Variables influencing types of public transportation based multimodal trips in The Netherlands

Using a Bayesian Belief Network

2023

C.A. Berendsen - 1634275
Construction Management & Engineering
Eindhoven University of Technology

This page is intentionally left blank

Colophon

Title:	Variables influencing types of public transportation based multimodal trips in The Netherlands
Subtitle:	Using a Bayesian Belief Network

Author:	C.A. Berendsen (Nina)
Student ID:	1634275
Email:	c.a.berendsen@student.tue.nl
University:	Eindhoven University of Technology

Graduation Program:	Master Construction Management and Engineering
Graduation date:	18-04-2022
Academic year:	2022-2023
Study load:	45 ECTS

Graduation committee:

Chairman:	Prof. dr. ir. B. de Vries
Fist supervisor:	dr. ing. P.J.H.J. van der Waerden
Second supervisor:	dr. ir. A.D.A.M. Kemperman

This master's thesis has been carried out in accordance with the rules of the TU/e Code of Scientific Integrity.

Frontpage figure: Multimodal trip (RideShark, n.d.)

Preface

In 2015 I passed my high school diploma and as a 16-year-old girl, I found every study interesting. However, the Bachelor of Architecture and Construction Engineering at the Avans University of Applied sciences in ‘s-Hertogenbosch was by far my favorite. After 5 years I graduated but I was not done with studying. Therefore, I applied for the Master of Construction Management and Engineering to deepen my knowledge. Despite the difficult pre-master, especially calculus, I was able to successfully start my master's. During the master, my enthusiasm concerning the built environment kept growing, and my results were better than I ever expected.

This research is the last step in the Master of Construction Management and Engineering (CME) at the Eindhoven University of Technology. In front of you lies the master thesis “Variables influencing types of public transportation based multimodal trips in The Netherlands”. The analyzing method of my thesis was already clear to me due to the course Big Data. Within this course, I learned about the basics of Bayesian belief networks, and my interest enlarged during my master's.

I learned a lot during this research process however, it was not always easy. Therefore, I am grateful for my family and friends that have supported me throughout. They read my chapters, listened, and brainstormed with me but also provided sometimes the needed distraction.

Subsequently, I would like to thank my first supervisor Peter van der Wearden, and my second supervisor Astrid Kemperman. They provided great guidance, feedback, and support throughout the process.

Dear reader, I hope you enjoy reading my thesis!

Nina Berendsen

Eindhoven, April 2023

Table of contents

Summary

Car ownership and use have significantly increased in the last few decades in most Western countries. Many environmental, economic, and social issues have arisen due to the increase in car ownership and use. To reduce the negative impact caused by excessive usage of private motorized transportation, many measures have been proposed and implemented by the government. It is needed to discourage private motorized vehicles and reduce dependency on them, and therefore promote the use of public transportation and/or active travel modes. Multimodal trips can be helpful in such a case, especially for longer-distance trips. In this study, a multimodal trip refers to the use of two or more modes, with at least one of the modes being a public transportation mode to complete a trip from origin to destination. Public transportation based multimodal trips (PTMTs), in particular, can reduce the environmental impact by using more sustainable modes. Therefore, multimodal trips are becoming more popular because individuals look for more effective and environmentally friendly transportation options. Notwithstanding the benefits of multimodal trips, the majority of trips are still conducted by one mode of transportation (unimodal car trips). However, multimodal trips have increased over the past 30 years in The Netherlands. The overall rise, however, only represents a small percentage of 4.5% in 2018 compared to all trips within The Netherlands.

Understanding individual travel behavior as well as how the transportation system is used to meet the derived travel demand is necessary for promoting PTMTs as an alternative to unimodal car trips. In particular, a phenomenon that requires more attention is the organization of transportation modes for the first, main, and last leg of PTMTs. Governments have a lack of knowledge regarding the current individuals that perform PTMTs and the reasons why, when or what kind of PTMTs they perform. Therefore, the purpose of this study is to fill in the knowledge gap about the variables influencing types of public transportation based multimodal trips. The results can help governments implement policies and reduce the load of private vehicles on the roads. From this, the following main research question is drawn up:

“Which variables are influencing types of public transportation based multimodal trips in The Netherlands?”

A Bayesian Belief Network (BBN) is estimated in order to explore and estimate the direct and indirect influencing variables of types of PTMTs. A BBN is a collection of variables that are linked in order to show their interdependencies and give information about their relationships. The variables that are included within this study’s network are categorized into personal, household, environmental, and trip characteristics. Two BBNs are estimated, the first network only includes unimodal car trips and PTMTs. The second network focuses solely on specific types of PTMTs. Network 1 – Trip type shows that the variables student public transportation smartcard, number of cars within the household, and motive are directly influencing the choice of trip type between a unimodal car trip and a PTMT. The indirectly influencing variables are urban density, age, social participation, driver’s license, household income, and household composition. Network 2 – Types of PTMTs illustrates that only the trip characteristics distance (directly), motive (indirectly), and travel time (indirectly) influence the types of PTMTs.

To conclude, network 1 – Trip type helps to give insight into which personal, household, environmental, and trip variables could be stimulated to conduct PTMTs and to decrease unimodal car trips. Network 2 – Types of PTMTs, on the other hand, already shows the persons that conduct PTMTs. This network helps governments not solely focus on decreasing unimodal car trips but also improving PTMTs. Governments should focus on the variables distance, motive, and travel time. They should look at shorting the distance of PTMTs to reduce the use of cars for the first leg. Thereby, the government

should stimulate to walk or cycle for the first or last leg for specific trip motives. For example, offices or schools should be easily accessible by walking or cycling. This could stimulate the use of active modes instead of the car. The last variable that the government should consider decreasing is travel time. By shorting the travel time the use of cars for the first leg will be reduced. Through this knowledge, governments can create suitable policies to reduce car ownership and use. Also, based on the results they can stimulate PTMTs within The Netherlands.

Samenvatting

Het bezit en gebruik van auto's is de afgelopen decennia in de meeste westerse landen aanzienlijk toegenomen. Door de toename van het autobezit en gebruik zijn er veel problemen ontstaan op ecologische, economische en sociale vlakken. Om de negatieve impact van overmatig gebruik van particulier gemotoriseerd vervoer te verminderen, zijn er door de overheid meerdere maatregelen voorgesteld en uitgevoerd. Het is nodig om particuliere gemotoriseerde voertuigen te ontmoedigen en de afhankelijkheid ervan te verminderen, en daarom het gebruik van openbaar vervoer en/of actieve vervoersmiddelen (zoals lopen or fietsen) te stimuleren. Multimodale verplaatsingen kunnen hierbij bijdragen, vooral voor verplaatsingen over langere afstanden. In deze studie verwijst een multimodale verplaatsing naar het gebruik van twee of meer vervoersmiddelen, waarbij ten minste één van de vervoersmiddelen een openbaar vervoermiddel is om de gehele verplaatsing te voltooien. Met name openbaar vervoer gebaseerde multimodale verplaatsingen (OVMV) kunnen de impact op het milieu verminderen door duurzamere vervoersmiddelen te gebruiken. Omdat individuen op zoek zijn naar effectievere en milieuvriendelijkere transportmogelijkheden worden OVMV steeds populairder. Ondanks de voordelen van OVMV, wordt het merendeel van de reizen nog steeds met één vervoermiddel uitgevoerd (unimodale auto verplaatsing). Des ondanks zijn de multimodale verplaatsingen de afgelopen 30 jaar toegenomen in Nederland. De totale stijging vertegenwoordigt echter slechts een klein percentage van 4,5% in 2018 ten opzichte van alle verplaatsingen binnen Nederland.

Inzicht in individueel reisgedrag en hoe het transportsysteem wordt gebruikt om aan de reisvraag te voldoen, is noodzakelijk om OVMV te promoten als alternatief voor unimodale auto verplaatsingen. Een fenomeen dat met name meer aandacht vraagt, is de organisatie van vervoersmiddelen voor het eerste, hoofd- en laatste deel van OVMV. Overheden hebben een gebrek aan kennis over de huidige individuen die OVMV maken en de redenen waarom, wanneer of wat voor soort multimodale verplaatsingen ze maken. Daarom is het doel van deze studie om inzicht te geven over de variabelen die van invloed zijn op specifieke soorten OVMV. De resultaten kunnen de overheid helpen bij het implementeren van beleid en het verminderen van auto's op de weg. Hieruit is de volgende hoofdonderzoeksvraag opgesteld:

> "Welke variabelen zijn van invloed op bepaalde typen openbaar vervoer gebaseerde multimodale verplaatsingen in Nederland?"

Een Bayesian Belief Network (BBN) wordt geschat om de directe en indirecte beïnvloedende variabelen van typen OVMV te formuleren en in te schatten. Een BBN is een verzameling variabelen die zijn gekoppeld om hun onderlinge afhankelijkheden te tonen en informatie te geven over hun relaties. De variabelen die zijn opgenomen in het netwerk van deze studie zijn onderverdeeld in persoonlijke, huishoudelijke, omgevings- en reiskenmerken. Er worden twee BBN's geschat, het eerste netwerk omvat alleen unimodale auto verplaatsingen en OVMV. Het tweede netwerk richt zich uitsluitend op specifieke typen OVMV. Netwerk 1 laat zien dat de variabelen studenten OV, aantal auto's binnen het huishouden en reis motief direct van invloed zijn op de keuze van het reistype tussen een unimodale auto verplaatsing of een OVMV. De indirecte invloeden zijn stedelijke dichtheid, leeftijd, maatschappelijke participatie, rijbewijs, huishoudinkomen en huishoudsamenstelling. Netwerk 2 illustreert dat alleen de reiskenmerken afstand (direct), motief (indirect) en reistijd (indirect) van invloed zijn op de typen OVMV.

Concluderend helpt netwerk 1 inzicht te geven in welke persoonlijke, huishoudelijke, omgevings- en reisvariabelen kunnen worden gestimuleerd om OVMV uit te voeren en unimodale auto

verplaatsingen te verminderen. Netwerk 2 toont daarentegen al de personen die OVMV uitvoeren. Dit netwerk helpt overheden zich niet alleen te richten op het verminderen van unimodale auto verplaatsingen, maar ook op het verbeteren van OVMV. Overheden zouden zich moeten richten op de variabelen afstand, motief en reistijd. Zij zouden moeten kijken naar het verkorten van de afstand van OVMV om het gebruik van auto's voor het eerste gedeelte van de reis te verminderen. Daarbij zou de overheid moeten stimuleren om bij specifieke verplaatsingsmotieven het eerste of laatste gedeelte van de reis te lopen of te fietsen. Zo is het belangrijk dat kantoren of scholen goed te voet of met de fiets bereikbaar zijn. Dit zou het gebruik van actieve vervoersmiddelen in plaats van de auto kunnen stimuleren. De laatste variabele die door de overheid gestimuleerd kan worden is reistijd. Door de reistijd te verkorten wordt het autogebruik voor de heenreis verminderd. Met de resultaten van deze studie kunnen overheden passende beleidsvormen creëren om the auto bezit en gebruik te verminderen. Daarnaast kunnen ze gebaseerd op de resultaten de OVMV in Nederland stimuleren.

Abstract

The rise of car ownership and use has resulted in various environmental, economic, and social issues. Through proposing and implementing policies, governments try to decrease car ownership and use. Public transportation based multimodal trips (PTMTs) can be helpful for this because they combine public transportation with active travel modes and/or the car for the first and last leg of the trip. Due to PTMTs less distance can be traveled by car. However, little previous research includes the first and last leg of PTMTs. Despite the first and last legs can influence the entire trip. Because of this, there is a knowledge gap of variables that influence types of PTMTs. Therefore, this study aims to analyze the variables that influence specific types of PTMTs within The Netherlands. Data concerning personal, household, environmental, and trip characteristics are gained from ODiN and OSM. Subsequently, two Bayesian Belief Networks (BBNs) are estimated to explore the direct and indirect relationships between the included variables and the choice of trip type (unimodal car trip or PTMT) and between the types of PTMTs. The results showed that the variables student public transportation smartcard, number of cars within the household, and travel motive are directly influencing the choice of trip type between a unimodal car trip and a PTMT. Focusing solely on the types of PTMTs, is seems that the variables distance (direct), travel motive (indirect), and travel time (indirect) have an influence. The two BBNs help governments to create policies to reduce car ownership and use and to stimulate PTMTs.

Keywords:

Unimodal car trip; Public transportation Multimodal trip; Bayesian Belief Network (BBN)

List of Figures

List of Tables

1. Introduction

This chapter introduces this study's subject. First, the context concerning the rise of private motorized transportation, policies proposed and implemented regarding private motorized transportation, and multimodal and unimodal trips are described. Subsequently, the problem and research question are addressed and lastly, the research design is explained.

1.1. Context

The dramatic rise in private motorized transportation has aroused the interest of the governments. In order to decrease car ownership and use several policies are proposed and implemented. Whereby, this study focuses on public transportation based multimodal trips (PTMTs). These trips can be beneficial to encourage alternatives to unimodal car trips and could reduce the reliance on private motorized transportation (Nes, 2002).

1.1.1. The rise of private motorized transportation

In the last few decades, car ownership and use have increased dramatically in most Western countries. The EU passenger car fleet increased by 1.2% in 2020 compared to 2019, with 246.3 million cars on the road (ACEA, 2022). The rise in car ownership and use has resulted in various environmental, economic, and social problems. Environmental issues arise from the release of toxic and harmful substances. These substances contribute to global warming, smog, and acid precipitation. Thereby, the production and use of cars require scarce raw materials and energy (Steg, 2003). Economic problems emerged from car use because it is related to decreased accessibility (Steg, 2003). Due to the dramatic growth more traffic congestions appear. Congestions have become a common and persistent phenomenon, particularly in more densely urbanized areas. Hence, peak hours have become longer and more intense, and traffic has come to a halt at times. This has made major city accessibility and travel time reliability prominent issues (Van Exel & Rietveld, 2009). Lastly, car usage poses a threat to the quality of urban life at a social level due to noise pollution, unpleasant odors, local air pollution, and traffic accidents (Steg, 2003). The rise of car ownership and use is also notable within The Netherlands. Kennisinstituut voor Mobiliteitsbeleid (2022) states that in 2020 1.5 million cars are owned by over 450,000 households. Households that own more than one car, together possess 56% of all cars in The Netherlands. Thereby, it seems that in 2021 70% of the total trip kilometers is conducted by car (CBS, 2022a). This is equivalent to each person driving more than 7,000 trip kilometers per year. These car trips have higher levels of emission than alternative transportation modes such as public transportation (Heinen & Mattioli, 2019). Therefore, to decrease the environmental, economic, and social problems that the rise of car ownership and use entails, it is needed to encourage the use of alternatives to car trips. Especially, for longer distances trips.

1.1.2. Policies concerning private motorized transportation

Numerous solutions have been proposed and implemented to decrease the negative impact of excessive use of private motorized transportation, with varying degrees of success (Loukopoulos, 2007). These solutions include investing in the development of private vehicles with fewer harmful emissions, city planning, and infrastructure strategies aimed at reducing congestion and offering alternatives to private cars, and policies aimed at influencing the demand for travel (Redman et al., 2013). Such as the government of London which implemented congestion charges to make people aware of the real costs of car ownership or use in densely populated areas (Tonne et al., 2008). Another example is the government of New York which redefined the streets to create more space for living and active transportation modes (Jonuschat et al., 2015). Likewise, cities have widely implemented parking policies to reduce car use and relieve the parking demand pressure (Yan et al., 2019). In The Netherlands for example, to stimulate drivers of private vehicles to use public transportation, park-and-ride facilities have been created. The car serves as a transportation mode in the first leg to access a train station (or bus station), which is then used for the remainder of the trip. Another example of a transportation service specifically designed to collect and distribute rail passengers is the dedicated shared taxi system 'Treintaxi' in the Netherlands. Also, the Transferium, which is a transfer facility at the city's borders where people can park their cars and access the city center via high-quality urban public transportation, is a concept that focuses on the quality and accessibility of city centers (Jong & Steen, 2001). Therewithal, a technological and behavioral shift is already underway toward a world in which privately owned and operated cars play a smaller role, with various forms of shared mobility taking their place (Hietanen, 2014). The concept of MaaS (Mobility as a Service) aims to provide a more convenient, sustainable, and cost-effective alternative to driving a private car (Jittrapirom et al., 2022; Eckhardt et al., 2018). MaaS combines various modes of transportation and transportation-related services into a single, all-encompassing, and on-demand mobility service (Hietanen, 2014). A MaaS operator provides a diverse menu of transport options to meet a customer's request, including (but not limited to) public transportation, active modes such as walking and cycling, ride/car/bike-sharing, taxi, and car rental or lease, or a combination thereof (Audenhove et al., 2021).

In order to develop policies and strategies that promote sustainable and efficient transportation systems, an understanding of travel patterns, preferences, and needs is acquired. Travel behavior research is essential for governments because it provides valuable insights into how people travel and make transportation decisions. Rietveld & Steg (2000) mention that if it is clear why people travel the way they travel, it is also clear which factors policymakers need to influence to change that travel behavior. Policy measures are therefore more effective if they are better geared to changing important underlying determinants of undesirable travel behavior. Moreover, people are more likely to accept measures if their wishes and possibilities are taken into account (Rietveld & Steg, 2000).

1.1.3. Multimodal and unimodal trips

To encourage the use of public transportation and/or active travel modes, and reduce reliance on private motorized transportation, it is necessary to discourage private motorized transportation. This is where multimodal trips can be beneficial. Multimodal trips are performed with two or more different transportation modes (Basheer et al., 2019; Nes, 2002). Multimodal trips can combine private and public transport since private transportation can be used as an access (first leg) or egress (last leg) mode to conduct a trip within public transport (Krygsman & Dijst, 2001). Krygsman & Dijst (2001) mention that trips conducted with a public transportation mode (train, bus, tram, or metro) are always considered as multimodal trips due to the access and egress part of utilizing the public transportation system. Krygsman & Dijst (2001) state the definition of a public transportation based multimodal trip (PTMT) as follows:

> "A <u>public transportation based multimodal trip</u> refers to the use of two or more modes, with at least one of the modes being a public transportation mode to complete a trip from origin to destination" (Krygsman & Dijst, 2001)

PTMTs are becoming more popular as people seek more efficient and sustainable travel options because they can reduce the environmental impact of transportation by using more sustainable modes (Nes, 2002). In particular, PTMTs can reduce the environmental impact (Heinen & Mattioli, 2019). PTMTs can also improve accessibility for people who may not be able to drive a car, such as the elder or disabled people.

Contrary to PTMTs, unimodal trips are performed with one transportation mode from the origin to the destination. This means that there are no transfers within the trip (Nes, 2002). Figure 1 illustrates the difference between unimodal and PTMTs in accordance with the definition of Krygsman & Dijst (2001). Unimodal trips, performed with a motorized transportation mode, are often considered less sustainable and more polluting than multimodal trips. Furthermore, they can also lead to traffic congestion and parking issues. Of course, this does not apply to unimodal trips performed by walking or cycling. On the other hand, unimodal trips have their advantages as well. They offer more flexibility and convenience as travelers do not have to rely on the schedules of public transportation, which can be more comfortable for long-distance trips (Nes, 2002).

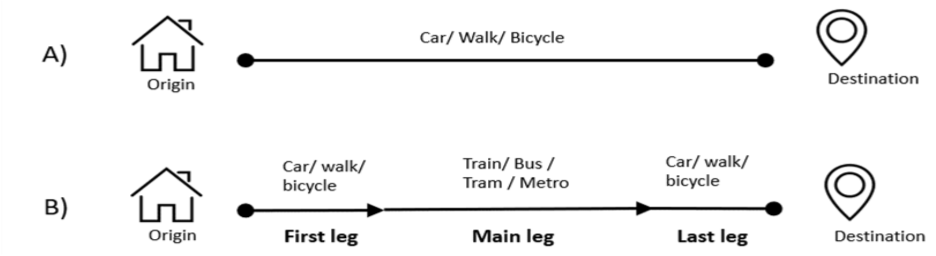
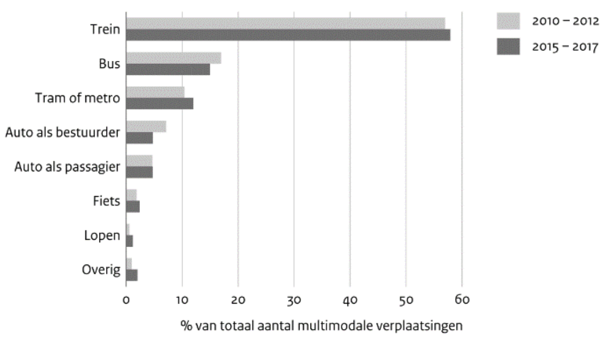


Figure 1: Examples of A) unimodal trip and B) Public transportation based multimodal trip (PTMT) (Own illustration)

Despite the advantages of PTMTs and disadvantages of unimodal trips, unimodal trips are still the majority conducted. However, in the past 30 years, there is a growth in conducted multimodal trips. Nes (2002) found that 2.9% of all trips are multimodal trips within The Netherlands. This means there is an increase of 25% compared to 1992. This growth in conducting multimodal trips is also confirmed by CBS & OViN (2020). They state that in 2010 3.1% of all trips (unimodal and multimodal) are multimodal and in 2018 this increased to 4.5%. When analyzing multimodal trips, the train seems to be the most used main transportation mode with more than 55% followed by the bus (17%) and metro (10%) which is illustrated in Figure 2 (CBS & OViN, 2020). Subsequently, CBS & OvIN (2020) describe that 15.3% of the total kilometers of all trips are conducted as multimodal trips. This is in comparison to the year 2010 an increase of 3.6%. this follows from Figure 3. Note, these multimodal trips can also be conducted without public transportation. CBS did not exclude other multimodal trips than PTMTs.



