred to as the socially optimal flows obtained by a typical minimum cost network flow program [1]:

$$\begin{aligned}
 & \underset{x_e \forall e \in E}{\text{minimize}} && C \\
 & \text{subject to} && \sum_p f_p^{st} = f^{st} \\
 & && x_e = \sum_s \sum_t \sum_p f_p^{st} \delta^{st}(p, e), \\
 & && x_e \geq 0, f_p^{st} \geq 0.
 \end{aligned} \tag{4.4}$$

where x_e refers to the flow on road e , f_p^{st} is the flow between the source s and target t on route p , and $\delta^{st}(p, e) = 1$ when road e lies on route p .

As drivers make selfish choices, the system settles into a suboptimal state. Although driver i only experiences and considers his/her own travel time, the cost the whole system incurs also includes the marginal cost driver i imposes on all other drivers on the road segments he/she takes. The set of flows that occur when every driver minimizes their own travel time is referred to as the *user equilibrium flows*. Theoretically, in the resulting system state, no driver can benefit from deviating from their route. This idea, essentially describing a Nash equilibrium in roads, is captured in Wardrop's principles in transportation [172]: the journey times on all the used routes for an origin-destination pair are equal, and are less than those which would

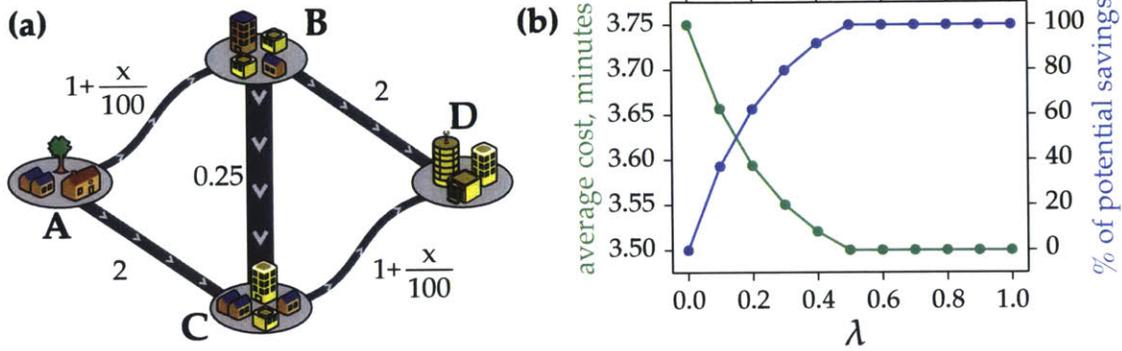


Figure 4-7: Illustration of routing equilibrium. (a) In this small network, 100 drivers are going from A to D. The road labels represent the costs of travel as a function of vehicle flows. User equilibrium allocates the flows between paths as $f_{ABD} = f_{ACD} = 25$ and $f_{ABCD} = 50$, and the average travel time is 3.75 minutes for all drivers. Socially optimal flows decrease total travel time to 3.5 by $f_{ABD} = f_{ACD} = 50$ and $f_{ABCD} = 0$, with road BC remaining unused. (b) Achieved percentage of potential savings for increasing values of social good weight λ : 10% and 20% social good weight results in 40% and 60% of potential savings, respectively.

be experienced by a single vehicle on any unused route. This routing game is solved through a potential function $\phi_e(x_e) = \int_0^{x_e} t_e(x)dx$ such that $\phi'_e(x_e) = t_e(x_e)$ [113]. The convex program for the user equilibrium problem has been formulated [19] as follows:

$$\begin{aligned} & \underset{x_e \forall e \in E}{\text{minimize}} && \sum_{e \in E} \phi_e = \sum_{e \in E} \int_0^{x_e} t_e(x)dx \\ & \text{subject to} && \text{constraints in Eq. 4.4.} \end{aligned} \tag{4.5}$$

Figure 4-7(a) depicts an example that captures solutions for equilibrium and optimal flows for a widely used toy network. For the demand of $d_{AD} = 100$, the user equilibrium flows allocate 50 drivers on path ABCD and 25 drivers on paths ABD and ACD each, resulting in a travel time from A to D of 3.75 regardless of the path chosen. The socially optimal configuration avoids allocating too much flow on the path ABCD, as its marginal cost is higher than those of paths ABD and ACD. By minimizing the marginal cost, path ABCD receives no flow and the average cost is minimized at 3.5.

In order to assess the benefits of different scenarios based on travel demand infor-

mation we make use of the formulation proposed in [42]. We reconfigure the utility function of a driver as a linear combination of the cost he/she will incur and the total marginal cost his/her choice imposes on everyone else:

$$\begin{aligned} c_e^\lambda(x_e) &= (1 - \lambda)t_e(x_e) + \lambda \frac{d[x_e t_e(x_e)]}{dx_e} \\ &= t_e(x_e) + \lambda x_e \frac{dt_e(x_e)}{dx_e} \end{aligned} \tag{4.6}$$

λ defines the weight towards social good; it is a parameter ranging between 0 and 1. A driver with $\lambda = 1$ chooses routes with respect to the marginal costs, thus moving the system closer to the system optimum. Conversely, a user with $\lambda = 0$ only considers the cost of his route and potentially moves the system away from optimality. The resulting convex program for the socially aware routing problem is as follows:

$$\begin{aligned} &\underset{x_e \forall e \in E}{\text{minimize}} && \sum_{e \in E} \int_0^{x_e} c_e^\lambda(x_e) \cdot x_e dx_e \\ &\text{subject to} && \text{constraints in Eq. 4.4.} \end{aligned} \tag{4.7}$$

For the city depicted in Figure 4-7(a), the user equilibrium configuration results in an average cost of 3.75 minutes per driver versus 3.5 minute for the system optimum, meaning solely by adjusting routing behavior to $\lambda = 1$, a benefit of 0.25 minutes can be achieved per driver. Figure 4-7(b) shows that for $\lambda = 0.1$, when the drivers begin valuing social good as well, the average cost drops to approximately 3.65, and almost 40% of potential savings are realized. In fact, the social optimum is achieved at $\lambda = 0.5$.

4.3.2 Travel times

To understand the relationship between travel demand and driving travel times, we begin by comparing our five cities during estimated morning peak period traffic conditions. The areas of analysis are significantly diverse: Rio is very highly populated over its large extensions; while Porto's population density considerably decreases af-

	City				
	Rio	Bay	Bos	Lis	Por
population (mil.)	12.6	7.15	4.5	2.8	1.7
area (1000 km^2)	4.6	18.1	4.6	2.9	2.0
# of total users (mil.)	2.19	0.43	1.65	0.56	0.47
# of calls (mil.)	1045	429	905	50	33
data period	5 months	3 weeks	2 months	14 months	14 months
data type	tower	tower	lat-lon	tower	tower
# of cell towers	1421	892	N/A	743	335
# of edges (th.)	40.9	24.3	21.8	28.1	15.1
# of nodes (th.)	22.1	11.3	9.6	16.1	8.6
# of tracts	381	1139	732	295	272
roads (th.miles)	6	20	12	7	3
all trips (mil.)	0.432	1.015	0.916	0.324	0.171
commutes (mil.)	0.183	0.353	0.401	0.151	0.084

Table 4.1: A comparison of the extent of the data involved in the analysis of the subject cities.

Table 4.2: A comparison of general properties of the subject cities.

	City				
	Rio	SF Bay	Boston	Lisbon	Porto
Population (millions)	12.6	7.15	4.5	2.8	1.7
Area (1000 km^2)	4.6	18.1	4.6	2.9	2.0
Demand (veh km hr^{-1})	3.1	9.1	5.4	2.9	1.1
Supply (veh km hr^{-1})	17.6	43.0	39.7	25.5	11.7
Demand-to-supply (Γ)	0.18	0.21	0.14	0.11	0.09
Expansion factor	890	100	32	96	164
Vehicle usage (veh person $^{-1}$)	0.25	0.67	0.67	0.56	0.62

ter $r > 20$ km from the most dense location. Rio de Janeiro, the Bay Area and Lisbon extend across Guanabara Bay, the Bay, and Tagus, respectively, and have many inhabitants commuting on few bridges.

As a consequence of their differences, cities demonstrate varying traffic conditions, as shown in Figure 4-8. The volume-over-capacity ratio (VOC) measures how successfully a road segment is able to cope with the assigned volume of vehicles, with high VOC values indicating more congestion. High VOCs are generally observed on highways, as they provide faster means of travel due to their wider roads, increased number of lanes, and higher speed limits. Additionally, bridges and roads that lie

	straight-line distance fit, KS statistic				
	Rio	Bay	Bos	Lis	Por
power-law	0.138	0.193	0.142	0.200	0.177
exponential	0.087	0.035	0.082	0.016	0.026
lognormal	0.049	0.023	0.028	0.021	0.018

Table 4.3: KS test statistics for lognormal, exponential and power-law distribution fits for the straight-line distances.

central in the network topology are typically congested due to a lack of alternative routes.

We begin by analyzing the efficiency of urban mobility for the five regions to understand the mechanisms underlying observed travel times. The main determinant of congestion is travel demand, which is heavily tied to commuting trip distances during weekday peak travel times. In Figure 4-9(a) we demonstrate that the straight-line (Euclidean) commuting distances, d , follow a lognormal distribution, $f(d) = \frac{1}{\sqrt{2\pi\sigma d_b}} e^{-(\ln(d)-\mu)^2/2\sigma^2}$ with means ranging from 5 to 8 kilometers ($\mu = 1.6 - 2.1$) and standard deviations ranging from 2 to 4 kilometers ($\sigma = 0.7 - 1.2$). It can be observed that majority of trips span relatively short distances, and trips over 25 kilometers are uncommon. However what makes a city more traversable are the speeds at which drivers can span these distances. In Figure 4-9(b) we investigate the effective speeds in both free and congested traffic conditions. It can be observed that cities exhibit similar free travel speed distributions, normally distributed with μ fluctuating around 50 km hr⁻¹ with mean values reported in the legend. The differences in road network supply $S = \sum_{x_e > 0, e \in E} l_e C_e$ (km vehicles hr⁻¹), where l_e and C_e are the length (km) and the flow capacity (vehicles hr⁻¹) of a road segment e , explains the slight differences in free flow speeds, as seen in Table 4.2. These differences are significantly more apparent in speed distributions under real traffic conditions: the effective OD travel speeds in Rio, the Bay Area and Boston decay considerably compared to those in free traffic conditions; while the speeds in Porto and Lisbon change less. We explore further these two different responses given the demand profiles of each city.

To that end we analyze the experienced travel times per distances traveled in Figure 4-9(c). We observe a strong yet very simple relationship that pronounces the

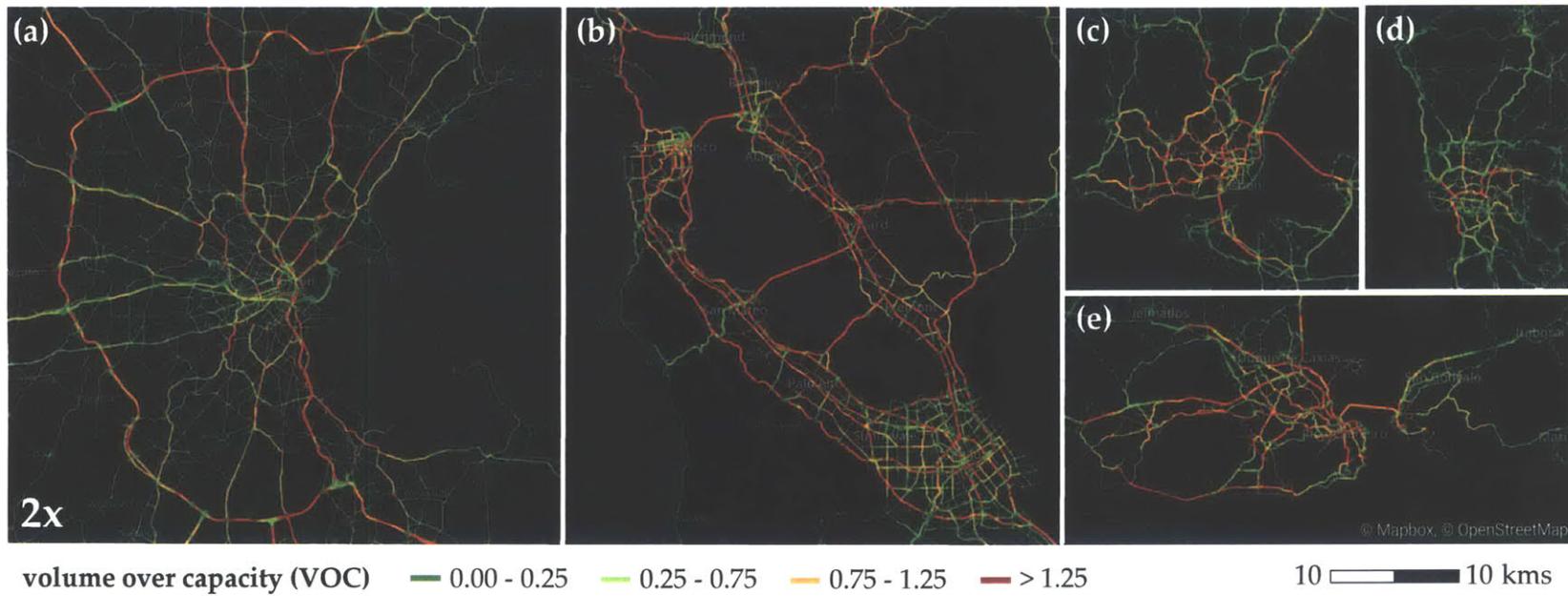


Figure 4-8: The maps of VOCs (volume over capacity) of the roads in the user equilibrium configuration. The depicted cities are (a) Boston, USA, (b) San Francisco Bay Area, USA, (c) Lisbon, Portugal, (d) Porto, Portugal, and (e) Rio de Janeiro, Brazil. Higher VOCs are generally observed in highways as they provide faster means of travel. (Boston is 2x the distance scale.)

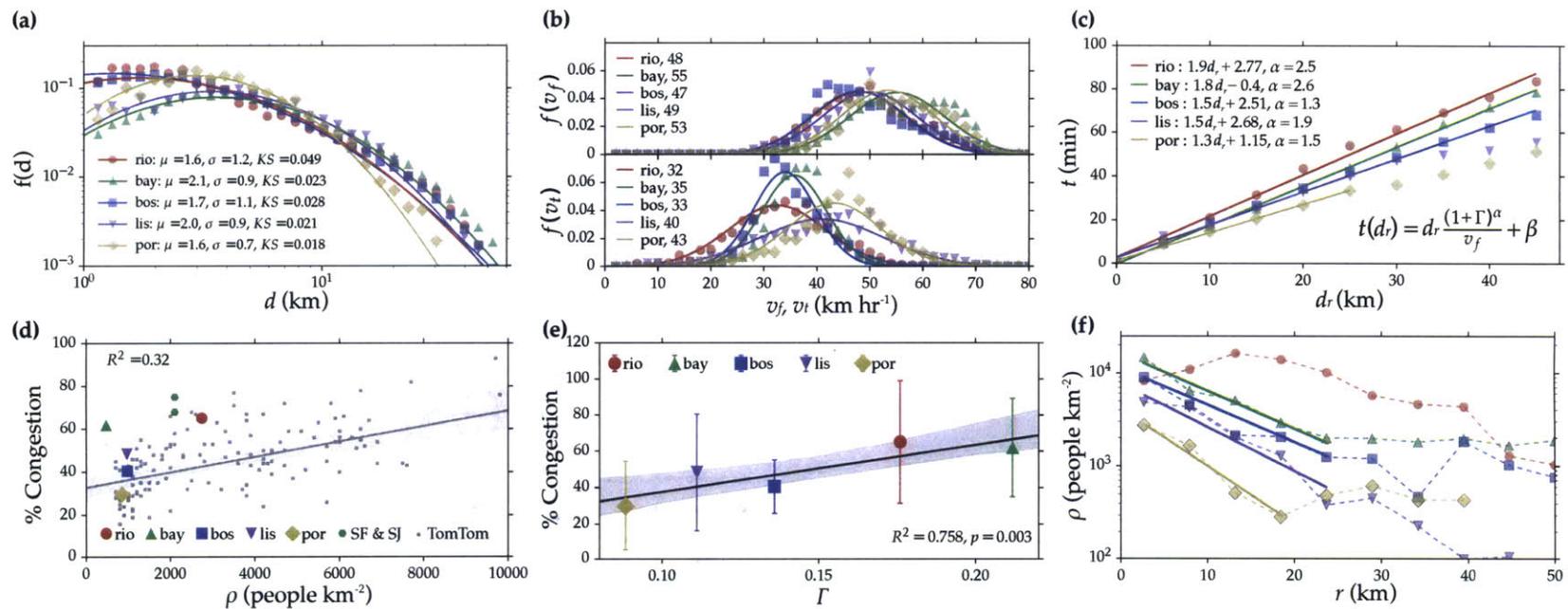


Figure 4-9: Comparisons of cities and their congested travel. (a) Distributions of commuting trip distances, d , in the morning peak period with parameters of the fitted lognormal distribution depicted in the legend. (b) Distribution of trip free flow speeds, v_f , and in traffic conditions, v_t . (c) Commuting travel times versus route distances of commuters, d_r . (d) Estimates of overall mean % of time lost in congestion versus population density p for TomTom Traffic Index estimates and our analysis. (e) Relationship of overall mean % congestion to the demand to supply ratio, Γ , for the five subject cities, with errorbars specifying the standard deviation. (f) Average population density ρ as a function of distance from most dense area in the region, r .

