UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-Graduação em Ciência da Computação

Átila Martins Silva Júnior

Cost-Benefit Analysis of a System that Integrates Mass Transportation with Ridesharing

Belo Horizonte
2020

Átila Martins Silva Júnior

Cost-Benefit Analysis of a System that Integrates Mass Transportation with Ridesharing

Final Version

Thesis presented to the Graduate Program in Computer Science of the Federal University of Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Jussara Marques de Almeida Gonçalves
Co-Advisor: Humberto Torres Marques Neto

Belo Horizonte
2020

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg Lucas Cruz CRB 6ª Região nº 819.

Silva Júnior, Átila Martins.

S586c Cost-Benefit analysis of a system that integrates mass transportation with ridesharing / Átila Martins Silva Júnior. — BeloHorizonte, 2020.
xx, 51f. ; il.; 29cm.

Dissertação (mestrado) - Universidade Federal de Minas Gerais – Departamento de Ciência da Computação.

Orientadora: Jussara Marques de Almeida Gonçalves.
Coorientador: Humberto Torres Marques Neto.

1. Computação – Teses. 2. Modelagem de desempenho – Teses. 3. Cidades inteligentes – Teses. 4. Transporte urbano – Teses. 4. Mobilidade urbana – Teses. I. Orientadora. II. Coorientador. III.Título.

CDU 519.6*24(043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Cost-benefit Analysis of a System that Integrates Mass Transit with Ridesharing

ÁTILA MARTINS SILVA JÚNIOR

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

Jussara Marques de Almeida Gonçalves

PROFA. JUSSARA MARQUES DE ALMEIDA GONÇALVES - Orientadora
Departamento de Ciência da Computação - UFMG

Humberto Marques Neto

PROF. HUMBERTO TORRES MARQUES NETO - Coorientador
Departamento de Ciência da Computação - PUC-MG

Artur Ziviani

PROF. ARTUR ZIVIANI
Ciência da Computação - Laboratório Nacional de Computação Científica

PROF. CLODOVEU AUGUSTO DAVIS JÚNIOR
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 23 de Junho de 2020.

"*É preferível a angústia da busca à calma da acomodação.*"
(Dom Resende Costa)

Resumo

O sistema de transporte metropolitano das grandes cidades é composto pelo sistema de transporte coletivo e pelos serviços de transporte particular. O sistema de transporte coletivo tem como principais características a alta ocupação dos veículos, a longa duração das viagens e o baixo custo para os passageiros. Por outro lado, os serviços de transporte particular, como o táxi, possuem baixa ocupação dos veículos, as rotas são flexíveis e o preço de longas viagens pode ser elevado se comparado com o transporte coletivo. Visando reduzir o preço das viagens do transporte particular, atualmente, esses serviços permitem o compartilhamento de viagens e divisão dos custos do trajeto compartilhado entre os passageiros adicionais. Embora as modalidades pública e privada de transporte se interceptem no tempo e espaço, a integração entre elas ainda é pouco explorada.

Esta dissertação tem como objetivo avaliar a viabilidade de integração dos sistemas de transporte coletivo e particular compartilhado. Sendo assim, as contribuições desse trabalho foram as seguintes. Primeiramente, foi realizada uma caracterização dos dados de viagens de táxi e transporte coletivo realizadas na cidade de Nova Iorque em um intervalo de tempo específico. Em seguida, foi proposto o TM-Sharing, um algoritmo que realiza o casamento das viagens desses diferentes modais. Além disso, foram propostas duas políticas de precificação para o sistema integrado de transporte. Por último, foram geradas quatro novas bases de dados sintéticas para avaliar o TM-Sharing em cenários com diferentes aspectos temporais e quantitativos. Os resultados mostram que no cenário de transporte integrado avaliado o passageiro de transporte coletivo pode fazer viagens mais rápidas, enquanto o passageiro de transporte particular realiza viagens mais baratas.

Palavras-chave: Cidades Inteligentes, Mobilidade Urbana, Sistemas de Transporte Multimodal, Compartilhamento de Viagens

Abstract

Taxi and mass transportation services (e.g., buses, subways) are key components of the transportation system of metropolitan areas. Mass transportation tends to be the cheapest alternative, as costs are amortized among many people on the same trip, but they offer the same fixed trajectory. On the other hand, taxi trips tend to be more flexible and faster as they carry fewer people, often in a single direct trip towards the passenger destination. As such, they also tend to be more expensive. Both services often operate completely independent, even though their trips may intercept each other spatially and temporally. Indeed, the integration of mass transportation and taxi services into a unified transportation system has been little explored.

In this thesis, we explored alternatives to combine these two systems focusing on cost and time reduction as the main metrics of interest. Specifically, we characterized data on mass transportation and taxi trips made in New York City in the same period. Then, we designed the TM-Sharing algorithm that matches trips considering temporal, spatial and economic aspects. Then, we proposed two pricing policies for the integrated system. Furthermore, we created four new datasets by inflating the original one to evaluate TM-Sharing in different scenarios. Results showed that in the evaluated scenario of integrated system, mass transportation passengers can save time, while taxi passenger saves money.

Keywords: Smart Cities, Urban Mobility, Multimodal Transport Systems, Ridesharing

Lista de Figuras

3.1 Diagram of the integration between mass transportation and taxisharing. . . . 24

4.1 Graphical result of a transit route computed by OpenTripPlanner. 37
4.2 Trip durations by different modes. 39
4.3 Distributions of computed straight line distances per mode. 40
4.4 Walking distances to access mass transportation modes. 41
4.5 Modes demand per hour of day. 41
4.6 Spatial distributions of mass transportation and taxi passengers origins and destinations. 42
4.7 Possibilities of integration. 44
4.8 Distinct trips varying payment policies. 46
4.9 Pricing divisions ($s\_initial$, s_detour, s_shared). 46
4.10 Segmented and proportional pricing policies. 47
4.11 Max Benefit for Segmented and Proportional Pricing Policies. 48
4.12 Max benefit and synthetic datasets. 50

Lista de Tabelas

4.1 Survey dataset: Number of trips per mass transportation service. 35
4.2 Synthetic Trips. 36
4.3 NYC taxi pricing. 38

Sumário

1 Introduction **11**

1.1 Motivation 12

1.2 Objective 13

1.3 Contributions 14

1.4 Organization 15

2 Related Work **16**

2.1 Ridesharing 16

2.1.1 Identification of Candidate Trips 16

2.1.2 Matching and Scheduling Trips 17

2.1.3 Pricing of Shared Trips 18

2.2 Multimodal Transportation Systems 19

2.2.1 Integrating Different Transportation Modes 19

2.2.2 Mass Transportation and Taxi 20

2.3 Discussion 20

3 TM-Sharing Mechanism **23**

3.1 Overview 23

3.2 Taxi and Mass Transportation Trip Integration 25

3.3 Pricing Policies 31

3.4 Summary 32

4 Evaluation **34**

4.1 Evaluation Methodology 34

4.1.1 Datasets 34

4.1.1.1 Survey Dataset 35

4.1.1.2 Synthetic Datasets 35

4.1.2 Trip Routing 36

4.1.3 Trip Pricing 38

4.2 Characterization of Trips 38

4.3 Spatiotemporal Matching 43

4.4 Maximum Benefit Matching and Pricing 45

4.5 Maximum Benefit and Synthetic Trips 49

5 Conclusion **51**

Referências Bibliográficas **54**

Capítulo 1

Introduction

Thousands of people use mass transportation and on-demand car services to get around in urban centers daily. mass transportation, including bus, subway, and train modes, is the cheapest alternative as costs are divided among many people sharing the same trip. However, they offer fixed trajectories aiming at serving a larger fraction of the population and not necessarily covering the shortest path between different locations. To travel by one of these modes, people need to walk from their origin to a location where a vehicle of the desired line stops for boarding. When disembarking, passengers may need to walk from the nearest line stop to their final destination. Thus, mass transportation services are often time-consuming.

The mass transportation system is composed of different lines, each with a default trajectory and timetable. Thus, it is possible to predict the time a vehicle of a certain line will reach a stop or station for boarding and disembarking of passengers. Some passenger trips demand more than one line to achieve the destination. In such cases, passengers should disembark on a mass transportation stop, walk to the nearest stop of the next line, and wait for boarding a different vehicle. In many cities, the fare paid by passengers on each boarding does not depend on the distance traveled nor on the duration of the trip. Thus, long distances traveled by mass transportation means tend to be cheaper than those traveled by other personalized transportation services.

On the other hand, on-demand car services (such as taxi and Uber[1]) tend to be faster but also more expensive than mass transportation. Due to their flexibility, drivers can choose the best route according to traffic conditions and passenger's desires. These transport services carry passengers between locations of their choice. Pricing is based mostly on the distance traveled and the total time the vehicle remained stopped on traffic, whereas additional costs are often charged during nighttime and peak hours. Currently, smartphone applications allow passengers to request a trip, estimate its cost, and pay the fare using a credit card. Mostly, all these conveniences come with higher costs compared with mass transportation modes. Moreover, the occupancy of such services is often low, further increasing the price per passenger of a trip.

Ridesharing services are cheaper mechanisms that offer passengers with (partially)

[1]https://www.uber.com/

similar routes the chance of sharing the trip and its costs. Yet, to save money, passengers may accept delays caused by the detour to catch and drop-off additional passengers. To work properly, a ridesharing system must keep track of the real-time positions of all registered vehicles, the number of available seats, origins, and destinations of each trip, as well as process passenger requests. By exploiting such data, ridesharing services can match similar routes and ideally quickly meet the demand. Uber Pool[2] and Lyft[3] are examples of such service where passengers request a shared trip through smartphones and the system looks for passengers with similar routes to put together on the same vehicle.

1.1 Motivation

The challenge of ridesharing is to find passengers with similar routes that would like to share their trips. In some cases, it is difficult to find candidate trips for ridesharing due to the lack of tracked trips in some areas at specific periods. Moreover, even if matchings of different trips are detected, the passengers may not accept delays in their trips to get additional passengers. More broadly, ridesharing incentives, such as environmental, monetary, and traffic-related ones, may not be enough to motivate participation.

On the other hand, combining ridesharing services with mass transportation modes can be a good alternative to minimize those drawbacks. Passengers can switch modes to save time and trip costs along their way. Buses, for example, have fixed itineraries. During periods of traffic congestion, when trip delays are expected, passengers may opt for disembarking at a particular station to join a shared car trip, aiming at taking alternative paths and arriving faster at their destinations. Downtown regions tend to be very congested during peak hours. Thus, passengers traveling by taxi services could choose to switch and continue their trips by subway to avoid traffic. In some cases, there might not be a station close to the passenger's destination. In that case, he could benefit from switching transportation mode and completing the last portion of the trip faster with a ridesharing service and drop-off closer to the destination.

The basic prerequisite for an integrated ridesharing system is to keep track of both cars and mass vehicle positions in real-time to be able to identify potential route sharings. Indeed, in many cities, both mass transportation and taxi services are already being tracked by central systems. For example, in large cities such as New York, the real-time position of mass vehicles are provided by the metropolitan traffic agency[4]. Individual

[2]https://www.uber.com/il/en/ride/uberpool/
[3]https://www.lyft.com/
[4]http://datamine.mta.info/

transport services are tracked as well and it is possible to know the occupation of each vehicle and the pick-up and drop-off positions[5]. The knowledge of these positions, in real-time, enables the design of services that combine mass transportation modes with individual transport services in single trips.

Previous studies have analyzed the viability and benefits of the integration between mass transportation and car modes. A preliminary study made by [8] showed that a transportation system that integrates buses with shared and fixed-route taxis is viable when taxi sharing is allowed in low-density areas. [16] designed a generic multimodal system that allows passengers with similar routes and the same mode to share it. Similarly, [5] proposed an optimal matching model to find the best combination of passenger trips for a multimodal transport system. [14] proposed a system that integrates ridesharing with mass transportation modes. However, the authors analyzed the system only from the perspective of the mass transportation passenger, disregarding the individual passenger viewpoint.

Although a variety of studies propose different means of ridesharing and multimodal systems, there is a lack of studies that investigate the costs and the benefits of transportation services that integrate ridesharing with mass transportation modes. Previous work is mostly focused on matching trips of the same mode and integrating different modes of transportation separately. Moreover, due to the scarcity of datasets that contemplate trips made by distinct modes in the same period, multimodal systems are often evaluated based only on synthetic data.

1.2 Objective

The objective of this work is to investigate to what extent the integration of on-demand individual car and mass transportation trips is feasible and economically viable. On one hand, the mass transportation passenger may find that ridesharing offers time savings that compensate for the extra costs (as compared to the mass transportation fares). The individual car passenger, on the other hand, may find it beneficial to share the costs despite a possible extra delay.