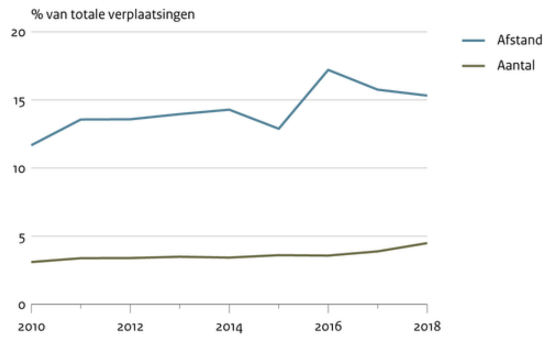


Figure 2: Split of the main mode of multimodal trips in the Netherlands (CBS & OViN, 2020)

Figure 3: Multimodal trips in the Netherlands (CBS & OViN, 2020)

Nes (2002) found that multimodal trips last on average 45 kilometers in total, which is more than 4.5 times longer than all types of unimodal trips. Thereby, the average distance of a unimodal car trip is 19 kilometers (Bakker, 2019). When traveling more than 10 kilometers, multimodal transportation seems a feasible option, and when traveling more than 30 kilometers, it becomes an intriguing choice in comparison with unimodal trips (Nes, 2002). Nevertheless, the overall increase of multimodal trips found by CBS & OvIN (2020) is only a small percentage of all trips within The Netherlands.

1.2. Problem statement and research question

This section will address the research problem, the main research question, and the research design. Furthermore, the relevance of the study is described followed by the reading guide.

1.2.1. Problem statement

Promoting multimodal transportation as an alternative to unimodal car trips requires an understanding of individual travel patterns as well as how the transportation system is used to meet the derived travel demand (Krygsman & Dijst, 2001). Most organizations that work on improving the public transportation system do not include the transfers of multimodal trips and focus mostly only on the main mode (public transport) (Guo & Wilson, 2011). In particular, the organization of transportation modes for the first, main, and last leg of multimodal trips is a phenomenon in travel behavior studies that seems to require more attention since the choice of travel mode and/or the available service at one leg of the trip influences the entire trip (Basheer et al., 2019; Nes, 2002; Krygsman & Dijst, 2001).

Despite the constraints of multimodal trips such as longer travel times, transfers, and long first and last leg distances (Krygsman & Dijst, 2001), multimodal trips (in particular when public transportation is used) can decrease car use and ownership that resulted in various environmental, economic, and social problems (Steg, 2003; Van Exel & Rietveld, 2009; Heinen & Mattioli, 2019). However, governments have a lack of knowledge regarding the current individuals that perform PTMTs and the reasons why, when or what kind of PTMTs they perform (Basheer et al., 2019). Therefore, this research aims to address the lack of knowledge concerning the variables that influence types of PTMTs. A Bayesian Belief Network (BBN) is estimated. This provides an overview of the relationships of variables influencing PTMTs. The network can also predict the probabilities when new information is entered within the network (Verhoeven, 2010). This study's results can help governments create and implement policies and reduce the load of private vehicles on the roads (Basheer et al., 2019).

1.2.2. Research question(s)

The following main research question is drawn up based on the problem statement:

> *"Which variables are influencing types of public transportation based multimodal trips within The Netherlands?"*

To answer the main research question, the following research questions are set up:

- *SQ1: What methods are used to research multimodal trips and what are the findings?*
- *SQ2: Which variables directly and indirectly influence the choice of trip type between a <u>unimodal car trip and a public transportation based multimodal trip</u>?*
- *SQ3: Which variables directly and indirectly influence <u>types of public transportation based multimodal trips</u>?*

1.3. Research design

To conduct this study, a quantitative research will be performed. The methodology will be discussed in the following sections. The complete research design is illustrated in Figure 4.

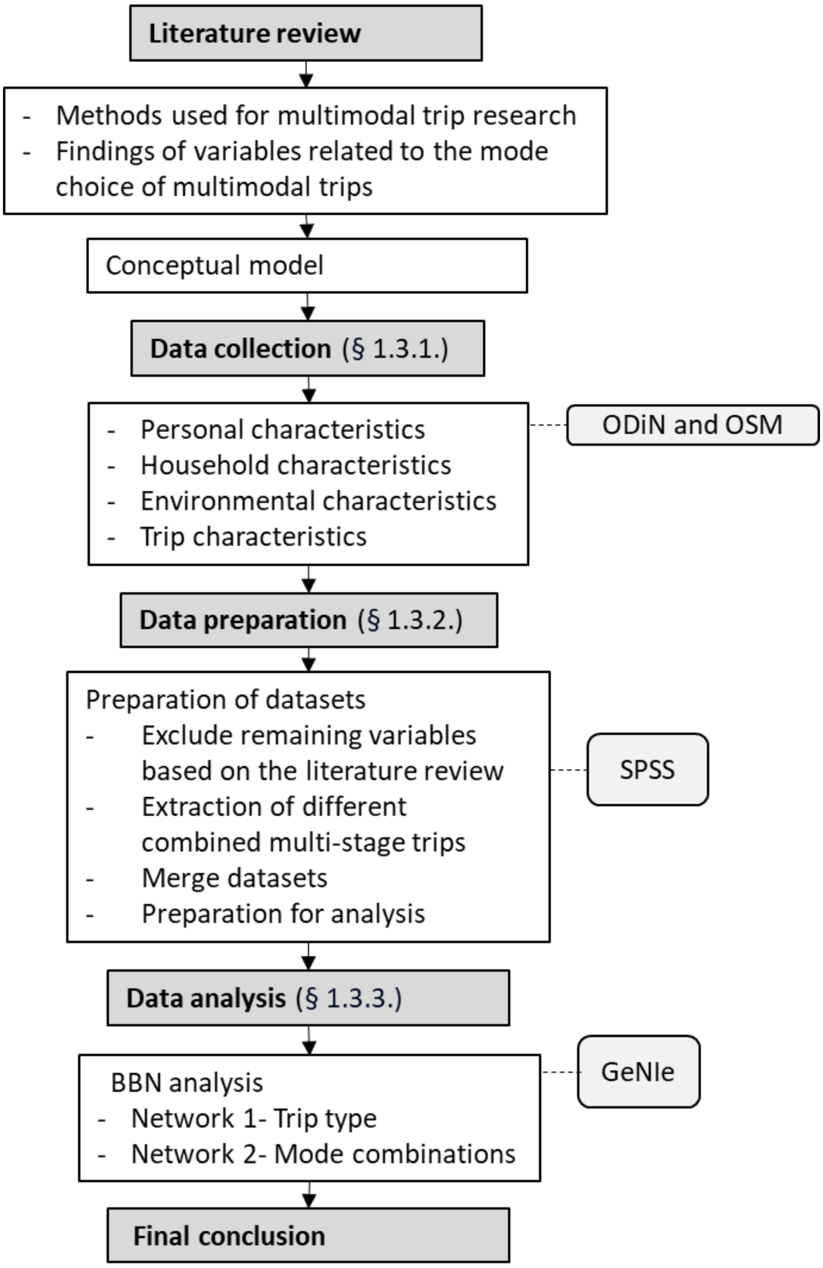


Figure 4: Research Design

1.3.1. Data collection

Within this study, secondary research is used to gain a large sample size. Hereby, no costs are made and no time is consumed for collecting primary data. The personal, household, and trip characteristics will be extracted from the data set of ODiN (Onderweg in Nederland) and the environmental characteristics will be received from OSM (OpenStreetMap).

The study area of this research is The Netherlands like the data set of ODiN. Centraal Bureau voor de Statistiek (CBS) researches annually the daily mobility of inhabitants of The Netherlands. ODiN is conducted by CBS on behalf of the Ministry of Infrastructure and Water Management, multiple policies and research organizations, society, and Statistics Netherlands. The data is used for improving traffic safety, public transportation, the environment, and preventing traffic jams (CBS, 2022b). For this research, the data set of ODiN 2021 is used, since this is the latest available data set. The data set is received from DANS EASY (Data Archiving and Networked Services). The environmental data of the infrastructure of The Netherlands, which comprises among other things as train stations and bus stops, will be received from OpenStreetMap (OSM)(http://www.openstreetmap.org). This is a collaborative project to create a free, editable map of the world. OSM provides a wide range of information, including street maps, building outlines, points of interest, and administrative boundaries. The data is widely used by many organizations, such as governmental, non-profit, or commercial organizations (Arsanjani et al., 2013).

1.3.2. Data preparation

Through a literature review, variables will be selected that could be influencing the mode choice of (public transportation based) multimodal trips. Characteristics such as personal, household, trip, and environmental are taken into account. The remaining variables will be deleted from the data set of ODiN and OSM.

The data preparation is executed in SPSS (Statistical Package for the Social Sciences). This is a statistical software platform for complex statistical data analysis. The respondents that do not conduct a unimodal car trip or a multimodal trip with public transportation as travel mode in the main leg will be excluded from the data set. Unimodal walk or cycle trips should not compete with PTMTs since they do not result in various environmental, economic, and social problems. Therefore, this study focuses solely on unimodal car trips and PTMTs. Furthermore, public transportation helps to reduce the negative effects of unimodal car trips hence, only PTMTs are included where the car can only be a first leg or last leg transportation mode. The last step of the data preparation is the preparation of the data for the analysis technique. The included variables will be categorized based on the variable distribution.

1.3.3. Data analysis

The data analysis will be performed after the data preparation. The modeling approach utilized involves the use of a Bayesian Belief Network (BBN). According to Neapolitan (1990), a Bayesian Belief Network is a graphical model that illustrates the conditional dependencies and interrelationships among a group of variables. This data mining analysis technique is used to estimate the variables that influence PTMTs and their direct and indirect impacts on types of PTMTs. The BBN is well-suited for addressing research questions and testing hypotheses related to travel mode choice (Kemperman & Timmermans, 2014a).

The analysis will be conducted with the software GeNIe (version 2.1) (BayesFusion, 2022). GeNIe Modeler is a commercial software package for creating and analyzing Bayesian networks however, a free version is available for academic or non-commercial use (GeNIe Academic) which is used for this research. The results of the analysis will be discussed and hereafter the conclusion of this research.

1.4. Reliability and Validity

This research only uses open data sources, which are publicly accessible. Due to this, large sample-sized data is gained, without making any effort in time and money. Also, publicly accessible data can be accessed and analyzed by other researchers, allowing for the replication of findings and increasing the confidence in the results (Kitchin, 2014). Thereby, Kitchin (2014) mentions that open data is often subject to scrutiny and review, which increases the transparency and reliability of the data. Within this study, the data is retrieved from ODiN and OSM as described in 1.3.1. Data collection. The data of ODiN is based on sample research. This could imply that statements at the population level can have a certain degree of uncertainty since only a part of the total population is included. However, the data is more reliable when using a larger sample size (CBS, 2022b). Thereby, the respondents have to report their daily mobility themselves. In conclusion, ODiN's reliability is based on how well respondents accurately remember the study's variables, including the number of trips, starting and ending location, travel time, distance traveled, etc. Hence, the respondents are provided with a 'help diary' (also known as a memory jogger). This increases the accuracy of the data (KiM, 2017). Looking at the data of OSM, the data reliability should be considered since the OSM project is a free open editable map for individuals. As the additions are not being monitored, it is difficult to determine with certainty that the data is trustworthy and of high quality (Arsanjani et al., 2013). The individual who entered the data is responsible for its validity and reliability (Jakobs & Mitchel, 2020). However, these individuals (OSM contributors) are identified into five groups; neophytes, interested amateurs, expert amateurs, expert professionals, and expert authorities. Studies (Arsanjani et al., 2013; Jakobs & Mitchel, 2020) find that more experienced individuals provide high-quality data that is reliable and accurate. With regard to this research, the validity of the extracted multimodal trips is based on public transportation. Thus follows that respondents have precise information regarding their conducted trip such as travel time and distance. Therefore it is assumed that the included variables are valid. Taking the sample size into account in this research, the data set concerns a sample of The Netherlands. However, the fewer cases of PTMTs in comparison with unimodal trips (car, bicycle, walk, etc.) should be considered. Though it represents the current population of The Netherlands that conducts PTMTs. The validity of this research results is based on the process of a Bayesian Belief Network. Based on the literature research variables are selected that are included in the BBN. Within GeNIe constraints can be set to the network. The created network excludes variables that do not seem to influence PTMTs. From this, it will be assumed that the network represents valid results. Thereby, the results are compared with existing literature, and conclusions are drawn up.

1.5. Relevancy

This thesis has a scientific contribution to the field of transportation engineering and transportation planning. Mode choice modeling involves the development of mathematical models and analytical tools to understand and predict the travel behavior of individuals and groups, specifically their choice of transportation mode (such as walking, cycling, driving, or using public transit) for a given trip. Mode choice modeling is an important area of research for transportation professionals, urban planners, and policymakers, as it can inform the design and evaluation of transportation systems, infrastructure, and policies, and help to improve mobility, accessibility, sustainability, and safety. The relevancy of PTMTs lies in their potential to enhance travel efficiency, convenience, and sustainability while promoting social and environmental well-being (Ulloa et al., 2018). This research provides the important variables that influence PTMTs within the Netherlands through a BBN (Bayesian Belief Network). From literature research these variables are selected and concern personal, household, environmental, and trip characteristics. With this knowledge, policymakers and urban planners can make informed decisions to create a well-designed public transportation system and implement specific strategies or policies to promote PTMTs and decrease the environmental, economic, and social problems caused by unimodal car trips.

In addition, a Bayesian Belief network (BBN) is used as analysis technique. This a mathematical model and analytical tool used for probabilistic reasoning and decision-making. It is a graphical model that represents the probabilistic relationships among a set of variables and their conditional dependencies (Verhoeven, 2010). Compared to other techniques commonly used for multimodal trip mode choice research, a Bayesian Belief Network has the potential to offer the advantage of being able to include direct and indirect patterns of causal relationships (Arentze & Timmermans, 2008). Although Bayesian Belief Networks have shown promise in mode choice research (Kemperman & Timmermans, 2014a, 2014b; Ma, 2015; Yankaya, 2010), they are not currently a commonly used analysis technique in the field since they are more computationally intensive and require a greater level of expertise to develop and interpret (D. Lee et al., 2018). Therewithal, MNL and nested logit models have a long history of use in transportation research (Basheer et al., 2019; Hu et al., 2018; Kim et al., 2007) and are well-established, which can make them more familiar and easier to use for many researchers. That being said, there is still little research done concerning BBNs and mode choice modeling (Zhu et al., 2018). However, they may become more commonly used in the future as their capabilities and applicability continue to be explored and refined. This research contributes to the slight literature on applying a BBN for formulating and estimating the choice of types of PTMTs.

1.6. Reading guide

This research is divided into five chapters. The first chapter is an introduction followed by a literature research (Chapter 2). The literature research covers the definitions of multimodal trips (Section 2.1.), previous research on multimodal trips (Section 2.2.), and related variables of mode choice (Section 2.3). Thereafter, the research's methodology, including data collection, preparation, and analysis, is described in Chapter 3. Subsequently, the results are presented in Chapter 4 and are divided into three subsections. First BBN network 1 is described in Section 4.1. In Section 4.2. BBN network 2 is described and the conclusion of these networks is discussed in Section 4.3. A discussion of the findings (Section 4.4.) brings Chapter 4 to a close. Lastly, Chapter 5 concerns the conclusion, limitations, and recommendations.

2. Literature research

Within this chapter, the definitions of multimodal trips, previous multimodal trip research, and related variables to mode choice and multimodal mode choice are reviewed. Lastly, a conceptual model is created that presents the included variables within this study.

2.1. Multimodal trips definitions

Within the literature, different definitions are used for multimodal trips. As mentioned in the introduction of this research, Krygsman & Dijst (2001) mentioned that trips conducted with a public transportation mode (train, bus, tram, and metro) are always considered as multimodal trips (Figure 2). This definition of public transportation based multimodal trips (PTMTs) is also applicable to this research, therefore:

> "A Public transportation based multimodal trip refers to the use of two or more modes, with at least one of the modes being a public transportation mode to complete a trip from origin to destination" (Krygsman & Dijst, 2001)

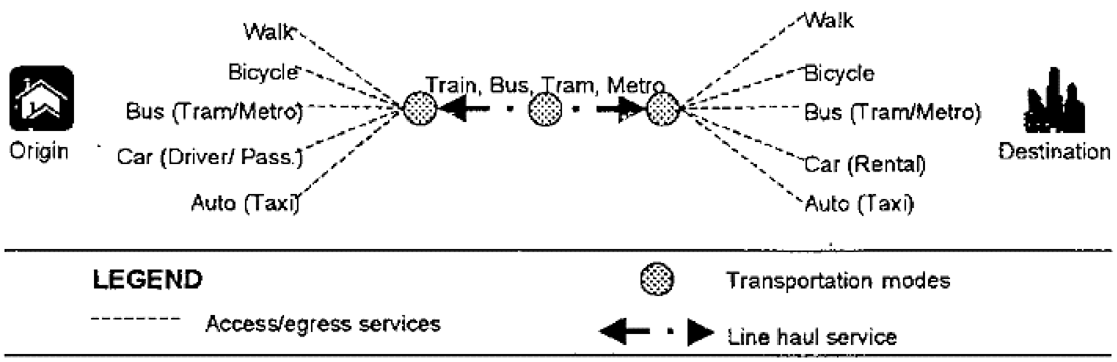


Figure 2: Diagrammatic outline of a multimodal trip (Krygsman & Dijst, 2001)

Krygsman & Dijst (2001) suggested that these trips are divided into several parts, and thus defined as a trip chain, due to the access (first leg) and egress (last leg) part of utilizing the public transportation system. This definition is also used by Bovy & Hoogendoorn-Lanser (2005) and Basheer et al. (2019) since they mention that multimodal trips refer to trips that involve a combination of either multiple public transportation modes or public transportation and private modes (such as car and bike). Arentze & Timmermans (2004) and Hoogendoorn-Lanser & Nes (2005) agree but with the note that the transportation mode is not necessarily motorized. However, Nes (2002) differs from the definition above. He states that a trip that involves transferring between two or more different modes of transportation is a multimodal trip whereas walking is not included as a change of mode. Nes (2002) states that walking is an integral part of most trips, in particular when traveling to and from transit stops or parking areas when using a car, even if the walking distance is short. Hence, it can be considered a universal component at the beginning and end of any trip and is not treated as a separate mode in defining a multimodal trip. For example, an individual who walks to the bus stop takes the bus, and then walks to their final destination is conducting a unimodal bus trip. In contrast, some studies such as Zumkeller et al. (2005), Kuhnimhof et al. (2006), and Blumenberg & Pierce (2014) do not focus on one trip, they look at different periods (one week or one day). These studies use these time periods because they are not interested in one trip. With this method, the likelihood of more persons using multiple modes is higher. Zumkeller et al. (2005) state the definition of this research of PTMTs as inter-modality which refers to the course of one trip which is performed with various transportation modes. Also, Basheer et al. (2019) and Gao et al. (2020) uses another designation for PTMTs, namely multi-stage trips. Despite the difference in name, the definition corresponds with the definition described earlier in this study.

Furthermore, the part of the (public transportation based) multimodal trip where one transportation mode is used is differently designated. Within this research, these parts are named trip legs as Nes (2000), Bovy & Hoogendoorn-Lanser (2005), Hoogendoorn-Lanser & Nes (2005), and Abenoza et al. (2019). The first, main, and last trip legs are also known as the access leg, main leg, and egress leg (Krygsman & Dijst, 2001). However, Chong et al. (2011) and Basheer et al. (2019) describe the legs as miles. The term "first mile" refers to the trip from the origin point (usually home) to a transit station, while the "last mile" refers to the trip from a transit station to the final destination (such as the workplace). The designation differs however the definition is similar.

