Title

Quantifying Sharing Potential in Transportation Networks and the Benefits of Mobility-on-Demand Services with Virtual Stops

Author

JAYASHANKAR SHOBHA, NAVJYOTH SARMA

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Quantifying Sharing Potential in Transportation Networks and the Benefits of Mobility-on-Demand Services with Virtual Stops

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil and Environmental Engineering

by

Navjyoth Sarma Jayashankar Shobha

Dissertation Committee:
Assistant Professor Michael F. Hyland, Chair
Professor R. Jayakrishnan
Professor Michael McNally

2023

DEDICATION

To

my mother, sister, and grandparents

for their continuous support throughout this journey

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

LIST OF ACRONYMS

C2C	:	Corner-to-Corner
D2D	:	Door-to-Door
MNFLOP	:	Maximum Network Flow Overlap Problem
MOD	:	Mobility-on-Demand
PUDO	:	Pickup and Dropoff
PU/DO	:	Pickup or Dropoff
RH	:	Ride-hailing
RP	:	Ride-pooling
SMM	:	Shared Mobility Mode
TSA	:	Transportation System Attributes

LIST OF DEFINITIONS

Corner-to-Corner (C2C) Mobility Service – A mobility service that picks up and drops off riders at nearby, predetermined locations — or Virtual Stops — close to their original pick-up and drop-off locations, which helps increase the service's efficiency (Moovit, n.d.). Riders are generally required to walk the first/last mile of their trip to/from these Virtual Stops. These are different from Door-to-Door (D2D) Mobility Services where riders are picked up and dropped off at their original PUDO locations (or their doorstep).

Mobility-on-Demand (MOD) – Transportation/Mobility services in which vehicles operate with flexible routes, schedules and/or PUDO locations based on rider demand, rather than using a fixed route or timetable.

Shared Mobility – Transportation/Mobility services that are shared among users, either simultaneously or one after another. This includes conventional fixed transit (local bus, Bus Rapid Transit (BRT), light rail, heavy rail); micromobility (bike-sharing, scooter sharing); and on-demand modes (carsharing, Ride-hailing, Ride-pooling, microtransit) (Moovit, n.d.).

Sharing Potential – In the context of transportation, refers to the level to which demand (either person trips or freight) in a region can be 'shared' or pooled together into fewer vehicles to minimize resource consumption (vehicles, physical road space etc.) and reduce congestion.

Virtual Stop - A location used as a potential PU/DO location in Corner-to-Corner (C2C) mobility services, which can be any identifiable location along a street, such as a

business, street corner, parking lot, or even a bus stop (Moovit, n.d.). Virtual Stops generally do not include any physical element of a conventional bus stop and hence are only visible on the mobile application of the Corner-to-Corner mobility service (Harmann et al., 2022).

ACKNOWLEDGMENTS

First and foremost, I extend my heartfelt gratitude to my advisor, Dr Michael Hyland. His mentorship has been a beacon of guidance, leading me through uncharted paths to tackle complex yet pertinent transportation problems with innovative solutions. His relentless pursuit of challenging and intriguing problems, coupled with an eagerness for unconventional approaches, has inspired me. I am deeply thankful for the privilege of working with him and for all the opportunities he provided me.

I am grateful to my committee members, Dr Radhakrishnan Jayakrishnan, and Dr Michael McNally, for their invaluable inputs and direction. Their expertise on various domains in transportation has genuinely helped me recognize the intricate and captivating interdisciplinary essence of the field. I am also thankful to all my professors at the Institute of Transportation Studies at UC Irvine, including Dr. Will Recker, Dr. Jean-Daniel Saphores, and Dr Wenlong Jin, whose intellectually stimulating lectures have enriched my knowledge and skills in the domain. I am also thankful to Dr Amelia Regan, Dr Gordon Fielding, and Dr Stephen Ritchie for their backing and encouragement.

I am thankful to the Grad division at UC Irvine and the Hein family for the generous assistance they provided me through the Gordon Hein Scholarship. This played a pivotal role in fueling my research program.

I extend my deep gratitude to all my research group members at the Institute of Transportations Studies. I would like to thank Younghun Bahk for his valuable assistance in creating plots for this dissertation. Additionally, I would like to appreciate the contributions of my cohorts Younghun Bahk and Zifan Wang, as well as my collaborators at Argonne National Laboratory - Dr Krishnamurthy Gurumuthy, Dr Felipe de Souza, and Dr Omer Verbas - for their instrumental role in the second study of my dissertation. I would also like to thank Argonne National Laboratory for partially funding the second study in my dissertation.

My heartfelt appreciation goes out to all my dear fellow graduate students, former students, and Postdoctoral researchers at ITS – Tanjeeb Ahmed, Pratiik Mailk, Eduardo Marino, Lu Xu,

Farzana Khatun, Irene Martinez, Felipe de Souza, Daisik Nam, Suman Mitra, De'von Jennings, Arash Ghaffar, and everyone – all of you made ITS a vibrant and memorable place for me, deeply enriching my personal and academic life. I am eternally indebted to Riju Lavanya, Koti Reddy Allu, Dhanya Pranab Kumar, and Dingtong Yang for being priceless sources of support and friendship throughout my highs and lows. I am very fortunate and honored to have friends like you.

I extend my gratitude to all the current and former staff at the Institute of Transportation Studies and Civil and Environmental Engineering, whose tireless efforts behind the scenes have contributed significantly to my academic journey.

I am thankful to the Disability Service Center at UC Irvine for their invaluable assistance in navigating the challenges of graduate school as an individual with a disability. The support, and resources provided by their staff and counselors have been pivotal in making UC Irvine a welcoming and accommodating place for students with disabilities to pursue their academic endeavors.

To all my friends near and far, thank you for providing an avenue for me to take a break from the rigors of doctoral studies. You all mean a lot to me.

Last but not the least, I would like to express my deepest gratitude and love to my family—my mother, Shobha Jaisankar, my sister Jyotsna Jaisankar, my grandparents, my cousins back in India—Vageesh K.S and Vignesh K.S, and my extended family. Their unwavering support, even from afar, has been my pillar of strength throughout this entire journey. I hold a special place of gratitude for my dear grandfather, Mr. SHK Sarma, who ignited my passion for transportation from a young age. His meticulous planning of journeys in an era predating widespread internet use and navigation tools sowed the seeds for my love for trains, buses, and road trips. I have vivid memories of my childhood looking in awe at my grandfather poring over huge paper road maps laid out on a table and over railway timetables to plan our journeys. His support and drive have been pivotal in shaping my academic pursuits and career path. I am forever grateful for his influence and encouragement.

VITA

Navjyoth Sarma Jayashankar Shobha

2012	B.Tech. in Computer Science and Engineering, Amrita University, Coimbatore, India
2012-13	Programmer Analyst, Cognizant Technology Solutions, Chennai, India
2014	P.G. Diploma in Geoinformatics, Centre for Development of Advanced Computing (C-DAC), Pune, India
2015	Diploma in Transportation Economics, Institute of Rail Transport, New Delhi, India
2016	Transportation Engineer Intern, Fehr and Peers Transportation Consultants, Orange County, CA
2017-2023	Graduate Student Researcher, Institute of Transportation Studies, University of California, Irvine
2020	M.S. in Civil and Environmental Engineering (Transportation Systems Engineering), University of California, Irvine
2023	Ph.D. in Civil and Environmental Engineering (Transportation Systems Engineering), University of California, Irvine

FIELD OF STUDY

Planning, Modeling, and Operations of Shared Mobility Systems, Transportation Modeling

PUBLICATIONS

Ahmed, Tanjeeb, Michael Hyland, Navjyoth J.S. Sarma, Suman Mitra, and Arash Ghaffar. 2020. "Quantifying the Employment Accessibility Benefits of Shared Automated Vehicle Mobility Services: Consumer Welfare Approach Using Logsums." Transportation Research Part A: Policy and Practice 141 (November): 221–47.

Navjyoth Sarma, J.S., Daisik Nam, Michael F. Hyland, Felipe de Souza, Dingtong Yang, Arash Ghaffar, and I. Omer Verbas. 2020. "Effective and Efficient Fleet Dispatching Strategies for Dynamically Matching AVs to Travelers in Large-Scale Transportation Systems." In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1–6. IEEE..

Yang, Dingtong, Navjyoth J.S. Sarma, Michael F. Hyland, and R. Jayakrishnan. 2021. "Dynamic Modeling and Real-Time Management of a System of EV Fast-Charging Stations." *Transportation Research Part C: Emerging Technologies* 128 (July).

CONFERENCE PRESENTATIONS

Navjyoth Sarma J.S, Michael Hyland. "Optimal Design of Mobility Service Networks Based on Trip Dispersion". INFORMS Annual Meeting, Anaheim, 2021

Navjyoth Sarma J.S, Michael Hyland. "Metrics for Quantifying Shareability in Transportation Networks: The Maximum Network Flow Overlap Problem". 102nd Annual Meeting, Transportation Research Board, Washington D.C., 2023

Navjyoth Sarma J.S, Krishnamurthy Gurumurthy, Michael Hyland, Younghun Bahk, Felipe de Souza, Zifan Wang, Omer Verbas. "On-demand ride-pooling with Walking Legs: Decomposition Approach for Dynamic Matching & Virtual Stops Selection", Bridging Transportation Research Conference (BTR5), 2023

ABSTRACT OF THE DISSERTATION

Quantifying Sharing Potential in Transportation Networks and the Benefits of Mobility-on-Demand Services with Virtual Stops

by

Navjyoth Sarma Jayashankar Shobha

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Irvine, 2023

Assistant Professor Michael F. Hyland, Chair

Cities around the world vary in terms of their transportation network structure and travel demand patterns, with implications for the viability of shared mobility services. Recently, the urban mobility sector has witnessed a significant transformation with the introduction of several new types of Mobility-on-Demand (MOD) services that vary in terms of their capacity and flexibility of routes, schedules, and user Pickup and Dropoff (PUDO) locations. This dissertation proposes models and algorithms to analyze sharing in transportation networks and Mobility-on-Demand (MOD) services in two comprehensive studies. The first study aims to quantify the sharing potential of travelers within a city or region's transportation network. The second study aims to measure trade-offs in user and operator costs when MOD services operate with Virtual Stops which refer to flexible PUDO locations requiring travelers to walk the first/last mile of their trip.

The first study addresses the lack of metrics that jointly characterize a region's travel demand patterns and its transportation network in terms of the potential for travelers to share trips. I define sharing potential in the form of person-trip shareability and introduce and conceptualize 'flow overlap' as the fundamental metric to capture shareability. The study formulates the Maximum Network Flow Overlap Problem (MNFLOP), a math program that assigns person-trips to network paths that maximize network-wide flow overlap. The results reveal that the shareability metrics can (i) meaningfully differentiate between different Origin-Destination trip matrices in terms of flow overlap, and (ii) quantify demand dispersion of trips from a single location considering the underlying road network. Finally, I validate MNFLOP's ability to quantify shareability by showing that demand patterns with higher flow overlap are strongly associated with lower mileage routes for a last-mile microtransit service.

The second study proposes a scalable algorithm for operating shared-ride MOD services with flexible and dynamic PUDO locations—called C2C (Corner-to-Corner) services—in a congestible network. I compare four MOD service types: Door-to-Door (D2D) Ride-hailing, D2D Ride-pooling, C2C Ride-hailing, and C2C Ride-pooling by evaluating operator and user costs. The results show that Ride-pooling reduces operator costs while slightly increasing user costs, whereas C2C reduces operator costs but significantly increases user costs. Combining Ride-pooling and C2C appears promising to reduce operator costs and to reduce vehicles miles traveled (VMT) in MOD systems.

Chapter 1. Introduction

1.1 Background and Motivation

The emergence and growth of the 'sharing economy' has been one of the biggest changes in the consumer market that has occurred since the turn of the millennium and especially over the past decade. The sharing economy promises to provide consumers with a cheaper and more convenient way to access services such as, transportation, rental cars, parking spaces, lodging, office space, recreational services as well as goods such as a hammer or a drill for short time usage.

Ideally, the sharing economy yields a win-win transaction for all the entities involved as well as for society in general. Consumers can now access goods and services during times of need, without having to own an asset, thereby reducing the cost. Suppliers who own an asset such as an empty seat in a vehicle, a parking spot, a room, or a drill can lend it to others when they are not using it, giving them an additional source of income from the asset, which would have otherwise been a wasted resource. At a societal level, the sharing economy has the potential to reduce consumption, reduce the amount of waste generated, and therefore reduce energy usage and emissions (Curtis and Lehner, 2019; Heinrichs, 2013).

The emergence of smart phones, near ubiquitous access to high-speed internet, online payment gateways, and social networks over the past decade have reduced transaction costs and made it much easier to share assets at a larger scale. In addition, advancements in big data and location-based services have made it possible to track and predict patterns of consumption and availability of service over time. The market size of the

sharing economy, which was valued at around $150 billion in 2022 is projected to grow around $800 billion by the end of the decade (Yahoo Finance, 2023).

1.1.1 Sharing Economy and the Urban Mobility Sector

The boom in the sharing economy along with advancements in computational and communication technology has transformed the transportation and mobility sector over the past decade. Earlier, urban mobility used to be a heavily regulated sector with modes such as rail, bus, taxicab, and paratransit services providing mobility services to travelers who did not have access to a personal vehicle. City and regional transit or planning agencies mostly controlled the planning and operation of these services in the form of the routes in which transit services operate, or the number of taxicab medallions to be issued in a city. Today, on the other hand, a plethora of new modes of different capacities and varying degrees of flexibility in route, schedule, and Pickup and Dropoff (PUDO) locations have entered the urban mobility sector. These include app-based Ride-hailing services such as Uber and Lyft; Ride-pooling services such as Uber Pool and Lyft Shared; Ride-pooling services with flexible Pickup and Dropoff (PUDO) locations such as Uber Express Pool; and other forms of flexible transit services such as those provided by Via. While these services accounted for less than 100 million riders in 2012, the number of passengers carried by such Transportation Network Companies (TNCs) grew exponentially to about 1.9 billion in 2016 and 4.2 billion in 2018 (Schaller, 2018).

Figure 1.1 shows a spectrum of shared mobility modes currently operating in the urban mobility arena. The left end of the spectrum is occupied by fixed modes such as heavy rail, light rail, BRT, and local bus. These services have a fixed route and schedule and have

high capacity. These services therefore typically require many passengers to be travelling in a similar direction or to the same destination, at the same time, for them to be viable. The rightmost end of the spectrum includes Ride-hailing services (like Uber, Lyft, Taxicabs) that provide point-to-point mobility-on-demand for a single passenger, and do not follow any route or schedule. In between these two extremes, there are other modes which are variations of fixed transit or on-demand ride sharing services that have varying degrees of route and schedule flexibility and varying capacities.

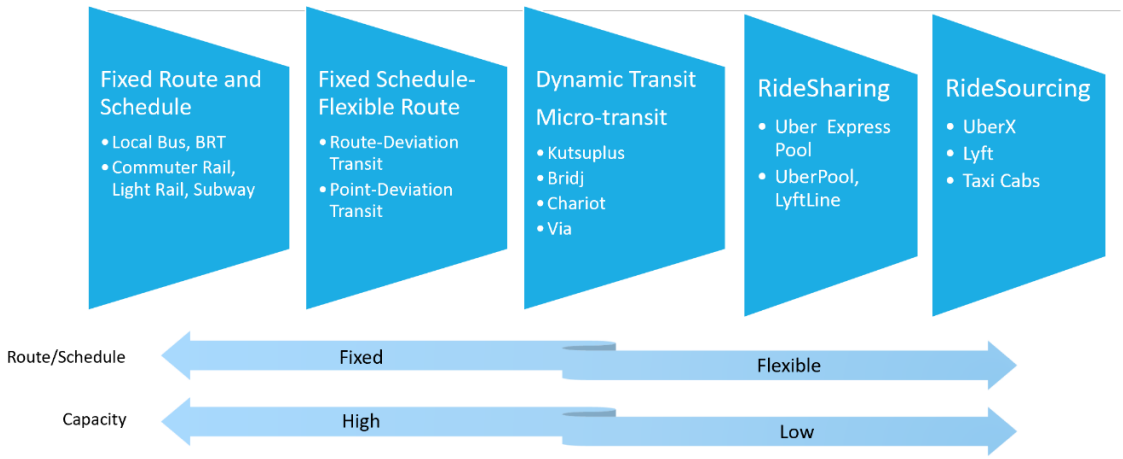

Figure 1.1 Spectrum of Shared Mobility Modes

Route-Deviation and Point-Deviation transit services, as their names suggest, represent transit services where the vehicle is allowed to take detours off a fixed route or a fixed set of locations, respectively. Examples of route-deviation transit services include the UTAFLEX services operated by Utah Transit Authority and 'Deviated Route' services operated by Tillamook County Transportation District (TLTD) in Oregon. Feeder services operating around transit stations or between multiple transit stations are examples of point-

deviation transit services. The above two services have some form of rigidity in terms of route and/or schedule in their operation.

Dynamic transit services such as the ones operated by Via and other van pool services operate on flexible routes and schedules trying to pool together many passengers heading in a similar direction. Ride-share services such as Uber Pool, Lyft Shared, and BlaBlaCar are a scaled down version of flexible transit services that try to pool together generally 2 or 3 passengers into a single vehicle. Other types of Mobility-on-Demand (MOD) services such as Uber Express Pool combine some attributes of fixed transit services with some attributes of a rideshare service, such as PUDO locations nearby the origin/destination or travelers, thereby requiring the passenger to complete one leg of the journey by foot.

The spectrum of shared mobility modes from mass transit on one end to Ride-hailing and taxicab services on the other end may provide a sustainable alternative to personal vehicle travel, as the shared mobility modes can pool together travelers heading in the same direction into a single vehicle and/or by reducing the requirement for parking in congested cities. Despite the potential of shared services to improve sustainability in the transportation sector, the actual results have been quite mixed over the past few years. Recent studies have shown that Ride-hailing services –- services where each vehicle serves a single passenger from/to their doorstep –- account for a significant proportion of trips operated by TNCs. These services have been found to be partially responsible for increasing congestion in cities, with 70% of all their trips being from the 9 largest densely populated metropolitan areas in USA (Schaller, 2018). Such services have also been found to have an adverse impact on transit ridership even in dense urban centers (Graehler et al., 2019; Schaller, 2018).

To address the inefficiencies and negative externalities associated with Ride-hailing services, operators run Ride-pooling services in which a single vehicle serves more than one trip request simultaneously, thereby reducing the number of vehicles required to serve the same number of travelers. Examples of such services include Uber Pool and Lyft Shared. There is also an emerging category of MOD services called Corner-to-Corner (or C2C) services wherein requests are picked up and/or dropped off at locations near the request origin or destination instead of conventional MOD modes that serve them at their doorstep (called as Door-to-Door or D2D services). The next subsection provides a brief background on such services.

1.1.1.1 Corner-to-Corner MOD Services

As the name implies, C2C MOD services require the users to walk the first/last mile of their trip to/from a location (or a 'street corner') close to their origin or destination. Such flexible PUDO locations are sometimes referred to as *Virtual Stops*. Virtual Stops can be any identifiable location along a street, such as a business, street corner, parking lot, or even a bus stop (Moovit, n.d.). Virtual Stops generally do not include any physical element of a conventional bus stop and hence are only visible on the mobile application of the C2C mobility service (Harmann et al., 2022). C2C services could either operate as Ride-hailing (RH) services or Ride-pooling (RP) services. Examples of such C2C services include Uber Express Pool, SacRT SmaRT Ride operated by Sacramento Regional Transit, COTA Plus operated by Central Ohio Transit Authority (COTA), UTA on Demand operated by Utah Transit Authority (UTA), and MetroNow operated by Southwest Ohio Regional Transit Authority (SORTA).

One of the primary advantages of C2C services, especially C2C Ride-pooling, is the potential to significantly enhance MOD service fleet productivity. By concentrating pickups and drop-offs at designated points instead of performing time-consuming door-to-door detours, service providers can optimize travel routes and reduce idle time for vehicles. This optimization translates into higher passenger throughput per unit of time, as more travelers can be accommodated within the same timeframe. Moreover, C2C services can contribute to improved vehicle occupancies, reducing the number of underutilized trips and making better use of available seating capacity.

Another compelling benefit of C2C Ride-pooling is the reduction in overall fleet distance traveled per passenger request. With predetermined PU and DO locations, vehicles can follow more direct and efficient routes, minimizing unnecessary detours and decreasing the total distance covered. This reduction in travel distance not only conserves energy and reduces emissions but also contributes to a more cost-effective operation for the service provider. However, transitioning from D2D to C2C services comes at the expense of inconveniences to users and also brings algorithmic challenges to the operator by increasing the problem complexity and computational costs. Therefore, a systematic analysis of different C2C and D2D MOD services is required to evaluate the trade-offs between operational, user and computational costs.

1.1.2 Heterogeneity in Urban Transportation Systems

The previous subsection described the plethora of shared mobility modes currently operational in cities around the world, the potential they have to address some of the mobility gaps, and the associated challenges and problems that arise. At the same time, cities

around the world also exhibit remarkable differences in terms of their transportation network structures, spatial and temporal travel patterns, travel behavior, population densities, and land use. This heterogeneity is shaped by a combination of historical development, policy choices, geographical constraints, and socio-economic factors. The variations in transportation demand and supply across cities or even within sub-regions of a city have huge implications on sharing potential and the viability of shared mobility services.

1.1.2.1 Mapping Transportation System Characteristics to Sharing Potential

The viability and operational efficiency of a shared mobility mode depends on the number of potential travelers that could be pooled into a single vehicle, and the extent of commonality between the routes of these potential travelers from their origins to their destinations. This in turn depends on multiple factors such as land use and density of the region, its underlying transportation network and the travel demand and mobility patterns in the region. The presence of large residential and employment zones in dense urban centers increases the likelihood of spatial and temporal commonalities between routes of trips made by a large number of travelers. This makes modes such as heavy rail, light rail, BRT, and high-frequency bus services viable for dense areas. However, in suburban areas where residential and employment centers are scattered throughout the region, the operation of high capacity and high frequency fixed transit services are not viable. Fixed transit services in such areas often have circuitous routes, see very low ridership, and do not provide automobile competitive level of services to their users. However, it may still be possible to operate shared mobility modes of lower capacities in such regions, based on the number of travelers that could be pooled together heading in a similar direction.

A remaining challenge for researchers involves determining generalizable associations between transport system attributes (TSAs) and the efficiency of specific Shared Mobility Modes (SMM) in Figure 1.1, where TSAs include the magnitude and directionality (i.e., the spatial distribution) of traveler demand and underlying road network structure. Only one stream of research aims to map the efficiency of SMMs onto TSAs, and these studies only consider one TSA—density of demand (or magnitude of demand) in a homogenous region—when comparing fixed and flexible transit (Li and Quadrifoglio, 2010a; Nourbakhsh and Ouyang, 2012; Quadrifoglio and Li, 2009). These studies do not consider the directionality of demand nor the underlying network structure.

The inability to map the efficiency of SMMs onto specific TSAs beyond demand density is a significant shortcoming in the literature that I believe stems from the challenges associated with jointly quantifying various TSAs. Figure 1.2 displays a roadmap for connecting TSAs to the efficiency of SMMs for a subregion of a metropolitan area (henceforth, just 'subregion'). While quantifying each attribute individually and then connecting these attributes to the efficiency of SMMs is theoretically possible, this approach has several challenges and shortcomings. First, it is difficult (or potentially meaningless, depending on the metric) to quantify the spatial distribution of demand independent of the underlying road network; hence, I propose the simultaneous consideration of road network structure and spatial demand in metric creation. Second, and similarly, measures of demand magnitude are more valuable in conjunction with the spatial demand distribution and road network structure. Hence, rather than connecting individual attributes to the efficiency of SMMs, I propose to create shareability metrics at the person-level (instead of vehicle- or mode-level) that combine all three TSAs.

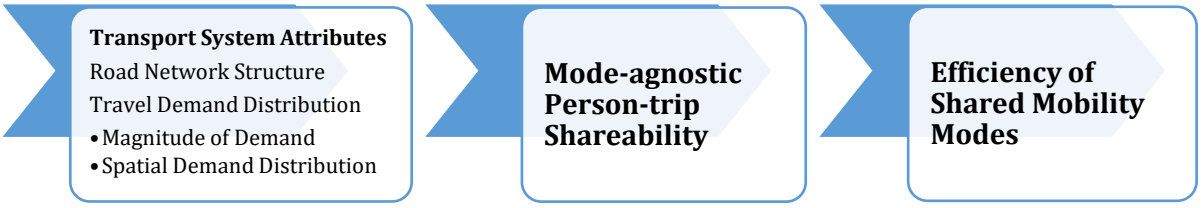

Figure 1.2 Roadmap to link TSAs, to the performance of SMMs through person-trip shareability

1.2 Research Objectives and Contributions

The Background and Motivation section presented an overview of the heterogeneity in urban transportation systems, the plethora of MOD services currently in service and the various challenges and problems that ensue. In this context my dissertation aims to achieve two major goals through two comprehensive studies:

1. To formulate and develop methods and metrics to quantify and analyze the sharing potential in transportation systems that can capture heterogeneity in demand patterns and network structure, and
2. To evaluate the trade-offs between user costs, operator costs, and computational costs for Corner-to-Corner (C2C) MOD services and Door-to-Door (D2D) MOD services.

The objectives of the first study that quantifies sharing potential of a city/region include:

1. Define and formulate sharing in passenger transportation systems in the form of overlaps of person-trip flows (or person-trip shareability).
2. Find the maximum extent of sharing possible in a city/region (or sharing potential) given its transportation network and mode-agnostic travel demand data.

3. Evaluate the trade-offs between maximizing overlaps and constraints on maximum permissible detours for person-trips, and
4. Derive flow-overlap-based shareability metrics at different levels of aggregation in a transportation system and validate their impact on the efficiency of Shared Mobility Mode (SMM) operations.

The objectives of the second study that evaluates C2C MOD services include:

1. Develop a scalable and effective decision policy and algorithmic approach for dynamically operating C2C MOD services, particularly C2C-Ride-pooling (or C2C-RP).
2. Compare the changes in user costs, operator costs, and computational costs between four different MOD service types, namely, Door-to-Door Ride-hailing (D2D-RH), Door-to-Door Ride-pooling (D2D-RP), Corner-to-Corner Ride-hailing (C2C-RH) and Corner-to-Corner Ride-pooling (C2C-RP), and
3. Evaluate the sensitivity of user, operator, and computational costs of C2C services to changes in first/last mile parameters and sequence of decision policy.

The contributions of the first study of this dissertation include the conceptualization of flow overlap to measure sharing potential and the definition and formulation of the Maximum Network Flow Overlap Problem (MNFLOP) as a new method to characterize and quantify the sharing potential of a transportation network-travel demand pattern pairing. This method adds a directional dimension to defining demand in a transportation network that may better characterize the extent of dispersion of trips in a network and therefore its sharing potential. To the best of my knowledge, I believe that this is the first study in

literature that characterizes the flow of travel demand in a transportation network. The MNFLOP formulation developed in the method can also be used to characterize flows in non-transportation networks, since the only inputs required are the network structure and aggregate flows in the network.

The contributions of the second study of this dissertation include the implementation of a scalable algorithm to operate C2C MOD services with dynamic and flexible PUDO locations (also known as Virtual Stops) in an agent-based transportation model for a congestible network. At the time of this study, I believe that this is the first such implementation involving dynamic PUDOs in an agent-based dynamic transportation network model with congestible links and travel time uncertainty. I also believe that this is the first study to comprehensively compare the cost trade-offs of operating MODs across two dimensions, namely sharing (Ride-pooling vs Ridesharing) and PUDO locations (D2D vs C2C).

1.3 Dissertation Outline

The rest of the dissertation is organized as follows. Chapter 2 and Chapter 3 present the first study in my dissertation involving sharing potential in transportation systems. Chapter 2 introduces the concept of flow overlap, formulates the MNFLOP, and derives metrics of shareability at different levels of aggregation in a region. This chapter also illustrates the usefulness of origin/node level shareability metrics using an example to show how different trip patterns and shareability levels can all have the same magnitude of demand density. Chapter 3 performs a comprehensive analysis of various OD demand patterns and how the metrics of shareability derived in Chapter 2 could meaningfully

differentiate between them. This chapter also performs sensitivity analysis of flow overlap with respect to maximum detour parameter, and also tests the statistical validity of flow-overlap-based shareability metrics in influencing the efficiency of a station based last-mile microtransit service.

Chapter 4 and Chapter 5 present the second study of my dissertation involving C2C MOD services. Chapter 4 states and formulates the C2C Ride-pooling (C2C-RP) problem and presents a scalable decomposition-based approach to solve the problem. Chapter 5 performs a comprehensive analysis of user, operator, and computational costs for both C2C and D2D variants of Ride-pooling (RP) and Ride-hailing (RH) services.

Chapter 6 concludes the dissertation, summarizing the key contributions and discussing the avenues to take this research forward in the future.

Chapter 2. Sharing Potential in Transportation Networks – Conceptual Framework, Models, and Metrics

2.1 Introduction

The overarching goal of this chapter is to develop metrics that quantify the sharing potential of transportation systems in the form of person-trip shareability in a manner that is agnostic of (shared mobility) mode, vehicle attributes, and operational strategy. For any subregion and data to describe its TSAs (See Figure 1.2), I aim to quantify the sharing potential of person-trips in the subregion. Put another way, if travelers (with fixed trip origins, trip destinations, and maximum willingness-to-detour but some flexibility in departure time) in a subregion are willing to, I aim to develop metrics that measure the extent to which it is possible for them to share space (i.e., overlap) in the transportation network.

2.1.1 Chapter Outline

The rest of the chapter is organized as follows: First, I define a new concept that I call person-trip shareability (Section 2.2.1) and discuss principles for operationalizing this definition (Section 2.2.2). Second, I mathematically formulate a fundamental metric called 'flow overlap' (Section 2.2.3). Third, I review in Section 2.3, the nascent literature related to shareability and delineate the academic contributions of this study in my dissertation. Fourth, I formulate the Maximum Network Flow Overlap Problem (MNFLOP) that assigns all person-trips to network paths to maximize network-wide flow overlap subject to constraints on detour distance (Section 2.4). Fifth, based on flow overlap concepts and MNFLOP output, I formulate a wide range of flow-overlap-inspired metrics at different levels of aggregation

in a transportation network, such as origin-destination (OD) pair level, location (i.e., node) level, network level, and link level (Section 2.5). Finally, Section 2.6 concludes the chapter.

2.2 Conceptual Framework

2.2.1 Defining Person-trip Shareability

This dissertation presents novel concepts and metrics related to the shareability of person-trips in a subregion. Before delving into shareability metrics, I begin with a definition of person-trip shareability.

Definition 1: Person-trip Shareability: The extent to which travelers within a subregion (with fixed trip origins and destinations) can overlap in time and space on physical links in the subregion's road network.

Notably, this definition includes person-trips instead of vehicle trips or trips with a specific mode. This distinction is in keeping with the proposed shift in transportation planning from an automobile-centric paradigm to a people/traveler-centric paradigm (Litman, 2013; Trombin et al., 2020). Moreover, the definition incorporates physical transportation links, as these links are where person-trips must overlap in time and space. The concept of overlap and the ability of trips to overlap in time and space on the road network underlies my definition of person-trip shareability.

The person-trip shareability definition also purposely uses the phrase 'can overlap,' as there is no guarantee that these person trips will overlap in time or space on physical transportation links (in shared vehicles) in the real-world. I seek to provide an upper bound on person-trip sharing potential in a subregion. Ultimately, sharing in the real world will

depend on at least two additional high-level factors, namely, the willingness of travelers to share with strangers (i.e., the behavioral side) and the combination of SMMs available to travelers in the subregion (i.e., the supply side).

Analysts can conceptualize person-trips (i.e., OD demand) in this study in two ways, depending on their needs and data availability: (i) person-trips from travelers willing to share rides with strangers or (ii) all person-trips independent of willingness to share rides. The former type of person-trips is relevant to shared mobility providers (e.g., transit agencies), as these trips represent their market. The latter set of person-trips, combined with our metrics, represent an upper bound on sharing in a subregion.

As a final note, the person-trip shareability definition does not imply the explicit pairing or matching of person-trips, in vehicles or to each other. As Section 2.3 discusses, this property differentiates our shareability framing, definition, and metrics from others in the existing literature.

2.2.2 Operationalizing Person-trip Shareability

Next, I seek to operationalize person-trip shareability consistent with the research roadmap in Figure 1.2 and the study's overarching goal. The key principles for operationalizing person-trip shareability in a subregion include:

- jointly considering the magnitude of travel demand, the spatial distribution of demand, and road network properties
- connecting demand and supply through maximizing the spatial and temporal overlap of person-trips (i.e., demand) on network links (i.e., supply)

- capturing the trade-off between maximizing overlap and constraints on maximum permissible detours for person-trips
- condensing person-trips onto as few links as possible
- considering partial sharing/overlaps, i.e., not requiring person-trips to be paired with other person-trips explicitly
- being mode-, vehicle-, and operational-policy-agnostic

2.2.3 Unit Flow Overlaps

Based on the principles for operationalizing person-trip shareability outlined in Section 2.2.2, I introduce the notion of flow overlaps. As a starting point, let us define flow overlap for a unit flow (i.e., a single person-trip) traveling from origin o to destination d on path k. Flow overlap for this unit flow is the link-length-weighted average number of other trips/flows with whom the unit flow overlaps on its path k during period t. Equation 1 displays the mathematical formulation of flow overlap for a unit flow on path $k \in K_p$ between OD pair p during time interval t, denoted Z_p^{kt}.

$$Z_p^{kt} = \sum_{a \in A} \frac{f_a^{pkt} c_a \delta_a^{pk}}{\sum_{a \in A} \left(c_a \delta_a^{pk} \right)} + \left(F_p^{kt} - 1 \right) \tag{1}$$

where, t represents the time interval for which overlap is measured; A is the set of links in the network, indexed by $a \in A$; P is the set of OD pairs in the network, indexed by $p \in P$ or $(o, d) \in P$; K_p represents a finite set of acyclic paths for OD pair p that are within a predefined maximum detour parameter Δ_{max}, indexed by $k \in K_p$; c_a gives the length of link a; δ_a^{pk} is a binary parameter that denotes the link-path incidence of link a on path k from OD

pair p; f_a^{pkt} provides the number of other person-trips on link a from all other paths and all other ODs, except for path k from OD pair p, during time interval t; and F_p^{kt} represents the number of person-trips from OD pair p on path k during time interval t. The minus one in the second term is the unit person-trip on path k for OD pair p for whom we are calculating flow overlap.

I define flow overlap for a single person-trip (i.e., a unit OD flow) during a fixed time interval t, because sharing must occur in space and time. As the metrics in this study aim to support high-level analysis of person-level sharing potential, as opposed to the analysis of low-level service operations, we are interested in the potential interaction between trips on links in the network during a reasonably large interval of time (e.g., 15-30 minutes). For simplicity, I drop the notation t from the expression for flow overlap and expressions derived from it in the rest of the dissertation.

With Equation 1, we can calculate the upper and lower bounds of overlap for a unit flow between OD pair p on path k. If the first term in Equation 1 is zero, it means that a unit flow between OD pair p on path k does not overlap with demand from any other OD pair or path associated with OD pair p. Hence, the only overlapping flows for this unit flow is $F_p^k - 1$. Hence, the lower bound of Z_p^k is zero which occurs when $F_p^k = 1$, and the first term in Equation 1 is also zero.

The upper bound of Z_p^k is $F - 1$, where F denotes the total demand (in terms of person flows) in the network. The upper bound for Z_p^k occurs when all demand in the

network is on a single path k from a single OD pair p, or when all demand, from all OD pairs in the network, use every link along path k for OD pair p.

2.2.4 Illustrative Example

While Equation 1 displays the flow overlap formula for a single trip, this subsection presents a simple network example to illustrate how to use this formula to calculate network-wide flow overlap for two different sets of path flows. I later formulate a mathematical program that maximizes cumulative network-wide flow overlap for a given network, OD demand pattern, and Δ_{max} (in Section 2.4).

Figure 2.1a and Figure 2.1b illustrate the notion of overlaps in a simple network and show how path assignment impacts flow overlaps. Figure 2.1 displays an undirected network with three origin nodes (A, B and C) and one destination node (D), with link distances as shown next to each link. Moreover, consider the case where three persons (P_1, P_2, and P_3) start their trips from origin nodes A, B, and C, respectively, and all have D as their destination node.

The shortest paths between each OD pair in Figure 2.1a are distinct and do not share any common links. Hence, assigning each person-trip to their shortest path produces zero flow overlap for each trip. In this study, link index $k = 0$ always represents the shortest path. By definition, this scenario involves zero units of detour distance.

In the second scenario, depicted in Figure 2.1b, person-trips P_1 and P_3 detour from their shortest paths and instead travel to destination node D via node B and link B-D (Path k = 1 for OD pairs AD and CD). Person-trip P_2 continues to take their shortest path to D (Path

k = 0). This network assignment of person-trips produces overlapping flows on link B-D. Person P_2 shares their trip with two others on the entirety of their trip. Persons P_1 and P_3 do not have any overlap on link AB and CB, respectively. However, from B to D (4/7th of their respective trip lengths), P_1 and P_3 share this link with two other person trips. Therefore, using Equation 1, I calculate flow overlap values for each unit trip in Figure 2.1b as follows:

$$Z_{AD}^{k=1} = Z_{CD}^{k=1} = 0 \times \frac{3}{7} + 2 \times \frac{4}{7} = 1.14$$

$$Z_{BD}^{k=0} = 2 \times \frac{4}{4} = 2$$

Hence, for the paths in Figure 2.1b, a unit trip between A and D (as well as C and D) shares its path with an average of 1.14 other person trips. Similarly, a unit trip between B and D shares its path with an average of two other person trips. Therefore, the weighted (by trips from each OD pair) average value for flow overlaps in the network is as follows:

$$Z = \frac{1 \times 1.14 + 1 \times 2 + 1 \times 1.14}{1 + 1 + 1} = \frac{4.28}{3} = 1.43$$

This value implies that an average unit trip in the network shares its path with 1.43 other person-trips in Figure 2.1b. The average detour in this case is 1.33 distance units (detour of two distance units for P_1 and P_3, and detour of zero distance units for P_2).

