

Assessing the Impact of Real-time Ridesharing on Urban Traffic using Mobile Phone Data

Lauren P. Alexander
Dept. of Civil and Environmental Engineering
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts, USA
lpalex@mit.edu

Marta C. González
Dept. of Civil and Environmental Engineering
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts, USA
martag@mit.edu

ABSTRACT

Recently, smart-phone based technology has enabled ridesharing services to match customers making similar trips in real-time for a reduced rate and minimal inconvenience. But what are the impacts of such services on city-wide congestion? The answer lies in whether or not ridesharing adds to vehicle traffic by diverting non-driving trips like walking, transit, or cycling, or reduces vehicle traffic by diverting trips otherwise made in private, single occupancy cars or taxis. This research explores the impact of rideshare adoption on congestion using mobile phone data. We extract average daily origin-destination (OD) trips from mobile phone records and estimate the proportions of these trips made by auto and other non-auto travelers. Next, we match spatially and temporally similar trips, and assume a range of adoption rates for auto and non-auto users, in order to distill rideshare vehicle trips. Finally, for several adoption scenarios, we evaluate the impacts of congestion network-wide.

Keywords

Ridesharing; on-demand ride services; human mobility; mobile phone data; data mining

1. INTRODUCTION

In 2014, ridesourcing services Uber¹, Lyft², and Sidecar³ launched ridesharing programs in the US that match customers making similar trips. Ridesharing offers monetary incentives to customers who pay a reduced rate for sharing some or all of their ride with another passenger, as well as drivers who are able to carry more passengers more efficiently. Ubiquitous technologies have allowed for the emergence of these real-time ridesharing services, with GPS providing driver and customer locations and route navigation,

¹www.uber.com

²www.lyft.com

³www.side.cr

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright is held by the author/owner(s).

UrbComp'15, August 10, 2015, Sydney, Australia.

smartphone apps affording real-time ride requests, and social networks establishing trust and accountability between customers and drivers. Further, advances in computing speed and data storage has enabled the development of platforms to run rideshare optimization algorithms in real-time.

Ridesharing has garnered support for its potential to reduce private automobile use by providing a convenient and affordable alternative to driving alone, translating to reduced roadway congestion and vehicle emissions in the short-term, and reduced automobile ownership in the longer term. A recent intercept survey of ridesharing customers in San Francisco reveals that although ridesharing often substitutes for longer transit trips, it otherwise complements transit, with many observed trip origins and destinations near transit stations [31]. However, ridesharing critics are skeptical about the likelihood that ridesharing will decrease vehicle congestion and emissions, due to the potential of ridesharing to divert trips from transit or other non-motorized modes and induce new trips altogether. Furthermore, safety and liability concerns tied to inadequate driver training and insurance, as well as direct competition with highly regulated taxi companies, has led some to call for regulations of the app-based ridesharing industry. As city leaders and policy makers are faced with decisions about regulating the growing rideshare market, it is becoming increasingly important that the overall impacts of ridesharing are understood.

Whether or not ridesharing adds to vehicle traffic depends on the balance of competing forces. On the one-hand, ridesharing may increase traffic by replacing non-driving modes such as transit, walking, or cycling or inducing new trips. On the other hand, ridesharing may decrease traffic by increasing vehicle occupancy, serving the first/last mile of transit trips, and reducing private car ownership and use. This research focuses on understanding the impact of two of these key drivers: adoption by non-drivers and adoption by travelers of private, single occupancy cars or taxis. The other factors are likely to occur on a longer time scale and therefore harder to quantify.

With the uncertainty surrounding the impacts of ridesharing in mind, this research aims to answer two questions unresolved in existing literature:

- What proportion of trips can be matched by a real-time ridesharing service given the temporal and spatial distribution of all urban trips and travel modes?
- What would be the change in the number of vehicles and traffic congestion given relative adoption rates of ridesharing from auto and non-auto travelers?

To answer these questions we develop a model of urban travel demand based on mobile phone data. In Section 2, we put our methods and work in context of previous literature on related topics. We first compare and contrast our data and methods with recent research in ridesharing, demonstrating how we differ from and add to this body of work. Then, we relate our proposed approach to methods developed in the urban computing and transportation domains to model human mobility and demand for transportation services and infrastructure.

The rest of the this paper follows with discussions of the data, methods, and results. We first estimate average daily origin-destination (OD) trips from mobile phone records, then estimate the proportions of these trips made by driving and other non-driving modes. Next, we match spatially and temporally similar trips, and explore a range of adoption rates for drivers and non-drivers, in order to distill rideshare vehicle trips, and by extension, total vehicle trips. Finally, we use algorithms to allocate these vehicle trips to a road network and evaluate the impacts of ridesharing on urban congestion.

2. RELATED WORK

2.1 Ridesharing

To-date, much of the research related to ridesharing has focused on understanding the characteristics of ridesharing trips and users. In a recent survey of app-based, on-demand rideshare users in San Francisco, researchers found that 45% of ridesharers stated they would have used a taxi or driven their own car had ridesharing not been available, while 43% would have taken transit, walked, or cycled [31]. The authors conclude there is a need for research to explore the impact of modal shift on vehicle congestion and emissions. This work is a step in that direction and leverages mobile phone data to understand underlying travel demand not captured by small-scale intercept surveys.

Santi et al. developed a framework to compute maximum matching of shareability networks constructed from an OpenStreetMaps⁴ (OSM) road network and 172 million taxi trips made in New York City in 2011 [33]. We add to this body of work by introducing a method that takes into account trip-making by all modes, in contrast to just taxis, which represent only a portion of potential rideshare demand. Further, authors explicitly assume that traffic conditions—which impact the travel time criteria used by their matching algorithms—will remain largely unaltered by the emergence of ridesharing. Given their finding that ridesharing could cut total taxi vehicle miles by 40%, we revisit this assumption in this work; using traffic assignment algorithms commonly used in transportation planning applications to route total vehicle demand, we assess the impact of ridesharing on network-wide congestion.