We envision the following scenario for such *bimodal* transport service: to request a shared trip, a mass transportation passenger must be registered in the system and connected to the Internet. Before boarding a mass vehicle, passengers must plan their trips, informing their current position and the desired destination location. Then, the integration system will look for route matches where individual on-demand trip passengers

[5]http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

save money while mass transportation passengers save time. The best match is the one in which the delay experienced by the individual car passenger and the extra cost imposed on the mass transportation passenger are minimized. Given such a target scenario, our driving questions are: *What are the costs and benefits for such a system for each party, i.e., mass transportation passenger and on-demand car passenger? Can we devise an integration strategy that delivers a win-win scenario?*

We break down these general questions into the following ones: 1) How are trips made by individual car services and mass transportation modes spatially and temporally distributed? 2) Are there significant overlaps among them that justify ridesharing? 3) How can we integrate ridesharing with mass transportation modes considering the trip duration and the costs imposed on passengers (trip prices)? 4) What would be a fair pricing policy for passengers in the integration scenario? 5) How would an increase in taxi trips impact the number of viable integrations? In this work, we would like to answer these questions analyzing trips from passengers that travel by different modes at the same time and in the same city.

1.3 Contributions

The contributions of this thesis are the following:

1. *Characterization of taxi and mass transportation trips:* To understand the spatial and temporal aspects of these two transportation modes, we analyze real datasets of taxi and mass transportation trips collected in New York City.

2. *An algorithm that matches mass transportation and taxi trips considering spatial, temporal and economical aspects:* To have a match, both taxi and mass transportation passengers should benefit. Taxi passengers must save money and mass transportation passengers must save time on their trips.

3. *Two pricing methods for the integrated system:* The first method is based on the New York City Taxi and Limousine Commission (TLC), composed by initial charge, rate per mile, rate per minute stopped and additional surcharges. In this policy, we define splitting factors to divide the cost of shared segments between taxi and mass transportation passengers. In the second method, the amount saved by taxi passengers is proportional to delay of their trip. Thus, the more taxi passengers deviate from their original route to get additional passengers from mass transportation modes, the more they will save in their trip cost.

4. *Evaluation of our integration method, composed of the proposed algorithm and pricing methods and referred to as TM-Sharing, in real and synthetic scenarios:* The real scenario is built from datasets collected from New York City. However, the total of taxi trips in the real dataset is fairly small. Thus we build synthetic (but realistic) scenarios by inflating the number of taxi trips. By doing so, we aim to evaluate how the proposed system performs as the number of available taxi trips increases.

1.4 Organization

The remaining of this thesis is organized as follows. Chapter 2 presents previous studies related to ours and introduces important concepts as well as algorithms and frameworks of ridesharing and multimodal systems of transportation. Chapter 3 presents our new TM-Sharing integration method. Chapter 4 presents our evaluation methodology and the results of trip data characterization, matching algorithm, and pricing policies evaluation. Finally, Chapter 5 summarizes our thesis and presents possible directions for future work.

Capítulo 2

Related Work

There have been many efforts to improve the efficiency of transportation systems in metropolitan areas. Some studies explore mechanisms to support ridesharing by passengers with similar routes (Section 2.1). Other studies investigate multimodal approaches by exploiting different modes of transportation on the same passenger's trip (Section 2.2). In this chapter, we review prior investigations of ridesharing and multimodal approaches, emphasizing how they deal with the aforementioned challenges.

2.1 Ridesharing

In all cases of ridesharing, any solution must tackle three basic challenges: the identification of candidate trips to be shared (Section 2.1.1), the matching of different trips and the scheduling of the matched trips (Section 2.1.2), and the pricing of shared and integrated trips (Section 2.1.3).

2.1.1 Identification of Candidate Trips

The basic condition for the success of ridesharing is the identification of multiple trips with similar routes (i.e., routes that share, both temporally and spatially, some segment that could be shared). Aiming at identifying candidate trips to be shared, some authors [7, 2] have analyzed mobile phone data (notably Call Data Records, or CDRs) to estimate origin and destination positions of different trips based on the locations of cell towers. Others have exploited data from social networks [7], data collected by GPS equipment at the taxi cars [10, 18, 3] as well as synthetic datasets [1, 9].

2.1.2 Matching and Scheduling Trips

After identifying candidate trips from a data source, the following steps consist of matching and scheduling these trips. To that end, [10] designed a real-time ridesharing algorithm with a service guarantee that considers waiting time and detour to get additional passengers. They analyzed one day of data collected from taxi trips in Shanghai and showed that the proposed kinetic tree algorithm is faster than branch-and-bound and integer programming approaches. However, the uncertainty of delays on traffic was not considered in their approach.

Similarly, [1] developed a framework to select the best candidates to share a ride evaluating constraints imposed by passengers and drivers related to waiting times and trip costs. They developed a dynamic model, named SHAREK, to match trips and used the Euclidean distance to prune candidate trips and reduce the search space. These pruning methods consider both time and costs of candidate trips and find those candidate drivers that are within a circular area around the passenger. The model selects the best driver, who dominates all the other candidates considering waiting time and price. The authors evaluated their method using the synthetic Minnesota Traffic Generator.

From a different perspective, [6] proposed approximation algorithms to assign passengers to drivers by considering that the satisfaction of riders is more important than vehicle travel costs. They formulated the problem of *Utility-Aware Ridesharing on Network Roads*, proved that it is NP-Hard, and designed approximate algorithms to tackle that problem. In general terms, these algorithms aim at maximizing the system's total utility, which is a function of riders' satisfaction, considering constraints of vehicle capacity and the maximum time of passenger's arrival at destination.

[15] developed an approach that considers the individual passengers' discomfort to compute the collective benefits of ridesharing. To that end, they introduced concepts of *shareability*, that is the maximum number of trips that can be shared in one route, and considered the time window each passenger could wait for sharing. They showed that their approach can be applied in scenarios in which the density of trips is high (e.g., which is the case of large cities such as New York City) and, by subsampling the dataset, they showed that good results can be achieved in cities with low-density taxi trips as well.

A big challenge of ridesharing systems is to meet all demand without compromising the global optimality. Due to the high computational effort required to supply all the ridesharing requests in real-time, systems tend to adopt greedy approaches that yield sub-optimal and faster solutions, resulting in missed requests. To tackle that, [13] used historical data to predict the real-time demand and make local decisions yielding to near-optimal global solutions. Thus, the number of unmatched trips reduced significantly.

2.1.3 Pricing of Shared Trips

The pricing policy is another important module of a ridesharing system, which determines how costs are divided among all involved parties and thus influences which candidate trips should be matched to bring benefits to all of them. In that direction, [18] developed a complete ridesharing system, including hardware and software design, and a win-win fare model where both drivers and passengers have monetary incentives to share a ride. They computed the ridesharing benefit based on shared and non-shared route distances. Initial charge and time stopped on traffic are not considered in the benefit computation. They showed that smaller ridesharing distances generate higher profits for drivers and lower costs for riders.

The framework proposed by [3], which is named Auction-based Price-Aware Real-Time (APART), chooses drivers with greater profits and compensates passengers for delays. To do that, both drivers and riders inform their constraints to the system, which tries to satisfy them while maximizing the revenue of the overall requests. APART runs in a distributed platform where each driver's schedule is processed in parallel and the one with the highest profit is selected in real-time.

More recently, [20] modeled an order dispatch system that selects drivers and passengers that maximize the ridesharing platform profit. They showed that this problem is NP-Hard and designed approximation methods. They also developed a framework of a ridesharing system based on real data and conducted experiments to validate their proposal.

To increase platform revenue, [4] proposed a pricing method based on future traffic conditions. For each trip, the method computes the demand at the origin position and estimates the destination one. Thus, a dynamic pricing scheme was designed to incentivize passengers who travel from low to high demand regions.

Beyond monetary incentives, ridesharing also has social and environmental benefits, which have been considered by some authors. For example, [7] considered friendship among passengers to select trips to be shared. They also evaluated the impact of ridesharing on the number of running vehicles. [2] explored the impact of ridesharing on urban traffic by considering the mobility of people by car and on foot. They analyzed how the number of people who choose to walk or to join a shared ride relate to the total number of running vehicles on traffic.

From a different perspective, [11] qualitatively evaluated the experiences of riders and drivers while sharing their trips. They examined how the ridesharing ecosystem contributes to the development of social and cultural capital. They evaluated semi-structured interviews of Uber users and analyzed the growth of social capital. Results showed that drivers and riders can gain informational, emotional, instrumental resources, and companionship by sharing their trips.

2.2 Multimodal Transportation Systems

Beyond sharing the same trip with different passengers, integrating different transportation modes on the same trip can be an economic and fast alternative to get around in big cities. Thus, this section presents works that evaluate the integration of different modes of transportation (Section 2.2.1) and those that focus on the integration of mass transportation with taxi modes, specifically (Section 2.2.2).

2.2.1 Integrating Different Transportation Modes

In a multimodal system, it is possible to combine the multiple transportation modes available in a metropolitan area on the same trip. In that direction, [19] modeled a multimodal transport system that integrates mass and individual modes, such as buses, private cars, bikes and even walking. They used a graph model and applied Dijkstra's routing algorithm, considering only temporal attributes to match trips. Despite the long computational time to run the algorithm, results show that it is possible to combine different modalities of transport into a street network.

Similarly, [5] designed a mathematical approach that allows the integration of private car sharing, ridesharing, buses and trains on the same trip. First, passengers and drivers schedule their trips informing their origin and destination. Then, the system searches for different transportation modes that could be integrated to fulfill the user requests. Results showed that even though the computation effort was high, their system could always find integrated trip solutions.

[16] designed a multimodal system that receives passenger transportation requests, decomposes them into segments according to their areas and modalities, assigns each segment to an available resource in an optimal way, and combines the segments formulating the entire passenger journey. Their work aims to answer passenger requests by creating itineraries composed of mass transportation, ridesharing, and free-use car modes. They modeled the problem as a Multi-Agent System, where the passengers were self-interested agents, having as a goal the maximization of the global utility of the system.

2.2.2 Mass Transportation and Taxi

Some authors have focused specifically on combining mass transportation and taxi services into a bimodal transportation system, as these are the most often used transportation modes (beyond private vehicles). For example, [8] designed a transit network that integrates fixed-route shared taxi and buses. They considered the costs for passengers and drivers as well as environmental and infrastructural costs in their design. The results showed that ridesharing services should be used in areas where population density is low and buses in areas with high population density (e.g., downtown).

[17] modeled the impact of bus passengers when a taxi stands near bus stop locations. They studied the possibility of designing an integrated station for taxi and bus passengers in such a way that a mode does not cause delay to the other. Then, they applied the model on a real road network and computed the probability of passenger queuing.

More recently, [12] developed a multimodal and context-aware transportation recommendation system. In a preliminary study, they analyzed data from different sources, including data related to user behavior, geographical and weather information as well as user profiles and built a framework that recommends unimodal and multimodal routes. The system generates feasible routes, constructs features from datasets, captures the user preference order, and then makes the recommendation.

Finally, [14] developed a multimodal trip planner that integrates the existing mass transportation network with a taxi sharing service. They validated their proposal using a mass transportation planner as a baseline and comparing the advantages to opt for ridesharing or mass transportation modes to complete the trip. Results show that integrating these modes can reduce the trip duration with an acceptable extra cost for passengers.

2.3 Discussion

In this section, we provide an overall discussion of prior studies, highlighting their limitations and how our present effort differs from them. As mentioned, one key component of ridesharing and multimodal systems is the identification of candidate trips to be shared, which requires datasets collected from different sources. Unlike prior ridesharing studies [18, 4], which mostly used real datasets in their evaluations, multimodal studies often lack trip-related data from different modes collected during the same time and thus have to resort to other strategies for evaluating the proposed approaches. For example, [19] used only data from mass transportation line stops and schedule, [8] evaluated their proposal analyzing demographic data, whereas [5] and [14] used randomly generated synthetic data. Moreover, [16] and [17] designed multimodal systems, but did not validate their proposals using real data, whereas [12] evaluated their recommendation system using data generated by users interaction with a map routing service.

In this thesis, we use data collected from real mass transportation and taxi services at the same time and covering the same region to evaluate our proposal. We reconstruct trips from an origin and destination survey made in New York City. The dataset contains trip characteristics such as locations, date and time, and mode of transportation informed by residents of that city. By doing so, we can assess the benefit of our approach TM-Sharing in a real scenario.