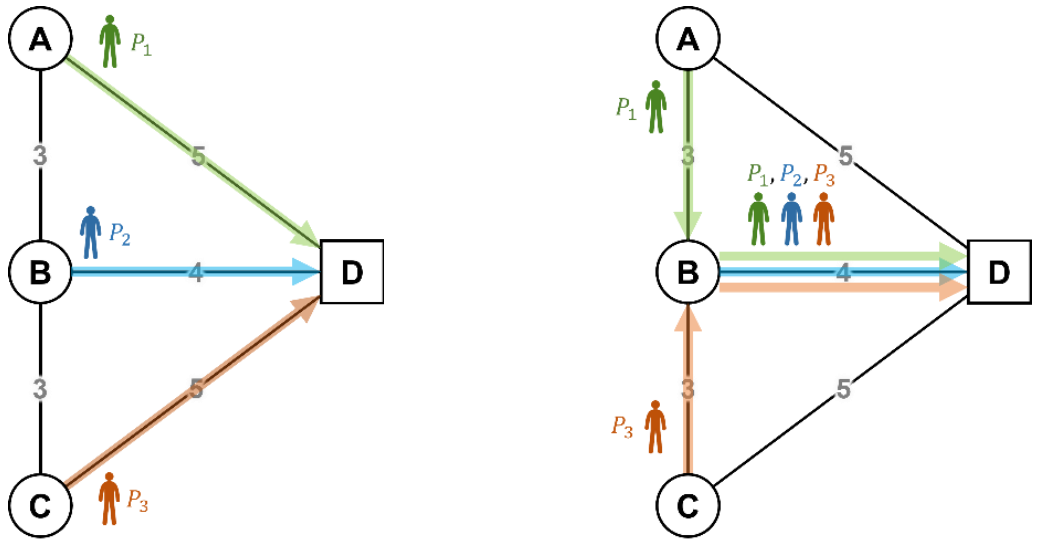

Figure 2.1 Flow assignment based on (a) shortest paths, and (b) maximum network-wide flow overlap

2.2.5 Unproductive Detours and the Least Cost Shared Path Subgraph

The increased overlap in Figure 2.1b is enabled only by allowing (or requiring) travelers to detour from their shortest paths. Hence, there is a fundamental trade-off between overlaps and detours when assigning person-trips to paths so as to maximize network-wide overlap. However, this is quite natural given that existing shared mobility services—including fixed-route public transit and on-demand Ride-pooling —often require travelers to detour in order to overlap with other travelers in time and space in the physical transportation network to permit sharing.

Fortunately, in Figure 2.1b, the detours that occur are productive detours, i.e., without detouring, P_1 and P_3 could not have overlapped with each other and with P_2. Unfortunately, it is also possible for person-trips to detour to increase network-wide flow overlap, but where the detours are unproductive, or even counter-productive. As an example, consider Figure 2.2, where there are ten person-trips from Node 2 to Node 16 and ten person-trips from Node 6 to Node 16. In Figure 2.2a, the person-trips detour onto a sub path between

Nodes 8 and 20 that is 8 miles in length compared to a shorter sub path that is 5 miles in length, shown in Figure 2.2b. In Figure 2.2a, the average flow overlap for a unit flow from Node 2 to Node 16 ($Z_{2,16}^{k}$) is 15.66 (14.83) person-trips, respectively. Despite the increase in flow overlap, clearly this is a counter-productive detour, as both sets of ten person-trips travel on longer paths, yet these detours do not produce additional overlaps.

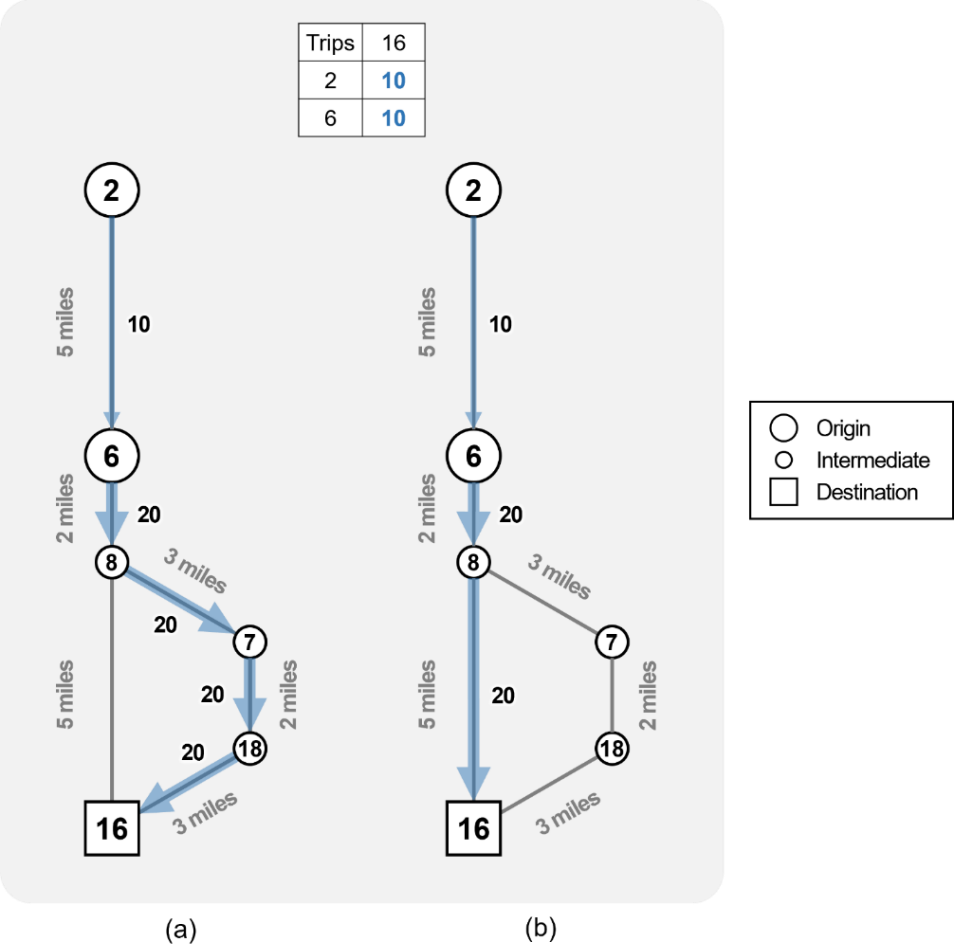

Figure 2.2 Network flows (a) with unproductive detours, and (b) without unproductive detours

To address this issue, we will only permit person trips on a subset of links in the network. I refer to this subset of links as the least cost shared path (LCSP) subgraph. The LCSP subgraph includes at least one connecting path for every OD demand pair with a detour distance less than Δ_{max}. Moreover, as the name suggests, the LCSP determines the links with

the lowest cumulative cost that satisfy the OD-pair path connection and detour distance constraints. Appendix A: Least Cost Shared Path (LCSP) Subgraph provides a mathematical formulation form the LCSP subgraph problem.

This LCSP subgraph prevents any and all unproductive detours because every link in the LCSP subgraph must have some non-negative flow. If this was not the case (i.e., the link could be excluded and every OD pair would still have a connected path with detour less than Δ_{max}), then the link would be removed, as removing it would decrease the cost of the subgraph. Hence, for travelers between OD pairs with multiple connecting paths, any detour from a shorter path will be productive as the detour will involve overlapping with travelers with which it could not otherwise overlap.

Notably, Δ_{max} has a significant impact on the LCSP subgraph. As Δ_{max} increases, more paths between each OD are feasible, thereby allowing for decreases in the cumulative link cost in the LCSP subgraph. In the case of the network and the OD demand pairs in Figure 2.1, if $\Delta_{max} < 2$, then the LCSP subgraph includes links AD, BD, and CD with LCSP cost equal to 14 distance units. However, if $\Delta_{max} \geq 2$, then the LCSP subgraph includes links AB, BC, and CD with LCSP cost equal to 10 distance units. Conversely, in the case of the network and the OD demand pairs in Figure 2.2, the LCSP is always links 2-6, 6-8, and 8-16, independent of Δ_{max}.

As a note, while it is possible to define K_p as the finite set of all acyclic paths between OD pair p, it is computationally demanding (or impossible, depending on the network size) to generate all paths. It is also unrealistic for a person trip to be assigned to a path $k \in K_p$ that detours too far away from their shortest path, unless they are given significant

compensation (König and Grippenkoven, 2020; Zhang et al., 2016). In the context of a specific shared-ride service, Lobel and Martin (2020) show that expected detours for travelers are small and penalizing detours is economically beneficial. Hence, in this study I restrict the paths in K_p to have a detour less than Δ_{max}. In practice, there are many ways to determine K_p, such as including all paths between OD pair p with an absolute detour less than x miles and/or less than y percent longer than the shortest path. In this study, I use Yen's algorithm to generate all acyclic paths between each OD pair limited to a distance threshold, as discussed in Section 2.4.4 (Yen, 1971).

2.3 Literature Review

The concept of 'shareability' has arisen in the transportation literature over the past decade with the emergence of SMMs beyond taxis and fixed-route and -schedule transit lines. Several research studies use the term shareability and/or present shareability metrics; however, these metrics and studies differ from the person-trip shareability metrics in this chapter of the dissertation. The first subsection below provides an overview of shareability metrics and studies in literature, while the second subsection differentiates my shareability metrics from those in the literature.

2.3.1 'Shareability' in the Literature

Several studies analyze the potential to reduce vehicle trips through having travelers share space in vehicles. Tsao et al. (1999) examine the potential for inter-household carpooling to reduce vehicle commuting trips. Similarly, Cici et al. (2013) estimate ride-sharing's potential to reduce vehicle trips when travelers are willing to share rides, where they use cell phone data to infer home and work locations.

Whereas Tsao et al. (1999) and Cici et al. (2013) focus on conventional ride-sharing and carpooling where both drivers and riders have their own trip origins and destinations, recent research analyzes sharing potential for mobility services with dedicated drivers (e.g. taxi or Ride-hailing drivers). Santi et al. (2014) introduce the concept of 'shareability networks', where network links connect pairwise shareable trips, to model taxi trip sharing in a deterministic setting. They apply methods from graph theory, after formulating the problem as a maximum matching problem, to determine the percentage of shareable trips in a subregion as a function of vehicle capacity and traveler willingness to detour from their preferred paths. Tachet et al. (2017) plot shareability network curves from Santi et al. (2014) for 4 cities, namely New York, Vienna, Singapore and San Francisco, where the x-axis is trip density and the y-axis is percent of shareable trips. The curves show similar patterns in all cities—sharing probability rises steadily at low densities and quickly saturates above 98%. Another recent study extends concepts from Santi et al. (2014) for a shared-ride system, wherein they jointly consider detour length and the economic value from allowing detours, and prove fundamental limits on the sum of detours and economic value for shared rides (Lobel and Martin, 2020)

Several studies focus on dynamic shared-ride mobility (or taxi) services and each find significant benefits of shared-rides over non-shared-rides. Gurumurthy & Kockelman (2018) use cell phone traces to infer OD locations in Orlando, Florida, and then they employ a dynamic ride-sharing algorithm within an agent-based simulation model to serve the OD trips. Alonso-mora et al. (2017) extend the shareability network concept to match feasible requests together and then match vehicle request sets to vehicles in the online case, for vehicles with 10-person capacities. Two other recent studies illustrate the benefits of

shared-ride over non-shared-ride services in a dynamic stochastic setting (Hyland and Mahmassani, 2020; Simonetto et al., 2019).

A pair of recent studies further extend the shareability networks in Santi et al. (2014). Kucharski and Cats (2020) develop several person-level shareability metrics using their Exact Matching of Attractive Shared rides (ExMAS) algorithm. They use total travel demand in Amsterdam and evaluate shareability purely from a demand perspective; they match shareable trips to each other rather than to fleet vehicles. Kucharski and Cats (2020) measure shareability across different scenarios by varying system parameters such as total demand, matching time window horizon, and price discount for sharing rides compared to ride hailing. Soza-Parra et al. (2022) apply the approach from Kucharski and Cats (2020) to examine shareability by varying spatial distributions of trip patterns for different number of urban centers, destinations, and trip length distributions.

Although the focus is not on shareability, there is another stream of research that aims to characterize urban forms, transportation networks, and/or travel patterns. Tsai (2005) and Oke et al. (2019) develop aggregate region level metrics and create typologies to classify cities based on their urban form and transportation system characteristics. Such studies provide valuable insights into the general characteristics of an urban area that may be valuable for the design of SMMs. For example, empirical research finds that fixed transit is more effective in highly clustered cities compared to sprawled-out cities (Cervero and Seskin, 1995; Dill et al., 2013; Guerra et al., 2018; McIntosh et al., 2014; Stead and Marshall, 2001). Similarly, studies such as Srinivasan (2002) and Ewing & Cervero (2010) show that urban form and the built environment have a strong influence on travel behavior (e.g. mode

choice). Other studies such as Yang et al. (2016) and Saberi et al. (2017) characterize urban mobility patterns using trip data and derive metrics to identify activity location clusters and their interactions. Finally, Schieber et al. (2017) focus on network topology and characterize transportation network structures, but they do not consider demand.

2.3.2 Spanning Trees and Subgraphs

Section 2.2.5 introduces LCSP subgraphs and Appendix A: Least Cost Shared Path (LCSP) Subgraph provides the associated math program. The LCSP subgraph problem is related to the well-known minimum spanning tree problem (Kruskal, 1956; Nesetril and Nesetrilová, 2023; Prim, 1957), and the Steiner tree problem (STP). The STP determines the least cost tree that connects a subset of vertices (called terminals) on a given undirected graph with non-negative edge weights (van Oudheusden, 1995). While there are extensions of the STP involving directed graphs (Charikar et al., 1998; Hsieh et al., 2006; Zelikovsky, 1997), and budget constraints on terminal-to-terminal travel times (Costa et al., 2009; Moss and Rabani, 2007), the LCSP formulation further extends the STP. The LCSP applies to directed graphs with multiple OD pairs (not just a list of vertices or terminals), and constraints on the maximum path cost (or distance) between each OD pair. While the output of the minimum spanning tree problem and the STP are trees, as their names suggest, the output of the LCSP subgraph problem is a directed subgraph. Hence, more than one path may connect an OD pair in an LCSP subgraph.

2.3.3 Contributions and Research Gaps Addressed

This study makes several contributions to the existing literature. First, I believe this is the first study to conceptualize, define, and formulate the person-trip sharing potential

(i.e., shareability) of a subregion (i) considering the overlap of person-trip flows on network links, and (ii) with partial-sharing between person-trips. No other studies capture either of these two features. I expand on this statement below.

In addition to not focusing on shareability, the studies that aim to characterize urban forms, transport networks, and traveler patterns in Section 2.3.1 either quantify the urban form and land use alone (Srinivasan, 2002; Tsai, 2005); the transportation network alone (Schieber et al., 2017); or travel patterns alone without considering the underlying transportation network (Yang et al., 2016). Even though Saberi et al. (2017) model mobility patterns as a complex network of flows and compute metrics using shortest network path distances, they do not capture how trips from different OD pairs might interact (and overlap) within the network.

Recent work by Kucharski and Cats (2020) and a follow-up study by Soza-Parra et al. (2022) are the most closely related to the current study. While these two studies measure shareability by analyzing person-trips independent of vehicles, their shareability metrics do not model travel demand as flows through a network. Moreover, their shareability metrics—which include reduction in VMT, number of shareable rides, and total passenger utility—are fundamentally different from the person-trip shareability metrics proposed in this dissertation. Kucharski and Cats (2020) shareability metrics are based on the number of requests that can be pooled together in a given time window, whereas the shareability metrics proposed in this dissertation are based on the potential spatial and temporal overlap of person-trip flows on network links.

The major practical difference between Kucharski and Cats (2020) and the current study is that I consider 'partial' shareability between person-trips in the network. For example, in my model-based metric, a traveler assigned to a network path with, for example, fifteen links (i.e., 15 street segments) may overlap (i.e., share space) with large numbers of person-trips from many OD pairs for only a link or two, before their paths divert (or end). While I capture this partial shareability/overlap explicitly, Kucharski and Cats (2020) and all other shareability metrics in the literature, including Santi et al. (2014) and Tachet et al. (2017), require explicit matching between two-or-more travelers. This difference stems from these other studies modeling individual trips, whereas I model OD demand flows.

The partial shareability property in our metrics is amenable to informing high-capacity fixed-route/-schedule transit and semi-flexible transit where travelers can and do transfer. In fact, my shareability metric is consistent with the concept of Mobility-as-a-Service, where travelers might need to transfer between modes, transit lines, and vehicles, as they travel from their origin to their destination through a physical network.

Second, the current study defines and formulates the MNFLOP—a novel mathematical program able to quantify the maximum network-wide flow overlap (i.e., this study's measure of shareability) of a network and OD demand, jointly, subject to hard constraints on maximum person-trip detour distance. As far as I know, this formulation represents a novel network model. Moreover, this study also analyzes the sensitivity of maximum shareability in a given network to changes in the maximum allowed detour. Another valuable feature of MNFLOP is that it applies to any network with a set of flows, not just transportation networks. Traditionally, constraints and objective functions in network

flow models try to avoid, either explicitly or implicitly, condensing flows onto few links, because most network flow applications involve link or node capacity constraints that are binding and/or flow-based link or node congestion penalty functions. However, as indicated in this study, there are potential network flow applications where condensing flows onto fewer and fewer links is an appropriate objective. I believe MNFLOP may apply outside of the transportation domain, as many other network models do.

Third, this study introduces a node-/location-level metric that jointly quantifies the spatial distribution of demand from one node/location and the underlying transportation network, termed *demand dispersion*. Much of the existing literature that aims to characterize demand for shared mobility systems only considers the magnitude of demand emanating from a single homogenous subregion, e.g., Li & Quadrifoglio (2010), Nourbakhsh & Ouyang (2012), and Quadrifoglio & Li (2009).

2.4 Mathematical Formulation

This section utilizes the concept and formulation of flow overlap to create a mathematical program that maximizes network-wide flow overlap, called the MNFLOP—Maximum Network Flow Overlap Problem. Cumulative network-wide flow overlap is one of the study's primary person-trip shareability metrics. In the following four subsections, I state the maximum network flow assignment problem, formulate the math program for MNFLOP, discuss computational complexity considerations, and describe the solution method, respectively.

2.4.1 Problem Statement

Let the directed graph $G = (N, A)$ with nodes N and directed street segments A, represent a road network. Moreover, let there be a set of origin-destination (OD) demand pairs P, indexed by $p \in P$ or $(o, d) \in P$, where $o \in O \subseteq N$ and $d \in D \subseteq N$. Each link $a \in A$ has a non-negative link cost, c_a, that measures the length of the link. Each OD pair has a demand flow (measured in person-trips), F_p, and a set of acyclic paths, K_p. Each path $k \in K_p$ has a detour distance, Δ_p^k, relative to the shortest path connecting OD pair p, which is less than the maximum detour distance for person-trips, Δ_{max}. Given this information, the problem is to determine (i) the flow assigned to each path $k \in K_p$ for every OD pair, $p \in P$, which I denote x_p^k, and (ii) whether each link $a \in A$ is included in the LCSP subgraph, where the binary variable y_a equals one if the LCSP subgraph includes the link and zero if not, such that: cumulative network-wide flow overlap is maximized; the sum of the link costs in the LCSP subgraph is minimized; all person-trips between each OD pair p are assigned to a path $k \in K_p$; and there is a connected path between each OD pair p.

2.4.2 Problem Formulation

Given the decision variables, parameters, constraints, and objective in the previous section, I formulate MNFLOP as a math program, as shown in Equation 3 through Equation 8. Below, I first explain the objective function (written in both Equation 2 and Equation 3), then each of the constraints (Equation 4 to Equation 8).

$$Max\ Z = \sum_{p \in P} \sum_{k \in K_p} x_p^k \left(Z_p^k - \epsilon \cdot \Delta_p^k \right) - M_1 \sum_{a \in A} c_a\, y_a \tag{2}$$

The first term in the objective, $\sum_{p \in P} \sum_{k \in K_p} x_p^k Z_p^k$, is the network-wide flow overlap. Equation 1 defines Z_p^k, the flow overlap for a unit flow between OD pair p on path k. Multiplying Z_p^k by the flow assigned to path k for OD pair p, and summing over all OD pairs P, and all paths K_p gives the network-wide flow overlap.

The second term in the objective function, $\sum_{p \in P} \sum_{k \in K_p} x_p^k \cdot \epsilon \cdot \Delta_p^k$, effectively breaks ties between two solutions with the same amount of network-wide flow overlap. Hence, ϵ is a small positive real number. Among alternative solutions with the same network-wide flow overlap, the second term ensures the selection of the one with the shortest cumulative detour.

The last term in the objective function, $M_1 \sum_{a \in A} c_a \, y_a$, denotes the total cost of all the links in the LCSP subgraph, multiplied by a large positive value, M_1. This term ensures that (i) the cumulative cost (distance) of used links is minimized, and, therefore, (ii) the final flow assignment does not include any unproductive detours—the issue identified in in Section 2.2.5.

I include this term in the objective function with the large weight M_1, as opposed to first determining the LCSP subgraph (and then determining the path flows on the LCSP subgraph that maximize flow overlap), because it is possible that there are multiple subgraphs with the same 'least cost'. Among the possible least cost subgraphs, we want the one that permits the most flow overlap.

The value for M_1 can be set as the largest possible increase in the value of the overlap term in the objective function between any two feasible solutions, which is $F \cdot (F-1)$ where $F = \sum_{p \in P} F_p$ denotes the total demand (person-trips) in the network.

Equation 3 replaces the Z_p^k in Equation 2 with the right-hand side of Equation 1. In Equation 1, f_a^{pk} and F_p^k are given; however, in Equation 3, f_a^{pk} is an auxiliary decision variable and the decision variable x_p^k replaces F_p^k. Equation 4 displays the formula for the auxiliary decision variables, f_a^{pk}. We compute the auxiliary decision variable f_a^{pk} for all links $a \in A$ that belong to each path $k \in K_p$ from all OD pairs $p \in P$. The variable f_a^{pk}represents the overlapping flows on link a for a particular path $k \in K_p$ from OD pair $p \in P$, where the sum is over every path for every OD pair, except for path $k \in K_p$ from OD pair $p \in P$. Since the path assignment for maximum flow overlap of an OD pair depends on the path assignment of other OD pairs, this implies that MNFLOP is a Quadratic Program.

$$Max\ Z = \sum_{p \in P} \sum_{k \in K_p} x_p^k \left(\sum_{a \in A} \frac{f_a^{pk} c_a \delta_a^{pk}}{\sum_{a \in A} \left(c_a \delta_a^{pk} \right)} + \left(x_p^k - 1 \right) - \epsilon \cdot \Delta_p^k \right) - M_1 \sum_{a \in A} c_a\, y_a \qquad (3)$$

s.t.

$$f_a^{pk} = \sum_{q \in P} \sum_{h \in K_q | k} x_q^h * \delta_a^{qh} \quad \forall a \in A, \forall p \in P, \forall k \in K_p \qquad (4)$$

$$\sum_{k \in K_p} x_p^k = F_p \quad \forall p \in P \qquad (5)$$

$$\sum_{p \in P} \sum_{k \in K_p} x_p^k \delta_a^{pk} \leq M_2 \cdot y_a \quad \forall a \in A \qquad (6)$$

$$x_p^k \geq 0 \quad \forall p \in P, \forall k \in K_p \tag{7}$$

$$y_a \in \{0,1\} \quad \forall a \in A \tag{8}$$

Equation 5 requires that the sum of flows across all paths $k \in K_p$ between an OD pair p, x_p^k, equal the total demand (person flows) for the same OD pair, F_p. Equation 7 mandates non-negative path flows for all paths from all OD pairs.

Equation 6 ensure that if a path $k \in K_p$ has a positive flow value (i.e., $x_p^k > 0$), then all the links on path k are included in the LCSP subgraph. M_2 is a large positive number whose value can be set to the maximum number of person-flows on any link in the network, which would be equal to the total demand flows between all OD pairs in the network, $M_2 = F = \sum_{p \in P} F_p$. Equation 8 enforces the binary constraint on decision variable y_a.

2.4.3 Computational Complexity Considerations

This section provides a brief overview of the computational complexity associated with the MNFLOP. If a network has $|P|$ OD pairs and |K| paths between each OD pair, then the number of path flow choice decision variables in the optimization problem is $|P| \times |K|$. The number of path flow decision variables is independent of the number of links in the network and depends only on the number of OD pairs and the paths between them. However, the quadratic nature of the objective function in Equation 3 means that an optimal path flow for an OD pair is strongly influenced by the optimal path flows for all other OD pairs with which there is an overlapping link on their paths. The number of quadratic terms in the objective function increases as the OD pairs increase in spatial closeness, as this results in more overlapping links between paths from multiple OD pairs. Naturally, the number of

quadratic terms also increases with network size, as larger networks have more links on which different OD pairs are likely to have overlapping paths.

For a network with |P| OD pairs and |K| paths between each OD pair, the MNFLOP could have up to a maximum of $\binom{|P|\times|K|}{2}$ quadratic terms in the objective function. $\binom{|P|\times|K|}{2}$ is the number of combinations of overlapping paths. More quadratic terms results in slower convergence while trying to find an exact optimal solution. The LCSP sub-problem in MNFLOP is also computationally intensive. Heuristics and approximation algorithms may be required to solve MNFLOP ensuring productive detours for large networks and a greater number of origin/destination nodes.

2.4.4 Solution Methodology

MNFLOP instances are solved as constrained quadratic programs using the Gurobi optimization package in Python programming language. I use the Networkx package to create the network and perform other network-related operations. As an input to the MNFLOP, a finite set of acyclic paths are pre-computed for each OD pair: $K_{od} = 0,1,2,..$ indexed by $k \in K_{OD}$, where $k = 0$ denotes the shortest path between OD pair (o, d), $k = 1$ denotes the second shortest path and so on. A variation of the Yen's algorithm (Yen, 1971) is used to find all acyclic paths between each OD pair with distance detours less than Δ_{max}. The maximum allowable detour distance could either be absolute (measured in miles) or relative percentage increase in detour distance (measured in percentage). Here detour distance for each path $k \geq 1$ is relative to the shortest path (k = 0).

Unfortunately, the number of paths increases exponentially with network size. Moreover, given the nature of MNFLOP, it is important for the set of k paths to be distinct for

each OD pair. As such, future research that applies MNFLOP to large networks must grapple with the challenge of determining a good set of paths for each OD pair.

2.5 Shareability Metrics from MNFLOP output

The MNFLOP model output can provide a variety of shareability metrics at various levels of aggregation, such as the OD level, link level, node level (origin or destination), and the network level. Table 2.1 provides a comprehensive list, along with descriptions and formulations, of shareability metrics.

Since flow overlap is defined at the OD level (for a unit trip), most of the OD-level metrics in Table 2.1 come directly from the Conceptual Framework section (Section 2.2). To clearly distinguish and show the relation between shareability metrics calculated at different levels of aggregation, I use the notation od instead of p to denote an OD demand pair (o, d). Flow overlap, Z_{od}^{k}, is from Equation 1; trip overlap percentage ($Z_{od}^{\mathrm{k,\%}}$) normalizes flow overlap for an OD pair by all OD demand; detour distance for a path k between an OD pair (Δ_{OD}^{k}) is straightforward and measured relative to the shortest path distance; the marginal overlap ($M_{Z_{od}^{k}}$) captures the ratio of overlap increase to detour distance when switching from the shortest path to path k for an OD; and finally, overlap distance (L_{od}^{k}) parallels flow overlap but instead of dividing the sum product of link flows and link distances by total path distance, overlap distance divides the sum product by total OD demand.

The node-level and network-level metrics parallel the OD-level metrics. In fact, all the node-level and nearly all the network-level metrics can be derived by summing the OD-level metrics.

At a node level, the metrics of shareability characterize the 'dispersion' of trips originating from or destined to a given node. Trips originating from a node exhibit a high degree of dispersion when their trip-ends are scattered to destinations in the network such that the paths to these destinations have limited overlaps. Conversely, when trips from a node are bound to a limited number of destinations and/or the paths to the destinations are highly overlapping, trips from the node exhibit low dispersion. The node overlap percentage metric ($Z_o^{\%}$) shown in Table 2.2 quantifies dispersion of trips from a single node for an illustrative example (Figure 2.3). A high value of node overlap percentage indicates that trips from the node are highly concentrated. Additionally, node overlap distance (L_o) denotes the average distance for which trips originating from a node that share paths with other trips originating from the same node. To understand the difference between node overlap and node overlap distance, consider nodes Node-A and Node-B that have similar overlap and overlap percentage. If Node-A has a higher node overlap distance, then trips from Node-A share paths for longer distances than trips from Node-B.

At the link level, shareability metrics include the number of person-trips on each link, as well as the number of links in the network that have non-zero flow values. The link flows can also be aggregated at the network level to find the link-length-weighted average number of person-trips on links with non-zero flows. A higher number of person-trips per link with non-zero flows indicate a higher concentration of trips in a region onto fewer links and potentially fewer corridors, the objective of MNFLOP.

Table 2.1 MNFLOP based Shareability metrics at different levels of aggregation in a network

Metric	Unit	Aggregation	Description	Formula
Trip Overlap	Person-trips	OD	Average number of other flows with which a unit trip from O to D on path k shares its path (Equation 1)	Z_{od}^{k}
Trip Overlap Percentage	%	OD	Percentage of other flows in the network with which a unit trip from O to D on path k shares its path, where F is total OD flow.	$Z_{od}^{k,\%} = \frac{Z_{od}^{k}}{F-1}100$
Detour	Miles	OD	Difference in distances between path k and the shortest path from O to D	Δ_{od}^{k}
Marginal Overlap	Person-trips/mile	OD	Marginal change in overlapping flows for every detour mile obtained from shifting from the Shortest Path to path k for an OD	$M_{Z_{od}^{k}} = \frac{Z_{od}^{k} - Z_{od}^{SP}}{\Delta_{od}^{k}}$
Overlap Distance	Miles	OD	Average distance a unit trip on path k from O to D shares with all other flows in the network	$L_{od}^{k} = \sum_{\forall a \in A} \frac{c_a * \delta_a^{odk} * (f_a^{odk} + F_{od}^{k} - 1)}{F-1}$
Node Overlap	Person-trips	Origin Node	Average number of other flows from origin O that a unit flow originating from the same node shares its path.	$Z_o = \frac{1}{F_o} \sum_{\forall d \in D} \sum_{\forall k \in K_{od}} Z_{od}^{k} F_{od}^{k}$
Node Overlap Percentage (Dispersion)	%	Origin Node	Percentage of demand that originates from a node that shares paths in whole or in part with other flows from the same node.	$Z_o^{\%} = \frac{1}{F_o} \sum_{\forall d \in D} \sum_{\forall k \in K_{od}} Z_{od}^{k,\%} F_{od}^{k}$
Node Overlap Distance	Miles	Origin Node	Average distance a person-trip overlaps with all other trips from the same origin.	$L_o = \frac{1}{F_o} \sum_{\forall d \in D} \sum_{\forall k \in K_{od}} L_{od}^{k} F_{od}^{k}$
Average Network Overlap	Person-trips	Network	Average number of other flows with which a unit trip in the network shares its path	$Z = \frac{1}{F} \sum_{\forall o \in O} \sum_{\forall d \in D} \sum_{\forall k \in K_{od}} Z_{od}^{k} F_{od}^{k}$
Average Network Overlap Percentage	%	Network	Percentage of total demand in the network that share paths	$Z^{\%} = \frac{1}{F} \sum_{\forall o \in O} \sum_{\forall d \in D} \sum_{\forall k \in K_{od}} Z_{od}^{k,\%} F_{od}^{k}$
Average Network Detour	Miles	Network	Demand weighted average of detour miles for all flows in the network	$\Delta = \frac{1}{F} \sum_{\forall o \in O} \sum_{\forall d \in D} \sum_{\forall k \in K_{od}} \Delta_{od}^{k} F_{od}^{k}$
Average Network Detour Ratio	-	Network	Ratio of average trip length in MNFLOP assignment to the average length of trips in Shortest Path assignment	$\Delta_r = \frac{Avg\ Trip\ Dist_{MNFLOP}}{Avg\ Trip\ Dist_{SP}}$
Marginal Network Overlap	Person-trips/mile	Network	Total increase in overlap in the network using MNFLOP assignment for a unit detour from Shortest Path	$M_Z = \frac{Z^{MNFLOP} - Z^{SP}}{\Delta}$
Marginal Network Overlap Percentage	%	Network	Total increase in percentage of overlapping flows in the network using MNFLOP assignment for a unit detour from Shortest Path	$M_{Z_\%} = \frac{Z_\%^{MNFLOP} - Z_\%^{SP}}{\Delta}$
Links Used	Number	Link/Network	Number of links with non-zero person-trip flows	-
Avg Link Flows	Person-trips	Link/Network	Link-length-weighted-average person-trips on links that have positive flows	-

2.5.1 Toy Example to Demonstrate Location/Node-Level Metrics

The node level metrics of shareability presented in this chapter consider the magnitude of demand, the spatial distribution of demand and the underlying road network, as well as the overlapping distance of demand emanating from a node, rather than just demand density (trips per time unit per area unit).

Figure 2.3 displays four different toy networks to illustrate the calculation of node-level measures and their potential value for characterizing shareability from a single node. Notably, the origin nodes in the four scenarios (Nodes A_1 to A_4 in Scenarios 1 to 4 respectively) have the same demand density originating from them: 90 person-trips per time unit per area unit. In scenario (1), all flows from Node-A_1 are bound to Node-B. In scenarios (2) to (4), flows from Nodes A_2, A_3, and A_4 are equally distributed among nodes B, C and D. A visual inspection of the scenarios in Figure 2.3, confirmed by the values in Table 2.2, indicate clear differences in flow overlap and dispersion, and therefore differences in shareability across the scenarios, despite the same magnitude of total demand in each case.

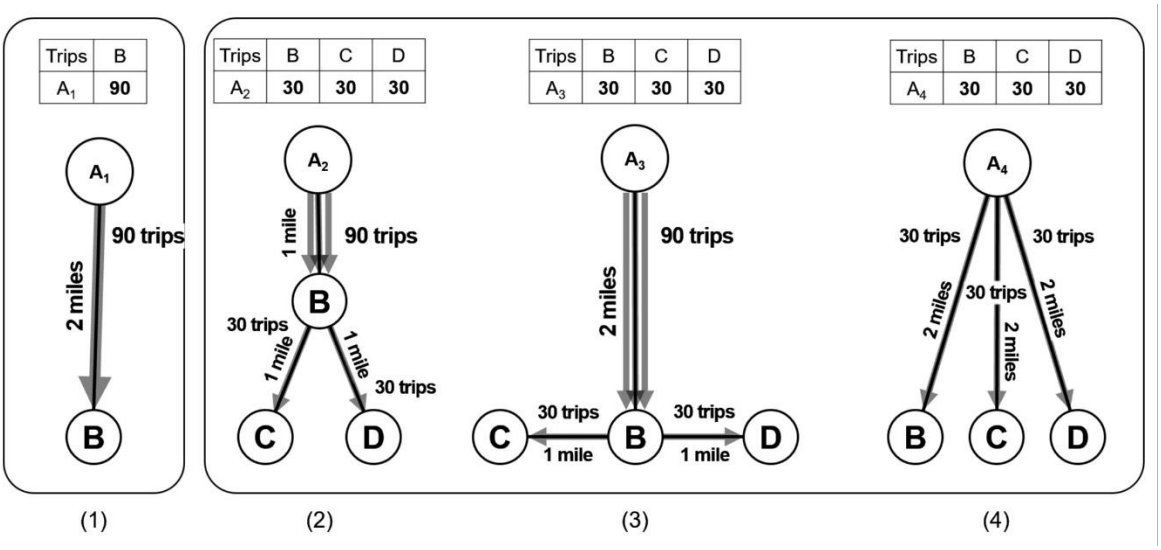

Figure 2.3 Example Scenarios to Illustrate Overlap/Dispersion and Overlap Distance at Node Level

In Scenario (1), all trips are bound to a single destination; hence, all flows share the same path, making the demand emanating from Node-A_1 highly concentrated; the overlap percentage is 100% (i.e., zero dispersion). A unit trip in this scenario overlaps with 89 other trips. Overlap distance in this scenario is the same as the average trip distance since a unit trip from Node-A overlaps with all other flows from Node-A_1 on its entire path.

Table 2.2 Comparison of Node Level Shareability Metrics for Example Scenario

Metrics of Shareability for Node	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Total Demand (Person-trips)	*90*	*90*	*90*	*90*
Avg Overlap (Person-trips)	89	69	75.7	29
Avg Overlap Percentage (%)	100	77.5	85	32.6
Avg Trip Distance (Miles)	2	1.67	2.67	2
Avg Overlap Dist. (Miles)	2	1.22	2.22	0.65

Scenario (4) exhibits the highest dispersion of trips because a unit trip from Node-A_4 on average shares paths with only 29 other trips (or 32.6% of all trips in the network) for an average overlap distance of 0.65 miles. The average overlap distance in this scenario being much less than the average trip distance indicates that each trip has a small overlap with other trips. Even though there are no overlapping paths between the three OD pairs in Scenario (4), a unit trip from Node-A_4 in this scenario still overlaps with 29 other trips heading to the same destination.

Scenarios (2) and (3) have similar network structures except for the length of the link connecting Node-A_2 and Node-A_3 respectively with Node-B. The values of overlap and overlap percentage are slightly higher for scenario (3) than scenario (2) because flows share paths for a longer distance on link A_3B in scenario (3) compared to link A_2B in scenario (2). The value of average overlap distance for scenario (3) is one-mile more than in scenario (2),

which means that the average trip from Node-A_3 in scenario (3) overlaps with all other flows for a mile longer than an average trip from Node- A_2 does in scenario (2).