Cici et al. used mobile phone and social network data to evaluate demand for ride-sharing between strangers, friends, and friends-of-friends in four cities in Spain and the US [18]. As in this work, the authors use mobile phone Call Detail Records (CDRs), however, they focus on commuting trips between home and work locations inferred from CDR, Twitter, and Foursquare data. We build on this research by (i)

using trips across all purposes, and (ii) estimating auto mode shares by origin and destination rather than selecting a portion of the trips as vehicle trips based on a single city-wide mode share, and (iii) estimating the impacts of ridesharing on urban congestion and travel times rather than using an online mapping service for static routing and travel time characteristics.

Researchers in [18, 33] and similar related work [25] focused on addressing the computational challenges of trip-matching—an NP-hard optimization problem—in real-time and developed heuristics to quantify potential ride-sharing demand. These algorithms re-route trips in order to match them with similar, overlapping trips, explicitly capturing demand for ridesharing relative to passenger’s willingness to experience prolonged travel time. We believe that development of such heuristics are crucial for the effective implementation of a real-time ridesharing system. In this work, however, we focus on other aspect that affect rideshare demand and urban congestion, namely:

- Total network-wide trips by mode
- Rideshare adoption by mode, and the impact of this mode-shift on the number of network-wide vehicles, and
- Dynamic relationship between demand and traffic congestion.

2.2 Mobility modeling

Data generated by the pervasive use of cellular phones has offered insights into characteristics of human mobility patterns. Recent work has found that individuals are predictable, unique, and slow to explore new places [21, 12, 20, 37, 36, 15, 14]. The availability of similar data across 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 massive, passive data for human mobility modeling have been realized in various contexts, including population movement [24], daily mobility motifs [34, 35], individual survey tracking and stay extraction [7], OD-estimation and validation [13, 27, 42, 22], traffic speed estimation [8, 43], and activity modeling [30, 32]. While these works have laid an important foundation, none have evaluated the impacts of a new transportation option on urban traffic. Moreover, with mobile phone data available in real-time, these methods can be adapted to support real-time rideshare matching applications. With our proposed framework, we therefore introduce a novel application to the field of urban computing.

Within the transportation domain, evaluating the demand for and impact of a new travel mode or option traditionally involves acquiring, adapting, and running a travel demand model. In particular, mode choice models are often used to estimate the diversion of trips to new modes or travel alternatives [6, 9, 10]. However, such models are expensive to develop and calibrate and their availability may be limited. Instead, we propose a framework to assess overall demand and congestion impacts of a new mode based on a range of hypothetical adoption levels and hourly OD trips inferred from mobile phone data. This approach offers an alternative end-to-end solution to quickly and economically perform transportation scenario analyses in any city for which mobile phone data is available.

⁴An open source mapping community supporting data on road networks all over the world. www.openstreetmap.org

3. DATA

3.1 Mobile Phone

To estimate travel demand patterns we utilize mobile phone CDR data in the Boston metropolitan area. The CDR dataset contains more than eight billion mobile phone records for roughly two million anonymized users over two months in the Spring of 2010. Each record contains an anonymous user ID, longitude, latitude, and timestamp at the instance of a phone call or other types of phone communication (such as sending SMS, etc.). The coordinates of the records are estimated by service providers based on a standard triangulation algorithm, with an accuracy of about 200 to 300 meters.

3.2 GIS/Survey

For this analysis, we rely upon a variety of spatial and survey data sources, summarized below.

- **Census tracts:** As detailed in Section 4.1, CDR trips are aggregated to the spatial resolution of 974 study area Census tracts, which contain roughly 5000 residents each. We use a GIS shapefile as well as population estimates from the American Community Survey to expand observed users to total population [1, 2].
- **Departure time:** As described in Section 4.1, we infer trip departure hour from the publicly-available 2009 National Household Travel Survey (NHTS) [40]. Filtering for respondents residing in a consolidated metropolitan statistical area (CMSA) or MSA with populations greater than or equal to 3 million, like in Boston, we generate hourly departure time distributions for weekdays and weekends and all trip purposes.
- **Commuting trips:** The 2006-2010 Census Transportation Planning Products (CTPP) Part 3 provides commuting characteristics between Census tract pairs [41]. This nationally-available dataset provides tabulations across 16 different travel modes, which we use to infer mode shares, as described in Section 4.2.
- **Communities:** In order to spatially match CDR trips in Section 4.3, we aggregate them to 174 areas (163 study area towns and 11 Boston neighborhoods), referred to as communities in this paper. We divided Boston into 11 neighborhoods to match rideshare trips at a finer resolution within the city of Boston. We used a GIS shapefile of town boundaries developed by MassGIS to map Census tracts to these communities, but further split Boston into neighborhoods using local knowledge of neighborhood boundaries [26].
- **Vehicle trips:** The 2010/2011 Massachusetts Travel Survey (MHTS) contains data on 153,099 trips made by 32,739 people across the state [28]. We utilize reported mode, departure time, travel time, and origin and destination tracts to calibrate the traffic assignment algorithms described in Section 4.4.
- **Road network:** For traffic simulation, as described in Section 4.4, we use a GIS shapefile from the local transportation authority containing road characteristics such as speed limits, road capacities, number of lanes, and classifications [17]. These attributes allow

us to compute travel times and capacities, key inputs of our traffic assignment procedure.

4. METHODS

4.1 Trip Estimation

Mobile phones offer good, but imperfect measurements of our whereabouts. First, coordinate locations are associated with uncertainty due to (i) tower-to-tower call balancing performed by the mobile service provider, creating the appearance of false movements, and (ii) inexact signal triangulation. Furthermore, observations are only recorded when an individual interacts with his or her device, resulting in an incomplete picture of daily behavior and heterogeneous sampling frequencies across users. Given these limitations, mobile phone data must be de-noised and processed to extract representative daily mobility patterns.