After the identification of trips from different modes, the next step is to match them in such a way that the passenger discomfort is minimized [15]. In the design of TM-Sharing, we consider costs and delays as sources of discomfort and exploit distances and trajectories traveled to find the best integration station for the passenger to leave the mass transportation mode and join a shared vehicle. Inspired by [1], we apply a filtering approach based on the aforementioned criteria to reduce the search space of candidate trips.

Another component is the scheduling, which determines the passenger pickup and drop-off sequence the driver should follow. [10] proposed an approach based on tree structures that process ridesharing requests on-the-fly and schedule trips considering the available seats th on vehicle and the maximum waiting time for passengers. Subjected to these constraints, [6] modeled the scheduling problem as a utility-aware ridesharing task that aims to maximize passenger satisfaction. In our approach, unlike previous ridesharing systems that consider only the origin and destination of passengers, the multimodal schedule considers the mass transportation lines and their stops as possible integration positions.

Finally, pricing is a key factor for the success of a ridesharing system. Passengers aim to save money while drivers need to maximize profits from trips. In that direction, [3] proposed a pricing scheme that takes into consideration the passengers' acceptable detours and expected discounts as well as the driver's expected costs (based on traveled distance and duration). In our proposed pricing schemes detours are divided among passengers and taxi drivers are paid proportionally over time and traveled distance. Similar to [18], we take into account the shared and non-shared route aspects in the benefit computation.

Despite the many ridesharing and multimodal transportation solutions available in the literature, alternatives that combine these modalities are scarce. Studies that most closely resemble ours are those by [14] and [12], which proposed alternatives that enable passengers to combine mass transportation modes with on-demand car services in the same route. Our proposal differs from them in the following aspects. In the multimodal trip planner proposed by [14], does not take the taxi passenger perspective into account to select candidate trips for ridesharing. In the multimodal recommendation system developed by [12], integrated taxi-bus trips are recommended, but ridesharing was not considered as an alternative.

Capítulo 3

TM-Sharing Mechanism

A service that combines mass transportation and ridesharing in single trips has the potential to deliver benefits to passengers from both individual transportation services (i.e., taxi services) and mass transportation modes. For taxi passengers, sharing the ride can be cheaper than individual rides [3]. For mass transportation passengers, the integration with taxi services may offer faster trips than a non-integrated system [14]. Moreover, trips from different modalities overlapping each other in both time and space may offer real opportunities for integration [12].

Motivated by these observations, we propose the TM-Sharing mechanism, which integrates taxi and mass transportation services aiming at offering real benefits to all passengers involved. In this section, we first present an overview of the integrated scenario (Section 3.1). We then describe two key components of TM-Sharing, explaining the algorithms that match taxi and mass transportation trips to be shared (Section 3.2) and the pricing policies (Section 3.3).

3.1 Overview

The main scenario for TM-Sharing usage is illustrated in Figure 3.1. The prerequisites are the following: (1) passengers as well as taxi drivers must have smartphones with Internet access and GPS to use our proposed service; (2) they should also accept to share their routes and (3) (preferably real-time) schedule of mass transportation lines in the target region must be available.

Using the a smartphone application, mass transportation (e.g., bus, subway, trolley) passengers may plan their trips choosing lines to travel from their origin to their destination (step 1 in the figure). Similarly, taxi passengers may request trips to available nearby drivers, informing the required number of passengers (or seats) (step 2). Either at trip planning/request time or after the trip has already started, both passengers may opt for integrating and sharing their trip with others to save time or money. Once the

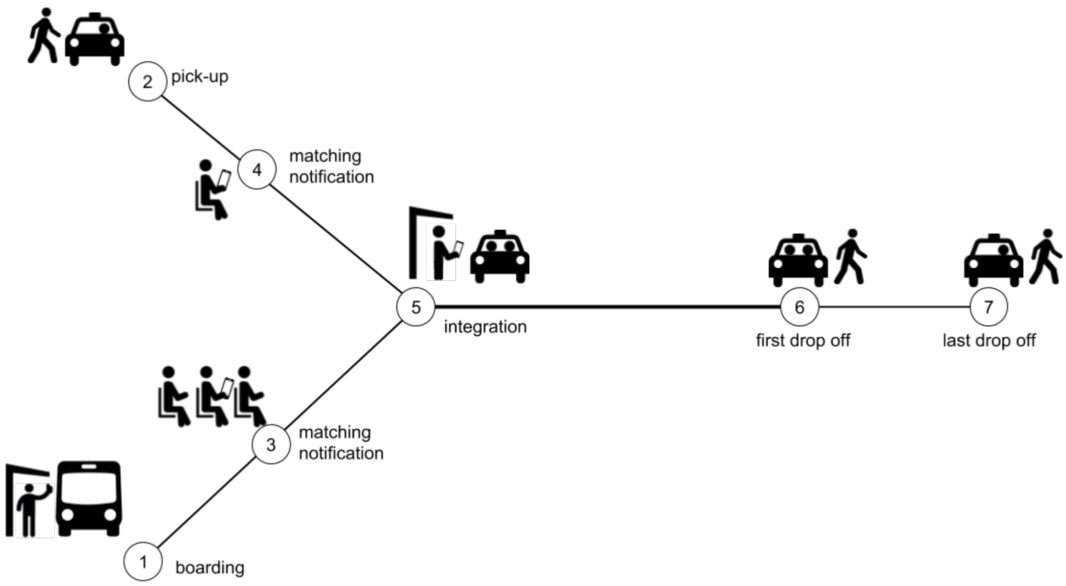


Figura 3.1: Diagram of the integration between mass transportation and taxisharing.

integration is requested, the system will search for trips from different modes (taxi and mass transportation) to be *matched*.

The goal of the TM-Sharing is to attend that specific scenario finding a pair of trips (one from a mass transportation and one from taxi), whose integration guarantees two conditions, namely, that mass transportation passengers save time and that taxi passengers save money. When a match is found, the system sends notifications to all involved passengers informing (estimated) costs and duration of the shared trip (steps 3 and 4). Part of the matching algorithm consists of computing a new shared route including an integration stop (where the mass transportation passenger should pick-up the taxi) and the sequence of destinations, i.e. passengers' drop-off locations. Once an integration option is determined, this will be shown to the taxi and mass transportation passengers. Acceptance from all parties is needed for the integration to be completed. If the driver and all passengers agree on sharing the trip, the system will inform the mass transportation passenger at which line stop (down the trip) she should drop-off and join the taxi ride (step 5) and it will also inform the taxi driver of the new route and where he should get the mass transportation passenger.

Whoever arrives first at the integration stop (mass transportation passenger or taxi driver) should wait for the other party. The ridesharing starts when the mass transportation passenger joins the shared taxi ride. The shared route is computed from the integration stop to the nearest passenger destination (step 6). Next, the trip to the final destination continues as a individual taxi trip (step 7). The drop-off order depends on the locations of the passengers' destinations specified at the time of the request.

Our goal in this thesis is to evaluate the extent to which a system could attend the scenario illustrated in Figure 3.1 bringing benefits (cost or trip duration reductions) to all parties. To that end, we developed the following procedure. Firstly, we characterized real data from passenger trips to understand how people move around in a large metropolitan area, notably New York City, by different modes of transportation and to which extent the opportunities for cost-effective sharing happen. The results of this characterization are discussed in the next chapter. Secondly, we ran our algorithms to match and integrate mass transportation trips with taxi ones. For the same pair of trips, we exploit all possible points of integration, taking the viable taxi trip that is near to a mass transportation stop as a candidate to the integration. We select pairs of trips to be matched where (1) taxi passengers would pay less than traveling only by taxi and (2) mass transportation passengers would travel faster than only by mass transportation modes. Therefore TM-Sharing exploits the trade-off between cost reduction for taxi passengers and time savings for mass transport passengers. If no pair of trips offers those gains, the integration is considered *not viable*[1].

The pricing scheme, that is, how the total costs are split among all passengers, is an important part of the matching algorithm. As a baseline, for individual taxi trips we estimate costs following rules of the Taxi & Limousine Commission (TLC) of the New York City[2]. Specifically, an initial amount is charged at the beginning of the trip, different amounts are charged for miles traveled and stoppeds minute on the traffic. There are additional rates at night and peak hours. We further elaborate on this pricing scheme in Section 4.1.

For shared taxi trips, we analyze different strategies to split the total price of the ride among all passengers sharing it, considering the proportion of the trip that is shared and the extra delay imposed on the (original) taxi passenger. We evaluate how the number of viable integrations (i.e., integrations that benefit all parties) vary for each pricing policy and considering different availabilities of taxi trips.

In the next two sections, we present a detailed description of our integration algorithm and of the pricing schemes we considered in the design of the TM-Sharing.

3.2 Taxi and Mass Transportation Trip Integration

In this section, we describe the main algorithms that compose the trip integration mechanism in TM-Sharing. While describing them, we assume that a set of mass transportation trips M and a set of taxi trips T are given. We assume that a mass trans-

[1]In a real setup, the taxi driver could choose to refuse the integration for personal reasons. In our study, we assume that all candidate integrations are accepted by all taxi drivers, although they are accepted by the passengers only if real benefits (money or time savings) are achieved.

[2]https://www1.nyc.gov/site/tlc/passengers/taxi-fare.page

portation trip is a passenger route composed by one or more mass mode of transportation (subway and/or train) that the passenger is interested in the integration service. Thus, each mass transportation trip is characterized by the origin and the destination of the associated passenger as well as by the stops of the corresponding transit line and the expected time of arrival at each stop. Each taxi trip, in turn, is characterized by an origin and a destination of its passenger as well as a real-time trajectory. In other words, we assume that the service has real-time information about the location of all taxi trips. Moreover, in our evaluation we assume that this information, as well as the arrival times of the mass transportation trips at all line stops, are precise[3]. We also assume that all taxis have seats for an extra (mass transportation) passenger and that the taxi passenger always accepts a request for integration (step 4 in Figure 3.1).

In our analysis, we focused on find incentives for passengers to share their trips and ensuring that taxi drivers will have no financial loss. Thus, our algorithm looks for means to make the trip integration viable for passengers, keeping drivers subject to the existing pricing rules. The consequence of our method is increasing the duration of taxi trips and making the taxi mode more accessible by increasing the occupancy of the vehicles.

In addition to the sets of trips T and M, another key input parameter is d, the maximum distance between the location of the taxi and the mass transportation stop where a potential integration can happen. The idea is that only taxis that are within such distance of a stop are candidates for integration (at that stop).

In general, our taxi-mass transportation trip integration mechanism, referred to as TM-Sharing, is composed of three steps. First, pairs of trips that could be integrated considering temporal and spatial attributes are matched by a *Spatiotemporal Matching* algorithm. Second, the integrations that are economically viable are filtered by a *Viability Filtering* algorithm. Third, from all integration possibilities of the same trip pair, the one with the maximum time and cost benefits is selected by the *Maximum Benefit* procedure.

The following algorithm matches mass transportation and taxi trips.

[3]We leave to future work an evaluation of the impact of imprecisions on the effectiveness of our approach.

Algorithm 1 Spatiotemporal Matching

```
procedure ST-MATCHING

    ▷ maximum integration distance d; set of taxi trips T; set of mass transportation trips M
    Input: d, T, M

    ▷ set of pairs of taxi and mass transportation trips B
    Output: B

    ▷ consider each mass transportation trip m
    for each m in M do

        ▷ consider each stop in the line corresponding to trip m
        for each stop in m do

            ▷ identify ongoing taxi trips
            T = taxis_on_route(stop.datetime, T)

            ▷ consider only taxi trips within d meters of stop
            T = taxis_around(stop, d, T)
            for each t in T do

                ▷ select the nearest taxi position in t from transit stop as
                ▷ the taxi passenger acceptance position
                acceptance = nearest_neighbourhood(stop, t)

                ▷ compute the distance between stop and the taxi acceptance position
                integration_distance = distance(acceptance.pos, stop.pos)

                ▷ compute the distance from acceptance position to taxi destination
                individual_distance = distance(acceptance.pos, t.destination.pos)
                if stop.datetime < t.destination.datetime
                and acceptance.datetime < m.destination.datetime
              and integration_distance < individual_distance then

                    ▷ compute the shared route b
                    b = OTP(stop, acceptance, m.destination, t.destination)

                    ▷ add shared route b to the solution set B
                    B = B + {b}

    Return: B
```

The *Spatiotemporal Matching* algorithm takes as input the parameter d, which is the maximum acceptable distance between candidate trips; as well as the sets T and M of taxi and mass transportation trips, respectively. It produces as output a set B of candidate trips to the integration, each integration b in B is a pair of taxi and mass transportation trips along with a location (mass transportation stop) where the integration is performed and the time as well. The generation of set B takes into account only spatial and temporal constraints. For example, only trips that overlap in time and space can be effectively integrated. Similarly, only integrations that can take place before the taxi passenger arrives at her destination are worth pursuing.