The node level measures of overlap provide some interesting implications for the efficiency of shared mobility services operated in a location. Even though a lower demand density in an area (trip requests per unit area per unit time) is likely to be less favorable for ridesharing services, such methods do not fully capture the direction of the demand and thus the extent of the demand's overlap and sharing potential (Li and Quadrifoglio, 2010; Mehran et al., 2020; Qiu et al., 2015; Quadrifoglio and Li, 2009; Ronald et al., 2013; Tong et al., 2017; Zheng, 2018). Similarly, existing studies on multi-modal design of transit and microtransit services generally pair a microtransit service to a fixed transit route and derive the optimal operational parameters of the vehicles(Chen and Nie, 2018, 2017a, 2017b; Luo and Nie, 2020; Pinto et al., 2019; Stiglic et al., 2018). However, such studies are applied to small sub-regions of a city without considering the overall travel demand patterns in the region. The routes of fixed transit services in many cities were designed before the wider availability of other shared mobility modes that could either substitute or complement fixed transit. The node level example scenario presented in this section indicates how the same total demand density in an area can lead to different levels of shareability and hence implications on the viability and operational efficiency of such services. This is further examined in detail (along with other scenarios) in the next chapter of the dissertation.

2.6 Conclusion

This chapter introduced the concept of flow overlap based person-trip shareability to measure sharing potential in transportation systems. The chapter used this conceptual

framework to formulate the Maximum Network Flow Overlap Problem (MNFLOP) that assigns person-trips (flows) between different OD pairs onto different paths that maximize network level overlap. This represents the ceiling for sharing in transportation systems, given the structure of the transportation street network, OD person-trip travel demand and the maximum detour distance parameters. I used path flow results of MNFLOP to derive metrics of shareability at different levels of aggregation in the network. I finally provided an illustrative example to demonstrate how shareability measures calculated using MNFLOP at a location/node level could distinguish between multiple scenarios with the same magnitude of demand density. The next chapter of this dissertation performs extensive analysis of MNFLOP-based overlap measures for different scenarios and also tests its statistcal validity in influencing the efficiency of operating shared mobility modes.

Chapter 3. MNFLOP based Shareability Metrics – Case Studies and Validation

3.1 Introduction

Overlap-based shareability metrics presented in Chapter 2 capture the sharing potential of trips in a region by capturing overlaps in path flows between different OD pairs in a network. Given the road network structure of a region, its OD person-trip travel demand, and the maximum detour distance willingness for trips between each OD pair, MNFLOP finds the optimal path flows between each OD pair that maximizes flow overlap (or shareability) in the network. The results of MNFLOP can determine the sharing potential of a region at different levels of aggregation. This chapter takes forward the conceptual framework, problem formulation and metrics defined in the previous chapter and applies it to several different scenarios. The objective of this chapter is to apply MNFLOP-based metrics to different scenarios to examine how meaningful the metrics are in capturing variations in travel demand patterns and test the sensitivity of shareability to maximum detour willingness. This chapter also aims to test the statistical validity of MNFLOP based shareability metrics in affecting the operational efficiency of a shared mobility mode.

The rest of this chapter is organized as follows. Section 3.2 briefly describes the MNFLOP implementation methodology and also describes the various scenarios for analysis. In Section 3.3, I **verify** the ability of MNFLOP (i) to increase flow overlaps compared to shortest path assignment; (ii) to differentiate between demand patterns in terms of their overlap; and (iii) to capture the magnitude and directionality of demand emanating from a single node/location while considering the road network. Moreover, in Section 3.4, I validate

maximum overlap as a useful measure of sharing potential and estimate the elasticity of vehicle fleet miles with respect to shareability metrics for a microtransit last-mile transit feeder service. Section 3.5 discusses the various inferences that can be drawn from the results and also lists some limitations of the study. Finally, Section 3.6 concludes the first study in my dissertation (Chapter 2 and Chapter 3).

3.2 Case Studies

3.2.1 Study Network

To verify MNFLOP and the associated metrics and to illustrate their usefulness for characterizing shareability, this study employs the Sioux Falls road network (Stabler, 2019). The Sioux Falls network includes 24 nodes and 76 directional links (Figure 3.1).

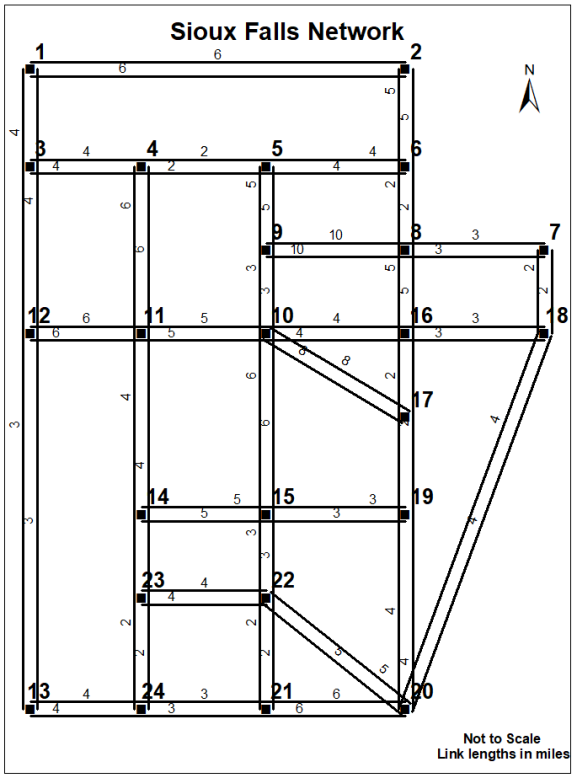

Figure 3.1 Sioux Falls Street Network

The study employs Gurobi's quadratic programming solver to solve the MNFLOP model. The solver runs on a system with 64 GB RAM and an Intel i9 processor with a clock speed of 3.60 GHz. The solver took at most 10 minutes to converge to the optimal solution across all scenarios (Sections 3.3.1 to 3.3.4) with a relative optimal gap value of 1.00e-04.

As described in the previous chapter, the solution of MNFLOP depends on Δ_{max}. I use Yen's algorithm to find all paths between each pair of demand nodes in the network with

detours less than Δ_{max} (Yen, 1971). I filter the generated paths based on the minimum of absolute and relative maximum detour distance constraints (denoted by Δ_{max}^{abs} and Δ_{max}^{rel} respectively). Unless specified otherwise, this study uses $\Delta_{max}^{abs} = 8$ miles and $\Delta_{max}^{rel} = 50\%$. I chose 8 miles because it represents approximately one-third of the maximum distance between any O-D pair in the study network (23 miles). I chose 50% because I believe it represents a reasonable value for travelers who are willing to share rides. In practice, these parameter values should be set based on a variety of behavioral and network factors. Travelers in one city may be more willing to take a detour than travelers in another city. Moreover, travelers may be more likely to accept longer detours if they are financially compensated for longer detours. Transportation planners would benefit from collecting data on the willingness of travelers to detour in their respective cities, before applying the MNFLOP.

3.2.2 Scenarios for Analysis

This chapter compares the output path flows and the resultant shareability metrics from these path flows based on two types of flow assignment approaches:

1. **Maximum Network Flow Overlap (MNFLOP) Assignment:** Trips between OD pairs are assigned to paths to maximize the total flow overlap in the network. The optimal path choice solution depends on Δ_{max}, thus evaluate the trade-offs between overlaps and detours by changing the Δ_{max} parameter and restricting the set of paths K_{od} for each OD pair with detours less than Δ_{max}.

2. **Shortest Path (SP) Assignment:** All trips between an OD pair are assigned to the shortest path between their OD pair. This is achieved by restricting the input set of

paths for each OD pair to only those with 0 detour. It should also be noted that the objective function formulation in Equation 3 ensures overlap maximizing optimal paths (path flows) for each OD pair if there are multiple paths between an OD pair with the same path length.

Table 3.1 provides a summary of the various scenarios analyzed in this study. The scenarios vary in terms of Origin and Destination nodes, total demand, maximum detour, and flow assignment methods. The objective of Scenario Set-1 (Section 3.3.1) is to evaluate how optimal path flows and shareability metrics vary for different maximum relative detour values (including the no-detour shortest path). Scenario Set-2 (Section 3.3.2) aims to demonstrate how the origin level overlap measures for trips starting from each of the 24 nodes varies using MNFLOP as well as SP assignment. Scenario Set-2 also aims to show how demand from a single location can be expressed in terms of both its magnitude as well as dispersion (in terms of overlap percentage) where the dispersion measure comes from the output of MNFLOP. The objective of Scenario Sets 3 and 4 (Sections 3.3.3 and 3.3.4 respectively) is to demonstrate that MNFLOP can differentiate between the shareability/overlap in a network under the same (or similar) magnitude of total demand but different spatial distributions of demand. Section 3.4 aims to validate the usefulness of MNFLOP-based origin-level shareability metrics by testing the hypothesis that these metrics are positively associated with the efficiency of a microtransit last-mile transit feeder service. Demand flows are expressed as person-trips.

Table 3.1 Summary of Scenarios for Analysis

Scenario	Description	Maximum Overlap (MNFLOP)	Shortest Path (SP)
Scenario 1	6 Origin Nodes in North to 6 Destination Nodes in South (Total Demand = 6,000 trips)	Yes (Δ_{max}^{rel} = 5 to 75%)	Yes (Δ_{max} = 0)
Scenario Set 2	Demand from each Origin Node to all 23 other Destination Nodes	Yes (Δ_{max}^{abs} = 8 miles, Δ_{max}^{rel}= 50%)	Yes (Δ_{max} = 0)
Scenario Set 3	Same Total Demand (1000 trips), Same Set of Origin Nodes, Different Destination Nodes	Yes (Δ_{max}^{abs} = 8 miles, Δ_{max}^{rel}= 50%)	No
Scenario Set 4	Same Total Demand (1000 trips) Distributed between Different OD pairs	Yes (Δ_{max}^{abs} = 8 miles, Δ_{max}^{rel}= 50%)	No

3.3 Results

3.3.1 Scenario Set 1: Comparing Assignment Methods in Baseline Scenario

In this scenario, I assign trips between 6 origin nodes in the North and 6 destination nodes in the South of the Sioux Falls network for three different values of Δ_{max}^{rel} – 50%, 25% and 0% (SP).

Figure 3.2 show the link flows on the Sioux Falls network for MNFLOP-50%, MNFLOP-25% and SP assignments (from left to right). Link thickness is proportional to the number of trips assigned on each link. Examining the link flows from the rightmost to the leftmost plots clearly demonstrates the ability of MNFLOP-based assignment to increase flow overlap as Δ_{max}^{rel} increases. Increasing Δ_{max}^{rel} leads to a higher chance of overlapping with more trips from other OD pairs. The plots also show that MNFLOP-50% concentrates trips between O-D pairs onto the fewest links, followed closely by MNFLOP-25%, whereas SP spreads person-trips across many links.

Table 3.2 provides the network level shareability metrics for Scenario 1. Many of the metrics directly confirm the conclusions derived from analyzing

Figure 3.2. Namely, MNFLOP-50% has the highest flow overlap, followed by MNFLOP-25%. However, also indicates that this increase in flow overlap comes at a small cost in terms of detour distance for the Sioux Falls network.

According to Table 3.2, MNFLOP-50% more than doubles the number of overlapping flows for an average unit flow in the network compared to SP (1740 person trips to 3613 person trips) for a mere 11% increase in the average detour. This suggests that small detours from shortest paths can significantly increase flow overlaps. As expected, MNFLOP-25% assignment produces slightly less overlap and lower detours than MNFLOP-50%. However, MNFLOP-25% still increases overlap by nearly 80% with only a 5% increase in distance compared to SP. The results indicate that path assignments based on MNFLOP-25% have a higher value for marginal increase in overlap per detour mile as well as a higher value for distance elasticity of overlap compared to MNFLOP-50% (12.3 vs 10.1).

The average overlap distance for a unit trip in the network increases from 4 miles to 9.6 miles with MNFLOP-50% compared to SP assignment. This implies that moving flows away from the shortest path not only increases the number of overlapping trips, but also the distance along the path in which the average person-trip shares path with all other trips.

Table 3.2 also shows that total cost of the LCSP subgraph (the sum of lengths of links with non-zero flows) decreases with increasing Δ_{max}^{rel}. The average flows on LCSP subgraph links are also highest for MNFLOP-50%, followed by MNFLOP-25% and SP. This metric, along with the number of links included in the LCSP subgraph, shows that MNFLOP effectively concentrates the person-trips onto fewer links.

Appendix B: Additional MNFLOP Results for Sioux Falls Scenario 1 includes further results for Sioux Falls Scenario Set 1, including Table B-1 that displays overlap/shareability metrics and path choice for every OD pair for all three assignment methods.

Finally, although the MNFLOP formulation in Section 2.4.2 permits demand from one OD to use multiple paths, the path flow results in this scenario (as well as all other subsequent scenarios) indicate that at optimality, the demand flow from an OD pair always uses one and only one path. This outcome is logical given the quadratic nature of the objective function and the lack of capacity constraints on links.

Table 3.2 Network Level Overlap Metrics for Scenario 1

Metric	**MNFLOP-50% (Δ_{max}^{rel} = 50%)**	**MNFLOP-25% (Δ_{max}^{rel} = 25%)**	**SP**
Avg Overlap, Z (person-trips)	3613	2859	1740
Avg Overlap Perc., $Z^{\%}$ (%)	56.4	44.7	27.2
Avg Trip Distance (miles)	17.28	16.44	15.62
Avg Overlap Distance (miles)	9.6	7	4
Avg Detour (miles)	1.66	0.82	-
Avg Detour (ratio)	1.11	1.05	-
Marginal Overlap (trips/detour-mile)	1128.3	1365	-
Marginal Overlap Perc. (% /detour-mile)	17.6	21.3	-
Detour Distance Elasticity of Overlap	10.1	12.3	-
Cost of LCSP (miles)	62	71	117
Avg Flow on LCSP Links (person-trips)	1784	1482	855
# Links with Non-zero Flows	18	21	31

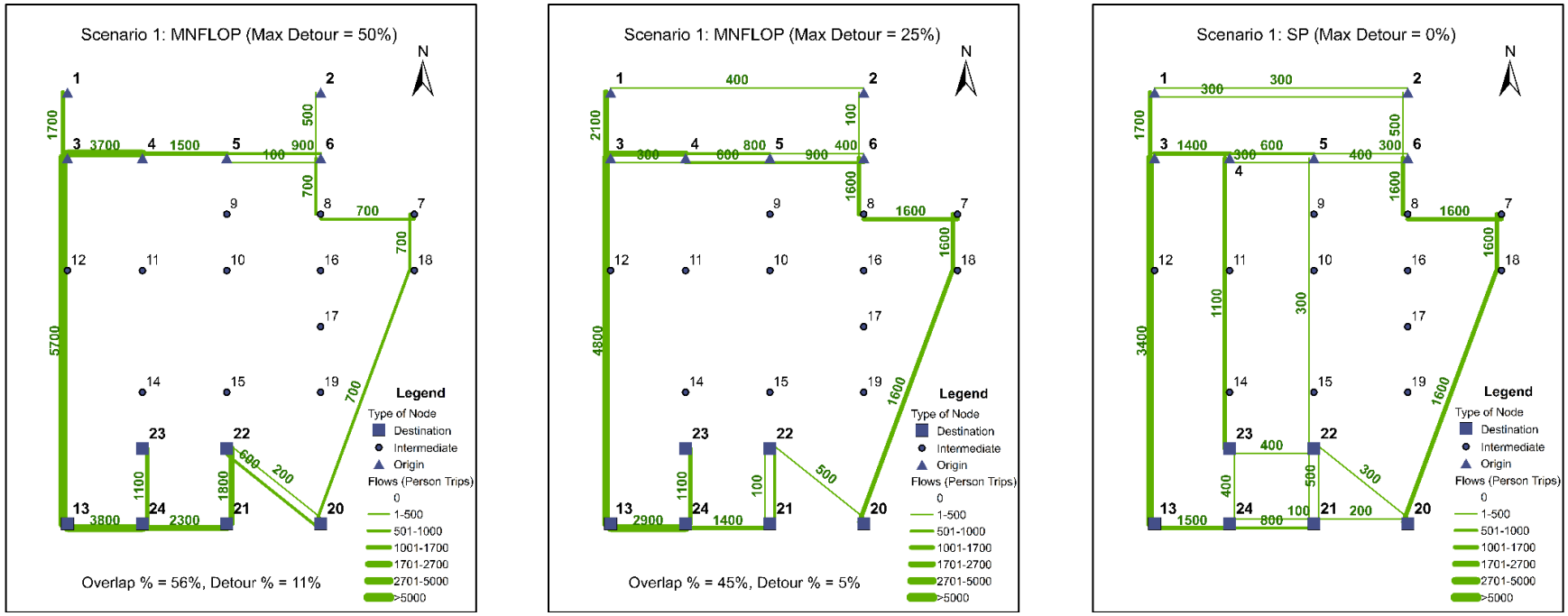

Figure 3.2 Link Flows (Person trips) for Scenario-1: MNFLOP - 50% Detour (left), MNFLOP- 25% Detour (middle), and SP (right)

3.3.1.1 Sensitivity Analysis for Maximum Detour Distance $\boldsymbol{\Delta_{max}}$

The results of MNFLOP are highly sensitive to Δ_{max}. This section examines how shareability metrics for Scenario 1 change with Δ_{max}, expressed in this instance in terms of relative percentage change in path length compared to the shortest path length (Δ_{max}^{rel}). MNFLOP is run for a range of Δ_{max}^{rel} values from 0% (SP) to 75% by restricting the path set K_{od} for each OD pair to those that are within Δ_{max}^{rel}.

Figure 3.3 displays average overlap percentage for the network and demand for Scenario 1 as a function of Δ_{max}^{rel}. As previously observed, increasing Δ_{max}^{rel} increases the number of overlapping flows for an average unit flow in the network. Average overlapping flows (and hence the percentage of overlapping flows) increases monotonically with increasing Δ_{max}^{rel} (Figure 3.3).

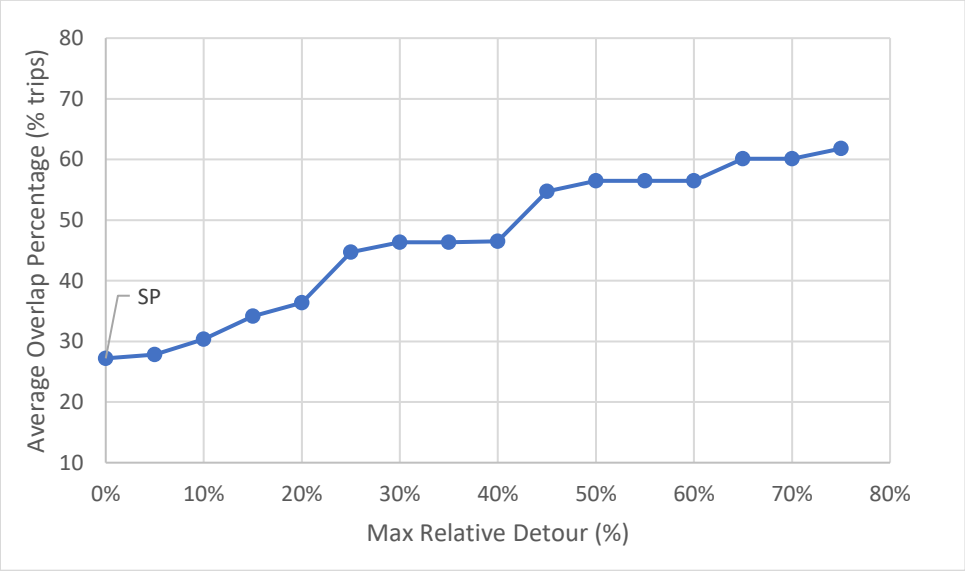

Figure 3.3: Maximum Relative Detour vs. Average Overlap Percentage for Scenario 1

Figure 3.4 displays average overlaps (person trips) and average detour (miles) for the above-mentioned range of Δ_{max}^{rel} values. Interestingly, Figure 3.4 shows there is a non-monotonic relationship between average detours and average overlaps because there is also

a non-monotonic relationship between average detours and Δ_{max}^{rel}. For example, when Δ_{max}^{rel} increases from 40% to 45%, overlapping flows increase by nearly 500 person trips. However, the average detour of the resultant MNFLOP path assignment decreases slightly from 1.5 miles to 1.4 miles. This result is possible because MNFLOP is attempting to maximize flow overlaps, and when more paths are feasible (i.e., there are more paths in K_p) it is possible that while the set of overlap-maximizing paths take advantage of the increase in Δ_{max}^{rel} from 1.4 to 1.5, the overlap-maximizing paths for many other OD pairs actually decrease in detour distance.

Figure 3.4 also shows that the change in average detour for the MNFLOP path assignment increases at a lower rate than Δ_{max}^{rel}. For instance, the average detour for Δ_{max}^{rel} = 75% is about 2 miles which is only a 12% increase in average trip distance compared to SP. This implies that even when Δ_{max}^{rel} is set to a high value, overlap-maximizing flows between most OD pairs still only detour slightly from their SP. Therefore, significant increases in overlap are possible even when flows from most OD pairs only need to make a slight detour from their shortest paths.

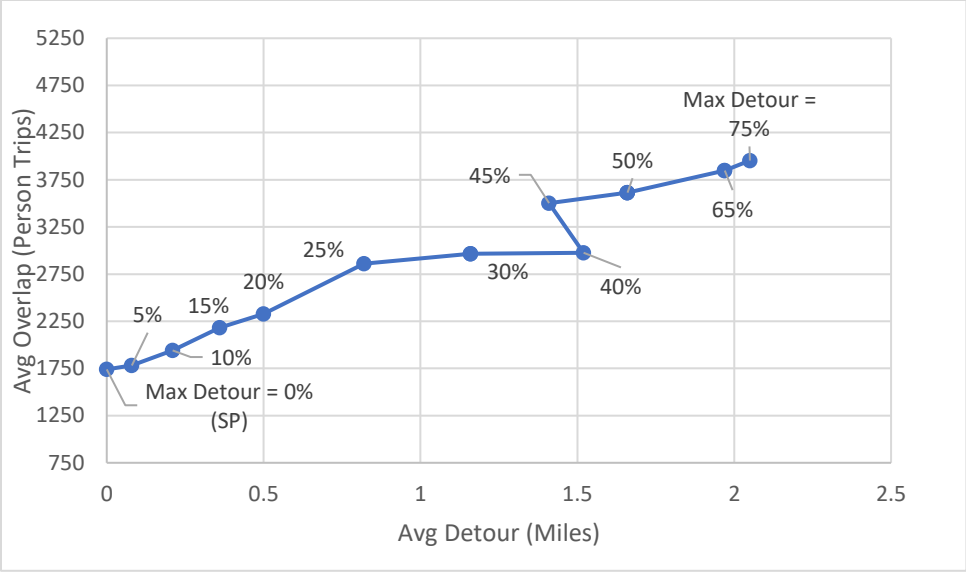

Figure 3.4: Trade-offs between Average Detour and Average Overlap for Scenario-1 (Each point is marked with the maximum relative detour parameter value for which it was produced)

3.3.2 Scenario Set 2: Origin-level Shareability Analysis

Scenario Set 2 analyzes the extent of overlaps between trips originating from each node of the Sioux Falls network by solving MNFLOP as well as the SP assignment problem, for each origin separately. The maximum detour parameter used in this scenario (and all subsequent scenarios) is the minimum of Δ_{max}^{abs} = 8 miles and Δ_{max}^{rel} = 50%. Moreover, for each scenario (i.e., each origin), we only consider the trips originating from said origin. This analysis gives insights into the relative dispersion of trips originating from an individual location.

Figure 3.5 displays the magnitude of demand along with the overlap percentage for all 24 origin demand nodes in the Sioux Falls network using MNFLOP and SP. The plots show that assigning trips using MNFLOP increases the overlap percentage for each origin node when compared to SP assignment. Overlap percentage for origin nodes such as Node-1 nearly doubles when trips are assigned based on MNFLOP compared to SP. On the other

hand, there is only a minor increase in overlap percentage for trips originating from Node-9, Node-10, Node-14, and Node-19.

Figure 3.5 also shows that trips originating from Node-2 exhibit the highest overlap percentage (i.e., the lowest dispersion) under both MNFLOP and SP assignment. Node-2 has a flow overlap percentage of over 60% for MNFLOP. This means that when flows are assigned based on MNFLOP, the average trip starting from Node-2 shares its paths with 60% of all other person-trips originating from Node-2. On the other hand, trips originating from Node-10 have the lowest overlap percentage—22% for MNFLOP—despite having a high magnitude of originating trips.

Figure 3.5 further reinforces that both the magnitude (indicated by total demand per unit time or total demand per unit area per unit time) as well as the dispersion of demand (indicated by percentage of overlap) are necessary to fully characterize of the shareability of trips from an origin location. Node-2 has a low magnitude of originating demand (4,000 person-trips), but a high overlap percentage value of 62%, meaning that trips from this location are still highly shareable. This might suggest that operating a shared mobility feeder mode from this location might be viable, despite its low magnitude of demand. While Node-16 has a lower overlap percentage value of 23%, since its magnitude of demand is high (26,000 person-trips), the average trip from Node-16 still shares paths with nearly 7,000 person-trips (26% of 26,000 person-trips). This means that even though trips from Node-16 are more dispersed, a unit flow from Node-16 is still likely to share a path in part or whole with nearly three times as many person-trips as compared to a unit person-trip starting from Node-2.

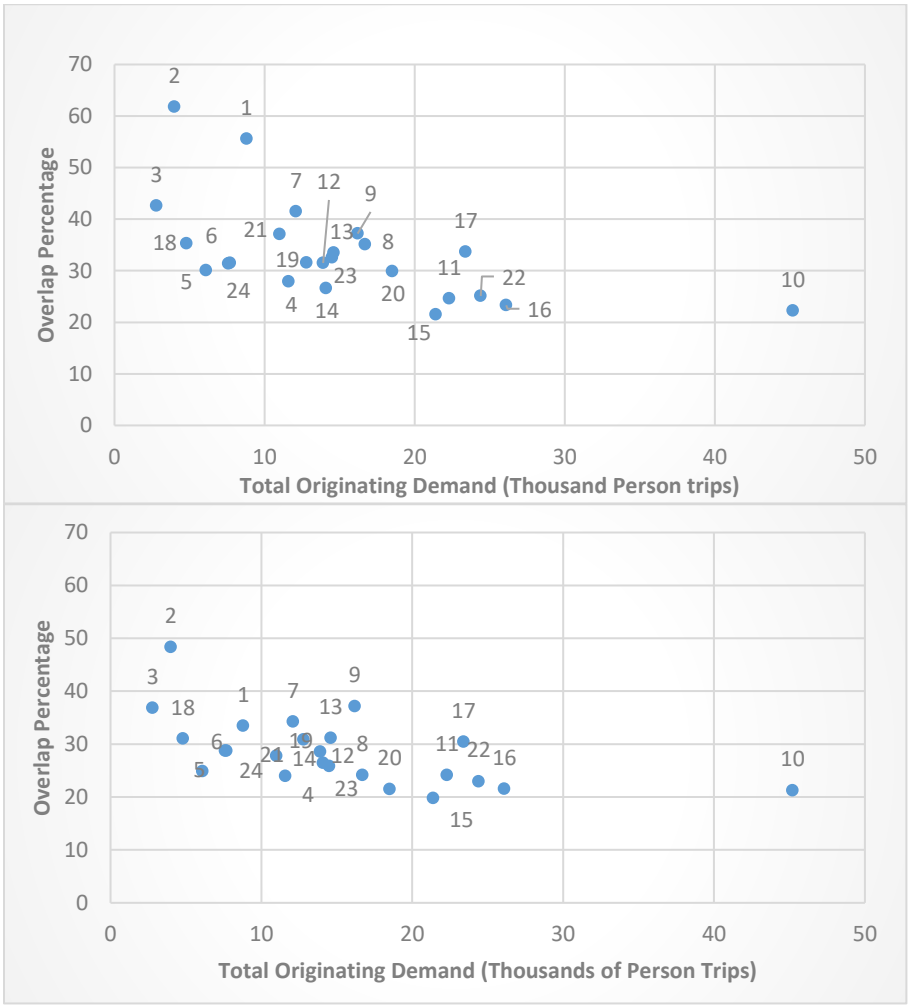

Figure 3.5 Demand Magnitude and Overlap Percentage for Origin Nodes in Sioux Falls with MNFLOP (top) and SP (bottom) Assignment

Figure 3.6 displays the link flows using MNFLOP for trips originating from Node-2 and Node-5. These two origin nodes have similar magnitudes of total originating demand, whereas the overlap percentage for Node-2 is twice as much as Node-5. The higher value of overlap percentage for Node-2 is evident in Figure 3.6, where flows have a tree structure with a root at Node-2 and flows to destination nodes only use a few links/branches, indicating high overlap between paths. On the other hand, Figure 3.6 displays a tree structure with flows emanating from Node-5 branching out across many links on the way to destination nodes.

The MNFLOP link flows for demand originating from Node-2 shown in Figure 3.6 also help to demonstrate partial shareability of trips captured in this dissertation. For example, a unit flow between OD pair (2,17) shares paths with person flows between OD pair (2,5) for $\frac{9}{14}$ of its path length; person flows between OD pairs (2,7) for $\frac{7}{14}$ of its path length; and with person flows between OD pair (2,10) for $\frac{12}{14}$ of its path length. The results in this case show that all flows between Node-2 and Nodes 5, 7, 10 and 17 are assigned on their respective shortest paths under MNFLOP assignment. Even though the shortest path from Node-2 to Node-9 is 2-6-8-9 (10 miles), MNFLOP assigns all flows between these nodes to the path 2-6-8-16-10-9 (12 miles) to maximize network wide overlap of flows. Thus, a unit flow between OD pair (2,17) shares paths with flows between OD pair (2,9) for $\frac{12}{14}$ of its path length. At the same time, it shares its entire path with person flows between OD pairs (2,19), (2,20) etc. The average overlap for a unit trip between (2,17) is close to 2613 person trips. This value captures the average amount of overlapping flows from all other OD pairs in the network that overlap in part or in whole with a unit trip between (2,17).

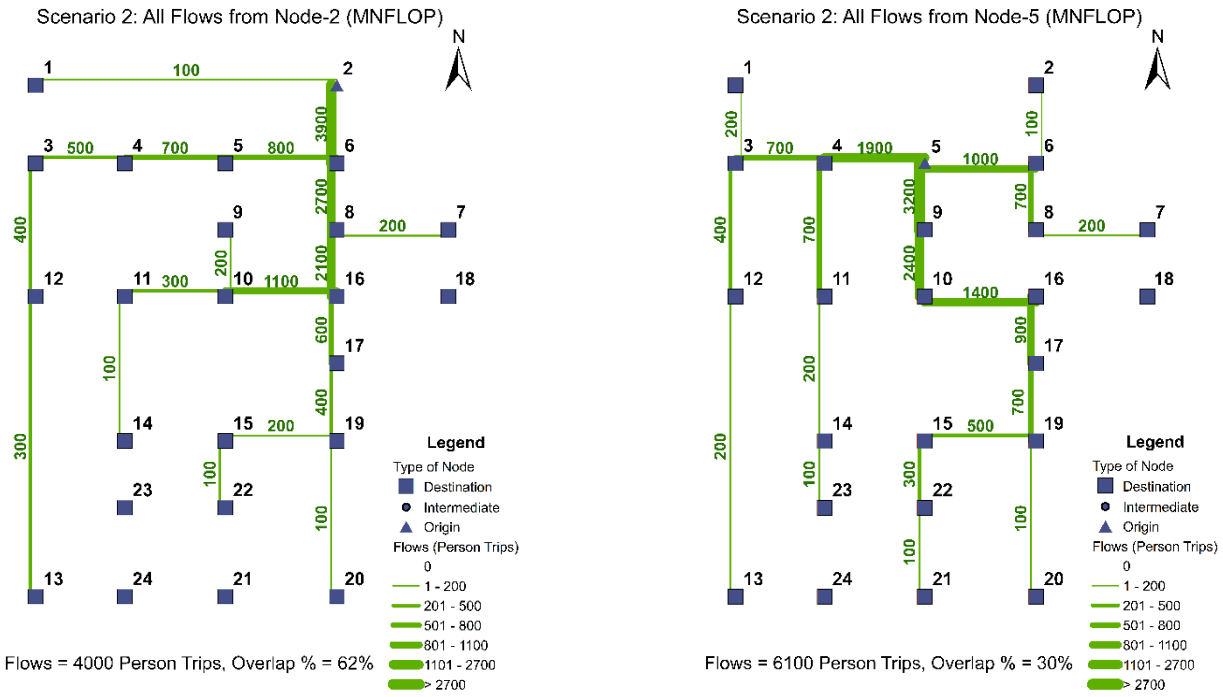

Figure 3.6: MNFLOP Assignment from Node 2 (left) and Node 5 (right)

3.3.3 Scenario Set 3: Same Origin, Different Destinations Shareability Analysis

This scenario set aims to highlight the difference in shareability for different demand patterns for the same (or similar) magnitude of total demand originating from the same set of origin locations in the Sioux Falls network. The Northern nodes in the Sioux Falls network (nodes 1 through 6) are the origin nodes for all cases in this scenario set. I assign a total demand of 1000 person-trips from these origin to three different sets of destination nodes to show how shareability depends on the spatial distribution of demand.

I obtain the demand between each OD pair in each case by first slicing the original OD table to include only the selected OD pairs relevant to each case. I multiply the remaining OD table cells by a single factor such that the total demand of the OD table adds up to 1000 trips. Then I round the OD flows to the nearest whole number. Hence, the total demand may vary

slightly from 1000 trips across the scenarios. This small variation does not affect the shareability metrics or inferences drawn.

Table 3.3 provides the shareability metric values for the three cases. All assignments are MNFLOP using Δ_{max}values as indicated in Table 3.1. The first case (N to CBD) includes trips from origin nodes in the North to the CBD (Nodes 10 and 16), as shown in Figure 3.7Figure 3.7a. The second case shows trips from origin nodes in the North to destination nodes in the South, as shown in Figure 3.7b. In the final case, we randomly assign trips from origin nodes in the North to destination nodes scattered throughout the network, as show in Figure 3.7c.

The results in Table 3.3 indicate that the magnitude of overlap, as well as overlap percentage, are highest when trips from the six northern origin nodes are sent to a few destination nodes that are close to each other (i.e., the N to CBD case). The number of overlapping trips and overlap percentage are the lowest when trips are bound to destination nodes scattered throughout the network (N to Scatter case). The higher marginal values of overlap for a unit detour for the first two cases compared to the third case (N to Scatter case) further indicate that when trip patterns are concentrated onto fewer destinations that are close to each other, there is significant potential for these trips to overlap on network links. The ratio of average overlap distance to average total trip distance is highest for the N to CBD case followed by N to S and the N to Scatter cases. This indicates that a unit trip in the N to CBD case overlaps with the largest fraction of other flows in the network along each link in its path. This is again due to trip destinations being clustered, enabling more overlaps. The highest overlap and overlap percentage values for the N to CBD case are also reflected in its

high value of average link flows and the fewest links used. Even though the number of links with non-zero flows is similar in both the N to S and N to Scatter cases, the average link flows are higher for the former, since the destinations are more clustered compared to the latter case. Even though the N to CBD case has the highest value of overlap and overlap percentage, as well as their marginal values, it does not have the highest value for detour elasticity of overlap. This is because the base SP assignment overlap is already quite high for the N to CBD case. MNFLOP results in the most improvement in overlapping flows in the N to S case compared to SP assignment.

Table 3.3 Network Level Shareability Metrics for Scenario Set-3 (MNFLOP)

Network level Shareability Metric	N to CBD	N to S	N to Scatter
Total Demand	*1001*	*1008*	*999*
# Origins	*6*	*6*	*6*
# Destinations	*2*	*6*	*6*
Avg Overlap Z (person-trips)	566.3	548.8	289.9
Avg Trip Distance (miles)	12.98	17.13	13.95
Avg Overlap Distance (miles)	7.3	9.2	3.8
Avg Overlap Percentage $Z^{\%}$ (%)	56.63	54.78	29.04
Avg Detour (miles)	0.58	1.5	1.66
Avg Detour (ratio)	1.05	1.1	1.14
Marginal Overlap (trips/detour-mile)	272.6	186.3	42.1
Marginal Overlap Perc. (% /detour-mile)	27.3	18.6	4.2
Detour Distance Elasticity of Overlap	8.3	10.8	2.4
Avg Link Flow (person-trips)	342	277	205
Cost of LCSP (miles)	38	62	68
# Links with Non-zero Flows	10	18	19

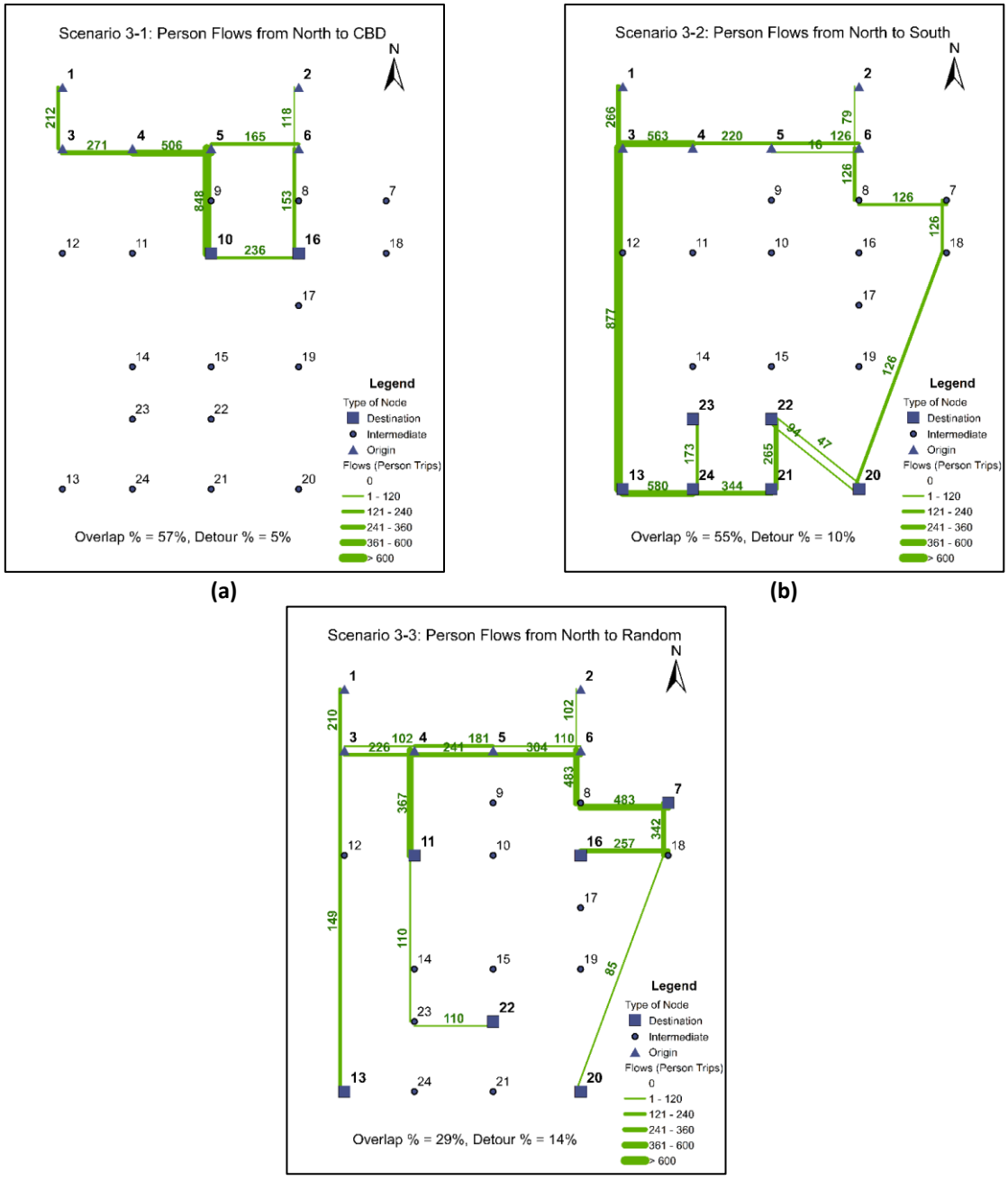

Figure 3.7: Link Flows using MNFLOP under (a) Scen. 3-1, (b) Scen. 3-2, and (c) Scen. 3-3

3.3.4 Scenario Set 4: Different Origins, Different Destinations Shareability Analysis

This scenario set aims to highlight the difference in shareability metrics when the same (or similar) magnitude of total demand in the network has different underlying trip

patterns. Unlike the prior scenario set, in Scenario Set 4 both the origins and destinations vary across the four cases. This section uses the same process as Scenario Set 3 for generating an OD table for each scenario. Once again, the OD table for each case has around 1000 person trips.