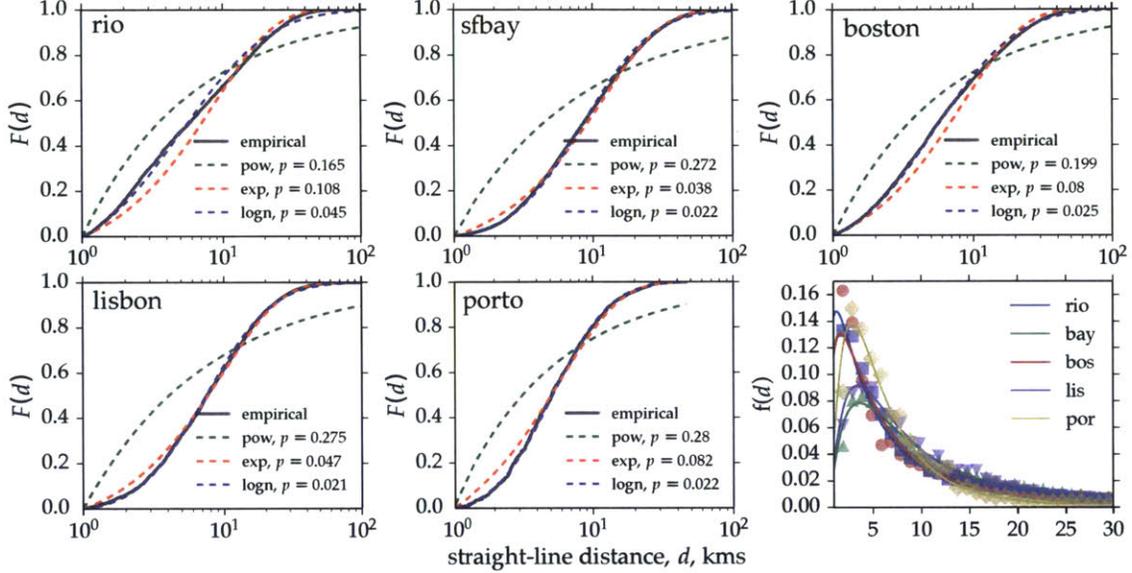


Figure 4-10: The probability distribution fits for the straight-line distances.

differences between the subject cities: Rio de Janeiro is the slowest city and is followed next by the Bay Area, and Porto is the fastest. All cities exhibit a linear relationship, with the exception of long distance trips in Porto and Lisbon where a different regime appears for longer distances. To explain this observation, we model travel times by city-specific parameters describing the demand, the capacity, and observed free traffic speeds. In doing so, we define demand-to-supply ratio of a city as,

$$\Gamma = \frac{\sum_{e \in E} l_e x_e}{\sum_{x_e > 0, e \in E} l_e C_e}. \quad (4.8)$$

This dimensionless measure is a simple ratio of the total distance traveled by all vehicles to the upper bound of the total vehicle kilometers the road network can support per hour, thus capturing the load on the road infrastructure by bringing together trip distances, trip magnitudes, road capacities, and the distances they span as shown in Table 4.2. Using this measure along with v_f , the average free travel speed of each city, we are able to better explain the linear relationship between travel time and distance by

$$t(d_r) = d_r \frac{(1 + \Gamma)^\alpha}{v_f} + \beta, \quad (4.9)$$

where α values vary between 1.3 – 2.5, essentially describing the sensitivity of the city to the stress imposed by travel demand on its road infrastructure.

In order to untangle the particular ordering of cities in terms of speed and understand why some cities are more congested than others, we investigate a typical relationship in Figure 4-9(d) to test the common conception that cities with higher population densities tend to exhibit more heterogeneity in their demand profiles, and therefore tend to be more congested. For this purpose, we measure the ratio of the time lost in traffic to the travel time under free flow conditions, known as the *traffic index*, along with those measured for many other urban areas by TomTom, a leading GPS company. We consider the percentage of congestion, defined as the percentage of additional travel time due to traffic compared to free flow conditions, for different population densities in these various cities. We observe that Boston, Lisbon and Porto fall on the fit model, whereas the Bay Area and Rio demonstrates a significantly higher level of congestion. The outlier appearance of the Bay Area is a consequence of the arbitrary definitions of urban areas and its influence in population density as pointed out in [5]. To account for this, we plot the subdivisions of San Francisco and San Jose which support the relationship as they lie closer to the fit. Interestingly, the dimensionless demand-to-supply ratio Γ lacks this problem and presents a better linear trend with congestion for the five analyzed urban areas as depicted in Figure 4-9(e), despite the broad behavior of the traffic response. The two most congested cities have the highest ratios, the Bay Area closely followed by Rio de Janeiro while Porto and Lisbon, the two least congested cities, have lower ratios.

To finalize our analysis, in Figure 4-9(f) we measure how population densities are spatially distributed from the most densely populated region in each of the subject cities based on the chosen administrative level. The results show different spatial distributions in the population density of the five cities. First, it verifies the expected effect of higher population densities in increasing congestion. It also highlights the importance of the spatial distribution around the highest density point. Lisbon and Porto present densities of population below 500 people km^{-2} for distances of $r > 20$ km; while the other three cities stabilize in values greater than 1000 people km^{-2} .

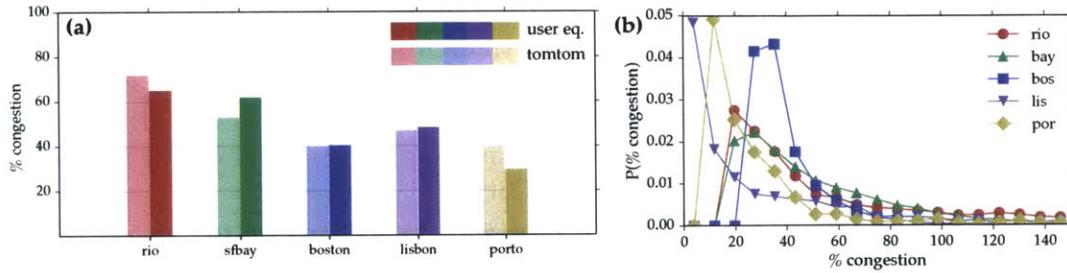


Figure 4-11: % congestion distributions for the five cities. (a) Overall % congestion levels obtained from user equilibrium compared to values from TomTom. (b) Distributions of % congestion for trips for the five cities.

These differences can explain the two types of responses in the effective travel speeds presented in Figure 4-9(b) where Lisbon and Porto belong to a city type of lower density, in agreement with the characteristics described in [24]. Taking these results together, we observe that congestion increases with Γ and appears to be influenced by the spatial distribution of population density and its gradient.

4.3.3 Selfish Routing

In this section, we compare the travel times for commuters in free flow, socially optimal and user equilibrium flow configurations. Our findings in the five subject cities are outlined in Table 4.4. Although the estimated free travel time averages are similar, congestion plays a significant role: Lisbon commuters lose 2.1 minutes on average by selfish routing preferences. Rio de Janeiro exhibits an average loss of 2.6 minutes on average incurred by selfish routing. The results show that on average 15-30% of total minutes lost in congestion is caused solely by selfish routing.

Although a more nuanced methodology incorporating stochastic traffic assignment and probabilistic origin destination matrices would likely improve validation, our formulation and central findings would remain robust as they are based on aggregate and endogenous, albeit simplified, behavior of our system. Furthermore a principled and singular validation source does not exist for our cities; we instead use an online map provider as a validation benchmark. Although the validation data is also the product of internal models and estimations, it is of value as they are obtained from

Table 4.4: Comparison of cost findings in the subject cities for the morning peak hour. Colored rows indicate the loss of travel times from free travel times to socially optimal flows, and from socially optimal flows to user equilibrium flows for commuters, respectively (FTT: free travel time, SO: social optimum, UE: user equilibrium, % S: percentage of total congestion attributed to selfish routing, $S = 100 * Benefit/Loss$)

(min)	City				
	Rio	SF Bay	Boston	Lisbon	Porto
FTT	20.6	21.1	19.3	22.4	15.3
Loss	14.1	12.5	8.2	8.0	4.0
UE	34.7	33.6	27.5	30.4	19.3
Benefit	2.6	2.6	1.3	2.1	1.1
SO	32.1	31.0	26.2	28.3	18.2
% S	18	21	16	27	28

	travel time distributions, regression statistics					
	coef	st. error	$P > t $	R^2	AIC	
Rio	0.7400	0.007	0.000	0.839	$1.239 * 10^4$	
Bay	0.6490	0.006	0.000	0.876	$1.129 * 10^4$	
Bos	0.5770	0.005	0.000	0.882	$0.904 * 10^4$	
Lis	0.6297	0.006	0.000	0.854	$1.232 * 10^4$	
Por	0.7602	0.005	0.000	0.922	$1.078 * 10^4$	

Table 4.5: Regression statistics for travel time estimations.

an independent data source to ours. In Figure 4-12(a), we compare the distributions of obtained travel times with those obtained from the map provider in the morning peak hour between 7:30 and 8:30 for 2000 origin-destination pairs with the highest commuting flows. There is an overall overestimation of travel times, which strengthens the notion that route choice in reality might not be a perfect user equilibrium or a social optimum, but somewhere in between. Neither the provider's nor our findings are expected to have accurate travel time variability as these comparisons are estimates of typical travel times for the given OD pairs and they act as a first step towards the validation of our estimated travel times based on the assigned traffic flows obtained from the phone data.

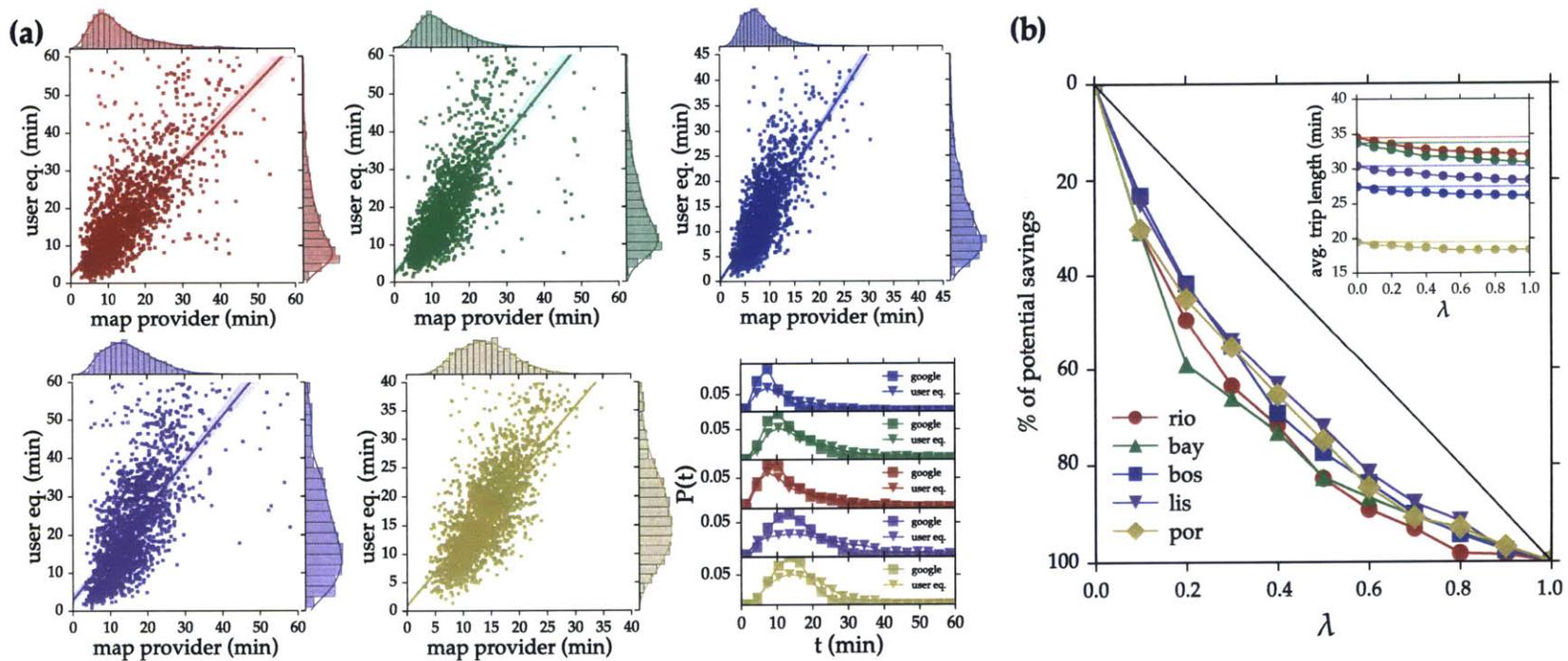


Figure 4-12: Travel time comparisons and potential savings. (a) Comparison of travel times and their distributions between user equilibrium versus routes obtained from the online map provider. Origin-destination samples consist of 2000 origin destination pairs with the highest commuting flow magnitudes for each city. (b) The percentage of potential savings in average commuting times for the five cities for varying levels of social good weight of routing. (inset: the travel time savings represented in actual minutes)

4.3.4 Weight of Social Good

In assessing the effects of socially aware routing behavior for the subject cities, we calculate the average commuting time for various levels of λ . The inset of Figure 4-12(b) depicts the decrease in average commuting travel times for increasing λ in all five cities, ranging from an average of 1 to 3 minutes. More importantly, the shape of the curves indicate that even modest social consideration weights can realize a significant portion of the potential savings. Figure 4-12(b) collapses these curves to represent realized potential savings as a percentage to exhibit a striking similarity between the five cities in terms of response to socially aware routing. To assess the economies of such routing behavior, we measure the Gini index of the obtained curves, by definition higher values of G indicate higher savings for smaller levels of social good weight. Our findings show that G ranges from 30 – 40%: $G_{\text{rio}} = 41\%$, $G_{\text{bay}} = 42\%$, $G_{\text{bos}} = 33\%$, $G_{\text{lis}} = 30\%$ and $G_{\text{por}} = 34\%$. These findings indicate congested cities benefit more from incorporating social good considerations into routing behavior.

4.3.5 Travel time benefit distributions

In the previous section we characterized the percentage of potential savings that can be obtained for increasing levels of social consideration. However these benefits are achieved at the expense of time of drivers who adjust their commute for the benefit of others. The unwillingness to give up time is the defining factor in drivers' failure to reach an optimal state on their own. This highlights the importance of fairness of the distribution of who has to sacrifice versus who benefits in terms of both the success potential of the implementation of policies or a reward/punishment reinforcement schema. Figure 4-13 (a) demonstrates one such schema, where drivers are shown a route that corresponds to a choice of λ which might result in a travel time sacrifice.