The first step to reliably infer trips is to extract meaningful origins and destinations, or *stays*, where users engage in an activity. To do so, we consolidate and filter CDR records in space and time using an agglomerative clustering algorithm described in detail in [5, 23]. The next step is to infer contextual information about each clustered *stay* location, using observation frequency, day of week, and time of day as described in [5, 16]. Specifically, we are interested in identifying each user's *home* location, which is the *stay* they are observed at most frequently between the hours of 8pm and 7am on weeknights. For users with too few stay locations, the CDR data may not fully represent their travel patterns. Accordingly, users with fewer than 8 (one per week, on average) visits to designated *home* stays are filtered out as in [5]. This filter also ensures with a reasonable degree of certainty that the designated stay is the user's home, a key assumption in our method of upscaling users to population, while still achieving a sample size that is an order of magnitude larger than in most household travel surveys.

With de-noised *stays* determined for our filtered users U , we next construct trips between these origins and destinations. As in [5, 16], we define *effective days* as periods between 3am one morning and 3am on the next consecutive morning, enabling us to capture trips to/from late night activities that occur past midnight. Moreover, we presume users start and end each *effective day* at home, such that we construct trips to/from home for users who are not observed at home for their first and last records of an effective day. Since the arrival time and duration at stay locations reflect the *observed* (based on phone usage) rather than *true* arrival time and duration, we probabilistically infer departure hour h , as in [5, 16]. We do so by assigning 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 is derived using National Household Travel Survey (NHTS) data on trips in major US cities. Lastly, we are interested in generalizing trip patterns between geographic areas so we convert origin and destination points to census tract IDs, resulting in a matrix of tract-pair trips t_{ij} .

Finally, we convert user-specific trips to average daily trips representative of all individuals in a city. To control for differences in mobile phone usage across users, we scale trips based on how often an individual u uses their phone. By dividing a user's trips t_{ij} by the number of days n in which we

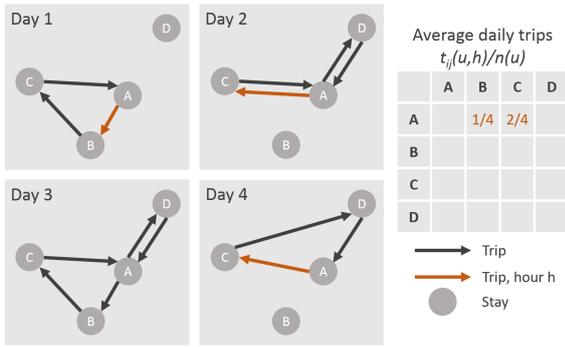


Figure 1: User u makes trips between four unique locations over four days. The user’s trips $t_{ij}(u, h)$ in hour h , shown in orange, are converted to average daily trips in hour h by dividing by the number of days $n(u) = 4$ we observe the user.

observe the user, we effectively measure the average number of trips a user makes between two locations in a day. This procedure is illustrated in Figure 1, with user u making three trips in hour h over $n(u) = 4$ days. Moreover, to control for differences in mobile phone market share across Boston, we scale trips based on the ratio of cell phone users to population residing in each tract. Accordingly, each user’s trips are multiplied by an expansion factor w computed as the ratio of the number of users assigned to their home tract and the census population of their home tract. Summing across all individuals, we compute average daily trip matrices T_{ij} , as summarized in Equation 1.

$$T_{ij}(h) = \sum_{u=1}^U w(u) * t_{ij}(u, h)/n(u) \quad (1)$$

4.2 Mode Share Estimation

We want to determine the number of trips currently made by three travel modes: drive-alone or taxi; carpool; and non-driving modes such as transit, cycling, or walking. The fraction of travelers for each mode is calculated using Census Transportation Planning Products (CTPP) of commuting data. Averaging this data across Boston, 70% of commuters drive alone or take a taxi, 8% carpool, and 22% take non-driving modes. Using this commuting data, we assign OD-pair mode shares to all CDR trips. Note that, although the fraction of travelers for each of these three modes is based on data for commuting trips only, we apply these shares to all trips due to the lack of available mode share data for non-commuting trips.

Next, we infer the proportion of travelers between each tract-pair using each of these three travel modes. We use CTPP data aggregated from Census tract pairs ij to community pairs IJ to minimize the effects of matrix sparsity and sample size. In other words, averaging data across larger geographic areas results in more representative mode shares, given that the smaller sample size of census data relative to CDR data results in many tract-pairs with zero or few CTPP trips for each mode. Even with this aggregation, however, some community pairs still have few total trips and/or zero auto trips due to their small sample size. For these pairs, we therefore assign average mode shares depending on their

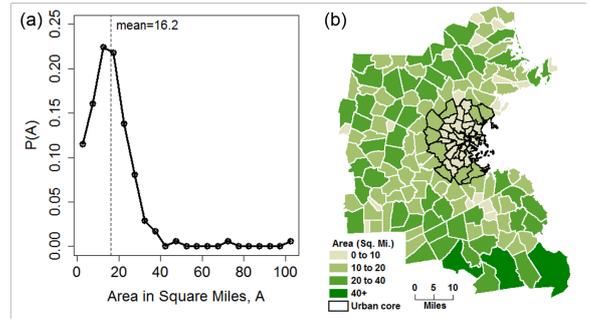


Figure 2: (a) Probability distribution of community area in miles² and (b) Spatial distribution of community area, with area increasing from light to dark green, and black borders denoting the Urban core.

geography: Urban-Urban, Suburban-Urban, or Suburban-Suburban. Communities lying within the I-95 highway ring, demarcated by black borders in Figure 2b are designated as Urban, and all other communities are designated Suburban. In this way, we adjust community-pair mode shares impacted by small sampling sizes to reflect more statistically significant mode shares of similar community-pairs.