The first case (Case 4-1) involves trips assigned from non-CBD nodes toward CBD nodes (Nodes 10 and 16). The second and third cases (Case 4-2 and Case 4-3) involve trips assigned between OD pairs scattered throughout the Sioux Falls network. The fourth case (Case 4-4) has trips assigned between a set of clustered origin nodes and a set of clustered destination nodes.

Table 3.4 outlines the shareability metrics for all four cases in this scenario set. The person-trips on links in each of the four cases are shown in Figure 3.8. Results show that, like what was observed in the previous subsection, the magnitude and percentage of overlapping trips are comparatively higher in cases where trips either originate from and/or terminate at nodes that are fewer or closer to each other. Case 4-4 (Figure 3.8d) has the highest magnitude (676 trips) as well as percentage (68%) of overlapping trips because both the origin nodes and the destination nodes are clustered. Clustered origins and destinations allow MNFLOP to increase overlapping flows by assigning trips to paths with a small detour (average detour of 17% of shortest path trip distance). This is further corroborated by the high distance elasticity of overlap observed in this case (8.3). The high ratio of average overlap distance to average trip detour distance in this case also reflects the fact that a unit trip shares paths with a high fraction of all flows in the network on each link in its path. This

case also has the highest average link flows as well as the fewest number of LCSP subgraph links, indicating that MNFLOP can assign more flows onto fewer links and fewer corridors.

Table 3.4 Network Level Shareability Metrics for Scenario Set-4 (MNFLOP)

Network level Shareability Metric	**4-1 Outskirts to CBD**	**4-2 Scattered Case A**	**4-3 Scattered Case B**	**4-4 Clustered O's to Clustered Ds**
Total Demand	*998*	*1000*	*995*	*998*
# Origins	*22*	*5*	*10*	*5*
# Destinations	*2*	*8*	*5*	*5*
Avg Overlap Z (trips)	164	114.6	137.7	676.2
Avg Overlap Perc. $Z^{\%}$ (%)	16.45	11.47	13.85	67.8
Avg Trip Distance (miles)	8.97	7.16	9.13	16.54
Avg Overlap Distance (miles)	1.4	0.8	1.2	10.7
Avg Detour (miles)	0.46	0.45	0.99	2.42
Avg Detour (ratio)	1.05	1.07	1.12	1.17
Marginal Overlap (trips/detour-mile)	44.1	50	36.9	163.8
Marginal Overlap Perc. (% /detour-mile)	4.4	5	3.7	16.4
Detour Distance Elasticity of Overlap	2.6	3.6	3	8.3
Avg Link Flow (person-trips)	105.4	79.6	80.4	412.7
Cost of LCSP (miles)	85	90	113	40
# Links with non-zero flows	24	23	31	11

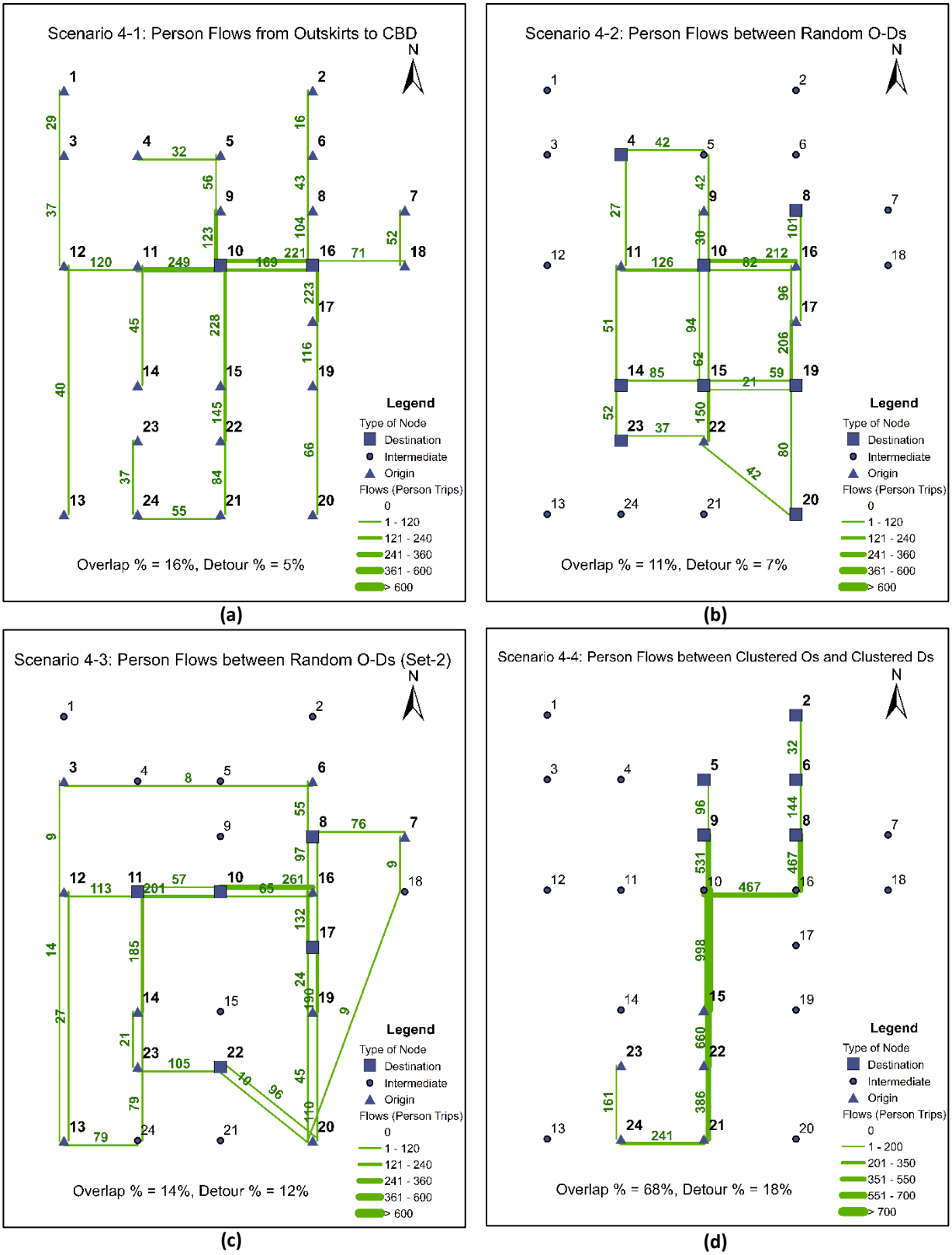

Figure 3.8: MNFLOP Link Flows (Person trips) for Scenarios (a) 4-1, (b) 4-2, (c) 4-3, (d) 4-4

Case 4-1 (Figure 3.8a) has the 2nd highest magnitude and percentage of overlaps because even though the origin nodes are scattered all over the region, all trips are bound to the CBD nodes (Node-10 and Node-16) that are close to each other. Case 4-2 (Figure 3.8b) and Case 4-3 (Figure 3.8c) have low shareability compared to the other two cases since their origin and destination nodes are scattered throughout the network. It is also interesting to note that even though Case 4-1 has a higher value for overlapping trips as well as overlap percentage compared to Case 4-2 and Case 4-3, it has lower values for marginal overlap, marginal overlap percentage, and distance elasticity of overlap. This is because MNFLOP is not able to significantly increase overlapping flows in Case 4-1 compared to when all flows are assigned on shortest paths in this case. A 1% increase in trip distance yielded only a 2.5% increase in overlapping flows for Case 4-1—the lowest value among all cases. Case 4-1 also has a high number of used links, due to the spatial extent of origin nodes being scattered all over the network.

3.4 Validation of Origin Level Shareability Metrics

The prior subsections verify that (i) the MNFLOP results are consistent with the flow overlap conceptualization and MNFLOP formulation and (ii) flow overlap and MNFLOP effectively operationalize my definition of person-trip shareability. I believe these are major contributions in (transportation) network science. However, this section aims to go one further step and validate the usefulness of MNFLOP in terms of (i) quantifying a subregion's shareability/sharing potential and (ii) connecting TSAs to the efficiency of a SMM, as displayed in Figure 1.2 Specifically, I test the hypothesis that as my person-trip shareability metrics increase, the SMM's efficiency metrics will increase. The null hypothesis, which

would invalidate my shareability metrics if we are unable to reject it, is that there is no association between our shareability metrics and SMM efficiency. To test the null hypothesis, I model a microtransit SMM operating as a transit feeder service from a commuter rail station to travelers' final destinations.

To validate the usefulness of MNFLOP and test this hypothesis, I designed a set of computational experiments. I assume that the commuter rail station is at Node-10 in the Sioux Falls network (Figure 3.1), and the microtransit vehicles serve batches of requests with origins at Node-10 and destinations at the other 23 nodes, before returning empty to Node-10. I create 10,000 problem instances, where the problem instances vary in terms of the distribution of destination nodes for the set of requests. In each instance, there are 10 requests, a fleet of 5 vehicles starting at Node-10, with a capacity of 5 seats each. I set the traveler drop-off time window constraints using the same absolute and relative maximum detour distance values used for generating shortest paths while solving MNFLOP (as specified in Section 3.2.1). Given one instance of ten traveler requests, I solve MNFLOP and a capacitated multi-vehicle routing problem with time-windows (CMVRPTW). I formulate CMVRPTW (representing the microtransit service) based on Kallehauge et al. (2005) and solve CMVRPTW to optimality using Gurobi. It is the CMVRPTW, not MNFLOP, that limits the problem size. I measure microtransit service efficiency in terms of fleet vehicle miles traveled (VMT) and average vehicle occupancy (AVO); hence, I hypothesize a negative association between flow overlap measures and VMT and a positive association between flow overlap measures and AVO. I estimate the following two empirical models using ordinary least squares regression:

$$ln(\boldsymbol{VMT}) = \beta_0 + \beta_1 \boldsymbol{Z_o^{\%}} + \beta_2 ln(\boldsymbol{PMT}) + \boldsymbol{\varepsilon}$$

$$ln(\boldsymbol{AVO}) = \beta_0 + \beta_1 ln(\boldsymbol{Z_o}) + \boldsymbol{\varepsilon}$$

where $\boldsymbol{VMT}$ and $\boldsymbol{AVO}$ are the vectors of VMT and AVO, respectively, for all problem instances. In the first model, I control for the cumulative direct origin (i.e., the rail station at Node-10) to destination distance for the 10 travelers ($\boldsymbol{PMT}$). Table 2.1 defines origin-level overlap percentage ($\boldsymbol{Z_o^{\%}}$) and origin-level overlap ($\boldsymbol{Z_o}$). $\boldsymbol{\varepsilon}$ is the vector of error terms for all problem instance that I assume is independently and identically normally distributed with mean zero. I log transform the variables, except $\boldsymbol{Z_o^{\%}}$, to directly obtain the elasticity between the shareability measures and mobility service efficiency measures. I do not log-transform $\boldsymbol{Z_o^{\%}}$ because its units are already in percentage form.

Table 3.5 and Table 3.6 show the estimation results for the two models defined above. As a note, I removed 50 scenarios (0.5% of all scenarios) as the CMVRPTW was infeasible for these instances.

The results in Table 3.5 allow us to reject the null hypothesis that there is no association between our shareability metrics and shared mobility efficiency. We can reject the null hypothesis from a statistical perspective (see the high t-value and low p-value) and practical perspective (see the coefficient value). In fact, the coefficient for $\boldsymbol{Z_o^{\%}}$ implies that VMT decreases by 0.71% for every unit (i.e., 1%) increase in $Z_o^{\%}$. A 0.71% elasticity value is quite high and confirms a strong connection between person-trip shareability and the efficiency of a SMM.

The results in Table 3.6 also allow us to reject the null hypothesis, from a statistical and practical perspective. The coefficient for Z_o indicates that AVO increases by approximately 0.27% for every 1% increase in Z_o. This lends support to the notion that when the OD travel patterns of person trips in the region exhibit greater overlaps (as measured using the flow overlap value for a unit trip in the region), a shared ride vehicle can effectively combine more person trips into the same vehicle, resulting in a higher AVO.

Table 3.5 Effect of Origin Overlap % ($Z_o^{\%}$) on VMT

Dependent Variable: **ln(VMT)**			*R Square*	0.3876
Variable	*Coefficients*	*Standard Error*	*t Stat*	*P-value*
Intercept	1.5752	0.0429	36.7431	0.0000
ln(PMT)	0.6870	0.0095	72.13787	0.0000
$Z_o^{\%}$	-0.0071	0.0001	-47.303	0.0000

Table 3.6 Effect of Origin Overlap (Z_o) on Avg Vehicle Occupancy

Dependent Variable: **ln(AVO)**			*R Square*	0.236
Variable	*Coefficients*	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.0964	0.004	-22.367	0.0000
ln(Z_o)	0.266	0.005	55.408	0.0000

3.5 Discussion

The results and analysis presented in the prior subsections show that (i) the MNFLOP model and a commercial solver can significantly increase overlaps in a transportation network compared to shortest path assignment; (ii) the MNFLOP and its associated shareability metrics capture the relative dispersion/concentration of trips emanating from/to and origin/destination node; (iii) the MNFLOP formulation and the optimal path flows can capture partial shareability of person trips in a region (iv) the MNFLOP and its

associated shareability metrics can effectively differentiate between the shareability potential of the same network under different OD trip tables, wherein clustered origins and/or clustered destinations increase flow overlaps and shareability potential relative to the same total demand scattered across the network; and (v) origin level measures of shareability such as average origin overlap (Z_o) and average origin overlap percentage ($Z_o^{\%}$) show a statistically and practically significant association with efficiency metrics for a microtransit last-mile transit feeder service.

The ability of MNFLOP and its associated metrics to achieve each of these five outcomes indicates that researchers can use it to meaningfully differentiate between subregions within cities in terms of the subregion's sharing potential at the person-trip level. Unfortunately, the computational complexity of the problem formulation limits city-scale analysis; hence, developing scalable formulations consistent with our person-trip shareability definition is an important future research direction. The next section, in addition to summarizing the study and discussing limitations, provides an extensive discussion of the potential uses of MNFLOP and its shareability metrics proposed in the first study of the dissertation.

Finally, I note that MNFLOP and its associated shareability metrics fully operationalize person-trip shareability according to the principles in Section 2.2.2.

3.5.1 Limitations

One shortcoming of the current study is the reliance on a commercial solver for the quadratic programming model. Future research should work to develop alternative exact, approximate, and heuristic solution algorithms to solve the problem for larger networks, as

well as consideration of alternative formulations and approximations of the original model. Approaches include reducing the problem size by clustering trip origin and destination locations or restricting the other candidate OD pairs that constitute overlapping flows when calculating overlap for a specific OD pair. In terms of heuristic solution approaches, genetic algorithms are potentially appealing as they can explore the solution space effectively and efficiently. Tailored heuristics that efficiently estimate, rather than fully evaluate, the objective function in sub-iterations may also be effective.

A second shortcoming of the shareability study is the path-based formulation of the problem. Even with better exact, approximate, or heuristic solution algorithms, the challenge of enumerating paths for OD pairs remains—path enumeration is exponential in network size. For large networks, even trying to identify a small number of k reasonably distinct paths is challenging. Hence, an important future research direction involves operationalizing person-trip shareability using alternative metrics and mathematical models. A link-based formulation is conceivable, but this is a significant challenge. As of now, the metrics and mathematical model in the current study are the only meaningful operationalization of person-trip shareability in the literature.

A third shortcoming is the lack of consideration of link capacities in the model. Despite wanting to maximize flow overlaps in the network by condensing flows onto fewer links, real-world street segments do have maximum capacities. To incorporate and estimate link capacity for shared mobility, one might consider the maximum person throughput of a heavy rail line as the upper limit on link flows in MNFLOP. Another option is to determine

the maximum person throughput of a bus rapid transit line, in the case where heavy rail lines are not feasible.

3.6 Conclusion

Chapter 2 and Chapter 3 of this dissertation contain a novel approach to quantify the shareability of person-trips in a region, given OD travel demand, a road network and maximum willingness to detour. The study conceptualizes and defines mathematically the notion of 'flow overlaps' for a unit person-trip in a network. The study employs the flow overlap concept to formulate the MNFLOP, a path- and flow-based Quadratic Program, to find the shareability for a given network and OD demand. A solution to the MNFLOP includes the optimal path flows (paths) between each OD pair that maximize shareability in the network. The study then uses the MNFLOP output to calculate several shareability metrics at various levels of aggregation: OD, origin, link, and network level. The study uses the MNFLOP model to analyze shareability in the Sioux Falls network, under multiple demand scenarios. The shareability metrics proposed in this dissertation add another dimension to measuring demand apart from its magnitude, specifically the dispersion (or overlap percentage) of trips from a node/location capture the directional component of demand.

The computational results verify the MNFLOP and associated shareability metrics can meaningfully distinguish between the shareability of a given street network across different spatial distributions of demand. The results also indicate that slight increases in detour for some OD pairs, compared to their shortest paths, can significantly increase flow overlaps in a network.

Finally, the study validates MNFLOP as a useful measure of person-trip shareability. I do so by performing a large number of computational experiments, wherein I vary the spatial demand patterns in a given network across the experiments, calculate MNFLOP and vehicle fleet miles for a microtransit last-mile transit feeder service, and test the null hypothesis that there is no association between MNFLOP-based shareability metrics and vehicle fleet miles. We reject the null hypothesis and find an elasticity of -0.71 between vehicle fleet miles and my shareability metric. The chapter also discussed the limitations of the MNFLOP based shareability metrics and suggested several directions of future research involving both the applications of MNFLOP based shareability metrics as well as enhancements to the methodology to overcome some of the discussed limitations.

The next pair of chapters (Chapter 4 and Chapter 5) presents the second study in my dissertation that deals with MOD services with Virtual Stops – called Corner-to-Corner services - that require users to walk the first/last mile of the trip to/from nearby Pickup and Dropoff (PUDO) locations. Chapter 4 introduces and defines the Corner-to-Corner Ride-pooling (C2C-RP) problem, offering a scalable decomposition-oriented method to address this problem. In Chapter 5, I present results from an extensive evaluation of user, operator, and computational expenses is conducted for both the C2C and D2D iterations of Ride-pooling (RP) and Ride-hailing (RH) services.

Chapter 4. Operating MOD Services with Virtual Stops – Problem Description and Scalable Solution Approach

4.1 Introduction

4.1.1 Background and Motivation

Mobility-on-Demand (MOD) services, enabled by smartphones and their applications, and offered by Transportation Network Companies (TNCs) such as Uber, Lyft, and Didi, emerged over a decade ago. The dominant MOD service option offered by TNCs is Ride-hailing (also known as ride-sourcing or e-hailing without shared rides), where vehicles can only carry one traveler request at a time. Several MOD service variants have emerged in recent years that aim to increase sharing and vehicle occupancies relative to Ride-hailing in order to decrease negative externalities from these systems.

The second most common variant is the shared-ride or Ride-pooling MOD service, such as Uber Pool and Lyft Line (now Lyft Shared) offered by Uber and Lyft. Ride-pooling MOD services involve pooling together sets of traveler requests that have similar, but not necessarily the same, origin and destination locations and request times into one vehicle. The envisioned benefits of Ride-pooling services include decreases in required fleet sizes (i.e., decreases in required drivers active on Uber and Lyft platforms at a given instant) and operational miles, potentially lowering the prices service providers can offer customers. Service providers also often market Ride-pooling as a service option that can reduce congestion, energy consumption, greenhouses gases, and local pollutants.

In conventional Ride-hailing and Ride-pooling MOD services, service providers offer and operate door-to-door (D2D) services, meaning that vehicles pick up and drop off travelers at their requested origin and destination locations, respectively. However, the need to pick up and drop off every traveler at their preferred origin and destination locations may significantly hamper MOD operational efficiency.

To address the operational inefficiencies associated with D2D MOD services, service providers now operate flexible Ride-pooling MOD services wherein travelers must walk to pickup (PU) locations from their trip origins, and from drop-off (DO) locations to their trip destinations. In this study of the dissertation, I refer to this MOD service as a Corner-to-Corner (C2C) Ride-pooling service. C2C Ride-pooling is quite similar to historical variants of dial-a-ride and flexible transit services, and several emerging on-demand microtransit services that require users to walk to/from their PU/DO locations instead of being served at their doorstep.

The premise behind C2C MOD services is that having travelers walk to nearby PUDO locations that are convenient for non-idle vehicles can increase MOD service fleet productivity measured in terms of travelers served per time unit, while also reducing fleet distance per request served and increasing vehicle occupancies. However, these operational improvements are likely to come at the cost of inconvenience for travelers, in terms of increased walk distances and increased request-to-destination travel times.

4.1.2 Research Goals, Problem Overview

Given this background on MOD services, the goals of the second study of my dissertation are two-fold. First, I aim to systematically analyze the trade-offs between

operator costs and user costs across four MOD services—D2D Ride-hailing (D2D-RH), C2C Ride-hailing (C2C-RH) D2D Ride-pooling (D2D-RP), and C2C Ride-pooling (C2C-RP), where every service responds to on-demand requests. Analyzing these four options requires decision policies and algorithmic frameworks for operating each MOD service, as well as a simulation environment. While the academic literature includes significant research on decision policies and solution algorithms for D2D-RH and D2D-RP services, the same is not true for C2C-RH and C2C-RP. Hence, the second goal of this study is to develop a scalable and effective decision policy and algorithmic approach for dynamically operating C2C MOD services, particularly C2C-RP. This chapter extensively lays out the formulation and a scalable solution methodology for the C2C-RP problem. The next chapter (Chapter 5) in the dissertation analyzes the cost trade-offs for MoD services with or without Ride-pooling and C2C options.

Developing a scalable and effective approach for C2C-RP services is a significant modeling and algorithmic challenge. The challenge stems from the size of the decision vector/space. A C2C-RP service provider needs to determine PUDO Virtual Stops for each traveler, while also matching travelers to vehicles, sequencing and scheduling PUDOs for each vehicle, and assigning vehicles to transportation network paths. Moreover, the fleet operator needs to make these decisions repeatedly in real-time, as new information (i.e., new requests and changes in network travel times) enters the system. To address the problem, I propose a decision policy that decomposes the full decision problem at each decision epoch into two subproblems that I solve sequentially and iteratively. The two subproblems are the PUDO locations selection subproblem and the vehicle-traveler matching subproblem, wherein the second subproblem embeds the vehicle sequencing and scheduling of traveler

PUDOs. Given the decomposed sub-problems, I test several decision policies for a C2C-RP service. Specifically, I compare the case where the fleet operator assigns travelers to PUDO locations first for each candidate vehicle, then matches travelers to vehicles vs. a policy where the operator matches travelers to vehicles first, then assigns PUDO locations for the traveler based on the matched vehicle.

4.1.3 Terminology and Abbreviations

Before going any farther, I want to clarify some key terminology and abbreviations that I will use throughout the rest of the dissertation. I use the abbreviation PUDO when referring to pickup *and* drop-off, whereas I will use the abbreviation PU/DO when referring to pickup *or* drop-off. Additionally, if I am only referring to a pickup, I will use the abbreviation PU, and if I am only referring to a drop-off, I will use the abbreviation DO. A request is associated with two PUDO locations (or links)—a pickup location (or link) and a drop-off location (or link).

Moreover, I want to differentiate between *locations* and *links* where PUDOs occur. I use the term PUDO *location* when discussing C2C-RP and C2C-RH in general, as in practice vehicles can pick up and drop off travelers anywhere (i.e., at nodes, or along links, or in designated PUDO spots). However, I use the term PUDO *link* to be precise in regard to the simulation model functionality in this study, as the simulation model allows PUDOs along network links.

4.1.4 Chapter Outline

The remainder of this chapter is structured as follows. Section 4.2 reviews the literature related to C2C MOD services and delineates the contributions of the current study.

Section 4.3 describes the C2C-RP operational problem. Section 4.4 presents the decision policy and solution algorithm for the C2C-RP operational problem. Section 4.5 concludes the chapter.

4.2 Literature Review

In this section, I present an overview of the existing literature related to C2C-RP. I focus my literature review on C2C-RP, as this is the focus of the second study of the dissertation. I refrain from providing a detailed review of D2D-RH and D2D-RP, as there is extensive literature in this area. However, for readers interested in D2D-RH and D2D-RP, there are several recent review articles worth mentioning.

Zardini et al., (2022) present a recent extensive review of automated MOD (AMOD) that covers operational-level and planning-level problems related to MOD services with automated vehicles. Many of the insights from this review article relate to MOD services without automated vehicles. In addition to characterizing different elements of AMOD modeling studies, the paper reviews solutions for matching, routing, and rebalancing for D2D-RH, and these three subproblems plus Ride-pooling for D2D-RP. Narayanan et al. (2020) provide a review of research on shared autonomous vehicles, which is another name for AMOD. Their review is quite broad, covering models of demand, parking, fleet sizing, vehicular traffic, matching, repositioning, pricing, and charging for electric vehicles. Mourad et al., (2019) survey shared mobility studies and focus on the operational problem. They consider travelers sharing rides with other passengers, and with parcels.

The remainder of this background and literature review section covers mobility services that include walking trips in general (Section 4.2.1), and mobility services with

walking legs wherein the fleet operator selects PUDO locations for travelers in real-time (Section 4.2.2). I conclude by delineating the contributions of my study (Section 4.2.3).

4.2.1 PUDO Locations Selection

One of the critical decisions in operating C2C MOD services is where to pick up and drop off travelers. PUDO locations selection impacts (i) the distance/time/cost required for a vehicle to serve each request; and (ii) each traveler's walking distance. Thus, the PUDO locations selection problem significantly impacts both operator and user costs.

Several studies develop strategies for selecting PUDO locations in C2C services. Wang et al. (2022) categorize these strategies into three approaches: 1) strict meeting points, 2) relaxed meeting points, and 3) no meeting points. As the simplest approach, strict meeting points determines a single set of PUDO points for each vehicle (Wang et al., 2022). All riders sharing the same vehicle walk from their respective origins to one common PU location and from one common DO location to their respective destinations (Aissat and Oulamara, 2014; Czioska et al., 2017; Stiglic et al., 2015). This approach entails few stops per vehicle trip and thus is convenient for the drivers. In fact, the strict meeting points approach is more common in conventional carpooling and ridesharing where the driver has their own trip origin and destination, rather than in MOD services with a dedicated driver. While being convenient for drivers, the strict meeting points approach limits the potential to match drivers and riders in a region.

To address the shortcomings associated with the strict meeting points approach, several studies enable multiple intermediate PUDO locations on a vehicle's route, which is the relaxed meeting points approach (Wang et al., 2022). The relaxed meeting points

approach clusters riders before assigning them a common location for PU/DO (Czioska et al., 2019; Martínez et al., 2015) or limits the candidate PUDO locations to a small subset of road nodes (Araldo et al., 2019; Gurumurthy and Kockelman, 2022). Studies using the relaxed meeting points approach, unlike the fixed meeting points approach, rarely assume that drivers have their own destinations (a special case is Miklas-Kalczynska and Kalczynski (2021)), rather, most studies focus on C2C-RP MOD services. A recent study by Lotze et al. (2022) is a special case of the relaxed meeting points approach, as they attempt to dynamically assign each new request to an existing planned vehicle stop. However, if the algorithm does not find a feasible existing PU and/or DO stop for a new request, then the algorithm has a vehicle pick up and/or drop off the traveler at the traveler's origin and/or destination.

Finally, the no meeting points approach does not restrict riders to share PU or DO points with other travelers (Wang et al., 2022). Some studies using the no meeting points approach (Balardino and Santos, 2016; Zheng et al., 2019) propose variants of the shortest covering path problem (SCPP) or generalized traveling salesman problem (GTSP) formulations for the carpooling or flex-route transit applications, where the selection of PUDO locations is a deterministic problem. Other studies dynamically select the PUDO locations for each stochastic request (Fielbaum et al., 2021; Li et al., 2020; Lyu et al., 2019). Given that the approach in my study involves no meeting points and dynamic selection of PUDO locations, I focus the remainder of my review on studies that dynamically select PUDO locations without explicit meeting points.

4.2.2 Dynamic PUDO Locations (Virtual Stops) Selection

Dynamic selection of PUDO locations without explicit meeting points theoretically permits better solutions in terms of service quality and operational than approaches with predetermined or limited meeting points (Wang et al., 2022). Table 4.1 compares C2C MOD studies with dynamic PUDO locations selection.

Most C2C MOD studies focus on Ride-pooling since the purpose of incorporating walking access and egress trip legs is often the same as incorporating pooled rides—to reduce fleet miles, operational costs, congestion, and emissions (Fielbaum et al., 2021; Li et al., 2020; Lyu et al., 2019). However, Martin et al. (2021) consider C2C in a Ride-hailing service with the objective of minimizing the riders' travel cost. With a fixed vehicle speed, Martin et al. (2021) test the performance of C2C-RH with a 400-meter walk range in the Manhattan road network. They also mainly compare static PUDO assignment with dynamic PUDO locations selection for C2C-RH and find that the impact on operational efficiency is relatively minor.

Other studies simulate C2C-RP services on various real-world city networks including Manhattan, Shanghai, and Chengdu (Fielbaum et al., 2021; Li et al., 2020; Lyu et al., 2019). These studies all assume deterministic link travel times or fixed vehicle speeds (i.e., the vehicles travel the same speed on all links). Notably, two studies that do incorporate stochastic travel times in large scale networks— (Gurumurthy and Kockelman, 2022; Zwick et al., 2021) in POLARIS and MATSim, respectively—use fixed or relaxed meeting points approaches. Hence, ours is the first study to incorporate stochastic link travel times (and congestible links) in a simulation, where the C2C service does not have fixed meeting points.

The number of MOD requests for the studies Table 4.1 range from 9,000 to 10.7 million. However, the hourly average request rates only range from 9,000 to 22,000 requests per hour. Fleet sizes in the literature range from 40-13,181. My study has the longest analysis period of 24 hours, and the number of requests and fleet size are consistent with prior research.

Vehicle capacity ranges from 1-20. As my study is the only study to consider C2C-RP and C2C-RH, it is also the only one to consider vehicle capacities of one and greater than one in the same study.

For maximum walking distance, parameter values range between 300 and 1000 meters. Similar to other studies, I vary the parameter between 250 and 1000 meters, to understand its impact on key performance metrics.

For walking speed, every study assumes 5 or 5.04 km per hour. In this study, I consider a walking speed of 5 km per hour, but also a 'walking' speed of 20 km per hour that is more akin to an electrified bike or scooter. I include this parameter to evaluate the benefits of electrified bikes and scooters on system performance metrics for C2C-RH and C2C-RP.

The third to last column in Table 4.1 refers to whether vehicles wait for travelers at PU locations. There are two interrelated aspects here, related to the simulation environment and the fleet's operational strategy. If the simulation does not include stochastic link travel times, then the service provider fully determines whether vehicles wait for travelers at PU locations (the second aspect of this service dimension). However, if the simulation does include stochasticity, then vehicle waiting is not fully in the operator's control. As ours is the

only study with stochastic travel times, it is also the only study where the arrival time of vehicles at PU locations is uncertain; hence, vehicles may arrive before requests and have to wait. Of course, this is consistent with a real-world MOD service, where travel time is uncertain. Interestingly, Lyu et al. (2019) also consider the case where the vehicles wait at PU locations. However, since their model is deterministic in link travel times, they intentionally have vehicles wait for travelers at PU locations. Their model explicitly captures the waiting cost for onboard passengers. The other studies do not specifically mention the possibility of vehicles waiting for travelers at PU locations (Fielbaum et al., 2021; Li et al., 2020).

Table 4.1 Summary of Existing Studies Analyzing C2C MOD Services (with dynamic PUDO locations selection)

Paper	MOD Services in Study	Road Network & Travel Time	Avg. Request Rate (req./hr.) and Simulation Period	Fleet Size (veh.) and Vehicle Capacity (seats)	Walk Range (m) and Walk Speed (km/h)	Vehicle Waiting for Request?	Procedure for Identifying PUDO Locations Candidates	Selecting PUDO Locations and Vehicle-Request (R-V) Matching
Fielbaum et al. (2021)	C2C RP D2D RP	Manhattan, NY with fixed link travel times	9,970 over 1 hour	2,000-3,000 and 6-9	417, 1000 and 5	No	Recursive search of neighboring PUDO locations within walk range that reduce vehicle travel cost	Joint optimization of PUDO locations selection and R-V matching
Li et al., (2020)	C2C RP	Shanghai, China with fixed vehicle speeds	10,000 over 2 hours	1,200 and 20	100, 200, 300, 400, 500 and 5.04	No	All PUDO locations within walk range	R-V matching first, PUDO locations selection second
Lyu et al. (2019)	C2C RP	Chengdu, China with fixed vehicle speeds	22,500 over 28 days and 17 hours per day	12,725-13,181 and 3	300 and N/A	Yes, planned early vehicle arrivals	All PUDO locations within walk range	PUDO locations selection for pooled requests first, R-V matching second
Martin et al. (2021)	C2C RH	Manhattan, NY with fixed vehicle speeds	40–160 total requests	40–400 and 1	400 and 5.04	No	All nodes from request's origin within walk range. DO location unchanged.	**Dynamic Case:** PU location selection for candidate R-V pairs first, R-V matching second. **Static Case:** PU location selection independent of vehicle
This study	C2C RH C2C RP D2D RP D2D RH	Bloomington, IL with dynamic stochastic congestion-dependent link travel times	9,200 over 24 hours	1,000–10,000 and 1 (RH), 4 (RP)	250, 500, 750, 1,000 and 5, 20	Yes, stemming from vehicle arrival time uncertainty	PUDO links selected based on walk range and vehicle travel direction	**Option 1:** R-V matching first, PUDO locations selection for matched R-V pairs second. **Option 2:** PUDO locations selection for candidate R-V pairs first, R-V matching second

Finally, I want to describe the different decision policies and algorithmic approaches for C2C-RP in the literature. Lyu et al. (2019) maintain a waiting queue for unassigned requests. When new requests arrive, their approach attempts to sequentially match each new request with unassigned requests already in the waiting queue to form companion candidate pairs. Then, they determine the 'maximum sharing satisfaction utility' set of companion candidate pairs where the PUDO locations of candidate pairs can be pooled together in one vehicle. Alternative PUDO locations are determined by finding walkable

nodes from each request's origin and destination in the companion pair, and then choosing the best set of PU and DO locations that minimize total tour distance. Next, the approach aims to insert additional requests into the previously identified best companion pairs. The seat capacity is four in their study, so only two more requests can be inserted into the original companion pair. Finally, their approach assigns the closest idle vehicle to the first stop in the stop sequence corresponding to the set of least cost requests. A main difference between my approach and that in Lyu et al. (2019) is that they do not allow the insertion of new requests into a vehicle's route plan, until the vehicle completes all its current PU and DO tasks. This almost certainly limits the operational effectiveness of the proposed decision policy and solution algorithm.

Li et al., (2020) select PUDO locations for each new request after matching each request to a vehicle. Li et al., (2020) sequentially assigns each new request to a vehicle, wherein the algorithm considers the cost of inserting the new request's origin and destination locations into the planned routes of each vehicle. After request-vehicle matching, the algorithm adjusts the traveler's PUDO locations to minimize the matched vehicle's detour. The approach evaluates all walkable PUDO locations near the request's origin and destination, respectively. Additionally, the approach includes a post-processing stage in which PU and/or DO locations of two requests are merged if they are close to each other. These requests are assigned to the same vehicle subject to time window constraints and vehicle not waiting for request constraint.