Our findings, in accordance with the results of the previous sections, indicate a net bias towards benefits, meaning the number of drivers that benefit outnumber those who sacrifice. Figure 4-13(b) summarizes the benefit distributions for the five cities for $\lambda = 0.1$ and $\lambda = 1$. The former exhibits a less spread distribution than the latter

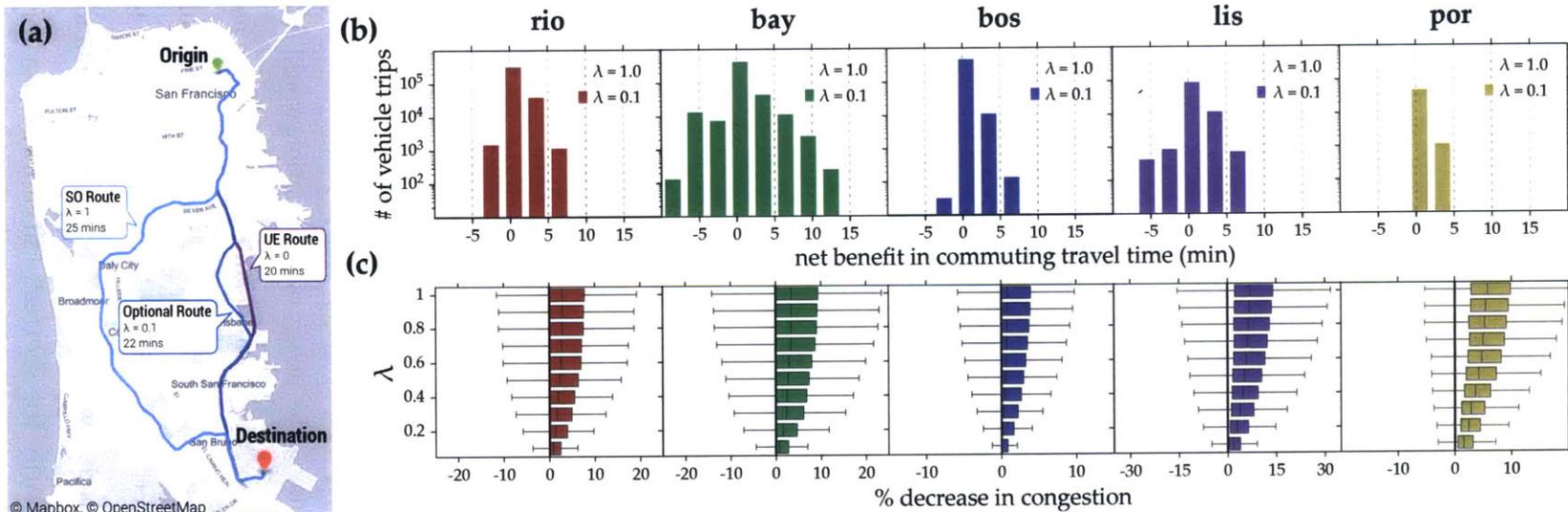


Figure 4-13: Benefit and congestion decrease distributions for different weights of social good. (a) A depiction of three route alternatives with the corresponding travel times for a trip from Union Square to San Francisco Airport for $\lambda = 0$; $\lambda = 0.2$ and $\lambda = 1$, respectively. (b) Counts of vehicle trips and observed travel time benefits for $\lambda = 1$ and $\lambda = 0.1$. Negative benefits refer to increase in travel times for vehicles sacrificing for the social good. The spread of the distributions increase for higher λ . (c) The response of distributions of percentage decrease in time lost to congestion to increasing values of λ . The skewness towards positive values of congestion decrease indicate movement towards more optimal configurations.

but the skewness remains inherent to the distributions. Although the average benefits described in the previous sections appear small, it should be noted that 10 minute benefits can be observed for tens of thousands of vehicles. Figure 4-13(c) describes in more detail how the positive skewness evolves for increasing social consideration. For higher λ , the % decrease in congestion distributions are shifted towards positive values, indicating a net benefit. This result demonstrates the potential of incentive schemes which could compensate the few drivers who sacrifice under consideration of social good.

4.4 Discussion

The economic and social costs of congestion are crippling. In addition to the overall loss of time, congestion underlies many major economic and urban issues such as increased gas consumption, infrastructure deterioration and CO_2 emissions. In this work we utilize massive amounts of data to estimate peak hour travel demand and understand travel times. We then explore the power of information based routing on congestion alleviation.

Our findings suggest very interesting similarities in the behavior of the five subject cities to explain congestion and potential benefits of social routing. Commuting distances follow a lognormal distribution and free travel speeds are normally distributed. A city's unique congestion fingerprint is strongly related to measurable characteristics. The population density and its spatial distribution together with the Γ parameter of demand-to-supply ratio are the two driving factors of the observed congestion in a diverse range of cities. Further, given the current state of traffic, we then estimate how centralized routing schemes using the power of information would reach possible benefits in travel times. Such information is important, as it allows the assessment of the upper bounds of routing policies, if effective in implementation, would influence the traffic on a city scale. In practice, this would imply that we could have similar routing applications that we use today with the incorporation of demand profiles to provide routes that are not necessarily the shortest, but also the best for decreasing

overall congestion.

We find that routing solutions that mimic socially optimal configurations, i.e. $\lambda = 1$, have a limit of decreasing time lost in congestion by up to 30%. This is in contrast with the effectiveness of direct and costly interventions where 1% target decrease in demand can achieve 18% decrease in travel times [171]. While in both scenarios the collective benefits for the whole city can be significant (15% – 30% decrease); the observed time benefits the average individual receives are marginal, ranging from 1 to 3 minutes. Furthermore, these times are below the travel time variability based on events, weather conditions, or traffic lights [27]. Our findings indicate that in the best-case scenario, time savings would be imperceptible for the majority of the drivers. From this, it is clear that such routing solutions can't fix the traffic problem for individual drivers but rather would contribute to the city as a whole. The advantage is that in the context of the implied routing application, the number of vehicles sacrificing their travel time is significantly smaller than the number of those that benefit. Lower levels of weight towards social good will also moderate the magnitude of benefits and losses, consequently making the policies fairer and easier to implement.

Open work in this subject contains, but is not limited to, a more generalized bottom-up approach to comparison of cities that includes various modes of transportation to demonstrate their similarities, differences and their consequences. As the volume, the variety and the resolution of data increase along with the expected disruptions from connected self-driving cars and similar technologies, this front of research will become more relevant to facilitate the study and planning for the future of urban mobility. With more updated demand models extracted from communication technologies, understanding the network effects on congestion will become easier to pinpoint and address. In addition, planning tasks on urban mobility previously difficult to tackle may now be addressed at lower costs and with much larger samples of the population. For example, a thorough analysis of how travel time and congestion is distributed among the population and its split by income and other sociodemographic characteristics remains an open front.

Chapter 5

Coupling Electric Vehicle Charging with Urban Mobility

Transportation electrification introduces a spatiotemporal tie between the traditionally independent power and transportation infrastructures through plug-in electric vehicle (PEV) charging. With the steadily increasing rates of PEV adoption in urban areas, the imminent paradigm shift in electricity consumption poses a challenge for researchers to untangle the relationship between mobility and electricity demand. In this work, we provide an understanding of PEV mobility by coupling origin-destination information obtained from mobile phone data with PEV charging session data for the San Francisco Bay Area, USA. We first lay out a methodology to extract mobility patterns of PEV drivers in the area. Next, we present the spatial and temporal characteristics of charging sessions and show that commuting patterns and PEV energy consumption are closely linked. We develop charging schemes that investigate the impact of arrival time scheduling and shifting of charging activity in reducing the power load in the peak hour. Finally, we quantify and evaluate the potential benefits of such schemes to demonstrate that substantial savings are achievable. Our results advance our ability to better manage the current and future state of electrified urban mobility by coupling the transportation and electricity needs.

This chapter is based on [46].

5.1 Introduction

With growing population and rapid urbanization, the demand for mobility and electricity in urban areas are not only increasing in magnitude, but also are becoming more concentrated spatially and temporally. The infrastructures that serve these needs, namely the road networks and the power grids, are under high levels of stress in efforts maintain reliable service. The low costs associated with the extraction and processing of petroleum and coal have led us to historically rely on these high emission fossil fuels to meet the ever-increasing transportation and electricity demand. As a consequence, in 2013, global CO_2 levels exceeded 400 ppm, a mark previously deemed as the critical threshold above which the effects on earth's climate would be irreversible [152]. Decades before reaching this mark, rising CO_2 emission levels had triggered the search for cleaner alternative fuels for transportation. Although the alternatives developed over the years failed to overtake conventional vehicles that use gasoline [109], today's plug-in electric vehicle (PEV) technology is the most promising candidate up to date. In the early stages, issues like range anxiety, charger unavailability and high prices hindered the adoption of PEV technology. However, these issues are overcome today with improvements in battery technology, tax breaks and subsidized charging programs. As a result, PEVs are becoming a more viable means to move and are being adopted by drivers at steadily increasing rates. According to the US Energy Information Administration, the number of PEVs in the USA doubled between 2013 and 2015 and is expected to reach 20 million by 2020 [165]. As PEVs become more ubiquitous, road networks and power grids will become tightly interlocked in their efforts to meet the mobility and electricity demand of the people. This coupling calls attention to an imminent need to understand the typical characteristics of the demand for electrified transportation to build solutions for its management.

Facilitating and expediting the movement of people and goods across cities has been a perennial goal of scientists, engineers and planners. To better plan for the mobility demand, transportation planners founded meticulous modeling practices that provide travel demand information at both aggregated and disaggregated levels.

These models contain modules that capture commuting patterns, activity profiles, mode and route choices, as well as short and long term trends such as changes in rates of car ownership or population growth [126]. In their implementation, travel demand models typically make use of synthetic populations obtained by household travel surveys to better mimic the behavior of all travelers. These surveys are costly and require a lot of labor, therefore can be carried out infrequently for small sample sizes. Recently, models that utilize mobile phone data have been introduced to the literature, providing cheaper and complementary means to generate travel demand information. Although the data used in these models lack the level of detail and precision of travel surveys, they make up for various disadvantages of surveys, as they capture the movement of millions as opposed to thousands and can be processed quickly at lower cost [28, 88, 3, 164, 45]. Despite the abundance of research on this front, there is currently a scarcity of works that aim to measure and assess the mobility of PEVs.

The planning for the mobility needs of PEVs are particularly important in the context of the vulnerability of the power grid to outages that can cascade drastically. This was most recently exemplified by the severe power outage observed in India in 2012 that affected approximately 650 million people. Although smaller in comparison, large power grid failures were also observed in Europe and in the United States over the last decade [80]. Failures of this scale signified a need to recontextualize the electricity infrastructure, and transdisciplinary approaches were promoted in tackling the complexity of the problem [34]. In addition to this already burdened landscape, the introduction of PEVs signal a substantial increase in total load as well as currently unpredictable changes in the norms of energy consumption. Therefore extending the current know-how of solutions that more efficiently manage the power grid at the urban scale has been of utmost interest to researchers. In this topic, a body of literature specifically focused on the nature of network reliability, the role of network topology on the spread of cascading failures [127, 108, 35, 81, 33, 58]. Other works analyzed urban microgrids [73], and the implications of the introduction of new clean sources to the energy market [115]. In the particular subject of PEVs and their

impact, methods of optimization and control of PEV electricity consumption [90, 18, 38, 114] is a rich avenue. Problems that researchers tackle on this front include measuring impact on the grid [44, 161, 74, 97, 134, 162, 82], developing accurate PEV models [184], energy management [137, 167, 166], smart charging strategies that probe centralized and decentralized approaches [103, 89], scheduling [160, 180], peak shaving, emissions, pricing models [185, 65], and joint optimization of power and transportation networks [4]. A common shortcoming observed in these works is the narrow scope in incorporating mobility information into the analyses, often limited to the estimation of arrival or departure hours. The literature currently lacks the incorporation of mobility patterns at the metropolitan scale into the models of electricity demand management in a systematic way.

In this work, we target these gaps in the literature to extend the current knowledge of transportation based electricity demand from a complex systems perspective. For this purpose, we bring together three independent data sources: (i) mobile phone activity of a large sample of the residents of the San Francisco Bay Area, (ii) charging sessions obtained from PEV supply equipment in the same region, and (iii) surveys on the use of conventional and electric vehicles, together with census data for income information at the zipcode level (see Methods). In the first part of the work, we estimate vehicular mobility in the Bay Area using the mobile phone activity of a large sample of residents (Figure 5-1). We then present a methodology to estimate PEV trips from the overall mobility patterns by utilizing information obtained from surveys regarding the income and travel distances of PEV drivers (see Methods). In the second part, we analyze the various aspects of charging activity to characterize the nature of electricity demand at charging stations. We present observations regarding visitation frequencies, arrival and departure hours, typical per session energy consumption patterns, and power levels. We observe that PEV charging patterns are highly regular with morning and evening peaks. Charging sessions have temporal flexibility, that is, PEVs are typically charged for only a portion of the time they stay plugged in. In contrast, the power consumption happens immediately upon arrival, and makes no use of this flexibility. In the third part, we explore the relationship be-

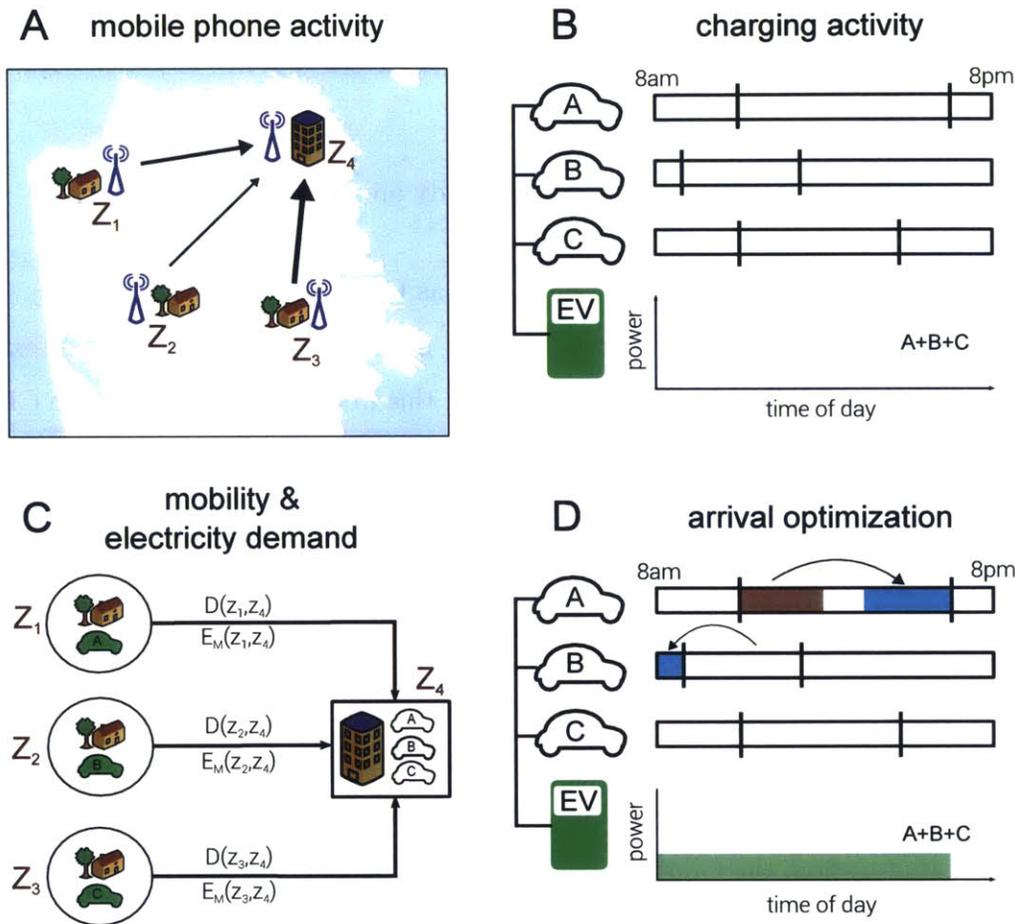


Figure 5-1: Coupling PEV charging with urban mobility. (a) Mobile phone trajectories are used to estimate mobility patterns. (b) Charging sessions used to characterize session and electricity demand curves. (c) These findings are combined to analyze the relationship between commuting patterns and electricity demand. (d) Charging activity is shifted to relieve peaks in demand and generate savings.

tween the travel distance of PEV commuters incoming to a charging station and the observed electricity demand at individual charging stations. We discuss how mobility patterns can provide insights towards a region's characteristic PEV based electricity consumption profile. In the fourth and final part, we lay out a charging optimization scheme that adjusts arrival times or delays charging activity to temporally distribute power consumption and mitigate the stress on the grid. We estimate the resulting effects on the commuting travel times and the monetary benefits to assess the value of such optimization solutions and their viability.