The algorithm iterates over each mass transportation trip m (line 3). For each stop of the mass transportation line trip, m will still pass through (line 5), the algorithm considers only taxi trips that are carrying passengers at the time the mass transportation vehicle passes by that stop (line 7) and that are at most within d meters (euclidean distance) from the stop position (line 9), we here assume $d = 3km$. Then, for each filtered taxi trip t (line 10), the algorithm selects the location along the taxi trip trajectory t that is nearest to the bus stop (line 13). Let's call the location the *acceptance location* as it will be the position where the taxi is when integration is accepted (in case it indeed is). The algorithm then computes two distances, namely, the distance between the acceptance location and the target bus stop (line 15) and the distance still to be covered by the taxi if it continues as a individual trip, i.e., if no integration occurs (line 17).

The taxi trip will be selected as a candidate for the integration if (i) the taxi trip does not arrive at its destination before the mass transportation reaches the stop; (ii) taxi passenger accepts the integration before the mass transportation passenger would arrive at his destination by the mass transportation line; and (iii) the distance between taxi acceptance and the nearest transit stop is shorter than the distance between the acceptance position to the taxi passenger's destination (line 18). If these three conditions are met, the integration route is computed (line 20). The algorithm returns a set of pairs of mass transportation and taxi trips (B) candidates to the integration (22).

Some candidate trips may not be viable considering trip costs and the duration of the integration. The *Viability Filtering* algorithm filter viable integrations taking as input the result from Algorithm 1 (B), a set T of individual taxi trips, and a set M of mass transportation trips. The algorithm returns a set C of trips pairs (mass transportation and taxi) that are temporal and economically viable to the integration.

Algorithm 2 Viability Filtering

1: **procedure** VIABILITY-FILTERING

▷ set of integrated trips B; set of taxi trips T; set of mass transportation trips M
Input: B, T, M

▷ set of viable shared trips C
Output: C

2: ▷ consider each shared trip b
3: **for** each b in B **do**

4: ▷ recover original taxi and mass transportation trips
5: $t = get\_trip(b.taxi.id, T)$

6: $m = get\_trip(b.transit.id, M)$

7: ▷ compute the original cost of the (individual) taxi trip
8: $taxi\_individual\_cost = compute\_taxi\_cost(t)$

9: **if** $b.taxi.duration < \alpha * t.duration$
and $b.transit.destination.datetime < m.destination.datetime$
and $bimodal\_costs.taxi < taxi\_individual\_cost$ **then**

10: ▷ add shared route b to the solution set C
11: $C = C + \{b\}$

Return: C

For each candidate integration b (line 3), the algorithm determines its viability as follows. A candidate integration b is viable if the following three conditions are met: (i) we consider an upper limit on the total duration of the integrated trip for the taxi passenger to avoid excessively long delays. We assume that the total duration of the integrated trip for the taxi passenger should be less than α times the duration of the individual trip, we here assume $\alpha = 2$, we here consider the waiting time at the integration station; (ii) the mass transportation passenger should save time in the integrated trip; (iii) the taxi passenger should save money in the integration trip (line 9). The algorithm returns a set of viable integrated trips (C) including repeated pairs with integrations occurring at different (all viable) stop positions (11).

The last procedure of the TM-Sharing Algorithm is responsible for selecting the best option of integration, out of all viable possibilities (set C). In other words, it selects the pair of mass transportation and taxi trips along with an integration stop (where the mass transportation passenger will disembark and join the shared taxi ride) with the maximum benefit to both parties. Thus, we define the utility of a candidate integration

c ($c \in C$) as the product of the total amount of time saved by the mass transportation passenger if she chooses c (compared to the original trip duration) and the total amount of money saved by the taxi passenger (compared to the original individual trip cost).

Algorithm 3 Maximum Benefit Trip

1: **procedure** MAXIMUM-BENEFIT

▷ set of viable integrated trips C
Input: C

▷ shared trip with maximum benefit to both passengers
Output: best_integration

2: ▷ initialize the maximum utility variable
3: $max\_utility = -\infty$
4: $best\_integration = NULL$

5: ▷ consider each integrated trip c
6: **for** each c in C **do**

7: ▷ compute mass transportation passenger saved time
8: $benefit_{mt} = get\_saved\_time(c)$

9: ▷ compute taxi passenger saved money
10: $benefit_t = get\_saved\_money(c)$

11: ▷ compute the integrated trip utility
12: $integration\_utility = benefit_{mt} * benefit_t$

13: ▷ select the best integration
14: **if** $integration\_utility > max\_utility$ **then**
15: $max\_utility = integration\_utility$
16: $best\_integration = c$

Return: best_integration

For each viable candidate integration, c in set C (line 6), the algorithm computes the benefit of c to the mass transportation passenger (line 8)), to the taxi passenger (line 10) as well as the overall utility of c (line 12). Out of all viable candidate integrations (set C), the algorithm selects and returns the one with maximum utility (line 14-16), or NULL if no such candidate exists ($C = \emptyset$).

3.3 Pricing Policies

Splitting the price of several route segments among the taxi and mass transportation passengers is an important task that impacts directly the viability of TM-Sharing. As shown in Figure 3.1, each trip segment should be priced in such a way that there is economy for taxi passengers, and the price paid for mass transportation passengers cannot be too high. In that direction, we propose two different pricing policies to enable the TM-Sharing mechanism.

We consider pricing policies that are based on the typical pricing schemes employed by individual taxis. The total price is the sum of a fixed initial amount, which is independent of trip duration and distance traveled, plus a variable amount that consists of a rate charged per unit of distance traveled. Moreover, the time stopped on traffic is charged, and additional costs are employed during peak and night hours. Then, we derive policies to distribute the total price of a shared trip among all participating passengers as follows.

Taking Figure 3.1 as a reference, we split the total price of a shared taxi trip into the following components: (a) **initial**, corresponding to the fixed amount charged for initiating the taxi ride; (b) **original route**, corresponding to the price of the taxi route until integration is established (segment between (2) and (4) in Figure 3.1); (c) **detour**, price of the (possible) deviation of the taxi and to pick up mass transportation passenger and the waiting time spent at integration station (segment between 4 and 5); (d) **shared route**, corresponding to the price of the shared route until first destination (segment between 5 and 6) and (e) **final destination**, corresponding to the price of trip between destinations (segment 6-7).

In both pricing policies, segmented and proportional, we assume that (b) is entirely charged to the original taxi passenger while (e) is fully charged to the last passenger to drop-off. Thus, we vary the way the segments (a), (c) and (d) are split between taxi and mass transportation passengers. Specifically, we consider two approaches.

In our first approach, we define splitting factors, $s_{initial}$, s_{detour}, and s_{shared} to be applied to the initial (a), detour (c) and shared route (d) segments, respectively, as follows. A fraction s_i of the price component i (i equal to initial, detour or shared) is charged to the passenger coming from the mass transportation, and the rest is charged to the original taxi passenger. Clearly, the values assigned to parameters $s_{initial}$, s_{detour} and s_{shared} directly impact the cost-effectiveness of the integration.

In our second policy, passengers from mass transportation modes pay proportionally to the taxi passenger extra delay due to integration imposed on the original taxi passengers. The system estimates the total taxi trip duration to the original passenger's destination with integration (t_{taxi}^{new}) and without the integration (t_{taxi}^{orig}). It then computes a splitting factor $f = \frac{t^{new}_{taxi} - t^{orig}_{taxi}}{t^{new}_{taxi}}$ that is a delay incurred in the taxi passenger route due to the integration. The price components (a), (c) and (d) are then split such that mass transportation passenger will pay $(s_{initial} + s_{detour} + s_{shared}) * f$ while the taxi passenger will pay for the rest of the shared trip such as $(s_{initial} + s_{detour} + s_{shared}) * (1 - f)$.

3.4 Summary

We have presented TM-Sharing, a mechanism to integrate mass transit and taxi trips. TM-Sharing takes into account spatial and temporal constraints as well as economical aspects to select viable integrations. TM-Sharing explores the one between time and cost of the integrated trips aiming to benefit both mass transit and taxi passengers. We have also proposed two schemes to split the price of the integrated trip amount both passengers: one considers fixed splitting factors while the other takes the delay imposed to the taxi passenger into account to decide how much should be changed to the mass transit passenger.

An example that illustrates the proposed scenario follows. Suppose that at 6:50 AM a taxi passenger requests for a trip informing her origin and destination position and an acceptable delay for ridesharing (15 minutes). At 6:55 an available taxi arrives at the origin of the passenger and she picks-up on the vehicle. Not far from there, another passenger board on a bus at 7:08 AM. At that time, both passengers are traveling towards their destination in different modes of transportation.

Unexpectedly, at 7:22, an accident slows down the traffic on the avenue where the bus of our passenger passes by. Then, to avoid getting late, the bus passenger requests to a shared trip using his smartphone. He opens the application and informs the bus line he is traveling, the position of his destination, and the max time he wants to arrive at his destination, 8:00 AM. The system then looks for on-rout taxi trips that are around the bus line route and that attends the bus passengers requisites.

The system sends requisitions for all taxi passengers and drivers that do not have their time constraints violated, and that generates acceptable money savings for taxi passengers. Generating and expected delay of 13 minutes and a saving of 5 dollars, our taxi passenger accepts to share her trip with the bus passenger. Therefore, the bus passenger could arrive at this destination at 7:46, saving around 14 minutes. To save that time, he should pay 21 dollars.

If both passengers and the taxi driver accept the ridesharing, the system recomputes the route of the taxi driver passing by a bus station to get the additional passenger. If the bus passenger arrives before the taxi at the station of integration, he should wait to pick-up. If the taxi arrives first, he should wait as well, but that waiting time is charged to passengers. In this example, the bus passenger disembarks at 7:28 and picks-up on the taxi at 7:30. The shared route is from the pick-up of the bus passenger to the drop-off of the passenger with the nearest destination. At this example, the taxi passenger drops-off first, at 7:37, and the bus passenger drops-off last, at 7:46. At this time the trip ends and the system charges the trip fees for both passengers and pays the taxi driver proportionally.

Given that example, in the next chapter, we evaluate our mechanism in the scenario of New York City using a dataset of real trips and generating synthetic trips derived from real ones.

Capítulo 4

Evaluation

In this chapter, we present our evaluation of the proposed TM-Sharing mechanism. We start by introducing the key aspects of our evaluation methodology in Section 4.1, notably the datasets used. Next, we present a temporal and spatial characterization of the trips in our dataset in Section 4.2. We show the results of the evaluation of each component of the proposed mechanism in the following sections.

4.1 Evaluation Methodology

In this section, we discuss the methodology adopted to evaluate the TM-Sharing mechanism. We present the datasets used in our study (Section 4.1.1), describe how we estimate the route taken by a given trip (Section 4.1.2) and how we compute the price of the shared trip (Section 4.1.3).

4.1.1 Datasets

We validated our method on real and synthetic datasets. The real dataset is built from a survey with passengers who traveled by different transportation modes during the same time at a given metropolitan area, i.e., New York City[1]. We used this dataset as the basis to assess the benefit of our proposal. We also used it to generate synthetic datasets aiming at assessing how the number of potential trip integrations increases as the number of taxi trips grows. In the following, we first give an overview of the survey dataset (Section 4.1.1.1) and then describe how we generate synthetic datasets (Section 4.1.1.2)).

[1] http://web.mta.info/mta/planning/data-nyc-travel.html

4.1.1.1 Survey Dataset

Our survey dataset is from the 2008 New York City Customer Travel Survey commissioned by the Metropolitan Transportation Authority (MTA). The survey method was to call by phone and mail NYC residents to ask them characteristics, such as origin, destination, modes, and the purpose of their previous day trips. This survey aimed to understand the travel patterns of New York City residents and to guide improvements on the transportation system. This survey collected 42,900 trips from 16,186 residents from May to November 2008. Out of all trips, 30,743 used some type of mass transportation service (mass transportation and taxi) as shown in Table 4.1[2].

Tabela 4.1: Survey dataset: Number of trips per mass transportation service.

Mode	Number of Trips
Subway	16,453
Bus	8,301
Subway + Bus	4,093
Taxi, car/van service	1,896

In our study, we focus on trips by mass transportation and taxi services only. Subway and buses are the most commonly used mass transportation services in the city. Indeed, trips by either bus or subway or by both modes (subway+bus) comprise the majority of all collected trips (94%). In contrast, taxi-related services correspond to a small part of the trips in our survey data, only 6%.

