REFINED CHOICE SET GENERATION AND THE INVESTIGATION OF MULTI-CRITERION TRANSIT ROUTE CHOICE BEHAVIOR

A Thesis

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Benjamin J. Tomhave

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Alireza Khani

December, 2019

Acknowledgements

This thesis represents the culmination of my studies and research as a student of the Civil Engineering program. Without the assistance and support of countless individuals, I would not be where I am today.

I would first like to offer my profuse thanks to my advisor and mentor, Dr. Alireza Khani. Through countless meetings and discussions, Dr. Khani has provided the best possible learning environment one could ask for–providing feedback, guiding directions, and support all along the way. Without his support and immeasurable knowledge on all things transit, this research would not have been possible. Second, my thanks to the other members of my thesis committee, Dr. Yingling Fan and Dr. Gary Davis, for their suggestions and input on this research. I would also like to thank my fellow peers in the Transit Lab for their insight and comradery. Throughout my time in the program, your expertise and input has helped me to succeed both in the classroom and in my research. Last, but certainly not least, I would like to extend the utmost thanks and gratitude to my family for their unparalleled support and enthusiasm for the research I have pursued as well as my general exuberance for transit. Your encouragement has meant the world to me!

Abstract

Transit route choice models play a crucial role in determining how passengers interact with the transportation system. The resulting route choice parameters are used to calibrate demand forecasting models to determine how system alterations and modifications affect transit ridership on a route-level basis. Despite the importance of route choice calibration, no known model is available that is more recent than 2004. In order to understand current passengers' interaction with the modern-day transit system, a new method for transit route choice estimation is proposed in which a forward label-setting schedule-based multi-criterion shortest path algorithm is combined with an iterative trip elimination methodology. This new methodology yields high quality transit path choice sets with detailed temporal information on all types of network links (in-vehicle, walking, and waiting). This increased specificity, in turn, heightens the validity and accuracy of the route choice model. Passenger information is sampled from a transit on-board survey containing origin-destination locations, demographic details, and trip-specific attributes. A multinomial logit model with stop-level path size correction term is estimated yielding a 67% match rate between the path with the highest estimated likelihood and the surveyed (taken) transit path. Furthermore, a transfer penalty of 28.8 minutes was estimated and coefficients' marginal rates of substitution are in close alignment to similar values in the literature for both walking and waiting time. Express routes were found to have a statistically significant negative impact on path utility for the lowest income thresholds while transitways (light rail, bus rapid transit, or commuter rail) had a positive associated perception for the highest household income class. Thus, support is found for the claim that transitways can potentially attract higher-income "choice" riders to the transit network. The merits and potential future applications of the new route choice model are analyzed through a case study investigating the impact of the A Line arterial bus rapid transit route on surrounding system ridership. The results of this research can be used to improve ridership projections and highlight areas for policy improvements that could have the largest impact on retaining and attracting new passengers to the transit system.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Each day, the average adult brain makes approximately 35,000 conscious choices, just over one decision every 3 seconds, according to researchers at the University of Cambridge [1]. Of these numerous choices, a significant number are devoted to the realm of transportation. For example, an individual has to decide, what time am I going to leave for work today?, Should I drive or take transit?, Which route is quickest?, Do I prefer a quicker or more scenic route?, as well as many other factors. Understanding the decisions that influence individuals behavior is critical as such knowledge can be used to improve transportation systems of the present and future as individuals' preferences become known.

While relevant in all sectors of the transportation industry, public transportation stands to gain the most from understanding passengers' decision making processes. This is particularly true given the number of choices users face in a public transportation (transit) system as opposed to a highway or car-oriented form of transportation. For auto driven modes, users have a limited set of choices between continuing straight, turning left, or turning right at an intersection. In the case of transit, however, users must initially choose which stop to begin their trip at and which route (of the available routes served by the stop) to initially board. After this decision is made, the individual must choose whether to stay on the transit route or transfer at every subsequent stop along the route. If they choose to get off, the individual is then faced with deciding whether to transfer to another route or whether to walk to the destination. As illustrated in this vastly simplified overview, decision making within a transit system occurs on multiple levels and depends on each of the previous decisions that were made.

1.2 Problem Statement & Contribution

Despite the importance of understanding transit riders' decision making process, little attention has been devoted to this issue in the Twin Cities metropolitan region of Minnesota. In fact, the only mention of such a transit route choice study is from 2004 as part of the Metropolitan Council's work on developing a Mode Choice Model based on data from the year 2000 [2]. As a result, what little research has been done on the topic occurred with data from one decade before the first regional light rail line (Blue Line) and nearly two decades before the introduction of the new arterial Bus Rapid Transit (aBRT) line. Therefore, regardless of the quality of this previous work, it is exceedingly likely that it does not capture the current transit route choice behavior of the region as higher level of service modes (Light Rail and aBRT) will draw passengers to them and away from other routes, fundamentally changing the route choice decision making process individuals employ.

As part of a larger body of research dedicated to multimodal modeling within the Twin Cities metropolitan region of Minnesota, this research pursues the calibration of an up-to-date and more specific transit route choice model. Specifically, the relationship between transit passenger preferences and the way in which they interact with the transit network will be analyzed in detail. The key question this research asks is whether passengers select the shortest path (minimizes total travel time), or if individuals subjectively select a path based on their personal preferences towards path attributes other than the total travel time. Additionally, this work will investigate whether passengers' route choice options are restricted based on limited service coverage across the region, or whether individuals have a wide variety of "attractive paths" they would consider taking to get to their destination. Finally, this research will examine if socio-demographic variables influence transit route choice and if transitways (light rail, bus rapid transit ,and commuter rail routes) are more desirable than paths only containing local buses.

In order to address these questions, this report introduces a new multi-criterion schedule based shortest path (SBSP) algorithm with single trip elimination method which is used to generate a robust set of between 2-15 attractive path choices for each user. A multinomial logit model is then estimated to compare recorded transit paths from an on-board transit survey to the attractive paths contained within each user's choice set. In answering the questions posed above and creating a robust transit route choice model using a new and innovative method, this study serves as a benchmark and foundation in the efforts to improve ridership projections and further inform policy decisions throughout the regional transit network.

1.3 Thesis Organization

The thesis presented here is divided into five chapters. **Chapter Two** contains a literature review outlining the current state of discrete choice modelings. Following the literature review, **Chapter Three** outlines the variety of data sources employed in the analysis. Next, in **Chapter Four**, the choice set generation research methodology will be provided connecting the theoretical precedent described in Chapter Two with the regional data summarized in Chapter Three. **Chapter Five** contains an overview of the discrete choice logit modeling methodology. After the methodology, the main results of the route choice model calibration will be presented and analyzed in **Chapter Six** before being applied to a local case study in **Chapter Seven**. Following this analysis, the research and contributions to the field will be summarized with a conclusion in **Chapter Eight**.

Chapter 2

Literature Review

Route choice modeling is a critical step in transportation planning and management as it weighs the relative importance of personal and trip-specific attributes on determining a user's selection of a particular transit path from his/her origin to destination. Calibrated models can be used to project future ridership, the system impacts of adding or removing a particular route, and in determining what aspects of the trip should be enhanced to attract existing and potential riders to the transit system.

Transit route choice models can be dissected into 2 key questions and components both of which have large bodies of related research.

1. What paths (set of transit routes and walking/waiting links) do users consider as potential options to get from their origin to destination?
2. Which personal and trip attributes are most influential in guiding a user's decision making process and with what probability will a user choose any of his/her considered paths?

2.1 Origin-Destination Estimation

Prior to determining route choice behavior, individuals' origin and destination locations must be determined. Despite the appearance of this being a straight-forward problem, a dichotomy exists within the literature between using passenger surveys or automatic fare card (AFC) data when estimating transit passengers' origin and destination locations. The key distinction between these two methods is that surveys rely on information from the passenger (which tends to be more precise but expensive to administer) while locations from

AFC data have to be estimated (less precise but very inexpensive). On the whole, AFC methods can save millions of dollars in data collection and editing costs while benefiting from continuously updating data repositories.

2.1.1 AFC Methods

AFC data is collected each time a passenger taps his/her payment card when boarding (and if required, when alighting) a route. This method, therefore, contains station identification numbers and the very specific time the passenger entered (or left) the station. Note that this method does not always record the route that was used. As such, if more than one route is served by a single tap-in pay station, route level information is missed. Additionally, if individuals do not need to swipe/tap their card when transferring, intermediary transfer routes are also not recorded.

Due to the stop-specific nature of this data acquisition system, most studies focus on analyzing travel behavior that occurs between transit stops and ignore the behavior individuals exhibit when accessing their first transit stop from their origin or egressing from the destination stop to their destination location [3–8]. When preforming these type of studies, it is assumed that a high percentage of riders “return to the destination station of their previous trip when beginning their next trip [and] end the last trip of their day at the station where they began their first trip of the day” [7]. Using a trip diary repository, researchers found that these assumptions were valid for 90% of subway users. Two drawbacks of this method, however, are that a minimum of two trips a day are needed to derive exit stations and AFC data does not represent the entire population of transit riders (only 80% in New York City) [7].

While the AFC approach captures a large percentage of users’ entire daily travel patterns, this wide network-level analysis also restricts the potential utility of these types of studies. In addressing the time spent walking, waiting, and in-vehicle using AFC data, various methods are proposed in the literature. One assumption is that walking and waiting times are random while in-vehicle time (IVT) is punctual and determined by the tap-in and tap-out times from the AFC data [3]. Still other researchers [4], however, select origin-destination (O-D) pairs that only have one effective route between them. This allows them to eliminate the uncertainty in timings due to the previous uncertainty over which route was chosen (as described above). In this method, trip travel time is then split into entry and exit walking times, waiting times, and in-vehicle time. These entry and exit times, however, only focus on the time from tapping the fare card and walking through the station to the actual boarding/alighting platform. As a result, these times are only significant in large METRO systems and, like other previously mentioned studies, this method ignores the time to reach the individuals first station from their origin and the time from their last station to destination location.

Many AFC-based studies focus on the O-D locations of passengers, these locations, however, are often the O-D stops rather than the individuals' actual origin and destination locations. In order to address this deficiency, select researchers analyze access walking distance, and the associated travel behavior, using the AFC tap-in stop location as well as the fare card's billing address [6]. This method resulted in over 25% of the sample dataset having an access distance greater or equal to 2 miles, much larger than is traditionally observed. As a result, the authors tossed this data from their analysis rationalizing that with the AFC data set, "it was impossible to distinguish" if the billing address was actually where the individual walked from when accessing the stop contained within the AFC data.

As shown from this review of AFC-related transit behavior studies, the availability of AFC data makes it a very appealing dataset for this type of study but this methodology critically neglects access and egress walking components of individuals' trips.

2.1.2 Transit Survey Methods

As an alternative to AFC data, many studies rely on the use of high resolution surveys which may include information pertaining to the access/egress walking distances, demographic information, a list of all routes on an individuals trip, trip preference, and/or detailed trip information for the route on which the passenger was surveyed [9–11]. While AFC analysis is the emerging method to gather transit route choice information, survey based approaches have been traditionally employed.

Chapleau et al., in their workshop presentation at the 8th International Conference on Survey Methods in Transport, outline the key surveys from which transportation data is typically extracted [12]. Specifically, the authors denote two key survey areas: (1) household travel surveys, and (2) on-board transit surveys. Household travel surveys, however, are inadequate for application to transit planning due to the infrequent administration of household surveys, insufficient focus on transit trips, and an adequate spatial and temporal resolution for transit planning [12]. As a result, a predominance of the transit route choice models in the literature employ research focused on data from transit on-board surveys. In fact a survey of 52 transit agencies found that 96% conducted on-board surveys between 2002-2004 [13]. These surveys are typically carried out once every 1-4 years and focus on questions related to who transit users are, where and when they made their trip, and why they rode transit.

2.2 Choice Set Generation

Regardless of how the transit passenger data is collected, route choice studies must determine which transit paths (sequence of routes) individuals consider when ultimately choosing their desired path. This is, perhaps, one of the most crucial steps of route choice generation as incorrect size or composition of the choice sets can lead to model biases causing the calibrated route choice parameters to be inaccurate. While this is a straightforward conceptual question, the literature is divided on the best way to generate this choice set. Choice set generation can largely be split into two categories–*deterministic* (where a set number of choices is generated for a given O-D pair typically using an iterative shortest path search), or *stochastic* (where the choice set is specific to the individual rather than the O-D pair) [14].

In order to generate deterministic choice sets, several methods have been proposed. K-shortest Path Algorithms, for example, generate the first “k” number of shortest paths between a specified O-D pair [15, 16]. Within this method, iterative shortest paths can be generated using link penalty and/or link elimination methods in which the cost of links within the current shortest path are increased or the link is removed from the network [15]. A second deterministic choice set generation method uses branch and bound techniques [17, 18]. Using this technique, choice sets are found by creating an additional level to a nested tree structure for each choice an individual faces. The paths within this tree are terminated/removed if they violate any of the several imposed constraints (i.e. certain time thresholds are obeyed and a trip after a transfer cannot leave before the person has arrived at the transfer point). Third, Dial, in his 1971 algorithm, proposes the concept of “reasonable options” defined as paths in which when traveling from node to node one always gets further from the origin and closer to the destination [19]. Additional studies have since proposed other methods by which to define “reasonable” or “attractive” paths based on time constraints [20].