5.2 Methods

5.2.1 Data

The three main sources of data used in this study are described below.

1. **Mobile Phone Activity:** Also referred to as Call Detail Records (CDRs), this data has been widely popular in the last decade, especially in the context of mobility modeling [28, 88, 3, 164, 45]. For this work, we make use of the CDRs for the Bay Area including approximately 430,000 users and about 429 million calls they made over 3 weeks. The spatial resolution is discretized to the service areas of 892 distinct cell towers. This information is used to estimate the travel demand for the Bay Area for a typical weekday.
2. **PEV Charging Sessions:** This data contains 580,000 records of PEV charging sessions in commercial PEV supply equipment (EVSE) locations across the Bay Area in 2013, including any vehicle with a battery that can be charged. For each charging session, the following information is available: (i) one-time information on the EVSE location type, unique driver ID, total energy transferred, and plug-in/plug-out times; and (ii) charging power readings obtained every 15 minutes. The locations of the charging stations are anonymized to zipcode level. As a preprocessing step, we filter out records lasting shorter than 1 minute, are not in 2013, or have erroneous power measurements exceeding typical cable capacity and maximum charging rates.
3. **Census and Survey Information:** The census data used in this study consists of shapefiles describing zip code regions, their population, and income information. The survey information is obtained from the California Plug-in Electric Vehicle Driver Survey carried out in 2013 [62]. This survey contains information on various sociodemographic characteristics and travel behavior of PEV drivers in California. We utilize information regarding income and average daily vehicle miles travelled in the estimation of PEV mobility.

Table 5.1: Characteristics of PEV drivers. Distribution of (a) average daily miles driven and (b) annual income by PEV drivers in California, USA [62].

(1000\$)	Conventional	PEV	(miles)	%
Unknown	20%	17%	< 15	14%
< 50	20%	2%	15-30	50%
50-100	30%	13%	30-45	28%
100-150	14%	20%	> 45	8%
> 250	15%	47%		

5.2.2 Electric Vehicle OD Estimation

We denote an OD-pair, a trip between a source od_s and a target od_t , as od . We denote EV as the random variable that represents the occurrence of the event that a car trip is made by a plug-in electric vehicle. I_{od} is the random variable that denotes the income of the trip maker, and $P(I_{od})$ follows a standard normal distribution centered at median income of the source zipcode, od_s . D_{od} is the random variable that denotes route distance and is equal to $D(od_s, od_t)$ between od_s and od_t , constant for the specific od . Our goal is to estimate $P(EV | I_{od}, D_{od})$ for each od , or in other words, the probability that a trip for a given OD pair is in fact made by a PEV. We assume that for a given od , I_{od} and P_{od} are independent, and $P(I_{od}, D_{od} | EV) = P(I_{od} | EV)P(D_{od} | EV)$, that is, I_{od} and P_{od} are also conditionally independent given PEV.

We begin by expressing the Bayesian relation,

$$P(EV | I_{od}, D_{od}) = \frac{P(I_{od}, D_{od} | EV)P(EV)}{P(I_{od}, D_{od})} \quad (5.1)$$

By imposing our aforementioned assumptions on Eq.5.1, we obtain

$$P(EV | I_{od}, D_{od}) = \frac{P(I_{od} | EV)P(D_{od} | EV)P(EV)}{P(I_{od})P(D_{od})} \quad (5.2)$$

In estimating this value, we assume $P(EV) = 0.62\%$ as the share of PEVs within all cars in the Bay Area. We make use of the PEV driver survey information regarding income and distance travelled, namely $P(I_{od} | EV)$ and $P(D_{od} | EV)P(EV)$,

respectively. We randomly assign an I_{od} to each od from $P(I_{od})$, and calculate the D_{od} by using a publicly available online API service for routing. Given that D_{od} is constant for all od , $P(D_{od}) = 1$. Once $P(EV | I_{od}, D_{od})$ is estimated, the probabilities are used to reweight the flow of each unique (z_s, z_t) pair. Figure 5-2 represents the D_{od} distribution of the posterior $P(D_{od} | EV)$.

5.2.3 Optimization Model

We begin by discretizing a day into 15-minute intervals such that each day starts at $t = 0$ and ends at $t = 95$ [89]. For each charging session i among N in a day at a charging station, we define t_a^i as the arrival time index, t_c^i as the time index where charging is complete, and t_d^i as the departure time index. We represent the time indices by the vector τ^i , and the power consumption by vectors \mathbf{P}^i and \mathbf{Q}^i , all defined as follows:

$$\begin{aligned}\tau^i &= [t_a^i, \dots, t_c^i]^\top \\ \mathbf{P}^i &= [P_0^i, \dots, P_{95}^i]^\top \\ \mathbf{Q}^i &= [P_{t_a^i}^i, \dots, P_{t_c^i}^i]^\top\end{aligned}\tag{5.3}$$

By shifting \mathbf{Q}^i within \mathbf{P}^i by an amount d^i for all sessions, we can modify the overall power demand curve. We define $M^i = (t_c^i - t_a^i) + 1$ as the total number of non-zero power measurements in this charging session (i.e. total number of elements in \mathbf{Q}^i), given that charging sessions start immediately upon arrival. We enforce continuity of the charging process, the non-violation of departure times, and amounts of session energy.

To capture the constraints proposed above, we introduce the following formal constraints:

$$\left. \begin{aligned}
\tau_j^i &\geq 0 \\
\tau_j^i &\leq 95 \\
\tau_j^i &\geq t_a^i + d^i \\
\tau_j^i &\leq t_c^i + d^i \\
t_d^i &\geq t_c^i + d^i \\
\tau_j^i &< \tau_{j+1}^i
\end{aligned} \right\} \begin{aligned}
&\forall i \in [1, N], \\
&\forall j \in [1, M^i]
\end{aligned} \quad (5.4)$$

We construct the proposed constraints using a binary decision matrix to represent charging or non-charging time slots within the optimization duration. To represent the candidate time slot at which Q_j^i can be positioned, we create binary row vectors \mathbf{x}_j^i each consisting of 95 binary decision variables: $x_{j,k}^i \in \{0, 1\}, \forall j \in [1, M^i], \forall i \in [1, N], \forall k \in [0, 95]$.

$$\mathbf{X}^i = \begin{bmatrix} \mathbf{x}_1^i \\ \vdots \\ \mathbf{x}_{M^i}^i \end{bmatrix} = \begin{bmatrix} x_{1,0}^i & \dots & x_{1,95}^i \\ \vdots & \ddots & \vdots \\ x_{M^i,0}^i & \dots & x_{M^i,95}^i \end{bmatrix} \quad (5.5)$$

Finally, we write the variables in the constraints given in (5.4) using the binary decision variable as follows:

$$\tau^i = \mathbf{X}^i \begin{bmatrix} 0 \\ \vdots \\ 95 \end{bmatrix} \quad (5.6)$$

The aggregate power vector \mathbf{AP} is given as follows:

$$\mathbf{AP} = \sum_{i=0}^N \mathbf{P}^i = \begin{bmatrix} Q^1 \\ \vdots \\ Q^N \end{bmatrix}^T \begin{bmatrix} \mathbf{X}^1 \\ \vdots \\ \mathbf{X}^N \end{bmatrix} \quad (5.7)$$

The resulting formulation is a mixed-integer linear program, with decision variables \mathbf{X}^i , P_{peak} , and d^i of which the latter two are integers. The problem can be proposed to minimize the daily peak load P_{peak} for a group of PEVs arriving to the

same zip code location:

$$\underset{X^i, P_{peak}, d^i}{\text{minimize}} P_{peak}$$

subject to (5.4) and the following additional constraints:

$$AP_t^i \leq P_{peak}, \quad \forall i \in [1, N], \quad \forall t \in [0, 95] \quad (5.8)$$

5.3 Results

5.3.1 Estimating Electric Vehicle Mobility

We begin by estimating the overall vehicular mobility of the Bay Area by following the methodology outlined in [164, 45, 3]. In doing so, we make use of the mobile phone logs of a large subsample of the population (see Methods). This process begins with the extraction of stay locations by cleaning the noise in the trajectories of each individual[88]. Each location is then labeled accordingly based on temporal properties of the call activity. Once home and work locations are successfully identified, sequences of trips are collected for each individual and categorized by their time and purpose as well origin and destination. When carried out for the whole sample, this process captures the observed mobility of the mobile phone users. Then, from the samples obtained for each region, the mobility patterns of the whole population in that region are estimated, in consideration of the ratio between the sample and the region population. The result is what is known as the origin-destination information (ODs): the number of car trips taken from and to different points in the area.

The next and key challenge in this section is the conversion of this OD information, representing the mobility of all vehicles, to that of PEVs. In order to statistically capture PEV drivers within all car drivers, we make use of the California Plug-in Electric Vehicle Driver Survey carried out in 2013[62]. This survey, while highlighting the increase in PEV adoption, presents various sociodemographic characteristics of PEV owners. One of the more noticeable results in this context is the comparison of the

household income distribution of PEV drivers compared to that of conventional car owners. PEV drivers' income distribution is skewed towards higher income segments, unlike conventional vehicle owners where the income distribution is relatively uniform. In particular, the percentage of those with average annual income above 150K\$ among conventional vehicle drivers is 15%, compared to the 47% observed among PEV drivers. The survey also highlights the typical distances PEV drivers travel: 64% of PEV drivers travel less than 30 miles per day (Table 1). Although newer generations of electric vehicles have increasingly higher ranges, PEV drivers typically have low daily travel distances. We make use of these survey findings to accurately subsample PEV trips from total vehicular ODs by implementing a Bayesian sampling procedure (see Methods). In this methodology, we use income distributions at the zipcode level and typical route distance of each individual origin-destination pair to estimate the probability of that trip being made by an PEV. Figure 5-2 summarizes our findings by comparing the distributions of route distance, D , and the commuting travel time under free flow conditions, T , for PEV trips to those of all vehicle trips. The applied methodology moderates the distribution of trip distances for PEVs, more visibly for trips shorter than 5 kms and longer than 30 kms, in agreement with the findings of the survey. The observed bimodality of the distance distribution is an outcome of the commuting patterns and the income distribution across the region.

5.3.2 Electric Vehicle Charging Session Profiles

In this section, we comprehensively analyze PEV charging by examining various aspects such as visitation patterns and adoption rates, temporal qualities of arrivals and departures, and typical energy and power consumption levels. PEV drivers display varying degrees of regularity in terms of how often they visit charging stations. Figure 5-3(a) reveals that for the majority of PEV drivers, the number of sessions per day, N_{day} , beginning from the day of their first record, is less than 1. The bottom left inset of Figure 5-3(a) displays the logarithmic distribution of the number unique PEV charging stations (EVSEs) visited, N_{EVSE} , by each PEV driver: 95.6% of PEV drivers have at least one charging activity in less than 20 distinct EVSEs. The top

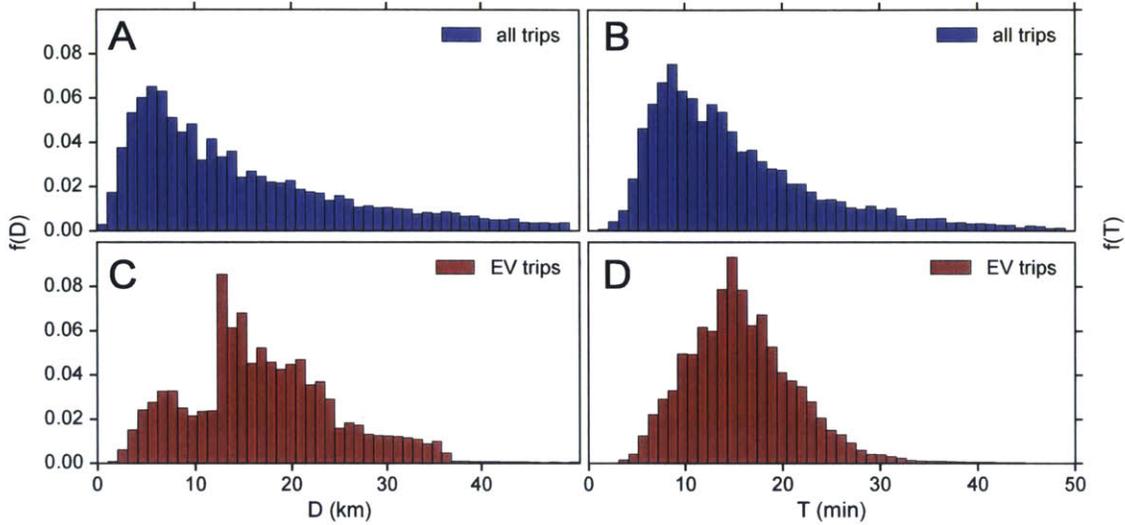


Figure 5-2: OD estimation for PEV drivers in the Bay Area. The probability distributions of commuting distances, D , and commuting travel times, T , of all vehicle trips and PEV trips estimated through income information and trip distances.

right inset of Figure 5-3(a) depicts the rate of PEV adoption observed throughout the year. The 3000 drivers observed in January 2013 increases by an average of 1000 per month, doubling twice over the course of 2013. The implication of this rate of adoption for the power grid is in fact more severe than at first glance, since the new PEVs entering traffic will not uniformly distribute the demand across the region but rather further increase the spatial concentration. This will result in a superlinear scaling of energy demand to number of PEVs in popular regions, reiterating the pressing need to understand electricity demand resulting from mobility needs and the development of efficient charging solutions.

Next, we look at the arrival and departure hours of charging sessions, h_a and h_d , in Figure 5-3(b). Approximately 50% of all arrivals take place in the 6am-11am morning period, and as expected, the morning and the evening peaks are highly pronounced. This points to the parallels between the temporal component of overall travel demand to electricity demand. The morning and evening demand peaks are abundant in all travel demand models, and this observation indicates that the temporal component of charging demand is directly translated from that of travel demand. To go into further detail regarding of arrivals and departures, in the inset of Figure 5-3(b) we

look at the distribution of interarrival and interdeparture times, Δh_a and Δh_d , i.e. the time between two consecutive charging sessions for the same driver ID. These distributions are peaked at multiples of 24 hours, pointing to the diurnal periodicity of PEV drivers' charging behavior. In conjunction with findings regarding arrival and departure times and visitation patterns, this finding reinforces the notion that commuting and charging behavior are highly related.

Next, we shift our focus to per session measures such as energy, duration, and power. Figure 5-3(c) exhibits the average energy consumption per session, E_S . The battery sizes of Nissan Leaf (24 kWh) and Chevrolet Volt (16 kWh), two of the most commonly used PEVs in the region are marked. Typically E_S are well below these capacities, indicating that PEV drivers typically stay within the range of their PEVs. The charging activity typically does not fill emptied batteries, in contrast to what range anxiety suggests. In fact, at these levels of flexibility, PEV drivers appear considerably free to choose whether to charge their vehicles at home or not, without having to fear not being able to complete their commute the next day. In line with this, the incentives in place that promote workplace charging [124] blur the one-to-one mapping between a single commuting trip and the electricity demand in the charging session that follows, as they enable PEV drivers to not necessarily start their commute at full capacity. On the other hand the distribution of session durations reveals that 98.4% of all charging sessions last less than a day (Figure 5-3(c)), in line with the strong ties to commuting previously mentioned. When findings regarding commuting-like temporal behavior and flexibility in terms of battery capacity are considered in conjunction, it is reasonable to expect that the session energy E_S will likely represent not a single commuting trip, but rather a number of them.