Each trip record has 136 variables, including census tract codes of origin and destination, mode traveled, date and time of departure and arrival at destination. We estimated the latitude and longitude of each origin and destination location as the centroid of its census tract.

4.1.1.2 Synthetic Datasets

In addition to using the real survey data, we also built synthetic datasets by artificially inflating the number of taxi trips, aiming at assessing how such inflation impacts the number of viable integrations. That scenario portrays the case when there are more trips made by different individual transport services. To keep the general mobility patterns

[2]The table does not account for trips using private vehicles.

observed in the real dataset, we inflated the number of taxi trips by simply replacing each taxi trip in the original data (real trip) by n new synthetic trips with the same spatial attributes (same origin and same destination) but slightly different departure times. Specifically, we replaced a real trip starting at time t by n other trips with departure time uniformly distributed in the period $[t - \delta; t + \delta]$. Then, we recomputed the destination arrival time of each synthetic trip using the OTP mechanism (Section 4.1.2). Such an approach mimics scenarios where the number of taxi passengers moving from place A to place B inflates by a factor of n, and each such passenger departs at around the same time (controlled by window 2δ), but not *exactly* the same.

Tabela 4.2: Synthetic Trips.

Name	n	δ	Total
$5x\_10min$	5	10	$9,480$
$5x\_20min$	5	20	$9,480$
$10x\_10min$	10	10	$18,960$
$10x\_20min$	10	20	$18,960$

As shown in Table 4.2, we built four synthetic datasets by varying n equal to 5 or 10 and by taking δ equal to 10 or 20 minutes. The total numbers of taxi trips in the synthetic datasets increased to 9,480 and 18,960 for n equal to 10 and 20, respectively.

4.1.2 Trip Routing

Only the origin and destination locations of each trip are available in our datasets. Yet, we need to estimate the complete route, with intermediate points on the way, to identify potential candidates for integration. To do so, we used the OpenTripPlanner service[3] which helps passengers planning their trips by offering itineraries that may combine transit, pedestrian, bicycle and private car segments (but no taxi services). OpenTripPlanner computes the best route based on a map of the metropolitan area extracted from OpenStreetMap[4] and the timetable of mass transportation extracted from General Transit Feed Specification (GTFS) files provided by metropolitan transport agencies (e.g. MTA). Figure 4.1 shows an example of a route computed by OpenTripPlaner (OTP). Note that it receives as input the latitude and longitude coordinates of the start and end points as well as departure time.

[3] http://www.opentripplanner.org/
[4] https://www.openstreetmap.org

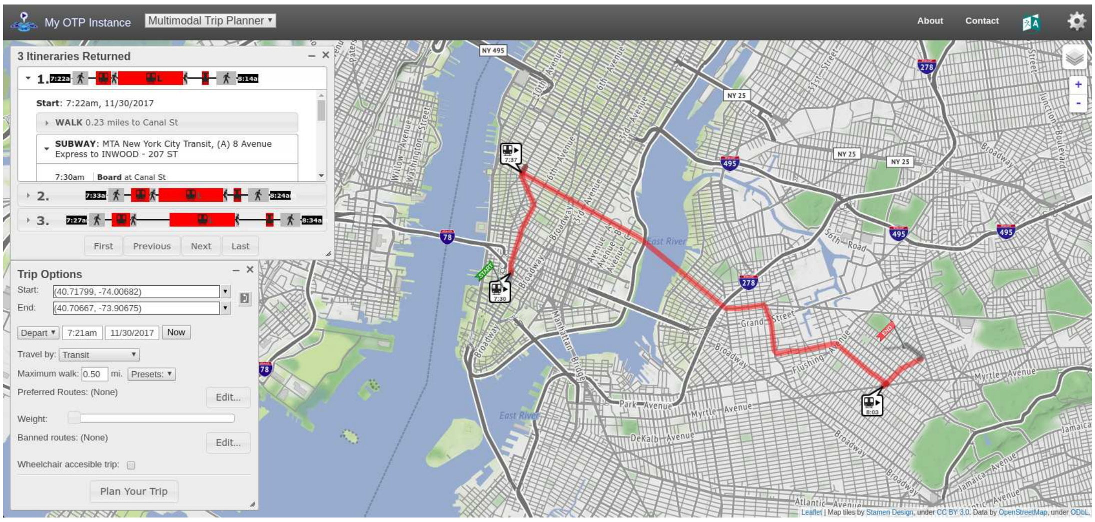


Figura 4.1: Graphical result of a transit route computed by OpenTripPlanner.

OTP was used to compute both taxi and mass transportation trip routes. To compute a route we need to inform the origin and destination positions, date and time of origin and which are the travel modes. The result is composed of a sequence of positions, time and mode. The last position and time are the destination and the arrival time, respectively. The OTP framework computes the best mass transportation route with the fewest integrations. From all possible routes OTP returns, we always choose the fastest one. OTP does not compute the integrated route (taxi and mass transportation) on the same route. To do that we need to compute separately one route for the taxi (car) trip and the other for the mass transportation mode. Then, the result is combined in one integrated route.

Default parameters of OTP were used to compute trip routes. Specifically, mass transportation routes were computed considering passengers' walking distances as short as possible, as few line integrations as possible, and fastest travel time. OTP implements several heuristics to determine the routes, thus they are not guaranteed to be always optimal. From departure time, mode of transportation, origin and destination positions OTP computes the entire route with its intermediate positions and estimated timestamp. Moreover, traffic congestion was not considered in the route calculation.

Initially, the detailed routes for all taxi trips are computed to identify all potential points of integration with mass transportation lines. Once matching is determined, OTP is also used to compute the shared route. For a fair comparison, we use OTP to estimate trip route and duration and, from that result, we compute the price (see next section) for all individual and shared taxi trips.

4.1.3 Trip Pricing

In general, mass transportation has a fixed price per trip while taxi service costs vary according to distance and duration of the trip. Thus, we here describe how we compute taxi trip costs. Our approach is based on the NYC Taxi and Limousine Commission (TLC) policies[5], presented in the table 4.3.

Tabela 4.3: NYC taxi pricing.

Description	Cost
Initial Charge	\$2.50
MTA State Surcharge	\$0.50
Rate per Mile	\$2.50
Rate per Minute Stopped	\$0.40
Peak Hours	\$1.00
Night Surcharge	\$0.50

Initially, a minimum amount is charged per trip, \$2.50, added to a State Surcharge of \$0.50 for all trips that end in New York City and nearby. Once the trip starts, there is a charge of \$2.50 per mile as well as a \$0.40 per minute the vehicle is stopped in slow traffic. In peak hours, from 4 pm to 8 pm, there is a surcharge of \$1.00. Finally, from 8 pm to 6 am, there is a night surcharge of \$0.50. These rules were used to compute the prices of individual trips and used as a reference for computing the taxi sharing prices (described in Section 3.3).

To compute the amount paid by public transport passengers in the integrated trip we do not consider the ticket costs for the following reasons. To have integration, mass transportation passengers must be onboard a public vehicle. Thus, they have already paid for the ticket. We consider that the transfer between transit modes is costless. Therefore, we only consider the additional costs they would be charged to join the shared taxi ride.

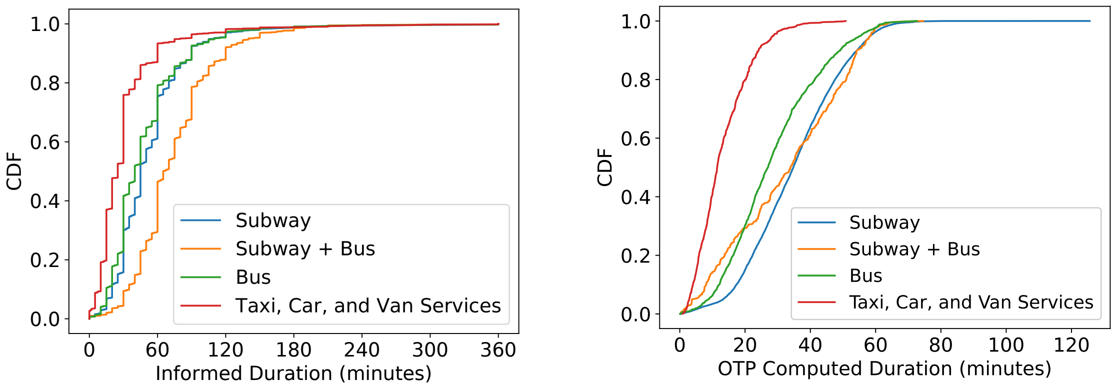
4.2 Characterization of Trips

In this section, we present a characterization of the trips in our dataset, emphasizing differences across transportation modes. We aim to provide an overview of the data used in the evaluation of our approach, which can be considered representative of

[5]http://www.nyc.gov/html/tlc/html/passenger/taxicab_rate.shtml

human mobility in a large metropolitan area. We analyze trip durations on the route and walking distances, as well as spatiotemporal distributions of origin and destination locations. We believe that this analysis can offer insights into opportunities to improve the transportation system as a whole. Specifically, we here search for opportunities for mode integration.

As mentioned, we focus on trips by bus and subway (combined or in isolation) as mass transportation trips, and we refer to trips by van, private car, and taxi services simply as taxi trips. We start by analyzing the distributions of trip durations. Figure 4.2 shows the cumulative distributions of trip durations for each aforementioned transportation mode. This figure shows distributions for durations *informed by passengers*, as captured in our dataset, as well as durations *computed* by running OTP with the origin and destination informed in the dataset.



(a) Informed Durations. (b) Computed Durations.

Figura 4.2: Trip durations by different modes.

As expected, Figure 4.2 shows that trips made by taxi services, serving a individual itinerary, tend to be faster than those made by mass transportation modes, which have more general, predefined routes. Moreover, mass transportation passengers often need to walk to the nearest line stop and wait for the vehicle of their preferred line to arrive. As the passenger takes multiple lines, trip duration may increase accordingly.

As shown in Figure 4.2, computed durations, in general, tend to be faster than informed ones. While the most time-consuming trip in Figure 4.2a took six hours, in Figure 4.2b the slowest one lasted for about two hours. Computed trips consider ideal conditions, delays caused by traffic and unforeseen events are not considered. Yet, we note that durations informed by passengers may not be accurate either, as they may vary depending on memory and perception of each passenger. Despite the large differences between informed and computed durations, for all transportation modes considered, the relative order of the distributions remains mostly the same, except for trips that combine subway and bus, which tend to be the slowest ones, if we rely on informed durations, but, according to computed durations, tend to be faster than trips by subway only.

In addition to duration, we also analyze the distance covered by each trip. To that end, we compute the shortest distance (based on a straight line) between the origin and destination of each trip. Figure 4.3 shows the distributions of these distances.

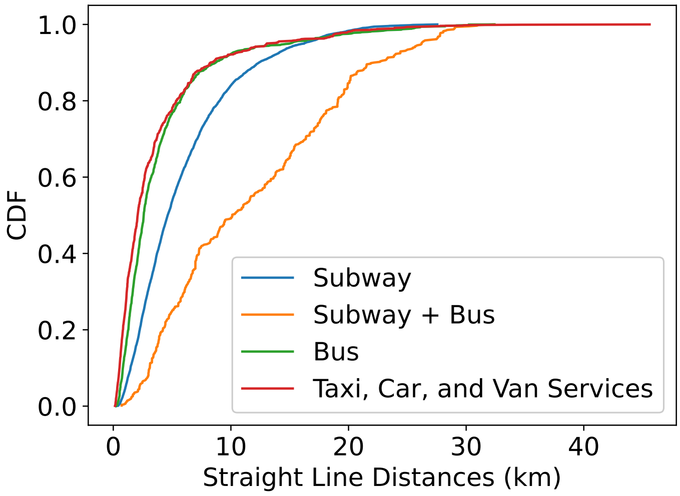


Figura 4.3: Distributions of computed straight line distances per mode.

As shown in Figure 4.3, trips by taxi and buses tend to have shorter distances, with very similar distributions. This observation can be justified by a general trend of passengers to take the subway (only or combined with buses) to cover longer distances. Subways avoid traffic and tend to be faster than buses while also cheaper than taxis, being thus a good alternative for such long trips.

We also analyze the walking distances the mass transportation passengers must cover to reach the stations nearest to their origin and destination. Figure 4.4 shows these distributions for the three mass transportation modes. We note that passengers tend to walk somewhat shorter distances when taking the buses. This may reflect the case that, in New York City, bus stops are often closer to origins and destinations of passengers than subway stations, despite the very long subway network available in the city. We note however that there are exceptions as the walking distances reach as long as 5 km when passengers traveled by bus mode.