Within stochastic generation, Freijinger et al. (2009) propose an approach where the set is generated probabilistically by using a random walk biased towards the shortest path. Furthermore, Michael Scott Ramming, in his 2001 PhD thesis, introduces a method by which choice sets are simulated using link costs from different probability distributions [21].

2.3 Discrete Choice Modeling

Regardless of the implementation method, an individuals’ set of paths that he/she considers as “attractive” are simulated, and from this set, one path must be chosen as the simulated path taken by the individual. Within the literature, several primary model classes are introduced to perform this function.

The first class of models employ multi-nomial logit (MNL) models which compare potential path choices from the choice set and maximize the likelihood of choosing the path that the passenger actually traversed. Within the logit classification, several sub-methods are proposed such as C-Logit [22], and Path Size Logit [23] which address shortcomings of the traditional MNL model when dealing with overlapping paths. The latter two methodologies adjust the utility of an overlapping path based on a measure proportional to the size of the overlapping paths.

To further capture the correlation between alternatives cross-nested logit (CNL) models are mentioned by Peter Vovsha [24]. The format of this model allows for correlation between specific choices through grouping the alternatives based on a commonality while also acknowledging combinations and cross similarities between different nests. For example, in a traditional nested logit model nests could be created for transitways, local buses, and cars. In the cross-nested logit model, these distinct nests still exist but now, combinations of the nests are allowed such that an individual could take a local bus *and* a transitway. Due to the structure of cross-nested logit models and the inherent complexity of transit route choice modeling, the majority of studies employing CNL methods are focused on mode-choice [24,25] although select studies have also concerned route-choice modeling [26,27].

An additional class encompasses models derived from the Generalized Extreme Value (GEV) theorem of McFadden [28]. The fundamental difference between route choice models in this class and MNL models is that "the similarity among routes is captured in the structure of the error component of the utility function" [29].

Regardless of the specific model used, several factors were consistently found to be key determinants in passengers ultimate route choice behavior. In particular, Vande Walle and Steenberghen (2006) summarize the literature stating that transfers are generally perceived to have a penalty of between 5-20 minutes while time spent out of the transit vehicle is perceived as being 1.5-2.3 times higher than in-vehicle time.

After a survey of the existing literature, only a few select studies were found that employed a schedule based shortest path (SBSP) algorithm in generating a choice set over which a multinomial logit model would be estimated. Additionally, no studies used a stop-level path correction term nor combined both SBSP and trip elimination algorithms when generating choice sets as is presented in this research. As such, this analysis will further enrich the transit route choice modeling field by expanding upon, and providing an alternative to existing methodologies.

Chapter 3

Data

At the aggregate level, the data employed in this research can broadly be classified into two categories: passenger-specific information and network configuration data. The following section will briefly introduce both sources of data including how the data was acquired and its primary purpose within the transit route choice model.

3.1 Passenger-Specific Data

3.1.1 Metropolitan Council 2016 On Board Survey

Passenger attribute data is extracted from the Metropolitan Council's "2016 On Board Survey." The survey was conducted from April 2016–February 2017 and consists of origin-destination records for "30,605 transit trips across all regional routes and providers" [30]. These records were garnered through personal interviews using handheld tablets while the passenger was on board a transit route. The survey was 30 questions in length and provides sociodemographic characteristics of the user, precise origin and destination locations, as well as many other attributes. The individual-specific variables utilized in this research are summarized in Table 3.1.

In order to analyze the most typical and constant travel behavior patterns, several simplifications/reductions are made to the dataset: (1) the time frame for analysis is in the autumn after schools had resumed, (2) only individuals traveling on Tuesdays are considered, (3) the Tuesday following Labor Day and Halloween, the Tuesday prior to Thanksgiving Weekend, and the last two Tuesdays of December are removed from the study period to

Table 3.1: **2016 Transit on-board survey variables**

Variable (s) Name	Variable Description
ID	Unique Identifier For Each Passenger
Date	Date survey was conducted
Route_Surveyed	Route survey was conducted on
Transfers_From	Number of transfers from origin before being surveyed
Transfers_From_ (First, Second, Third, Fourth)	Route number for each transfer from the origin before survey
Origin_Place_Type, Destination_Place_Type	Type of place respondent is coming from (going to) now
Origin_Lat, Origin_Lon	Latitude, Longitude coordinates of nearest intersection to origin
Destination_Lat, Destination_Lon	Latitude, Longitude coordinates of nearest intersection from destination
Access_Mode, Egress_Mode	Mode of access to (egress from) transit
Transfers_To	Number of transfers taken after surveyed route to destination
Transfers_To_ (First, Second, Third, Fourth)	Route number for each transfer from the surveyed route to destination
Time_On	One hour range when the respondent boarded the route on which he/she was surveyed
Payment_Method	Payment method of the trip (Cash, Card, Pass, Mobile, etc.)
Fare_Type	Type of fare paid (Regular, Limited Mobility, Senior, Student/Youth)
Trip_Purpose	Purpose of making trip (Meal, Work, Recreation/Religious, School, Shopping, Errands, Other)
Age	Age of respondent
Race	Race of respondent
Income	Total annual household income of respondent
Linked_Weight_Factor	Estimated number of trips per day between given origin-destination

reduce the irregularities and potential impact from holiday travel. It is rationalized that Tuesdays have the highest chance of avoiding atypical travel behavior seen at the beginning and ends of the week, while fall is chosen in order to minimize the adverse effects of weather [31,32]. As such, the specific dates of analysis are September 13th, 20th, and 27th, October 4th, 11th, 18th, and 25th, November 8th, 15th, and 29th, as well as December 6th, and December 13th.

Following the selection of the analysis dates, the set of studied passengers are further filtered to only include individuals whose access and egress links are walked rather than made by car, bike, or other method. This step is taken in order to ensure that variables not specific to the transit network are controlled for and identical for all passengers involved.

After all refining measures are performed on the On Board Survey (OBS) data, 2,787 trip records (individual passengers) are left in the sample population. The majority of this population is white and between the ages of 18 and 34. To gain a greater understanding of the sample population, several figures and tables are displayed below. The racial spread seen within the sample population is representative of the entire survey but when compared to the greater Twin Cities region, it is much more diverse (Table 3.2). Furthermore, the annual household income for survey respondents is fairly evenly distributed with the majority of respondents reporting an annual household income of between $25,000-$100,000 as illustrated in Figure 3.1. Overall, the sample population is 46% female and 54% male.

Table 3.2: **Racial distribution comparison between sample population and greater the Twin Cities**

	Sample Survey Population	Twin Cities Metro Area
White	53 %	77%
Black	27%	10%
Asian	9%	7%
Latino/Hispanic	7%	6%
Native American	4%	1%
Pacific Islander	0%	0%

Note: Twin Cities data extracted from 2013-2017 ACS 5-Year Estimates

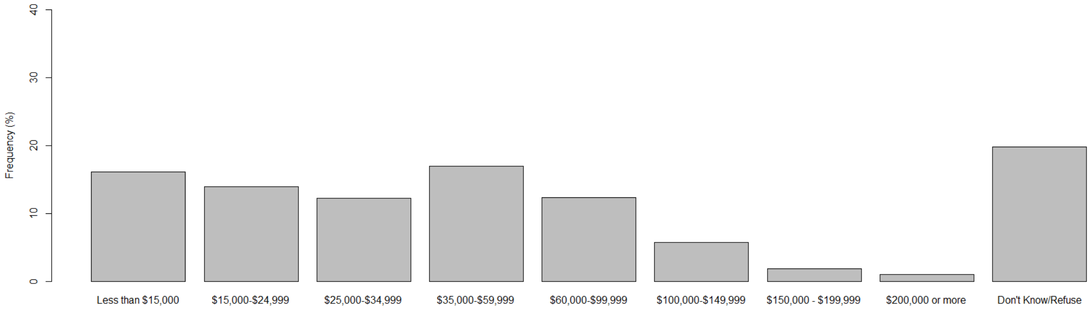


Figure 3.1: **Annual household income distribution of sample population**

Turning toward trip-based summary statistics of the sample population, the majority (54%) of surveyed transit trips are direct and involve no transfers, 37% have one transfer, and only 10% of the sample population takes more than 1 transfer. Passengers' trip purpose is fairly evenly distributed with only a slightly larger portion using the transit trip as a means of getting to work or a restaurant (Figure 3.2). Finally, when paying for their transit trip, passengers appear to have an affinity towards stored value "Go-To" Cards and cash while the University of Minnesota U-Pass is the most represented pass type (Figure 3.3).

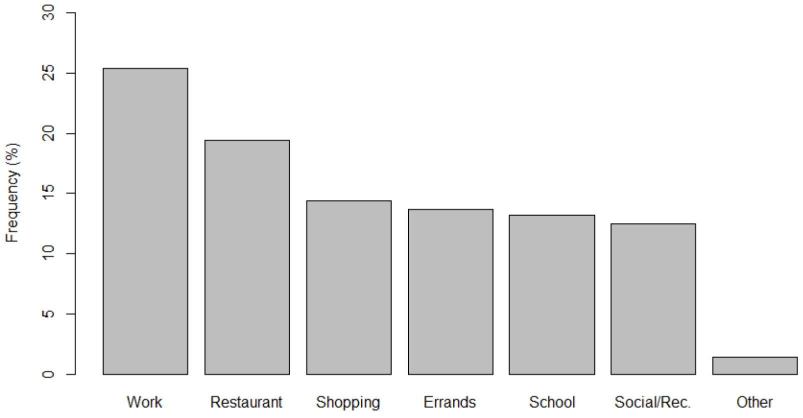


Figure 3.2: **Trip purpose distribution**

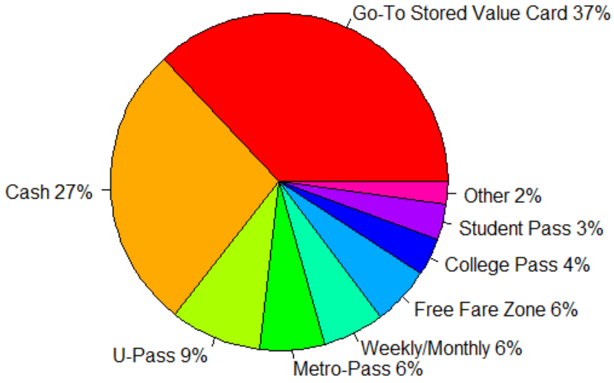


Figure 3.3: **Payment type distribution**

3.2 Network Configuration Data

3.2.1 GTFS Dataset

General Transit Feed Specification (GTFS) is a widely utilized data specification that standardizes the presentation of public transit data. GTFS data is provided in the form of several text files each containing information related to one attribute of the transit network: agency, stops, routes, trips, stop-times, and calendar (Table 3.3). Together, these files provide a detailed and complete representation of the transit network such that one can determine the precise scheduled arrival of a bus on any route at any given stop. The entire transit network is comprised of 190 routes (including 84 express routes, 2 light rail lines, 2 Bus Rapid Transit routes, and one heavy rail route) across 13,579 stops.

GTFS File	File Contents
agency.txt	Lists the agencies that provide service to the region. **Fields:** agency_id, agency_name, agency_url, ...
calendar.txt	Identifies a set of dates and the day(s) of the week a route with the given service_id is in service **Fields:** service_id, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday,Sunday, Start_Date, End_Date
calendar_dates.txt	Identifies a set of dates when a service exception occurs for one or more routes **Fields:** service_id, date, exception_type
routes.txt	Lists all transit routes in region **Fields:** route_id, agency_id, route_short_name, route_long_name, route_desc, route_type, ...
shapes.txt	Defines the visual path a vehicle travels. Consists of connecting a sequence of points **Fields:** shape_id, shape_pt_lat, shape_pt_lon, shape_pt_sequence, shape_dist_traveled,
stop_times.txt	Lists the arrival/departure time of each bus (trip) to a each stop **Fields:** trip_id, arrical_time, departure_time, stop_id, stop_sequence, ...
stops.txt	Lists all the transit stops in the network **Fields:** stop_id, stop_code, stop_name, stop_desc, stop_lat, stop_lon, ...
trips.txt	Lists all transit trips in the network **Fields:** route_id, service_id, trip_id, trip_headsign,direction_id, block_id,shape_id,wheelchair_accessible

Table 3.3: **Information contained within GTFS network**

3.2.2 OpenStreetMap Data

While travelers' in-vehicle movements are captured by the transit network, their out-of-vehicle movements occur along a walking network. For this study, a walking network is obtained from the open-source website OpenStreetMap. While the open-source nature of this database presents potential for errors, as anyone can edit and alter the map, the key benefit of this database is the provision of a highly specific sidewalk network. Using this sidewalk dataset, the Python package OSMnx developed by Geoff Boeing (2017) allows for the calculation of precise network walking distances (and times) between any two points [33].

Chapter 4

Choice Set Generation Methodology

To model and analyze transit users' route choice behavior, passengers must have a set of paths from which to choose between. When selecting a path to take, individuals are unlikely to consider every possible transit option connecting their origin and destination locations. Some paths may take an exceedingly long time, others may involve large amounts of waiting, and even more may have numerous transfers. Therefore, when making decisions, passengers, instead, are only likely to consider a select handful of potential transit paths. This small number of "attractive" paths are what comprise users' choice set.

The first step of this research, therefore, lies in generating a choice set of transit paths for each passenger. The construction of these sets, however, is not a trivial matter. In fact, choice set generation is arguably the most important, and most difficult component of formulating a route choice model.