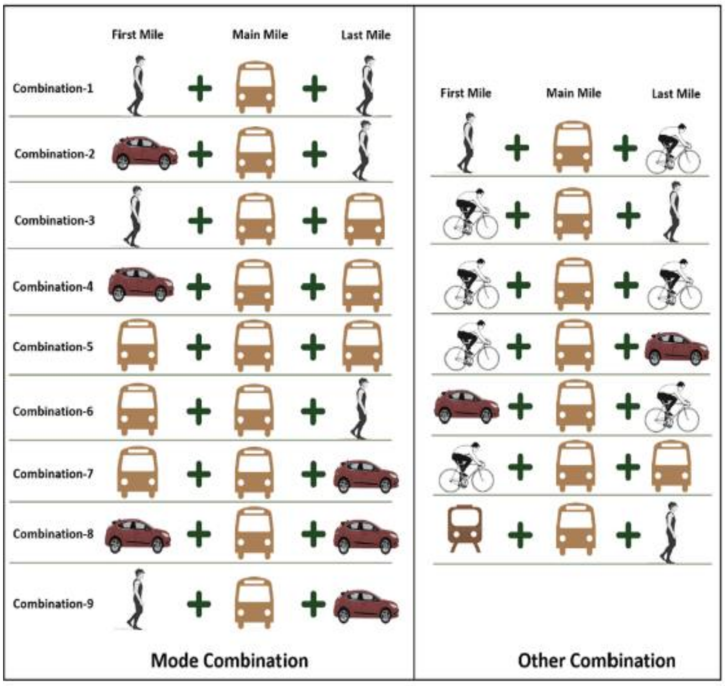
2.2. Previous multimodal trips research

Previously done research concerning (public transportation based) multimodal trips is reviewed. First, the difference approaches between the three selected case studies are discussed, and afterward the used analyzing techniques within the literature.

2.2.1. Case studies

Most studies conducted concerning (public transportation based) multimodal trips are case studies. A case study is a research method that involves an in-depth examination and analysis of a particular individual, group, organization, or situation (Crowe et al., 2011). Crow et al. (2011) mentioned that case studies can be used to investigate a range of topics, from understanding the behavior of individuals to examining the effectiveness of interventions or policies. A case study approach can be useful for multimodal trip research because it allows for an in-depth investigation of the travel behavior and experiences of individuals or groups in a real-life context. By analyzing a specific case or example, researchers can gain a better understanding of the complexity of multimodal travel behavior and identify opportunities for improving the design and implementation of multimodal transportation systems (Crowe et al., 2011). Below, three case studies are discussed to highlight the different approaches and their results. The study of Basheer et al. (2019) is similar to this research focusing on mode choice combinations of PTMTs. Therefore, it is useful to know which variables they included, what they found for England, and if this differs from the Netherlands. Kim et al. (2007) study is chosen to review since it focuses only on the first and last trip leg of train passengers. It could be interesting to see the difference or similarities of findings between studies that include the entire trip including the first, main, and last leg, and studies that include only the first and last leg. The last case study selected is from Ma (2015). He included unimodal car trips as well as multimodal trips (car+ public transport). This study also compares unimodal car trips and multimodal trips however, the PTMTs are differently defined and only concern the car in combination with public transport, which is interesting to see if there are different findings. Thus, the three case studies all include a part of this research and are therefore chosen to be reviewed.

Basheer et al. (2019) conducted a case study of England. They researched variables that could influence mode choice combinations. The used data was extracted from the National Travel Survey (NTS) and the variables trip purpose, trip distance, household income, household structure, age, vehicle ownership, gender, job status, and education are included as independent variables in the model analysis. The dependent variable mode choice combinations is generated from the NTS. They select only trips with public transportation as a mode for the main leg. Thereby, the mode combinations that have less than 50 cases are joined into the category other. The generated mode combinations can be seen in Figure 3. Thereby, Figure 4 shows the descriptive statistics of the mode combinations. The results of this research show that the variable trip purpose is significantly related to the mode choice of the combinations. They found that people do not prefer to conduct PTMTs for work-related trips. Thereby, as the trip distance increases the probability of a combination involving a car at the first or last leg of the public transportation based multimodal trip increases as well. From this can be stated that trip distance is significantly related to the mode choice at the legs of PTMTs. The variable household income influences the probability of a combination involving public transport or car for the first and last leg. In Addition, families with children do prefer the car or public transport for the first and last leg instead of walking as a transportation mode. Basheer et al. (2019) also found that gender and age influence the choice of a mode combination. Individuals younger than 29 years are less likely to choose any of the mode combinations, which is also applicable for males compared to females. Furthermore, the variable job status is too related to the mode combinations. Individuals that are not working full or part-time have a greater probability to choose a mode combination. Subsequently, vehicle ownership is strongly related to the probability of a mode combination. The probability of choosing a combination involving walking declines when the individual has access to a car. Lastly, the variable education is too associated with certain mode combinations. It seems that higher educated individuals are more likely to choose mode combinations that involve walking in the first or last trip leg. Remarkable is that cycling is not used or limited for the first or last trip leg and is therefore included as others.



Mode Combination	N	Percent
W + PT + W (i1)	594	28.9
C + PT + W (i2)	130	6.3
W + PT + PT (i3)	132	6.4
C + PT + PT (i4)	197	9.6
PT + PT + PT (i5)	423	20.6
PT + PT + W (i6)	129	6.3
PT + PT + C (i7)	182	8.9
C + PT + C (i8)	99	4.8
W + PT + C (i9)	87	4.2
Others (i10)	269	4.0

Figure 3: Mode combinations (Basheer et al., 2019)

Figure 4: Descriptive statistics for mode combinations (Basheer et al., 2019)

Kim et al. (2007) researched variables that influence the mode choice of light rail riders in the St. Louis metropolitan area (US). They conducted a survey focusing on trips from the transit station to home or vice versa. The independent variables included in this study are age, gender, licensed driver, employment status, race, household income, time of day, vehicle available for the trips, trip purpose, station location, station neighborhood land use, bus availability, distance to transit station by mode, and crime. The dependent variable mode from/ to transit station is divided into pick-up/drop-off, drive and park, bus, and walk. The descriptive statistics of the independent variable are shown in Figure 5. They found that individuals younger than 25 are more likely to be picked up or dropped off at stations however, the ratio between bus, drive, and parking is not influenced by age. The variable gender does significantly influence mode choice. It seems that females are associated with the mode choice bus, except when they have access to a private vehicle then there is a reduction in the mode choice of walking and taking the bus. Thereby, the mode choice of females depends on the time of the day. The picked-up or dropped-off share increases at nighttime (7 p.m. and later). This is also linked to the crime level and even for higher crime rate stations, females favor pick-up/drop-off independent of the variable time of the day. However, the station's reported crime rate is linked to a decrease in the share of people who choose to use pick-up/drop-off services as opposed to driving and parking, taking the bus, or walking. The variables of vehicle availability and driver's license are significantly related to a rise in the percentage of people who opt for driving and parking over pick-up/drop-off, taking the bus, or walking, indicating a strong preference for private vehicle usage among individuals. However, when there is a bus stop within 805 meters of the residence of the individual, the share of the bus used increases. Regarding racial influences, African-Americans are linked to a higher proportion of bus usage and, to a lesser degree, an increase in pick-up/drop-off usage compared to driving and parking or walking, in comparison to other racial groups. However, Kim et al. (2007) mentioned that this could also be due to income differences between household sizes. It appears that lower-income increases the share of buses used, and vice versa for other modes and higher income. Also, According to the findings, full-time employees who use transit have a higher percentage of pick-up/drop-off usage in comparison to other modes, indicating that they are likely daily commuters who do not require a vehicle during the day, as other family members can utilize it. Conversely, full-time students are linked to a greater proportion of bus and walking usage, as opposed to driving and parking or pick-up/drop-off. Subsequently, two variables appear not significantly influence the mode choice of light rail transit (LRT) riders. The study analyzed whether the trip purpose, either work or school-related vs. personal or recreational, influenced the mode choice of LRT riders between their homes and stations. However, the variable was not statistically significant, indicating that it did not significantly impact their mode choice. Additionally, the study examined whether the station's location in an urban or suburban area affected the riders' mode choice. However, this variable was also found to be statistically insignificant, largely due to the multicollinearity between this variable and the remaining land use variables included in the model.

Variable		Number of observations	Percent
Mode from/to transit station	Pick-up/drop-off	37	9.1
	Drive and Park	219	53.8
	Bus	117	28.7
	Walk	34	8.4

Figure 5: Descriptive statistics (Kim et al., 2007)

Ma (2015) examines the daily mobility mode choice behavior of cross-border workers in Luxembourg, with a particular emphasis on their use of multiple modes of transportation (such as park and ride) and their behavior of multimodal trips. Within this study, mode choice is classified into walking, bike/motor, car, public transport, and multimodal (car+ public transport) as shown in Figure 6. The includes independent variables in this study are: gender, full or part-time job, relationship status, number of children within the household, number of cars within the household, possession of driver's license, household income after tax, trip purpose, number of trips, departure time of the trip, trip distance, trip destination, and travel time. He found that mode choice directly is influenced by travel time, travel distance, parking pace, number of cars in the household, possession of a driver's license, and presence of children in the household. In addition, he stated that individuals with full-time jobs are more likely to use cars as their mode of transportation. Factors such as being in a couple, owning multiple cars, having complex travel itineraries, and taking long-distance trips have a positive correlation with car usage. Conversely, households without children, higher household incomes, longer travel times, and difficulty finding parking spaces are more likely to use public transportation.

Variable	Definition	Mean
Mode_work	Walk	5.74%
	Bike and motorcycle	0.18%
	Car	79.7%
	Public transport	4.1%
	Car + Public transport	7.5%

Figure 6: Descriptive statistics of the sample (Ma, 2015)

From this review, it is shown that these three studies all include different variables based on the scope and area of the study. Looking at the dependent variable, they all define the variable differently based on their study scope and their descriptive statistics. For example, Basheer et al. (2018) did not separately research cycling as a transportation mode for the first and or last leg of the trip since there were too few cases. Also, Kim et al. (2007) did not include cycling as a transportation mode for the first or last trip leg. It is assumed for The Netherlands this will differ due to the Dutch cycle culture. Therefore, it is needed to generate this study's dependent variable also on the descriptive statistics like the other three studies. Thereby, the independent variables included in these three studies are stated in Table 1. It is necessary for this research to not exclude variables on the fore hand based on findings of other studies since this can significantly differ. More in-depth literature research is needed concerning variables that could influence travel mode choice independently of the scope of the research. This will be elaborated on in 2.3. Related variables mode choice.

Table 1: Independent variable included by Basheer et al., 2019; Kim et al., 2007; Ma, 2015

Personal characteristics	Household characteristics	Environmental characteristics	Trip characteristics	Other
- Age - Gender - Vehicle ownership - Job-status - Education - Driver's license - Race - Relationship status	- Income - Structure - Vehicle availability - Number of children - Number of cars	- Land use of destination - Bus availability - Parking availability at destination	- Purpose/ motive - Distance - Number of trips - Departure time - Destination of trip - Travel time	- Crime

2.2.2. Analyzing Methods

Primary methods used for many years in travel mode choice modeling have been the multinomial logit (MNL) model and its variations, such as the nested logit model (e.g. Basheer et al., 2019; Dissanayake & Morikawa, 2010; Hu et al., 2018; Kim et al., 2007; Lu et al., 2015; Wen et al., 2012). A multinomial logit (MNL) analysis is a statistical method used to model and predict the choices that individuals make among several alternatives in a given situation. It is a type of regression analysis that estimates the probability of choosing each alternative based on various independent variables. The MNL model assumes that individuals make choices based on the utility they derive from each alternative and that the probability of choosing an alternative is proportional to its utility relative to the other alternatives (Hausman & McFadden, 1984). The MNL model offers several advantages, such as its elegant closed-form mathematical structure and the ability to interpret model estimation results based on random utility theory (Wang & Ross, 2018). However, Wang & Ross (2018) also mentioned that the primary drawback of the MNL model is its stringent statistical assumptions, such as the independence of irrelevant alternatives (IIA). These assumptions often necessitate meticulous model specification and well-organized data structures. Another limitation of the MNL model is that it may result in biased estimation when the data set is highly imbalanced, meaning that different classes are represented unequally. This may lead to a higher prediction error for classes with smaller shares.

The nested logit model is considered a generalization of the MNL model (Hausman & McFadden, 1984). That allows for correlation among alternatives within subsets or "nests", which solves the drawback of the IIA of an MNL. The nested logit model assumes that the alternatives within each nest are more closely related to each other than they are to alternatives in other nests, and thus the utility derived from each alternative is influenced not only by its own attributes but also by the attributes of the other alternatives within the same nest. This allows for a more flexible and realistic representation of decision-making behavior in situations where alternatives can be grouped into subsets or categories (Hoffman & Duncan, 1988).

A Bayesian Belief Network (BBN) is a less-used analyzing technique. According to Neapolitan (1990), a Bayesian Network is a graphical model that represents the conditional dependencies and interrelationships among a set of variables. Yankaya (2010) mentioned that this technique is useful when a researcher wants to model and analyze complex relationships among variables that may be uncertain or incomplete. BBNs allow for the incorporation of expert knowledge and can handle situations where the relationships among variables are not well understood or the data are incomplete. BBN is useful for decision-making situations where there is a need to understand the potential impact of different variables and the uncertainty associated with those variables. Till now BBNs have found applications in various fields, including environmental modeling and management, pattern recognition and classification, and medical diagnoses (Aktas et al., 2007; Bromley et al., 2005; Kahn et al., 1997; S. Lee & Abbott, 2003), operational risk management in banks (Cornalba & Giudici, 2004), as well as resource planning and management. However, within transportation modeling BBNs are rather limited used (Yankaya, 2010). Kemperman et al. (2019) mention that it is a considerable challenge to include a large number of variables in a network and identify meaningful relationships. Typically, the structure of these relationships is unclear, such as mediating or interaction effects, and the variables are often correlated. Therefore, selecting variables and determining an appropriate structure for explanatory variables can be difficult. However, using a Bayesian Belief Network (BBN) approach can help overcome these difficulties by simultaneously deriving and representing all direct and indirect relationships among the variables in the set. It is worth noting that all variables used in the BBN estimation are categorized and discrete, which is an advantage over other methods like structural equation modeling that can only handle continuous variables (Theo Arentze & Timmermans, 2008). This is particularly useful when dealing with a mix of continuous and discrete variables, such as gender and household type, which are often included in a model.

Despite these advantages, a BBN is less used. BBNs are a relatively newer approach that has not been widely adopted in transportation research, while MNL and nested logit models have been widely used in this field for several decades and are well-established methods (Yankaya, 2010). However, the purpose of a BBN is fundamentally different compared to a Discrete Choice Model such as MNL. BBNs aim to analyze complex relationships between variables. It is a statistical technique used to test theories about causal relationships between multiple variables by estimating the strength and direction of relationships among them (Yankaya, 2010). Discrete Choice Models on the other hand aim to analyze choices made by individuals or groups among a set of alternatives. These models aim to explain the variables that influence decision-making and predict the likelihood of choosing a particular option (Hausman & McFadden, 1984). Thereby, BBNs can be more computationally intensive than MNL or nested logit models, particularly when the network structure becomes more complex. If a large set of variables is used within the BBN, a complex network could be found. Hence, it could be hard to predict the outcomes of a dependent variable (Kemperman & Timmermans, 2014a). Also, the interpretation of BBNs can be more challenging than that of MNL or nested logit models. BBNs provide probabilistic relationships between variables, which may not be as straightforward to interpret as the coefficients in MNL or nested logit models (Chen, 2014). This can make it more difficult to explain the results of a BBN to decision-makers or stakeholders who are not familiar with probabilistic modeling approaches (Chen, 2014). On the other hand, Ma (2015) found similar percentages of the observed and predicted choices when comparing an MNL and BBN. Though, he mentioned that further research is needed to improve the performance. He suggests including different discretization schemes for continuous variables, and other relevant variables or using a data-driven causal structure learning approach (Ma, 2015).

2.3. Related variables mode choice

As presented in the case studies several variables could be related to the mode choice of (public transportation based) multimodal trips. This paragraph will discuss the variables more in-depth. First, an overview is given of possible related variables of travel mode choice and subsequently, variables especially concerning (public transportation based) multimodal trips are reviewed.

2.3.1. Travel mode choice

The variables that could be related to travel mode choice are classified into personal, household, environmental, and trip characteristics and will be reviewed below. This classification can help to identify the key factors that are most important in influencing mode choice and to develop models that accurately predict and explain mode choice. Studies often classify the variables like Basheer et al. (2018) into personal, household, vehicle, and trip characteristics or Ma (2015) into socio-demographic, spatial, and trip characteristics.

2.3.1.1. Personal characteristics

Travelers' personal characteristics (e.g. gender, income, age, availability of the personal vehicle, and access to alternative modes) can significantly influence travelers' mode choice behavior (Cho, 2013; Nes, 2002; Racca & Ratledge, 2003). Many studies include the following variables: gender (Kim et al., 2007; Senbil et al., 2009), age (Almasri and Alraee, 2013; Schlossberg, 2013; Liu et al., 2016; Racca and Ratledge, 2003), income (Jong, 2008; Racca and Ratledge, 2003; Tyrinopoulos & Antoniou, 2013), education, and employment (Liu et al., 2016). For example, dependency on car travel is related to the work status of individuals (Ha et al., 2020) and Ababio-Donkor et al. (2020) observed a significant impact of age, income, and education on the likelihood to use active travel modes.

2.3.1.2. Household characteristics

Variables that could influence mode choice include household structure, household size, household income, and the availability of a personal vehicle (Bhat, 1997; Cho, 2013; Li et al., 2015; and Ratledge, 2003). It seems that when traveling with a family, a private vehicle is preferred over public transportation (Javid et al., 2016). Ababio-Donkor et al. (2020) found that the likelihood of traveling by public transport reduces with an increasing level of household income. Similarly, car availability increases the utility for private motorized modes and the likelihood of traveling by car which is also confirmed by Kim et al. (2007).

2.3.1.3. Environmental characteristics

The most common environmental variables that are used in studies are urban intensity (Hu et al., 2018; Krygsman & Dijst, 2001; Limtanakool et al., 2006; Schlossberg, 2013; Tyrinopoulos & Antoniou, 2013), land use mix/ percentage of residential/ commercial/ industry (Hu et al., 2018; Kim et al., 2007; Koh & Wong, 2016; Limtanakool et al., 2006; Racca & Ratledge, 2004; Schlossberg, 2013; Senbil et al., 2009), presence of transport hub (G. C. De Jong, 2008; Kim et al., 2007; Koh & Wong, 2016), distance to the city center (Scheiner, 2010; Schlossberg, 2013; Senbil et al., 2009), parking availability destination (Ma, 2015; Racca & Ratledge, 2004; Schlossberg, 2013; Tyrinopoulos & Antoniou, 2013), road density (Hu et al., 2018; Schlossberg, 2013), number of traffic signals (Schlossberg (2013), intersection density (Senbil et al.,2009; Hu et al., 2018; Schlossberg, 2013), and number of bus services (Hu et al., 2018; Kim et al., 2007; Koh & Wong, 2016; Li et al., 2015; Schlossberg, 2013; Senbil et al., 2009). Schlossberg (2013) found that the environmental characteristics around the transit are important for mode choice. He found that street connectivity is positively related to the use of public transportation as a travel mode.

2.3.1.4. Trip characteristics

Trip characteristics such as travel time, travel distance, origin, destination, purpose, and time of department are most common within travel mode choice research (De Jong, 2008; Ma, 2015; Racca & Ratledge, 2004). The majority of studies show that total trip travel time (Almasri & Alraee, 2013; Racca and Ratledge, 2003), trip type/purpose (Almasri and Alraee, 2013; Cho, 2013; Limtanakool et al., 2006), and trip distance (Ashiabor et al., 2007; Cho, 2013; Nes, 2002) influence travelers' mode choice decisions. There is also some disagreement; a study by Kim et al. (2007) found that trip purpose is not related to mode choice. In addition, the variable trip costs (e.g. fuel) is also included in studies (De Jong, 2008; Racca & Ratledge, 2004; Tyrinopoulos & Antoniou, 2013) however, only De Jong (2008) mentioned that trip costs affect mode choice for short distance trips. Thereby, Chowdhury and Ceder, (2016) and Rietveld (2000a) found that travelers choose a private car over public transportation due to the travel time dependability, waiting time, poor connection, and low speed of entry modes of public transportation.