There are three key differences between my study's decision policy and solution algorithm compared to Li et al., (2020). First, I perform request-vehicle matching using bi-

partite matching after batching requests together over a time interval. Second, my methodology allows for the flexibility to adjust PUDO locations either before or after request-vehicle matching. Third, I shortlist PU/DO candidates for a request-vehicle pair based on the vehicle's planned travel direction, in addition to walk distance restrictions. Experiments suggest that this added condition improves operational efficiency and computational run time.

Fielbaum et al. (2021) batch new requests received between decision epochs and solve an integer-linear program (ILP) assignment problem that may assign multiple requests to the same vehicle in a single decision epoch. They extend the request-vehicle ILP matching algorithm in Alonso-Mora et al. (2017) to incorporate the selection of PUDO locations for each request-vehicle match. The extended model and algorithm shortlists PUDO candidates for a request-vehicle match by evaluating neighboring network nodes around the request's origin and destination—if a neighboring node can reduce the matching cost for the request-vehicle pair it is put on the shortlist. This procedure continues by evaluating the neighbors' neighbors until no remaining PUDO location candidates reduce the total cost of matching. The solution approach then jointly optimizes request-vehicle assignment and selection of PUDO locations for each request. The joint optimization of an ILP request-vehicle matching with PUDO locations selection is computationally intensive, limiting its scalability. Fielbaum et al. (2021) mention that incorporating PUDO locations selection increases request-vehicle matching by approximately 10 times.

Our approach differs from Fielbaum et al. (2021) in several ways. First, I decompose PUDO locations selection and request-vehicle matching to decrease computational

complexity and increase scalability. My proposed approach also considers the directionality of non-idle vehicles when shortlisting PU and DO locations for each request—this is not the case in Fielbaum et al. (2021). Finally, unlike Fielbaum et al. (2021), my approach considers the resequencing of all planned PUDO stops assigned to each vehicle, after inserting each new request into a vehicle's planned route, and after choosing each new request's PUDO locations.

4.2.3 Contributions and Research Gaps Addressed

This study makes several contributions to academic literature. The first set of contributions relate to the evaluation of C2C-RP against alternative MOD services and the design of C2C-RP services. The second set of contributions relates to the decision policy and solution algorithm I develop to solve the C2C-RP problem.

This study compares four MOD services (C2C-RH, C2C-RP, D2D-RH, D2D-RP) in terms of operator and user costs. Prior research analyzes at most one pair of these MOD services. Conversely, I systematically compare all four services in terms of operator and user costs, under a unified modeling and algorithmic framework, and a common simulation environment. Assuming D2D-RH is the baseline service, I can isolate the operational benefits (and user costs) stemming from (i) pooling rides and (ii) incorporating walking legs. We can also evaluate the synergistic benefits (and costs) between (i) and (ii).

Moreover, I perform the comparison in a state-of-the-art agent-based transportation system simulation model—POLARIS—that captures road network congestion dynamics. This is another contribution of the study. The simulation permits analysis of a severely overlooked problem related to operating C2C MOD services, namely, vehicles needing to wait

for travelers at PU locations (Wang et al., 2022). In the real world, and in ABM simulations that capture link travel time stochasticity, it is frequently the case that vehicles arrive at PU locations before travelers. This negative outcome significantly impacts the efficiency and productivity of vehicles in the fleet. Capturing link travel time uncertainty in the simulation environment permits a much more realistic assessment of the benefits and costs of C2C service relative to D2D service.

My study also evaluates the impacts of maximum walking distance on fleet performance for a dynamic C2C-RP service without meeting points. This is an important service design parameter for C2C-RP that impacts customer and operator costs. My study is also the first to consider a 'walking' speed in a C2C MOD service that is reflective of travelers having access to an electrified scooter or bike.

Additionally, I propose novel decision policies and algorithmic strategies for C2C-RP that are scalable and operationally effective. Ultimately, as described above, the problem of dynamically operating a C2C-RP service is a highly complex decision problem, where the enormous size of the decision space for even small problem instances essentially precludes elegant models, algorithms, and decision policies, which appears to have stymied research related to C2C-RP. This is particularly discouraging given the potential of a C2C-RP to meet societal and transportation system goals, such as reducing vehicle miles traveled, traffic congestion, energy consumption, and harmful emissions, while providing lower cost (compared to D2D-RH) and high-quality mobility. This study tackles the C2C-RP service problem as the complex set of engineering problems that it entails. The proposed solution framework considers the trade-off between computational run time and solution quality

within a particular decision epoch. I decompose the decision problem in each epoch into a subproblem that matches vehicles to travelers, sequences, and schedules PUDOs for each vehicle, and a second subproblem that assigns travelers to PUDO locations. This decomposition permits the use of optimization techniques for the first subproblem, and a flexible solution algorithm that explores the solution space for the second subproblem. Moreover, I propose an algorithm that efficiently solves these two subproblems sequentially and iteratively.

My proposed solution approach includes several important algorithmic contributions. I only allow PUDOs on links that are in the planned direction of travel of non-idle vehicles. Links in conflicting directions will either never be chosen in the PUDO selection step, or if they are chosen, will likely harm the operational efficiency of the fleet. Moreover, like several other studies, my approach decomposes the decision problem into the request-vehicle matching subproblem and the PUDO selection problem. However, unlike other studies in the literature, my approach allows PUDO selection before or after request-vehicle matching. This permits the service provider to explicitly trade-off between computational run time and solution quality, as selecting PUDOs for candidate request-vehicle matches before the matching step improves solution quality but it increases computational run time. Importantly, I also batch requests over a given time interval and solve the batched request-vehicle matching problem using the bi-partite assignment problem as the engine. My solution approach leverages the benefits of bi-partite matching (i.e., it is highly scalable) while addressing its major limitation for Ride-pooling (i.e., it can only assign one request per vehicle) by iteratively matching requests to vehicles within one decision epoch, such that multiple requests can ultimately be assigned to one vehicle in each decision epoch.

4.3 Problem Description

4.3.1 Nomenclature

The below list describes the notations used for all sets, indexes, variables, and parameters used throughout the rest of this chapter and the following chapter in the problem formulation, methodology and results sections.

R: Set of all requests, indexed by $r \in R$
V: Set of all vehicles, indexed by $v \in V$
L: Set of all links in the road network, indexed by $l \in L$
T: Set of time steps in simulation, indexed by $\tau \in T$
Δ: Time between decision epochs
(r, v): Request r, Vehicle v pair (match or match candidate)
o_r: Original PU link of request r [(x,y) coordinate representing average location along the link]
d_r: Original DO link of request r [(x,y) coordinate]
o_r': Adjusted PU link of request r [(x,y) coordinate]
d_r': Adjusted DO link of request r [(x,y) coordinate]
L_{o_r}: Set of PU links that are within walk range from o_r, indexed by $o_r' \in L_{o_r}$
L_{d_r}: Set of DO links that are within walk of d_r, indexed by $d_r' \in L_{d_r}$
τ_r: Request initiation time of request r [Simulation time (sec.)]
V_r: Set of feasible candidate vehicles for unassigned request r
R_v: Set of requests in v's PUDO sequence (Empty for idle vehicles)
c_{veh}: Vehicle capacity [Seats]
k_{veh}: Maximum number of candidate vehicles for each unassigned request
k_{veh}^{idle}: Minimum number of candidate *idle* vehicles for each unassigned request
θ_{max}: Maximum angle between a request's Euclidean path vector and a vehicle's average future path vector (see Figure 4.2)
D_{dir}^{max}: Maximum remaining distance for a candidate vehicle to complete its current tour so as to be exempt from directionality and detour compatibility constraints [Meters]
$D_{rev}(r, v)$: Euclidean distance between vehicle v's current link and the projection of o_r onto the average future path of vehicle v [Meters]
D_{rev}^{max}: Maximum distance a vehicle can travel in a direction opposite to vehicle's average future path to pick up a new request (see Figure 4.3) [Meters]
$D_{detour}(r, v)$: Euclidean detour distance from vehicle v's current average path to o_r (see Figure 4.3) [Meters]
D_{detour}^{max}: Maximum Euclidean detour distance from vehicle v's current average path to o_r (see Figure 4.3) [Meters]
$\tau_{r,v}^{PU}$: Time at which request r gets picked up by vehicle v at the finalized PU link [Simulation time (sec.)]
$\tau_{r,v}^{DO}$: Time at which request r gets dropped off by vehicle v at the finalized DO link [Simulation time (sec.)]
$t_{r,v}^{w}$: Total wait time for request r to be picked up by vehicle v [Duration (sec.)]
t_{max}^{w}: Maximum PU wait time for a request [Duration (sec.)]
$t(o_r, d_r)$: Approximate direct travel time from o_r to d_r, calculated based on Euclidean distance and average hourly speeds in the zones associated with o_r and d_r during time of matching [Duration (sec.)]
$t_{r,v}^{iv}$: In-vehicle travel time (IVTT) for request r in vehicle v from the finalized PU link location to finalized DO link location [Duration (sec.)]
$t_{r,v}^{iv_{max}}$: Maximum IVTT for request r in vehicle v from the finalized PU link location to finalized DO link location (initially o_r and d_r, respectively) [Duration (sec.)]
$t_{max_{abs}}^{iv}$: Maximum allowable increase in IVTT for any request [Duration (sec.)]
$t_{max_{rel}}^{iv}$: Maximum allowable increase in IVTT relative to a request's direct travel time [%]
$t_{S_{r,v}}$: Time it takes for vehicle v to complete PUDO sequence $S_{r,v}$ after insertion of request r [Duration (sec.)]
$C_{v,r_{(o_r,d_r)}}$: Cost of inserting request r into vehicle v with PU at o_r and DO at d_r [Duration (sec.)]

w_{wt}: Weight parameter for wait time in insertion cost function
w_{ivtt}: Weight parameter for IVTT in insertion cost function
D_{walk}^{PU}, D_{walk}^{DO}: Maximum effective walk ranges from o_r to o_r' and d_r' to d_r, respectively [Meters]
D_{walk}^{max}: Maximum walk range for PU leg, and Maximum walk range for DO leg [Meters]
s_w: Average walk speed [Kilometers per hour]
θ_{buf}: The buffer angle around coordinate-axes to shortlist candidate PUDO links based on their link bearings and the vehicle's travel direction (Section 4.4.3.3) [Degrees]
k_{PUDO}: Maximum number of candidate feasible PU/DO links to be chosen for each new (r, v) match or match candidate for PUDO links adjustment
γ_{C2C}: Boolean parameter to set sequence of PUDO links adjustment and request-vehicle matching. TRUE indicates PUDO links are adjusted before matching stage. FALSE indicates PUDO links are adjusted after matching stage

4.3.2 Problem Statement

In this section I describe the C2C-RP problem. I model a central controller operating a fleet of homogeneous vehicles $V = \{1,2,..,v,...|V|\}$. These vehicles provide service to a set of requests $R = \{1,2,..,r,...|R|\}$ over the analysis/simulation period $T = \{1,2,..,\tau,...|T|\}$, where each request $r \in R$ has an origin o_r and a destination d_r within a geographical service region or network. All requests have a walking speed of s_w. Each request, r, also has a request initiation time τ_r, as well as set of feasible PU locations L_{o_r} and feasible DO locations L_{d_r}. Most importantly, the τ_r values for each request r are unknown to the fleet controller before τ_r. Similarly, the fleet controller only knows L_{o_r} and L_{d_r} for each request r at τ_r. Hence, the fleet controller faces a stochastic dynamic decision problem.

The fleet controller's problem is multi-objective; the objectives include maximizing requests served, minimizing request in-system time, and minimizing total fleet miles. To meet its objectives, the fleet controller must dynamically assign or match requests to vehicles, where (r, v) denotes a request-vehicle match or a candidate request-vehicle match, and $X_{v,r}$ is the binary decision variable equal to one if (r, v) is a match. In this study, I batch requests that arrive every Δ seconds, which is the time between decision epochs. After

matching a request r to a vehicle v, the fleet controller cannot later reject the request r or serve it with another vehicle. Embedded inside the request-vehicle matching subproblem is the sequencing and scheduling of traveler PUs and DOs. Moreover, the fleet controller must determine PUDO locations for requests, where $o'_r \in L_{o_r}$ and $d'_r \in L_{d_r}$ denote the selected PU and DO locations of request r, respectively. In this study, PUDO locations are at the upstream node of each link. Therefore, the fleet controller must determine PUDO *links*, rather than a generic PUDO *location*.

The fleet controller also faces several hard constraints. The first is the vehicle capacity constraint, where c_{veh} is the parameter representing vehicle capacity. Other constraints and their associated parameters include maximum PU wait time for requests (t^{w}_{max}), maximum in-vehicle travel time detour ($t^{iv}_{max_{abs}}$) and maximum percent increase in in-vehicle travel time relative to the request's shortest path travel time ($t^{iv}_{max_{rel}}$). I also enforce a maximum walk distance for both the access and the egress walking legs (D^{max}_{walk}). I also enforce a constraint on non-idle vehicles that they can only travel in the opposite direction of their planned route for a given maximum distance (D^{max}_{rev}).

4.4 Solution Methodology

4.4.1 Overall Flow and Sequence

Figure 4.1 shows the flowchart of the decision policy and iterative solution algorithm at each decision epoch for the C2C-RP problem. The time between decision epochs is Δ, a fixed input parameter. The inputs at each decision epoch are the current status of requests, vehicles, and the network inclusive of link, node, and zonal information. The outputs at each decision epoch are the new assignments of requests to vehicles, the unmatched requests, the

updated PUDO links for every matched request, and the updated vehicle routes and vehicle schedules. Each iteration of each decision epoch involves 3 main stages: (i) finding feasible (r, v) match candidates, (ii) optimizing request-vehicle matches, and (iii) adjusting PUDO links. The following three subsections describe these three main stages in detail.

However, I first want to motivate and describe the iterative nature of my decision policy and solution algorithm. The algorithm is iterative because I use bi-partite matching as the engine of the optimal request-vehicle matching module, similar to Hyland and Mahmassani (2020); Navjyoth Sarma et al. (2020); and Simonetto et al. (2019). Bi-partite matching is highly efficient because dropping the integrality constraint in the math program and using an exact solution method still returns binary solutions. Unfortunately, this property, stemming from the constraint matrix being totally unimodular, comes at a cost—a vehicle can only be matched to at most one request per call to optimal R-V matching module. To partially address the shortcoming of the bi-partite matching approach, we call the bi-partite matching module multiple times in each decision epoch. After solving one instance of the bi-partite matching problem, we then (i) insert the PU and DO links of the newly assigned requests into their matched vehicles and (ii) find a new set of feasible match candidates composed of remaining unassigned requests and vehicles with updated planned routes and schedule. Next, we call the bi-partite matching algorithm again. This iterative process terminates at each decision epoch when there are no feasible request-vehicle matches remaining.

I test two different decision policies, wherein the only difference is whether we adjust PUDO links for *candidate* request-vehicle matches *before* optimal request-vehicle matching,

or we adjust PUDO links for *finalized* request-vehicle matches *after* optimal matching. If the γ_{C2C} parameter is equal to one, the algorithm adjusts PUDO links before matching.

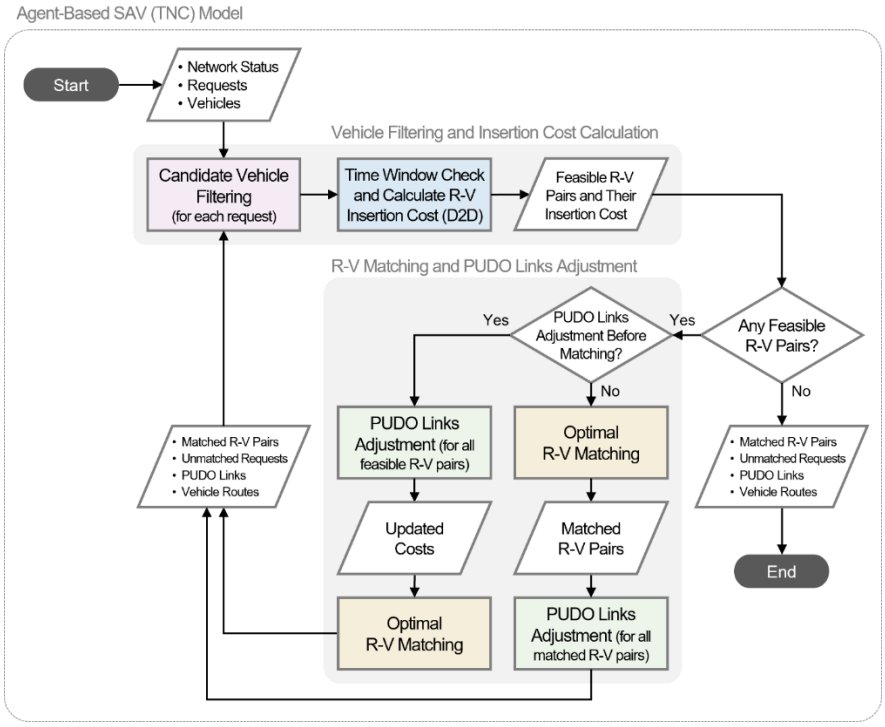

Figure 4.1 Overview of C2C-RP decision policy and solution algorithm (R-V = Request-Vehicle)

4.4.2 Finding Feasible Request-Vehicle Match Candidates

Algorithm 1 describes the overall procedure for finding feasible vehicle candidates, for a single unassigned request. The solution approach repeats Algorithm 1 for each unassigned request at the beginning of each iteration of the iterative optimal matching procedure. Importantly, as we determine candidate vehicles for each request independently, we can easily parallelize across requests.

The following subsections detail the requirements each candidate vehicle must meet to be a feasible match candidate for an unassigned request. These requirements relate to the directionality of vehicle and request travel, the maximum detour for each new request and the requests inside each non-idle vehicle, and time-window constraints. Moreover, in order

to properly determine whether a request-vehicle pair meets the maximum detour and time window requirements, it is necessary to determine the optimal sequence and schedule of traveler PUDOs for a given vehicle (considering each new potential request).

Algorithm 1 – Finding feasible vehicle candidates for unassigned request r

Input: Set of Vehicles V, Unassigned request r
Output: Set of feasible vehicles for request r, $\boldsymbol{V_r}$;
Optimal PUDO sequence for all requests in v, including new request r, for all vehicles $v \in V_r$, $\boldsymbol{S^*_{r,v}}$;
Cost of inserting r into each vehicle $v \in V_r$, $\boldsymbol{C_{v,r_{(o_r,d_r)}}}$
Procedure:
$V_r = \emptyset$
$k = 0$
Find k_{veh}^{idle} idle vehicles near request r and store in set V_r^{idle}
$V_r = V_r \cup V_r^{idle}$
$k = n(V_r)$
Let V_r^{other} be the set of all other available vehicles (idle/non-idle with capacity available) near request r within the t_{max}^{w} time range.
for each $v \in V_r^{other}$ **do**
  if $k == k_{veh}$ **then**
    break
  end if
  if v is idle **and** $v \notin V_r$ **then**
    $V_r = V_r \cup v$
    $k = k + 1$
    continue
  end if
  else
    Check direction and detour compatibility for r with non-idle vehicle v (Algorithm 2)
    if TRUE **then**
      $S^*_{r,v}$ **=** Find optimal PUDO sequence after inserting r into v's PUDO sequence (Algorithm 3)
      Evaluate time-window constraints based on $S^*_{r,v}$ (Section 4.4.2.3)
      if TRUE **then**
        Calculate insertion cost based on $S^*_{r,v}$ (Section 4.4.2.4)
        $V_r = V_r \cup v$
        $k = k + 1$
      end if
    end if
  end else
end for
return V_r; $S^*_{r,v}$ and $C_{v,r_{(o_r,d_r)}}$ $\forall v \in V_r$

4.4.2.1 Direction and Vehicle Detour Checks

This step involves checking the direction compatibility and vehicle path detour constraints for matching a new request r with a non-idle vehicle v that has capacity available

to serve the new request. A vehicle is considered non-idle if it has a non-empty PUDO sequence due to unserved requests matched to it either from a previous decision epoch or from a previous iteration of the optimal matching procedure in the same decision epoch. The procedure to evaluate direction and detour compatibility of a candidate (r, v) pair is described in Algorithm 2. Figure 4.2 and Figure 4.3 provide illustrations of the procedure. Direction and vehicle detour constraints are not evaluated for a non-idle vehicle if it is close to completing its current PUDO sequence (Based on the D_{dir}^{max} parameter). Further description of this step is added in Appendix C.

Algorithm 2 – Evaluating direction and detour compatibility of request r with non-idle vehicle v

Input: Unassigned request r, Candidate non-idle vehicle v
Output: Direction Compatibility TRUE/FALSE
Procedure:
Calculate D_v – The remaining distance in vehicle v's current tour (PUDO sequence)
if $D_v \leq D_{dir}^{max}$ **then**
    return TRUE
end if
Calculate angle θ between vectors representing average future path of vehicle v and Euclidean path between o_r and d_r (See Figure 4.2 and Appendix C for description)
if $\theta > \theta_{max}$ **then**
    return FALSE
end if
Calculate u parameter value (Equation $C2$, Figure 4.3, Appendix C)
if $u > 1$ **then**
    return TRUE
end if
if $u < 0$ **then**
    v needs to travel in reverse direction
    Calculate $D_{rev}(r.v)$ (Equation $C3$, Figure 4.3, Appendix C)
    if $D_{rev}(r.v) > D_{rev}^{max}$ **then**
        return FALSE
    end if
else if $u \leq 1$ **then**
    Calculate $D_{detour}(r, v)$ (Equation $C4$, Figure 4.3, Appendix C)
    if $D_{detour}(r, v) \leq D_{detour}^{max}$ **then**
        return FALSE
    end if
end if
return TRUE

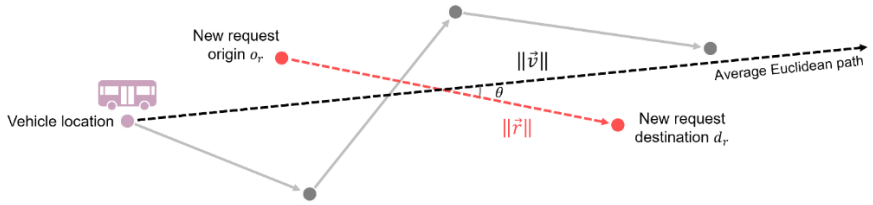

Figure 4.2 Directionality check for finding feasible request-vehicle pairs.

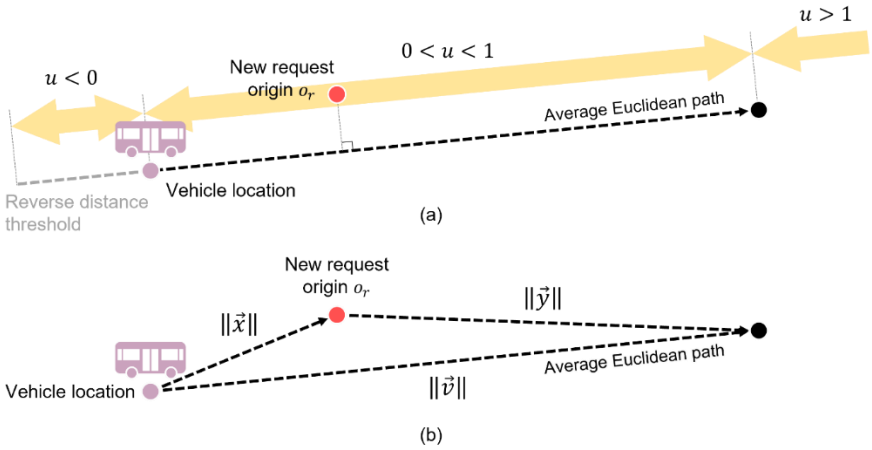

Figure 4.3. Vehicle path detour check for finding feasible request-vehicle pairs, considering (a) PU travel direction threshold and (b) PU travel detour threshold.

4.4.2.2 Finding Optimal PUDO Sequence after Insertion

This step involves finding the optimal order of PUDOs after the new unmatched request is inserted into each candidate vehicle obtained in the previous stage. The insertion location and the revised optimal PUDO sequence is found using a greedy algorithm based on R-tree similar to Gurumurthy and Kockelman (2022). The procedure is described in Algorithm 3 and illustrated in Figure 4.4.

Algorithm 3 – Find optimal PUDO sequence after inserting request r into (candidate) vehicle v

Input: Unassigned request r, Candidate vehicle v
Output: Optimal PUDO Sequence $S^*_{r,v}$, Time to complete sequence $t_{S^*_{r,v}}$
Procedure:
Insert v's current link location coordinates into R-tree
Insert coordinates of v's current PUDO links into R-tree
Insert o_r and d_r into R-tree
Let x_v be the current link location coordinates of v
Let $S^*_{r,v}$ be an empty FIFO queue denoting the optimal PUDO sequence after inserting r into v
while R-tree is not empty **do**
    y_v = Query from R-tree the nearest location to x_v that satisfies precedence constraints
    Push y_v to $S^*_{r,v}$
    Delete y_v from R-tree
    $x_v = y_v$

end while
return $S^*_{r,v}$, $t_{S^*_{r,v}}$

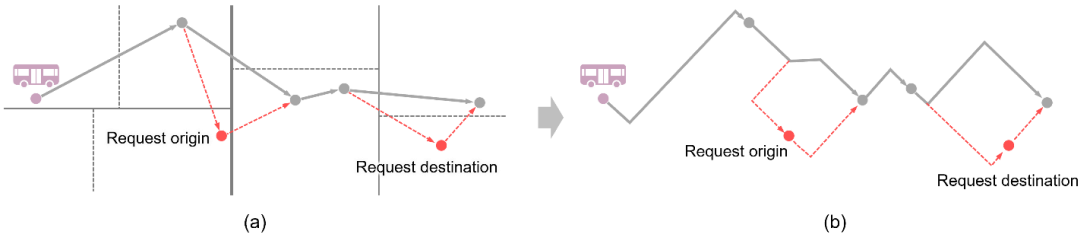

Figure 4.4 R Tree Query to find optimal PUDO sequence: (a) PUDO Sequencing and (b) Updated Path

4.4.2.3 Evaluation of Time-Window Constraints

In this step, time window constraints are checked for each request (new and previously assigned) in the revised PUDO sequence obtained from Algorithm 3. Starting from the vehicle's current link location and the current simulation time, I estimate arrival times at each PUDO link using a Euclidean distance approximation for each trip leg in the candidate vehicle's revised PUDO sequence. I use the average hourly zonal speeds at the origin and destination zones of each trip leg in the tour (at the current simulation time) to estimate the arrival times at each PUDO link. The subroutine checks the following two time window constraints for all requests in the revised PUDO sequence for each (r, v) match candidate:

1. Latest PU time constraint - The latest PU time constraint mandates that the vehicle arrival time at the PU link for any request must be no later than the latest allowable PU time of the request corresponding to that PU link. The PU link is assumed to be the origin link for the new request that is inserted into the candidate vehicle (before PUDO links adjustment). For requests that have already been assigned to the vehicle, PU links in the PUDO sequence could either be their origin link or an adjusted origin based on the decision made in Section 4.4.3 while matching the request to the vehicle. The latest allowable PU time of a request is the request time plus the value of the t^w_{max} parameter. The constraint is expressed in Equation 9 and Equation 10.

$$t^w_{r,v} = \tau^{PU}_{r,v} - \tau_r \tag{9}$$

$$t^w_{r,v} \leq t^w_{max} \tag{10}$$

2. Maximum in-vehicle travel time delay constraint – This constraint mandates that the in-vehicle travel time for each request in the revised PUDO sequence (both previous assigned and new request) from PU time to DO time should not be more than a delay threshold. The subroutine includes two delay thresholds, one is an absolute delay value, and the other is a relative delay value. The constraint is expressed in Equation 11 to Equation 13.

$$t_{r,v}^{iv} = \tau_{r,v}^{DO} - \tau_{r,v}^{PU} \tag{11}$$

$$t_{max}^{iv} = min\left(t(o_r, d_r) + t_{max_{abs}}^{iv}, t(o_r, d_r) \times \left[1 + t_{max_{rel}}^{iv}\right]\right) \tag{12}$$

$$t_{r,v}^{iv} \leq t_{max}^{iv} \tag{13}$$

Euclidean distance along with the average zonal speed at the origin and destination of the request at the time of matching are used to approximate the direct automobile travel time. (r, v) match candidates that fail to meet either of the time window constraints after inserting the new request are discarded in this step.

4.4.2.4 Insertion Cost Calculation for Feasible Candidates

In this step, the cost of adding the new request to a candidate vehicle's tour is calculated for each feasible (r, v) match candidate that fulfilled all time window constraints, and directionality and detour related constraints listed in the previous steps. I use the cost associated with each (r, v) match candidate in the optimal matching stage (described in Section 4.4.4). The insertion cost is calculated as a factor of change in total request wait time and in-vehicle travel time for all requests associated with the candidate vehicle based on the revised optimal PUDO sequence obtained upon inserting the new request into the candidate vehicle. This is expressed in the following equation:

$$C_{v,r_{(o_r,d_r)}} = w_{wt} * t^{w}_{r,v} + w_{ivtt} * \left(t^{iv}_{r,v} - t(o_r, d_r)\right) + \sum_{\forall r' \in R_v} \left[w_{wt} * \Delta t^{w}_{r',v} + w_{ivtt} * \Delta t^{iv}_{r',v}\right] \quad (14)$$

The first term in the expression represents the PU wait time for the new request r if matched with vehicle v. The second term represents the increase in in-vehicle travel time for the new request r based on match candidate (r, v) compared to the direct travel time between the request origin and destination if the person chose to drive alone instead of choosing a C2C-RP service. The terms in the expression after S represent the change in wait times and in-vehicle travel times for all other requests that are in vehicle v's PUDO sequence during the time of matching other than the newly inserted request r.

4.4.3 PUDO Links Adjustment

As the name suggests, the PUDO links adjustment subroutine, involves selecting PU and DO links for a request-vehicle match or a candidate request-vehicle match, depending on whether the PUDO links adjustment subroutine occurs before ($\gamma_{C2C} = 1 - TRUE$) or after ($\gamma_{C2C} = 0 - FALSE$) the optimal request-vehicle matching subroutine. The purpose of the PUDO links adjustment subroutine is to decrease vehicle detours and detours for in-vehicle travelers. Figure 4.5 displays an overview of the PUDO links adjustment subroutine for a request-vehicle. The following five subsections describe the components of the subroutine in more detail.

As a small note, the subroutine does not consider very short trips, i.e., trips with a trip origin to destination Euclidean distance less than the total maximum walking range ($2 \cdot D^{max}_{walk}$). Additionally, the PUDO links adjustment procedure sequentially adjusts PU links and then DO links for a request-vehicle pair.

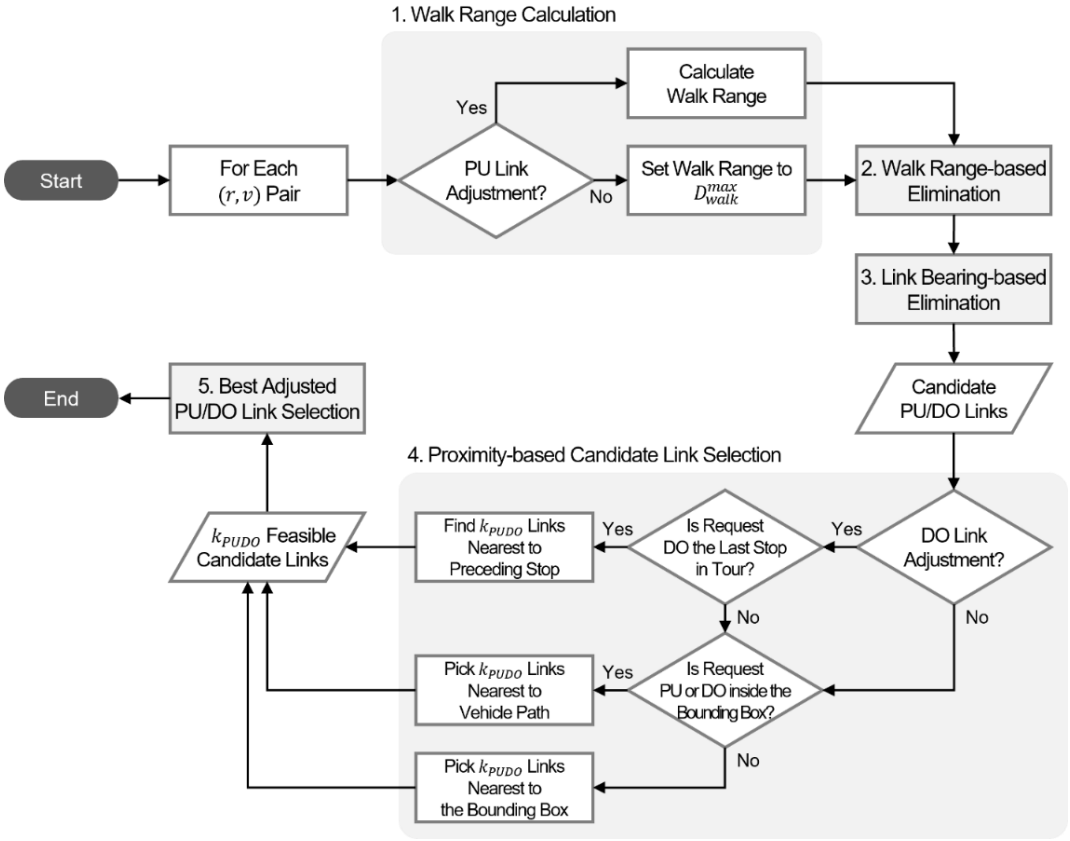

Figure 4.5. Overview of PUDO links adjustment procedure (Repeated sequentially for PU link adjustment and DO link adjustment)

4.4.3.1 Walk Range Calculation

The walk range calculation subroutine determines the maximum distance a traveler can walk to PU links from their trip origin (D_{walk}^{PU}) and from DO links to their destinations (D_{walk}^{DO}), respectively, for a given request-vehicle pair. While we have a hard upper-bound for the maximum walk distance to PU links and from DO links (D_{walk}^{max}) to prevent travelers from having to walk long distances, there is another important consideration when considering PU links for a request-vehicle pair. This consideration is how far away the vehicle is from the request origin location at the current simulation time for a request-vehicle pair. If the vehicle is very close to the request origin, then 'allowing' the request to walk 300-500 meters to a PU location would almost certainly require the vehicle to wait for the request at the PU location for several minutes. Having vehicles, particularly non-idle vehicles, wait at PU

locations for travelers to arrive via walking can significantly reduce the efficiency (and therefore productivity) of the vehicle fleet. Although not captured in the decision model or simulation environment, having vehicles wait at PU locations can also negatively impact traffic flow on the PU link.

The walk range for DO links is always D_{walk}^{max} because a vehicle will never need to wait for a request after dropping them off.

$$D_{walk}^{DO} = D_{walk}^{max} \tag{15}$$

The estimated PU walk range for a new request D_{walk}^{PU} depends on the (i) initial distance between the request origin link and the vehicle's current link, (ii) the average walk speed, and (iii) the average vehicle speed during the matching time step. The initial distance between the request origin and the vehicle's current position is approximated using Euclidean distance. Average walk speed (s_w) is an input parameter. The average vehicle speed (s_v) is approximated as the average of the zonal speeds at the request origin and the current vehicle link at the current simulation time.

Figure 4.6a illustrates the approach used to estimate the PU walk range for the new request. The Euclidean path assumption between the request origin and the vehicle's current link at the current decision epoch is made irrespective of the PUDO sequence of the vehicle. I make this assumption so as not to over-estimate the initial distance of separation between the request origin and vehicle position, and hence minimize vehicle wait time at the adjusted PU link by not setting a high value for D_{walk}^{PU}. Based on this assumption, the following

equation describes the initial Euclidean distance between the vehicle position and the request origin (D):

$$D = s_w t + s_v t$$

where t is the approximate time that the request and vehicle will occupy the same point in Euclidean space if they were to travel directly toward each other. From the above equation, $s_w t$ is the walk range of the request such that the vehicle does not have to wait at the adjusted PU link for the request to arrive. Rearranging the above equation to solve for t yields the following:

$$t = \frac{D}{s_v + s_w} \tag{16}$$

Therefore, the PU walk range for the new request is obtained as follows:

$$D_{walk}^{PU} = s_w t = \frac{D \cdot s_w}{s_v + s_w} \tag{17}$$

PU walk range is capped at a maximum value of D_{walk}^{max} set as an input parameter. Therefore, the effective maximum PU walk range for a request in a request-vehicle pair:

$$D_{walk}^{PU} = min(D_{walk}^{PU}, D_{walk}^{max}) \tag{18}$$

4.4.3.2 Walk Range-based PU/DO Link Elimination

The walk range-based PU/DO link elimination subroutine is the first among several PU/DO link elimination subroutines. The initial set of PU and DO link candidates for each

(potential) request-vehicle match is the set of all walkable links in the network. The link elimination subroutines remove links from this set for each request-vehicle pair.