4.1 Access/Egress Link Generation

When conceptualizing users' route choice, it is easy to become focused on their decisions only at the level of the transit route. While certainly a crucial element in passengers' choices, one must also consider their behavior with higher resolution at the level of the individual transit stop. The stops, as argued by Nassir et al. (2015), are fundamental to route choice as not all stops are served by the exact same route [34]. Therefore, while an individual may have a choice between 5 attractive routes at one stop, this choice set could be reduced to only one or two routes simply by the act of choosing a different stop. In order to maximize the quality or perceived attraction of the routes within an individuals choice set, the stops accessible to an individual must be carefully considered. If the transit stops accessible to a user are too narrowly defined, enticing path options may be missed. If, on the other hand, no restrictive assumptions are made to the stops an individual can access, his/her choice set will be much too large and lead to inaccurate model results. With this framework in mind, several principles and restrictions are implemented to restrict the number of accessible transit stops, thereby filtering the access and egress links used as an input to the choice set generation algorithm to be described in the next section.

4.1.1 Network vs. Euclidean Distance

The first restrictive measure implemented on the generation of access and egress links is the use of a network, rather than straight-line (euclidean), walking distance. Throughout the route choice literature [35, 36], euclidean and network distances are often used interchangeably as several studies have found a high degree of correlation between the two methods concluding that the "substitution of one [method] for the other is unlikely to have a substantial impact on analytic results" [37–40]. The primary drawback of euclidean methods, however, is that they are focused on roadway networks which are extremely dense within American metropolitan areas. When turning to less-dense transit networks, walking-distance scales, and regions with large amounts of lakes, rivers, and uncrossable highways, however, the literature notes that euclidean and network distances can no longer be used interchangeably [40, 41].

Following this review of the literature, the two methods are tested in the Twin Cities region where OpenStreetMap sidewalk data is used for the network walking distances. Due to the high prevalence of rivers and major roads/highways without frequent bridges or other opportunities to cross, euclidean distance (Fig. 4.1) is an inaccurate estimation of network distance (Fig. 4.2) in the Twin Cities as shown by the area reachable within 0.25 miles of the same origin point. Therefore, in order to heighten the walking distance and timing accuracy of the access/egress links, network distances are used in this research.

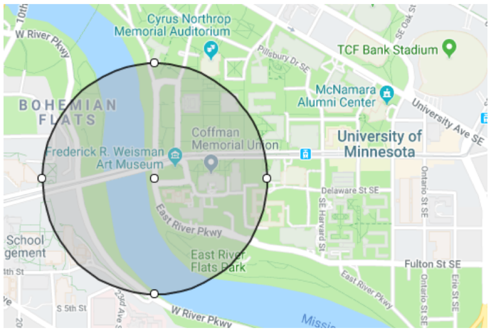


Figure 4.1: **Euclidean distance**

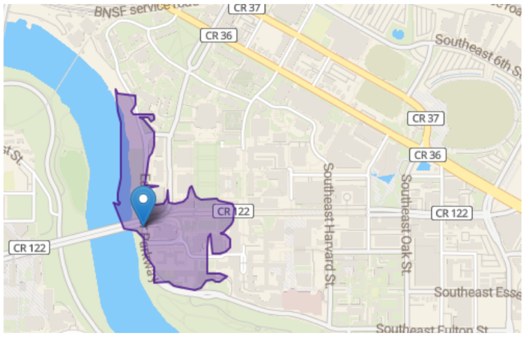


Figure 4.2: **Network distance**

4.1.2 Maximum Distance Thresholds

In addition to using network distances, further restrictions are imposed on the access and egress links. While the majority of the sample population walks less than 0.3 miles when accessing or egressing from a transit stop, some individuals walk over 1.5 miles. Given the decision to use the more computationally expensive network distance method, it is unreasonable to generate access and egress links for all stops that are within 1.5 miles of each individuals origin and destination location. Neither, however, is it acceptable to only include individuals with short access/egress walking links as one assumes that route choice behavior is fundamentally different for passengers who choose to walk long rather than short distances. As a result, the decision was made to select walking distance thresholds for both access and egress links that include 95% of the sample population. In this manner, computational time can be saved while still including the vast majority of the sample population. The resulting restrictions are thus determined to be 1.1 miles for access distances and 0.71 miles for egress links as shown in Table 4.1.

Table 4.1: **Sample population included by access/egress distance threshold**

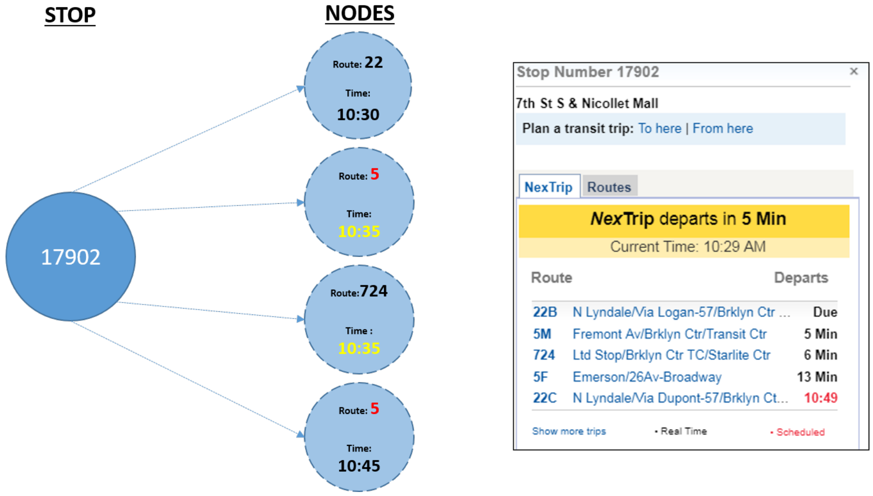
Percent Population Included	Access Distance (Miles)	Egress Distance (Miles)
50%	0.29	0.22
75%	0.49	0.40
85%	0.69	0.50
95%	1.10	0.71
100%	1.60	1.75

4.2 Schedule Based Shortest Path (SBSP)

The transit network is not a static entity in space due to the fact that service fluctuates throughout the course of the day. In order to capture the precise intricacies and timings of the various transit paths that are available to each passenger, a Schedule-Based Shortest Path (SBSP) algorithm adapted from the previous work of Khani et al (2015) [42] is implemented.

4.2.1 SBSP Network Typology

While other approaches such as frequency-based methods use stops and links as the fundamental components of their transit network, the SBSP employs *nodes* and links. Nodes, in the SBSP algorithm, represent both the physical and temporal location of a transit stop are identified by a tripID and stopID. In order to transform the transit stops into nodes, one must create an individual node for each transit vehicle that arrives at an individual stop. In the regional context of the Twin Cities, this can, perhaps, be best thought of as taking a single physical stop and creating a node for each departure listed on the NexTrip arrival screen. Therefore, even if buses from two different routes arrive at a physical stop at the same time (as shown by the two 10:35 arrivals denoted in yellow in Fig. 4.3) or the same route serves the individual stop but at different times in the day (as illustrated by the two arrivals of route 5 denoted in red in Fig. 4.3), individual nodes will be created. Thus, for this small example, one physical stop expands into a multitude of nodes.



Stop Number 17902

7th St S & Nicollet Mall

Plan a transit trip: To here | From here

*Nex*Trip departs in **5 Min**

Current Time: 10:29 AM

Route		Departs
22B	N Lyndale/Via Logan-57/Brklyn Ctr ...	Due
5M	Fremont Av/Brklyn Ctr/Transit Ctr	5 Min
724	Ltd Stop/Brklyn Ctr TC/Starlite Ctr	6 Min
5F	Emerson/26Av-Broadway	13 Min
22C	N Lyndale/Via Dupont-57/Brklyn Ct...	10:49

• Real Time • Scheduled

Figure 4.3: **Network transformation from stops to nodes**

Each transit link in the SBSP algorithm uniquely connects two of the aforementioned nodes and can be categorized into one of four sub-types: (1) a walking-transfer link (in which a transfer is made between two different physical stops and time passes), (2) waiting-transfer links (in which a transfer is made to a different route at the same physical stop), (3) in-vehicle transit links, and finally (4) access/egress walking links. With this detailed categorization, travel time can be more precisely segregated leading to a multi-criterion approach to the definition of a shortest path, which therefore allows for greater model flexibility and accuracy.

4.2.2 SBSP Algorithm

Using an individual's origin and destination coordinates provided by the on-board survey, a "Dijkstra's shortest path" label setting path algorithm is implemented using the schedule-based time-dependent transit network.

First, a list of possible access and egress links (connecting the origin and destination to the transit network respectively) are created using the passengers' origin/destination coordinates and a network distance that is within the distance thresholds described above. Following the initial loading of the transit & passenger data as well as the access/egress links, each node is initialized with two labels (l_i and p_i). The label l_i indicates the length of time a passenger has traveled to reach the specified node while p_i denotes the last node the passenger was at.

Following the initialization step, the algorithm iteratively selects the node with minimum time label (l_i), removes it from the selection list (S), and updates all nodes that are connected to the current node and have a larger time label than the sum of the selected node and the time to traverse the link connecting the two nodes. This label setting process continues until all nodes have been scanned (S is empty). As soon as the selection list is empty, the algorithm, beginning at the destination node, uses the predecessor labels (p_i) to trace the shortest time path back to the origin. After reversing the order of this path, the shortest path between the passenger's origin and destination is printed and the algorithm terminates once all passengers have been processed. The following logic details the algorithm steps of this traditional SBSP process.

Notation

U Set of all passengers

N Set of all nodes

l_i Time label of node i

S Selection list of nodes

$\bar{S}$ Previously selected nodes list

A Set of all links

a_{ik} Link connecting nodes i and k

t_{ik} Time to traverse link connecting nodes i and k

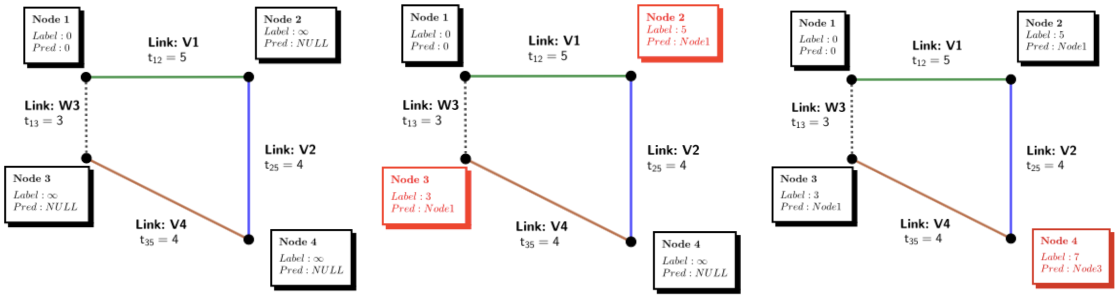
Algorithm 1 SBSP Path Generation

```
 1: for $u_i$ in U do
 2:     Inputs:
        Create N and A
        $n_{origin}, n_{dest} \leftarrow$ Survey Data
        Create access & egress links for $u_i$ and add to A
 3:     Initialize:
        $l_i \leftarrow \infty$
        $l_{origin} \leftarrow 0$
        $p_i \leftarrow \emptyset$
        $p_{origin} \leftarrow -1$
        $S \leftarrow \{N\}$
        $\bar{S} \leftarrow \{\}$
 4:     for $n_i \in N$ do
 5:         if $n_i$ has minimum $l_i$ then
 6:             Select the node $n_i$ as current node.
 7:             Remove $n_i$ from $S$ and add it to $\bar{S}$

 8:             for All links $a_{ik}$ emanating from current node $n_i$ do
 9:                 if $l_k > l_i + t_{ik}$ then
10:                     $l_k = t_{ik} + l_i$
11:                     $p_k = n_i$

12:     if $L = \{\}$ OR $\bar{L} = \{N\}$ then
13:         Continue
14:     else
15:         Return to Line 3
16:     Start at $n_{dest}$ and record subsequent predecessor nodes ($p_i$) until origin ($n_{origin}$)
17:     Reverse list order to obtain the shortest path between the $n_{origin}$ and $n_{dest}$ for $u_i$
```

SBSP Path Generation: Algorithm Example

In order to more easily understand the specific steps of the algorithm above, a simple example is provided describing how the SBSP algorithm is used on the sample network shown in Figure 4.4 for one passenger whose origin is located at node one. The sample network contains 4 nodes (numbered 1-4), 1 walking link ($W3$), and 3 in-vehicle links ($V1$, $V2$,$V4$) where the link travel times are listed adjacent to the respective link.

Figure 4.4a shows the initialization stage of the algorithm, while Figure 4.4b and Figure



(a) **Initialize** (b) **Update Nodes 2 & 3** (c) **Update Node 4**

Figure 4.4: **Example algorithmic labeling of nodes and links**

4.4c illustrate the remaining steps of the algorithm resulting in a shortest path traveled from the origin to destination by traversing Node 1 → Node 3 → Node 4. Due to the network representation composed of detailed nodes and links, this shortest path includes walking (W3) and in-vehicle links (V4), and for the actual Twin Cities network this shortest path would also include waiting links.

4.3 Trip Elimination Sub-Algorithm

As previously mentioned, passenger route choice behavior can only be analyzed and modeled if the transit user in question actually has a choice between attractive transit paths connecting his or her origin and destination locations. In its current state, however, the SBSP algorithm only produces a single path per person. Therefore, an additional module is added to the SBSP algorithm whose purpose is to iteratively generate additional attractive paths following the logic below.