2.3.2. Multimodal mode choice

The personal, household, environmental, and trip characteristics mentioned above, influence also (public transportation based) multimodal mode choice. But in addition, according to Chowdhury and Ceder (2016), mode selection within multimodal trips is not determined by one or two variables. However, Nes (2002) found that among the trip-related variables, trip distance, type of destination, and trip purpose appear to have the greatest impact on multimodal trips. Based on these three variables, nearly 83 percent of multimodal trips can be correctly classified (Nes, 2002). Thereby, the first and last leg of a PTMT are influential hence they contribute to an increase in travel time and discomfort (Rietveld, 2000b). For example, Keijer & Rietveld (2007) found a link between the distance from the railway station and the mode choice behavior for the first and last legs of a PTMT. The results show that the distance between home and the railway station has a strong influence on mode choice. It has been observed that for short-distance trips, people prefer to ride their bikes or walk, whereas public transportation is mostly used to travel long distances between origins and destinations. Furthermore, when the distance from a railway station was compared to the use of the train as a mode of transport, it was discovered that people living within 500 m of a railway station use 20% more the train as a mode of transportation than people living within a range of 500-1000 meters (Keijer & Rietveld, 2007). Thereby a study was conducted in The Netherlands to investigate the importance of walking, bicycling, and other modes for the first and last leg of PTMTs (Rietveld, 2000b). Walking and bicycling are viable alternatives for origins and destinations closer to the railway station, according to the findings of this study. Bicycling is most commonly used for the first trip leg, whereas walking is the preferred mode for the last trip leg (Rietveld, 2000b).

2.4. Conclusion

The literature mentioned a lot of variables that could be related to (public transportation based) multimodal trips. Following from the literature study, a conceptual model is proposed. The variables that will be included are also presented in the conceptual model, illustrated in Figure 7.

The personal characteristics variables that could be related stated in the literature are gender, age, ethnicity, social participation, education, and driver's license. The variable income is only included as household characteristics since the data set of ODIN does not include the variable income per individual and cannot be traced afterward. Thereby, the variable motorcycle driver's license is included despite it is not mentioned within the reviewed studies. This could infer that it is not related or it is multicollinear to the variable driver's license. Nevertheless, the variable will be taken into account in this study to confirm or reject this assumption.

The household characteristics will contain the following variables found in the literature: size, structure, and income. The variable availability of a personal vehicle is subdivided into cars, motorcycles, scooters 45km/h or 25km/h, and electric bicycles. Because this research focuses on types of modes it seems logical to subdivide also the availability of the transportation mode and not only include the availability of a car within a household like other studies.

The environmental characteristics discussed in the literature are urban density, land use mix, presence of transport hub, distance to the city center, availability of parking spots, road density, number of traffic signals, intersection density, and the number of bus stops. Included in the conceptual model are urban density, number of traffic signs, number of crossings, number of parking spaces for cars and bicycles, number of bus stops, number of railway stations, number of tram/metro stops, and kilometers of pedestrian, bicycle, and motorist path. The only variable that is not included is land use mix because no data was found. In addition, Kim et al. (2007) included the variable crime rate since it applies to the case study of Kim et al. (2007) because the LRT stations are significantly associated with crime. For this research, it is assumed that the crime rate will not influence PTMTs within The Netherlands because The Netherlands has among others (Austria and Switzerland) of Europa the lowest fear of crime rate (Kujala et al., 2019).

The trip characteristics include motive, distance, and travel time. The literature mentions also the origin and destination however this data could not be used for the data set because the multimodal trips are not specified and due to reforming the data set this data is lost. However, the variable urban density is assumed to be sufficient since it concerns the urban density of the residential municipality and the trips are mostly conducted from or to home. Subsequently, the variable travel costs is not included as well since this is not included in the data set and cannot be traced afterward.

Within this research, two networks are created following this conceptual model whereby the dependent variable differs. The first network concerns the choice of trip type between a unimodal car trip or a PTMT. The second network focuses solely on the types of PTMTs. For the dependent variables, this study will consider walking as a transportation mode for the first and last trip leg of a PTMT which is in contrast to Nes (2002). This is because it is assumed that people living close to a bus stop or train station probably will choose to walk or cycle to the transit stop. For governments, this could be important to know, for creating new policies or for example redesigning the infrastructure to stimulate walking or cycling.

The specific types of PTMTs are generated from the data set's descriptive statistics. This is because it can differ per research what kind of types of public transportation based multimodal are conducted as mentioned in Section 2.2.1 Case studies.

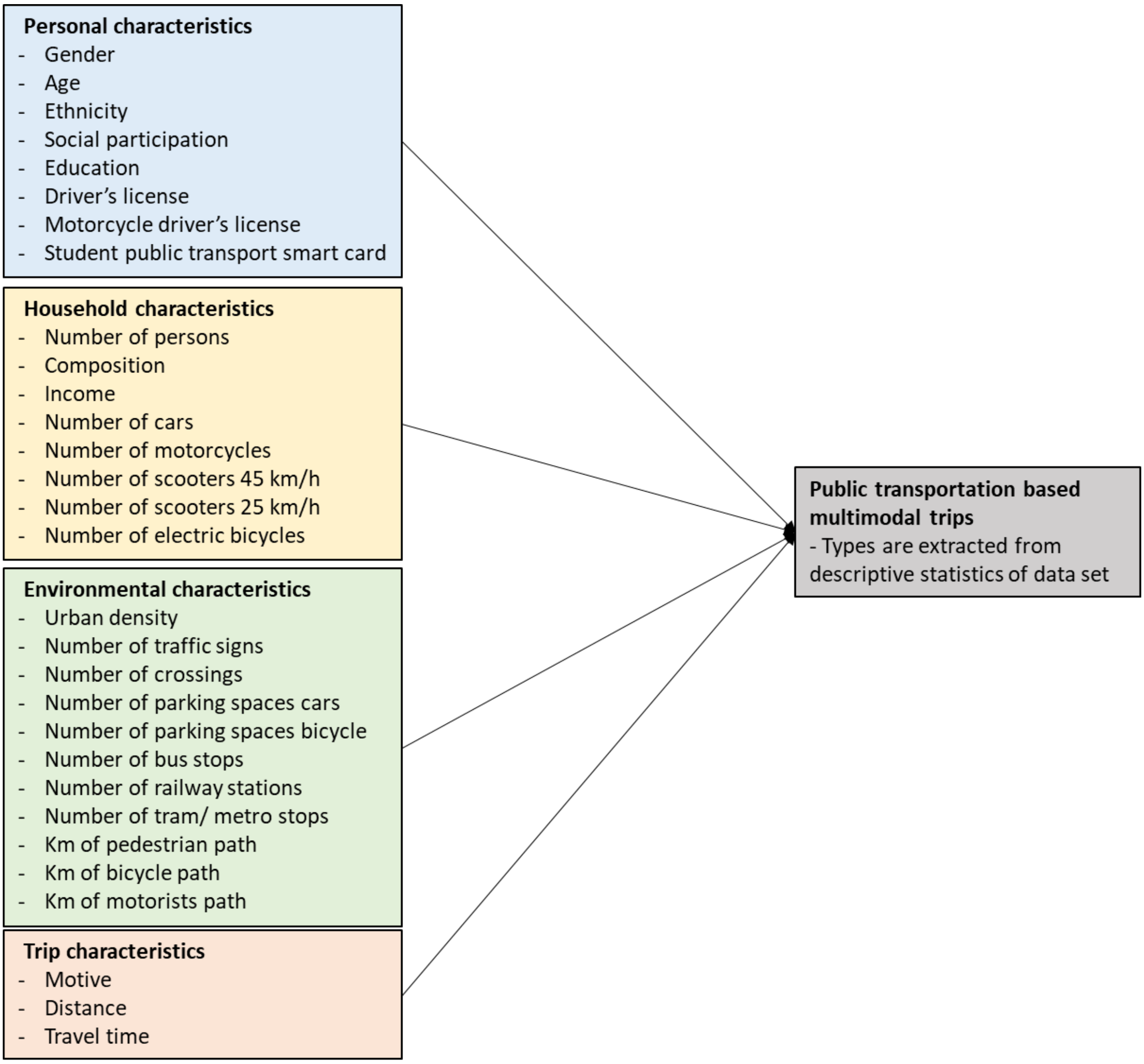


Figure 7: Conceptual model

3. Methodology

This chapter describes the methodology used in this study. The data collection, preparation, and analyzing technique will be addressed in detail.

3.1. Introduction

To answer the research question "Which variables are influencing types of public transportation based multimodal trips within The Netherlands?" various steps are taken. Based on the literature research, data is collected on potential variables that influence types of PTMTs. Subsequently, the data is prepared for statistical analysis and the analysis is performed.

The included variables are based on the literature research which can be found in Chapter 2 Literature research. The variables that will be included are categorized into personal, household, environmental, and trip characteristics. The personal characteristics are gender, age, ethnicity, social participation, education, driver's license, motorcycle driver's license, and student public transportation smartcard. The household characteristics are the number of persons, composition, income, number of cars, motorcycles, scooters 45 km/h, scooters 25 km/h, and electric bicycles within the household. The environmental characteristics are urban density, number of Traffic signs, crossings, parking spaces for cars, parking spaces for bicycles, bus stops, railway stations, tram/metro stops, and kilometers of pedestrian, bicycle, and motorist paths. Lastly, the trip characteristics are motive, distance, and travel time. These variables are potential influencing the dependent variable 'Public transportation based multimodal trips'.

The collected data of these variables is prepared in SPSS and the analysis, a Bayesian Belief Network (BBN), is estimated with the software GeNIe (version 2.1) (BayesFusion, 2022). A Bayesian Belief Network is created to explore the relationships between a set of variables that influence the choice of a unimodal car tip or a public transportation based multimodal trip, and to provide insight into the relationships between the variables influencing types of PTMTs.

3.2. Data collection

This research search for relationships between the above-mentioned variables and PTMTs. Therefore, the data used within this research is quantitative. The values of these variables are coming from a third party (ODiN and OSM) which is therefore secondary data. Secondary data is data that is already collected through a third party and is publicly accessible for other researchers to conduct their research. Moreover, publicly available data can be accessed and scrutinized by other researchers, enabling the replication of findings and boosting confidence in the outcomes (Kitchin, 2014).

The data set of the Dutch National Travel Survey, 'Onderweg in Nederland' (ODiN) is used for this research, as mentioned in 1.3.1. Data collection, ODiN concerns data on the daily mobility and travel behavior of inhabitants of The Netherlands. Annually, a representative sample of the Dutch population is formed through the participation of 30,000 to 50,000 respondents. In order to gather information, based on (stratified) random sampling participants are asked to document their activities on a specific day of the year, including details such as destination, purpose, mode of transportation, and travel duration (CBS, 2022b). Additionally, the survey seeks information on general personal and household characteristics, as well as possession of a driver's license and transportation. For the personal, household, and trip characteristics, the data set of ODiN is used. Thereby, one of the environmental variables 'urban density' is also included within ODiN and used for this research. The data set encompasses the year 2021. This is the latest available data set, providing the most recent representation of the current situation. To gain access to the data set it is needed to register via Dans Easy (Data Archiving and Networked Services), this is without any conditions or costs.

Thereby, data is retrieved from OpenStreetMap (OSM) (http://www.openstreetmap.org/). The OpenStreetMap project seeks to establish a comprehensive geographic database of the world at no cost. The ultimate objective is to document every geographic element on the planet, extending beyond the initial focus on street mapping to include details on footpaths, buildings, pipelines, waterways, beaches, trees, postboxes, and other features. Mappers, referred to as contributors within OpenStreetMap, create the database by gathering information through walking, cycling, or driving along streets and paths, and capturing their movements via GPS receivers. This data is used to establish a set of points and lines that can be used for navigation or map creation. The database employs a wiki-like system, allowing any mapper to add or modify any feature in any area, with a complete editing history maintained for each object. This ensures that any errors or deliberate tampering can be reversed, preserving data accuracy (Arsanjani et al., 2013). For this research, the environmental data (except the variable urban density) is obtained from OpenStreetMap (OSM). This concerns data on the number of traffic signs, crossings, bus stops, railway stops, and tram/ metro stops. Also, the possibility of car and bicycle parking and the number of kilometers sidewalk, bicycle path, and motorists path are included. The retrieved data of OSM is based on the infrastructure of 2018. For this research, it is assumed that the infrastructure did not significantly change between 2018 and 2021. Also, the retrieved data of OSM is obtained without any conditions or costs.

3.3. Data preparation

After the data collection, the data is cleaned and prepared for analysis. The two data sets (ODiN and OSM) are merged into one data set with all relevant variables. This is done with Merge Files within SPSS. The data merging is based on the key variable 'Zip code'. Because of this, it is clear which environmental variables' data apply per respondent.

The data set of ODiN 2021 has 196,768 cases. These cases represent trip legs, so for unimodal trips, one case is defined and for multimodal trips, multiple cases are defined. This depends on the number of legs used for the trip. For example, if a respondent uses three different transportation modes, so three legs in one trip, three cases represent the complete trip. These multimodal trips are converted to one row (case) so it has the same format as the unimodal trips.

For the analysis, two data sets are prepared. The first data set (Data set 1 – Trip type) only includes cases of unimodal car trips and PTMTs. Unimodal trips such as cycling and other active travel modes should not be replaced by PTMTs since they do not result in various environmental, economic, and social problems (Steg, 2003). Also, all cases that do not have public transport as the main mode of the multimodal trip are excluded since the aim is to shift unimodal car trips to PTMTs so the environmental, economic, and social problems decrease (as mentioned in Chapter 1. Introduction). Because of the comparison between unimodal car trips and PTMTs, trips shorter than 30 kilometers are excluded. The literature stated that a multimodal trip is an intriguing choice if the travel distance is longer than 30 kilometers (Nes, 2002). Due to this, the focus is only on reducing unimodal car trips and increasing PTMTs. By converting the multimodal trips and excluding the above-mentioned cases, 56,312 cases of unimodal car trips and 3,895 cases of PTMTs are retained (Table 2). Note, the cases presenting thus one complete trip since the PTMT legs are converted to one trip. This distribution shows already a growth of 2% in PTMTs compared to the data from 2018 of multimodal trips (CBS & OvIN, 2020).

Table 2: Data set 1: Distribution of the variable Trip type

Trip type	Cases	Percent
Unimodal car trip	56,312	93.5
Public transportation based multimodal trip	3,895	6.5
Total	60,207	100

The second data set (Data set 2 – Types of PTMTs) focuses only on specific types of public transportation based multimodal trips. The most common types are defined and based on descriptive statistics of the PTMTs (Table 3). The type of PTMT 'other' does include legs conducted by the following transportation modes: speed pedelec, touring car, truck, camper, taxi, agricultural vehicle, Motorcycle, scooter 45/25 km/h, skates, and boat. Also, the trip type 'car + public transport + car' is included in others because it includes very few cases. This is in contrast with the study of Basheer et al. (2019). Thereby, based on Krygsman & Dijst (2001) it is assumed that a PTMT first leg could not start with a public transportation mode, since travelers should always walk or cycle to a transit stop (e.g. train, bus, or metro). Within this study, these specific cases are recoded with walking as the first leg mode. This differs from Basheer et al. (2019), since they did not recode these trips but included them as a trip type. Subsequently, in comparison with the descriptive statistics of the dependent variable of Basheer et al. (2019), the types of PTMTs with cycling as the first or last leg mode are more conducted within The Netherlands. Therefore, these trips are included as a specific type in this study instead of included in the type 'Other'.

Table 3: Data set 2: Distribution of the variable Types of PTMTss

Types of PTMTs	Cases	Percent
Walking + Public transport + Walking	2,046	52.5
Cycling + Public transport + Cycling	204	5.2
Cycling + Public transport + Walking	500	12.8
Walking + Public transport + Cycling	556	14.3
Car + Public transport + Walking	139	3.6
Car + Public transport + Cycling	39	1.0
Other	411	10.6
Total	3,895	100

Data sets 1 and 2 contain the same independent variables, except the dependent variable per data set (1 – Trip type and 2 – Types of PTMTs) as discussed above. Within the ODiN 2021 data set, multiple variables are excluded. The included variables are selected based on the literature research. Thereafter, in SPSS the remaining variables which will be included in the analysis are recoded into categories. The number of categories depends on the variable and its distribution. Table 4 shows the included variables with the description, code, descriptive, and frequencies of data set 1 and 2. The frequencies of data set 1 and 2 will not be reflected on differences.

Table 4: Description of included variables

Variable	Description	Code	Descriptive	Frequencies (%) Data set 1	Frequencies (%) Data set 2
gender	Gender of the respondent	0	Men	52.6	43.7
		1	Women	47.4	56.3
Age	Age of the respondent	0	< 21	8.4	16.9
		1	21-35	17.3	46.9
		2	36-50	24.2	17.5

		3	51-65	26.4	12.0
		4	> 65	23.7	6.6
Ethnicity	Ethnicity of the respondent	0	Dutch background	79.7	62.7
		1	Western migration background	9.3	12.1
		2	Non-western migration background	11.0	25.2
Social participation	Social participation of the respondent	0	Employed	62.3	48.9
		1	Unemployed	2.9	2.6
		2	Student	12.6	40.6
		3	Other	22.2	7.9
Education	Highest education completed by the respondent	0	No education	0.6	1.1
		1	Pre-vocational secondary education	15.5	12.7
		2	secondary vocational education	33.1	33.6
		3	higher professional education/ university education	43.0	46.2
		4	Other	7.7	6.4
Driving license	The respondent's Possession of a driver's license	0	No	13.6	43.7
		1	Yes	86.4	56.3
Motorcycle driver's license	The respondent's possession of a motorcycle driver's license	0	No	87.5	96.8
		1	Yes	12.5	3.2
Student public transportation smartcard	The respondent possesses a student public transportation smartcard	0	No	94.2	66.3
		1	Yes	5.8	33.7
# of Persons	Number of persons in the household of the respondent	0	1 person	14.9	25.6
		1	2 persons	36.3	27.6
		2	3 persons	16.0	16.9
		3	4 persons	22.8	18.6
		4	5 or > persons	10.0	11.3
Composition	Composition of the household of the respondent	0	Single household	14.9	25.6
		1	Couple	32.8	21.7
		2	Couple/ single + child(ren)	51.9	51.0
		3	Other	0.4	1.6
Income	Disposable income of the household of the respondent (ODiN divided it into 10% groups)	0	Below average income 0-40 %	16.9	29.1
		1	Average income 50-70%	44.5	35.6
		2	Above average income 80-100%	38.7	35.3
# of Cars	Number of cars in the household of the respondent	0	0	7.1	41.2
		1	1	45.6	35.1
		2	2 or more	47.3	23.7
# of Motorcycles	Number of motorcycles in the household of the respondent	0	0	91.4	95.3
		1	1 or more	8.6	4.7
		0	0	95.5	96.8

# of Scooters 45 km/h	Number of scooters 45km/h in the household of the respondent	1	1 or more	4.5	3.2
# of Electric bicycles	Number of electric bicycles in the household of the respondent	0	0	65.3	81.5
		1	1 or more	34.7	18.5
# of Scooters 25 km/h	Number of scooters 25km/h in the household of the respondent	0	0	92.3	94.9
		1	1 or more	7.7	5.1
Motive	The motive for traveling of the respondent	0	Work	27.2	30.5
		1	Education	3.5	23.6
		2	Shopping/ grocery	17.7	7.8
		3	Leisure	17.2	15.1
		4	Other	34.3	23.1
Travel distance	Travel distance of the trip conducted by the respondent	0	30-50 km	23.6	6.0
		1	51-100 km	23.9	13.0
		2	101-250 km	27.0	31.1
		3	> 250 km	25.5	49.9
Travel time	Travel time of the trip conducted by the respondent	0	> 30 min	9.0	0.4
		1	31-60 min	21.0	6.9
		2	61-90 min	20.9	14.4
		3	91- 120 min	17.0	19.6
		4	> 120 min	32.0	58.7
Urban density	The urban density of the residential municipality of the respondent based on address density	0	Very highly urban (2500 or more)	24.9	52.5
		1	Highly urban (1500 – 2500)	29.7	25.6
		2	Moderately urban (1000 – 1500)	16.0	8.2
		3	Little urban (500 – 1000)	22.2	10.2
		4	Not urban (less than 500)	7.2	3.4
# of Traffic signs	Number of traffic signs within the residential municipality of the respondent	0	0	32.3	17.4
		1	1-5	21.8	20.0
		2	6-15	23.3	25.5
		3	16-30	13.8	21.2
		4	> 30	8.7	15.9
# of Crossings	Number of crossings within the residential municipality of the respondent	0	0-2	40.2	29.5
		1	3-10	32.3	32.3
		2	> 10	27.5	38.2
# of car Parking	Number of car parking spaces within the residential municipality of the respondent	0	0	33.3	32.2
		1	1-3	41.3	41.7
		2	> 3	25.3	26.1
# of bicycle Parking	Number of bicycle parking spaces within the residential municipality of the respondent	0	0	57.2	50.0
		1	1-2	26.1	26.3
		2	> 2	16.7	23.6
# of Bus stops		0	0	34.3	37.6
		1	1-2	50.4	48.4

	Number of bus stops within the residential municipality of the respondent	2	> 2	15.4	14.0
# of Railway stations	Number of railway stations within the residential municipality of the respondent	0	0	77.5	71.9
		1	1 or more	22.5	28.1
# of Tram Metro stops	Number of trams/metro stops within the residential municipality of the respondent	0	0	87.7	70.8
		1	1 or more	12.3	29.2
Km of Pedestrian	Kilometers of a pedestrian path within the residential municipality of the respondent	0	< 10 km	10.5	11.1
		1	11 – 25 km	19.5	21.9
		2	26 – 80 km	35.8	35.7
		3	> 80 km	34.1	31.4
Km of Bicyclists	Kilometers of bicycle path within the residential municipality of the respondent	0	< 10 km	9.3	9.8
		1	11 – 25 km	43.6	39.4
		2	26 – 80 km	36.8	39.8
		3	> 80 km	10.4	11.0
Km of Motorists	Kilometers of motorist's path within the residential municipality of the respondent	0	< 10 km	9.6	8.9
		1	11 – 25 km	41.4	48.1
		2	26 – 80 km	42.0	35.2
		3	> 80 km	7.0	7.8
Trip type (dependent variable data set 1)	Type of trip conducted by respondent	0	Unimodal car trip	93.5	-
		1	Public transportation based multimodal trip	6.5	-
Multimodal trips (dependent variable data set 2)	Specific combined public transportation multimodal trips	0	Walking + Public transport + Walking	-	52.5
		1	Cycling + Public transport + Cycling	-	5.2
		2	Cycling + Public transport + Walking	-	12.8
		3	Walking + Public transport + Cycling	-	14.3
		4	Car + Public transport + Walking	-	3.6
		5	Car + Public transport + Cycling	-	1.0
		6	Other	-	10.6

3.4. Analysis

To explore and estimate the direct and indirect variables influencing types of PTMTs, a Bayesian Belief Network (BBN) is estimated. A BBN is a collection of variables that are linked together to show their interdependencies and provide information about their relationships (Arentze & Timmermans, 2008).