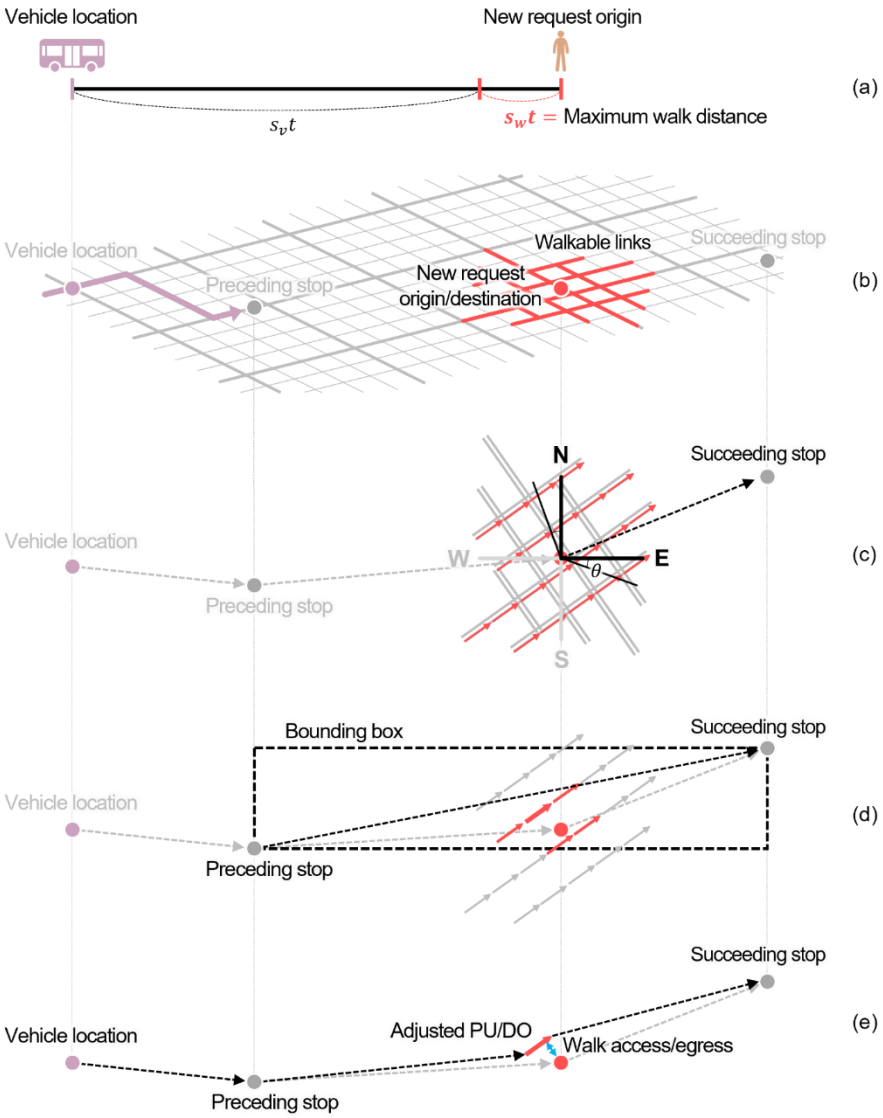

Figure 4.6. PU/DO Links Adjustment: (a) determining the maximum walk range for PUs, (b) eliminating links considering walk range, (c) eliminating links based on their bearing and the vehicle's current planned path, (d) determining k_{PUDO} candidate links nearest to the vehicle path when the request origin/destination is within the bounding box, and (e) choosing a pair of PU and DO links that minimize cost

Figure 4.6b displays the subroutine. The process applies to both PU locations and DO locations, so I will only describe the process for PU locations. Prior to the simulation (i.e., offline), I create one-Origin to many-destination Dijkstra trees for every walk link in the

network. The Dijkstra trees only extend from each link to the links within D_{walk}^{max}. Given the PU walk range (D_{walk}^{PU}) for a request-vehicle pair determined in the walk range calculation subroutine, the walk range-based PU/DO link elimination subroutine eliminates all links from the request origin link's Dijkstra walk tree that are not within D_{walk}^{PU}.

4.4.3.3 Link Bearing-based PU/DO Link Elimination

The link bearing-based PU/DO link elimination subroutine further eliminates PU and DO link candidates for a request-vehicle pair. Figure 4.6c displays this link elimination subroutine. The overall solution algorithm does not call this link elimination subroutine for the new request's DO location if the request's destination would be the vehicle's last planned stop at the current iteration of the current decision epoch.

The link bearing-based PU/DO link elimination subroutine considers the future direction of travel of the vehicle after picking up or dropping off a new request. The future travel direction of the vehicle is obtained by measuring the bearing of the vector connecting the new request's link (origin o_r for PU link adjustment and destination d_r for DO link adjustment) and the matched vehicle's preceding stop. The subroutine classifies the vector into one of 4 directional quadrants based on its relative direction with respect to the North-South and East-West direction axes. Based on this classification, the subroutine eliminates links that do not have a bearing within the same directional quadrant as the future vehicle path, plus a bearing buffer of θ_{buf} in the upper and lower bounds of the quadrant.

4.4.3.4 Proximity-based Candidate Link Selection

The proximity-based candidate link selection subroutine involves selecting up to k_{PUDO} feasible candidate PU links for a request's origin and up to k_{PUDO} feasible candidate

DO links for a request's destination. For each origin and destination, the available k_{PUDO} are those links remaining after walk range-based and link bearing-based link eliminations.

As stated earlier, the PUDO links adjustment algorithmic step adjusts PU and DO links sequentially for an (r, v) match or candidate match, in which it adjusts PU links before DO links. The procedure to choose k_{PUDO} candidate links for PU and DO link adjustment for a request-vehicle (candidate) match is described as follows:

1. Construct an R-Tree of all feasible candidate PU/DO links (PU links for PU link adjustment, DO links for Dropoff link adjustment) that have fulfilled the walk distance range and link bearing constraints as described in Sections 4.4.3.1 and 4.4.3.2.
2. Perform a find nearest links search query on this R-Tree to find k_{PUDO} candidate links based on the following criteria:
    a. For DO link adjustment where the new request's destination is the last stop in the vehicle's tour, query the k_{PUDO} nearest feasible candidate links from the R-Tree closest to the stop preceding the new request's DO link. This is illustrated in Figure 4.7.
    b. For all other cases, construct a minimum bounding box rectangle using the maximum and minimum coordinates of the stops preceding and succeeding the new request's link (origin link for PU link adjustment, destination link for DO link adjustment) in the vehicle's tour. Also construct a polyline object denoting the vehicle's Euclidean path connecting the stops preceding and succeeding the new request's link. The k_{PUDO} nearest link search query is performed based on the location of the new request link with respect to the bounding box:
        i. If the new request link is within the bounding box, then query the k_{PUDO} nearest candidate links from the R-Tree closest to the vehicle path. This is illustrated in Figure 4.6d.
        ii. If the new request link is outside the bounding box, then query the k_{PUDO} nearest candidate links from the R-Tree closest to the bounding box. This is illustrated in Figure 4.8.

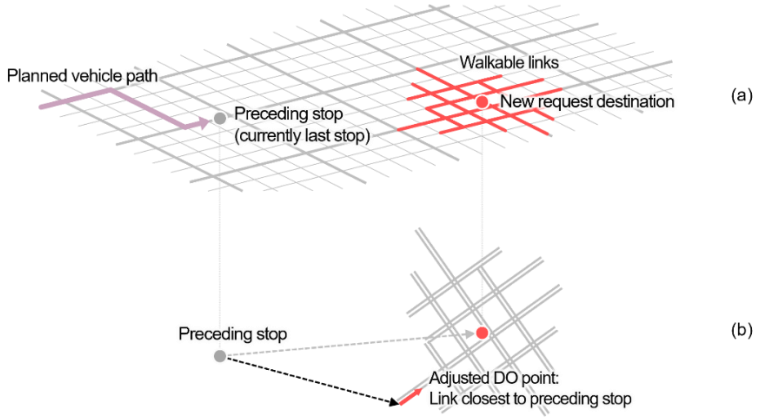

Figure 4.7. DO Links Adjustment in Case of Last DO in vehicle tour: (a) Finding Walkable Links and (b) Finding the Closest Link from Preceding Stop

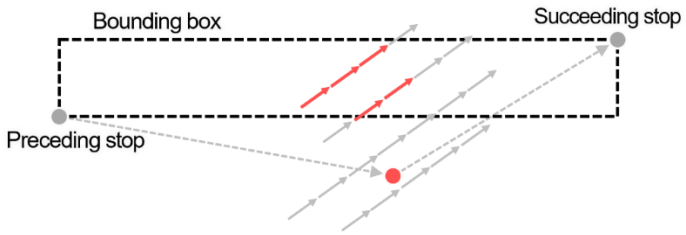

Figure 4.8. PU/DO Links Adjustment in the Outside Bounding Box Case

4.4.3.5 Best adjusted PU/DO Link Selection

This step involves choosing the best link for PU/DO link adjustment by evaluating from the k_{PUDO} feasible candidate links returned from the R-Tree query in the previous step. The PU/DO link is adjusted if it results in a reduction in total vehicle travel time to complete the sequence of PUDOs in its tour (including serving the new request) as known during the matching time step. PU/DO link adjustment is not performed if none of the candidate PU or DO links returned from the previous step results in a travel time reduction for the vehicle. The procedure to select the best candidate PU/DO link is described below in Algorithm 4. Sections 4.4.3.1 to 4.4.3.5 are performed first for PU link adjustment for a (r, v) pair and repeated for DO link adjustment.

Algorithm 4 – Optimal Adjusted PU/Dropoff Link for New (r, v) pair

Input: Unassigned request r, Vehicle v, PU adjustment or DO adjustment (Boolean)
Output: Adjusted PU/DO Location, PUDO Sequence, Insertion Cost and Travel time after PU/DO adjustment
Procedure:
Let $S^*_{r,v}$ be the Optimal PUDO Sequence for vehicle v after inserting request r (from Algorithm 3)
Let $t_{S^*_{r,v}}$ be the minimum total travel time for the vehicle v to complete the optimal PUDO sequence $S^*_{r,v}$
Let $C_{v,r_{(o_r,d_r)}}$ be the insertion cost of adding request r to trip sequence of vehicle v before PUDO links adjustment
Let $C^*_{v,r_{(o'_r,d_r)}}$ be the insertion cost of adding request r to trip sequence of vehicle v after PU Link adjustment
Let $C^*_{v,r_{(o'_r,d'_r)}}$ be the insertion cost of adding request r to trip sequence of vehicle v after DO Link adjustment
Let $L_{PU/DO}$ be the set of optimal PU/DO links returned from Section 4.4.3.4
Let L be the set of links to evaluate
Let l^* be the optimal adjusted PU/DO link
Initialize values based on whether PU or DO link is being adjusted
if PU Link Adjustment **then**
    $S^*_{r,v}$ = $S^*_{r,v}$ returned by Algorithm 3 before PU Link adjustment
    $t_{S^*_{r,v}}$ = $t_{S^*_{r,v}}$ returned by Section before PU Link adjustment
    L = L_{PU} feasible alternative PU links returned by Section 4.4.3.4
    $l^* = o_r$
    $C^*_{v,r_{(o'_r,d_r)}} = C_{v,r_{(o_r,d_r)}}$ returned by Section 4.4.2.4 before PU Link adjustment
end if
if Dropoff Link Adjustment **then**
    $S^*_{r,v}$ = $S^*_{r,v}$ returned by Algorithm 4 after PU Link Adjustment
    $t_{S^*_{r,v}}$ = $t_{S^*_{r,v}}$ returned by Algorithm 4 after PU Link Adjustment
    L = L_{DO} feasible alternative Dropoff links returned by Section 4.4.3.4
    $l^* = d_r$
    $C^*_{v,r_{(o'_r,d'_r)}} = C_{v,r_{(o'_r,d_r)}}$ returned by Algorithm 4 after PU Link adjustment
end if
for each $l \in L$ **do**
    if PU Link Adjustment **then**
        $o'_r = l$
    end if
    if Dropoff Link Adjustment **then**
        $d'_r = l$
    end if
    Find updated optimal PUDO sequence $S'_{r,v}$ and updated vehicle travel time $t_{S'_{r,v}}$ using Algorithm 3
    Find updated request travel times $t^{w}_{r',v}$ and $t^{iv}_{r',v}$ $\forall r' \in v_r \cup \{r\}$ from Section 4.4.2.24.4.2.2 for the updated PUDO sequence $S'_{r,v}$
    Check for violation of time window constraints $\forall r' \in v_r \cup \{r\}$ based on the new optimal PUDO sequence $S'_{r,v}$ after adjusting PU link (Section 4.4.2.3)
    Choose the PU/DO link that results in the least vehicle travel time without violating time window constraints
    if no time window constraints violated **and** $t_{S'_{r,v}} < t_{S^*_{r,v}}$ **then**
        $t_{S^*_{r,v}} = t_{S'_{r,v}}$
        $l^* = l$
        Update insertion cost for the (r, v) pair after adjusting PU or DO link based on Section 4.4.2.4
        if PU Link Adjustment **then**
            $C^*_{v,r_{(o'_r,d_r)}} = C'_{v,r_{(o'_r,d_r)}}$

end if
if Dropoff Link Adjustment **then**
$C^*_{v,r_{(o'_r,d'_r)}} = C'_{v,r_{(o'_r,d'_r)}}$
end if
end if
end for
Finalize adjusted PU/DO link, Insertion Cost, PUDO sequence corresponding to $t_{S^*_{r,v}}$
if PU Link Adjustment **then**
$o'_r = l^*$
$C_{v,r_{(o'_r,d_r)}} = C^*_{v,r_{(o'_r,d_r)}}$
return o'_r, $C_{v,r_{(o'_r,d_r)}}$, $S^*_{r,v}$, $t_{S^*_{r,v}}$
end if
if Dropoff Link Adjustment **then**
$d'_r = l^*$
$C_{v,r_{(o'_r,d'_r)}} = C^*_{v,r_{(o'_r,d'_r)}}$
return d'_r, $C_{v,r_{(o'_r,d'_r)}}$, $S^*_{r,v}$, $t_{S^*_{r,v}}$
end if

4.4.4 Optimal Request-Vehicle Matching

The optimal matching of requests to vehicles is performed either before or after PUDO links adjustment based on the γ_{C2C} parameter. If PUDO links adjustment is performed before matching, then the cost value associated with each candidate (r, v) match is the insertion cost $C_{v,r_{(o'_r,d'_r)}}$ returned by Algorithm 4 after adjusting the PU and DO links for r if assigned to v. Thus, the PUDO links adjustment is also a deciding factor in the optimal (r, v) matching process since the objective function reflects the change in insertion cost for each (r, v) candidate upon adjusting the PUDO links. If PUDO links adjustment is performed after matching, then the cost value for each (r, v) match candidate is the insertion cost before PUDO links adjustment calculated in Section 4.4.2.4. In this case $C_{v,r_{(o'_r,d'_r)}} = C_{v,r_{(o_r,d_r)}}$. The objective of each iteration of the optimal request-vehicle matching procedure is to match unassigned requests to available vehicles such that the total insertion cost is minimized:

$$Z = min \sum_{\forall r \in R^u} \sum_{\forall v \in V_r} \left(C_{v,r_{(o'_r, d'_r)}} - P \right) \cdot X_{v,r} \quad (19)$$

where R^u is the set of all unassigned requests at the beginning of each iteration of the optimal matching stage, and V_r is the set of all feasible vehicle candidates for request r, where $V_r \subset V$. $C_{v,r_{(o'_r, d'_r)}}$ is the insertion cost of assigning request r to vehicle v with the following PUDO links (o'_r, d'_r). P is a penalty cost incurred for not making an assignment; P is a large positive number. $X_{v,r}$ is the binary decision variable denoting whether request r is matched with vehicle v.

The matching problem is subject to two sets of constraints. The set of constraints in Equation 20 suggest that each request $r \in R^u$ can be assigned to at most one vehicle. The set of constraint in Equation 21 limit each vehicle to be assigned to at most one unassigned request.

$$\sum_{v \in V} X_{v,r} \leq 1 \qquad \forall r \in R^u \quad (20)$$

$$\sum_{r \in R^u} X_{v,r} \leq 1 \qquad \forall v \in V \quad (21)$$

The C2C-RP problem is solved iteratively (Sections 4.4.2 to 4.4.4) at each decision epoch until there are no unassigned requests left ($R^u = \emptyset$) or no feasible candidate vehicles for each unassigned request ($V_r = \emptyset, \quad \forall r \in R^u$).

4.5 Conclusion

This chapter introduced, formulated, and presented a solution methodology for the Corner-to-Corner Ride-pooling problem (C2C-RP) that involves MOD services with dynamic and flexible PUDO locations called Virtual Stops. This chapter includes several algorithmic contributions related to addressing the C2C-RP operational problem, delineated in Section 4.2 relative to the existing literature and described in Section 4.4. The chapter describes the challenges involved in solving the C2C-RP problem in a scalable manner due to the number of sub-problems involved that increase the decision space. The chapter presents a decomposition-based solution methodology to solve the C2C-RP problem that involves matching requests to vehicles, updating PUDO order sequences for each vehicle and also adjusting PUDO locations (or links) for each request to nearby locations (or links). The solution methodology presented is different from the existing literature in several ways, including but not limited to the procedure to identify viable alternative PUDO locations, dealing with early vehicle arrivals at PU locations, and the sequence of decision making.

The next chapter (Chapter 5) implements the solution algorithm proposed in this chapter in an Agent Based Modeling framework (POLARIS) and performs extensive analysis of user, operator, and computational costs for C2C and D2D variants of Ride-pooling and Ride-hailing services.

Chapter 5. Analyzing Cost Trade-offs in MOD services with Virtual Stops

5.1 Introduction

In the previous chapter, I described a scalable solution algorithm for the C2C-RP problem, which involved MOD services with dynamic and flexible PUDO locations (called as Virtual Stops). The most commonly used type of MOD service today is the Door-to-Door Ride-hailing service (or D2D-RH) that serves a single request from/to their doorstep. Significant gains in operator efficiency could be obtained when multiple requests are served by the same vehicle at the same time (Ride-pooling or RP) and/or when flexibility is introduced to the Pickup and Dropoff locations of the request (C2C services). However, such gains in operator efficiency could come at the expense of i) increased inconvenience to travelers due to increases in travel time and/or having to walk a short distance for the first/last mile of their trip, and ii) increased computational cost for the operator as C2C services introduces additional sub-problems for the operator to solve apart from matching and PUDO sequencing in D2D services. The objective of this chapter is to systematically analyze the user, operator, and computational cost trade-offs for D2D and C2C variants of MOD services.

The rest of the chapter is organized as follows: Section 5.2 describes the simulation environment in which I test the proposed C2C-RP decision policy and compare the four MOD service options (C2C-RP, D2D-RP, C2C-RH, D2D-RH). Section 5.2 also describes the computational experiments, including the performance metrics and scenarios to assess the operator and user costs associated with the four MOD service options. Section 5.3 presents and discusses the results of the computational experiments and sensitivity analysis. Section

5.4 concludes the second study of my dissertation (Chapter 4 and Chapter 5) with a summary as well as limitations and discussion of future research directions.

5.2 Computational Experiments Setup

To evaluate the performance of the proposed C2C-RP decision policy and algorithmic approach, I constructed a large number of computational experiments. This section describes the simulation environment wherein I embed the proposed C2C-RP decision policy and algorithmic approach (Section 5.2.1), the road network model (Section 5.2.2), the performance metrics for analysis (Section 5.2.3), and the scenarios for testing (Section 5.2.4). The four MOD service types evaluated are C2C-RP, C2C-RH, D2D-RP, D2D-RH. To operate D2D services I use the same decision policy and solution algorithm as the C2C service, except that PUDO links are not adjusted (i.e., Section 4.4 excluding section 4.4.3). Similarly, to operate Ride-hailing services, I use the same decision policy and solution algorithm as the Ride-pooling service, except that I only consider idle vehicles in the matching subproblem.

5.2.1 POLARIS Simulation Environment

I use the large-scale agent-based simulation framework called POLARIS (Auld et al., 2016) to compare the four MOD service types and evaluate the C2C-RP decision policy and algorithmic approach. POLARIS integrates supply and demand and allows the simulation of MOD services in a congestible network with full feedback (Gurumurthy et al., 2020). POLARIS and its modules can simulate activities and trips in a large metropolitan region with over 10 million people in under 5 hours.

POLARIS is a high-performance C++ codebase for agent-based modeling of transportation demand and supply. The tool consists of several modules for population

synthesis, long-term and short-term planning, vehicle routing, traffic flow, and has functionality to model transit, MOD services, and freight at a high-level of detail. The population synthesis module creates person agents for the target region based on underlying demographic information as sourced from the Census and American Community Survey (ACS) in the United States. Consistent with agent-based modeling, each individual makes travel and activity decisions based on its individual characteristics and information available to the individual regarding the state of the system (e.g., prevailing travel times, modal attributes, destination attributes, etc.).

5.2.1.1 Corner-to-Corner Routing Module in POLARIS

The MOD module in POLARIS models operations through a centralized operator. The operator maintains control over all fleet vehicles, tracks their real-time information, and runs algorithms to determine the next set of instructions to pass to the vehicles. To enable easy incorporation of new algorithms, a generic structure exists in POLARIS for different aspects of MOD control: such as matching, repositioning, charging (in the case of EVs), and parking. These strategies are custom coded to either serve a single purpose like a matching algorithm or the modeler can control several aspects of operation simultaneously like the joint control of matching, repositioning, and charging of fleet vehicles (Dean et al., 2022).

I now briefly describe the general flow of information and control in POLARIS for a MOD request to be received and served. The fleet operator is informed of requests that originate from person agents. Requests encompass information on time of request, PU location, DO location, and the estimated fare, travel time, and costs associated with serving it. The matching subroutine receives trip requests from the operator and can either

immediately assign an available vehicle or batch the requests over a pre-defined duration and then solve an optimization-based request-vehicle matching problem. Once the matching subroutine is complete, an update regarding the match is passed on to both the person agent and the assigned vehicle (if matched). Assuming the match is successful, the vehicle is instructed to route itself to the PU location and then, on picking the traveler, proceed to the DO location. While en-route, new requests can be added to the existing vehicle trip, with the vehicle reporting available seats remaining.

The C2C service in this study requires travelers to walk to their PU locations and from their DO locations. Trips in POLARIS are typically modeled to start and end at activity locations. Activity locations represent buildings and places that are typical origins and destinations. Depending on network density incorporated in the model, some level of aggregation may exist in the number of activity locations used to represent the underlying origins and destinations. Route computation considers all possible links associated with the origin and destination activity locations as candidates for the shortest path, and finally results in a link-to-link vehicle trajectory. Once in the network, vehicles follow the trajectory subject to constraints of the traffic flow model. Trips are completed when the vehicle enters the last link of its trajectory. While Gurumurthy and Kockelman (2022) show the benefits of aggregation at the activity location level for C2C MOD service, their approach does limit flexibility in assigning travelers to PUDO locations in the region. Hence, in the current study, we allow traveler PUDOs in every link in the network. This approach both reduces overall VMT and improves traveler experiences compared to the approach in Gurumurthy and Kockelman (2022), as the later study requires all travelers to walk to PUDO locations, even if these walking trips do not improve vehicle utilization and productivity. Changes in PUDO

links are transmitted to the person and vehicle agents at the end of each batching interval once the iterative procedure terminates.

5.2.2 Study Network

I use the default network in POLARIS – the Bloomington IL network (Figure 5.1Figure 5.1) – to evaluate the C2C-RP algorithm. The network includes 3057 intersections, 4527 directional drive links, 6885 walk links and 185 Traffic Analysis Zones (TAZs). The walk network is broken down into links with a maximum length of 250 meters. There are 2,833 activity locations in the network which represent request origins and destinations. The default PU link and DO link for each new request is the link closest to the request origin and destination activity locations respectively.

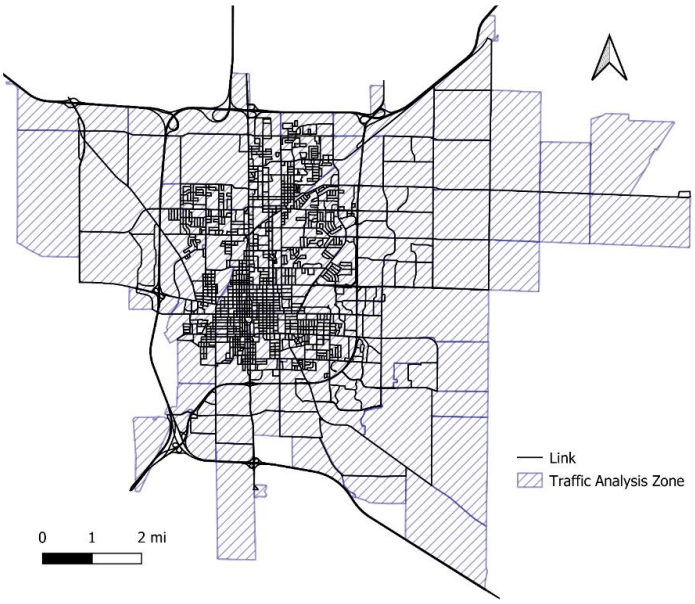

Figure 5.1. Bloomington, IL Network

5.2.3 Metrics for Analysis

To compare the four MOD service types, I consider two cost dimensions, operator costs and user costs. I use average vehicle kilometer traveled (VKT) per served request as the main operator cost metric. I normalize by 'served requests' to prevent outcomes where

an efficient and productive service generates more VKT while serving a large number of requests than an inefficient and unproductive service that generates lower VKT while serving significantly fewer requests.

The main customer cost metric is request-to-destination time, which as the name suggests, measures the total time between a traveler requesting a vehicle and the same traveler arriving at their destination location. In C2C services, request-to-destination time includes time waiting to be assigned, access walking time, waiting time at the PU link, in-vehicle time, and egress walking time. The results will explicitly delineate the time travelers spend in each state.

I also include matching rate, the percentage of all MOD traveler requests who are served by the MOD service, as a key performance metric. This metric is necessary for a holistic comparison of MOD service types; without matching rate, a service type or decision policy that prioritizes *easy-to-serve* requests may perform well in terms of VKT per served request and request-to-destination time for served request, compared with other policies that attempt to serve all requests.

The final metric I use is average vehicle occupancy (AVO) per VKT. AVO per VKT includes empty vehicle kilometers as well as vehicle kilometers with one passenger, two passengers, three passengers, and four passengers. This metric does not perfectly align with either operator costs or user costs, but it is a metric of interest to transportation planners and policymakers.

5.2.4 Scenarios

This section describes the scenarios developed to help answer the study's two main research questions. I first introduce the baseline scenarios, specifically which parameters I vary in the baseline scenarios and which parameters are fixed. Next, I describe the additional scenarios developed to provide insights into various service design decisions, the inefficiencies associated with travel time uncertainty, and various algorithmic strategies and decision policies.

Table 5.1 shows that in the set of baseline scenarios, the only two parameters that vary are the MOD service type and the fleet size. Maximum walk range, walking speed, and the sequencing of matching and PUDO links selection are fixed. These baseline scenarios aim to compare and contrast the four MOD service types under various supply-demand ratios, where the demand is fixed, and the supply varies across scenarios. The results of the baseline scenarios are described in Section 5.3.1. The number of available seats for Ride-pooling services (C2C-RP and D2D-RP) is fixed at 4.

After the baseline scenarios, I create several additional sets of scenarios. The first additional set of scenarios involves varying the maximum walk range parameter D_{walk}^{max} alongside the MOD service type and fleet size to perform sensitivity analysis on user and operator costs. I vary D_{walk}^{max} between 250 and 1000 meters in an increment of 250 meters. The results are presented in Section 5.3.2.

The second additional set of scenarios varies 'walking' speed s_w alongside MOD service type and fleet size. In these scenarios, there I use a baseline s_w of 5 km per hour

speed, and the 'fast walking' (or e-scooter) speed of 20 km per hour. Results of this analysis are provided in Section 5.3.3.

The third additional set of scenarios varies the sequence of Request-Vehicle matching and PUDO links selection sub problems, alongside MOD service type and fleet size. This is done by enabling/disabling the Boolean flag γ_{C2C}. The two alternative orders are adjust PUDO links after matching and adjust PUDO links before matching. Section 5.3.4 describes the results of this analysis.

Finally, I also run a set of scenarios to analyze the computation time and scalability of the implemented solution methodology for the C2C-RP problem (Section 5.3.5). This is done by analyzing the total computational time and stage-wise computational time by varying D_{walk}^{max}, s_w and γ_{C2C} parameters across different MOD service types for 2 different fleet sizes (5,000 and 10,000 vehicles).

To ensure consistency while comparing across several scenarios, the total demand, origin, destination as well as the request initiation times are kept fixed across all scenarios. Repeating the same scenario with all parameters unchanged yielded a less than 0.1% change in values of the metrics evaluated. The spatial distribution of the initial location of vehicles at the beginning of simulation is also same across scenarios. Fleet repositioning is also disabled so as to control for idle vehicle movements from biasing the output metrics evaluated across the four different MOD types.

Table 5.1 Fixed and Variable Parameters for Baseline Scenario (Section 5.3.1)

Fixed or varying?	Parameter	Parameter Values
Varying Params	MOD Service Type	D2D-RH ($c_{veh} = 1$)
		D2D-RP ($c_{veh} = 4$)
		C2C-RH ($c_{veh} = 1$)
		C2C-RP ($c_{veh} = 4$)
	Fleet Size (veh.)	1000, 1500, 2000, 3000, 4000, 5000, 6000, 7500, 9000, 10000
Fixed Params for PUDO Links Adjustment (C2C-RP and C2C-RH)	D_{walk}^{max}	1,000 m
	γ_{C2C}	FALSE (Adjust PUDO link after matching)
	s_w	5 km/h
	k_{PUDO}	5 link candidates
	θ_{buf}	30°
Fixed Params for R-V Matching (All Four MOD Services)	k_{veh}	8 vehicles
	k_{veh}^{idle}	2 vehicles
	θ_{max}	30°
	D_{dir}^{max}	3000 m
	D_{rev}^{max}	3000 m
	D_{detour}^{max}	6000 m
	t_{max}^{w}	1200 seconds
	$t_{max_{abs}}^{iv}$	900 seconds
	$t_{max_{rel}}^{iv}$	50%
	w_{wt}	1.0
	w_{ivtt}	1.0
Simulation Parameters	Analysis Period	24 hours
	Δ	30 seconds
	# Requests	221,711 (Fixed) 1 request = 1 traveler
	Initial Fleet Location	Distributed inversely proportional to TAZ area
	Fleet Repositioning	FALSE

5.3 Results and Discussion

5.3.1 Operator-User Cost Trade-off in MOD Systems – Baseline Scenario

This section aims to answer this study's main research question: what are the trade-offs between operational costs and user costs across D2D-RH, D2D-RP, C2C-RH, and C2C-RP? Figure 5.2b shows significant operator cost differences between the four MOD services, particularly when the supply (i.e., vehicle fleet size) is low. With a fleet size of 2000 vehicles,

C2C-RP has an average VKT per request of nearly 5.5, which is lower than the D2D-RP at 6.0, and substantially lower than C2C-RH at 7.0 and D2D-RH at over 7.6. The operator cost benefits of pooled-rides at low fleet sizes are even more impressive when considering the significantly higher matching rates for Ride-pooling than Ride-hailing. Hence, there are clear operational cost benefits associated with pooled rides, as found in many other studies (Hyland and Mahmassani, 2020).

Figure 5.2b also shows there is undoubtedly an operational benefit associated with C2C service over D2D service. Interestingly, this gap in VKT per request seems to remain steady, or even increase, as the fleet size increases. This is an important finding, particularly the quantification of the VKT per request gap between C2C and D2D services. While the gap is not as large as the Ride-pool vs. Ride-hail gap for matching rate in Figure 5.2a, at low to medium fleet sizes, C2C service does have a significantly higher matching rate than D2D, especially at lower values of fleet size.

Figure 5.2c shows the clear downsides of Ride-pooling and C2C service, relative to Ride-hailing and D2D service, respectively in terms of total user travel time. Given the low matching rates for both Ride-hail services at fleet sizes below 3000, I will compare the services in terms of request-to-destination time for fleet sizes 3000 and larger. At a fleet size of 3000 vehicles, D2D-RH service has an average request-to-destination time of slightly over 8 minutes, whereas the C2C-RH service is around 13 minutes. The gap between these two services is nearly all due to egress walk distance/time. There is also an approximately 4-minute gap for request-to-destination time between C2C-RP and D2D-RP.

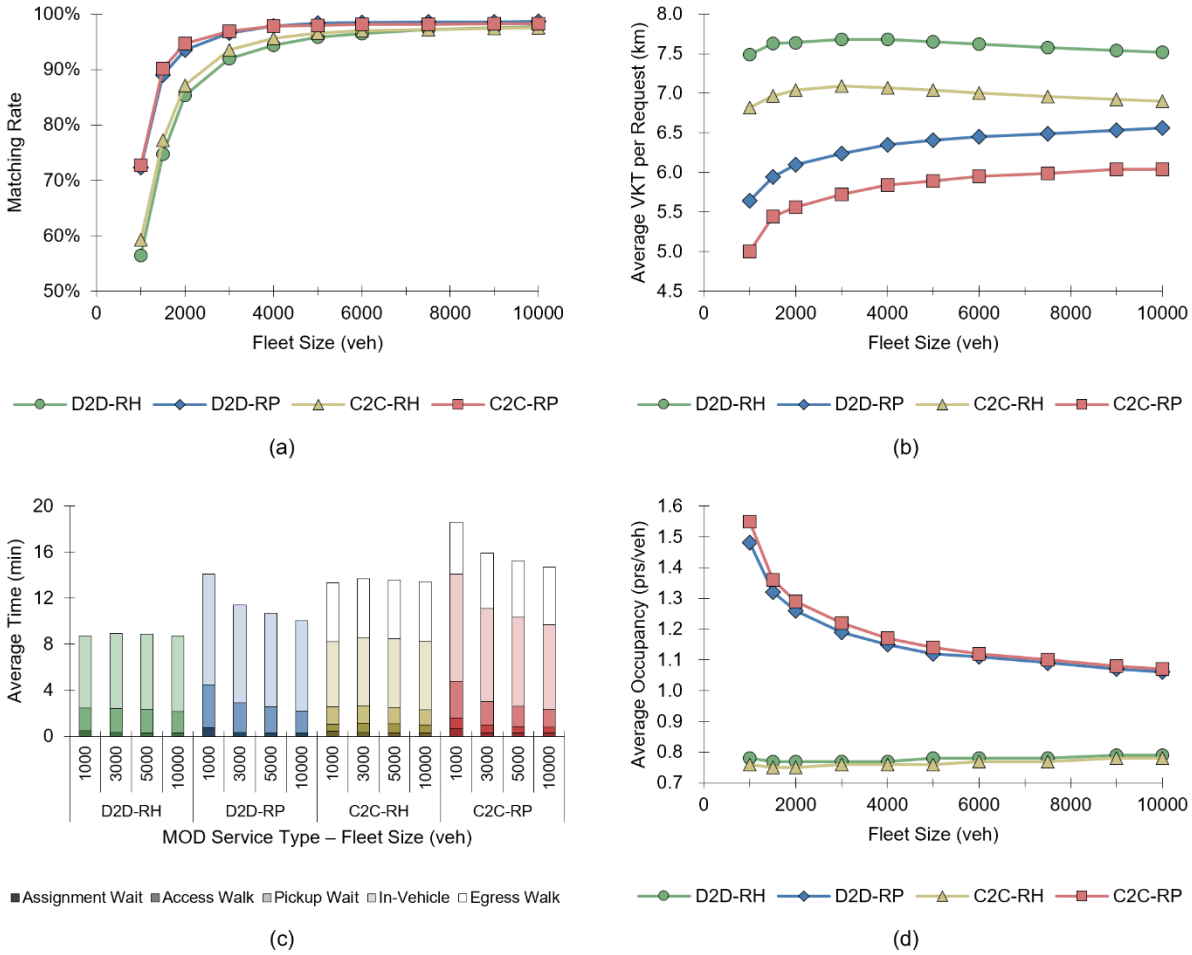

Figure 5.2. Performance metrics for MOD service types: (a) Matching Rate, (b) Average VKT per Request, (c) Average Travel Time broken down by Segment, and (d) Average Vehicle Occupancy

Figure 5.2d is consistent with Figure 5.2a and Figure 5.2b and the features of each service design—the AVO per request is significantly higher for Ride-pool compared to Ride-hail. There is also a gap between C2C and D2D, especially at low fleet sizes, but it is noticeably smaller than the gap between Ride-pool and Ride-hail.

Additionally, Figure 5.2d shows that AVO decreases steadily as fleet size increases. This is simply the result of the decision policy using all the vehicles in the fleet to reduce traveler wait times as shown in Figure 5.2c. Interestingly, AVO per request of C2C-RP is slightly greater than D2D-RP for low fleet sizes, whereas AVO per request of C2C-RH is slightly lower than D2D-RH (or nearly the same) for all fleet sizes. This could potentially be

because adjusting PUDO links for a Ride-hailing service does not increase sharing of trips in the system since a Ride-hail vehicle serves only one request at a time.

In summary, regarding the original research question, there is a clear trade-off between operator cost and user costs in these four MOD service designs. The gap between Ride-hail and Ride-pool is the most significant in terms of operator costs, but there is also a significant gap between C2C and D2D (5 to 10% reduction in VKT per served request for C2C services compared to D2D). In terms of user costs, there is not a huge gap between Ride-hail and Ride-pool, but there is a significant gap between C2C and D2D. However, the gap between C2C and D2D is almost entirely due to the egress walking distance/time.

The implications of the gap between C2C and D2D being from egress walking distance are several. First, from a behavioral perspective, some users may find this egress walking to be both highly inconvenient and onerous, while other users may find a 3-or 4-minute walk to be only slightly inconvenient and even pleasant, depending on their schedule, the weather, and various other factors. The heterogeneity of travelers in terms of the willingness to walk suggests there is likely a role for multiple service offerings from the same MOD service provider.

Second, from a decision policy/algorithm and service design perspective, it is possible to put more weight on the disutility of walking in the algorithm, in order to decrease average egress walking distances. However, this algorithmic change would likely increase the operator costs for a C2C service. Similar to applying more weight to the disutility of walking in the decision policy function, it is also possible to explicitly limit the maximum total and/or

egress walking distance for travelers. The next section analyzes variations in this parameter in terms of both operator and user costs.

Figure 5.3 illustrates the effect on user and operator costs if MoD operators offer pooled C2C services with only PU Link adjustments (C2C-RP-PU) or only DO Link adjustments (C2C-RP-DO) compared to C2C-RP and D2D-RP for the baseline scenario parameters as listed in Table 5.1. Figure 5.3a shows that VKT per served request for C2C-RP-PU is slightly less than D2D-RP while it is slightly more for C2C-RP-DO compared to C2C-RP. This is also reflected in the segment wise travel times shown in Figure 5.3, with total travel time of C2C-RP-PU being slightly more than D2D-RP, while total travel time of C2C-RP-DU being slightly less than C2C-RP. This trend can be explained when comparing the % of requests with PU Links adjusted with those with DO links adjusted (Figure 5.3c and Figure 5.3d) as well as the average access and egress walk distances in their respective cases (Figure 5.3e and Figure 5.3f). Even though the maximum walk range is 1,000 m for both C2C-RP-PU and C2C-RP-DO, the average access walk distance for adjusted PU links in C2C-RP-PU is less than half of the egress walk distance for adjusted DO links in C2C-RP-DO. This is because, the algorithm proposed in this study calculates the effective maximum PU walk range (D_{walk}^{PU}) for each request-vehicle pair in such a way that instances of vehicles having to wait for the request at the adjusted PU link is reduced (Section 4.4.3.1). Whereas this does not apply for DO links adjustment. This also explains the higher instances of requests with DO links adjustment (Figure 5.3d) compared to those with PU links adjustment (Figure 5.3c).