After this aggregation and adjustment, each community-pair is assigned three mode shares: (i) drive-alone and taxi d_{IJ} , (ii) carpool c_{IJ} , and (iii) non-auto o_{IJ} . These mode shares are mutually exclusive and exhaustive, such that $(d_{IJ} + c_{IJ} + o_{IJ}) = 1$. Detailed results and validation of these methods can be found in [5, 19, 39].

4.3 Rideshare Vehicle Estimation

Given OD trips inferred from the CDR data and mode shares inferred from census data, we next estimate drive-alone, carpool, and non-auto trips and match candidate ridesharing trips together. We explore different adoption rates for the ride sharing service by travelers that use taxi or drive-alone a_d and travelers that use non-driving modes a_o . While total rideshare trips will increase with increasing values of a_d and a_o , vehicle traffic will reduce for $a_d \gg a_o$, and increase under $a_d \ll a_o$.

We assume existing carpools would not adopt ridesharing since they already coordinate their trip with at least one other traveler; however, we take into account the contribution of carpool vehicles to total vehicle traffic. Carpoolers in the context of all trip purposes, as we use the term here, can be thought of as anyone making a trip in a vehicle with at least one other person.

We first define the spatial resolution for which ridesharing trips can be matched. Because Census tracts are the size of a few city blocks in downtown Boston, it is too restrictive of an assumption to only match trips beginning and ending in the same Census tracts. Accordingly, we match potential ridesharing trips based on the study area communities. Figure 2a illustrates the probability distribution function of community area, with a median of 15.0 miles² and a mean of 16.2 miles² (by comparison, for a circle, this implies a radius of 2.27 miles). Figure 2b illustrates the spatial distribution of community areas, with the majority of those in the urban core (demarcated with black borders) having areas less than 10 miles², while communities with the greatest area lie on the southern border of the study area.

It should be noted that when we refer to ridesharing we explicitly mean *end-to-end ridesharing*⁵. However, by using spatial resolution of communities we also implicitly capture *en-route ridesharing*⁶ trips with the origins and destinations of matched riders falling within the same communities.

We also require finer temporal resolution to match potential ridesharing trips than departure hour as estimated in Section 4.1. We assume that trips occur uniformly throughout each hour, and compute the number of trips made within time window Δ . For $\Delta = 6$ minutes, for example, hourly demand is split into 10 intervals, and potential ridesharing trips are matched within each interval. From the rider's perspective, Δ represents the maximum allowable change in departure time they would be willing to incur to take ridesharing. A larger Δ will enable more trips to be matched by a ridesharing service, but will impose higher level of inconvenience to customers, which may hinder adoption.

Lastly, we introduce a parameter s , representing the number of rideshare customers that can be matched within a vehicle. In this study, we assume $s = 2$ since existing rideshare services currently match two trip requests. However, if and when rideshare services allow larger rideshare passenger occupancy, we can increase s to match more potential riders in fewer rideshare vehicles. This parameter also gives the flexibility to assess the potential of dynamic shuttle services, such as Bridj⁷, which currently serves Boston-area commuters using sprinter vans with 12 passenger seats.

Finally, we estimate the number of vehicles $V_{IJ}(h)$ needed to satisfy trips $T_{IJ}(h)$ by first mapping tract pair trips to community pairs, $T_{ij}(h) \rightarrow T_{IJ}(h)$. Next, driver adopters $f_{D,IJ}(h)$ per Δ are calculated using the driver adoption rate a_d and share of drivers d_{IJ} as shown in Equation 2. Similarly, non-driver adopters $f_{O,IJ}(h)$ per Δ are calculated using the non-driver adoption rate a_o and share of non-drivers o_{IJ} as shown in Equation 3. Potential riders $f_{IJ}(h)$ per Δ are simply the sum of driver and non-driver adopters (Equation 4).

$$f_{D,IJ}(h) = T_{IJ}(h) * a_d * d_{IJ} * \Delta / 60 \quad (2)$$

$$f_{O,IJ}(h) = T_{IJ}(h) * a_o * o_{IJ} * \Delta / 60 \quad (3)$$

$$f_{IJ}(h) = f_{D,IJ}(h) + f_{O,IJ}(h) \quad (4)$$

In practice, however, a ridesharing system may not be able to capture all of these potential adopters $f_{IJ}(h)$. Given the number of adopters traveling between two communities within a time step Δ , we will reject adopters if: (i) there are less than s travelers, or (ii) there is a residual in the division of the number of trips by s . Accordingly, we can measure the efficiency $e_{IJ}(h)$, or fraction of potential demand that can be realized, of the rideshare system.

We refer to riders that are unable to be matched in a rideshare vehicle as rejected riders $r_{IJ}(h)$. Further, we assume that these rejected riders will instead take their *original* mode, meaning that driver adopters who are rejected will drive just as they would if ridesharing were not available. Accordingly, single-occupancy vehicles carrying driver adopters who are rejected $v_{X,IJ}(h)$ are accounted for

in the calculation of total vehicles in subsequent steps.