In order to produce additional attractive paths for each user, the transit network, over which the SBSP is run, must fundamentally be altered as without this change, the algorithm will never produce more than one unique path. This sub-algorithm component, therefore, alters the transit network input to the SBSP algorithm by taking an individual's shortest path and removing, one at a time, each transit trip (and all associated nodes and links) from the transit network before reiterating the body of the SBSP described in the previous sub-section. In this manner, the network will have fundamentally been altered forcing the next generated path to be different than the original "shortest" path. The trips of each newly generated path will be added to a nested tree structure as children of the excluded trip (parent). Therefore, each time a new path is generated from an existing path, the excluded trips are all the trips excluded when generating the existing path as well as one of the trips on the existing path. The tree structure is defined as shown in Figure 4.5, adopted from Adrian Mejia [43].

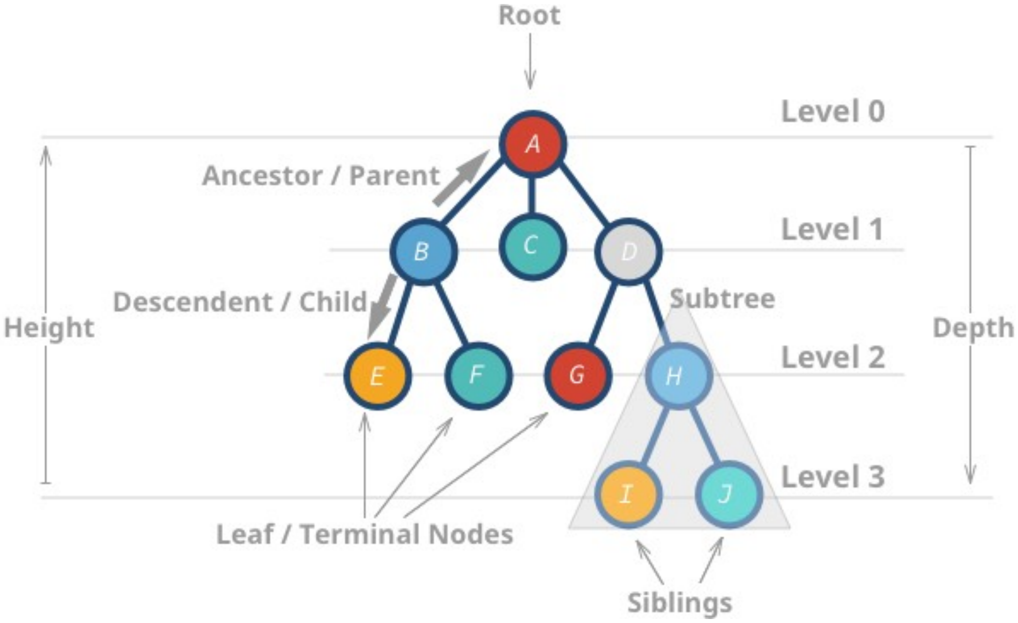


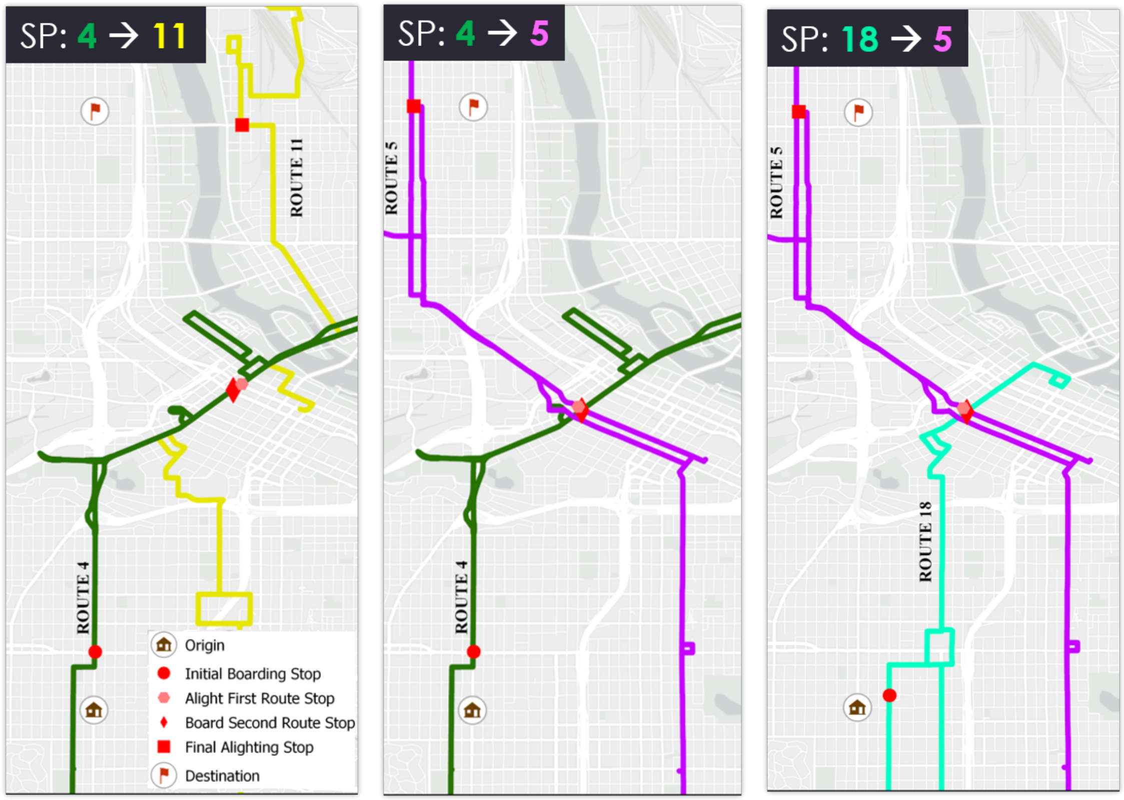
Figure 4.5: **Tree structure terminology**

Algorithm 2 SBSP With Trip Elimination

```
 1: Inputs:
        Create N and A (Including Access,Egress,Origin, and Dest)
        List of All Trips
        Sample Survey Population
 2: for passenger in passengerList do
 3:     previousPathList = $\emptyset$
 4:     newPathList = $\emptyset$

 5:     run SBSP Code to output shortestPath            ▷ Generate passengers' first path
 6:     append shortestPath to previousPathList

 7:     for initialTripID in shortestPath do                ▷ Each trip on SP is tree root
 8:         create ancestor tree with initialTripID as root

 9:     for selectedPath in previousPathList do:            ▷ Generate subsequent paths
10:         for tripID in selectedPath do
11:             excludedTripList $\leftarrow \emptyset$
12:             Load Complete (Un-Filtered) GTFS Network
13:             append tripID to excludedTripList

14:             for ancestorTrip of tripID do                ▷ Also exclude ancestor trips
15:                 append ancestorTrip to excludedTripList

16:             for excludedTrip in excludedTripList do              ▷ Filter GTFS Network
17:                 exclude nodes and links on excludedTrip
18:                 exclude excludedTrip from list of network trips
19:                 run SBSP code with filtered GTFS network
20:                 append newPath to newPathList

21:                 for newTripID in newPath do                ▷ Append new child trip
22:                     append newTripID as new child to excludedTrip

23:     if number of desired unique paths is reached then
24:         PASS
25:     else
26:         previousPathList $\leftarrow$ newPathList
27:         loop back to line 9
```

SBSP With Trip Elimination: Algorithm Example

Due to the inherently nested and looped structure of this new algorithm an example is provided to further clarify the algorithm steps. In this example, a single individual's choice set will be generated. In visualizing the steps of the trip elimination choice set generation, the brown home icon is the origin, the red flag symbol is the destination, and the other red geometric shapes correspond to boarding and alighting locations as defined in the legend of Figure 4.6a.



(a) **Shortest path**

(b) **Exclude Route 11**

(c) **Exclude Route 4**

Figure 4.6: **Example choice set generation with trip elimination**

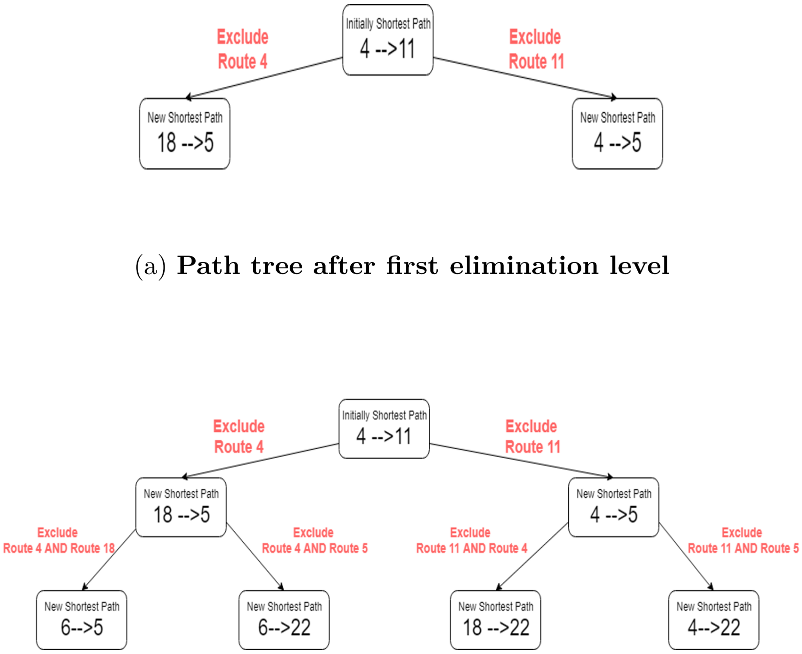
Figure 4.6a, illustrates the results of the traditional SBSP algorithm (Lines 1-6 in the SBSP With Trip Elimination Algorithm) where no trip information is excluded from the algorithm. In this case, the shortest (quickest) path between the selected passenger's home and destination location is to board Route 4 → Route 11 with a transfer in downtown Minneapolis. Next, as detailed in lines 7-8 of the algorithm, an ancestor tree is created for *each* trip component on the shortest path. Through subsequent iterations of the algorithm, each of these tree structures will be filled with descendant nodes. Due to the fact that the shortest path has two trips (Route 4 and Route 11), lines 9-22 will be iterated over twice–once for each of the corresponding trip ID's.

Figure 4.6b, illustrates the results of algorithm lines 9-22 when the selected *tripID* corresponds to the Route 11 trip. In this case, the Route 11 trip is added to the list of trips to exclude from the transit network (line 13). Because this is the first level of trip eliminations, there are no parent trips to exclude (lines 14-15) and as a result, the algorithm moves to lines 16-22 at which point all nodes, links, and trips associated with the specific Route 11 *tripID* are excluded from the GTFS transit network. Then, in line 19, the body of the SBSP code described in Algorithm 1 is run again this time using this filtered transit network to produce a new attractive path which is the quickest path available to the passenger in the filtered network. As shown in Figure 4.6b this new attractive path results in the passenger boarding Route 4 at the same stop as the initial shortest path, transferring downtown and taking Route 5 to his/her destination. The trips on this new path are then added (lines 21-22) as "child" trips of the original shortest-path trip that was excluded.

After this process has been completed for the exclusion of the Route 11 trip, the algorithm re-loads the GTFS network in its entirety (line 12) before, in line 13, excluding Route 4 (the remaining *tripID* on the original shortest path in Figure 4.6a). Again, because this is the first level of trip eliminations, there are no parent trips to exclude (lines 14-15) so the algorithm moves to lines 16-22 and excludes all nodes, links, and trips associated with the the Route 4 *tripID* resulting in the new attractive path Route 18 → Route 5 (Fig. 4.6c).

At the conclusion of this phase, the algorithm moves to lines 23-27, finds that the desired number of unique paths has not been reached, sets the *previousPathList* to include only the two new paths generated from lines 16-22, and returns to line 9 having completed the first level of trip eliminations. This path generation (lines 9-27) will continue for a given passenger until a specified number of unique paths has been reached at which point the algorithm will move on to the next passenger.

The resulting paths within the example passenger's choice set after one level of trip elimination, described above, are illustrated in Figure 4.7a. Arrows indicate which trip (listed as a route number for simplicity) has been excluded in order to generate the new path. If a second level of trip elimination were to have been described for the passenger above, two trips would have been excluded for each iteration: one from the initially shortest path, and a second from the newly generated paths found from the previous level of elimination. The resulting choice set and the corresponding trips (displayed as routes) eliminated to obtain a specific path are visualized in Figure 4.7b which only differs from Figure 4.7a in the bottom row of paths as it is a iterative progression.



(a) **Path tree after first elimination level**

(b) **Path tree after second elimination level**

Figure 4.7: **Trip elimination path tree**

4.4 Constraints, Assumptions, and Data Preprocessing

In addition to the restrictions placed on the access and egress walking links as previously mentioned in Section 4.1, several additional assumptions and preprocessing-measures must be made before running the aforementioned algorithm.

4.4.1 Preferred Departure Time (PDT) Assumptions

The most significant assumption deals with individuals' preferred departure time (PDT) from their origin. Within the 2016 On Board Survey, a PDT was not given. What is provided, instead, is a one hour timeframe in which the passenger was surveyed. This time range, however, is problematic for two reasons: first, it is not specified to single-minute resolution, and second, the time range is not necessarily when the passenger left the origin, but rather when he/she was surveyed.