A BBN is a Directed Acyclic Graph (DAG) and is formulated by Heckerman, Mandani, & Wellman (1995) and Pearl (1988) as follows:

$$BBN = (V, E)$$

where *V* is a set of nodes, and *E* is a set of directed arcs. Nodes are variables (*x,y,z,..*) and the directed arcs present a link between for example *X*→*Y*, where variable *X* is called the parent of variable *Y* and variable *Y*, is called the child of variable *X*. Each variable in the network has a conditional probability table (CPT), which specifies the likelihood of the variable for each state configuration of its parent nodes (if any). These CPTs collectively constitute the network's parameters (Arentze & Timmermans, 2008; Kemperman & Timmermans, 2014a; Ma, 2015). A BBN can deal with discrete variables such as gender and household structure, in contrast to for example structural equation modeling, (Kemperman et al., 2019).

The process of learning a network involves two primary tasks (BayesFusion, 2022):

1) Learning the network's structure
2) Estimating the parameters, which are the conditional probability tables (CPTs).

3.4.1. Learning the network's structure

Structure learning determines conditional dependency/independency between the variables. For structure learning, a Greedy Thick Thinning (GTT) algorithm is used. This algorithm is used within the software GeNIe since compared to the other algorithms, the GTT can achieve statistical significance and large gains for large data sets (BayesFusion, 2022). The Greedy Thick Thinning (GTT) algorithm begins with computing the mutual information (*I*) between each pair of variables (*X, Y*) and using it as a measure of closeness. The definition of mutual information between two variables, *X* and *Y*, is as follows (Pearl, 1988):

$$I\ (X, Y) = \sum_{X,Y} P\ (x, y) \log \frac{P(x, y)}{P(x)\ P(y)}$$

The mutual information between variables *X* and *Y* is calculated using the joint probability *P(x,y)* and the unconditional probabilities *P(x)* and *P(y)*, where *P(x)* represents the probability of *X* taking the value *x*, and *P(y)* represents the probability of *Y* taking the value *y*. This metric is used to quantify the amount of information gained about *Y* when the value of *X* is observed. In BBN mutual information is useful for determining the dependence between two variables to define the network structure. If two variables are dependent and the value of one variable is known, it can provide information about the value of the other variable (Theo Arentze & Timmermans, 2008). Subsequently, the algorithm iteratively adds arcs (without creating a cycle) that maximally increase the marginal likelihood *P(D|S)* until further arc addition does not yield a positive increase (this is the thickening phase). The marginal likelihood *P(D|S)* represents the probability of the data given the network, integrated over all possible values of the network parameters (BayesFusion, 2022). Hereafter, it removes arcs repeatedly until no further arc deletion will result in a positive increase in *P(D|S)* (this is the thinning phase) (Cheng et al., 1997). The result of learning the network's structure is a network consisting of nodes and arcs. The learned network of Ma (2015) is shown as an example of a learned network in Figure 8.

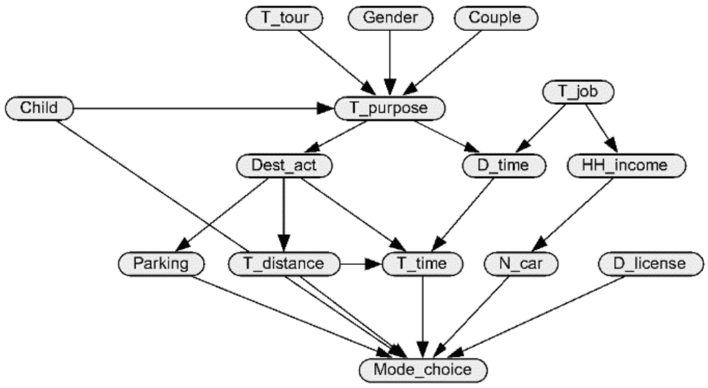


Figure 8: Example of a learned network's structure (Ma, 2015)

3.4.2. Estimating the network's parameters

For parameter learning, meaning the estimation of the conditional probability tables (CPTs), the Expectation–Maximization (EM) algorithm is used. This algorithm uses an iterative approach to search for the conditional probability distributions. It begins with a candidate BBN, performs an expectation (E) step on it to discover a better one, and then performs a maximizing step (M). The process is repeated until the log-likelihood values stop rising (according to a tolerance that is specified) (Lauritzen, 1995). With the estimated CPTs it is possible to make predictions about the probabilistic changes in other variables, and it is also possible to simulate changes under specific conditions. Whenever new findings, also called evidence, are added to the network, the CPTs of all related variables will be updated. The result of estimating the network's parameters (CPTs) is the learned network with the estimated CPTs. The network of Ma (2015) is shown as an example of a learned network with the estimated parameters in Figure 9.

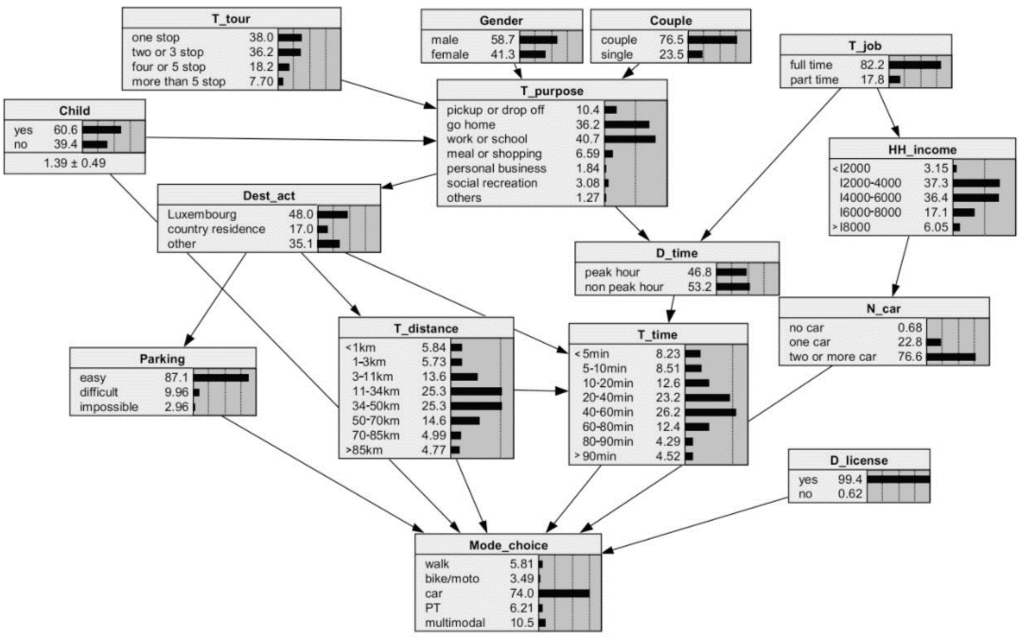


Figure 9: Learned network's structure + parameters (CPTs) (Ma, 2015)

The network of this research is created with the software GeNIe, which can learn, display, and apply the BBN. GeNIe generates a network overview of the variables and illustrates the impact of categories within variables on other variables by mutual information. The BBN shows which variables have relationships and how strong these relationships are. When some variable categories in the network are set to 100%, meaning evidence is added to the network, changes in the probabilities of the categories of other related variables can be noted. This makes it clear how one variable influences the probabilities of the categories of other variables. These influences can be direct (between two variables) or indirect (one variable influences another, causing a change in a third or fourth variable).

3.4.3. Constraints

In advance of learning the structure of the BBN, constraints can be defined on variables and certain relationships between variables. Within GeNIe constraints can be set due to the max parent count and the background knowledge.

Within this research, the max parent count is set to 3, meaning a child node can only be influenced by 3 parent nodes. Because, when there are too many parent nodes influencing a child node, their effect is averaged out, creating an equal distribution in the child variable. With the max parent count set to 3, the most important relationships are shown. The max parent count was also set at a higher number (5). However, this did not influence the network's structure, since none of the variables had more than 3 parent nodes. From this, it is assumed that the network is stable.

The background knowledge within GeNIe can force and forbid relationships and also assign variables to temporal tiers. For this research, it is chosen to limit the constraints to look for the best-fitting network based on the provided data without any prior influence of the researcher. However, the variables that cannot influence each other by nature (e.g. age and gender) are forbidden. The forbidden arcs are only applied for the personal variables to limit the constraints. Thereby, the variables are also set to different temporal tiers. These tiers are establishing the chronological sequence of the variables: there will be no arcs originating from variables set at higher tiers to nodes that are set in lower tiers. The variables are logically arranged into temporal tiers; for instance, gender and age are assigned to tier 1 because they are naturally unaffected by other variables. Figure 10 shows the background knowledge used for data set 1 – Trip type. The background knowledge is also used for data set 2 – Types of PTMTs however, the dependent variable (which can be found in temporal tier 4) differs.

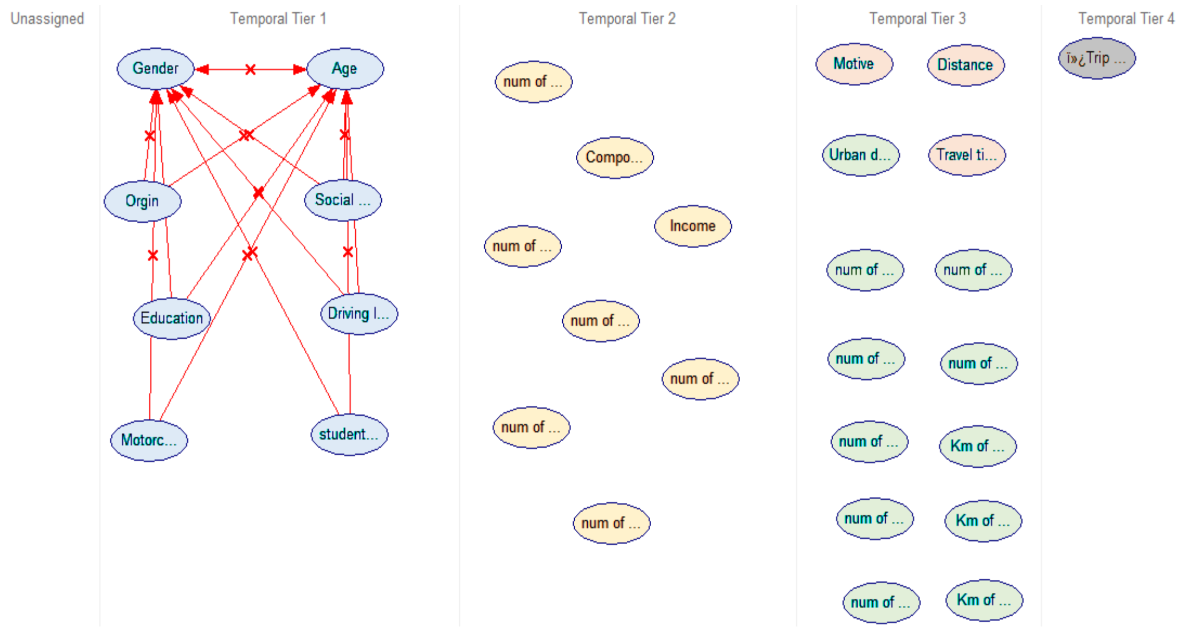


Figure 10: Applied background knowledge: Temporal tiers and forbid arcs

To summarize the process of creating a BBN, the steps are presented in Table 5. Thereby, the steps that are taken within this research are described, including the paragraph where the explanation and process of the taken steps are elaborated.

Table 5: Summary of steps of creating a BBN

Steps of creating a BBN	Steps taken regarding this research	Paragraph reference
1. Select the variables of interest from the database	Variables of the literature research (ODiN & OSM)	2.4. Conclusion 3.2. Data collection
2. Solve missing values, if any	Data preparation within SPSS	3.3. Data preparation
3. Where needed, re-classify variables and recode cases. Discretize continuous variables Merge categories of a variable if there are too many	Data preparation within SPSS	3.3. Data preparation
4. Specify constraints on the network (if there are any)	Background knowledge → Forbid arcs and Tiers	3.4. Analysis
5. Specify the threshold value for identifying links	Max parent count: 3	3.4. Analysis
6. Run a structural-learning algorithm	Greedy Thick Tinning algorithm	3.4. Analysis
7. Manually orient links where needed	Not needed since the network is assumed stable	3.4. Analysis
8. Run a parameter-learning algorithm	Expectation–Maximization algorithm	3.4. Analysis

4. Results

This part will discuss the two created BBNs, the strength of influences, and findings within the networks. Also, the conceptual model shown in Section 2.4. Conclusion is reviewed based on the results.

4.1. BBN 1 – Trip type

The first network learned is the network to answer the following sub-question: *Which variables directly and indirectly influence the choice of trip type between a unimodal car trip and a public transportation based multimodal trip?*

4.1.1. The learned network 1 – Trip type

Within GeNIe the GTT algorithm is used to learn the BBNs structure. The parameter for this algorithm 'max parent count' is set to 3. In this setting, the most important relationships are shown since changing the setting to a higher number did not influence the variable trip type significantly. From this, it is assumed that the network is stable. Thereby, the background knowledge is set as follows: the personal characteristics are set to tier 1, the household characteristics are set to tier 2, the trip and environmental characteristics are set to tier 3, and the dependent variable trip type is set to tier 4. These tiers indicate that there will not be a relationship between variables that are stated in higher tiers and variables that are stated in lower tiers. Thereby, only forbid arcs are set to some personal characteristics since they cannot influence each other by nature. The forbid arcs are set as follows: none of the personal characteristics can influence the variable gender, which also applies to the variable age. To look for the best-fitting network, no other constraints are set.

From this, the network's structure is learned of network 1 – trip type (Figure 11 and Appendix A). The variables are visible as nodes and the relationships between the variables are presented as arcs. It can be seen that the variable trip type has a direct relationship with the variables student public transportation smartcard, motive, and number of cars within a household. Though, this network is not very clear due to the many variables and relationships between the variables. When looking at the network, there are a few remarkable points. First, the environmental variables are mainly influencing each other however, a couple of environmental variables are influenced by the variables number of electric bicycles within the household, ethnicity, and number of cars. Thereby, the variable urban density directly (and indirectly) influence all environmental variables. This implies that the variable urban density includes and represents all environmental variables. Therefore, it is chosen solely to include urban density within the network. The second remark is that some variables are presenting similar information. Looking at the personal variables this concerns the variables of driving license and motorcycle driver's license. The variable driving license does indirectly influence the dependent variable trip type which does not apply to the variable motorcycle driver's license. Hence, the variable motorcycle driver's license is excluded from the network. For the household variables, this applies to household composition and the number of persons within the household. It is chosen to include only household composition since it comprises the number of persons within a household sufficient for this research. The last remark concern variables that do not significantly influence (direct and indirect) the variable trip type. This regards the following variables: gender, ethnicity, education, number of motorcycles, number of scooters 45, number of scooters 25, and number of electric bicycles. Therefore, these variables are excluded from the network since they are not of interest to this research. By clearing the network of these variables the network is more explicit and only important variables are taken into account.

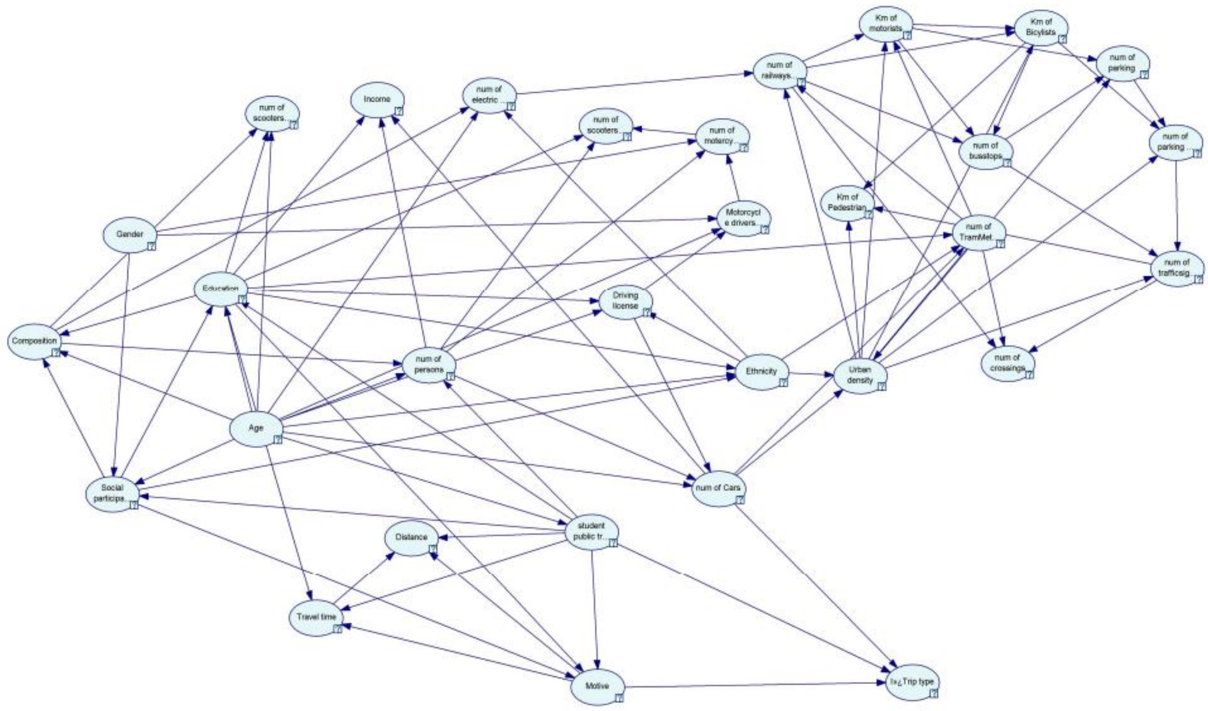


Figure 11: Learned network structure 1 – trip type

4.1.2. Final learned network 1 – Trip type

After excluding the variables as mentioned above in 4.1.1. The learned network, the following network is learned with the network's parameters (CPTs) (Figure 12). The network is also visible in Appendix B. For this network, the same settings are used as described above regarding the max parent count parameter and the background knowledge. It can be seen that the variables directly influencing the dependent variable trip type are still the variables student public transportation smartcard, motive, and the number of cars within a household. Within this network, the relationships (arcs) are bolder based on the strength of influence between the variables through the function 'Strength of influence'. Also, all influence strengths between the variables within the network are extracted. Table 6 shows the extracted influence strengths between the variables. The measuring scale is from 0 to 1, whereby 1 is the highest possible value and thus a strong relationship. The strongest relationships (above 0.300) are marked in bold within the Table like the arcs in the learned network.