Under this framework, rejected riders $r_{IJ}(h)$ are calculated as the remainder of the potential riders $f_{IJ}(h)$ per Δ divided by group size s (Equation 5). By extension, matched riders $m_{IJ}(h)$ are calculated as the difference between the potential and rejected riders (Equation 6), and the efficiency of ridesharing $e_{IJ}(h)$ can be computed as the ratio of matched to potential riders (Equation 7).

$$r_{IJ}(h) = f_{IJ}(h) \bmod s * 60 / \Delta \quad (5)$$

$$m_{IJ}(h) = f_{IJ}(h) * 60 / \Delta - r_{IJ}(h) \quad (6)$$

$$e_{IJ}(h) = m_{IJ}(h) / f_{IJ}(h) * \Delta / 60 \quad (7)$$

The total number of vehicles traveling between two communities in a given hour $V_{IJ}(h)$ is the sum of all types of vehicles, namely: rideshare vehicles ($v_{W,IJ}(h)$), un-matched drivers ($v_{X,IJ}(h)$), drivers not adopting ($v_{Y,IJ}(h)$), and carpool vehicles ($v_{Z,IJ}(h)$), as shown in Equation 8. Since vehicles carrying matched riders have a vehicle occupancy of s , the number of rideshare vehicles $v_{W,IJ}(h)$ is simply the number of matched riders $m_{IJ}(h)$ divided by group size s (Equation 9). Single-occupancy vehicles carry rejected driver adopters $v_{X,IJ}(h)$, equal to the ratio of driver adopters to total potential riders $f_{D,IJ}(h) / f_{IJ}(h)$ multiplied by rejected riders (Equation 10), as well as by drivers who did not adopt $v_{Y,IJ}(h)$ (Equation 11). Lastly, carpool vehicles $v_{Z,IJ}(h)$ are computed using the average vehicle occupancy of carpool vehicles in the study area (in Boston, $p = 2.18$), as shown in Equation 12.

$$V_{IJ}(h) = v_{W,IJ}(h) + v_{X,IJ}(h) + v_{Y,IJ}(h) + v_{Z,IJ}(h) \quad (8)$$

$$v_{W,IJ}(h) = m_{IJ}(h) / s \quad (9)$$

$$v_{X,IJ}(h) = r_{IJ}(h) * f_{D,IJ}(h) / f_{IJ}(h) \quad (10)$$

$$v_{Y,IJ}(h) = T_{IJ}(h) * d_{IJ} * (1 - a_d) \quad (11)$$

$$v_{Z,IJ}(h) = T_{IJ}(h) * c_{IJ} / p \quad (12)$$

4.4 Traffic Assignment

On most city roads, free-flow speeds are rarely achieved due to congestion. As a result, traffic patterns may significantly change the time costs associated with using a particular route. In conventional travel demand models, vehicle trips are allocated to road networks using a traffic assignment algorithm that captures the impact of congestion on travel time and route choice. In order to simulate traffic under various rideshare adoption scenarios, which enables us to assess the impact of ridesharing on urban travel conditions, we distribute trips on the roadway network using Incremental Traffic Assignment (ITA) [3, 29].

Our ITA 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.

⁵ride-sharing of users with similar origins and destinations

⁶ride-sharing of users sharing portions of their paths between dissimilar origins and/or destinations, such that additional passengers can be picked up en-route

⁷www.bridj.com

Although this incremental approach allows us to capture the impact of congestion on travel times of each subsequent batch, once a driver has been assigned a route it does not change. Consequently the approach does not converge to Wardrop’s equilibrium⁸ even for very small increment sizes. To reach an optimal routing solution, an equilibrium-based approach is necessary. Despite its shortcomings, however, ITA is attractive for its ease of implementation.

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 [11, 38, 4]. 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. 13. The Bureau of Public Roads’s (BPR) default guidelines use $\alpha = 0.15$, $\beta = 4^9$.

$$t_{current} = t_{freeflow} \cdot (1 + \alpha(V/C)^\beta) \quad (13)$$

As often done in traffic assignment modeling, we modify the default coefficients of the BPR function in order to better represent local traffic patterns. By comparing the reported travel time from the Massachusetts Household Travel Survey (MHTS) with that of MHTS vehicle trips assigned using ITA, we select $\alpha = 0.85$ (compared with the default value of 0.15), but maintain the default value for β (4). $\alpha = 0.85$ is in line with transportation literature, which typically increases the value of α for highways and major roads. Lastly, we underestimate total travel time using ITA because we do not apply time penalties for intersection and traffic light stops and queues. To roughly account for such delays, we add two minutes to all travel times. Detailed results and validation of these methods can be found in [39].

5. RESULTS

5.1 Change in Vehicles

Identifying aggregate trends across all hours of the day and OD pairs enables us to draw conclusions that allow us understand the impacts of ridesharing in other cities with different travel patterns and behavior. With this in mind, we estimate a model to capture aggregate daily impacts of ridesharing on the number of network-wide vehicle trips.

To help define the functional form of this model, we first analytically derive the percent change in vehicles $\Delta V_{IJ}(h)$ for a given OD pair and hour relative to the baseline scenario with no rideshare adoption. Specifically, the percent change in vehicles is derived using Equations 8 through 12, with the numerator equal to the change in the number of vehicles due

to ridesharing, and the denominator equal to the number of baseline vehicles, as shown in Equation 14. This percent change in vehicles simplifies to the formulation shown in Equation 15.

$$\Delta V_{IJ}(h) = \frac{V_{IJ}(h) - (T_{IJ}(h) * d_{IJ} + v_{Z,IJ}(h))}{T_{IJ}(h) * d_{IJ} + v_{Z,IJ}(h)} * 100 \quad (14)$$

$$\Delta V_{IJ}(h) = \beta_{IJ}(h) * [(s - 1) * a_d * d_{IJ} - a_o * o_{IJ}] \quad (15)$$

$$\text{where, } \beta_{IJ}(h) = \frac{e_{IJ}(h)}{-s * (d_{IJ} + c_{IJ}/p)} * 100$$

In other words, Equation 15 shows that the change in vehicles for a given hour and OD pair is proportional to the difference between the number of driver and non-driver adopters of the ridesharing service, and a parameter $\beta_{IJ}(h)$. $\beta_{IJ}(h)$ describes the relationship between the efficiency of the rideshare system $e_{IJ}(h)$ (as defined in Equation 7) and the share of vehicle trips given no rideshare adoption. Accordingly, $\beta_{IJ}(h)$ will be larger for OD pairs and time periods with higher rideshare matching efficiency and higher shares of non-driving trips, and result in a greater change in vehicles due to ridesharing. As the number of existing vehicle trips approaches zero, $\Delta V_{IJ}(h)$ and $\beta_{IJ}(h)$ tend to infinity, however, for most US cities the majority of trips are made in vehicles these values are bounded in practice.