In New York City bus stations are closer to the origins and destinations of passengers than subway ones. This result shows that passengers tend to walk less opting to travel by bus then when traveling by subway. Despite having a large subway network, bus infrastructure is easy to be accessed. Passengers that combined bus and subway on the same trip had intermediate walking distances.

Next, we analyze the temporal and spatial distributions of the demand for transportation services through several hours of the day and different regions of the metropolitan area. Figure 4.5 shows the hourly demand for different transportation services, i.e., for

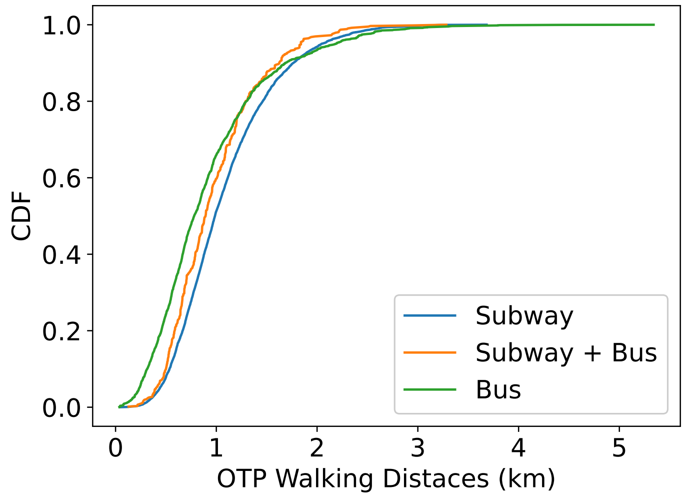


Figura 4.4: Walking distances to access mass transportation modes.

each hour of the day and each transportation mode, it presents the average percentage of the daily trips (by that particular mode) that initiated (boarding time) sometime during that hour. According to the figure, for all four transportation modes, the trips present peaks early morning and late afternoon, most probably reflecting patterns of passengers commuting to work and back home. Note that peaks by taxi tend to be somewhat later, possibly due to the shorter durations. However, we note that, unlike trips by subway (only subway or jointly with a bus), the demands for taxi and bus services tend to remain high, with some variation, throughout the day, dropping at the evening (for taxis, the drop occurs much later in the evening).

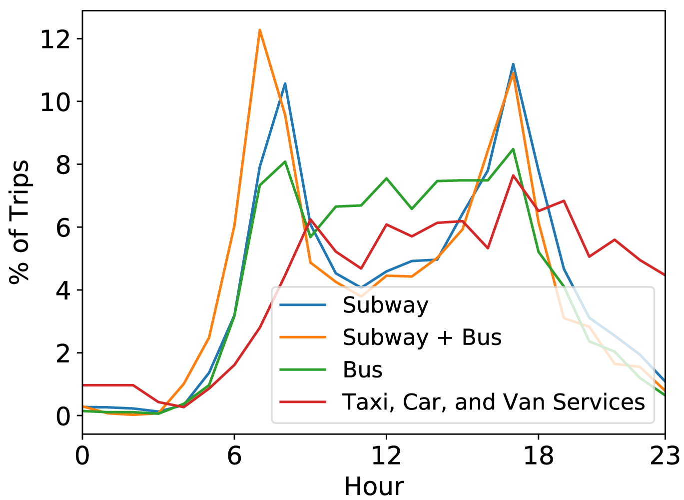
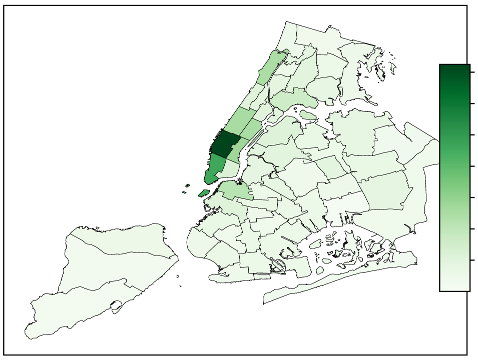
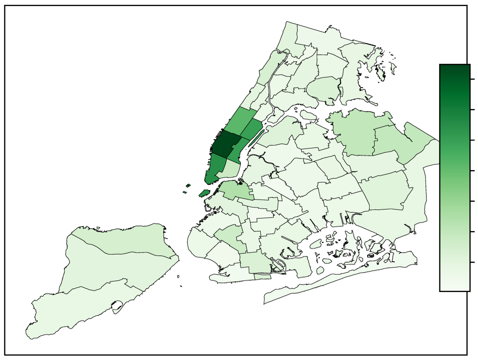


Figura 4.5: Modes demand per hour of day.

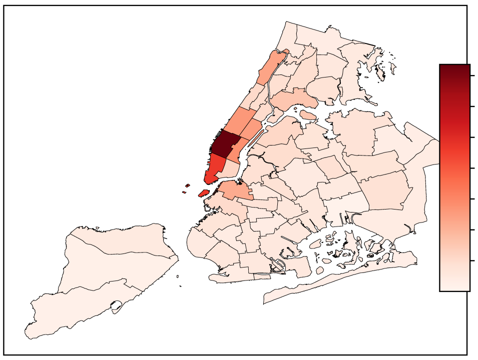
The spatial distribution of origins and destinations enables us to identify regions of greater and smaller demand. Figure 4.6 shows how origin and destination positions are distributed in New York City for mass transportation and taxi trips. The map of NYC was divided into PUMAs (Public Use Microdata Areas) which are regions marked by The United States Census Bureau to provide statistics and demographic information. Each PUMA was designed to contain one hundred thousand people.



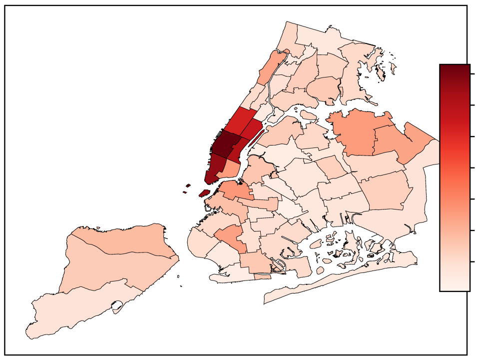
(a) Mass transportation origins.



(b) Taxi origins.



(c) Mass transportation destinations.



(d) Taxi destinations

Figura 4.6: Spatial distributions of mass transportation and taxi passengers origins and destinations.

Figures 4.6a and 4.6c present the spatial distribution of mass transportation passenger origins and destinations. Similarly, the distributions of taxi passenger pick-ups and drop-offs are shown in Figures 4.6b and 4.6d, respectively. For each transportation mode, the regions with more concentration of origins and destinations coincide. This suggests that passengers tend to use the same model both ways. Moreover, we observe much similarity in the distributions for mass transportation and taxi passengers, which may reflect characteristics of the demography in different regions of the metropolitan area (e.g., the greater concentration of population in Manhattan).

In sum, our characterization suggests that the integration of mass transportation and taxi services is a viable and effective alternative in New York City as, despite differences, the trips made by both models often overlap, both in time and space. By matching similar trips we may be able to offer a service that combines the benefits while softening the drawbacks of each modal. The design of TM-Sharing was driven by that goal. In the following sections, we present the main results of the evaluation of each step of the method to show that it indeed can be effective.

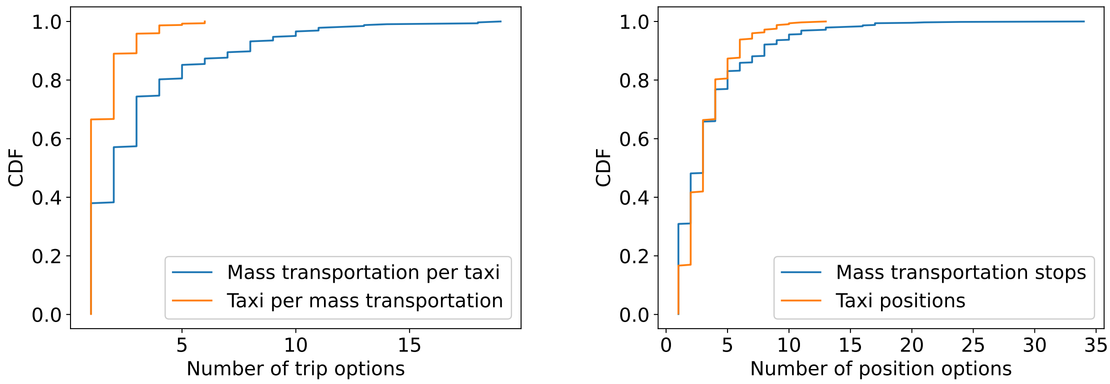
4.3 Spatiotemporal Matching

Having analyzed the characteristics of our dataset, we now turn to the evaluation of our TM-Sharing method, performed by applying it to the analyzed dataset. We aim at integrating mass transportation trips (i.e., subway or bus trips) with taxi trips (i.e., van, private car, and taxi services). In this section, we focus on the first step of our method, i.e., the spatiotemporal matching of candidate trips.

Recall that to integrate a pair of trips we need to match those that benefit both taxi and transit passengers. Initially, we consider spatial and temporal attributes, and then economic aspects. As the input of the Spatiotemporal Matching Algorithm (Algorithm 1), we selected trips computed by OpenTripPlanner including their intermediate positions. Additionally, we considered the parameter d as the maximum detour acceptable to the taxi driver get the mass transportation passengers. We assume d = 3km (euclidean distance) to reduce the search space and eliminate possible unfeasible matchings.

The goal of this first step is to find pairs of candidate trips whose integration benefit mass transportation passengers, i.e., they can save trip time. Later those trips will be filtered out to keep only those that *also* benefit the taxi passenger. We focused our analysis only on the perspective of passengers, ensuring that the driver will not have a financial loss. Therefore, given a mass transportation trip, there may be multiple taxi trips that can be integrated, as there may be multiple mass transportation stops where the integration could take place. Similarly, a given taxi trip may have with multiple mass transportation trips options to integrate.

Figure 4.7 shows the number of possibilities for integration in terms of trips (Figure 4.7a) and stations (Figure 4.7b). Figure 4.7a, in particular, shows the cumulative distribution of the number of candidate taxi trips that could be integrated to each mass transportation passenger (blue curve), as well as the distribution of the number of mass transportation trips that could be integrated to each available taxi trip (orange curve). Figure 4.7b shows, for each candidate match, the cumulative distribution of the number of possible mass transportation stop positions (blue curve) where the integration can take place and the possible taxi trip positions (orange curve) where the trip can deviate from the individual route to the shared route.



(a) Trip options.

(b) Position options.

Figura 4.7: Possibilities of integration.

Figure 4.7a shows in blue the cumulative distribution of candidate taxi trips that could attend a transit passenger, and in orange are the number of transit trips that could attend a taxi passenger integration request. In total, there are 10,173 possible integrations, built from 324 unique taxi trips and 673 unique transit trips. As Figure 4.7a shows, for around 60% of the taxi trips, there is more than one integration option, and for around 20% of them, there are more than five options. In contrast, for almost 70% of the mass transportation trips no integration option is available in the dataset, that is, there is no taxi trip available that could meet the criteria for integration. This result reflects the natural asymmetry of mass transportation trips regarding the number of taxi trips.

Differently, Figure 4.7b present the cumulative distribution of integration position options given pairs of candidate trips. Taxi positions are those where passengers would be when they accept the integration requisition, from there they will deviate from the original route and get additional transit passengers. To speed up the computation we considered the minimum distance of five hundred meters between each position. Transit positions are stations where the chosen vehicle stops and the passenger could disembark and wait to pick-up a shared taxi. Results show that more than 80% of taxi trips have more than one position where the integration could begin while for transit trips these options are in 70% of all trips.

4.4 Maximum Benefit Matching and Pricing

We now turn to the analysis of possibilities of integration from the perspective of the taxi passenger. Starting with the candidate integrations identified in the previous section, we now filter them into only those that benefit the taxi passenger, i.e., we identify the viable integrations from the perspective of both parties. The main incentive for taxi passengers to share their trip is saving money. Thus, we propose two pricing policies where the shared trip cost is proportionally divided between mass transportation and taxi passengers.

Our first pricing policy is based on the segments of the trip. The total cost of the shared trip is divided in three parts: (i) initial cost, which is the initial charge; (ii) integration cost, which is the cost of traveling from the location where the taxi passenger accepted the integration to the mass transportation station pick-up; and (iii) sharing cost, which is the cost of shared trip from mass transportation passenger pick-up on the shared taxi until the first passenger drops off. We assume that these costs can be either equally divided between both passengers or paid in a great proportion by mass transportation passengers, as the latter aims to save time.