Since the actual charging activity does not last as long as the session duration δ_S , in Figure 5-3(d) we look at how the power changes as the session continues. We divide sessions into four categories based on their session duration, and plot the average power consumption for each segment at various percentages of the total duration. At this point, it should be mentioned that there are three power rating levels observed, namely Level 1 (L1), Level 2 (L2) and Level 3 (L3). The first two deliver 120V

and 240V, typically corresponding to 3.3kW and 6.6kW, respectively. L3 chargers are mainly for fast charging at 480V and are relatively uncommon. In fact, L1 and L2 chargers make up 99.9% of all the sessions in the dataset. This composition of power ratings explains the 4 kW upper limit to average power consumption observed in Figure 5-3(d). For sessions lasting less than 4 hours, average power stays above 3 kW up to 80% of the duration into the session. Conversely, for sessions that last longer than 12 hours but less than a day, only in the starting 25% of the session duration there is active charging. This corresponds approximately to 3-6 hours, and the power remains zero thereafter. This is consistent with constant-current constant-voltage battery charging behavior and it suggests that currently there is no strategy to charging involved: charging begins as soon as a session begins and ends when the battery is full. Despite their flexibility in terms of session durations and battery capacity, PEVs are charged immediately upon arrival.

5.3.3 Energy and Travel Demand Relationship

As we established in the previous section, there currently is no strategy in place to better manage the load PEVs impose on the power grid. In order to develop effective strategies for this purpose, understanding the unique electricity demand of each region is a necessity. In estimating a single PEV's electricity demand, it is possible to make use of the fact that the charging session energy demand is an outcome of the trips that PEV has previously taken. Similarly, at the spatial resolution of a zipcode, the mobility patterns of commuters to that zipcode can be used to profile the average electricity demand at that zipcode. Following this reasoning, our goal in this section is to analyze the extent to which aggregate mobility patterns can be used to estimate average energy demand of a region. For this purpose, we look at how the average electricity demand at a zipcode, measured by using charging sessions, is related to the average commuting distance to that zipcode inferred from our EV mobility estimates.

We begin by analyzing d , T , T_c , and E_S , representing trip distance, trip travel time under free flow conditions, trip travel time in traffic, and session energy, respectively. The probability distributions of these metrics are depicted in Figure 5-4(a), with the

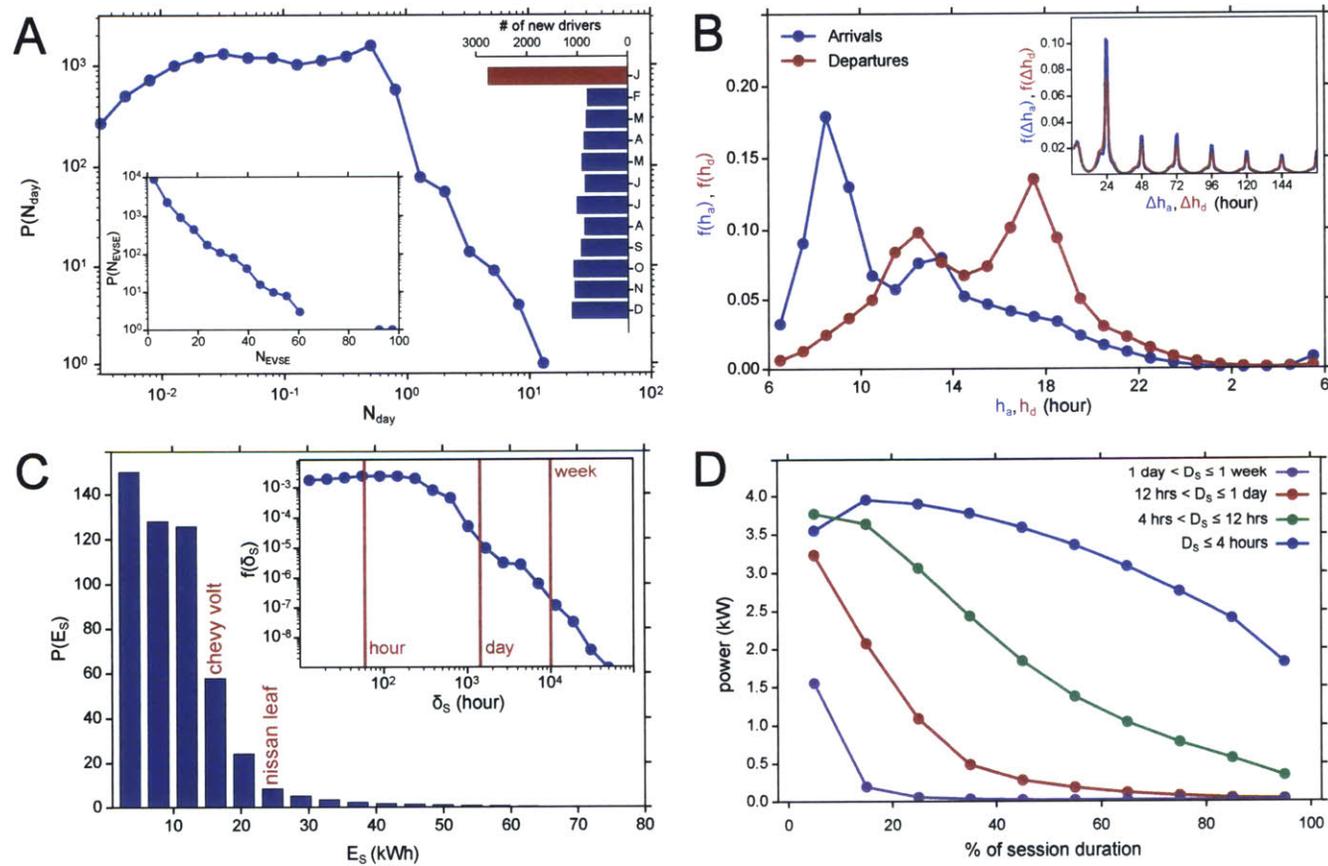


Figure 5-3: PEV charging session profiles. (a) Distribution of N_{day} the number of sessions per day for each driver ID starting from the day of first record (right inset: number of new driver IDs added every month, left inset: distribution of N_{EVSE} , the number of unique EVSEs visited by every driver ID.) (b) The distributions of h_a and h_d , the arrival and departure hours to and from an EVSE. (inset: the distributions of Δh_a and Δh_d , the interarrival and interdeparture times for a driver ID visiting a specific EVSE.) (c) The distribution of E_s , the total energy withdrawn per session (inset: the distribution of δ_s , session durations.) (d) The power consumption as a function of the normalized session duration segmented by total duration groups.

cumulative distributions in the inset. The peaks in the distribution of E_S demonstrate the heterogeneity in the electricity demand, as well as the battery capacities of various PEVs. The 3-4 kWh peak is a combination of low energy demand as a consequence of short commuting trips and plug-in hybrid electric vehicles (PHEVs) that typically have a battery capacity around 4 kWh [182]. The similarity of these distributions are in agreement with the known fact that the energy consumption of a single PEV is a direct outcome of the characteristics of the preceding trip such as distance and speed. More interestingly however, it can be observed that this relationship is conserved for a population of PEVs rather than at the individual level. In other words, a set of the trips represented by the OD information of PEVs and the resulting distribution of energy consumption of the same population of PEVs follow similar distributions. This finding signifies the potential in combining mobility and energy information to provide insights at the urban scale.

Next, we further explore the relationship between energy demand and travel demand. For this purpose, we aggregate the session energy readings of the charging sessions from each zipcode to obtain E_S^z for that zipcode. Similarly, we average the distance traveled by all morning commuters to that zipcode to find D^z . In order to better assess our findings, we refer to models in the literature that estimate the electricity consumption of a single PEV trip given speed and distance information of that trip, generally referred to as drivetrain models. To serve as a benchmark, we implement the drivetrain model presented in [143]. We denote this estimate as E_M , and we aggregate it for each zipcode to obtain E_M^z . To select Nissan Leafs in our charging session, we select the sessions of only the vehicles that have not had any charging sessions with total energy readings that exceed the nameplate battery capacity of a Nissan Leaf. We compare how both E_S^z and E_M^z change with respect to D^z in Figure 5-4(b). E_S^z exhibits a relatively low slope with a positive energy intercept. In other words, PEVs appear to have a fixed energy demand even for $D^z = 0$. In comparison to E_S^z readings around 9kWh, E_M^z stay within the range of 2.5 kWh to 3.5 kWh. Although the overall trends are in agreement, the scales of E_S^z and E_M^z are different. These results stem from findings from previous sections. We previously es-

established that within the commuting distance ranges we observe, PEV drivers do not always start their day at full battery capacity. This notion is strengthened by heavy subsidies supporting workplace charging. A recent report by Department of Energy suggests that 80% of partner workplaces in their Workplace Charging Challenge program provide free PEV charging, compared to 20% who charge their employees a fee [124]. In support of this finding, it is widely known that the energy added to the PEV battery and the energy for the EVSE meter often differ in measurements around 10-15%. The commonality for these factors is that they are uniform across vehicles and stations, hence are constant biases throughout. This explains the constant of proportionality between E_S^z and E_M^z , energy values obtained from the sessions and the model. To account for this proportionality, we analyze the two models by their first differences. By analyzing the additional session electricity demand for the additional average commuter trip distance to a zipcode, we are able to more robustly assess the quality of mobility information as an estimator of energy demand at a zipcode. For this purpose, we sort the zipcodes by the average incoming commuter travel distance, and then estimate the change in energy demand, ΔE_S^z , for each consecutive change in sorted distances ΔD^z . We fit the following linear models $\Delta E_S^z = \alpha_S \Delta D^z + \beta_S + \epsilon_S$ for sessions and $\Delta E_M^z = \alpha_M \Delta D^z + \beta_M + \epsilon_M$ for the model using least squares, where α , β and ϵ denote the slope, the intercept and the error term, respectively. As can be observed from the inset of Figure 5-4(b), The slope obtained from the model is in agreement with that obtained from charging sessions. This indicates that OD information and energy demand resulting from commuting patterns are closely linked, and commuting patterns into a region can be used in laying the foundation to provide the electricity demand fingerprint of a region.

5.3.4 Assessment of the Impact of Charging Timeshifts

As mentioned, the spatial and temporal concentration makes electricity demand difficult to meet in certain regions at certain times of the day. This led to more dynamic and complex rate structures that not only bill customers for the total energy demanded but also incorporate the maximum power demand, time of day, and season

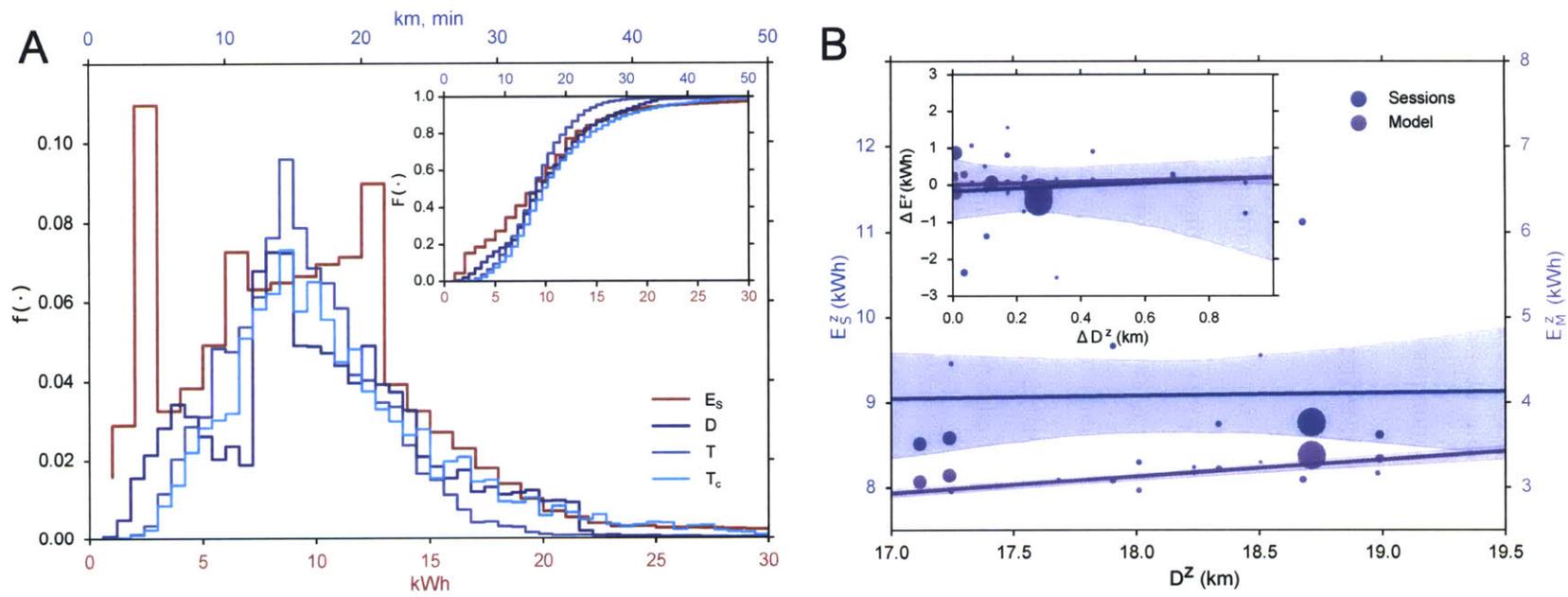


Figure 5-4: Relationship between OD trips and energy consumption. (a) The probability distributions of session energy E_S obtained from charging sessions compared to those of distance, D , free commuting travel time T and commuting travel time in traffic, T_c , estimated from ODs. (inset: the cumulative distributions) (b) Linear regressions of the average distance of Nissan Leaf commuting trips to a zipcode, D^z , versus the average morning session energy for these PEVs for that zipcode, E_S^z and E_M^z . The bands represent the 95% confidence intervals, and point sizes are proportional to the number of sessions in that zipcode. (inset: a regression of the first differences of the energy and distance measurements)

of year. These pricing policies counteract consumption behavior that pushes the power load curve towards nonuniformity. However in a future where the number of PEVs will be magnitudes more than today, achieving a temporally homogeneous load curve at the urban scale will pose a tougher challenge. In this section, we propose a methodology that aims to mitigate the peak load imposed by PEVs on the grid by minimizing peak power at the EVSE level. By adjusting vehicle arrival times or delaying charging activity of various sessions, we move the aggregate charging activity towards hours of the day when there is less demand. This enables the transformation of the load curve into one that is more uniformly distributed across the day. Finally, we explore the monetary benefits of the potential savings and the implications of arrival time adjustment on commuting times to discuss the viability of this class of power management approaches.

To investigate the impact of arrival time scheduling and charging activity delay on the overall peak of the PEV charging demand, we cast problem as a mixed-integer linear program with discrete shifts in arrival times and charging end times as inputs. The program modifies the total power P_t measured through the day resulting from the overlapping charging activities of a population of PEVs in a way that minimizes the peak power while keeping the total energy consumed constant. That is, all PEVs are charged by the same amount of energy as they used to, the only difference being the temporal distribution of how that energy was transmitted. Moreover, the charging activity is never allowed to be interrupted, and the departure times are not modified (see Methods).

In this context, we test three different approaches. The first fixes the arrival times for PEVs and delays the charging by d^i , an amount specific to session i within the interval $[0, d]$. The drivers are free to arrive as they wish, however the PEV charging can be delayed to minimize P_{peak} . We refer to this model as *start bound*. The second model, referred to as *end bound*, offers modifications to the arrival times by proposing drivers to arrive earlier by d^i minutes in the interval $[-d, 0]$. In this approach, PEV charging begins immediately upon arrival. The start bound and end bound models represent a tradeoff: the latter generates savings at the expense of having at most

one hour worth of charging less capacity when in a need to leave earlier, whereas the former is at the expense of possibly more inconvenient arrival times. The third and the last model combines the first two models and enables the adjustment of both arrivals and charging activity, referred from here on as the *flexible* model. In this model, the charging activity is shifted in the interval $[-d, d]$. We implement these three models for a typical day at a zipcode that contains 493 sessions (see Methods).

Figure 5-5(a) illustrates an instance of the flexible model for $d = 2$ (30 minutes) and how a sample of charging sessions would have been modified. To minimize the peak power, morning sessions have their charging shifted to earlier, whereas the charging of afternoon sessions are deferred. To test our models, we selected a zipcode and all 439 sessions recorded in a specific weekday. Figure 5-5(b) demonstrates how aggregate power curves are modified under these models. The flexible model is able to push P_{peak} down 38% from 973 kW to approximately 600 kW for $d = 4$, or namely 1 hour. Start-bound and end-bound models, as expected, require higher values of d to achieve comparable savings. The wider domain achieved by the combination of arrival adjustment and delay of charging provides strong flexibility for the flexible model, in agreement with its name, enabling the attainment of more substantial savings.