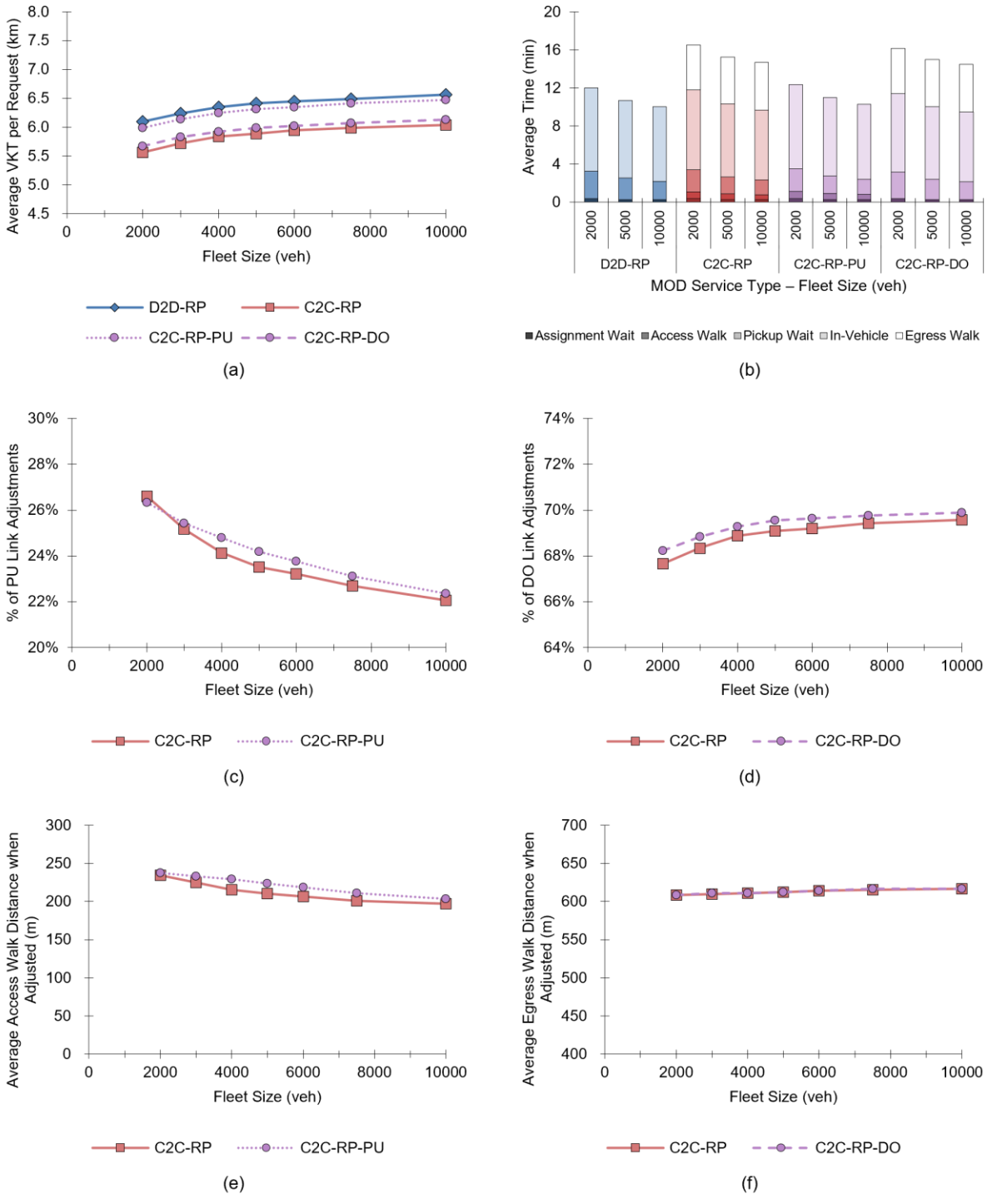

Figure 5.3. Performance metrics with selective 'PU Only' or 'DO Only' link adjustments (a) Average VKT per Request, (b) Average Travel Time broken down by Segment, (c) % of Requests with PU Links Adjusted, (d) % of Requests with DO Links Adjusted, (e) Average Access Walk Distance when PU Link is Adjusted, and (f) Average Egress Walk Distance when DO link is Adjusted

The percentage of instances as well as average access walk distance for adjusted PU links in C2C-RP-PU scenario is slightly more than C2C-RP. This could be because, since DO

links are not adjusted in C2C-RP-PU, it offers slightly more flexibility for the operator to adjust PU links without violating time window constraints. The prevalence of higher instances of DO links adjustment in C2C-RP-DO compared to C2C-RP can also be explained similarly. Additionally, the access walk range and percentage of requests with PU links adjustment decreases with increasing fleet size for both C2C-RP-PU and C2C-RP. This is because an unassigned request is closer to an available vehicle as fleet size increases. On the other hand, % of requests with DU links adjustment increases with increasing fleet size (Figure 5.3d) as more vehicles means less waiting time giving more flexibility to adjust DO links. It is also interesting to note that the average egress walk distance for requests with DO links adjustment remains more or less flat across fleet sizes. This could be because unlike calculating the effective maximum PU walk range, DO walk range calculation is not affected by proximity of requests to a vehicle.

5.3.2 Sensitivity Analysis with respect to Walk Range

Figure 5.4 displays the computational results for variations in maximum walking distance D_{walk}^{max} (i.e., maximum walk range for PU as well as DO), across three fleet sizes and both C2C MOD services. The figure also shows the results for the two D2D MOD services, but these services always have zero walk range. Walk speed (s_w) is kept fixed at 5 kmph.

The matching rate results in Figure 5.4a indicate that for these three fleet sizes, walking range does not have significant practical impact on matching rate, although longer walking ranges do slightly increase matching rate.

Figure 5.4b and Figure 5.4c show that as walking range decreases from 1000m to 250m, VKT per request increases and request-to-destination time decreases for both C2C-

RH and C2C-RP. Interestingly, the relationship is strongly linear for both C2C services, both performance metrics, and all three fleet sizes. However, the results indicate that there is a step change improvement in VKT per request when moving from D2D to *some walking*, i.e., a 250m walking range. This latter result suggests that even allowing a small walking distance can produce significant operational efficiencies in the system. This should not be surprising, but it is important, as having travelers walk even a short distance can prevent the worst-case PUDO links for vehicles. A small amount of walking can also ensure the PUDO links are on links with the same bearing as the vehicle's direction of travel after picking up or dropping off a request.

Figure 5.4d shows that walking range does not have a significant impact on AVO for Ride-pool or Ride-hail nor for any fleet size. Even though Figure 5.4d showed that the AVO for C2C-RP with a baseline walk range of 1000 meters is slightly more than D2D-RP for low fleet sizes, Figure 5.4d denotes that change in AVO for C2C-RP services for every 250 meter increment in maximum walk range is very minimal.

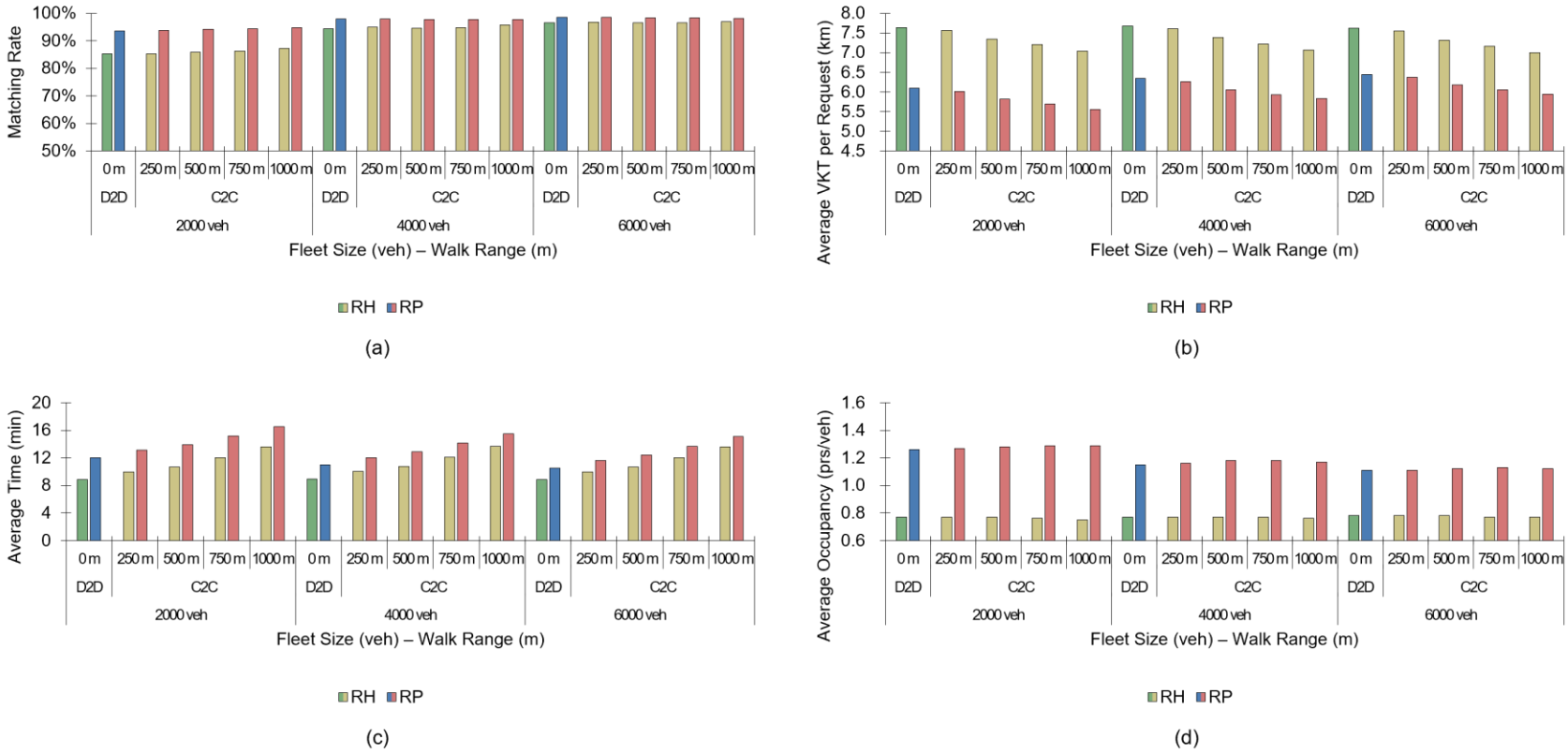

Figure 5.4. Sensitivity Analysis with respect to Walk Range: (a) Matching Rate, (b) Average VKT per Request, (c) Average Request-to-destination Time, and (d) Average Vehicle Occupancy

5.3.3 What about e-scooters? Sensitivity Analysis with respect to Walk Speed

This section analyzes the change in fleet performance and user cost with respect to changes in walk speed (s_w). I compare the baseline walking speed of 5km/h with a much faster 'walking' speed of 20km/h. This analysis serves two purposes, namely, to assess the potential benefits of using ubiquitous—personal or shared—e-scooters as access and egress modes to/from adjusted PUDO links, and to illustrate the role of travel time uncertainty across service performance metrics.

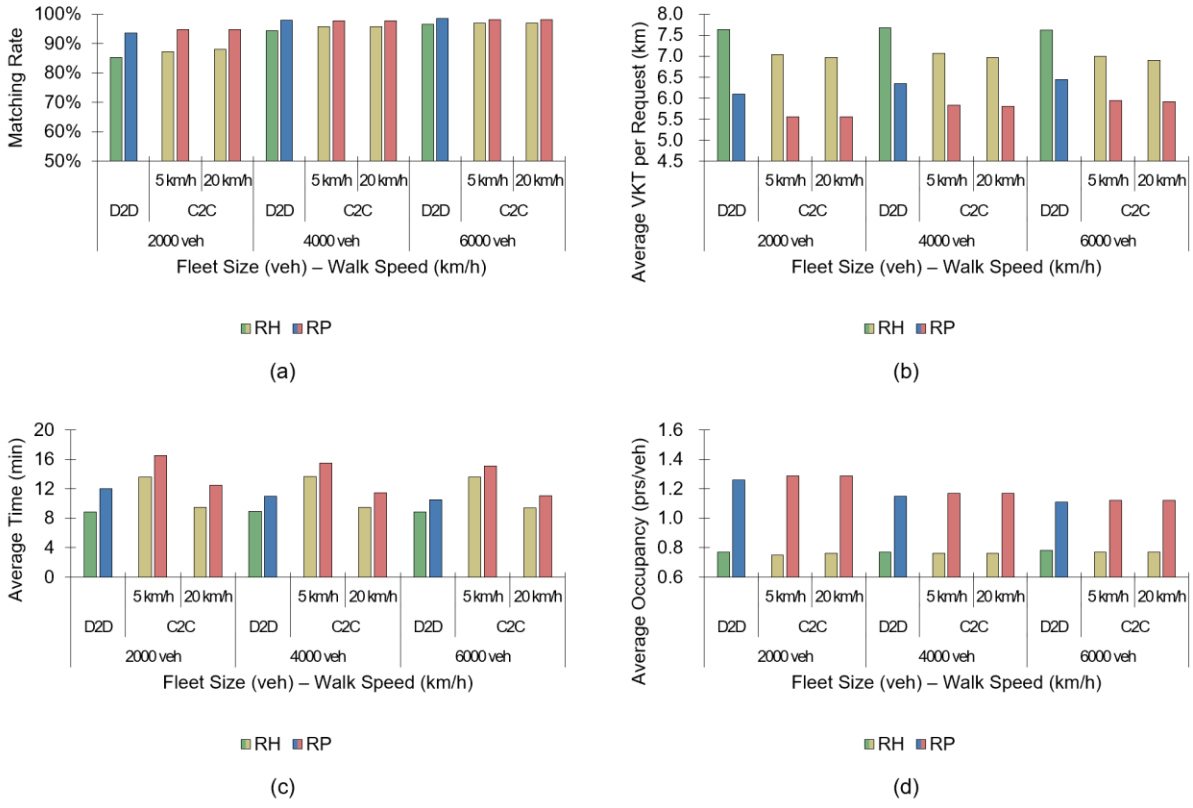

Figure 5.5. Sensitivity Analysis with respect to Walk Speed: (a) Matching Rate, (b) Average VKT per Request, (c) Average Request-to-destination Time, and (d) Average Vehicle Occupancy

Figure 5.5 shows that increasing walking speed increases matching rate slightly, decreases average VKT per request, significantly decreases request-to-destination time, and does not significantly impact AVO, compared to results presented for baseline walk speed in Figure 5.2. Figure 5.5c shows that the increased walking speed decreases request-to-

destination time by around 5 minutes. This five-minute reduction in travel time effectively makes C2C competitive with D2D for both Ride-hailing and Ride-pooling along this dimension, while retaining a significant advantage in terms of VKT per request.

Figure 5.6 aims to provide insights on how/why faster walking speeds dramatically improve request-to-destination time, while also improving VKT per request and matching rate. Figure 5.6a shows the matching rate as a function of walking range, walking speed and fleet size. Naturally, as all three of these input parameters increase, particularly fleet size, the matching rate increases. Figure 5.6b shows the frequency (i.e., percentage) of users with adjusted PU links (i.e., a different PU link than their trip origin). As walk range and walk speed increase, the frequency of PU link adjustments increases. The relationship between fleet size and PU link adjustments is not monotonic. Matching rate and frequency of PU link adjustments are relevant background information, for the two main sets of results in Figure 5.5c and Figure 5.5d.

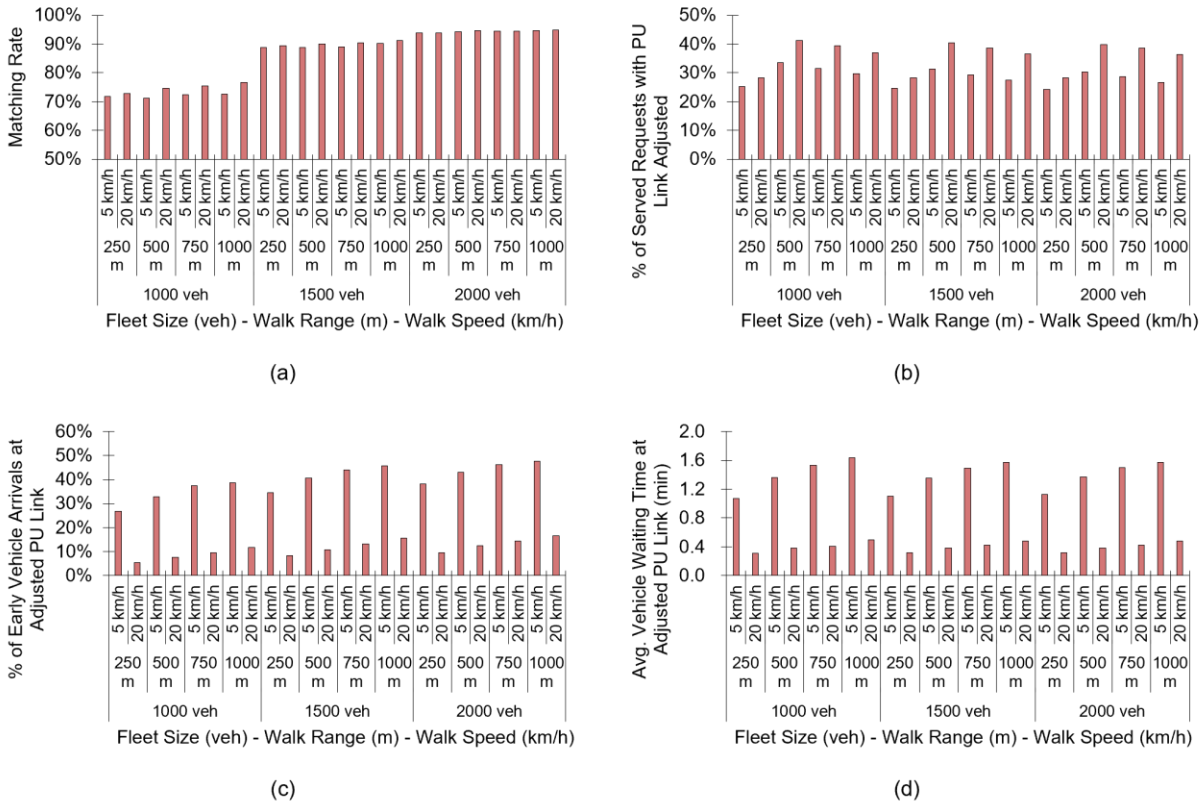

Figure 5.6. PU Link Adjustments and Vehicle Waiting at PU Link (C2C-RP Scenarios): (a) Matching Rate, (b) PU Link Adjustment Rate, (c) Rate of Vehicles Waiting for Requests with Adjusted PU links, and (d) Average Vehicle Wait Time at Adjusted PU Link for Early Arrivals

Figure 5.6c shows the probability a traveler's assigned vehicle had to wait at the traveler's PU link for the traveler to arrive, conditional on the traveler having an adjusted PU link. In such instances, the vehicle arrived earlier than the traveler at the adjusted PU link and hence has to wait for the traveler to complete their access walk trip to the adjusted PU link. The results clearly indicate that faster walking speeds drastically reduce the probability of a traveler's vehicle having to wait for them to arrive at the adjusted PU link. Similarly, Figure 5.6d shows that even when a vehicle waits for a traveler in the faster walking speed case, which is much less likely to happen, the vehicle waiting time is significantly lower than the case with slower walking speeds.

Figure 5.6c and Figure 5.6d collectively illustrate why faster walking speeds can improve operational efficiency and significantly reduce request-to-destination travel time—the faster walking speeds significantly reduce instances where vehicles wait for travelers, and in the case where vehicles do wait for travelers, the wait time is quite short. This combination reduces the amount of lost or unproductive vehicle time in the system. This could also be the reason why matching rate shown in Figure 5.5a is slightly higher for 20 kmph walk speed compared to 5 kmph walk speed, especially for low fleet sizes.

As mentioned in the Wang et al. (2022) review article, the issue of vehicles waiting for travelers at PU locations in C2C systems, is an overlooked issued. It is also an issue that does not or should not arise when travel times in the simulation environment are deterministic. However, travel times in congested real-world networks are not deterministic, they are uncertain. The POLARIS model used in this study enables us to capture this critical feature of C2C systems that is less relevant in D2D systems. I am concerned that other road network simulation models that do not capture travel time uncertainty may overestimate the benefits of C2C services. Similarly, as highlighted in Section 4.4.3.1, a good estimate of the effective pick up walk range (D_{walk}^{PU}) value is required to reduce instances where vehicles have to wait for the request to arrive at the adjusted PU link. Using the network routed travel times would give a more accurate value for vehicle speed s_v in Equation 17 to estimate D_{walk}^{PU}, however it is computationally more intensive to perform this for each (r, v) pair compared to using a heuristic. Overestimating s_v would restrict the value of D_{walk}^{PU} thereby curtailing performance gains that could have been attained with a longer PU walk range. On the other hand, underestimating the value of s_v would lead to overestimating D_{walk}^{PU} thereby resulting in vehicles arriving early at the

adjusted PU link and waiting for the traveler to arrive. Vehicles having to wait too long at the curbside to pick up a traveler may also impact curb space utilization as well as congestion on the adjacent links. This study could be extended in the future to evaluate such significant externalities.

5.3.4 What about the sequence of R-V Matching and PUDO Links Adjustment?

This section compares two different algorithmic approaches (i.e., decision policies) for solving the C2C-RP problem. This is controlled by the γ_{C2C} parameter which determines whether PUDO Links adjustment is performed either after or before Request-Vehicle matching. In the baseline approach—labelled C2C-RP-A where the 'A' stands for 'after'—we assign travelers to PUDO links *after* we assign them to vehicles (For each (r, v) match). In the alternative approach—labelled C2C-RP-B where the 'B' stands for 'before'—we assign travelers to PUDO links for each (r, v) match candidate *before* we assign them finally to vehicles. The C2C-RP-B approach is computationally more intensive than the C2C-RP-A approach. This is because, the PUDO links adjustment procedure described in Section 4.4.3 is repeated for every feasible (r, v) candidate for each request in C2C-RP-B, whereas the procedure is performed only for optimal (r, v) match pairs in C2C-RP-A. Since the effect of PUDO Links adjustment is also factored into the insertion cost used in the optimal Request-Vehicle matching module (Section 4.4.4) for the C2C-RP-B strategy, it should be pareto-improving in terms of operator and user costs. However, this depends on the accuracy of changes in insertion cost caused due to PUDO links so as to lead to a better optimal (r, v) match after PUDO Links adjustment.

Figure 5.7a shows that there is basically no difference between the two algorithmic approaches in terms of matching rate. At low fleet sizes, C2C-RP-A does match slightly more travelers to vehicles. Figure 5.7b shows that C2C-RP-B produces significantly lower VKT per request than C2C-RP-A across all fleet sizes. This is because the optimal Request-Vehicle matching module is able to make better optimal (r, v) pair matches since the best PUDO links have been chosen for each (r, v) candidate. The gap between PUDO links adjustment before match, and PUDO links adjustment after match is not significant for the Ride-hailing case in Figure 5.7b. Figure 5.7c shows that at low fleet sizes, C2C-RP-B slightly increases request-to-destination time relative to C2C-RP-A. This change seems to mostly arise from slightly higher access and egress walk times (or distances) for C2C-RP-B strategy compared to C2C-RP-A. This is because the current formulation of the optimal matching objective function (Equation 19) does not include PUDO walk distances. Hence the optimal minimum insertion cost matching solution may have slightly longer access and/or egress walk distances. For Ride-hailing, the gap between the two algorithmic policies is insignificant. Finally, Figure 5.7d shows that vehicle occupancy is higher for C2C-RP-B than C2C-RP-A, which could also be explained as C2C-RP-B factoring in PUDO links adjustment to make better optimal matching choices.

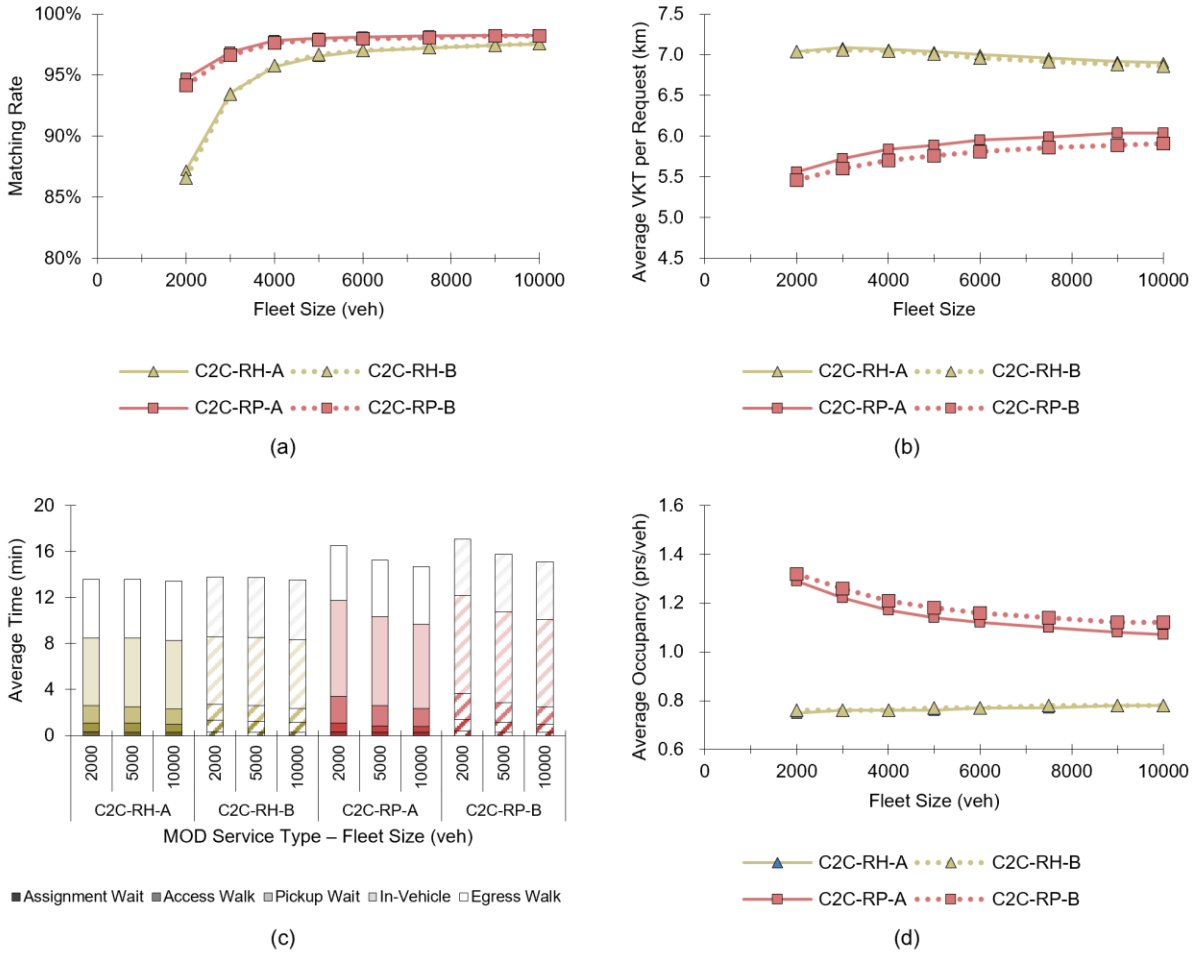

Figure 5.7. PUDO Links Adjustment after/before Traveler-Vehicle Matching: (a) Matching Rate, (b) Average VKT per Request, (c) Average Travel Time broken down by Segment, and (d) Average Vehicle Occupancy

5.3.5 Computational Time Results

This section aims to illustrate the scalability of the proposed decision policy and algorithmic approaches for the C2C-RP problem. All scenarios in this section were run on University of California Irvine's High Performance Computing Cluster (HPC3). The system resources allocated to run the simulation on the cluster are 32 logical processors and 128 GB of RAM. Figure 5.8 shows that for up to 10,000 vehicles serving up to 220,000 requests over the course of a day, in a detailed mesoscopic transportation system simulation model of Bloomington IL, model run times do not exceed 40 minutes. Additional test runs in the

Chicago network with similar fleet sizes and traveler requests indicate that model run time does not explode in a bigger network.

Figure 5.8 further shows how computational run time changes with respect to several key parameters. As expected, larger fleet sizes require longer run times than shorter fleet sizes. Similarly, longer walk ranges require longer run times. The reason for this finding is that longer walk ranges increase the number of feasible PUDO links, thereby increasing the size of the PUDO links selection problem. In contrast, Figure 5.8 shows that run time increases slightly with higher walk speed. This can be attributed to the increase in effective PU walk range value D_{walk}^{PU} with a higher value of s_w which increases the search space of feasible PU link candidates.

Finally, Figure 5.8 shows that the pareto-improvement in operator and user costs when using the adjust PUDO links before match decision policy (C2C-RP-B), does come at the expense of computational time (i.e., a third 'cost'). The computational run time is significantly larger for C2C-RP-B than C2C-RP-A, since PUDO links evaluation is performed for each (r, v) candidate instead of just (r, v) match pairs (Explained in Figure 5.9). Hence, in practice, fleet operators can choose between C2C-RP-A and C2C-RP-A depending on the computational power they have available and the size of individual problem instances. For the largest problem instances (10,000 vehicles, 1000 meters walk range and 5 km/h walk speed), C2C-RP-A takes nearly 20% more computation time compared to D2D-RP (26 minutes vs 21 minutes). C2C-RP-B on the other hand takes nearly 70% more computation time compared to D2D-RP (36 minutes vs. 21 minutes).

Figure 5.9 displays a breakdown of computational run time in each submodule of the C2C-RP solution implemented in POLARIS, for the largest problem instance—10,000 vehicles, 5 km/h walk speed, 1,000-meter walk range. As the figure shows, much of the computational time is spent in the module for determining feasible request-vehicle match candidates ('R-V Candidate Prep Time, Section 4.4.2). For the adjust PUDOs before matching case, the time spent adjusting PUDOs for each request is also quite high. Compared to the adjust PUDO links after matching case, the former case takes eight times longer, because PUDO links are adjusted for all eight candidate vehicles associated with each request (k_{veh} parameter in Section 4.4.2), rather than just the single matched vehicle. The computational time will also increase if the number of candidate PU/DO links being evaluated (k_{PUDO}) is increased.

Fortunately, it is straightforward to parallelize the feasible request-vehicle search module as well as the PUDO links adjustment module in the code. The algorithm determines candidate vehicles for each request independently. Similarly, the algorithm adjusts PUDO links for each (r, v) match or match candidate independent of the other (r, v) matches/match candidates. The only C2C-RP module that cannot be parallelized is the request-vehicle matching optimization, but this module consumes very little computational resources compared to other stages as shown in Figure 5.9.

The 'other processing time' incorporates the rest of the modules in POLARIS including traffic simulation and vehicle pathfinding. Much of this run time is independent of the C2C-RP module.

Given that the runtime is relatively short to begin with, and the two MOD submodules that take the most time can be parallelized, it is clear that the proposed decision policy and algorithm strategy for the C2C-RP problem is highly scalable.

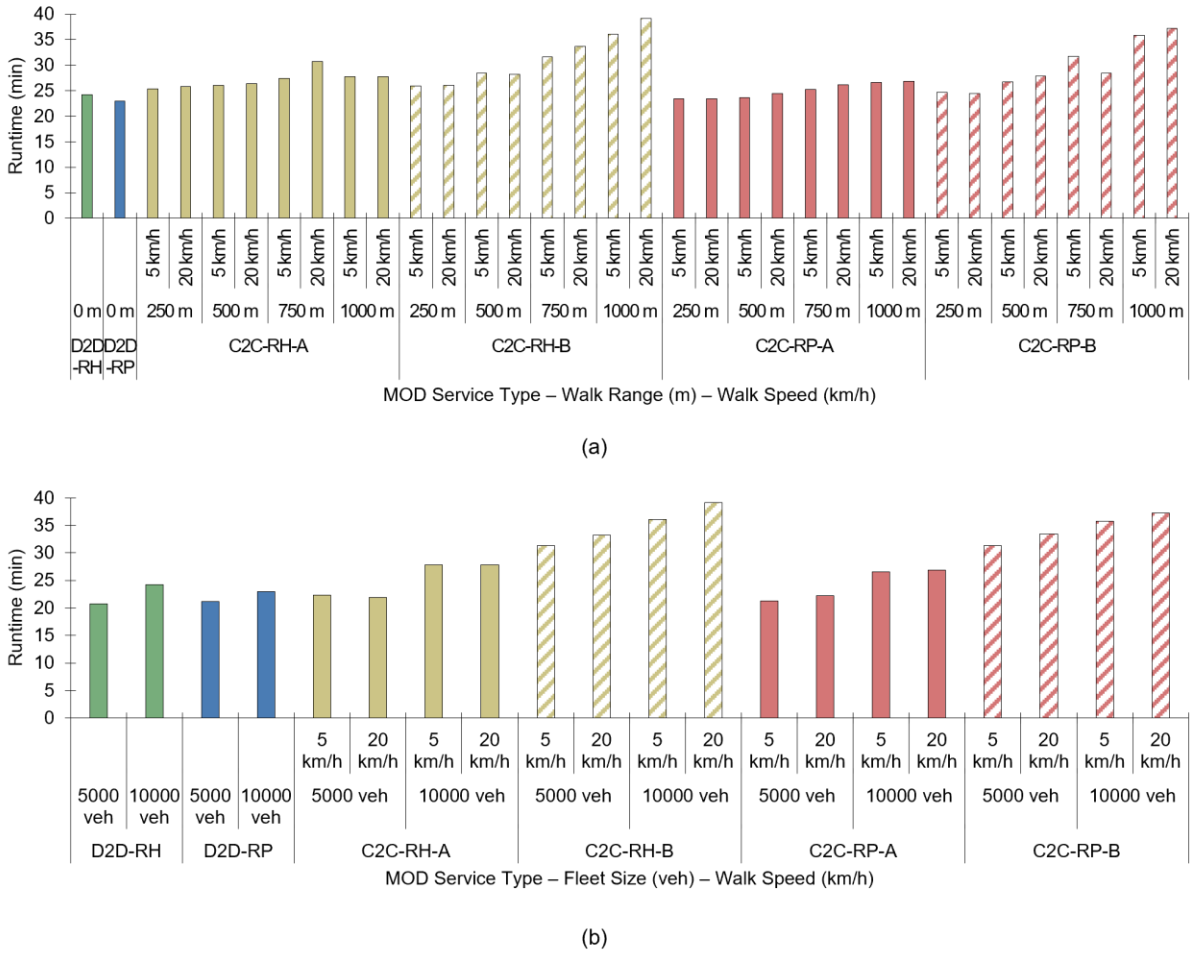

Figure 5.8. Computational Time by (a) Walk Range and (b) Fleet Size

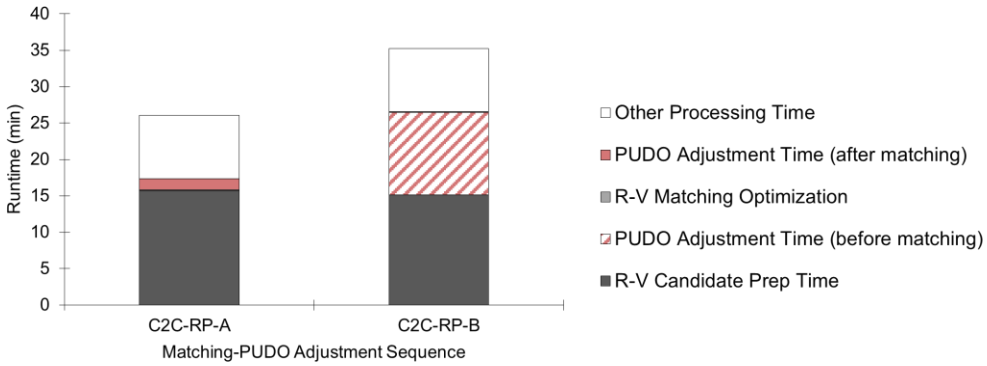

Figure 5.9. Computational Time by Matching-PUDO Links Adjustment Sequence (10,000 vehicles, 5 km/h walk speed, 1,000 m walk range)

5.4 Conclusion

MOD services serve a sizable portion of demand in many urban areas. The D2D-RH service still dominates the MOD market, which is problematic because Ride-hailing includes significant deadheading miles and low vehicle occupancies. Several MOD service variants offer many of the benefits of D2D-RH with significantly fewer vehicle miles and higher vehicle occupancies. These variants include D2D-RP, C2C-RH, and C2C-RP. One of the two goals of this study is to compare these four MOD service variants in terms of user costs and operator costs. The second goal of this study is to develop an effective and scalable decision policy and solution algorithm to solve the C2C-RP operational problem.

The C2C-RP operational problem is a complex highly dynamic sequential decision problem with a very large decision space. The C2C-RP operator needs to frequently assign new requests to vehicles and to PU and DO links. I propose a decision policy that decomposes the decision problem into two subproblems, the request-vehicle matching problem and the PUDO links adjustment problem. The decision policy utilizes the location and planned itineraries of every vehicle in the fleet, the status of each request in the system, forecasted

link travel times, and geospatial information to solve both subproblems in every decision epoch. The proposed decision policy dynamically assigns new requests to vehicles, sequences and re-sequences traveler PU and DO tasks, schedules and re-schedules traveler PU and DO tasks, and selects PU and DO locations for each request-vehicle match or candidate match. I also vary the order in which we iteratively solve the two subproblems.

I use the POLARIS agent-based transportation systems simulation model to test the proposed decision policy for the C2C-RP problem and to compare the four MOD service variants. The computational results illustrate that the proposed decision policy is both operationally effective and scalable. The most computationally demanding components of the decision policy and solution algorithm can easily be parallelized to further reduce run time.