To address the first issue, an additional constraint is placed upon the SBSP algorithm to allow for a pseudo-initial wait time of up to one hour. In practice, the algorithm sets the PDT to be the first minute of the surveyed time window and then scans all departing routes a passenger could reach within one hour that successfully connect to the passenger's destination. Following the termination of the algorithm, the pseudo-access wait time is set to zero and if multiple identical paths exist, only one of these paths is kept in the choice set. For example, if the surveyed range was 2:00 p.m.-3:00 p.m., the algorithm would set the PDT equal to 2:00 p.m. but find any valid trips that depart between 2:00 p.m. and 3:00 p.m. and are within the maximum allowable access walking distance (1.1 miles). Then, for instance, if a valid trip were found that departed at 2:45 p.m. the passenger's access wait time is set to zero minutes rather than the the difference in time between 2:45 p.m. and the time at which the passenger arrived at the stop after departing his/her origin at 2:00 p.m. Actual access wait time is then calculated using the method described in the next sub-section (Section 4.4.2).

Of the two issues related to PDT estimation, the second issue (provision of surveyed time range rather than first route boarding time), is particularly troubling as it is possible that an individual could be surveyed on a route other than the first route on his or her trip. Without correction, this issue would thus have led to inaccurate PDT estimations. In order to translate this surveyed time into a PDT, the sum of the average unlinked in-vehicle time (IVT) and the average transfer time is subtracted from the surveyed time. This fixed time adjustment is then used as a proxy for the average time associated with walking and transferring from each additional route. Thus, 23.2 minutes are subtracted from the surveyed time window for each transfer that occurred before the surveyed route.

4.4.2 Access Waiting Time Calculations

Given that the PDT is estimated rather than directly extracted from survey responses, time spent waiting at the boarding bus stop of the initial route also has to be estimated. The method of estimation follows the headway-based procedure described by Fan and Machemehl (2009) and formulated in Equation 4.1 [44].

$$AccessWaitTime = min\ (0.5 * Headway, 2.28 + 0.29 * Headway, 13.3) \tag{4.1}$$

This piece-wise equation is based on the assumption that passengers arrive at random for short headways (given in minutes) and strategically coordinate their arrival at bus stops served by routes with longer headways. In order to best capture the fluctuating nature of route headways throughout the day, headways are calculated as the time between the simulated first route taken and the next bus of that same route. If the simulated bus is the last of its route for the day, the difference in time between the preceding and simulated bus is used as the headway.

4.4.3 SBSP Algorithm Constraints and Parameters

In order to best capture passenger behavior, an additional set of algorithm parameters and constraints are introduced (Table 4.2). A total time threshold (160% of the shortest path's travel time) was implemented to make sure all paths are seen as "reasonably attractive" to the passenger. For example, if a passengers shortest total travel time is 10 minutes, he/she would consider trips with total times of 16 or less minutes. The 160% threshold percentage is chosen to include 95% of the current passenger set. Walking speed is assumed to be 3 miles per hour and is used to calculated walking link times. In regards to the transfer distance threshold it is assumed, given the fact that the vast majority of the system is composed of high-stop density local-bus routes, that individuals transfer at the stop on their current route that is closest to their next boarding stop which, more often than not, is directly across the street. Finally, it is assumed, based on values within the literature, that individuals associate a 15 minute disutility for each transfer they take and that they will not consider transferring if they have to wait for more than 20 minutes for the next route.

Table 4.2: **Algorithm constraints**

Parameter	Constraint
Transfer Distance	≤ 0.1 miles
Access(Egress) Walk	≤ 1.1 (0.72) miles
Assumed Transfer Penalty	15 min
Transfer Wait Time	≤ 20 min
Walking Speed	3 mph
Path Length	≤ 160 % Shortest Path Time

Chapter 5

Logit Model Methodology and Formulation

Following the generation of a set of attractive path choices for each passenger within the sample population, logit models are estimated using version 0.2.2 of the PyLogit estimation tool [45]. In performing the logit estimations, 70% of the passengers are used as the training dataset and the remaining 30% are used as the testing dataset. Two types of model structures are considered: Multinomial Logit (MNL), and Mixed MNL. The Mixed MNL model is tested in order to determine the significance of variation in parameter perception amongst individuals. In this formulation, the parameters are treated as normal random variables and the logit function is formulated in the same form as is described below in Equation 5.1.

5.1 Path Overlap Correction Factor

One of the key features of the MNL model is the assumption of independence of irrelevant alternatives (IIA). This assumption states that when individuals are asked to choose from a set of alternatives, the likelihood of them picking one choice over another should not depend on an additional alternative. In other words, it is assumed that the choices are not correlated. Unfortunately, in the case of the transit route choice problem, many of the path choices can be, in fact, correlated. In particular, it is often the case that two routes traverse the same exact stretch of roadway and are only different in their route number. As a result of this potential correlation, a path-size correction term, described by Tan et al. (2015), is introduced to the model (16). The MNL formulation with an overlapping path size correction factor is described by Equation 5.1,

$$P(i|C_n) = \frac{e^{V_{in}+\beta_{PS}PS_{in}}}{\sum\limits_{j\epsilon c_n} e^{V_{jn}+\beta_{PS}PS_{jn}}} \tag{5.1}$$

where:

$P(i|C_n)$ = Probability of taking path **i** given choice set **C** for person **n**

V_{in}, V_{jn} = Utilities for path **i** and **j** for person **n** respectively

PS_{in}, PS_{jn} = Path-size correction for path i and j in choice set C_n respectively

β_{PS} = Estimated coefficient for the path size correction term.

The path size correction term, as defined by Tan et al. (2015) [46] and adapted for stop overlap rather than link overlap, is shown in Equation 5.2,

$$PS_{in} = \sum_{s\epsilon\Gamma_i} \frac{1}{N_i} ln(\sum_{j\epsilon c_n} \delta_{sj}) \tag{5.2}$$

where:

Γ_i = Set of all stops for path **i**

N_i = Total number of stops served by path **i**

δ_{sj} = 1 if stop **s** is on path **j** and 0 otherwise.

While the path size correction term is typically calculated based on link length, for transit, it is reasoned that stops are more indicative of the actual overlap experienced by passengers. For example, when comparing bus rapid transit (BRT) service with local bus service, the paths may overlap for a long stretch of distance but only share a small number of actual stops due to the wider stop spacing present in BRT networks. Additionally, Nassir et al. (2015) argue that modeling at the stop level is more consistent with the decision transit users typically make [34]. Therefore, it is at the stop level that overlapping becomes most critical. Due to the path size correction term formulation within the MNL formulation, it is expected that β_{PS} will be negative in value indicating that increased path overlap leads to smaller relative utilities for each path.

Chapter 6

Results and Discussion

6.1 Overview of Passenger and Network Features

The analyzed network contains a total of 487,045 nodes and 2,949,325 links (not counting access/egress links which are generated on a per person basis). A more detailed glance into the network structure is portrayed in Table 6.1.

Table 6.1: **Network composition**

Network Feature	Count
Stops	13,579
Trips	9,211
Nodes	487,045
In-Vehicle Links	478,063
Waiting Transfer Links	468,373
Walking Transfer Links	2,002,889

After the initial filtering of the survey data to only include passengers with walk access/egress links and who made their journey on the specified Tuesday's, the initial passenger dataset contains 2,754 passengers. Following the completion of the choice set generation and filtering to include only passengers who had at least two paths, 1,938 passengers, or 70.5% of the original dataset, have a choice set that successfully includes the actual surveyed path that the individual took. This dataset is then further filtered to only include passengers who have at least two paths with a total travel time less than 160% of the shortest total time path (as described in the restriction parameters above). After this final filtering 1,724 passengers remained in the sample population with the average person having 3.1 attractive paths to choose from (Figure 6.1). Given that only 214 passengers (11%) did not have more than two attractive alternatives, it can be concluded that Twin Cities transit users almost always have multiple options to consider when selecting a path to reach their destination.

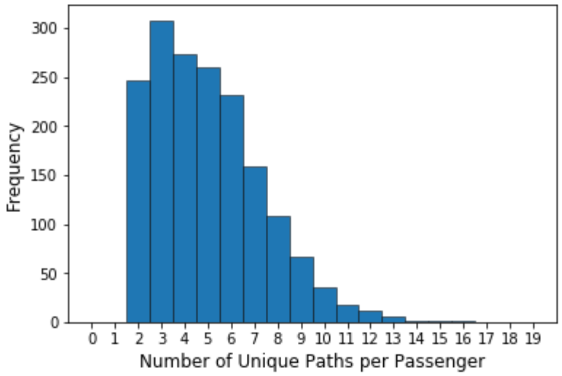


Figure 6.1: **Choice set size per passenger histogram**

One hallmark of this research is that the sample dataset, even after the aforementioned filtering, is extremely representative of the total transit rider population. When comparing the racial composition of the sample dataset to the full population, all races/ethnicities were within 1% of the entire surveyed population except for whites (+ 3.4 percentage points from total population) and Asians (+ 1.7 percentage points). Females comprise 45.5% of the sample population, only 2.7 percentage points lower than the population average, and all age categories are represented to within 5% of their actual occurrences within the total population.

Prior to examining the results of the multinomial logit (MNL) model, the paths actually taken by the transit passengers (as recorded in the survey) are studied. From an aggregate analysis of these surveyed paths, as compared with the other paths within an individuals choice set, broad qualitative conclusions can be drawn about riders' behavior that can be supported (or contradicted) by the logit results.

On average, the sampled passengers' surveyed (taken) path has a mean total travel time of 40.3 minutes, 16.2 minutes walking, 5.4 minutes waiting, and 22.2 minutes in a transit vehicle. Additionally, there is a fairly strong aversion to transfers as 66.37% of passengers' actual paths have no transfers, 27.7% have one transfer, 1.7% have 2 transfers, and only 1 person out of 1,724 sample passengers transfers three times.

Five attributes are created for each passenger indicating if the individuals' surveyed (taken) path followed a particular strategy, when compared to the other attractive paths in the simulated choice set (Table 6.2).

Table 6.2: **Surveyed (taken) path strategy frequency**

Surveyed Path Strategy	Percent of Sample Population
Minimum # of Transfers	86.0%
Minimum Wait Time	49.5%
Minimum Total Time	46.4%
Minimum Walking Time	39.6%
Minimum In-Vehicle Time	26.7%
Minimum in ALL of the above 5 categories	7.3%

From this aggregation, it appears that the number of transfers is the most critical factor involved in deciding which path to take as 86% of the sample population are surveyed on a path that has the minimum number of transfers. Furthermore, the minimum walk time appears to be a weaker predictor of route choice than waiting time (39.6% versus 49.5%). As a result, it is expected that the transfer penalty will be quite large and that the magnitude of the logit model coefficients (all of which are expected to be negative) will be largest for the wait time followed by walk time, and then in-vehicle time.

6.2 Multinomial Logit Results

Both traditional MNL models and mixed logit models are tested on the training dataset (defined as a random 70% of the sample passenger set) with varying interactions and combinations between the variables listed below. The mixed logit models, in which parameters are assumed to be normally distributed, however, are found to not improve the model fit in comparison with MNL models. Due to this lack of improvement and the greater simplicity of the MNL models, only the MNL model results will be presented. After comparing many MNL models, the two best models (chosen as having the lowest Bayesian information criteria (BIC) value and highest McFadden's rho-squared value) are presented where all presented variables are statistically significant. The considered variables are as follows:

- Path Size Correction Factor
- Route Type (Transitway, Express Bus, Local Bus)
- Number of Transfers per Path
- Categorical Annual Household Income
- Race
- Gender
- Timing Parameters (IVT, Access\Egress Walk Time, Access Wait Time)
- Trip Purpose (Work, School, Shopping, Meal, Other).

6.2.1 Expanded MNL Model Results

The model results presented in Table 6.3, encompass all significant variables and their interactions from the list of variables above. Given the number of observations in the sample training dataset (1,207), the MNL model's adjusted rho-squared value of 0.423 is relatively large. When analyzing the marginal rate of substitution column, in which all parameter coefficients have been normalized to the scheduled non-transitway IVT coefficient, several interesting conclusions can be extracted. First, passengers perceive 41 seconds (0.687 minutes) on board a transitway route as the equivalent of 1 minute on a non-transitway route indicating that Twin Cities transit passengers have a preference towards transitway routes when given the choice. Additionally, given the marginal rates of substitution for walking and waiting time, passengers perceive 2.65 minutes of non-transitway in-vehicle time (IVT) as one minute of waiting time at their initial boarding stop, 2.77 minutes IVT as 1 minute of egress walking, and 1.65 minutes IVT as 1 minute of access walking time. An interesting takeaway from this comparison is that individuals are more likely (1.6 times) to walk a greater distance when accessing their initial transit route, then they are when walking from the alighting stop on their last transit route to their final destination.