We first use Equation 15 to calculate the change in vehicles for a given OD pair and time period, then generalize this relationship (for $s = 2$) to model aggregate results in any city using the model described by Equation 16. With the change in vehicles calculated in Boston for all hours and OD pairs, we empirically estimate $\beta = -59.22$, minimizing the mean squared error between the data and model predictions. Equation 17 describes this model, such that the total percent change in vehicles ΔV can be calculated for adoption rates a_o and a_d , given $\beta = -59.22$, and aggregate Boston mode shares of $d = 0.7003$ and $o = 0.0846$. Figure 3 illustrates the fit of the model to the data.

$$\Delta V = \beta * (a_d * d - a_o * o) \quad (16)$$

$$\Delta V \approx -59.22 * (a_d * 0.7003 - a_o * 0.0846) + 1.75 \quad (17)$$

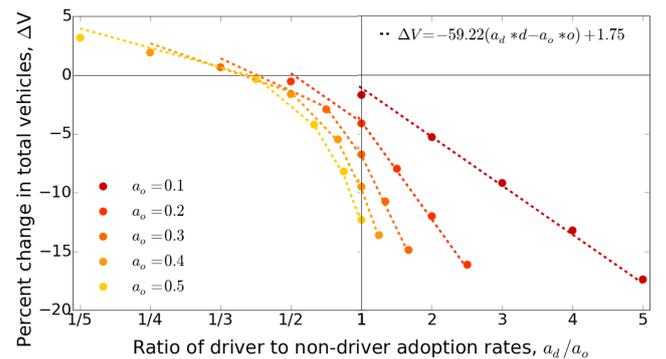


Figure 3: Percent change in total vehicles ΔV relative to the ratio of driver and non-driver adoption rates a_d/a_o . ΔV is proportional to the difference between driver and non-driver rideshare trip shares ($a_d * d - a_o * o$), as described by the model: $\Delta V = -59.22 * (a_d * d - a_o * o) + 1.75$.

⁸The journey times on all used routes are equal and less than those which would be experienced by a single vehicle on any unused route. In other words, each user seeks to minimize his cost of travel, and equilibrium is reached when no user may lower his travel cost through unilateral action.

⁹Travel Demand Modeling with TransCAD 5.0, User’s Guide (Caliper., 2008).

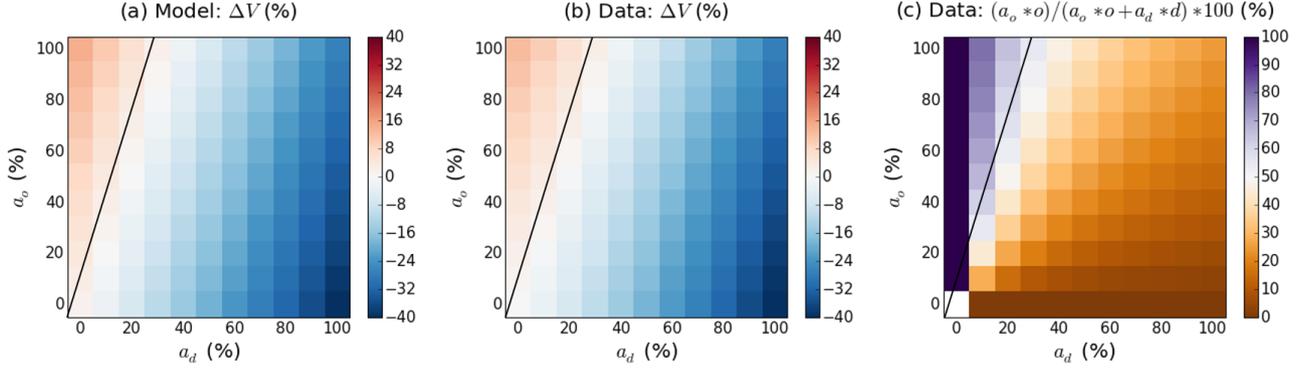


Figure 4: (a) Percent change in total vehicles ΔV relative to the ratio of driver a_d and non-driver adoption rates a_o as estimated by the model $\Delta V = -59.22 * (d * a_d - o * a_o) + 1.75$. (b) Percent change in total vehicles ΔV relative to the ratio of driver a_d and non-driver a_o adoption rates from the data. (c) Percentage of ridesharers that diverted from non-driving modes $(o * a_o) / (d * a_d + o * a_o) * 100$ relative to the ratio of driver a_d and non-driver a_o adoption rates from the data. The black line on each plot is described by $a_o = 3.26 * a_d$, approximately representing $\Delta V = 0$.

Note that the model described by Equation 17 and illustrated in Figure 3 includes an intercept parameter, which increases the estimated percentage change in vehicles by 1.75%, providing a better fit of the Boston data. At the extremes, the model suggests a 43% decrease in vehicles for 100% driver adoption and 0% non-driver adoption, and a 14% increase in vehicles for 0% driver adoption and 100% non-driver adoption in Boston (as compared with -40% and +13% change in vehicles from the data, respectively).

As shown in the right side of Figure 3, when the number of ridesharing adopters from drivers is greater than from non-drivers ($a_d * d / a_o * o > 1$), there is a reduction in vehicles ($\Delta V \leq 0$). Given the average mode shares of drivers ($d = 70.0\%$) and non-drivers ($o = 21.5\%$) in Boston, this relationship results in an overall reduction in vehicles for $a_o \lesssim 3.26 * a_d$, as illustrated by the point on the x-axis at which the data crosses the $\Delta V = 0$ in the left side of the figure.