To evaluate the segmented policy, we consider different values of parameters $s_{initial}$, s_{detour}, and s_{shared}, which represent the factors applied to split the cost of each shared segment (see Section 3.3). A value equal to 0.5 implies that the corresponding cost is equally divided into both passengers, whereas larger values represent the fraction of the cost imposed on the mass transportation passenger. We consider 5 different scenarios, as shown in the x-axis of Figure 4.8 which shows the numbers of distinct mass transportation, taxis and viable taxi-mass transportation integrations per pricing scheme.

We designed scenarios where either the cost of a segment is equally shared by both passengers or entirely paid by the mass transportation passengers, to analyze the impact of the splitting factors in extreme cases. There are in total eight different scenarios, but we considered that the shared segment of the trip should be split between those passengers, except in the extreme case where they would pay for the entire trip. Focusing on the five remaining possibilities it can be seen that the number of viable integrations increases when mass transportation passengers pay more for the shared trip. This is because the larger the share paid by the mass transportation passengers, the larger the chance of the taxi passenger also benefiting (financially) from the integration. Thus, the number of integration options increases as only those that benefit *both parties* are considered viable.

We now analyze the benefit each passenger received for each pricing division scheme. Figure 4.9 shows the cumulative distributions of the total amount of money saved by the taxi passenger (Figure 4.9a) and the total amount of time saved by the mass transportation passenger (Figure 4.9b) for each scheme. In both figures the x axis cannot

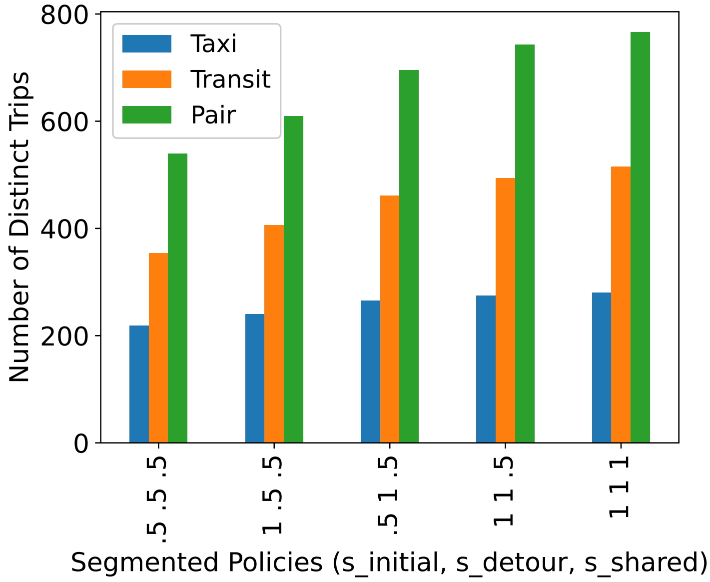
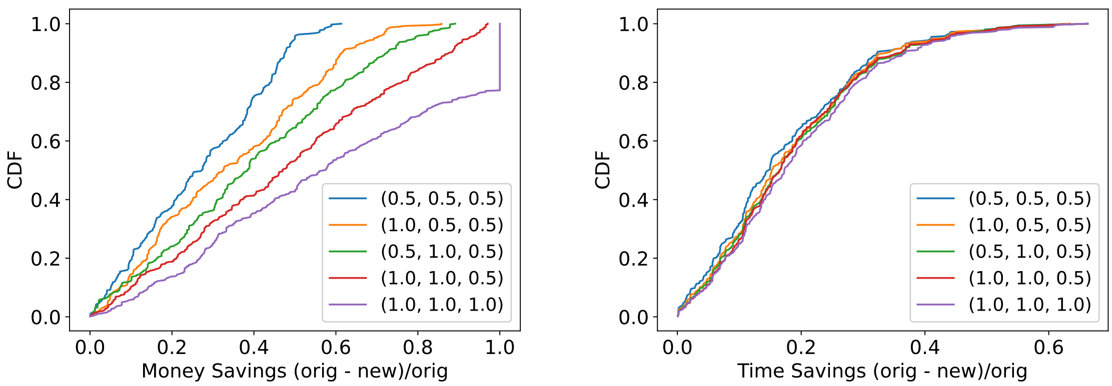


Figura 4.8: Distinct trips varying payment policies.

assume negative values ($orig > new$) because Algorithm 2 allows only integrations with money saving for taxi passengers and time saving for mass transportation ones.

Figure 4.9a shows that the greater the portion paid by the mass transportation passenger, the greater the taxi passenger savings, as expected. When mass transportation passengers pay half of the initial charge and half of the shared portion of the trip ($s\_initial = 0.5$, $s\_detour = 0.5$, $s\_shared = 0.5$), in about 75% of all trips the taxi passengers save up to 40% in the total cost of their trips. If the mass transportation passenger pays for the whole shared trip, taxi passengers save at least 40% of the trip ($s\_initial = 1$, $s\_detour = 1$, $s\_shared = 1$), which happens in 65% of all integrated trips.



(a) Taxi passenger saving money.

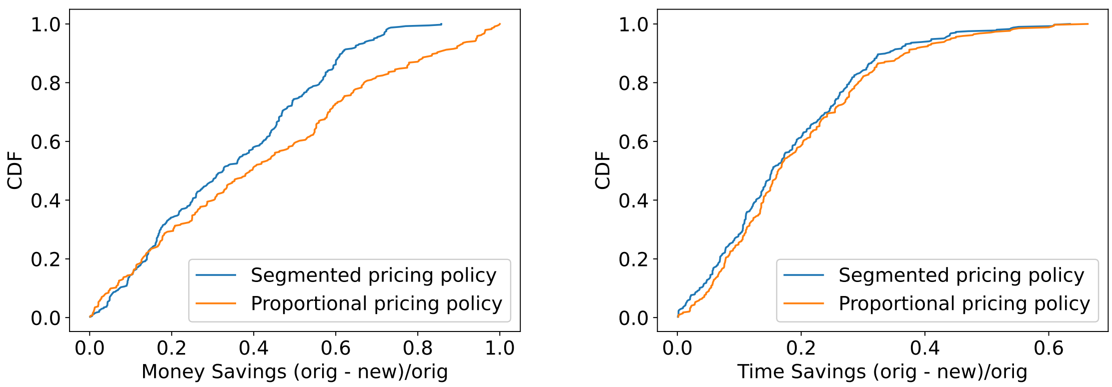
(b) Mass transportation passenger saving time.

Figura 4.9: Pricing divisions ($s\_initial$, s_detour, s_shared).

Figure 4.9b presents the cumulative distribution of the amount of time the mass transportation passenger saves for different pricing divisions. It can be seen that the proportion paid by the passengers does not affect substantially their time savings. In general, in 40% of the shared trips, mass transportation passengers save at least 20% of their trip time.

In the second pricing policy, the mass transportation passenger pays proportionally to the taxi passenger's extra delay. This value is computed considering the extra time incurred in the taxi passengers' trip when they opt to integrate their trip (see Section 3.3). Thus, the longer the trip delay, the greater the discount in the value paid by the taxi passenger in the integrated trip.

In Figure 4.10 we compare the proportional pricing policy with the segmented one. In the segmented pricing, we consider the division where transit passenger pays entirely for the initial cost, the detour, and the shared segment is divided equally to mass transportation and taxi passenger ($s\_initial = 1, s\_detour = 0.5, s\_shared = 0.5$).

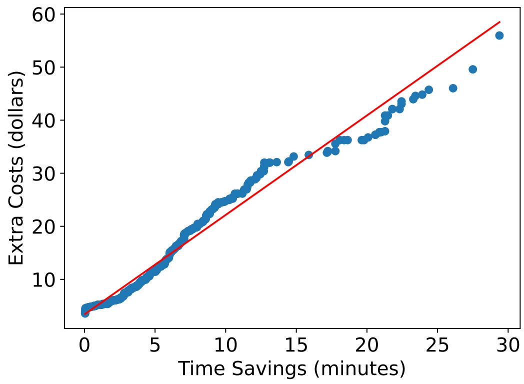


(a) Taxi passenger saving money. (b) Mass transportation passenger saving time.

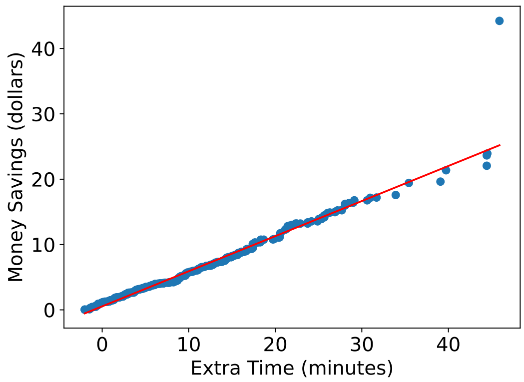
Figura 4.10: Segmented and proportional pricing policies.

Figures 4.10a and 4.10b present the cumulative distribution function of taxi saving money and mass transportation saving time for the segmented and proportional pricing policies. These results show that although the way to compute the integrated trip is different, one based on trip segments and the other proportional to the taxi passenger's extra time, the curves show similar behaviors. Figure 4.10a shows that for both policies in 30% of matchings taxi passengers save about 20% for integrating their trip. Observing 90% of the integrated trips, while in the segmented pricing policy the saved amount is about 60%, in the proportional pricing, this amount grows to around 80% savings. Figure 4.10b shows that no significant difference can be observed in the time savings for mass transportation passengers when the policies vary.

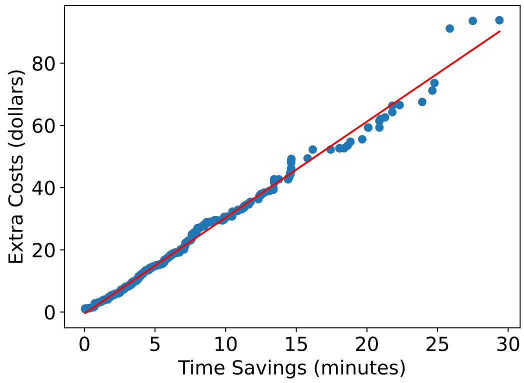
To show the difference between segmented and proportional policies, we computed the function that represents the gain of mass transportation and taxi passengers in each policy and compared them. Thus, from the first policy we generated Figures 4.11a and 4.11b, and from the proportional policy Figures 4.11c and 4.11d.



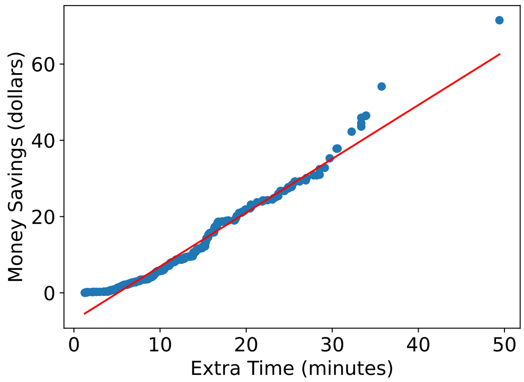
(a) Segmented Pricing: mass transportation passenger trips. $y = 1.87x + 3.39$



(b) Segmented Pricing: taxi passenger trips. $y = 0.54x + 0.54$



(c) Proportional Pricing: mass transportation passenger trips. $y = 3.09x - 0.49$



(d) Proportional Pricing: taxi passenger trips. $y = 1.41x - 7.18$

Figura 4.11: Max Benefit for Segmented and Proportional Pricing Policies.

In these figures, each blue dot represents a pair of trips that could be integrated generating the maximum benefit for both taxi and mass transportation passengers. The red line is the tendency curve, which the function that describes it is in the respective label. Figures 4.11a and 4.11c present the perspective of mass transportation passengers, where x-axis are the saving time and y-axis the extra cost paid to integrate their trip with a taxi passenger. On the other hand, Figures 4.11b and 4.11d present the perspective of taxi passengers, where the x-axis is the extra time and y-axis represents how much money taxi passengers save to share their trips with mass transportation ones.

Considering the mass transportation perspective, in the segmented pricing policy, Figure 4.11a, mass transportation passengers pay more for saved time than in the proportional pricing policy (Figure 4.11c). For instance, saving ten minutes in segmented pricing, public passengers should pay about 22.09 dollars while in proportional pricing they should pay about 33.41 dollars. In the taxi passenger's perspective, the proportional policy pricing (Figure 4.11d), generates more savings than the segmented one (Figure 4.11b). Considering an extra time of ten minutes in the taxi passenger's trip, they save about 6,92 dollars in the proportional pricing policy and in the segmented one they save 5.94 dollars.