One key reason why an PEV driver might not want to comply with an earlier arrival schedule would be its negative influence on the travel time. To assess how realizable the benefits of these models are, we investigate the consequences of the flexible model in terms of how it affects commuting travel times. In Figure 5-5(c), we look at how the proposed changes for varying values of d for the flexible model would affect the commuting travel times of all trips into the subject zipcode using the OD flows. The results show that the peak power reductions can be achieved without causing major discomfort to commuters in terms of travel times. The most negatively influenced drivers end up losing a maximum of 20 minutes in the case of $d = 4$ (1 hour), and are far less than those who are unaffected by the proposed changes. In fact, there is a substantial number of drivers that achieve travel time savings in efforts to minimize peak power.

Next, we evaluate the monetary outcomes of these models. As mentioned previ-

ously, rate structures have charges associated with peak power, referred to as demand charge. For our models, we use the E-19 rate structure for the region to calculate the change in demand charge as a proxy of the cost in terms of dollars [66, 89]. More specifically, we use the max part-peak demand summer rates. The peaks we observe fall in the shoulder period of 8:30am-12:00pm, for which the demand charge rate is 4.07\$/kW. This enables us to gauge the magnitude of power shaving in monetary terms. When implemented, the possible benefits of the schemes we proposed are displayed in Figure 5-5(d): monthly potential savings in the demand charge can reach up to 1500\$, which considering the 439 sessions in the subject weekday, correspond to roughly 3\$ per month per session. Without managing charging, these savings remain unrealized, and are paid by PEV drivers or the companies that subsidize the charging activity. As a sum the savings are substantial, yet for the number sessions on a typical weekday considered here, the amount per individual is relatively small, making uniform distribution of savings a relatively unexciting reward for cooperation. However, recent studies have suggested that gamified systems are successful in promoting behavior that help achieve social good [112]. More specifically, these systems encourage engagement by building raffles in which each participant has a chance to win a bigger reward with a probability proportional to their cooperation level. This type of mechanisms will make incentivization highly viable in the context of PEVs and managing their electricity demand.

5.4 Discussion

Rising CO_2 emission levels are increasingly threatening the moderate nature of earth's climate. In efforts to solely sustain the quality of urban life in today's cities, decreasing fossil fuel dependency is a must. In this context, technological innovations led to PEVs becoming economically and socially more viable everyday and being adopted by citizens at increasing rates. This paradigm shift in the interplay of electricity demand, mobility and environmental concerns creates an urgency to rethink PEV mobility. In order to devise schemes that optimize or control electricity demand, we

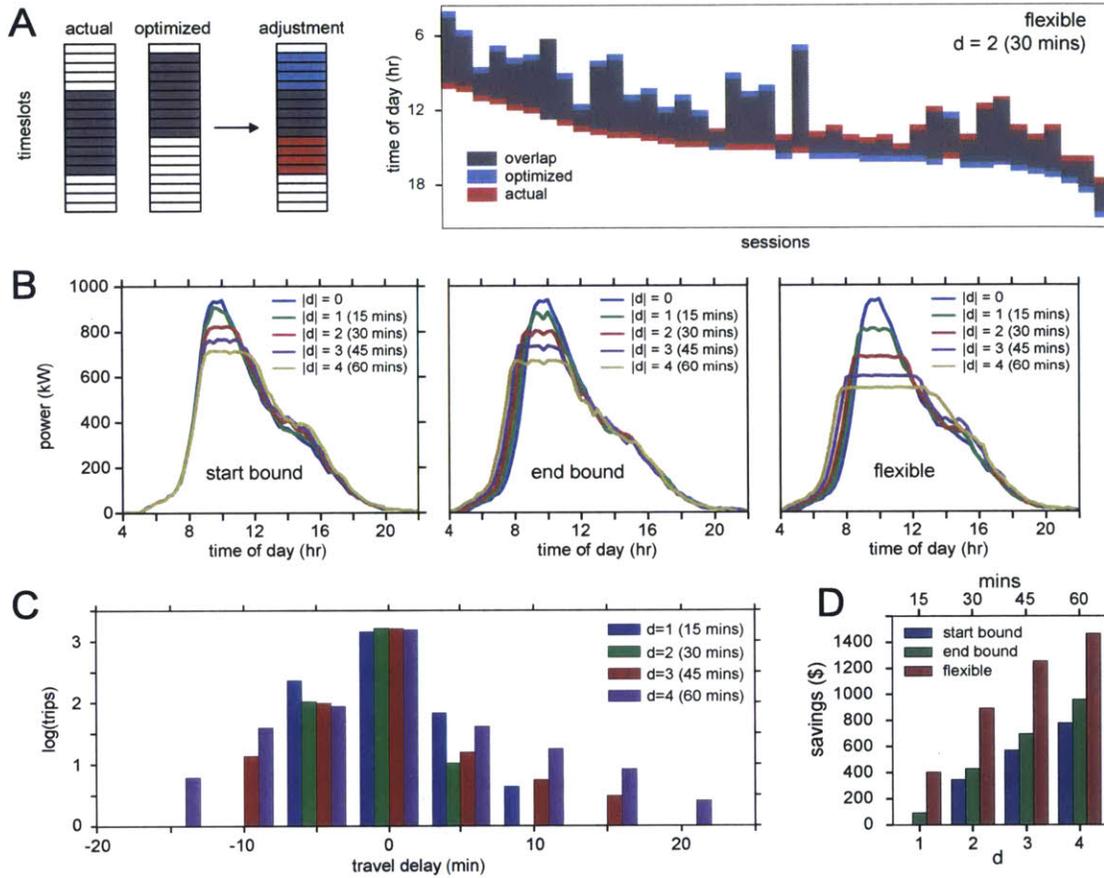


Figure 5-5: Assessing the benefits of minimizing peak power. (a) An illustration of a sample of sessions in the flexible model. The charging event within a session is shifted such that the overall peak power is minimized. (b) The decrease in peak power measurements for each model and varying d . Peak shaving of up to 38% can be achieved. (c) The influence of the proposed changes on travel times. Majority of drivers are not influenced, the worst case is a few individuals suffering from an additional 20 minute delay for $d = 4$. (d) The demand charge portion of the monthly bill and distribution of savings for varying values of d . The flexible model is able to generate savings worth approximately 1400\$ for $d = 4$.

need to focus on the methods that estimate PEV demand.

This work presents, to the best of our knowledge, the first exploratory analysis that couples two unique large datasets on urban mobility and electric vehicle energy consumption. We first present methods to estimate PEV mobility patterns using mobile phone records and appropriate sampling methods. In tandem, we dissect PEV charging sessions in the same area to look at the spatiotemporal properties of electricity demand. We observe that drivers visit few charging stations and charge their vehicles in diurnal periodic fashion, thus there is spatiotemporal regularity in PEV charging. Moreover, PEV charging typically begins immediately, and as expected, more often on peak hours. However, charging sessions demonstrate high temporal flexibility, in other words, the vehicles are parked for longer amounts of time than what is required to fully charge them. Therefore there is significant room for improvement in the scheduling of PEV charging. Upon combining our insights from the two datasets, we also observe that mobility is a precursor to electricity demand, and the two are closely interrelated.

Building on this, we provide a method to shave the daily peak power to alleviate the load on the power grid. We find that even with simple charging delay and arrival hour adjustments that do not impose any constraints on departure times and do not violate the charging continuity, peak power values can be shed by up to 40%. This class of solutions typically perform better with higher levels of participation from drivers. To incentivize cooperation, every driver needs to be presented with a balanced composition of benefits and costs, a key determinant of the success of these types of schemes. In an effort to further strengthen our analysis regarding the applicability of the proposed and similar solutions, we estimate the possible monetary benefits and the travel time losses resulting from the proposed schemes. Although the resulting savings are not large at the daily individual level, they are certainly substantial enough for implementation of gamification and similar reward based incentivization schemes to induce cooperation and raise awareness. On the other hand, the travel time losses are almost imperceptible to the majority of the drivers, and a substantial number of drivers in fact benefit from the adjustment of their arrival times as it aids them to

escape morning traffic. These findings demonstrate that in the current setting and medium-term future, energy management in the context of electric vehicles is highly viable.

There are various avenues in which this work can be extended. A meticulous methodology to more accurately estimate electric vehicle mobility, especially with higher temporal resolution, is of prime importance. In energy management, the dynamics of power demand plays a very important role, therefore obtaining dynamic ODs for electric vehicles is a necessary next step. For this purpose, collecting and utilizing the data generated by PEVs are crucial. As drivetrain models are continually being improved, a stronger comprehension of the tie between mobility and energy demand at varying levels of spatial resolution is necessary to create bottom-up solutions. Finally, the efficient implementation of these energy management solutions in real time remains an open front.

Chapter 6

Conclusion

In this final chapter, we summarize the work and elaborate on possible avenues of future work in this line of research.

6.1 Summary

In overall, this dissertation presented methods to process large spatiotemporal datasets and extract aggregate mobility information, which in turn is used to generate knowledge regarding congestion from both topological and supply and demand point of view and power consumption along with analyses of potential benefits and their distribution under varying levels agent cooperation.

More specifically, in Chapter 1 we presented an overview of the literature regarding this dissertation. Human mobility research is new and novel, and is currently used to update older transportation models as well as forming the basis for building newer ones. Transportation modeling has achieved strong simulators built in leading research labs, and is updating itself to introduce newer data sources into traditional methodologies. Traffic flow is a relatively older line of research, revamped with improved GPS data and increasingly efficient map-matching algorithms. At the same time, the abundance of data has also pushed forward a newer science of cities to better understand their evolution and growth. The four chapters that followed lies in the intersection of these four distinct lines of research.

Chapter 2 delved into understanding the influence of network topology and space in the emergence of congestion. In uniting the knowledge generated in both physics and transportation literatures, we implemented a point queue model and analyze how cities get congested from a phase transition perspective. We showed that network topology and spatial constraints are determining factors in the nature of this transition. That is, the transition became a first order when there is lack of space, as spillbacks dominated the spread of congestion. The next step was to tackle the missing component that is the influence of realistic travel demand. We aimed to incorporate various routing schemes, and see how these results could be applied to other cities.

Chapter 3 was a step in this direction. We used large mobile phone datasets and provided a framework to extract origin-destination data from it. We demonstrated that by using the right methodology and algorithms on these mobile phone traces, it is possible to generate results similar to those obtained from models used by practitioners. Moreover, the procedures proved to be specifically more valuable for developing countries where mobile phone data is available, but travel surveys are low in quality, old, or sometimes even non-existent. To that end, our results were in good agreement with models used in industry, and were also well received in practice ready form. To build on this work, we aimed to use the generated realistic and validated origin-destination information to better understand the complexities of cities, specifically in the context of understanding travel times and congestion. We aspired to carry the same portability of the presented system to understand congested urban travel, with the longer term goal of comparing different cities and analyzing routing solutions.

In Chapter 4, we used the demand information in a generalized routing model that takes into account socially aware routing behavior. We showed that a city's congestion fingerprint is related to measurable characteristics, namely a ratio of total travel demand to total supply. We demonstrated that under a centralized and socially optimal routing scheme, there were significant potential travel time savings for all cities. Moreover, low levels of social awareness in routing were enough to achieve

a substantial portion of these savings, as well as moderating benefits and losses or drivers, making proposed adjustments fairer and easier to implement. As for future work, our goal was to understand better the interplay of travel, benefits and optimization, but this time from a rather new perspective: electrification of transportation, or namely, electric vehicles.

In Chapter 5, we began by estimating the travel demand information for electric vehicle drivers in the Bay Area. By combining this information with that obtained from charging session data from 2013 in the same region, we showed that commuting distances to a location are good indicators of energy demand at that location, hence can be helpful in providing solutions to ease the load on the power grid. We showed the potential benefits that can be achieved by a generalized schema that can optionally implement smart charging and arrival hour adjustment, given proper incentivization.

6.2 Future Work

6.2.1 Analytics Engines for Mobility

An overarching theme in this dissertation has been data. With technological innovation and the significant decrease in hardware and computational costs, we find ourselves immersed deeper in the age of internet of things. Smart phones, smart cars, smart thermostats and similar devices will soon be able to not only generate the wealth of data they currently do in higher precision and volume, but will also be able to communicate with each other.

Ideally, we expect to use data to understand phenomena in ways we previously did not. However data in its raw form is far from being useful. There is substantial know-how associated with building efficient pipelines for data processing. Extraction of relevant information from the data extracted from such pipelines requires a different set of skills. After pipelines are built and data is processed, it is still has not realized its potential to the fullest: A platform has to exist for people who have the expertise of knowing which questions to ask the data.

Chapter 3 demonstrates one such platform that takes in raw mobile phone data, parses, cleans, processes it and presents it on a browser. It is portable and scalable, therefore it is possible to include various data sources and generate the same information for a new city very quickly. The resulting browser interaction is able to answer questions like "Which roads do people departing from Boston Common use?", or conversely, "Where are the origins of the trips that use Memorial Drive?", among others. Future work in this avenue is building better and more efficient engines that can convert raw data to relevant information. The challenge here lies in bringing together the various skills necessary to create this framework.

6.2.2 Data-Driven Policies

Transportation policies typically have very complex consequences often difficult to foresee. There are many parties involved, including state, city and local departments of transportation, metropolitan planning organizations, transit agencies, investors, construction companies, worker syndicates, and citizens. Let's consider the case when the local department of transportation shuts down an avenue for, say, renovation. First, the project has to be approved. Environmental factors need to be considered, reports on impact assessment need to be prepared, read, negotiated and agreed on. Budgeting has to be prepared to forecast costs as accurately as possible while spending as little taxpayer money as possible. In doing so, the rights of the workers and demands from syndicates have to be taken into account. Solutions regarding traffic redirection have to be proposed so that citizens are minimally affected. Stores on that road are affected and their owners might need to be compensated. All these dimensions that need attention make policy making a difficult task.

When it comes to making a policy decision in a complex setting, and in the case of transportation this can be a multi billion dollar infrastructure project, all plans, estimates, projections, or shortly the information that feeds the decision, needs to make use of all data available. This is where the ideas proposed in the previous section can come into play: with analytics engines that can provide answers with a few clicks, new bus lines can be introduced after a more thorough analysis of tap-in

tap-out data. Road closures and traffic management under extreme events can be carried out by revisiting the data from similar cases in the past. Even for relatively minor decisions such as when a planning agency is considering providing covered bus stops so riders can avoid getting soaked, the information of how many person-minutes have been spent under the rain at each bus stop will be obtained within seconds.

Scientists and engineers are getting increasingly more adept at building the tools to answer such questions. In the context of transportation, various companies and research groups are working on building analytics engines for mobility and transportation that can help policy makers make more informed and better targeted decisions. The next step is the permeation of these tools into the political process.

6.2.3 Mobility and Sociodemographics

Briefly before the 2013 Confederations Cup and the 2014 World Cup, bus, train and metro prices were increased in various cities in Brazil, mostly around 10%. This was a tipping point for most Brazilians already upset about the spending for mega projects for sport organizations soon to be carried in the country. Over the course of two months, more than three hundred thousand residents of Rio de Janeiro and two million Brazilians in various cities flocked to the streets to protest. The protesters not only requested the reversal of the price hikes for public transportation, but also reforms regarding education funding, fiscal responsibility, and corruption. The results were astounding, as in a few months the government complied with most of these demands.

It is no coincidence that the last straw was increasing the bus fares: millions of people depend on transportation systems on a daily basis. This points to a need to understand better how sociodemographic characteristics such as income or education are intertwined with ease of mobility. It can be argued that lower income segments of the population are pushed further outside city centers, yet are cornered into commuting into the city since that is where most job opportunities are. Commutes plagued with congestion into an increasingly far downtown not only decrease the overall quality of life for citizens, but also has long lasting detrimental effects on

social mobility.

In this context, the tools that we presented in Chapter 3 form a basis for furthering the understanding of the relationship between income, education and mobility. When combined with points highlighted in the previous two sections, namely building analytics engines for mobility and using them to generate data-driven policies, a bigger picture emerges: we can build tools that use data to inform decision makers and lead them to build appropriate policies that will correct for unjust sociodemographic biases such as income inequality or residential segregation.