Table 6.3: **Expanded MNL coefficient (beta) values and marginal rate of substitution w.r.t. non-transitway in-vehicle time**

Parameter	Coefficient	Std. Error	Confidence Interval (95%)	Marginal Rate of Substitution (With Respect to Hours of IVT)
Non-Transitway IVT (Hours)	-6.486	0.597	[-7.655, -5.316]	1.000
Transitway IVT (Hours)	-4.458	0.736	[-5.901, -3.014]	0.687
Access Walk Time (Hours)	-10.712	0.692	[-12.069, -9.355]	1.652
Egress Walk Time (Hours)	-17.938	0.886	[-19.675, -16.202]	2.766
Access Wait Time (Hours)	-17.210	1.616	[-20.378, -14.042]	2.654
# of Transfers per Path	-2.766	0.153	[-3.065, 2.466]	0.426 (25.6 min penalty)
Stop Overlap Correction	-0.646	0.226	[-1.087, -0.202]	—
Transitway: Access Walk Time (Hours)	2.013	0.740	[0.563, 3.463]	-0.310
Transitway: Annual HH Income >$150,000	2.310	1.073	[0.208, 4.412]	-0.356
Express: Annual HH Income <$35,000	-1.655	0.423	[-2.484, -0.825]	0.255
Express: Annual HH Income $35,000-$60,000	-2.470	0.635	[-3.715, -1.224]	0.381

Number of Observations	1,207 (70% of Sample Population)
Initial Log-Likelihood	-1,755.6
Final Log-Likelihood	-1,001.9
BIC	2,082.0
Adjusted ρ^2	0.423

Due to the fact that all marginal rates of substitution for the time-related parameters are not 1.0 (with respect to non-transitway IVT), it is apparent that all time components are not perceived equally as individuals' appear more likely to choose alternative paths with longer in-vehicle times in order to minimize the time spent walking and waiting for transit vehicles. The marginal rates of substitution in Table 6.3, are found to be in close alignment with the respective values published by other scholars [47]. Furthermore, after converting the marginal rate of substitution for the number of transfers in a path to minutes, it is determined that a transfer is perceived to cost approximately 25.6 minutes of in-vehicle time. While this value is at the high end of other values reported in the literature [47], most of which are centered between 5 and 20 minutes, this behavior appears to accurately capture regional behavior as two thirds of the sample population takes a path with no transfers and 86% of the population follow a route choice strategy that minimizes their total number of transfers. Additionally, the average trip headway (minutes between buses of the same route) for the entire transit network is 22.8 minutes. Therefore, if an individual misses his/her connecting route he/she, on average, must wait an additional 22.8 minutes. As a result, a transfer penalty of 25.6 minutes appears to potentially address the nearly 23 minute cost associated with the chance of missing the transfer as well as the actual inconvenience of having to make the transfer itself.

Beyond the time-related parameters, several interesting takeaways can be noted. First, one must remember that only significant parameters are included in the model so parameters missing from the model (such as Transitway: Annual HH Income $35,000-$60,000) are insignificant and are therefore assumed to have a coefficient value of 0. With this framework in mind, the coefficient values are compared amongst the categorical interaction parameters. Due to the fact that the "Transitway: Access Walk Time" coefficient is positive, transit passengers prefer to walk further from their origin to their initial transit route if this increased walking time results in a transitway anywhere along their path. In other words, access walking is perceived less negatively if the passenger can then ride on a transitway. When comparing the interaction between route type and individuals' annual household income, it is apparent that the highest income earners perceive transitways the most positively while low income individuals have a negative association with express routes (given that all other income categories are insignificant and therefore have a coefficient value of 0 while low income groups have a negative coefficient). However, within the two lowest income categories (<\$35,000 and \$35,000-\$60,000), the relative disutility of express routes decreases for individuals in the lowest income category (<\$35,000). One potential explanation for this variance is that many of the passengers in the lowest income category may receive discounted fares on express buses if traveling outside peak rush hours (\$0.75 rather than \$2.25). This reduced fare may, therefore, make express buses more favorable for the lowest-income individuals.

6.2.2 Simplified MNL Model Results

A simplified MNL model (Table 6.4) is estimated in addition to the expanded model (Table 6.3) in order to have an accurate model that can more easily be input into transit assignment models. Despite the reduction in the number of parameters, this simplified model still has a relatively large adjusted rho-squared value of 0.415.

Comparing the marginal rates of substitution between the expanded and simplified models, one finds that they are, for the most part, in close approximation with one another. In fact, the only parameters with a change in marginal rate of substitution greater than 0.2 are the Transitway IVT and Access Wait Time. In the simplified model, transitway IVT has a smaller relative marginal rate of substitution while the associated value with the access wait time parameter is higher. In other words, passengers described by the simplified model have a more positive perception of transitways and a more negative association with waiting at their initial boarding stop. Additionally, the transfer penalty per transfer is 2.2 minutes larger than in the expanded model.

Overall, however, the differences between the models are minute and so the simplified

Table 6.4: **Simplified MNL coefficient (beta) values and marginal rate of substitution results**

Parameter	Coefficient	Std. Error	Confidence Interval (95%)	Marginal Rate of Substitution (With Respect to Hours of IVT)
Non-Transitway IVT (Hours)	-5.867	0.547	[-6.939, -4.796]	1.000
Transitway IVT (Hours)	-2.187	0.613	[-3.388, -0.986]	0.373
Access Walk Time (Hours)	-10.246	0.651	[-11.523, -8.970]	1.746
Egress Walk Time (Hours)	-16.971	0.849	[-18.635, -15.308]	2.893
Access Wait Time (Hours)	-18.227	1.577	[-21.318, -15.135]	3.107
# of Transfers per Path	-2.814	0.152	[-3.113, -2.516]	0.480 (28.8 min penalty)
Stop Overlap Correction	–0.779	0.221	[-1.213, -0.345]	—
Number of Observations	1,207 (70% of Sample Population)			
Initial Log-Likelihood	-1,755.6			
Final Log-Likelihood	-1,019.3			
BIC	2,088.0			
Adjusted ρ^2	0.415			

model is used henceforth. A fairly high degree of matching between the simulated path and the passengers' surveyed path is found when using the simplified estimated multinomial logit model and applying in to the testing dataset (remaining 30% of the total sample passenger set). In particular, the average simulated probability of the chosen path was 53.4%. Additionally, the simulated path with highest probability matched the surveyed path for 66.5% of the sample passengers indicating that the model is "correct" nearly two thirds of the time.

6.3 Sample Logit Results Visualized

In order to best illustrate the importance these calibrated route choice parameters have on simulating passengers route choice behavior, recall the example choice set from Figure 4.7b. This example passenger is traveling from his/her home location near uptown to a destination in North Minneapolis and is presented with 7 potential paths (Fig.6.2 and Table 6.5). Without the route choice parameters and a conceptual understanding of how the passenger makes decisions, there is no way to ascertain which of the paths in Figure 6.2 is most attractive to the passenger. Only after applying the route choice parameters and simulating the path likelihood does the passengers' choice making behavior become apparent.

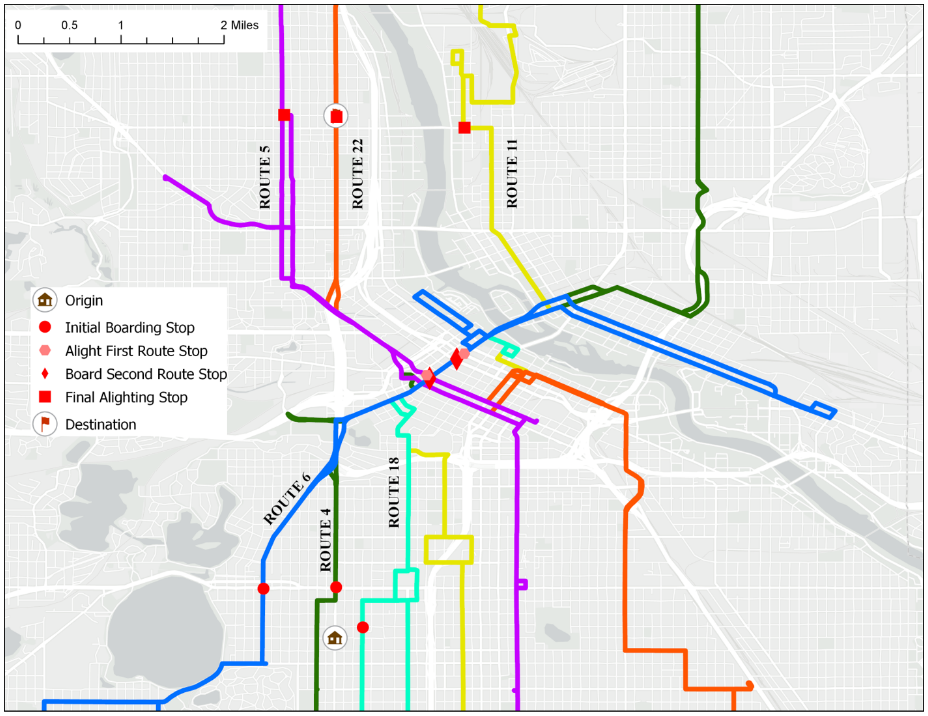


Figure 6.2: **Sample choice set mapped**

As listed in Table 6.5, the most attractive alternative path (and thus the path the passenger was assigned to) is "Alternative A." While the total travel time of this path is not markedly different from the other alternatives, and, in fact, is approximately average in length, the true difference between the chosen alternative and all other paths lies in the walking and waiting components of the trip. Hearkening back to the route choice model (Table 6.4), the largest marginal rates of substitution, and therefore the time components the passenger seeks to minimize most are, in order, access wait time, egress walk time, and then access walk time. Returning to Table 6.5, any alternatives with long access wait times and non-zero egress walk times can be eliminated. This criteria therefore leaves only Alternatives A and B remaining between which the largest differentiator is that Alternative A has an access walk time that is half that of Alternative B. In this rough example, the route choice parameters and their relative value compared with one another can, therefore, quickly give order to a choice set and conceptually illustrate how path probabilities are calculated.

Table 6.5: **Sample choice set probability table**

Alternative	Attractive Path	Prob.	Total Time (Min.)	Access Walk (Min.)	Access Wait (Min.)	IVT (Min.)	TR Wait Time (Min.)	Egress Walk (Min.)
A	4 → 22	63.2%	64	10	4	31	19	0
B	6 → 22	19.6%	65	20	3	31	11	0
C	18 → 22	5.5%	73	6	11	40	16	0
D	4 → 5	5.3%	64	10	4	32	10	7
E	4 → 11	4.4%	58	10	4	32	2	10
F	6 → 5	1.6%	64	20	3	32	2	7
G	18 → 5	0.4%	70	6	9	41	7	7

Chapter 7

Model Application Case Study

Route choice parameter coefficients form the backbone of transit ridership projections. Therefore, now that these parameters have been estimated for the current state of the regional transportation network (including the addition of the light rail, commuter rail, and bus rapid transit lines since 2000), transit ridership projections with increased accuracy can be performed. The key benefit of performing these transit assignments is that the route choice parameters are calibrated to the route choice behavior from the Twin Cities rather than adopting fixed values from other cities.

To illustrate the value of the calibrated route choice parameters, the Flexible Assignment and Simulation Tool for Transit and Intermodal Passengers (FAST-TrIPs) model developed by Alireza Khani (2013) [48] will be implemented to ascertain the ridership impact the introduction of the A Line arterial Bus Rapid Transit (aBRT) route had on ridership on all routes across the Twin Cities region.

7.1 Assignment Methodology

7.1.1 Demand Scaling

Similar to the SBSP algorithm described in the route choice methodology above, the FAST-TrIPs model requires origin and destination locations as well as the GTFS transit network data. Unlike the SBSP methodology, however, this assignment methodology requires an input demand file for all transit riders in the metro region. Thus, while a sample of approximately 2,000 passengers is used for the route choice calibration this input number of passengers is scaled up to be nearly 212,000. The input demand is again taken from the 2016 On-Board survey records for passengers that have access/egress links that are walked rather than traversed by any other mode. Within the on-board survey, each individual record (passenger) has an associated linked weight factor which estimates "the number of trips per day" that are made between the origin and destination location. As such, the input demand file for the FAST-TrIPs model was generated by creating a number of duplicate passengers (same origin and destination coordinates) equal to the linked weight factor. In order to avoid overloading individual routes and to provide a more accurate and smooth demand function with respect to the time of day, each of the duplicate origin-destination passengers was randomly assigned a preferred departure time (PDT) within 1 hour of the PDT of the base passenger contained within the survey record. The result (Fig. 7.1) is a fairly typical demand distribution plot with peaks in the morning and evening rush hours.

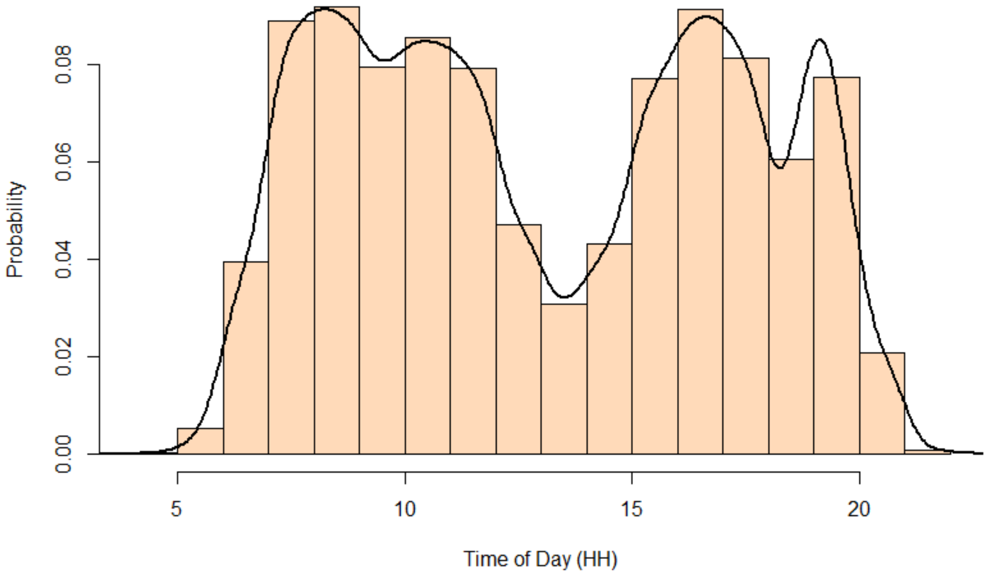


Figure 7.1: **Preferred departure time probability distribution**

7.1.2 K Means Spatial Demand Clustering

Aside from the demand profile, the only other significant departure from the SBSP methodology lies in the method by which the access and egress links were generated for the FAST-TrIPs algorithm. Due to the small number of passengers (2,000) in the SBSP sample population, access and egress links were generated connecting each passenger's unique origin and destination to the transit network. Unfortunately, due to the much larger number of passengers that are input to the FAST-TrIPs model, replicating the method for producing access and egress links at a latitude-longitude level is not feasible as this simulation would have taken 27 days to complete.