The computational results also indicate that Ride-pooling provides significant fleet operational benefits over Ride-hailing services, in terms of VKT per request served, while only slightly increasing traveler request-to-destination times. The benefits for C2C services compared to D2D services are relatively smaller, with improvements in VKT per request coming at a large increase in traveler request-to-destination time. Moreover, combining Ride-pooling and C2C service appears to provide additive benefits, compared to D2D-RH.

Additional computational experiments indicate that increasing the maximum customer walking range does not provide significant additional benefits in terms of VKT per request. Rather it is allowing walking legs at all, that provides much of the operational benefit, as allowing some walking can prevent highly inefficient PU and DO locations for vehicles.

The next chapter (Chapter 6) concludes the dissertation, summarizing the key objectives and takeaways from the two studies as well as providing various directions in which the research presented in my dissertation could be carried forward.

Chapter 6. Conclusion

6.1 Summary

Cities around the world exhibit huge variations in their transportation network structures, travel demand patterns and travelers' willingness to share trips. The urban mobility sector also has several mobility modes that vary in capacity and flexibility of service. This dissertation presented two studies that comprehensively evaluated sharing and shared mobility services in urban transportation systems.

The first study discusses a novel method for quantifying the shareability of person-trips in a region based on travel demand, road network structure, and willingness to detour. The concept of 'flow overlaps' for individual person-trips is introduced and used to create a Quadratic Program called MNFLOP. This math program calculates optimal path flows between origin-destination pairs that maximize shareability. Various shareability metrics are derived from the MNFLOP results, calculated at different levels such as OD pairs, origins, links, and the entire network. The research applied this approach to the Sioux Falls network, demonstrating its ability to assess shareability under different demand scenarios. Computational results reveal that even short detours for some trips increase total flow overlap in the network. The study validates MNFLOP's efficiency through numerous experiments that show a meaningful link between shareability metrics and vehicle fleet miles for a microtransit last-mile service. The elasticity between vehicle fleet miles and the shareability metric is found to be -0.71. This indicates that a 1% increase in shareability leads to a 0.71% reduction in VMT implying that shared mobility services operate more efficiently in regions with high shareability. The results also show that the magnitude of demand in an

area alone does not capture the viability or efficiency of operating shared mobility services in the area. Location level metrics of shareability proposed in this study capture both the magnitude and dispersion/concentration of demand in a location.

The second study introduced a scalable algorithm for operating MOD services with flexible and dynamic PUDO locations, called Corner-to-Corner (C2C) services, within a congested network. A comparative analysis of four MOD service types—Door-to-Door (D2D) Ride-hailing, D2D Ride-pooling, C2C Ride-hailing, and C2C Ride-pooling —was conducted based on operator and user costs. The algorithm used a decomposition approach to tackle the increased complexity of the C2C-RP problem with optimal decisions to be made on request-vehicle matching, PUDO sequencing for each vehicle and alternative PUDO locations selection for each request-vehicle pair. The findings indicate that Ride-pooling leads to reduced operator costs while slightly increasing user costs, whereas C2C decreases operator costs but significantly elevates user costs. Combining Ride-pooling and C2C emerges as a promising strategy to mitigate operator costs and decrease vehicle miles traveled (VMT) within MOD systems.

6.2 Future Research

This subsection includes an extensive discussion of future research directions related to MNFLOP and Corner-to-Corner MOD services presented in my dissertation. I believe that MNFLOP and the shareability metrics presented in the first study of the dissertation have significant potential to provide insights into shared mobility system planning and design decisions, beyond the experiments in Chapter 3. Similarly, future research could also be conducted on the C2C-RP decision policy, implementation, its impacts on the transportation

system and also for infrastructure planning practice to designate optimal PUDO locations in dense city centers.

6.2.1 Shareability as Input for Shared Mobility System Design

The original impetus for the study on shareability in this dissertation was that the existing flexible transit design literature did not effectively characterize demand dispersion alongside demand density in a service region, nor did the existing literature consider the region's underlying transportation network. I first sought to determine the optimal SMM (i.e., fixed-route transit, flexible transit, ridesharing, or ridesourcing) or SMM combination for each subregion of a city given its TSAs. This study takes one step toward connecting TSAs and SMMs, but additional research is needed to fully connect subregions and the optimal SMM or SMM combination.

Considering the study's background, I believe the MNFLOP and associated shareability metrics can provide valuable insights for planning multi-modal shared mobility networks and even function as an input to planning and design models for multi-modal shared mobility networks. While researchers and practitioners understand the broad generalizations that (i) high-capacity, high-frequency fixed-route transit services work well in dense areas, and (ii) flexible, demand-adaptive door-to-door services work better in lower density areas, real-world urban areas do not fit neatly into these two extremes. Hence, deciding (i) where to operate specific SMMs ranging from heavy rail to ridesourcing and (ii) the optimal combination of SMMs in a city, are ongoing challenges that the MNFLOP and associated shareability metrics can help address.

In addition to helping assess the viability of specific SMMs, I believe the MNFLOP output, specifically the link flows in the network visualizations (e.g.,

Figure 3.2), can identify street segments and corridors wherein high-capacity, high-frequency transit lines are most viable. Similarly, for shared-ride mobility-on-demand services with Virtual Stops and walking legs, these network visualizations illustrate where service providers may want to cluster Virtual Stops for customer pickups and drop-offs.

Finally, the origin level analysis and overlap percentage metric proposed in this study capture the magnitude and the extent of directional overlap of demand starting from an origin. Having a high percentage of flow overlap from a node, notwithstanding a low magnitude, indicates that it is possible to operate shared modes even in low density areas when there is high spatial and temporal overlap of flows starting from the location. These origin-level shareability metrics should provide considerable value when planning first- and last-mile feeder services around transit stations, as indicated in Section 3.3.2 and validated in Section 3.4. Future research involves developing a robust empirical model for microtransit service performance metrics (e.g., VMT, AVO, etc.) as a function of the shareability metrics and other relevant factors.

6.2.2 Shareability Metrics as Predictors of Transit Ridership

Transit ridership and transit mode share vary across cities. Moreover, demand for transit varies across spatial areas within cities. Similarly, the demand for ridesourcing and shared-ride mobility services vary across and within cities. Transit operators, mobility service providers, and transportation planners are quite interested in understanding the factors that cause these variations in transit usage across and within cities. The authors

believe that the shareability metrics presented in this study can be used, along with various other factors, to explain and forecast the usage of SMMs like public transit and shared-ride services. After controlling for sociodemographic attributes, a proxy for willingness to use shared-ride options, as well as transit expenditures and other factors, cities with higher shareability are likely to have higher transit ridership. Future research can test this hypothesis.

6.2.3 Future Research on C2C-RP Services

The C2C-RP problem studied in my dissertation is one of only a few studies that address the problem at scale, in a congestible network and in an Agent Based Transportation model. Hence, there are several remaining areas for future research. First, in the proposed decision policy and solution algorithm, I chose to avoid making extensive calls to a pathfinding algorithm to determine high-quality estimates of travel times between vehicles and requests. However, future research should consider the added operational effectiveness benefits and computational costs of employing a pathfinding algorithm. This could also increase productivity in the system by reducing instances where the vehicle has to wait too long at the Pickup location for the traveler to walk to caused due to poor estimates of travel time. Second, in the current study, I do not consider vehicles blocking lanes of traffic or utilizing curb space when picking up and dropping off travelers. Future research should consider this possibility, as curb space utilization and double-parking violations especially in dense city centers are a hot topic among transportation planners, engineers, and policymakers. Third, future research should consider incorporating predictive models of supply and demand into the fleet operator's decision policy at each decision epoch. Fourth, while the current study performs extensive computational experiments in Bloomington, IL,

future research should compare the four MOD services and evaluate the proposed decision policies effectiveness in different regions/networks. Preliminary results for Chicago suggest that the benefits of C2C-RP relative to D2D services are greater in Chicago, IL than Bloomington IL.

6.2.4 MNFLOP Path Flows for C2C Service Planning

The visualization of network flow overlap maximizing flows for various scenarios shown in Chapter 3 indicate that MNFLOP concentrates demand flows between various OD pairs onto fewer links to create corridors of overlapping flows within a given detour threshold for flows between each OD pair. These results provide interesting implications and possibilities to advance the research on C2C services. As noted in Chapter 1, there exist several variants of MOD services with varying flexibility in terms of route, schedule and PUDO locations. MNFLOP path flow results could be used to analyze travel patterns in a region to find optimal corridors to run Route-Deviation transit (Section 1.1.1) to serve passengers who are within a short detour from a maximum overlap corridor. Similarly, MNFLOP path flow results could be used to determine alternative PUDO locations or meeting points for travelers (nodes or intersections in the network from/to where overlapping flows merge/diverge).

6.2.5 Alternative formulations and heuristic solutions for MNFLOP

Using MNFLOP based methods to plan C2C or any form of MOD services requires the development of heuristics to solve MNFLOP in large networks. Another focus area of future research could be on creating different methods, including exact, approximate, and heuristic algorithms, to address the problem for larger networks. This involves exploring alternative

formulations to represent the original model. Other potential strategies to reduce computational complexity include shrinking the problem by grouping similar trip origins and destinations or limiting the considered OD pairs that create overlapping flows when assessing overlap for a particular OD pair. Genetic algorithms offer promise as a heuristic solution approach due to their ability to navigate the solution space with efficiency. Additionally, customized heuristics that quickly estimate the objective function rather than fully analyzing it during sub-iterations could prove to be successful.

REFERENCES

Aissat, K., Oulamara, A., 2014. A Priori Approach Of Real-Time Ridesharing Problem With Intermediate Meeting Locations. J. Artif. Intell. Soft Comput. Res. 4, 287–299. https://doi.org/10.1515/jaiscr-2015-0015

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D., 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proc. Natl. Acad. Sci. U. S. A. 114, 462–467. https://doi.org/10.1073/pnas.1611675114

Araldo, A., Maria, A. Di, Stefano, A. Di, Morana, G., 2019. On the Importance of demand Consolidation in Mobility on Demand, in: 2019 IEEE/ACM 23rd International Symposium on Distributed Simulation and Real Time Applications (DS-RT). IEEE, pp. 1–8. https://doi.org/10.1109/DS-RT47707.2019.8958669

Auld, J., Hope, M., Ley, H., Sokolov, V., Xu, B., Zhang, K., 2016. POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. Transp. Res. Part C Emerg. Technol. 64, 101–116. https://doi.org/10.1016/j.trc.2015.07.017

Balardino, A.F., Santos, A.G., 2016. Heuristic and Exact Approach for the Close Enough Ridematching Problem. pp. 281–293. https://doi.org/10.1007/978-3-319-27221-4_24

Cervero, R., Seskin, S., 1995. An Evaluation of the Relationships Between Transit and Urban Form, Transit Cooperative Research Program.

Charikar, M., Chekuri, C., Cheung, T. yat, Dai, Z., Goel, A., Guha, S., Li, M., 1998. Approximation

algorithms for directed Steiner problems, in: Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 192–200.

Chen, P.W., Nie, Y.M., 2018. Optimal design of demand adaptive paired-line hybrid transit : Case of radial route structure. Transp. Res. Part E 110, 71–89. https://doi.org/10.1016/j.tre.2017.12.006

Chen, P.W., Nie, Y.M., 2017a. Connecting e-hailing to mass transit platform : Analysis of relative spatial position q. Transp. Res. Part C 77, 444–461. https://doi.org/10.1016/j.trc.2017.02.013

Chen, P.W., Nie, Y.M., 2017b. Analysis of an idealized system of demand adaptive paired- line hybrid transit 102, 38–54. https://doi.org/10.1016/j.trb.2017.05.004

Cici, B., Markopoulou, A., Frías-Martínez, E., Laoutaris, N., 2013. Quantifying the potential of ride-sharing using call description records. ACM HotMobile 2013 14th Work. Mob. Comput. Syst. Appl. https://doi.org/10.1145/2444776.2444799

Costa, A.M., Cordeau, J.F., Laporte, G., 2009. Models and branch-and-cut algorithms for the Steiner tree problem with revenues, budget and hop constraints. Networks 53, 141–159. https://doi.org/10.1002/net.20274

Curtis, S.K., Lehner, M., 2019. Defining the Sharing Economy for Sustainability. Sustainability 11, 567. https://doi.org/10.3390/su11030567

Czioska, P., Kutadinata, R., Trifunović, A., Winter, S., Sester, M., Friedrich, B., 2019. Real-world meeting points for shared demand-responsive transportation systems. Public Transp.

11, 341–377. https://doi.org/10.1007/s12469-019-00207-y

Czioska, P., Mattfeld, D.C., Sester, M., 2017. GIS-based identification and assessment of suitable meeting point locations for ride-sharing. Transp. Res. Procedia 22, 314–324. https://doi.org/10.1016/j.trpro.2017.03.038

Dill, J., Schlossberg, M., Ma, L., Meyer, C., 2013. Predicting Transit Ridership at the Stop Level: The Role of Service and Urban Form. Annu. Meet. Transp. Res. Board. https://doi.org/10.1016/j.sbspro.2013.11.183

Ewing, R., Cervero, R., 2010. Travel and the built environment. J. Am. Plan. Assoc. 76, 265–294. https://doi.org/10.1080/01944361003766766

Fielbaum, A., Bai, X., Alonso-Mora, J., 2021. On-demand ridesharing with optimized pick-up and drop-off walking locations. Transp. Res. Part C Emerg. Technol. 126, 103061. https://doi.org/10.1016/j.trc.2021.103061

Fu, L., n.d. Planning and Design of Flex-Route Transit Services 59–66.

Graehler, M., Mucci, R.A., Erhardt, G.D., 2019. Understanding the Recent Transit Ridership Decline in Major US Cities: Service Cuts or Emerging Modes?, in: 98th Annual Meeting of the Transportation Research Board. pp. 1–19.

Guerra, E., Caudillo, C., Monkkonen, P., Montejano, J., 2018. Urban form, transit supply, and travel behavior in Latin America: Evidence from Mexico's 100 largest urban areas. Transp. Policy 69, 98–105. https://doi.org/10.1016/j.tranpol.2018.06.001

Gurumurthy, K.M., de Souza, F., Enam, A., Auld, J., 2020. Integrating Supply and Demand

Perspectives for a Large-Scale Simulation of Shared Autonomous Vehicles. Transp. Res. Rec. J. Transp. Res. Board 2674, 181–192. https://doi.org/10.1177/0361198120921157

Gurumurthy, K.M., Kockelman, K.M., 2022. Dynamic ride-sharing impacts of greater trip demand and aggregation at stops in shared autonomous vehicle systems. Transp. Res. Part A Policy Pract. 160, 114–125. https://doi.org/10.1016/J.TRA.2022.03.032

Gurumurthy, K.M., Kockelman, K.M., 2018. Analyzing the dynamic ride-sharing potential for shared autonomous vehicle fleets using cellphone data from Orlando, Florida. Comput. Environ. Urban Syst. 71, 177–185. https://doi.org/10.1016/j.compenvurbsys.2018.05.008

Harmann, D., Yilmaz-Niewerth, S., Jacob, C., 2022. Methodological Distribution of Virtual Stops for Ride-pooling, in: Transportation Research Procedia. pp. 442–449. https://doi.org/10.1016/j.trpro.2022.02.055

Heinrichs, H., 2013. Sharing economy: A potential new pathway to sustainability. GAIA. https://doi.org/10.14512/gaia.22.4.5

Hsieh, M.-I., Tsai, M.-F., Hsiao-Kuang, E., 2006. FasterDSP: A Faster Approximation Algorithm for Directed Steiner Tree Problem. FasterDSP: A Faster Approximation Algorithm for Directed Steiner Tree Problem *. Artic. J. Inf. Sci. Eng. 22, 1409–1425.

Hyland, M., Mahmassani, H.S., 2020. Operational benefits and challenges of shared-ride automated mobility-on-demand services. Transp. Res. Part A Policy Pract. 134, 251–

270. https://doi.org/10.1016/J.TRA.2020.02.017

Kallehauge, B., Larsen, J., Madsen, O.B.G., Solomon, M.M., 2005. Vehicle routing problem with time windows, in: Column Generation. Springer US, pp. 67–98. https://doi.org/10.1007/0-387-25486-2_3

König, A., Grippenkoven, J., 2020. Travellers' willingness to share rides in autonomous mobility on demand systems depending on travel distance and detour. Travel Behav. Soc. 21, 188–202. https://doi.org/10.1016/J.TBS.2020.06.010

Kruskal, J.B., 1956. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. Source Proc. Am. Math. Soc. 7, 48–50.

Kucharski, R., Cats, O., 2020. Exact matching of attractive shared rides (ExMAS) for system-wide strategic evaluations. Transp. Res. Part B Methodol. 139, 285–310. https://doi.org/10.1016/J.TRB.2020.06.006

Li, N., Kong, L., Shu, W., Wu, M.-Y., 2020. Benefits of Short-Distance Walking and Fast-Route Scheduling in Public Vehicle Service. IEEE Trans. Intell. Transp. Syst. 21, 3706–3717. https://doi.org/10.1109/TITS.2019.2931798

Li, X., Quadrifoglio, L., 2010a. Feeder transit services: Choosing between fixed and demand responsive policy. Transp. Res. Part C Emerg. Technol. 18, 770–780. https://doi.org/10.1016/j.trc.2009.05.015

Li, X., Quadrifoglio, L., 2010b. Feeder transit services : Choosing between fixed and demand responsive policy. Transp. Res. Part C 18, 770–780.

https://doi.org/10.1016/j.trc.2009.05.015

Litman, T., 2013. The new transportation planning paradigm. ITE J. (Institute Transp. Eng. 83, 20–28.

Lobel, I., Martin, S., 2020. Detours in Shared Rides. SSRN Electron. J. https://doi.org/10.2139/SSRN.3711072

Lotze, C., Marszal, P., Schröder, M., Timme, M., 2022. Dynamic stop pooling for flexible and sustainable ride sharing. New J. Phys. 24, 023034. https://doi.org/10.1088/1367-2630/ac47c9

Luo, S., Nie, Y. (Marco), 2020. Paired-line hybrid transit design considering spatial heterogeneity. Transp. Res. Part B Methodol. 132, 320–339. https://doi.org/10.1016/j.trb.2019.04.007

Lyu, Y., Lee, V.C.S., Ng, J.K.-Y., Lim, B.Y., Liu, K., Chen, C., 2019. Flexi-Sharing: A Flexible and Personalized Taxi-Sharing System. IEEE Trans. Veh. Technol. 68, 9399–9413. https://doi.org/10.1109/TVT.2019.2932869

Martin, S., Taylor, S.J., Yan, J., 2021. Trading flexibility for adoption: Dynamic versus static walking in ridesharing. SSRN Electron. J. https://doi.org/10.2139/ssrn.3984476

Martínez, L.M., Viegas, J.M., Eiró, T., 2015. Formulating a New Express Minibus Service Design Problem as a Clustering Problem. Transp. Sci. 49, 85–98. https://doi.org/10.1287/trsc.2013.0497

McIntosh, J., Trubka, R., Kenworthy, J., Newman, P., 2014. The role of urban form and transit

in city car dependence: Analysis of 26 global cities from 1960 to 2000. Transp. Res. Part D Transp. Environ. 33. https://doi.org/10.1016/j.trd.2014.08.013

Mehran, B., Yang, Y., Mishra, S., 2020. Analytical models for comparing operational costs of regular bus and semi-flexible transit services. Public Transp. 12. https://doi.org/10.1007/s12469-019-00222-z

Miklas-Kalczynska, M., Kalczynski, P., 2021. Self-organized carpools with meeting points. Int. J. Sustain. Transp. 15, 140–151. https://doi.org/10.1080/15568318.2019.1711468

Moovit, n.d. On-Demand Microtransit Glossary [WWW Document]. URL https://moovit.com/on-demand/on-demand-glossary/ (accessed 8.21.23).

Moss, A., Rabani, Y., 2007. Approximation algorithms for constrained node weighted steiner tree problems. SIAM J. Comput. 37, 460–481. https://doi.org/10.1137/S0097539702420474

Mourad, A., Puchinger, J., Chu, C., 2019. A survey of models and algorithms for optimizing shared mobility. Transp. Res. Part B Methodol. 123, 323–346. https://doi.org/10.1016/J.TRB.2019.02.003

Navjyoth Sarma, J.S., Nam, D., Hyland, M.F., Souza, F. de, Yang, D., Ghaffar, A., Verbas, I.O., 2020. Effective and Efficient Fleet Dispatching Strategies for Dynamically Matching AVs to Travelers in Large-scale Transportation Systems, in: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1–6. https://doi.org/10.1109/ITSC45102.2020.9294340

Nesetril, J., Nesetrilová, H., 2023. The origins of minimal spanning tree algorithms – Borůvka and Jarník, in: Optimization Stories. pp. 127–141. https://doi.org/10.4171/dms/6/17

Nourbakhsh, S.M., Ouyang, Y., 2012. A structured flexible transit system for low demand areas. Transp. Res. Part B 46, 204–216. https://doi.org/10.1016/j.trb.2011.07.014

Oke, J.B., Aboutaleb, Y.M., Akkinepally, A., Azevedo, C.L., Han, Y., Zegras, P.C., Ferreira, J., Ben-Akiva, M.E., 2019. A novel global urban typology framework for sustainable mobility futures. Environ. Res. Lett. 14, 095006. https://doi.org/10.1088/1748-9326/AB22C7

Pinto, H.K.R.F., Hyland, M.F., Mahmassani, H.S., Verbas, I.Ö., 2019. Joint design of multimodal transit networks and shared autonomous mobility fleets ☆. Transp. Res. Part C 1–19. https://doi.org/10.1016/j.trc.2019.06.010

Prim, R.C., 1957. Shortest Connection Networks And Some Generalizations. Bell Syst. Tech. J. 36, 1389–1401. https://doi.org/10.1002/J.1538-7305.1957.TB01515.X

Qiu, F., Shen, J., Zhang, X., An, C., 2015. Demi-flexible operating policies to promote the performance of public transit in low-demand areas. Transp. Res. Part A 80, 215–230. https://doi.org/10.1016/j.tra.2015.08.003

Quadrifoglio, L., Li, X., 2009. A methodology to derive the critical demand density for designing and operating feeder transit services. Transp. Res. Part B 43, 922–935. https://doi.org/10.1016/j.trb.2009.04.003

Ronald, N., Thompson, R., Haasz, J., Winter, S., 2013. Determining the viability of a demand-responsive transport system under varying demand scenarios, in: IWCTS 2013 - 6th

ACM SIGSPATIAL International Workshop on Computational Transportation Science. pp. 7–12. https://doi.org/10.1145/2533828.2533831

Saberi, M., Mahmassani, H.S., Brockmann, D., Hosseini, A., 2017. A complex network perspective for characterizing urban travel demand patterns: graph theoretical analysis of large-scale origin–destination demand networks. Transportation (Amst). 44, 1383–1402. https://doi.org/10.1007/s11116-016-9706-6

Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S.H., Ratti, C., 2014. Quantifying the benefits of vehicle pooling with shareability networks. Proc. Natl. Acad. Sci. U. S. A. 111, 13290–13294. https://doi.org/10.1073/pnas.1403657111

Schaller, B., 2018. The new automobility: Lyft, Uber and the future of American cities. Schaller Consult. 41.

Schieber, T.A., Carpi, L., Díaz-Guilera, A., Pardalos, P.M., Masoller, C., Ravetti, M.G., 2017. Quantification of network structural dissimilarities. Nat. Commun. 8, 1–10. https://doi.org/10.1038/ncomms13928

Simonetto, A., Monteil, J., Gambella, C., 2019. Real-time city-scale ridesharing via linear assignment problems. Transp. Res. Part C Emerg. Technol. 101, 208–232. https://doi.org/10.1016/j.trc.2019.01.019

Soza-Parra, J., Kucharski, R., Cats, O., 2022. The shareability potential of Ride-pooling under alternative spatial demand patterns. Transp. A Transp. Sci. https://doi.org/10.1080/23249935.2022.2140022

Srinivasan, S., 2002. Quantifying spatial characteristics of cities. Urban Stud. 39, 2005–2028. https://doi.org/10.1080/0042098022000011335

Stabler, B., 2019. GitHub - bstabler/TransportationNetworks: Transportation Networks for Research [WWW Document]. GitHub. URL https://github.com/bstabler/TransportationNetworks (accessed 9.9.20).

Stead, D., Marshall, S., 2001. The Relationships between Urban Form and Travel Patterns . An International Review and Evaluation.

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2018. Enhancing urban mobility: Integrating ride-sharing and public transit. Comput. Oper. Res. 90, 12–21. https://doi.org/10.1016/j.cor.2017.08.016

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2015. The benefits of meeting points in ride-sharing systems. Transp. Res. Part B Methodol. 82, 36–53. https://doi.org/10.1016/j.trb.2015.07.025

Tachet, R., Sagarra, O., Santi, P., Resta, G., Szell, M., Strogatz, S.H., Ratti, C., 2017. Scaling law of urban ride sharing. Sci. Rep. 7, 1–6. https://doi.org/10.1038/srep42868

Tong, L. (Carol), Zhou, L., Liu, J., Zhou, X., 2017. Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing. Transp. Res. Part C Emerg. Technol. 85. https://doi.org/10.1016/j.trc.2017.09.022

Trombin, M., Pinna, R., Musso, M., Magnaghi, E., De Marco, M., 2020. Mobility Management: From Traditional to People-Centric Approach in the Smart City. Stud. Syst. Decis. Control

242, 165–182. https://doi.org/10.1007/978-3-030-22773-9_11/TABLES/3

Tsai, Y.H., 2005. Quantifying urban form: Compactness versus “sprawl.” Urban Stud. 42. https://doi.org/10.1080/0042098042000309748

Tsao, H.J., Lin, D., Jacob, H., Berkeley, U.C., 1999. Spatial and Temporal Factors in Estimating the Potential of Ride-sharing for Demand Reduction. Optimization.

van Oudheusden, D., 1995. The Steiner tree problem. Eur. J. Oper. Res. 81, 221. https://doi.org/10.1016/0377-2217(95)90155-8

Wang, Z., Hyland, M.F., Bahk, Y., Sarma, N.J.S., 2022. On Optimizing Shared-ride Mobility Services with Walking Legs. arXiv.

Yahoo Finance, 2023. Sharing Economy Market Size 2023, Share | Growing Report [2028] [WWW Document]. URL https://finance.yahoo.com/news/sharing-economy-market-size-2023-141551534.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAACg7b6a9Tm5JTLmv7ovbiu1vir1acIrEP6HsNdSMx8Kau6Yo16IH88MN6_Rj4uLFwMQ56JeZ0sNr2YISfGll5uAlvYw9hk- (accessed 8.22.23).

Yang, X., Zhao, Z., Lu, S., 2016. Exploring spatial-temporal patterns of urban human mobility hotspots. Sustain. 8. https://doi.org/10.3390/su8070674

Yen, J.Y., 1971. Finding the K Shortest Loopless Paths in a Network, Source: Management Science.

Zardini, G., Lanzetti, N., Pavone, M., Frazzoli, E., 2022. Analysis and Control of Autonomous

Mobility-on-Demand Systems. https://doi.org/10.1146/annurev-control-042920-012811 5, 633–658. https://doi.org/10.1146/ANNUREV-CONTROL-042920-012811

Zelikovsky, A., 1997. A Series of Approximation Algorithms for the Acyclic Directed Steiner Tree Problem 1 18, 99–110.

Zhang, W., He, R., Xiao, Q., Ma, C., 2016. Research on Strategy Control of Taxi Carpooling Detour Route under Uncertain Environment. Discret. Dyn. Nat. Soc. 2016. https://doi.org/10.1155/2016/4702360

Zheng, Y., 2018. A Methodology for Choosing between Route Deviation and Point Deviation Policies for Flexible Transit Services 2018.

Zheng, Y., Li, W., Qiu, F., Wei, H., 2019. The benefits of introducing meeting points into flex-route transit services. Transp. Res. Part C Emerg. Technol. 106, 98–112. https://doi.org/10.1016/j.trc.2019.07.012

Zwick, F., Kuehnel, N., Moeckel, R., Axhausen, K.W., 2021. Agent-based simulation of city-wide autonomous ride-pooling and the impact on traffic noise. Transp. Res. Part D Transp. Environ. 90, 102673. https://doi.org/10.1016/j.trd.2020.102673

Appendix A: Least Cost Shared Path (LCSP) Subgraph

As mentioned in Section 2.2.5 while introducing the concept of maximizing flow overlap, we only allow detours onto paths/links between OD pairs that are part of the least cost shared path (LCSP) subgraph. The LCSP subgraph problem is formulated in Equations $A1$ to $A5$.

$$Min\,\theta = \sum_{a \in A} c_a\, y_a \qquad (A1)$$

s.t.

$$\sum_{k \in K_p} w_p^k \geq 1 \quad \forall p \in P \qquad (A2)$$

$$\sum_{p \in P} \sum_{k \in K_p} w_p^k \delta_a^{pk} \leq M_A \cdot y_a \quad \forall a \in A \qquad (A3)$$

$$w_p^k \in \{0,1\} \quad \forall p \in P, \forall\, k \in K_p \qquad (A4)$$

$$y_a \in \{0,1\} \quad \forall\, a \in A \qquad (A5)$$

Where:

A: set of links/arcs in the graph representing the road network, indexed by $a \in A$

P: set of OD node pairs that must be connected by the LCSP subgraph, indexed by $p \in P$

K_p: set of paths for OD pair $p \in P$ with detour distance less than θ_{max}, indexed by $k \in K_p$

y_a: binary variable equal to one if link $a \in A$ is included in the LCSP

w_p^k: binary variable equal to one if path $k \in K_p$ connects OD pair $p \in P$

c_a: cost or distance of link $a \in A$

δ_a^{pk}: link-path incidence matrix, with values equal to one if path $k \in K_p$ for OD pair $p \in P$ contains link $a \in A$

M_A: a large positive number

The objective, Equation $A1$, minimizes the sum of the cost (or distance) of the links in the LCSP subgraph. The first constraint, Equation $A2$, ensures that there is at least one path connected each OD pair $p \in P$. The second constraint, Equation $A3$, ensures that if a link $a \in A$ is included in any path $k \in K_p$ for any OD pair $p \in P$ then link $a \in A$ must be included in the LCSP subgraph, i.e., if $\sum_{p \in P} \sum_{k \in K_p} w_p^k \delta_a^{pk} \geq 1$, then $y_a = 1$. The last two constraints ensure the two decision variables, y_a and w_p^k take on binary values.

Appendix B: Additional MNFLOP Results for Sioux Falls Scenario 1

Table B-1 displays each OD pair, the demand for each OD pair, as well as the optimal path, trip overlap, overlap percentage, detour distance, and marginal overlap for three separate assignment methods described in Section 3.3.1—MNFLOP-50%, MNFLOP-25%, and SP.

Table B-1 Comparison of OD Level Overlap Metrics for Sioux Falls Scenario - 1

OD Pair	Demand (F_p)	MNFLOP-50%						MNFLOP-25%		SP	
		k	Optimal Path	Trip Overlap (Z_{od})	Overlap % ($Z_{od}^{\%}$)	D (Miles)	Marginal Overlap (M_{Z_p}) (Trips/Mile)	Trip Overlap (Z_p)	Overlap % ($Z_p^{\%}$)	Trip Overlap (Z_p)	Overlap % ($Z_p^{\%}$)
1-13	500	0	1-3-12-13	4244.5	66.3	0	-	3817.2	59.7	2780.8	43.5
1-20	300	4	1-3-12-13-24-21-22-20	3015.0	47.1	3	673.5	1279.0	20.0	994.5	15.5
1-21	100	0	1-3-12-13-24-21	3821.2	59.7	0	-	3210.1	50.2	2165.7	33.8
1-22	400	0	1-3-12-13-24-21-22	3619.0	56.6	0	-	2989.0	46.7	1999.0	31.2
1-23	300	0	1-3-12-13-24-23	3769.6	58.9	0	-	3281.4	51.3	2199.0	34.4
1-24	100	0	1-3-12-13-24	4125.7	64.5	0	-	3572.3	55.8	2439.0	38.1
2-13	300	1	2-6-5-4-3-12-13	2899.0	45.3	5	198.8	2610.8	40.8	1904.9	29.8
2-20	100	0	2-6-8-7-18-20	636.5	10.0	0	-	1130.2	17.7	1255.2	19.6
2-22	100	23	2-6-5-4-3-12-13-24-21-22	2886.1	45.1	10	185.9	2391.3	37.4	1027.6	16.1
3-13	100	0	3-12-13	5699.0	89.1	0	-	4799.0	75.0	3399.0	53.1
3-22	100	0	3-12-13-24-21-22	4099.0	64.1	0	-	3211.5	50.2	2074.0	32.4
3-23	100	0	3-12-13-24-23	4406.7	68.9	0	-	3645.2	57.0	2352.8	36.8
4-13	600	0	4-3-12-13	4971.7	77.7	0	-	4035.4	63.1	2671.7	41.8
4-20	300	17	4-3-12-13-24-21-22-20	3335.0	52.1	8	271.4	1316.6	20.6	1163.7	18.2
4-21	200	0	4-3-12-13-24-21	4265.7	66.7	0	-	3343.4	52.3	2099.0	32.8
4-22	400	4	4-3-12-13-24-21-22	4019.0	62.8	2	1590.6	3109.0	48.6	943.4	14.7
4-23	500	1	4-3-12-13-24-23	4240.2	66.3	3	1085.2	3422.5	53.5	1099.0	17.2
4-24	200	0	4-3-12-13-24	4659.0	72.8	0	-	3732.3	58.3	2359.0	36.9
5-13	200	0	5-4-3-12-13	4437.5	69.4	0	-	3537.5	55.3	2352.8	36.8
5-20	100	0	5-6-8-7-18-20	539.0	8.4	0	-	1412.3	22.1	1279.0	20.0
5-21	100	1	5-4-3-12-13-24-21	3989.0	62.3	1	3647.9	3089.0	48.3	277.9	4.3
5-22	200	6	5-4-3-12-13-24-21-22	3789.9	59.2	5	684.1	1184.0	18.5	299.0	4.7
5-23	100	1	5-4-3-12-13-24-23	3951.6	61.8	3	1005	3146.4	49.2	1036.5	16.2
6-13	200	0	6-5-4-3-12-13	3604.9	56.3	0	-	2799.0	43.7	1869.6	29.2
6-20	300	0	6-8-7-18-20	699.0	10.9	0	-	1599.0	25.0	1599.0	25.0
6-21	100	12	6-5-4-3-12-13-24-21	3474.0	54.3	7	338.4	1126.8	17.6	1104.9	17.3
6-22	200	0	6-8-7-18-20-22	542.8	8.5	0	-	1255.2	19.6	1192.8	18.6
6-23	100	4	6-5-4-3-12-13-24-23	3420.7	53.5	3	870.6	2668.6	41.7	889.0	13.9
6-24	100	1	6-5-4-3-12-13-24	3641.9	56.9	1	2687.9	2818.0	44.0	954.0	14.9

Appendix C

Direction and Detour Checks in Request-Vehicle Matching (Section 4.4.2)

The first step in finding feasible vehicle candidates for each new unassigned request r involves subjecting each vehicle v to an elimination process based on direction compatibility and vehicle detour constraints. The procedure is described as follows:

- If v is idle and $v \notin V_r$ then add v to V_r
- If v is not idle and has seats available, then add v to V_r if all the following conditions are also met:
    - Directionality check: Check if the angle between the vectors representing the average `future path of vehicle v and the Euclidean path from the origin to the destination of request r is within a threshold. The average future path of v is formed by creating a unit vector from the current vehicle location in the direction of the average coordinates of all future PUDO links that the vehicle currently plans to visit (to pick up or drop off passengers), scaled by the total Euclidean distance of the remaining tour.

$$\theta = \cos^{-1} \frac{\vec{r} \cdot \vec{v}}{||\vec{r}|| \; ||\vec{v}||} \qquad (C1)$$

Where θ is the angle between the vehicle's current average Euclidean path and the requests Euclidean path, $\vec{r}$ denotes the vector connecting the origin and destination of request r and $\vec{v}$ is the vector denoting the current average planned path of vehicle v. Vehicles that are on the final leg of their current tour (i.e., the total distance remaining for v to complete its current tour is less than a minimum distance, are exempt from the directionality check.

    - Vehicle path detour check: This is done to ensure that the candidate vehicle v does not detour beyond a certain threshold to pick up the new request. This is done based on the relative location of the request origin with respect to the vehicle's current average planned path, found by calculating the below parameter:

$$u = \frac{\vec{x} \cdot \vec{v}}{||\vec{v}||^2} \quad (C2)$$

where $\vec{x}$ denotes the vector that starts from the vehicle's current location and ends at request's origin, and $||\vec{v}||$ denotes the Euclidean length of the vehicle's current average path vector. The vehicle path detour check is done based on the u parameter as follows:

- If $u < 0$, this means that the vehicle v needs to travel in a direction opposite to its current planned path to PU the new request r. Vehicle v is skipped if

$$|u \cdot v_x| + |u \cdot v_y| > dist_{rev}^{max} \quad (C3)$$

where v_x, v_y denote the x and y components of $\vec{v}$ respectively, and $veh\_rev\_dist\_thresh$ denotes the maximum distance between the vehicle's current location and the projection of the request origin onto the upstream of vehicle path $\vec{v}$. If the above reverse distance threshold is met, the vehicle v is also subjected to the next detour constraint.

- If $u \leq 1$, then add v to V_r if

$$||\vec{x}|| + ||\vec{y}|| - ||\vec{v}|| \leq dist_{detour}^{max} \quad (C4)$$

Where $\vec{y}$ denotes the vector connecting the request origin and the end point of the average current vehicle path vector $\vec{v}$, and $||\vec{x}||$, $||\vec{y}||$ and $||\vec{v}||$ respectively denote

the Euclidean distance between current vehicle location and request origin, request origin and end point of $\vec{v}$ and vehicle's current average Euclidean path length.

If $u > 1$, vehicle v is added to the candidate vehicle list V_r for request r, since the request origin is in downstream of the vehicle's current path end.