Table 6: Strength of influence network 1 – Trip type

	Age	Social particpation	Driving license	Student PT smartcard	Income	Num of cars	Composition	Urban density	Motive	Distance	Travel time	Trip type
Age	x	**0.435**	0.225	0.135	-	0.191	0.263	0.095	0.108	-	-	-
Social particpation	-	x	0.120	-	0.190	-	0.160	-	0.135	-	0.144	-
Driving license		-	x	-	-	**0.371**	-	-	-	-	-	-
Student PT smartcard	-	**0.475**	0.273	x	-	-	0.259	-	0.210	0.127	0.158	0.234
Income	-	-	-	-	x	-	-	0.086	-	-	-	-
Num of cars	-	-	-	-	0.220	x	-	0.182	-	-	-	0.212
Composition	-	-	-	-	0.237	**0.325**	x	-	-	-	-	-
Urban density	-	-	-	-	-	-	-	x	-	-	-	-
Motive	-	-	-	-	-	-	-	-	x	0.153	0.122	0.173
Distance	-	-	-	-	-	-	-	-	-	x	-	-
Travel time	-	-	-	-	-	-	-	-	-	0.252	x	-
Trip type	-	-	-	-	-	-	-	-	-	-	-	x

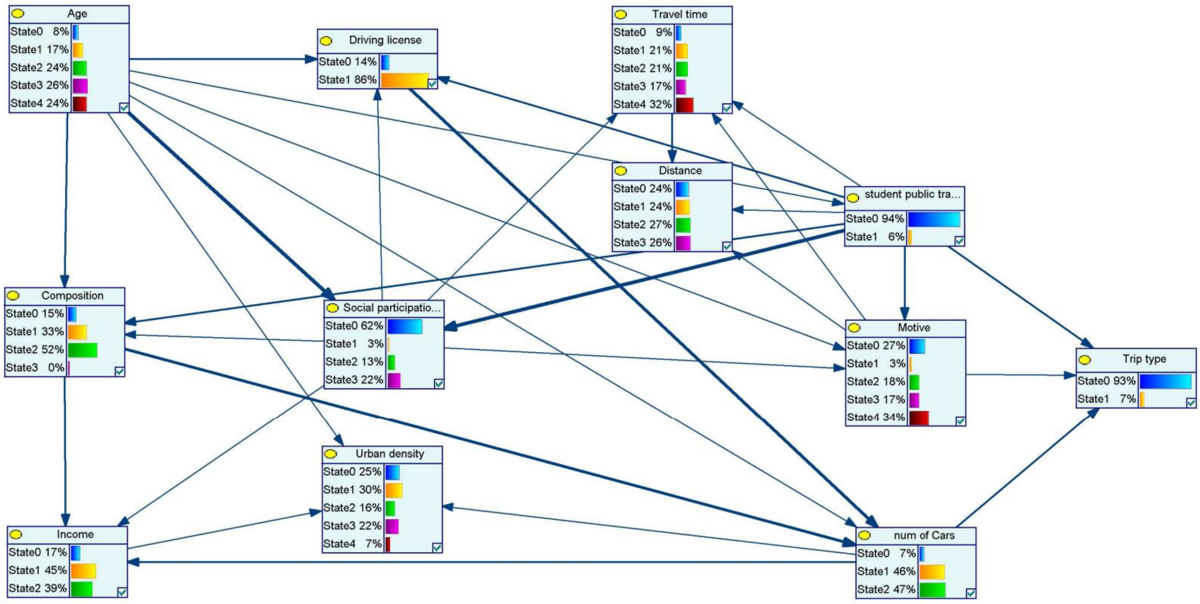


Figure 12: Final learned network 1 with CPTs – Trip type

4.1.3. Findings final network 1 – Trip type

Per characteristic, the relationships will be discussed between a certain variable and the dependent variable trip type.

First, the personal characteristics. It seems that only the variable student public transportation smartcard has a direct influence on the dependent variable trip type. With the evidence of having a smartcard shows that 64% will conduct a unimodal car trip and 36% a public transportation based multimodal trip. This shows a significant positive influence since without set evidence the probabilities are stated as 93% for a unimodal car trip and 7% for a public transportation based multimodal trip. Table 7 shows the probability distribution with and without set evidence within the variable public transportation smartcard.

Table 7: Updated probabilities hard evidence student public transportation smartcard

Trip type	No evidence	Evidence No smartcard	Evidence Yes smartcard
Unimodal car	0.935	0.953	0.642
PTMT	0.065	0.047	0.358

Thereby, the variables age, social participation, and driving license influence the variable trip type indirectly. It seems that 15% of the people aged between 21-35 will conduct a multimodal trip (Table 8). This is for the other age groups significantly lower. Looking at the variable social participation, it appears that students significantly conduct more multimodal trips than employed and unemployed people. Between employed and unemployed people there is no difference compared to the choice of trip type (Table 9). For the variable driving license it applies that if a person does not have a driving license, the probability of conducting a unimodal car trip is 87% and a PTMT 13% (Table 10). If a person has a driving license the probability of conducting a PTMT reduces by 8%.

Table 8: Updated probabilities hard evidence age

Trip type	No evidence	Evidence Age < 21	Evidence Age 21-35	Evidence Age 36-50	Evidence Age 51-65	Evidence Age > 65
Unimodal car	0.935	0.886	0.849	0.953	0.963	0.965
PTMT	0.065	0.114	0.151	0.047	0.037	0.035

Table 9: Updated probabilities hard evidence social participation

Trip type	No evidence	Evidence Employed	Evidence Unemployed	Evidence Student	Evidence Other
Unimodal car	0.935	0.947	0.949	0.820	0.935
PTMT	0.065	0.053	0.051	0.180	0.065

Table 10: Updated probabilities hard evidence driver's license

Trip type	No evidence	Evidence No driver's license	Evidence Yes driver's license
Unimodal car	0.935	0.824	0.932
PTMT	0.065	0.176	0.068

Second, the household characteristics. These include income, composition, and number of cars. The only direct influence on the dependent variable trip type is the variable number of cars within a household. It seems that with no cars in a household, the probability of conducting a multimodal trip is 36%, which is significantly lower if there are one or more cars within a household. These probability distributions are presented in Table 11.

Table 11: Updated probabilities hard evidence # of cars within a household

Trip type	No evidence	Evidence No cars in Household	Evidence 1 car in Household	Evidence 2 or > cars in Household
Unimodal car	0.935	0.639	0.948	0.966
PTMT	0.065	0.361	0.052	0.034

Looking at the variables that only have an indirect influence on the dependent variable trip type the following results are shown in the network. Households with below-average income tend to make more multimodal trips (11%) than households with average and above-average income (both 6%). Also, a household that consists of one person tends to conduct more multimodal trips in comparison with a household consisting of a couple with or without children. See Table 12 for the relationship between income and trip type, and see Table 13 for the relationship between composition and trip type.

Table 12: Updated probabilities hard evidence household income

Trip type	No evidence	Evidence Below average income	Evidence Average income	Evidence Above average income
Unimodal car	0.935	0.893	0.943	0.944
PTMT	0.065	0.107	0.057	0.056

Table 13: Updated probabilities hard evidence household composition

Trip type	No evidence	Evidence Single household	Evidence Couple	Evidence Couple/ single + child(ren)	Evidence Other
Unimodal car	0.935	0.883	0.957	0.937	0.837
PTMT	0.065	0.117	0.043	0.063	0.163

The environmental characteristics only consist of the variable urban density as mentioned in Section 4.1.1. The learned network 1 – Trip type. Urban density is not directly related to the variable trip type. However, the probabilities distribution of the variable trip type does slightly change when setting hard evidence. It seems that PTMTs are conducted more within a very highly urban environment (10%). For highly urban environments this is only 6% and for the lower urban densities, it is 5% as shown in Table 14.

Table 14: Updated probabilities hard evidence urban density

Trip type	No evidence	Evidence Very highly urban	Evidence Highly urban	Evidence Moderately urban	Evidence Little urban	Evidence Not urban
Unimodal car	0.935	0.903	0.940	0.949	0.949	0.949
PTMT	0.065	0.097	0.060	0.051	0.051	0.051

The last characteristics are the trip characteristics (motive, distance, and travel time). Only the variable motive is directly influencing the choice of trip type (Table 15). It seems that most PTMTs (43%) are used to travel for education. For work, this is 8%, for leisure 6%, and only 3% conduct a PTMT for shopping/grocery.

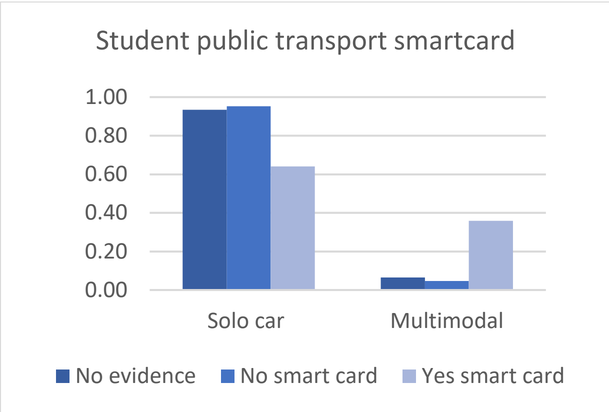
Table 15: Updated probabilities hard evidence motive

Trip type	No evidence	Evidence Work	Evidence Education	Evidence Shopping/ grocery	Evidence Leisure	Evidence Other
Unimodal car	0.935	0.924	0.571	0.969	0.943	0.959
PTMT	0.065	0.076	0.429	0.031	0.057	0.041

The variable travel time and distance are not directly related to the variable trip type and do not significantly change the probability distribution of the variable trip type. The probability distributions can be seen in Appendix C.

4.1.4. Conclusion of network 1 – Trip type

The learned network is slightly corresponding with the conceptual model as stated in Section 2.4. Conclusion. Within the included characteristics (person, household, environment, and trip) at least one variable significantly influences the dependent variable trip type. Regarding personal characteristics, the only direct influence is the variable student public transportation smartcard. People that have a smartcard are 29% more likely to conduct a PTMT compared to people who do not have a smartcard (Figure 13). Looking at the household characteristics, it seems that the variable number of cars within a household influences the variable trip type the most. If there is no car available within a household 36% use a PTMT to get to their destination. When there are two or more cars available this percentage decreases to 3% (Figure 13). The solely included environmental characteristic, urban density, is not directly related but influences the trip type indirectly. Within a very high urban area, 10% conduct a PTMT, this is for a highly urban area 6 %, and for lower urban areas this is only 5% (Figure 15). The variable motive seems to be the most important trip characteristic. It appears that most of the PTMTs are conducted for education (43%). For shopping/grocery, it seems that 97% is conducted with a unimodal car trip (Figure 16). So, to conclude there are no characteristics that do not influence the variable trip type. The strongest relationship (0.234) with the dependent variable trip type is with the variable student public transportation smartcard. Thereby, the variables motive (0.173) and number of cars (0.212) are also significantly related to the choice of conducting a unimodal car trip or a PTMT.



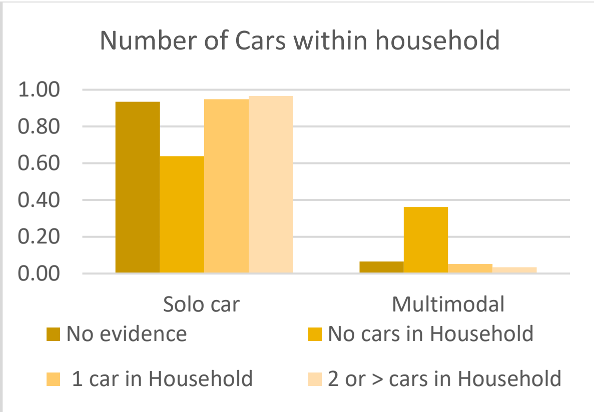
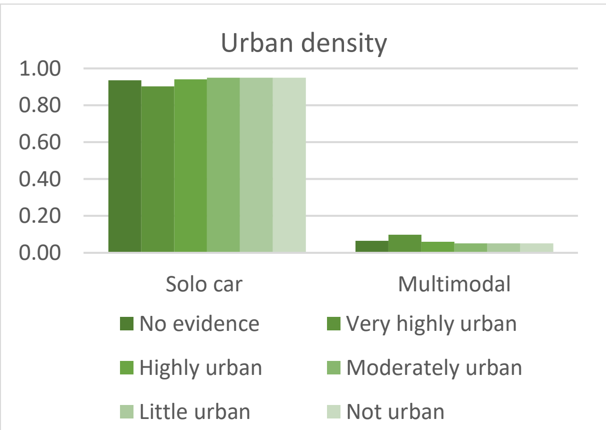


Figure 13: Update probabilities hard evidence student public transportation smartcard

Figure 14: Update probabilities hard evidence number of cars within household



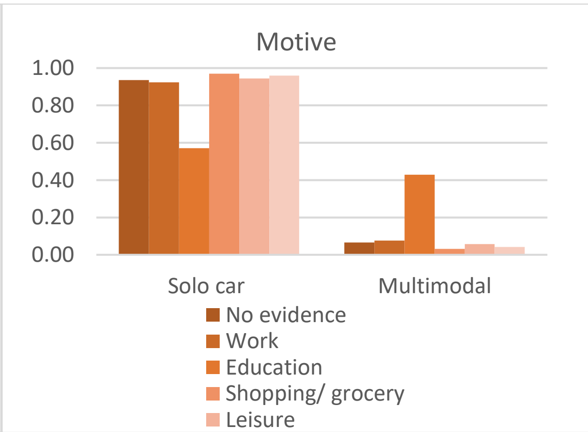


Figure 15: Update probabilities hard evidence urban density

Figure 16: Update probabilities hard evidence motive

4.2. BBN 2 – Types of PTMTs

The second network learned is the network to answer the following sub-question: *Which variables directly and indirectly influence types of public transportation based multimodal trips?*

4.2.1. The learned network – Types of PTMTs

Also for this network, the GTT algorithm is used to learn the BBNs structure. The settings are the same as in network 1. The parameter 'max parent count' is set to 3. In this setting, the most important relationships are shown since changing the setting to a higher number did not influence the variable trip type significantly. From this, it is assumed that the network is stable. Thereby, the background knowledge is set as follows: the personal characteristics are set to tier 1, the household characteristics are set to tier 2, the trip and environmental characteristics are set to tier 3, and the dependent variable types of PTMTs is set to tier 4. These tiers indicate that there will not be a relationship between variables that are stated in higher tiers and variables that are stated in lower tiers. Thereby, only forbid arcs are set to some personal characteristics since they cannot influence each other by nature. The forbid arcs are set as follows: none of the personal characteristics can influence the variable gender, which also applies to the variable age. To look for the best-fitting network, no other constraints are set.

From this, the network's structure is learned of network 2 – Types of PTMTs (Figure 17 and Appendix D). The variables are visible as nodes and the relationships between the variables are presented as arcs. It can be seen that the dependent variable types of PTMTs solely has a direct relationship with distance. Also, this network is not clear due to the many variables and relationships between the variables. When looking at the network, there are a few remarkable points. First, the environmental variables are mainly influencing each other however, a couple of environmental variables are influenced by the variables number of electric bicycles within the household, ethnicity, and number of cars. The environmental variables do not influence other variables. Although the variable number of tram/metro stops indirectly influences the dependent variable types of PTMT it is chosen solely to include urban density within the network. Urban density directly (and indirectly) influence all environmental variables. This implies that the variable urban density includes and represents all environmental variables. The second remark is that some variables are presenting similar information. Looking at the personal variables this concerns the variable driving license and motorcycle driver's license. The variable driving license does indirectly influence the variable types of PTMTs which does not apply to the variable motorcycle driver's license. Hence, the variable motorcycle driver's license is excluded from the network. For the household variables, this applies to household composition and the number of persons within the household. It is chosen to include only household composition since it comprises the number of persons within a household sufficient for this research. The last remarks concern variables that do not significantly influence (direct and indirect) the variable types of PTMTs. This regards the following variables: gender, ethnicity, education, number of motorcycles, number of scooters 45, number of scooters 25, and number of electric bicycles. Therefore, these variables are excluded from the network since they are not of interest to this research. By clearing the network of these variables the network is more explicit and only important variables are taken into account.

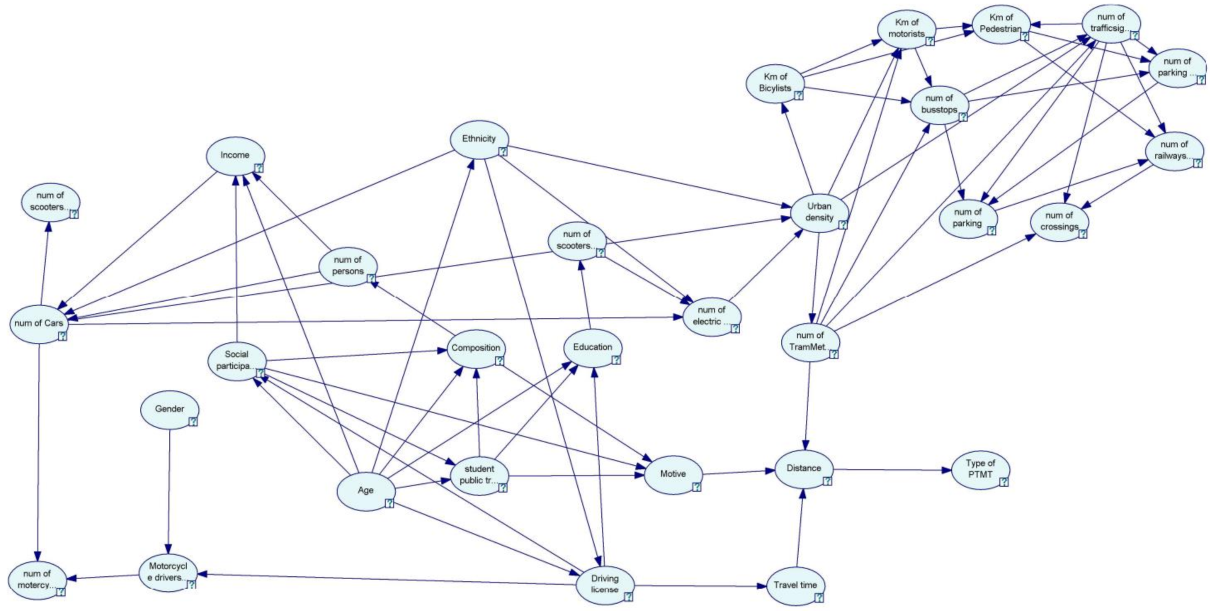


Figure 17: Learned network's structure 2 – Types of PTMTs

4.2.2. Final learned network – Types of PTMTs

After excluding the variables as mentioned above in 4.2.1. The learned network, the following network is learned (Figure 18). The network is also visible in Appendix E. For this network, the same settings are used as described above regarding the max parent count parameter and the background knowledge. It can be seen that the only variable directly influencing the variable types of PTMTs is the variable distance. Within this network, the relationships (arcs) are bolder based on the strength of influence between the variables through the function 'Strength of influence'. Also, all influence strengths between the variables within the network are extracted. Table 16 shows the extracted influence strengths between the variables. The measuring scale is from 0 to 1, whereby 1 is the highest possible value and thus a strong relationship. The strongest relationships (above 0.300) are marked in bold within the Table like the arcs in the learned network.