In both policies, the amount paid by public passengers is considerably more than the amount saved by taxi passengers. That asymmetry is due to the detour of the integrated trip that is charged mostly to the mass transportation passengers. In the segmented pricing, they pay entirely the initial cost of a taxi trip plus a half of the detour route to get them. In the proportional pricing, the public passenger should pay proportionally to the delay in the taxi passenger's route. Furthermore, if mass transportation passenger drops-off after the taxi one, they should pay fully entirely for the segment between drop-offs.

4.5 Maximum Benefit and Synthetic Trips

Naturally, more people are traveling by mass transportation modes than by taxi (as seen in Table 4.1). Then, we inflated the taxi dataset to analyze how the number of viable integrations increases when the taxi offer is greater (see Section 4.1.1.2). Thus, integrations were recomputed considering that synthetic taxis picked up passengers at the same place as the real ones but at different moments. To reduce the search space of potential integrations we matched trips that the shortest distance between them is at most three kilometers ($d = 3,000$ in Algorithm 3). Both pricing policies were considered, the one that is based on route segments ($s\_initial = 1$, $s\_detour = 0.5$, $s\_shared = 0.5$) and the other that is proportional to the delays incurred in the route of taxi passengers. Figure 4.12 shows results of this analysis.

Results show that the segmented pricing policy generates more viable integrations than delay proportional ones. The increase in available taxis generates more possible integrations, as well. When the number of available taxi increases five times ($n = 5$), the number of viable integrations increases around two times. And when the increase is ten times ($n = 10$) of the number of available taxis integrations grows up to three times. Additionally, the variation of synthetic trips pick-up time t considerably affects the number of potential integrations. In both policies, when the range goes from ten minutes ($\delta = 10$) to twenty minutes ($\delta = 20$) around the original pick up time, integrations grow around 40%.

Varying two parameters, n and δ, we can conclude that the increase of pick-up time window δ generates more viable integrations than the increase of available trips n. In the dataset *5x_10min* ($n = 5$ and $\delta = 10$) the number of viable integrations in the segmented policy is 570. When we double the number of trips and keep the time interval,

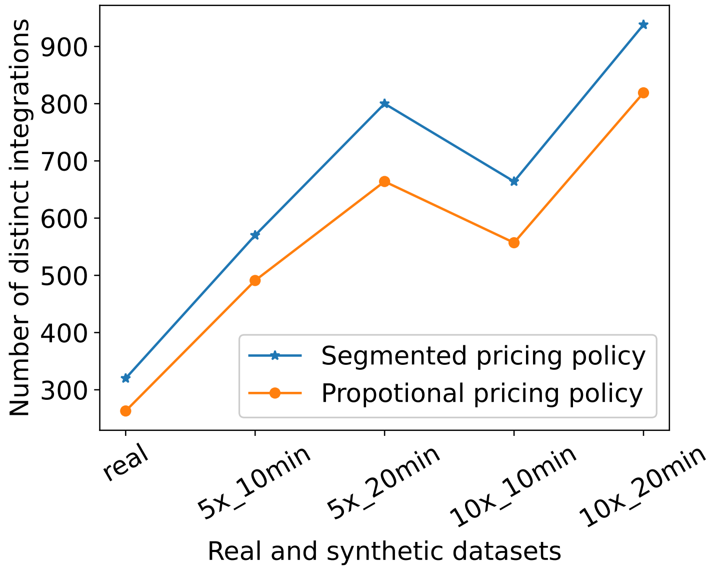


Figura 4.12: Max benefit and synthetic datasets.

dataset *10x_10min*, the number of integrations increases to 664 (16%). Keeping $n = 5$ and increasing the window time to $\delta = 20$, dataset *5x_20min*, the number of integrations increases to 800 (40%). Therefore, in a scenario where there are more available individual trips, the chances of viable matches increases significantly.

Capítulo 5

Conclusion

In large urban centers, the metropolitan transportation network is composed of different means that enable passengers to commute and move around for different purposes. However, there are unexplored options for transportation that could integrate existing transport modes to offer cheaper and fast trips for passengers. The mass transportation system, for example, is composed of different modes, such as subway, train, and bus that integrate with each other but do not communicate with existing private car services such as taxi. Therefore, we evaluate a mechanism that integrates taxi with mass transportation modes in such a way that taxi passengers share their trip with mass transportation ones and both passengers benefit someway. In the proposed integration strategy, mass transportation passengers save time, while taxi passengers save money if they opt to share their trip. Furthermore, to save time passengers coming from the mass transport system should pay a little more and taxi passengers should accept some delays in their trip to save money.

To evaluate our proposal of integration, we analyzed real data from an origin-destination survey on residents of New York City. The survey collected trip characteristics such as the origin and destination, date time of origin, and the transportation mode used by interviewed residents in the day before the survey. From this data, we reconstructed all passenger trip routes using the Open Trip Planner framework. Then, from all passenger trips composed of the complete route with intermediate trip positions and the mass transportation stops, we characterized trip data considering temporal and spatial aspects to understand the specificity of each transport modal. Results show that trips made by taxi and mass transportation modes often overlap in time and space, suggesting that route sharing is a viable alternative.

Then, we propose the TM-Sharing mechanism that integrates mass transportation with ridesharing. In the service, passengers that are initially in a mass vehicle can request a shared taxi trip using their smartphone. If there is a match, the system computes the shared route and request the passenger to disembark in a station and pick-up in a shared taxi. To evaluate the viability of that alternative of transport, we designed an algorithm that computes the integrated trips considering temporal and economic aspects. The TM-Sharing algorithm is composed by three functions: *ST-Matching*, *Viability-Filtering* and

Maximum-Benefit. First, *ST-Matching* selects pairs of trips that could be integrated considering temporal and tpatial aspects. Second, *Viability-Filtering* selects those pairs that are economically viable. Third, *Maximum-Benefit* selects those pairs of trips with maximum benefit, as well as, the best integration station considering temporal and economic aspects.

To be viable, an integrated transport system should have a fair pricing scheme. Thus, we propose two different strategies to compute the fare of TM-Sharing. One is based on the segment of the integrated trip, and the other on the taxi passenger trip delay. In the segment-based policy, we divided the integrated trip cost into three parts: *s_initial* (the taxi initial charge), *s_detour* (cost of the deviation to get the passenger), and *s_shared* (cost of the shared portion of the route), the cost of each part was divided among mass transport and taxi passengers. In the second pricing policy, the amount of the integrated trip paid by the mass transport passenger is proportional to the delay in the taxi passenger trip. The results of this pricing comparison show that the segment-based policy generates 21% more viable trips than the proportional one. In the segmented policy to save ten minutes in their trip, the mass passenger should pay on average \$22.09 while the taxi passenger saves \$5.94. In the proportional policy, to save ten minutes passengers from mass transportation pay \$30.41, while taxi passengers save \$6.92. In sum, the segmented policy showed to be more viable because it charges less mass transportation passengers and generates more opportunities of integration.

Finally, we generated four synthetic datasets based on the real one by increasing the number of available taxi trips to evaluate how the increase of taxi trips and its temporal aspects would impact in the number of viable integrations. For each real taxi trip, we generate other n trips with the same spatial aspects and varying the pick-up time around a time window of t minutes. Results show that longer values of t generates more viable integrations than the increase of the number of trips (n). Considering the segmented pricing policy, when $n = 5$ and we increase the value of t from 10 to 20 minutes the number of viable integrations increases by 40%. When we keep $t = 10$ and vary n from 5 to 10 times, the number of viable trips increases by only 16%.

In this work, we explore the trade-off of time and money and we suppose that mass transportation passengers would like to save money while individual transport ones would like to save money, which is not always true. There are other different trade-offs that could be explored like comfort and money and, comfort and time. Additionally, the results of this work cannot be generalized to other cities, New York is a large city that is almost entirely covered by the subway network, which is complemented by the bus mode. Understand the local passenger's behavior and the geography of the city is crucial to the design of an integrated and shared transportation system. Furthermore, the success of such a transportation system depends on the ability to schedule different modes and adapting to adversities like user cancellations, traffic delays, and unavailability of the

Internet. However, an integrated system that includes other different modes like bikes, motorcycles, private and autonomous vehicles could make the transport of large cities more efficient and adaptable, but the complexity of such a system increases substantially.

Referências Bibliográficas

[1] Louai Alarabi, Bin Cao, Liwei Zhao, Mohamed F. Mokbel, and Anas Basalamah. A demonstration of sharek: An efficient matching framework for ride sharing systems. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '16, pages 95:1–95:4, New York, NY, USA, 2016. ACM.

[2] Lauren P. Alexander and Marta C. Gonzalez. Assessing the impact of real-time ridesharing on urban traffic using mobile phone data. In *Proceedings of the 4th ACM SIGKDD International Workshop on Urban Computing*. ACM, 2015.

[3] Mohammad Asghari, Dingxiong Deng, Cyrus Shahabi, Ugur Demiryurek, and Yaguang Li. Price-aware real-time ride-sharing at scale: An auction-based approach. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '16, pages 3:1–3:10, New York, NY, USA, 2016. ACM.

[4] Mohammad Asghari and Cyrus Shahabi. Adapt-pricing: A dynamic and predictive technique for pricing to maximize revenue in ridesharing platforms. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '18, pages 189–198, New York, NY, USA, 2018. ACM.

[5] E. Cangialosi, A. Di Febbraro, and N. Sacco. Designing a multimodal generalised ride sharing system. *IET Intelligent Transport Systems*, 10(4):227–236, 2016.

[6] Peng Cheng, Hao Xin, and Lei Chen. Utility-aware ridesharing on road networks. In *Proceedings of the 2017 ACM International Conference on Management of Data*, SIGMOD '17, pages 1197–1210, New York, NY, USA, 2017. ACM.

[7] Blerim Cici, Athina Markopoulou, Enrique Frias-Martinez, and Nikolaos 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, 2014.

[8] Ali Gholami and Afshin Shariat Mohaymany. Analogy of fixed route shared taxi (taxi khattee) and bus services under various demand density and economical conditions. *Journal of Advanced Transportation*, 46(2):177–187, 4 2012.

[9] Can Huang, Defu Zhang, Yain-Whar Si, and Stephen C. H. Leung. Tabu search for the real-world carpooling problem. *Journal of Combinatorial Optimization*, pages 1–21, 2016.

[10] Yan Huang, Favyen Bastani, Ruoming Jin, and Xiaoyang Sean Wang. Large scale real-time ridesharing with service guarantee on road networks. 7(14):2017–2028, 2014.

[11] Vaishnav Kameswaran, Lindsey Cameron, and Tawanna R. Dillahunt. Support for social and cultural capital development in real-time ridesharing services. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 342:1–342:12, New York, NY, USA, 2018. ACM.

[12] Hao Liu, Yongxin Tong, Panpan Zhang, Xinjiang Lu, Jianguo Duan, and Hui Xiong. Hydra: A personalized and context-aware multi-modal transportation recommendation system. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pages 2314–2324, New York, NY, USA, 2019. ACM.

[13] Yifang Liu, Will Skinner, and Chongyuan Xiang. Globally-optimized realtime supply-demand matching in on-demand ridesharing. In *The World Wide Web Conference*, WWW '19, pages 3034–3040, New York, NY, USA, 2019. ACM.

[14] Ali Masri, Karine Zeitouni, and Zoubida Kedad. Retry: Integrating ridesharing with existing trip planners. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'17, pages 47:1–47:10, New York, NY, USA, 2017. ACM.

[15] Paolo Santi, Giovanni Resta, Michael Szell, Stanislav Sobolevsky, Steven H. Strogatz, and Carlo Ratti. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111(37):13290–13294, 2014.

[16] Zhanjun Wang, Khaled Mesghouni, and Slim Hammadi. Agent-based coalition formation in a co-modal transport system. In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 03*, pages 310–317. IEEE Computer Society, 2014.

[17] Jiangfeng Xi, Quan Wang, Tongqiang Ding, Lili Zheng, Shengli Wang, and Wei Li. Research on the coordinated design of bus and taxi station. *Mathematical Problems in Engineering*, 2015(372496):5, 2015.

[18] Desheng Zhang, Tian He, Fan Zhang, Mingming Lu, Yunhuai Liu, Haengju Lee, and Sang H. Son. Carpooling service for large-scale taxicab networks. *ACM Trans. Sen. Netw.*, 12(3):18:1–18:35, August 2016.

[19] Jianwei Zhang, Feixiong Liao, Theo Arentze, and Harry Timmermansa. A multi-modal transport network model for advanced traveler information systems. *Procedia Computer Science*, 5:912 – 919, 2011.

[20] Libin Zheng, Lei Chen, and Jieping Ye. Order dispatch in price-aware ridesharing. *Proc. VLDB Endow.*, 11(8):853–865, April 2018.