6.2.4 Incentivization for Overall Benefit

Sections 4 and 5 of this work made analogous arguments: if controlled, there is a substantial amount of savings to be made from our infrastructure systems, namely the road networks and the power grid. More specifically, Section 4 argued that an overall decrease in congestion and thus travel times can be achieved, if few individuals sacrifice their own travel time. Similarly, Section 5 demonstrated that in case some users are willing to leave their home early to arrive to work early, the peak power can be more efficiently shaved to ease the load on the power grid. However, to reach optimality or even only to generate more efficiency, we need people to adjust their behavior, and this is not an easy problem.

This problem is by no means a new one, economists studied the concepts of individual utility, social welfare, behavioral economics for decades, often through game theory and mechanisms. Various implementations of gamification and lotteries were proposed in efforts to generate tools that modify behavior. The theories developed look more plausible every passing day. With smartphones and easier communication, these programs will require less effort from the user. In addition, automation will in fact will create another paradigm shift: people will have more time and will not be required to put in extra effort to accept sacrifices. For example, a traveler will board his/her self-driving electric car to go from A to B. The routing engine will calculate a route as well as the burden imposed on the society for that route and estimate the energy demand at the destination charging station and how it will influence the power

load at B at the time of arrival. From the traveler's perspective, the perception is a screen in a machine that takes you places. The traveler will be asked the question if they would be willing to accept a rebate for a cup of coffee in exchange for 5 minutes of extra travel time a slightly longer route takes. The lack of effort on the traveler's part makes incentivization not only significantly easier to implement, but also makes it highly profitable as many avenues for advertising can be made use of.

This example is not that far from reality: Google's self-driving electric cars are expected to be on the roads by 2020 and many other similar companies like Apple, Tesla or Uber are reportedly not that far behind. Real time routing engines are readily available and have improved to a point where they can dynamically provide you a faster alternative route along a trip. Considering this setting, such complex incentivization programs will form the basis of increasing the operational efficiency of our already burdened infrastructures and overall quality of life for citizens.

Bibliography

- [1] Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1 edition, February 1993.
- [2] Rahmi Akcelik. Travel time functions for transport planning purposes: Davidson's function, its time dependent form and alternative travel time function. *Australian Road Research*, 21(3), 1991.
- [3] Lauren Alexander, Shan Jiang, Mikel Murga, and Marta C González. Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58:240–250, 2015.
- [4] Mahnoosh Alizadeh, Hoi-To Wai, Mainak Chowdhury, Andrea Goldsmith, Anna Scaglione, and Tara Javidi. Joint management of electric vehicles in coupled power and transportation networks. *arXiv preprint arXiv:1511.03611*, 2015.
- [5] Elsa Arcaute, Erez Hatna, Peter Ferguson, Hyejin Youn, Anders Johansson, and Michael Batty. Constructing cities, deconstructing scaling laws. *Journal of The Royal Society Interface*, 12(102):20140745, 2015.
- [6] Alex Arenas, Albert Díaz-Guilera, and Roger Guimerà. Communication in networks with hierarchical branching. *Phys. Rev. Lett.*, 86(14):3196–3199, 2001.
- [7] Yasuo Asakura and Eiji Hato. Tracking survey for individual travel behaviour using mobile communication instruments. *Transportation Research Part C: Emerging Technologies*, 12(3):273–291, 2004.
- [8] Andrea Asztalos, Samcet Sreenivasan, Boleslaw K Szymanski, and G Korniss. Distributed flow optimization and cascading effects in weighted complex networks. *Eur. Phys. J. B*, 85(8):1–10, 2012.
- [9] M Balmer, K Meister, M Rieser, K Nagel, Kay W Axhausen, Kay W Axhausen, and Kay W Axhausen. *Agent-based simulation of travel demand: Structure and computational performance of MATSim-T*. ETH, Eidgenössische Technische Hochschule Zürich, IVT Institut für Verkehrsplanung und Transportsysteme, 2008.
- [10] Jayanth R Banavar, Amos Maritan, and Andrea Rinaldo. Size and form in efficient transportation networks. *Nature*, 399(6732):130–132, 1999.

- [11] Hillel Bar-Gera. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from israel. *Transportation Research Part C: Emerging Technologies*, 15(6):380–391, 2007.
- [12] Jaume Barceló et al. *Fundamentals of traffic simulation*, volume 145. Springer, 2010.
- [13] Marc Barthélemy. Betweenness centrality in large complex networks. *Eur. Phys. J. B*, 38(2):163–168, 2004.
- [14] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.
- [15] Marc Barthélemy and Alessandro Flammini. Modeling urban street patterns. *Phys. Rev. Lett.*, 100(13):138702, 2008.
- [16] Marc Barthélemy and Alessandro Flammini. Co-evolution of density and topology in a simple model of city formation. *Networks and spatial economics*, 9(3):401–425, 2009.
- [17] Michael Batty. The size, scale, and shape of cities. *Science*, 319(5864):769–771, 2008.
- [18] I Safak Bayram, George Michailidis, Michael Devetsikiotis, Fabrizio Granelli, and Subhashish Bhattacharya. Smart vehicles in the smart grid: Challenges, trends, and application to the design of charging stations. In *Control and Optimization Methods for Electric Smart Grids*, pages 133–145. Springer, 2012.
- [19] MJ Beckmann, CB Mc Guire, and CB Weinstein. *Studies in the Economics of Transportation*. Yale University Press, 1956.
- [20] Vitaly Belik, Theo Geisel, and Dirk Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X*, 1(1):011001, 2011.
- [21] Michael GH Bell. The estimation of origin-destination matrices by constrained generalised least squares. *Transportation Research Part B: Methodological*, 25(1):13–22, 1991.
- [22] Moshe Ben-Akiva, Haris N Koutsopoulos, Constantinos Antoniou, and Ramachandran Balakrishna. Traffic simulation with dynamit. In *Fundamentals of traffic simulation*, pages 363–398. Springer, 2010.
- [23] Moshe E Ben-Akiva and Steven R Lerman. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.
- [24] Alain Bertaud. The spatial organization of cities. *Deliberate Outcome or Unforeseen Consequence, Background Paper to World Development Report*, 2003.
- [25] Luis M. A. Bettencourt. The origins of scaling in cities. *Science*, 340(6139):1438–1441, 2013.

- [26] Luís MA Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States*, 104(17):7301–7306, 2007.
- [27] Sebastien Blandin, Dan Work, Paola Goatin, Benedetto Piccoli, and Alexandre Bayen. A general phase transition model for vehicular traffic. *SIAM journal on Applied Mathematics*, 71(1):107–127, 2011.
- [28] Vincent D Blondel, Adeline Decuyper, and Gautier Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4:10, 2015.
- [29] David E Boyce, Hani S Mahmassani, and Anna Nagurney. A retrospective on beckmann, mcguire and winsten’s studies in the economics of transportation. *Papers in Regional Science*, 84(1):85–103, 2005.
- [30] Dietrich Braess, Anna Nagurney, and Tina Wakolbinger. On a paradox of traffic planning. *Transportation Science*, 39(4):446–450, 2005.
- [31] David Branston. Link capacity functions: A review. *Transportation Research*, 10(4):223–236, 1976.
- [32] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.
- [33] Charles D Brummitt, Raissa M D’Souza, and EA Leicht. Suppressing cascades of load in interdependent networks. *Proceedings of the National Academy of Sciences*, 109(12):E680–E689, 2012.
- [34] Charles D Brummitt, Paul DH Hines, Ian Dobson, Cristopher Moore, and Raissa M D’Souza. Transdisciplinary electric power grid science. *Proceedings of the National Academy of Sciences*, 110(30):12159–12159, 2013.
- [35] Sergey V Buldyrev, Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.
- [36] N Caceres, JP Wideberg, and FG Benitez. Deriving origin destination data from a mobile phone network. *Intelligent Transport Systems, IET*, 1(1):15–26, 2007.
- [37] Francesco Calabrese, Mi Diao, Giusy Di Lorenzo, Joseph Ferreira Jr, and Carlo Ratti. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies*, 26:301–313, 2013.
- [38] Duncan S Callaway and Ian A Hiskens. Achieving controllability of electric loads. *Proceedings of the IEEE*, 99(1):184–199, 2011.

- [39] Julián Candia, Marta C González, Pu Wang, Timothy Schoenharl, Greg Madey, and Albert-László Barabási. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224015, 2008.
- [40] S. Carmi, Z. Wu, S. Havlin, and H. E. Stanley. Transport in networks with multiple sources and sinks. *Europhys. Lett.*, 84(2):28005+, 2008.
- [41] Ennio Cascetta. Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research Part B: Methodological*, 18(4):289–299, 1984.
- [42] Po-An Chen and David Kempe. Altruism, selfishness, and spite in traffic routing. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 140–149. ACM, 2008.
- [43] Debashish Chowdhury, Ludger Santen, and Andreas Schadschneider. Statistical physics of vehicular traffic and some related systems. *Phys. Rep.*, 329(4):199–329, 2000.
- [44] Kristien Clement-Nyns, Edwin Haesen, and Johan Driesen. The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. *Power Systems, IEEE Transactions on*, 25(1):371–380, 2010.
- [45] Serdar Çolak, Lauren P Alexander, Bernardo G Alvim, Shomik R Mehndiratta, and Marta C González. Analyzing cell phone location data for urban travel. *Transportation Research Record: Journal of the Transportation Research Board*, 2526:126–135, 2015.
- [46] Serdar Colak, Emre C. Kara, Scott J Moura, and Marta C. Gonzalez. Coupling electric vehicle charging with urban mobility. Under review at Science Advances.
- [47] Serdar Colak, Antonio Lima, and Marta C. Gonzalez. Understanding congested travel in urban areas. *Nat Commun*, 7, 03 2016.
- [48] Serdar Colak, Christian M Schneider, Pu Wang, and Marta C González. On the role of spatial dynamics and topology on network flows. *New Journal of Physics*, 15(11):113037, 2013.
- [49] José R Correa, Andreas S Schulz, and Nicolás E Stier-Moses. A geometric approach to the price of anarchy in nonatomic congestion games. *Games and Economic Behavior*, 64(2):457–469, 2008.
- [50] Carlos F Daganzo. Optimal sampling strategies for statistical models with discrete dependent variables. *Transportation Science*, 14(4):324–345, 1980.
- [51] Carlos F Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transp. Res. B*, 28:269–287, 1994.

- [52] Carlos F Daganzo. The cell transmission model, part ii: network traffic. *Transp. Res. B*, 29(2):79–93, 1995.
- [53] Carlos F Daganzo and Nikolas Geroliminis. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transp. Res. B*, 42(9):771–781, 2008.
- [54] Carlos F Daganzo and Yosef Sheffi. On stochastic models of traffic assignment. *Transportation Science*, 11(3):253–274, 1977.
- [55] D. De Martino, L. Dall’Asta, G. Bianconi, and M. Marsili. Congestion phenomena on complex networks. *Phys. Rev. E*, 79(1):015101, 2009.
- [56] Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3:1376, 03 2013.
- [57] Robert B Dial. A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B: Methodological*, 40(10):917–936, 2006.
- [58] Ian Dobson, Benjamin A Carreras, Vickie E Lynch, and David E Newman. Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 17(2):026103, 2007.
- [59] Omar Drissi-Kaïtouni and Abdelhamid Hamed-Bencheikroun. A dynamic traffic assignment model and a solution algorithm. *Transp. Sci.*, 26(2):119–128, 1992.
- [60] Pablo Echenique, Jesús Gómez-Gardeñes, and Yamir Moreno. Dynamics of jamming transitions in complex networks. *Europhys. Lett.*, 71(2):325, 2005.
- [61] Joseph Ferreira, Mi Diao, Yi Zhu, Weifeng Li, and Shan Jiang. Information infrastructure for research collaboration in land use, transportation, and environmental planning. *Transportation Research Record: Journal of the Transportation Research Board*, 2183(1):85–93, 2010.
- [62] California Center for Sustainable Energy. California plug-in electric vehicle driver survey results - may 2013. Technical report, California Center for Sustainable Energy, May 2013.
- [63] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.
- [64] Masao Fukushima. A modified frank-wolfe algorithm for solving the traffic assignment problem. *Transportation Research Part B: Methodological*, 18(2):169–177, 1984.

- [65] J García-Villalobos, I Zamora, JI San Martín, FJ Asensio, and V Aperribay. Plug-in electric vehicles in electric distribution networks: A review of smart charging approaches. *Renewable and Sustainable Energy Reviews*, 38:717–731, 2014.
- [66] Pacific Gas and Electric Company. Electric schedule e-19:medium general demand- metered tou service, 2010.
- [67] Nikolas Geroliminis and Carlos F Daganzo. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transp. Res. B*, 42(9):759–770, 2008.
- [68] Edward L. Glaeser, Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer. Growth in cities. Working Paper 3787, National Bureau of Economic Research, July 1991.
- [69] K.-I. Goh, B. Kahng, and D. Kim. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.*, 87:278701, Dec 2001.
- [70] Marta C González, Cesar A Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
- [71] Roger Guimerà, Alex Arenas, Albert Díaz-Guilera, and Francesc Giralt. Dynamical properties of model communication networks. *Phys. Rev. E*, 66(2):026704+, 2002.
- [72] Roger Guimerà, Albert Díaz-Guilera, Fernando Vega-Redondo, Antonio Cabrales, and Alex Arenas. Optimal network topologies for local search with congestion. *Phys. Rev. Lett.*, 89(24):248701, 2002.
- [73] Arda Halu, Antonio Scala, Abdulaziz Khayami, and Marta C González. Data-driven modeling of solar-powered urban microgrids. *Science Advances*, 2(1):c1500700, 2016.
- [74] Chioke B Harris and Michael E Webber. An empirically-validated methodology to simulate electricity demand for electric vehicle charging. *Applied Energy*, 126:172–181, 2014.
- [75] Shlomo Havlin and Daniel Ben-Avraham. Diffusion in disordered media. *Adv. Phys.*, 36(6):695–798, 1987.
- [76] Martin L Hazelton. Estimation of origin–destination matrices from link flows on uncongested networks. *Transportation Research Part B: Methodological*, 34(7):549–566, 2000.
- [77] Martin L Hazelton. Inference for origin–destination matrices: estimation, prediction and reconstruction. *Transportation Research Part B: Methodological*, 35(7):667–676, 2001.

- [78] Martin L Hazelton. Some comments on origin–destination matrix estimation. *Transportation Research Part A: Policy and Practice*, 37(10):811–822, 2003.
- [79] A Hernando, R Hernando, and A Plastino. Space–time correlations in urban sprawl. *Journal of The Royal Society Interface*, 11(91):20130930, 2014.
- [80] Paul Hines, Jay Apt, and Sarosh Talukdar. Large blackouts in north america: Historical trends and policy implications. *Energy Policy*, 37(12):5249–5259, 2009.
- [81] Paul Hines, Eduardo Cotilla-Sanchez, and Seth Blumsack. Do topological models provide good information about electricity infrastructure vulnerability? *Chaos: An interdisciplinary journal of nonlinear science*, 20(3):033122, 2010.
- [82] Paul Hines, Jeff Frolik, Jeffrey Marshall, Pooya Rezaei, Andrew Seier, Andrew Fuhrmann, Jonathan Dowds, and Alexander Hilshey. Understanding and managing the impacts of electric vehicles on electric power distribution systems. Technical report, University of Vermont, 2014.
- [83] Haye Hinrichsen. Non-equilibrium critical phenomena and phase transitions into absorbing states. *Adv. Phys.*, 49(7):815–958, 2000.
- [84] Md Shahadat Iqbal, Charisma F Choudhury, Pu Wang, and Marta C González. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74, 2014.
- [85] Jane Jacobs. *The death and life of great American cities*. Vintage, 1961.
- [86] Bruce N Janson. Dynamic traffic assignment for urban road networks. *Transportation Research Part B: Methodological*, 25(2):143–161, 1991.
- [87] R Jayakrishnan, Wei T Tsai, Joseph N Prashker, and Subodh Rajadhyaksha. A faster path-based algorithm for traffic assignment. *University of California Transportation Center*, 1994.
- [88] Shan Jiang, Gaston A Fiore, Yingxiang Yang, Joseph Ferreira Jr, Emilio Frazzoli, and Marta C González. A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, page 2. ACM, 2013.
- [89] Emre C. Kara, Jason S. Macdonald, Douglas Black, Mario Bérge, Gabriela Hug, and Sila Kiliccote. Estimating the benefits of electric vehicle smart charging at non-residential locations: A data-driven approach. *Applied Energy*, 155:515 – 525, 2015.
- [90] Willett Kempton and Steven E Letendre. Electric vehicles as a new power source for electric utilities. *Transportation Research Part D: Transport and Environment*, 2(3):157–175, 1997.