Table 16: Strength of influence network 2 – Types of PTMTs

	Age	Social particpation	Driving license	Student PT smartcard	Income	Num of cars	Composition	Urban density	Motive	Distance	Travel time	Types of PTMTs
Age	x	**0.551**	0.205	0.264	-	0.199	0.235	0.146	-	-	-	-
Social particpation	-	x	0.176	0.239	0.216	-	0.198	-	0.238	-	-	-
Driving license		-	x	-	-	0.236	-	-	-	-	0.078	-
Student PT smartcard	-	-	0.182	x	-	-	0.218	0.221	0.236	-	-	-
Income	-	-	-	-	x	-	-	-	-	-	-	-
Num of cars	-	-	-	-	0.206	x	-	0.165	-	-	-	-
Composition	-	-	-	-	0.232	**0.309**	x	-	0.177	-	-	-
Urban density	-	-	-	-	-	-	-	x	-	-	-	-
Motive	-	-	-	-	-	-	-	-	x	0.191		
Distance	-	-	-	-	-	-	-	-	-	x	-	0.266
Travel time	-	-	-	-	-	-	-	-	-	**0.307**	x	-
Types of PTMTs	-	-	-	-	-	-	-	-	-	-	-	x

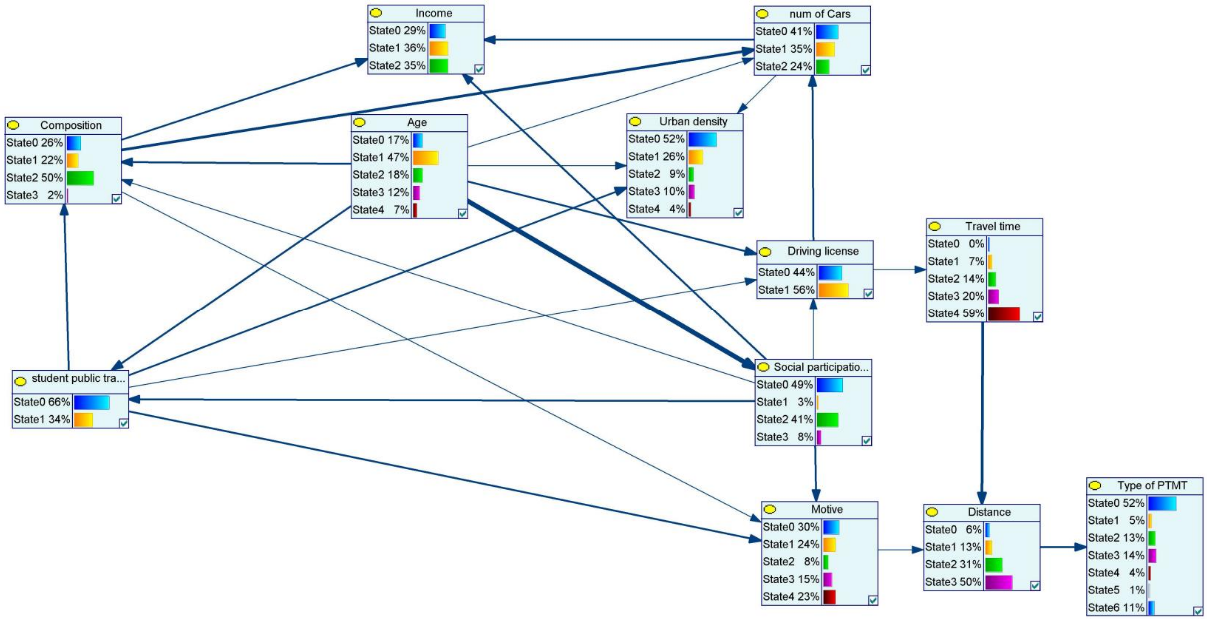


Figure 18: Final learned network 2 with CPTs – Types of PTMTs

4.2.3. Findings final network 2 – Types of PTMTs

Per characteristic, the relationships will be discussed between a certain variable and the dependent variable types of PTMTs.

Looking at the personal characteristics (age, student public transportation smartcard, driving license, and social participation) there are no direct relationships with the variable types of PTMTs. Also when setting hard evidence within age, student public transportation smartcard, driving license, and social participation the probabilities of the variable types of PTMTs are not changed or not significantly (Appendix F). So, none of the personal characteristics influences directly or indirectly the dependent variable types of PTMTs.

The household characteristics include household composition, income, and the number of cars. These variables do not have a direct relationship with the variable types of PTMTs. Thereby there is also no significant indirect relationship based on the set evidence (Appendix G). So, also none of the household characteristics influences directly or indirectly the dependent variable types of PTMTs.

For the environmental variables, only urban density is included. Also, this variable does not directly or indirectly significantly influence the variable types of PTMTs (Appendix H).

On the other hand, the trip characteristics appear to be influencing the dependent variable types of PTMTs. Solely the variable distance is directly influencing the variable types of PTMTs. Table 17 shows the probability distribution with and without hard evidence. It seems that shorter distance trips (30-50 km) are 96% conducted according to the type of PTMT: walking (first leg) + Public transport (main leg) + walking (last leg). For the distance 51-100 kilometers, this is decreased to 83% and decreases further with longer-distance trips. People tend to use a bicycle as the mode for the first or last leg of their trips if it is longer than 100 kilometers. The car seems to be a less used transportation mode for the first leg as it is only 5% in combination with walking and only 2% in combination with cycling.

Table 17: Updated probabilities hard evidence distance

Types of PTMTs	No evidence	Evidence 30-50 km distance	Evidence 51-100 km distance	Evidence 101-250 km distance	Evidence > 250 km distance
Walking + Public transport + Walking	0.523	0.958	0.830	0.530	0.384
Cycling + Public transport + Cycling	0.053	0.004	0.006	0.039	0.080
Cycling + Public transport + Walking	0.128	0.017	0.037	0.143	0.157
Walking + Public transport + Cycling	0.142	0.008	0.057	0.168	0.166
Car + Public transport + Walking	0.036	0.004	0.016	0.022	0.055
Car + Public transport + Cycling	0.011	0.004	0.002	0.003	0.019
Other	0.106	0.004	0.053	0.096	0.138

The variable motive only influences the dependent variable types of PTMTs indirectly. For shopping/grocery, 61% tend to walk as first and last leg. For work, this is 50%, and for leisure 55%. The type of public transportation and cycling (first and/ or last leg) is slightly less used for the motive of shopping/ grocery. These findings of the variable motive are presented in Table 18.

Table 18: Updated probabilities hard evidence motive

Types of PTMTs	No evidence	Evidence Work	Evidence Education	Evidence Shopping/ grocery	Evidence Leisure	Evidence Other
Walking + Public transport + Walking	0.523	0.499	0.516	0.613	0.547	0.515
Cycling + Public transport + Cycling	0.053	0.056	0.054	0.040	0.049	0.056
Cycling + Public transport + Walking	0.128	0.134	0.131	0.105	0.123	0.129
Walking + Public transport + Cycling	0.142	0.149	0.146	0.119	0.137	0.142
Car + Public transport + Walking	0.036	0.038	0.036	0.029	0.034	0.039
Car + Public transport + Cycling	0.011	0.012	0.011	0.008	0.010	0.012
Other	0.106	0.111	0.107	0.087	0.101	0.108

The last variable to be discussed is travel time. This variable is indirectly influencing the dependent variable types of PTMTs. Travel time is also like the variable distance. When the travel time increases the transportation mode of walking for the first and last leg decreased (Table 19). So, for a travel time shorter than 60 minutes, 74% use walking as the first and last mode. This is for trips between 60 and 90 minutes 64%, for trips between 90 and 120 minutes 55%, and for trips more than 120 minutes, this is 46%.

Table 19: Updated probabilities hard evidence travel time

Types of PTMTs	No evidence	Evidence > 30 min travel time	Evidence 31-60 min travel time	Evidence 61-90 min travel time	Evidence 91- 120 min travel time	Evidence > 120 min travel time
Walking + Public transport + Walking	0.523	0.718	0.746	0.633	0.550	0.460
Cycling + Public transport + Cycling	0.053	0.025	0.021	0.034	0.046	0.064
Cycling + Public transport + Walking	0.128	0.078	0.071	0.103	0.124	0.143
Walking + Public transport + Cycling	0.142	0.090	0.082	0.119	0.141	0.156
Car + Public transport + Walking	0.036	0.020	0.018	0.024	0.031	0.044
Car + Public transport + Cycling	0.011	0.005	0.005	0.006	0.009	0.014
Other	0.106	0.063	0.058	0.082	0.098	0.120

4.2.3. Conclusion network 2 – Types of PTMTs

The network does not correspond with the proposed conceptual model as stated in Section 2.4. Conclusion. Looking at the types of PTMTs without any evidence it seems that the car is less used for the first leg of a multimodal trip for shorter distances. Walking appears to be the dominant transportation mode for the first and last leg when conducting a multimodal trip whereas the main mode is public transport. Thereby, the bicycle is also frequently used for the first leg (13%) and the last leg (14%) in combination with walking compared to the car. It seems that only the trip characteristics (motive, distance, and travel time) significantly influence the types of PTMTs. It appears that the type: walking (first leg) + public transport (main leg) + walking (last leg) is mostly (96%) used for shorter distance trips (30-50 km) (Figure 19). This decreased to 83% for longer trips and for longer distances the bicycle become more used as first or last leg mode, however, walking stays dominant over cycling. This also applies to the variable travel time. Walking stays the dominant mode for the first and last leg however, it decreases when the travel time increases (Figure 21). When setting evidence within the variable motive on each category the type of PTMT walking (first leg) + Public transport (main leg) + walking (last leg) stays above 50% however, it seems that shopping/grocery is the biggest motive with 61% to conduct this type of PTMT (Figure 20). The personal, household, and environmental characteristics do change the trip variable motive when setting hard evidence however, indirectly this does not influence the types of PTMTs. To conclude, solely the trip characteristics (motive, distance, and travel time) appear to influence the types of PTMTs. Whereas the variable distance is the strongest (0.266) and is the only direct relationship with the dependent variable types of PTMTs.

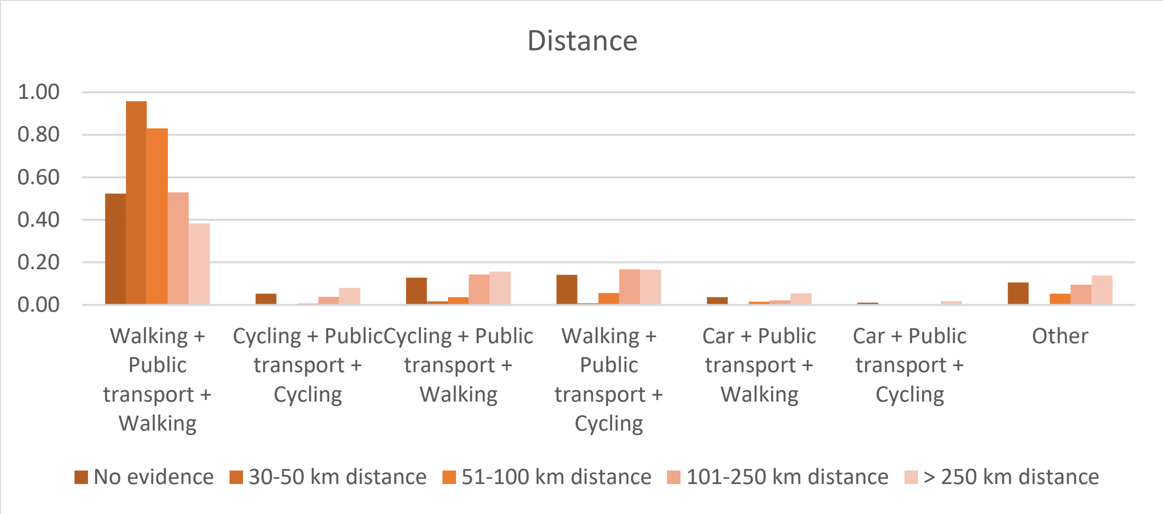


Figure 19: Updated probability hard evidence distance

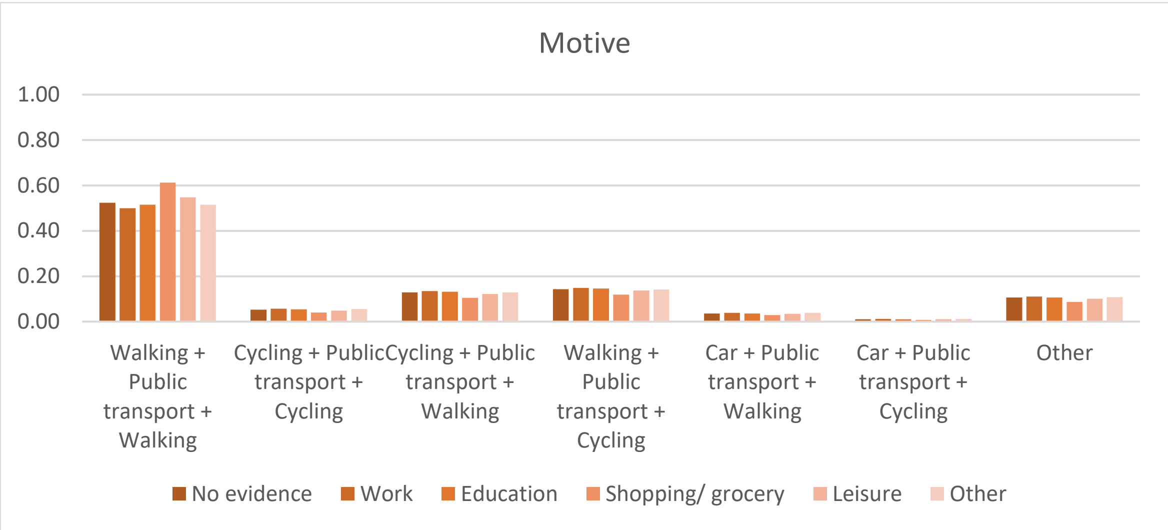


Figure 20: Updated probability hard evidence motive

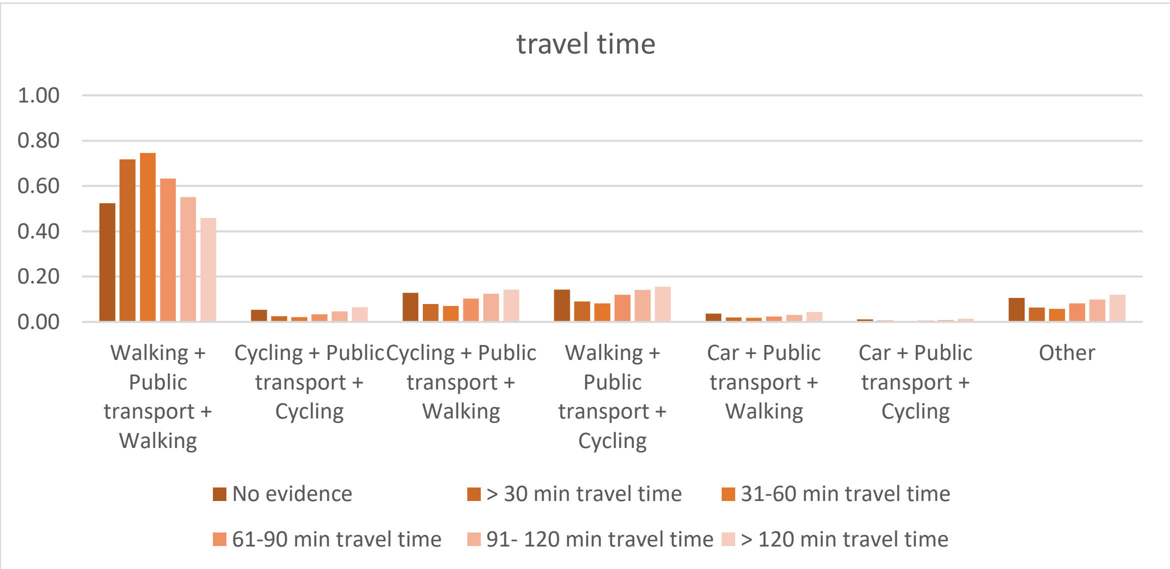


Figure 21: Updated probability hard evidence travel time

4.3. Conclusion results

Within this section, the proposed conceptual model based on the literature is revised based on the results of the BBNs. Due to these revised conceptual models, sub-questions 2 and 3 will be answered.

Sub-question 2 is stated as follows: *Which variables directly and indirectly influence the choice of trip type between a unimodal car trip and a public transportation based multimodal trip?* The variables directly influencing the choice of trip type between a unimodal car trip and a public transportation based multimodal trip within The Netherlands are student public transportation smartcard, motive, and number of cars within a household. These variables were subdivided into the personal, trip, and household characteristics. The environmental characteristic urban density, the personal characteristics age, social participation, driver's license, and the household characteristics income, and composition are indirectly influencing the choice of a trip between a unimodal car trip and a PTMT trip. These relationships are displayed in a revised conceptual model and are presented in Figure 22 and Appendix I. Note, this conceptual model is an illustration of the estimated BBN. The direct relationships are illustrated with a bold line, the indirect relationships are illustrated with a black line, and the other relationships within the BBN are illustrated as black dotted lines.

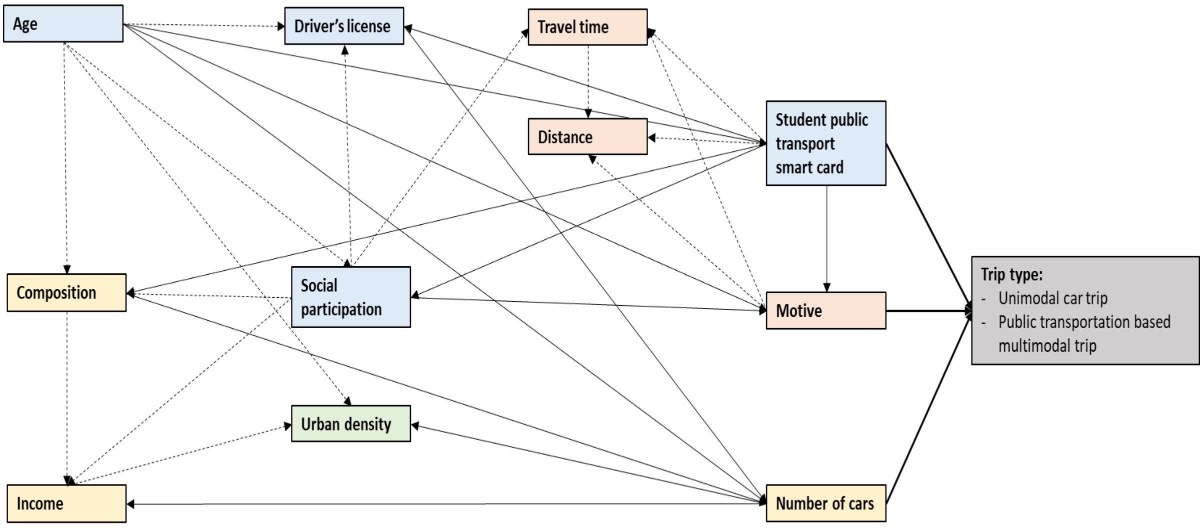


Figure 22: Revised conceptual model – Trip type

Sub-question 3 is stated as follows: *Which variables directly and indirectly influence types of public transportation based multimodal trips?* The trip characteristics (motive, distance, and travel time) are influencing the types of PTMTs within The Netherlands. Only the variable distance is directly influencing the dependent variable types of PTMT. The motive for conducting a type of PTMT and the travel time of the trip are indirectly related to the type of PTMT through the variable distance. The other characteristics (personal, household, and environment) do not significantly influence the types of PTMTs. These relationships are displayed in a revised conceptual model and are presented in Figure 23 and Appendix J. Note, this conceptual model is an illustration of the estimated BBN. The direct relationships are illustrated with a bold line, the indirect relationships are illustrated with a black line, and the other relationships within the BBN are illustrated as black dotted lines.

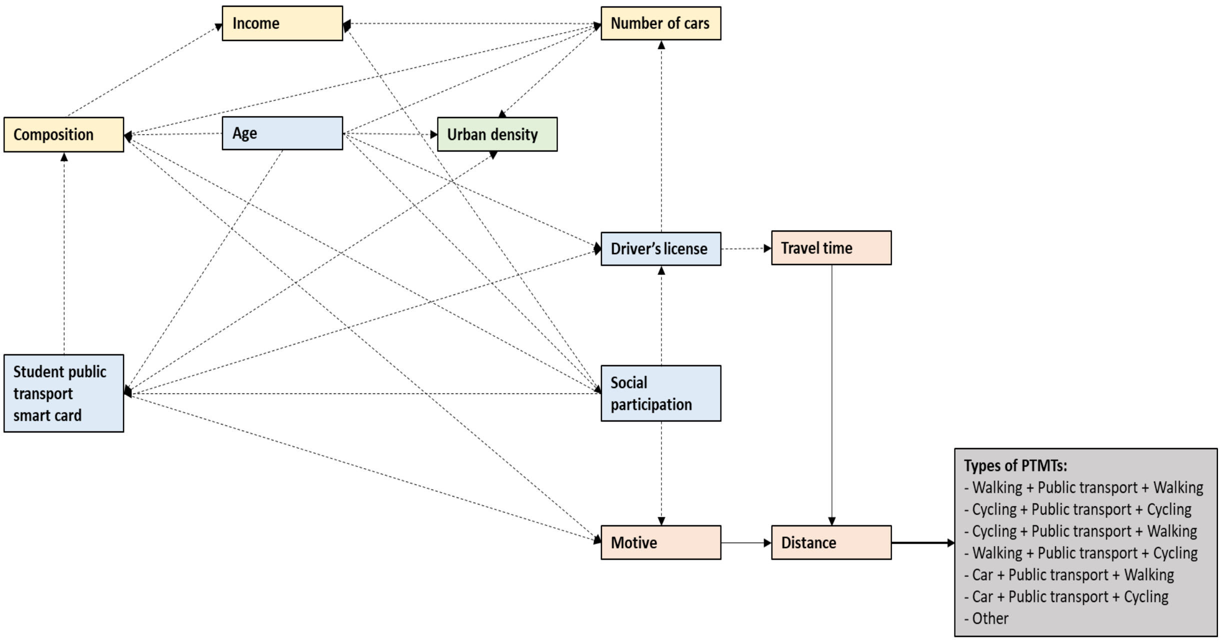


Figure 23: Revised conceptual model – Types of PTMTs

4.4. Discussion

Within this section, the results of the Bayesian Belief Networks are evaluated and interpreted. Also, a comparison to the literature will be discussed. First network 1 – Trip type is addressed and thereafter Network 2 – Types of PTMTs.