While transportation analysis zones (TAZs)–geographic units usually consisting of one or more census blocks, block groups, or census tracts–are often used as a proxy by which to categorize transportation demand, such a method is fraught with problems in the context of Twin Cities transit assignment. In practice, TAZ methodology most frequently assigns the location of all individuals contained within the zone to all be artificially placed at the center of the zone. As a result, the number of origin and destination locations is drastically reduced becoming a function of the zones rather than the number of sampled passengers. While this may seem like the perfect solution to the problem of access/egress link generation for increased demand, this method is best suited for auto rather than transit-based trips. The reason this method is inapplicable to the transit route choice context is the inherent inaccuracy in this method's walking distance calculations as measured from the TAZ centroid. As described in Section 6.1, individuals have a strong aversion to walking relative to in-vehicle time. Therefore, in order to minimize walking distances, the transit assignment model will most likely assign the passenger to the closest stop to the TAZ centroid.

As shown in Figure 7.2, due to the short distance between local bus stops (small purple dots), local bus stops have a much higher likelihood of being the closest stop to the TAZ

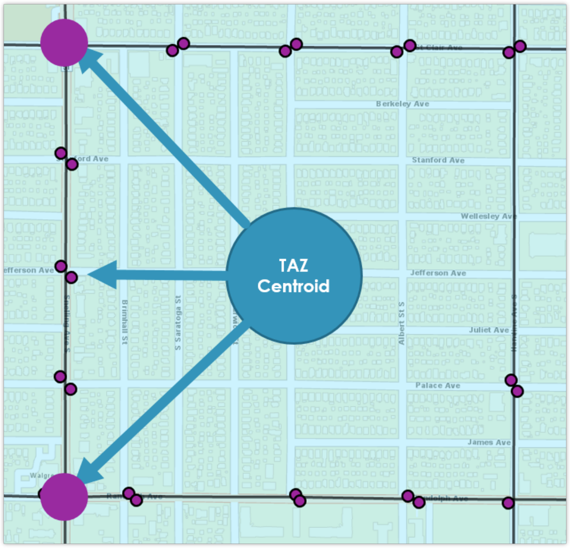


Figure 7.2: **Incorrect passenger assignment due to TAZ centroid methodology**

centroid when compared with BRT stops (large purple circles) which have larger stop spacing. As a result, although passengers have a higher relative utility for transitways (BRT, LRT, and Commuter Rail), the assignment algorithm is more likely to assign passengers to local routes when using a TAZ method to create access and egress links. This incorrect assignment is especially prevalent considering that TAZ boundaries (shown as black lines in Figure 7.2) often are drawn down the middle of key local streets and arterials which contain a high percentage of the fixed route transit service in the Twin Cities region.

Clearly, a method must be implemented to generate access and egress links that combines the increased accuracy of origin/destination locations, as found in the route choice calibration method, with the fast computational times of the TAZ approach. The proposed method that accomplishes both goals and produces higher accuracy passenger assignment is a spatial clustering approach.

When researching the existing literature on spatial clustering methodologies, two primary clustering methodologies can be found based on the inputs they require. The first clustering methodology relies on a maximum intra-cluster distance threshold between points while the second method relies only on a total number of clusters to create. Within the context of origin and destination clustering, however, the distance threshold method does not produce high-resolution clusters. Instead, because there are so many origin and destination locations within the downtown regions, these methods create many clusters in the suburbs while only generating one cluster for the downtown area because each of the downtown points are within the maximum distance threshold of at least one other downtown point. Due to this over-aggregation in downtown areas, the method used for this case study is K Means Clustering in which the total number of desired clusters is input.

K Means clustering first creates "k" number of unique cluster points. Then, the algorithm determines measures the distance between the demand location and each of the k cluster points before assigning the demand location to the nearest cluster point. After all locations have been assigned to a cluster point, the algorithm re-calculates the location of the cluster point (hereafter referred to as the cluster centroid) to be the average position of all locations within that cluster. This process continues until no centroid changes position after recalculation.

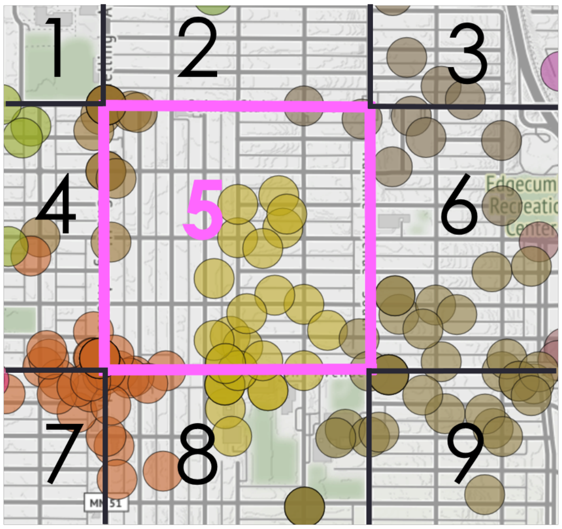


Figure 7.3: **K Means clustered locations overlaid with TAZ grid**

The results of the K Means clustering of passengers' origin and destination location are shown in Figure 7.3 for a section of Snelling Avenue North. This figure depicts the traditional TAZ grid overlaid above the demand locations (colored circles). Additionally, TAZ 5 (boundaries highlighted in pink) is the same central TAZ shown in Figure 7.2. Comparing these two figures one can see that there is, indeed, a large number of demand locations near the centroid of the old TAZ (as shown by the off-yellow color). However, the large pockets of demand to the southwest and northwest corners of TAZ 5 (orange and dark brown respectively) now are assigned a cluster (rather than TAZ) centroid that is almost exactly at where their true location lies rather than being removed by at least three city blocks. As a result, the transit assignment for these individuals, and the greater metropolitan region as a whole, is in much closer agreement with the actual paths these passengers were surveyed on.

In order to determine the validity and accuracy of the K Means clustering method, the ensuing transit assignment results for the full 2016 demand were compared with the *Transit Stops Boardings and Alightings* data from 2016 provided by the Metropolitan council and uploaded on the Minnesota Geospatial Commons (`https://gisdata.mn.gov/dataset/us-mn-state-metc-trans-stop-boardings-alightings`). When analyzing the average weekday ridership on select transitway routes (Blue Line, Green Line, and A Line), it was found that the average values from the Geospatial Commons were 9-12% higher than the corresponding values reported on the Metro Transit website (`https://www.metrotransit.org/metro-transit-ridership-tops-826-million-in-2016`). As a result, the average weekday ridership on these three routes was fixed as the value reported on the Metro Transit website. It should be noted, however, that it is unknown whether the other non-transitway routes are similarly overestimated in the Minnesota Geospatial Commons data used for the methodology validation shown in Figure 7.4.

All routes with over 1,000 daily riders are included within this figure with the average weekday ridership simulated by the route choice transit assignment model on the x-axis

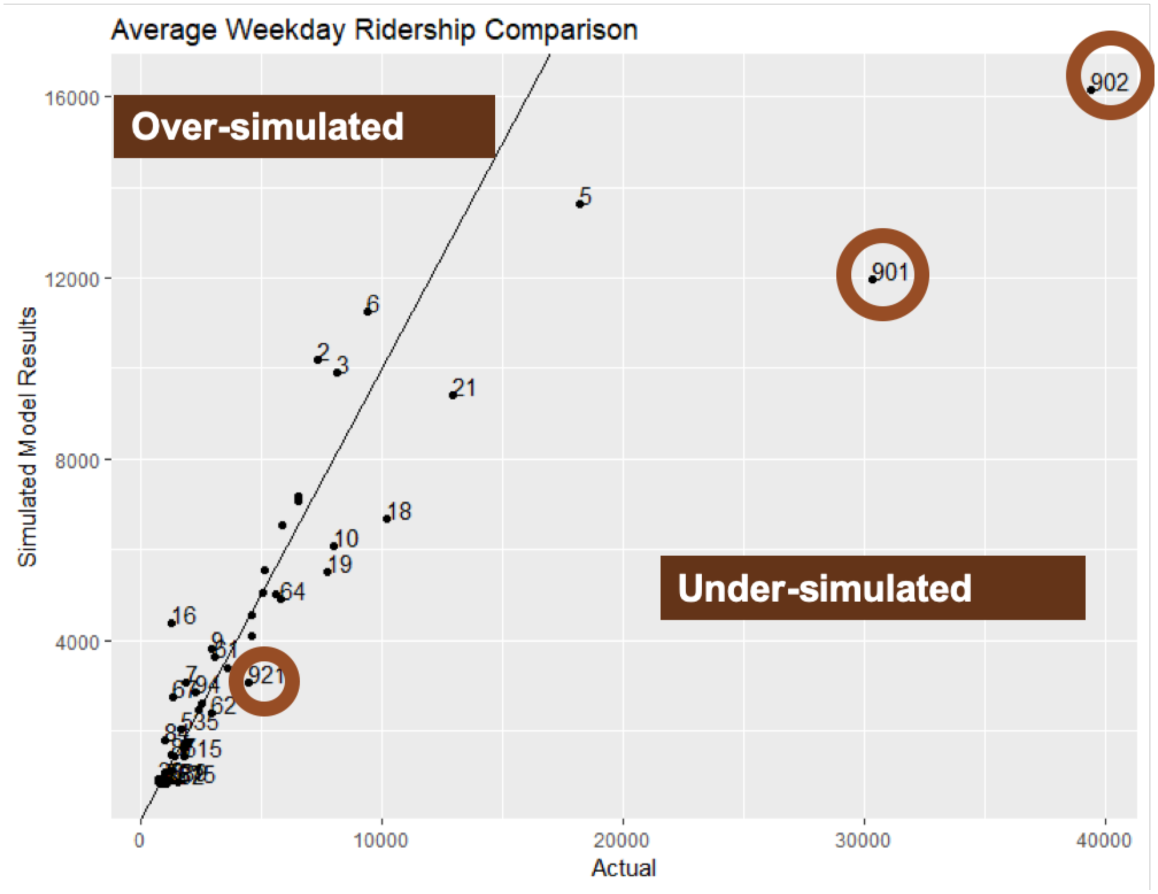


Figure 7.4: **Comparison between simulated and realized daily route ridership**

and the corresponding reported Geospatial Commons data on the y-axis. Additionally, route number labels have been provided for any routes that deviate by more than 15% from the "actual" ridership. Perfect agreement between the modeled ridership and Geospatial Commons data would be realized if all points fell along the diagonal line. From this figure, it is apparent that the majority of routes, including the A Line (denoted as Route 921) are in close approximation with the actual weekday ridership observed in 2016. The two significant outliers, even with the demand location clustering methodology, however, are the Blue Line (Route 901) and Green Line (Route 902). At present, the best explanation for this significant deviation is that the positive relative utility associated with rail transitways is not fully encompassed within the model. When including the two light rail lines, the fitted linear regression line's slope between simulated and actual ridership is 0.75 (where 1.00 would indicate perfect agreement on average) while the slope value is 0.86 when excluding the light rail lines. As a result, it appears that the route choice model with K Means clustering assigns transit demand fairly precisely for all routes except the Blue and Green Lines.

While the under-simulation of the light rail lines remains an ongoing investigation, when comparing the transit ridership before and after the introduction of the A Line service, the methodology will be the same. Therefore, because the relative ridership changes between the two time periods is being investigated, rather than absolute numbers, this inconsistency will not cause inaccurate results.

7.1.3 Ridership Comparison Methodology

When comparing ridership changes before and after A Line implementation, the only methodological change that was necessary was a change in the transit network GTFS files. As a result, for both the transit assignment before and after the introduction of the A Line, the same aforementioned route choice model parameters and input demand were used, with just the A Line trip, stop, and stop time information removed from the network files. For easy visualization of ridership changes between the two time periods at both the regional, route, and stop levels, an interactive R Shiny application was created. Additionally, comparisons were made only for routes that existed in both time periods.

7.2 Case Study Results

Using the R Shiny application and the output load profiles generated for the before and after A Line time periods several conclusions can be made. First, as shown in Table 7.1, only six routes had a daily ridership change in magnitude of greater than 4% of the ridership levels from before the A Line. Of these routes, Route 84 (the route that the A Line was largely designed to supplement) saw the greatest decrease in average daily ridership–falling by nearly 90%. Furthermore, only one route across the entire metro region (excluding the A Line as it did not exist in both time periods) saw an increase in daily ridership greater than 5%. This route, Route 32, travels east-west from the northern terminus of the A Line (Rosedale) to Robbinsdale.