Moreover, $\beta = -59.22$ captures the inefficiency of the ridesharing system in Boston. Generalized from $\beta_{IJ}(h)$ in Equation 15, the β parameter in Equation 16 is described by Equation 18. Assuming perfectly efficient ($e = 1$) rideshare matching in Boston, it follows that $\beta_e \approx -67.65$, as shown in Equation 19. As expected, $\beta = -59.22$ is smaller in magnitude than $\beta_e = -67.65$, equating to an average efficiency e of approximately 88%.

$$\beta = \frac{e}{-2 * (d + c/p)} * 100 \quad (18)$$

$$\beta_e \approx \frac{1}{-2 * (0.7003 + 0.0846/2.18)} * 100 \approx -67.65 \quad (19)$$

The relationship between a_d , a_o , and ΔV is further illustrated by Figure 4, with ΔV as estimated by the model and calculated from the data shown in Figure 4a and Figure 4b, respectively. White cells have no change in vehicles ($\Delta V = 0$), and the black line with a slope equal to 3.26 is a contour line approximately representing scenarios with no change in vehicles from the model. Figure 4c illustrates the percentage of ridesharers that diverted from non-driving modes for each combination of driver and non-

driver adoption rates. Here, white cells illustrate scenarios where ridesharers diverted from driving and non-driving modes equally.

5.2 Change in Traffic

Using the methods described in Section 4.4, we simulate traffic patterns to assess the network-wide impacts of ridesharing in the peak weekday evening hour for a few adoption scenarios. Table 1 summarizes the resulting percent change in vehicles, vehicle miles traveled (VMT), vehicle hours traveled (VHT), and congested travel time (minutes spent in non-free flow driving conditions). These percentage differences are relative to the base case of no ridesharing, with 874,324 vehicles traveling an average of 6.58 miles in 18.64 minutes (4.78 of which are spent in congestion).

Again, the actual percent change in vehicles for the three adoption scenarios shown in Table 1 (5.99%, -1.83%, and -19.17%) are similar to those we can estimate using the model in Equation 17 (4.26%, -1.90%, and -18.99%).

We see a smaller change in total VMT than total vehicles, suggesting that ridesharing is more efficient in shorter distance, urban markets. Meanwhile, the percent changes in VHT are more significant than for VMT, suggesting that the increase in rideshare efficiency in these markets is somewhat counteracted by the fact that they experience more congestion than longer distance markets.

a_d (%)	a_o (%)	Vehicles (%)	VMT (%)	VHT (%)	Congested TT (%)
0	50	5.99	1.83	3.02	7.16
10	10	-1.83	-0.85	-1.43	-2.98
50	0	-19.17	-11.57	-17.55	-37.30

Table 1: Percent change in vehicles, vehicle miles traveled (VMT), vehicle hours traveled (VHT), and congested travel time (TT) relative to drive-alone/taxi and other non-auto adoption rates $a_d, a_o = 0$. Results are for peak hourly evening (3-7pm) trips, $s = 2$, and $\Delta = 6$.

Lastly, the percent changes in congested travel times reflect the relationship between road segment volume and travel time as captured by the BPR volume delay function; changes in vehicle demand have an exponential impact on travel times under congested conditions. This trend is demonstrated in the 50% driver adoption scenario, with a decrease in congested travel time (37%) nearly double the decrease in vehicles (19%).

In general, these trends suggest that under moderate to high levels of rideshare adoption, ridesharing services would have a noticeable impact on urban traffic conditions.

6. CONCLUSIONS

This research explores the extent to which ridesharing services impact network-wide congestion using mobile phone records. To-date, other research efforts have used partial travel demand (i.e. of taxi or commuting trips) to estimate the proportion of trips that can be pooled for ridesharing under explicitly-defined spatio-temporal constraints. In contrast, we estimate aggregate, total daily travel patterns using Call Detail Records (CDRs) and explore different scenarios of adoption rates to estimate ridesharing demand.

Further, we assess the impact of relative levels of rideshare adoption from auto and non-auto travelers on vehicle usage and traffic congestion. When the number of ridesharing adopters from drivers is greater than from non-drivers, there will be a reduction in total vehicles, and vice versa. In Boston, given the aggregate mode shares of drivers and non-drivers, this translates into a reduction in vehicles when the non-driver adoption rate is less than about three times the driver adoption rate.

The magnitude of this change in vehicles varies spatially and temporally, depending on the distribution of trips and mode shares. This stems from the fact that a ridesharing service will not be able to match all potential trips with one another, resulting in a system that cannot operate at perfect efficiency. In this work, we assume that any customer who cannot be matched will be turned away, representing uncaptured rideshare demand. Using data for Boston, we estimate a parameter to capture the average efficiency of the rideshare service across all OD pairs and hours, enabling us to define a model to estimate the total change in vehicles given auto and non-auto rideshare adoption rates and aggregate mode shares. Future research should explore this relationship in other cities with different demand and travel mode distributions. For cities with land use characteristics that result in heterogenous trip patterns, such as multiple major employment centers and sprawling or sparse residential development, we expect rideshare efficiency to decrease.

Lastly, by simulating traffic for several rideshare adoption scenarios, we evaluate the impact of ridesharing on cumulative vehicle travel time and distance. We find that under moderate to high adoption rate scenarios, ridesharing would likely have noticeable impacts on congested travel times, indicating the importance of incorporating traffic simulation into ridesharing studies. Future work could explore variable adoption rates dependent on trip attributes, such as distance or time of day, as well as socioeconomic characteristics of trip origins. Further, the sensitivity of rideshare demand to variable waiting times would provide important insight into the impact of this threshold on the efficiency of trip matching.