- [91] Jon M Kleinberg. Navigation in a small world. *Nature*, 406(6798):845–845, 2000.
- [92] Masao Kuwahara and Takashi Akamatsu. Decomposition of the reactive dynamic assignments with queues for a many-to-many origin-destination pattern. *Transp. Res. B*, 31(1):1–10, 1997.
- [93] Masao Kuwahara and Takashi Akamatsu. Dynamic user optimal assignment with physical queues for a many-to-many od pattern. *Transp. Res. B*, 35(5):461–479, 2001.
- [94] Stefan Lämmer, Björn Gehlsen, and Dirk Helbing. Scaling laws in the spatial structure of urban road networks. *Physica A: Statistical Mechanics and its Applications*, 363(1):89–95, 2006.
- [95] Larry J LeBlanc, Edward K Morlok, and William P Pierskalla. An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research*, 9(5):309–318, 1975.
- [96] Xiao Liang, Jichang Zhao, Li Dong, and Ke Xu. Unraveling the origin of exponential law in intra-urban human mobility. *Scientific reports*, 3, 2013.
- [97] Zhenhong Lin. Optimizing and diversifying electric vehicle driving range for us drivers. *Transportation Science*, 48(4):635–650, 2014.
- [98] HP Lo, N Zhang, and William HK Lam. Estimation of an origin-destination matrix with random link choice proportions: a statistical approach. *Transportation Research Part B: Methodological*, 30(4):309–324, 1996.
- [99] Thomas Louail, Maxime Lenormand, Miguel Picornell, Oliva García Cantú, Ricardo Herranz, Enrique Frias-Martinez, José J Ramasco, and Marc Barthelemy. Uncovering the spatial structure of mobility networks. *Nature Communications*, 6:6007, 2015.
- [100] Rémi Louf and Marc Barthelemy. How congestion shapes cities: from mobility patterns to scaling. *Scientific Reports*, 4:5561, July 2014.
- [101] Chung-Cheng Lu, Xuesong Zhou, and Kuilin Zhang. Dynamic origin-destination demand flow estimation under congested traffic conditions. *Transportation Research Part C: Emerging Technologies*, 34:16–37, 2013.
- [102] Xin Lu, Linus Bengtsson, and Petter Holme. Predictability of population displacement after the 2010 haiti earthquake. *Proceedings of the National Academy of Sciences*, 109(29):11576–11581, 2012.
- [103] Zhongjing Ma, Duncan S Callaway, and Ian A Hiskens. Decentralized charging control of large populations of plug-in electric vehicles. *Control Systems Technology, IEEE Transactions on*, 21(1):67–78, 2013.

- [104] MJ Maher. Inferences on trip matrices from observations on link volumes: a bayesian statistical approach. *Transportation Research Part B: Methodological*, 17(6):435–447, 1983.
- [105] George L Mamede, Nuno AM Araújo, Christian M Schneider, José Carlos de Araújo, and Hans J Herrmann. Overspill avalanching in a dense reservoir network. *Proc. Nat. Acad. USA*, 109(19):7191–7195, 2012.
- [106] Traffic Assignment Manual. Bureau of public roads. *US Department of Commerce*, 1964.
- [107] A Paolo Masucci, Joan Serras, Anders Johansson, and Michael Batty. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E*, 88(2):022812, 2013.
- [108] Thomas C. McAndrew, Christopher M. Danforth, and James P. Bagrow. Robustness of spatial micronetworks. *Phys. Rev. E*, 91:042813, Apr 2015.
- [109] Noel Melton, Jonn Axsen, and Daniel Sperling. Moving beyond alternative fuel hype to decarbonize transportation. *Nature Energy*, 1:16013, 2016.
- [110] GA Mendes, LR Da Silva, and HJ Herrmann. Traffic gridlock on complex networks. *Phys. A*, 391(1):362–370, 2012.
- [111] Deepak K Merchant and George L Nemhauser. A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science*, 12(3):183–199, 1978.
- [112] Deepak Merugu, Balaji S Prabhakar, and N Rama. An incentive mechanism for decongesting the roads: A pilot program in bangalore. In *Proc. of ACM NetEcon Workshop*. Citeseer, 2009.
- [113] Dov Monderer and Lloyd S Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996.
- [114] S. J. Moura, H. K. Fathy, D. S. Callaway, and J. L. Stein. A stochastic optimal control approach for power management in plug-in hybrid electric vehicles. *IEEE Transactions on Control Systems Technology*, 19(3):545–555, May 2011.
- [115] Mario Mureddu, Guido Caldarelli, Alessandro Chessa, Antonio Scala, and Alfonso Damiano. Green power grids: How energy from renewable sources affects networks and markets. *PloS one*, 10(9):e0135312, 2015.
- [116] Kai Nagel. Particle hopping models and traffic flow theory. *Phys. Rev. E*, 53(5):4655, 1996.
- [117] Kai Nagel and Michael Schreckenberg. A cellular automaton model for freeway traffic. *J. Phys. I*, 2(12):2221–2229, 1992.

- [118] John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.
- [119] Mark EJ Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005.
- [120] Xiaojian Nie and H Michael Zhang. A comparative study of some macroscopic link models used in dynamic traffic assignment. *Networks and Spatial Economics*, 5(1):89–115, 2005.
- [121] Yu Nie, HM Zhang, and WW Recker. Inferring origin–destination trip matrices with a decoupled gls path flow estimator. *Transportation Research Part B: Methodological*, 39(6):497–518, 2005.
- [122] Yu Marco Nie. A class of bush-based algorithms for the traffic assignment problem. *Transportation Research Part B: Methodological*, 44(1):73–89, 2010.
- [123] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PloS one*, 7(5):e37027, 2012.
- [124] U.S. Department of Energy. Workplace charging challenge, mid-program review: Employees plug in. Technical report, U.S. Department of Energy, December 2015.
- [125] Toru Ohira and Ryusuke Sawatari. Phase transition in a computer network traffic model. *Phys. Rev. E*, 58:193–195, Jul 1998.
- [126] J de Ortúzar and Luis G Willumsen. *Modelling transport*. John Wiley & Sons, Chichester, England, 1994.
- [127] Sakshi Pahwa, Caterina Scoglio, and Antonio Scala. Abruptness of cascade failures in power grids. *Scientific reports*, 4, 2014.
- [128] P. Patriksson. *The Traffic Assignment Problem: models and methods*. V.S.P. Intl Science, 1994.
- [129] Santi Phithakkitnukoon, Teerayut Horanont, Giusy Di Lorenzo, Ryosuke Shibasaki, and Carlo Ratti. Activity-aware map: Identifying human daily activity pattern using mobile phone data. *Human Behavior Understanding*, pages 14–25, 2010.
- [130] Arthur C Pigou. The economics of welfare, 1920. *McMillan&Co., London*, 1932.
- [131] Louis A Pipes. Car following models and the fundamental diagram of road traffic. *Transportation Research*, 1(1):21–29, 1967.
- [132] Carlo Giacomo Prato. Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, 2(1):65–100, 2009.

- [133] Zhen Sean Qian and HM Zhang. On centroid connectors in static traffic assignment: Their effects on flow patterns and how to optimize their selections. *Transportation Research Part B: Methodological*, 46(10):1489–1503, 2012.
- [134] Sumedha Rajakaruna, Farhad Shahnia, and Arindam Ghosh. *Plug In Electric Vehicles in Smart Grids*. Springer, 2015.
- [135] Jonathan Reades, Francesco Calabrese, and Carlo Ratti. Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 36(5):824–836, 2009.
- [136] Yihui Ren, Mária Ercsey-Ravasz, Pu Wang, Marta C. González, and Zoltán Toroczkai. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nature Communications*, 5:5347, 2014.
- [137] Pooya Rezaei, Jeff Frolik, and Paul DH Hines. Packetized plug-in electric vehicle charge management. *Smart Grid, IEEE Transactions on*, 5(2):642–650, 2014.
- [138] Anthony J Richardson, Elizabeth S Ampt, and Arnim H Meyburg. *Survey methods for transport planning*. Eucalyptus Press Melbourne, 1995.
- [139] Martin Rosvall, Ala Trusina, Petter Minnhagen, and Kim Sneppen. Networks and cities: An information perspective. *Physical Review Letters*, 94(2):028701, 2005.
- [140] Tim Roughgarden. *Selfish routing and the price of anarchy*. MIT press, 2005.
- [141] Tim Roughgarden and Éva Tardos. How bad is selfish routing? *Journal of the ACM (JACM)*, 49(2):236–259, 2002.
- [142] Tim Roughgarden and Éva Tardos. Bounding the inefficiency of equilibria in nonatomic congestion games. *Games and Economic Behavior*, 47(2):389–403, 2004.
- [143] Samveg Saxena, Caroline Le Floch, Jason MacDonald, and Scott Moura. Quantifying ev battery end-of-life through analysis of travel needs with vehicle powertrain models. *Journal of Power Sources*, 282:265 – 276, 2015.
- [144] Christian M Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, 2013.
- [145] Morton Schneider. Gravity models and trip distribution theory. *Papers in Regional Science*, 5(1):51–56, 1959.
- [146] Michael Schreckenberg, Andreas Schadschneider, Kai Nagel, and Nobuyasu Ito. Discrete stochastic models for traffic flow. *Phys. Rev. E*, 51(4):2939, 1995.

- [147] Andres Sevtsuk and Carlo Ratti. Does urban mobility have a daily routine? learning from the aggregate data of mobile networks. *Journal of Urban Technology*, 17(1):41–60, 2010.
- [148] Yosef Sheffi. *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods*. Prentice Hall, 1985.
- [149] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.
- [150] Michael E Smith. Design of small-sample home-interview travel surveys. *Transportation Research Record*, 701:29–35, 1979.
- [151] MJ Smith. The marginal cost taxation of a transportation network. *Transportation Research Part B: Methodological*, 13(3):237–242, 1979.
- [152] Susan Solomon, Gian-Kasper Plattner, Reto Knutti, and Pierre Friedlingstein. Irreversible climate change due to carbon dioxide emissions. *Proceedings of the national academy of sciences*, pages pnas–0812721106, 2009.
- [153] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.
- [154] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
- [155] Heinz Spiess. A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B: Methodological*, 21(5):395–412, 1987.
- [156] Heinz Spiess. Technical note — conical volume-delay functions. *Transportation Science*, 24(2):153–158, 1990.
- [157] Sameet Sreenivasan, Reuven Cohen, Eduardo López, Zoltán Toroczkai, and H Eugene Stanley. Structural bottlenecks for communication in networks. *Phys. Rev. E*, 75(3):036105, 2007.
- [158] Peter R Stopher and Stephen P Greaves. Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, 41(5):367–381, 2007.
- [159] Samuel A Stouffer. Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, 5(6):845–867, 1940.
- [160] Ananth Subramanian, Manuel J Garcia, Duncan S Callaway, Kameshwar Poolla, and Pravin Varaiya. Real-time scheduling of distributed resources. *Smart Grid, IEEE Transactions on*, 4(4):2122–2130, 2013.

- [161] Gil Tal, Michael Nicholas, Jamie Davies, and Justin Woodjack. Charging behavior impacts on electric vehicle miles traveled. *Transportation Research Record: Journal of the Transportation Research Board*, 2454:53–60, 2014.
- [162] Michael A Tamor, Paul E Moraal, Briana Reprogle, and Miloš Milačić. Rapid estimation of electric vehicle acceptance using a general description of driving patterns. *Transportation Research Part C: Emerging Technologies*, 51:136–148, 2015.
- [163] Elad Tomer, Leonid Safonov, Nilly Madar, and Shlomo Havlin. Optimization of congested traffic by controlling stop-and-go waves. *Phys. Rev. E*, 65(6 Pt 2):065101, 2002.
- [164] Jameson L. Toole, Serdar Colak, Bradley Sturt, Lauren P. Alexander, Alexandre Evsukoff, and Marta C. González. The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58:162 – 177, 2015.
- [165] Tali Trigg, Paul Telleen, R Boyd, F Cuenot, D D’Ambrosio, R Gaghen, JF Gagné, A Hardcastle, D Houssin, AR Jones, et al. Global ev outlook: understanding the electric vehicle landscape to 2020. *Int. Energy Agency*, pages 1–40, 2013.
- [166] Konstantina Valogianni, Wolfgang Ketter, and John Collins. A multiagent approach to variable-rate electric vehicle charging coordination. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1131–1139. International Foundation for Autonomous Agents and Multiagent Systems, 2015.
- [167] Konstantina Valogianni, Wolfgang Ketter, John Collins, and Dmitry Zhdanov. Effective management of electric vehicle storage using smart charging. In *AAAI*, pages 472–478, 2014.
- [168] John B Van Huyck, Raymond C Battalio, and Richard O Beil. Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, pages 234–248, 1990.
- [169] Henk J Van Zuylen and Luis G Willumsen. The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, 14(3):281–293, 1980.
- [170] William S Vickrey. Congestion theory and transport investment. *The American Economic Review*, pages 251–260, 1969.
- [171] Pu Wang, Timothy Hunter, Alexandre M Bayen, Katja Schechtner, and Marta C González. Understanding road usage patterns in urban areas. *Scientific reports*, 2:1001, 2012.

- [172] J.G. Wardrop and Institution of Civil Engineers (Great Britain). *Some theoretical aspects of road traffic research*. Institution of Civil Engineers, 1952.
- [173] D.C.: Transportation Research Board Washington. Highway capacity manual 2010. Technical report, Washington, D.C.: Transportation Research Board, 2010.
- [174] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.
- [175] 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*, 338(6104):267–270, 2012.
- [176] HJ Wootton and GW Pick. A model for trips generated by households. *Journal of Transport Economics and Policy*, pages 137–153, 1967.
- [177] Zhenhua Wu, Lidia A. Braunstein, Shlomo Havlin, and H. Eugene Stanley. Transport in Weighted Networks: Partition into Superhighways and Roads. *Phys. Rev. Lett.*, 96(14):148702+, April 2006.
- [178] Gang Yan, Tao Zhou, Bo Hu, Zhong Q. Fu, and Bing H. Wang. Efficient routing on complex networks. *Phys. Rev. E*, 73(4):046108+, 2006.
- [179] Xiao-Yong Yan, Chen Zhao, Ying Fan, Zengru Di, and Wen-Xu Wang. Universal predictability of mobility patterns in cities. *Journal of The Royal Society Interface*, 11(100):20140834, 2014.
- [180] Lei Yang, Junshan Zhang, and H Vincent Poor. Risk-aware day-ahead scheduling and real-time dispatch for electric vehicle charging. *Smart Grid, IEEE Transactions on*, 5(2):693–702, 2014.
- [181] 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, 2014.
- [182] M. Yilmaz and P. T. Krein. Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles. *IEEE Transactions on Power Electronics*, 28(5):2151–2169, May 2013.
- [183] Hyejin Youn, Michael T Gastner, and Hawoong Jeong. Price of anarchy in transportation networks: Efficiency and optimality control. *Phys. Rev. Lett.*, 101(12):128701, 2008.
- [184] Tugce Yuksel and Jeremy J Michalek. Effects of regional temperature on electric vehicle efficiency, range, and emissions in the united states. *Environmental science & technology*, 49(6):3974–3980, 2015.

- [185] Alireza Zakariazadeh, Shahram Jadid, and Pierluigi Siano. Multi-objective scheduling of electric vehicles in smart distribution system. *Energy Conversion and Management*, 79:43–53, 2014.
- [186] Xianyuan Zhan, Samiul Hasan, Satish V Ukkusuri, and Camille Kamga. Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C: Emerging Technologies*, 33:37–49, 2013.
- [187] Liang Zhao, Ying-Cheng Lai, Kwangho Park, and Nong Ye. Onset of traffic congestion in complex networks. *Phys. Rev. E*, 71:026125, Feb 2005.
- [188] Yu Zheng and Xing Xie. Learning travel recommendations from user-generated gps traces. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):2, 2011.