This ridership increase may therefore signify that the increased service quality associated with the A Line has helped to facilitate a marginal increase in east-west connections that were not present before the A Line. Using the exact same demand size and locations between the two time periods (therefore excluding the impacts of induced demand), this case study has indicated, through use of the new route choice model parameters, that the A Line has had a significant impact on the ridership as riders from the nearest and most parallel routes have transitioned from their previous route to the A Line.

Table 7.1: **Average weekday ridership change due to A Line service**

Route Number	Route Description	Daily % Change	Daily Count Change
84	Rosedale - Snelling - Sibley Plaza	-89%	-1,585
87	UMN St. Paul-Cleveland-Highland	-19%	-285
65	Dale Street-County Rd B-Rosedale	-10%	-118
74	46th St.-Randolph-W 7th St.-Sunray	-6%	-302
67	W Minnehaha-Raymond Station-Franklin Ave	-4%	-113
32	Robbinsdale-Lowry Ave-Rosedale	+5%	+95

Chapter 8

Conclusion and Contributions

The primary aim of this research was to create a new route choice model for the Twin Cities metropolitan region that included all the new routes and transitways introduced in the last two decades. In creating this model, a new choice set generation methodology was proposed that systematically generated choice sets of between 2 and 15 attractive paths for each passenger by iteratively excluding previous paths' trip components from the transit network. Using these choice sets, a multinomial logit estimation model with path size correction factor was used to maximize the likelihood that the simulated transit path matched the actual path taken by each passenger.

On-board transit survey data was used to generate a preferred origin departure time, precise origin and destination coordinates, the actual path taken by the passenger, and the passenger specific demographic and trip purpose information. In addition to being the first known transit route choice model to encompass transitways in the region, a major contribution of this study resides in the choice set generation method. This research demonstrated that by including an iterative trip elimination method embedded in the schedule based shortest path algorithm, a more robust and complete choice set can be generated without having to continuously alter input parameters.

A second aim of this analysis was to ascertain which attributes (both network specific and passenger specific) had the strongest influence on a passengers' choice of a transit path and if passengers chose the shortest (quickest) path. Based on the results of the multinomial logit model, it can be concluded that passengers do not perceive the passage of time uniformly with only 46% of passengers taking the shortest path. More specifically, passengers perceive the relative disutility of waiting to be three times larger than local bus in-vehicle time. Additionally, the relative disutility of local bus in-vehicle time is nearly three times that of transitway in-vehicle time and Twin Cities passengers associate a penalty of nearly 29 minutes with each transfer. This strong aversion to transfers is further evidenced based on the observation that when selecting a path from a set of attractive paths, nearly 90% of the population chooses a path with the minimum number of transfers.

Together, the route choice parameter results allow for heightened accuracy and understanding of transit riders within the Twin Cities. By using existing route-level average weekday ridership counts as a baseline comparison, the previously mentioned route choice parameters have been used to calibrate transit assignment models. Specifically, as illustrated with the A Line case study (Section 7), the coefficient values and relationships amongst the variables relating to the different components of time can be used to further improve the precision and validity of ridership projections and to highlight the impact service additions (or subtractions) can have on system-wide ridership.

8.1 Future Work

Future avenues of work to be pursued are the inclusion of a transfer reliability factor as well as an analysis on the perception of transit delays at both the stop and route level and the subsequent impacts on transit route choice. Furthermore, a cross-nested logit model could be created and compared with the multinomial logit model. Additionally, further sensitivity analysis could be conducted on the trade-off between the number of K Means demand clusters and the computational time to determine the number of clusters that yields the best results. While time-intensive, an additional direction for future work could center around the iterative regeneration of choice sets. Specifically, the entire procedure of generating a choice set for each passenger and then estimating the route choice behavior with a MNL model could be iterated over where, for each new iteration, the time-related route choice coefficient inputs to the choice set generation algorithm would be set as the previous iterations' MNL parameters. Through this iterative, but very time intensive, procedure, one could examine if the model converged to a "best" set of estimated route choice parameters.

While many interesting and exciting avenues exist for future research, the main contributions of this research can currently best be summarized as creating a new choice set generation method and path size overlap factor calculation as well as implementing a route choice model that describes both traditional local and express routes as well as the implementation of transitways in the years since 2000. The results of this research, therefore, provide a glimpse into the numerous transportation decisions individuals make every day. By increasing our understanding of these decisions, transit systems and policy can, therefore, be better tailored to meet the needs and desires of passengers both in the present and future.

Bibliography

[1] B. J. Sahakian and J. N. Labuzetta, “Bad moves: How decision making goes wrong, and the ethics of smart drugs,” 2013.

[2] PB Consultants, Inc, “Twin Cities Regional Model: Development of the Twin Cities 2000 Mode Choice Model,” 2004.

[3] Y. Sun and R. Xu, “Rail transit travel time reliability and estimation of passenger route choice behavior,” *Transportation Research Record*, no. 2275, pp. 58–67, 2012.

[4] Y.-S. Zhang and E.-J. Yao, “Splitting Travel Time Based on AFC Data: Estimating Walking, Waiting, Transfer, and In-Vehicle Travel Times in Metro System,” 2015.

[5] J. B. Gordon, N. H. M. Wilson, and J. P. Attanucci, “Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2343, pp. 17–24, 2013.

[6] M. Utsunomiya, J. Attanucci, and J. Wilson, “Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1971, pp. 119–126, 2006.

[7] J. J. Barry, R. Newhouser, S. Sayeda, J. J. Barry, R. Newhouser, A. Rahbee, and S. Sayeda, “Origin and Destination Estimation in New York City with Automated Fare System Data,” *Transportation Research Record Journal of the Transportation Research Board*, vol. 1817, pp. 183–187, 2002.

[8] I. Kim, H.-C. Kim, D.-J. Seo, and J. I. Kim, “Calibration of a transit route choice model using revealed population data of smartcard in a multimodal transit network,” *Transportation*, pp. 1–24, 2019.

[9] A. Khani, T. J. Beduhn, J. Duthie, S. Boyles, and E. Jafari, “Using Transit ITS Data For Modeling Network Reliability And The Impact On Passenger Route Choice,” 2014.

[10] S.-H. Lam and F. Xie, “Transit Path-Choice Models That Use Revealed Preference and Stated Preference Data,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1799, pp. 58–65, 2002.

[11] "Estimating Weights of Times and Transfers for Hyperpath Travelers," *Transportation Research Record*, vol. 2284, no. 1, pp. 89–99, 2012.

[12] R. Chapleau, M. Trèpanier, and K. K. Chu, "The Ultimate Survey For Transit Planning: Complete Information With Smart Card Data and GIS," in *8th International Conference on Survey Methods in Transport: Harmonisation and Data Comparability*, 2008.

[13] B. Schaller, *On-board and intercept transit survey techniques.* No. 63, Transportation Research Board, 2005.

[14] E. Frejinger, M. Bierlaire, and M. Ben-Akiva, "Sampling of alternatives for route choice modeling," *Transportation Research Part B: Methodological*, vol. 43, no. 10, pp. 984–994, 2009.

[15] S. Bekhor, M. E. Ben-Akiva, and M. S. Ramming, "Evaluation of choice set generation algorithms for route choice models," *Annals of Operations Research*, vol. 144, no. 1, pp. 235–247, 2006.

[16] H. Kato, Y. Kaneko, and M. Inoue, "Comparative analysis of transit assignment: Evidence from urban railway system in the Tokyo Metropolitan Area," *Transportation*, vol. 37, no. 5, pp. 775–799, 2010.

[17] M. Friedrich, I. Hofsaess, and S. Wekeck, "Timetable-based transit assignment using branch and bound techniques," no. 1752, pp. 100–107, 2001.

[18] C. G. Prato and S. Bekhor, "Applying Branch-and-Bound Technique to Route Choice Set Generation," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1985, no. 1, pp. 19–28, 2006.

[19] R. B. Dial, "A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration," *Transportation Research*, vol. 5, pp. 83–111, 1970.

[20] J. Swait, "Choice set generation within the generalized extreme value family of discrete choice models," *Transportation Research Part B: Methodological*, vol. 35, no. 7, pp. 643–666, 2001.

[21] M. S. Ramming, "Network Knowledge and Route Choice," 2002.

[22] E. Cascetta, A. Nuzzolo, F. Russo, and A. Vitetta, "A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks," in *Proceedings of the international symposium on transportation and traffic theory*, (Lyon, France), pp. 697–711, 1996.

[23] M. E. Ben-Akiva and M. Bierlaire, "Discrete choice methods and their applications to short term travel decisions," in *Handbook of transportation science*, pp. 5–33, Springer, 1999.

[24] P. Vovsha, “Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area,” Tech. Rep. 1607, 1997.

[25] E. Cascetta and A. Papola, “A joint mode-transit service choice model incorporating the effect of regional transport service timetables,” *Transportation Research Part B: Methodological*, vol. 37, no. 7, pp. 595 – 614, 2003.

[26] A. Papola, “Some developments on the cross-nested logit model,” *Transportation Research Part B: Methodological*, vol. 38, no. 9, pp. 833 – 851, 2004.

[27] P. Vovsha and S. Bekhor, “Link-nested logit model of route choice: Overcoming route overlapping problem,” *Transportation Research Record*, vol. 1645, no. 1, pp. 133–142, 1998.

[28] D. McFadden, “Modeling the choice of residential location,” *Transportation Research Record*, no. 673, 1978.

[29] S. Bekhor, T. Toledo, and J. N. Prashker, “Effects of Choice Set Size and Route Choice Models on Path-Based Traffic Assignment,” *Transportmetrica*, vol. 4, no. 2, pp. 117–133, 2008.

[30] Metropolitan Council, “Travel Behavior Inventory (TBI) 2016 Transit On Board Survey - Resources - Minnesota Geospatial Commons,” 2016. `https://gisdata.mn.gov/dataset/us-mn-state-metc-society-tbi-transit-onboard2016`, Accessed 2019-10-29.

[31] R. Kitamura and T. Van Der Hoorn, “Regularity and irreversibility of weekly travel behavior,” *Transportation*, vol. 14, no. 3, pp. 227–251, 1987.

[32] M. Wei, J. Corcoran, T. Sigler, and Y. Liu, “Modeling the influence of weather on transit ridership: A case study from brisbane, australia,” *Transportation Research Record*, vol. 2672, no. 8, pp. 505–510, 2018.

[33] G. Boeing, “OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks,” *Computers, Environment and Urban Systems*, vol. 65, pp. 126–139, 2017.

[34] N. Nassir, M. Hickman, A. Malekzadeh, and E. Irannezhad, “Modeling transit passenger choices of access stop,” *Transportation Research Record*, vol. 2493, no. 1, pp. 70–77, 2015.

[35] W. Jang, “Travel time and transfer analysis using transit smart card data,” *Transportation Research Record*, vol. 2144, no. 1, pp. 142–149, 2010.

[36] T. Brands, E. De Romph, T. Veitch, and J. Cook, “Modelling public transport route choice, with multiple access and egress modes,” *Transportation research procedia*, vol. 1, no. 1, pp. 12–23, 2014.

[37] D. Martin, H. Wrigley, S. Barnett, and P. Roderick, “Increasing the sophistication of access measurement in a rural healthcare study,” *Health & Place*, vol. 8, no. 1, pp. 3–13, 2002.

[38] D. Wood and A. Gatrell, “Equity of geographical access to inpatient hospice care within north west england: A geographical information systems (GIS) approach,” *Lancaster, UK: Institute for Health Research, Lancaster University*, 2002.

[39] H. Jordan, P. Roderick, D. Martin, and S. Barnett, “Distance, rurality and the need for care: access to health services in south west england,” *International journal of health geographics*, vol. 3, no. 1, p. 21, 2004.

[40] F. P. Boscoe, K. A. Henry, and M. S. Zdeb, “A Nationwide Comparison of Driving Distance Versus Straight-Line Distance to Hospitals,” *The Professional Geographer*, vol. 64, no. 2, pp. 188–196, 2012.

[41] J. Gutiérrez and J. C. García-Palomares, “Distance-measure impacts on the calculation of transport service areas using GIS,” *Environment and Planning B: Planning and Design*, vol. 35, no. 3, pp. 480–503, 2008.

[42] A. Khani, M. Hickman, and H. Noh, “Trip-based path algorithms using the transit network hierarchy,” *Networks and Spatial Economics*, vol. 15, no. 3, pp. 635–653, 2015.

[43] A. Mejia, “Tree Data Structures in JavaScript for Beginners ,” 2019. `https://adrianmejia.com/data-structures-for-beginners-trees-binary-search-tree-tutorial`, Accessed 2019-10-31.

[44] W. D. Fan and R. B. Machemehl, “Do transit users just wait for buses or wait with strategies?: Some numerical results that transit planners should see,” *Transportation Research Record*, vol. 2111, no. 1, pp. 169–176, 2009.

[45] T. Brathwaite and J. L. Walker, “Asymmetric, closed-form, finite-parameter models of multinomial choice,” *Journal of Choice Modelling*, vol. 29, pp. 78 – 112, 2018.

[46] R. Tan, M. Adnan, D.-H. Lee, and M. E. Ben-Akiva, “New path size formulation in path size logit for route choice modeling in public transport networks,” *Transportation Research Record*, vol. 2538, no. 1, pp. 11–18, 2015.

[47] S. Vande Walle and T. Steenberghen, “Space and time related determinants of public transport use in trip chains,” *Transportation Research Part A: Policy and Practice*, vol. 40, no. 2, pp. 151–162, 2006.

[48] A. Khani, “Models and solution algorithms for transit and intermodal passenger assignment (development of fast-trips model),” 2013.