7. ACKNOWLEDGMENTS

This work was partially funded by the U.S. Department of Transportation New England UTC Y25, the Austrian Institute of Technology, and the MIT Portugal Program. We thank Airspace for data and technical support.

8. REFERENCES

- [1] *2014 TIGER/Line Shapefiles*. www.census.gov/cgi-bin/geo/shapefiles2014/main.
- [2] American Community Survey, 2006-2010 5-year estimates. http://www.census.gov/acs/www/data_documentation/2010_release/.
- [3] Transportation Research Board National Cooperative Highway Research Program. Report 716: Travel Demand Forecasting: Parameters and Techniques, 2012.
- [4] R. 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.
- [5] L. P. Alexander, S. Jiang, M. Murga, and M. C. González. Validation of origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies (2015)*, in press, 2015.
- [6] A. Anas. Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological*, 17(1):13–23, 1983.
- [7] Y. Asakura and E. Hato. Tracking survey for individual travel behaviour using mobile communication instruments. *Transportation Research Part C: Emerging Technologies*, 12(3):273–291, 2004.
- [8] H. 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.
- [9] M. Ben-Akiva and M. Bierlaire. Discrete choice methods and their applications to short term travel decisions. In *Handbook of transportation science*, pages 5–33. Springer, 1999.
- [10] M. E. Ben-Akiva and S. R. Lerman. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.
- [11] D. Branston. Link capacity functions: A review. *Transportation Research*, 10(4):223–236, 1976.
- [12] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.
- [13] N. Caceres, J. Wideberg, and F. Benitez. Deriving origin destination data from a mobile phone network. *Intelligent Transport Systems, IET*, 1(1):15–26, 2007.
- [14] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira Jr, and C. Ratti. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26:301–313, 2013.
- [15] J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási. Uncovering individual and collective human dynamics from mobile phone

- records. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224015, 2008.
- [16] S. Çolak, L. P. Alexander, B. G. Alvim, S. R. Mehndiretta, and M. C. González. Analyzing cell phone location data for urban travel: Current methods, limitations, and opportunities. *Transportation Research Record: Journal of the Transportation Research Board (2015)*, in press., 2015.
- [17] Central Transportation Planning Staff (CTPS). Model-based highway data. http://www.ctps.org/Drupal/data_resources, 2010.
- [18] B. Cici, A. Markopoulou, E. Frias-Martinez, and N. Laoutaris. Assessing the potential of ride-sharing using mobile and social data: a tale of four cities. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 201–211. ACM Press, 2014.
- [19] S. Colak, L. Alexander, B. G. Alvim, S. R. Mehndiretta, and M. Gonzalez. Analyzing cell phone location data for urban travel: Current methods, limitations and opportunities. *Transportation Research Records*, 2015.
- [20] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3, 2013.
- [21] M. C. González, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
- [22] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74, 2014.
- [23] S. Jiang, G. A. Fiore, Y. Yang, J. Ferreira Jr, E. Frazzoli, and M. 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.
- [24] X. Lu, L. Bengtsson, and P. Holme. Predictability of population displacement after the 2010 haiti earthquake. *Proceedings of the National Academy of Sciences*, 109(29):11576–11581, 2012.
- [25] S. Ma, Y. Zheng, and O. Wolfson. T-share: A large-scale dynamic taxi ridesharing service. In *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*, pages 410–421, April 2013.
- [26] MassGIS. Community Boundaries. <http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/towns.html>, 2014.
- [27] Y. Nie, H. Zhang, and W. Recker. Inferring origin–destination trip matrices with a decoupled gis path flow estimator. *Transportation Research Part B: Methodological*, 39(6):497–518, 2005.
- [28] NUSTATS. *Massachusetts Department of Transportation: 2010/2011 Massachusetts Travel Survey*, 2012.
- [29] J. D. Ortúzar and L. G. Willumsen. *Modelling transport*. John Wiley & Sons, Chichester, England, 1994.
- [30] S. Phithakitnukoon, T. Horanont, G. Di Lorenzo, R. Shibasaki, and C. Ratti. Activity-aware map: Identifying human daily activity pattern using mobile phone data. *Human Behavior Understanding*, pages 14–25, 2010.
- [31] L. Rayle, S. Shaheen, N. Chan, D. Dai, and R. Cervero. App-Based, On-Demand Ride Services: Comparing Taxi and Ridesourcing Trips and User Characteristics in San Francisco. University of California Transportation Center Working Paper, August 2014. www.uctc.net/research/papers/UCTC-FR-2014-08.pdf.
- [32] J. Reades, F. Calabrese, and C. 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.
- [33] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(37):13290–13294, 2014.
- [34] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, 2013.
- [35] A. Sevtsuk and C. 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.
- [36] C. Song, T. Koren, P. Wang, and A.-L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.
- [37] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
- [38] H. Spiess. Technical note—conical volume-delay functions. *Transportation Science*, 24(2):153–158, 1990.
- [39] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González. The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 2015.
- [40] U.S. Department of Transportation Federal Highway Administration. 2009 National Household Travel Survey. <http://nhts.ornl.gov/download.shtml>, 2011.
- [41] U.S. Department of Transportation Federal Highway Administration. CTPP 2006-2010 Census Tract Flows. http://www.fhwa.dot.gov/planning/census_issues/ctpp/data_products/2006-2010_tract_flows/index.cfm, 2013.
- [42] P. Wang, T. Hunter, A. M. Bayen, K. Schechtner, and M. C. González. Understanding road usage patterns in urban areas. *Scientific reports*, 2(1001), 2012.
- [43] X. Zhan, S. Hasan, S. V. Ukkusuri, and C. Kamga. Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C: Emerging Technologies*, 33:37–49, 2013.