4.4.1. Network 1 –Trip type

Network 1 presents the variables that influence the choice of trip type between a unimodal trip and a PTMT. Within the used data set (ODiN 2021) about 7% of all trips in 2021 are PTMTs. This shows a growth of 2.5% compared to the findings of CBS in 2018 which included all types of multimodal trips. However, this is still a small percentage in comparison with the conducted unimodal car trips. This distribution shows that there is excessive use of unimodal car trips, which confirms that there is a need of reducing car use and ownership to decrease environmental, economic, and social problems. When looking at the distribution of the variables driver's license and student public transportation smartcard there is an extreme difference. This is logical since only students, which is 13% of the respondents, receive a smartcard within The Netherlands. For the variable driver's license, this is also reasonable since 8% of the respondents are younger than the age of 18. Within the literature, this difference within variables is also found. For example Ma (2015) with the variable driver's license and Kemperman & Timmermans (2014a) with the variable possession of a bike. These distributions do not influence the functioning of the network. However, it is considerable if the predictions are reasonable when setting hard evidence on these categories. This depends on the number of cases that are left, which are the basis for predicting the probabilities of other variables when setting evidence.

Looking at the results and the literature, the results confirm that the trip characteristic motive influences the choice of trip type between a unimodal car trip and a PTMT (Almasri & Alraee, 2013; Basheer et al., 2019; Cho, 2013; Nes, 2002). This applies also to the number of cars within a household (Basheer et al., 2019; Kim et al., 2007; Ma, 2015). However, the trip characteristics travel time and distance do not significantly influence the trip type (unimodal car trips or PTMTs). This result differs from the literature (Cho, 2013; Nes, 2002; Racca & Ratledge, 2003). This could be through the exclusion of trips shorter than 30 kilometers and the exclusion of other unimodal mode trips than unimodal car trips. Also, the variables gender and education are not included in the final network. According to the

Bayesian Belief Network, these variables did not influence the dependent variable trip type. This is also in contrast to the studies of Kim et al. (2007), Senbil et al. (2009), and Liu et al. (2016). This difference may appear due to the different categorizations used. This research does not categorize the public transportation modes, this is all assigned into public transport instead of train, bus, metro, and tram. Thereby, the scope of only including unimodal car trips and PTMTs could be a cause that some personal characteristics appear not to be related to the choice of trip type.

4.4.2. Network 2 –Types of PTMTs

Network 2 presents the variables that directly and indirectly influence types of PTMTs. The multimodal trips are extracted from the data set of network 1. Therefore, within this network, only 56% of the respondents have a driver's license. Network 1 shows that 35% of all multimodal trips are students and 55% are under the age of 35 which can explain the distribution of the variable driver's license within network 2. Also, the differences in the conditional probability tables (CPTs) of some other variables are due to the different case numbers of both models. Thereby is the distribution of the dependent variable types of PTMTs also non-normal. However, the types of PTMTs are all specified based on the frequencies within the data set as Basheer et al. (2019) and Ma (2015) did in their studies.

Looking at the variables researched within the literature the following results are noticeable. Keijer & Rietveld (2007) found that the distance to a railway station influences the mode choice. This variable is researched within this study with the variable number of railway stations within the respondent's zip code because the data concerning the distance to a railway was lost due to the reformed data set as mentioned in 3.3. Data preparation and solely data was found of the number of railway stations within the respondent's zip code. However, this variable is excluded from the BBN since there was no relationship found with the dependent variable types of PTMTs. This could be due to the difference in scope of both researches since Keijer & Rietveld (2007) focus solely on multimodal trips where the train is the main transport mode. Subsequently, Rietveld (2000b) states that cycling is used mostly for the first leg and walking for the last trip. This is not confirmed by the conditional probability table of the variable types of PTMTs without evidence set. The apparent difference is also possible due to the difference in scope (excluding other public transportation modes than the train as the main mode). Thereby within this research, it is assumed that all PTMTs cannot begin with the mode public transport due to the access and egress part of utilizing the public transportation system which is not done in the research of Rietveld (2000b). In addition, Nes (2002) found that the trip characteristics: trip distance, type of destination, and trip purpose have the greatest impact on PTMTs. The results from network 2 confirm that the variable distance is the greatest impact on the variable types of PTMTs. Also, the variable motive is significantly related to the types of PTMTs as Nes (2002) states. However, the variable type of destination is not included in this research. He subdivided the destination area the same as the variable urban density within this research. The difference between these variables is that for this research urban density is included as an environmental characteristic based on the zip code of a respondent's residence and not as a trip characteristic that is based on the destination zip code. The variable urban density (of the destination location) was not possible to include as trip characteristics since the destination zip codes are lost due to the reforming process of the ODiN data set to extract PTMTs.

5. Conclusion & recommendations

This chapter discusses the conclusion of this thesis by answering the stated research question in 1.2.2. Research question(s). Also, the limitations and recommendations will be discussed.

5.1. Conclusion

The rise in car ownership and use has resulted in various environmental, economic, and social problems. Public transportation based multimodal trips (PTMTs) can help to decrease the negative impact of excessive use of private motorized transportation. Promoting PTMTs as an alternative to unimodal car trips requires an understanding of individual travel choices and the variables that influence the mode choice. In particular, the organization of transportation modes for the first, main, and last leg of PTMTs is a phenomenon in travel behavior studies that seems to require more attention. Because the choice of travel mode and/or available service at one leg of the trip influences the entire trip. Therefore, this study analyzes the direct and indirect influence of variables (categorized in personal, household, environmental, and trip characteristics) on types PTMTs.

Two Bayesian Belief Networks (BBNs) are estimated. The first network (network 1 – Trip type) formulate and estimates the relationships between the included variables and the choice of trip type (unimodal car trips or PTMTs). As, the aim is to shift unimodal car trips to PTMTs to decrease environmental, economic, and social problems. The second network (network 2– Types of PTMTs) formulate and estimates the relationships between the included variables and the types of PTMTs. This network helps governments to improve or create policies concerning types of PTMTs.

The results of network 1 – Trip type show that the variables student public transportation smartcard, motive, and number of cars within a household are directly influencing the trip type. It seems that if a person has a student public transportation smartcard the probability of conducting a PTMT increase significantly. Thereby, 43% of all PTMTs are conducted for the motive education. This is 8% for work, for leisure 6%, and only 3% for shopping/grocery. Subsequently, it seems that if there are no cars within a household the probability of conducting a PTMT is 36%, which is significantly lower if there is one car within a household (5%). The indirect influencing variables are age, social participation, driver's license, income, composition, and urban density. The variables travel time and distance do not significantly influence the choice of trip type. The relationships are illustrated in Figure 22: Revised conceptual model – Trip type (Section 4.3. Conclusion results).

The results of network 2 – Types of PTMTs show that the types of PTMTs in The Netherlands are influenced by the trip characteristics: motive, distance, and travel time. Among these variables, only distance has a direct relationship with the dependent variable types of PTMTs. On the other hand, the variables motive and travel time have an indirect relationship with the types of PTMTs through distance. It appears that for shorter distance trips (30-50 km) the type walking (first leg) + public transport (main leg) + walking (last leg) is mostly used (96%). For longer trips, this decreases to 83%. Also, the bicycle is used more for longer trips however, walking continued to be preferred over cycling. This is the same for the variable travel time. For the first and last leg, walking continues to be the dominant mode, but as the travel time increase, fewer people choose to walk. Thereby, it seems that shopping/grocery is the strongest motive to conduct the type walking (first leg) + Public transport (main leg) + walking (last leg). Furthermore, the personal, household, and environmental characteristics do not have a significant impact on the types of PTMT. The relationships are illustrated in Figure 23: Revised conceptual model – Types of PTMTs (Section 4.3. Conclusion results).

Based on the reviewed literature, several variables have been found to influence the mode choice of PTMTs. However, the results of this study reveal some deviations from the findings of previous research. For instance, while Keijer & Rietveld (2007) found that the distance to a railway station influences mode choice the most, this variable (number of railways stations within the respondent's zip code) did not show any significant relationship with types of PTMTs in this study. This is in contrast to the results of network 2 which shows that the variable distance has the greatest impact on the dependent variable types of PTMTs. This finding is in line with the findings of Nes (2002) that found that the variables trip distance, type of destination, and trip purpose have the greatest impact on PTMTs. The variable motive is also found in this thesis to be significantly influencing the types of PTMTs, which supports Nes's (2002) findings. The variable type of destination is not included in this research so no conclusion can be drawn up regarding if the type of destination influences the dependent variable types of PTMTs.

5.2. Limitations

This study assigns all public transportation modes (train, bus, tram, and metro) into one mode. This limits the results of the transportation modes separately. It is plausible that individuals that use transportation modes such as a tram or metro for the main leg within PTMTs travel less distance. Trips with a tram or metro are often conducted within a city and therefore compete with the transportation modes such as walking or cycling. For the transportation mode train, this is less likely because the train is often used for longer distances for example between two cities. By assigning all the public transportation modes into one mode, the recommendations are only applicable to public transportation in general. Thereby, this study only includes trips that are longer than 30 km since then it becomes an intriguing choice (Nes, 2002) and due to this, the focus is only on reducing unimodal car trips and increasing PTMTs. However, trips with the tram or metro as the mode in the main trip leg are probably less than 30 km. So, no results can be given if these trips can reduce shorter-distance unimodal car trips.

Subsequently, another limitation is the setup of the data set of ODiN. The legs of the multimodal trips are reported as multiple cases as mentioned in Section 3.3. Data preparation. Therefore for this research, it was needed to reform these trips into one case. Due to this, data of the trips' main and last leg of variables such as travel time, arrival and departure times, and the destination location is lost. So, this study was only able to include the complete travel time of the public transportation based multimodal trip and not the travel time per leg. However, multimodal trips are influenced indirectly by the variable travel time so not knowing the travel time per trip leg could have limited the results. Moreover, a limitation is therefore also the use of secondary data because the researcher is depending on the setup of the questionnaire and cannot gain more or other information about the respondents and their trips.

The last limitation concerns the uses of data from the year 2021. In December 2019 a virus, COVID-19, broke out and was promulgated in 2020 as a global pandemic (WHO, 2020). Within The Netherlands, measures were taken by the government to limit travel. This resulted in more private mobility and less shared mobility like public transportation (Przybylowski et al., 2021). This could imply that the results of this study could differ from pre-COVID-19 circumstances. Despite this, this study shows that a Bayesian Belief Network is suitable to formulate the direct and indirect relationships between the variables and estimate the conditional probabilities. The relationships among the variables seem to be rather complex and due to the learned network, these relationships are identified and clearly visible within a network. Also, entering hard evidence within the network helps governments to understand the potential impact of different variables. This provides support to the decision-making process concerning new or improved policies.

5.3. Recommendations

For future research, it is interesting to research the mode choice of public transportation modes of multimodal trips separately instead of assigning all public transportation modes into one mode. This will provide a more in-depth understanding of the different mode choices within the multimodal trip and how they affect the entire trip. Thereby, this study could not include the trip characteristics of the first, main, and last trip legs, solely the trip characteristics of the complete PTMT. It would be interesting to see if there is a difference in results when this data is also included. Lastly, it would be an option to use primary data instead of secondary data. By using primary data the researcher can include all variables of interest in the desired setup. Variables that are excluded within this research due to the reforming process or unavailable data such as individuals income, distance to the city center, trip costs, land use mix, and destination urban density are interesting to include in future research to see if they are direct of directly influencing the types of PTMTs. Thereby, it is interesting to see if there is a difference between pre- during- and post-COVID-19 and the variables that influence the unimodal car trips and the types of PTMTs. Due to the taken measures, people traveled less and therefore other variables could be influencing the choice of trip type.

For the practice, the Bayesian Belief Network concerning unimodal car trips or PTMTs (network 1 – Trip type) helps to give insight into what kind of individuals, households, or trips could be stimulated to conduct PTMTs and to decrease the unimodal car trips. It seems that namely the student public transportation smartcard, number of cars in a household influence, and motive, influence the choice of trip type. For example, governments can create policies concerning the number of cars in a household regarding taxes per car, or in very highly urban areas car-free zones so households are perhaps less likely to own a car (Gonzalez et al., 2021; Rietveld, 2000b; Tonne et al., 2008). Another option is to stimulate businesses to provide a public transportation smartcard for employees to decrease the work-related unimodal car trips since of all trips 92% is a unimodal car trip and is related to work. So to conclude, the network shows all direct and indirect variables that are influencing the trip types. Based on these results governments have insight into variables that can stimulate PTMTs and policies can be created or optimized. The second Bayesian Belief network created focuses solely on specific types of PTMTs (network 2 – Types of PTMTs). Despite this network already showing the individuals that conduct PTMTs, governments should not only focus on decreasing unimodal car trips but also improving PTMTs. The network shows that the variable distance is directly influencing the types of PTMTs and the variables motive and travel time indirectly influence the types of PTMTs. For example, governments can improve the public transportation system by focusing on reducing the travel time of PTMTs to make them a more appealing option. This could be done by creating better connections or developing more or efficient transit stops and train stations, so the travel distance and time would decrease. This results in fewer PTMTs with the car as the mode for the first or last leg, and more active travel modes.

Reference

Abenoza, R. F., Cats, O., & Susilo, Y. O. (2019). How does travel satisfaction sum up? An exploratory analysis in decomposing the door-to-door experience for multimodal trips. Transportation, 46(5), 1615–1642. https://doi.org/10.1007/s11116-018-9860-0

ACEA. (2022). Vehicles in use in Europe 2022. January, 1–20. https://www.acea.auto/publication/report-vehicles-in-use-europe-2022/#:~:text=In 2020%2C the EU passenger,on the road in total.

Aktas, E., Ülengin, F., & Sahin, S. önsel. (2007). A decision support system to improve the efficiency of resource allocation in healthcare management. 41, 130–146. https://doi.org/10.1016/j.seps.2005.10.008

Almasri, E., & Alraee, S. (2013). Factors Affecting Mode Choice of Work Trips in Developing Cities — Gaza as a Case Study. 2013(October), 247–259.

Arentze, T., & Timmermans, H. (2004). Multistate supernetwork approach to modelling multi-activity, multimodal trip chains. International Journal of Geographical Information Science, 18(7), 631–651. https://doi.org/10.1080/13658810410001701978

Arentze, Theo, & Timmermans, H. (2008). Regimes in social-cultural events-driven activity sequences: Modelling approach and empirical application. Transportation Research Part A: Policy and Practice, 43(4), 311–322. https://doi.org/10.1016/j.tra.2008.11.010

Arsanjani, J. J., Barron, C., Bakillah, M., & Helbich, M. (2013). Assessing the Quality of OpenStreetMap Contributors together with their Contributions.

Ashiabor, S., Baik, H., & Trani, A. (2000). Logit Models for Forecasting Nationwide Intercity Travel Demand in the United States. https://doi.org/10.3141/2007-01

Audenhove, F.-J. Van, Rominger, G., Arby, H., & Tauvel, M. (2021). *How to realize the promise of Mobility-as-a-Service*. September.

Bakker, N. (2019). *Gemiddelde autorit is 19 kilometer en duurt een kwartier.* Recieved from: https://www.verkeerskunde.nl/artikel/gemiddelde-autorit-is-19-kilometer-en-duurt-een-kwartier#:~:text=Als%20het%20totaal%20aantal%20afgelegde,uit%20andere%20bronnen%20bekend%20is.

Basheer, M. A., van der Waerden, P., Kochan, B., Bellemans, T., & Raheel Shah, S. A. (2019). Multi-stage trips: An exploration of factors affecting mode combination choice of travelers in England. Transport Policy, 81(June), 95–105. https://doi.org/10.1016/j.tranpol.2019.06.007

BayesFusion. (2022). GeNIe Modeler.

BayesFusion. Genie 2.1. 2022. Available online: https://www.bayesfusion.com/news/ (accessed on 24 October 2022)

Bhat, R. (1997). COVARIANCE HETEROGENEITY IN NESTED LOGIT MODELS : ECONOMETRIC STRUCTURE AND APPLICATION TO. 31(I), 11–21.

Blumenberg, E., & Pierce, G. (2014). Multimodal travel and the poor: Evidence from the 2009 National Household Travel Survey. Transportation Letters, 6(1), 36–45. https://doi.org/10.1179/1942787513Y.0000000009

Bovy, P. H. L., & Hoogendoorn-Lanser, S. (2005). Modelling route choice behaviour in multi-modal transport networks. Transportation, 32(4), 341–368. https://doi.org/10.1007/s11116-004-7963-2

Bromley, J., Jackson, N. A., Clymer, O. J., Giacomello, A. M., & Jensen, F. V. (2005). The use of Hugin Ò to develop Bayesian networks as an aid to integrated water resource planning *. 20, 231–242. https://doi.org/10.1016/j.envsoft.2003.12.021

CBS. (2022a). Hoeveel reisden inwoners van Nederland en hoe ? https://www.cbs.nl/nl-nl/visualisaties/verkeer-en-vervoer/personen/hoeveel-reisden-inwoners-van-nederland-en-hoe-

CBS. (2022b). Onderweg in Nederland (ODiN) 2021 - Onderzoeksbeschrijving Over deze publicatie. 1–36. https://www.cbs.nl/nl-nl/longread/rapportages/2022/onderweg-in-nederland--odin---2021-onderzoeksbeschrijving?onepage=true

CBS & OViN. (2020). Multimodale verplaatsingen, 2010-2018. september, 3–7. https://www.clo.nl/indicatoren/nl2189-ontwikkeling-multimodale-verplaatsingen

Chen, R. (2014). Simulation Modeling and Application of Travel Mode Choice Based on Bayesian Network. 19–25.

Cheng, J., Bell, D. a., & Liu, W. (1997). An Algorithm for Bayesian Belief Network Construction from Data. Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics (AI & STAT ’97), 83–90. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.4.9611

Cho. (2013). The factors that affect long-distance travel mode choice decisions and their implications for transportation policy.

Cornalba, C., & Giudici, P. (2004). Statistical models for operational risk management. 338, 166–172. https://doi.org/10.1016/j.physa.2004.02.039

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach.

Dissanayake, D., & Morikawa, T. (2010). Investigating household vehicle ownership , mode choice and trip sharing decisions using a combined revealed preference / stated preference Nested Logit model : case study in Bangkok Metropolitan Region. Journal of Transport Geography, 18(3), 402–410. https://doi.org/10.1016/j.jtrangeo.2009.07.003

Gao, Y., Rasouli, S., Timmermans, H., & Wang, Y. (2020). Prevalence of alternative processing rules in the formation of daily travel satisfaction in the context multi-trip, multi-stage, multi-attribute travel experiences. Transportation, 47(3), 1199–1221. https://doi.org/10.1007/s11116-018-9941-0

Gonzalez, J. N., Perez-doval, J., Gomez, J., & Vassallo, J. M. (2021). What impact do private vehicle restrictions in urban areas have on car ownership ? Empirical evidence from the city of Madrid. Cities, 116(August 2020), 103301. https://doi.org/10.1016/j.cities.2021.103301

Guo, Z., & Wilson, N. H. M. (2011). Assessing the cost of transfer inconvenience in public transport systems : A case study of the London Underground. Transportation Research Part A, 45(2), 91–104. https://doi.org/10.1016/j.tra.2010.11.002

Ha, J., Lee, S., & Ko, J. (2020). Unraveling the impact of travel time , cost , and transit burdens on commute mode choice for different income and age groups. Transportation Research Part A, 141(May), 147–166. https://doi.org/10.1016/j.tra.2020.07.020

Hausman, J., & McFadden, D. (1984). Specification Tests for the Multinomial Logit Model. 52(5), 1219–1240.

Heckerman, D., Mandani, A., & Wellman, M. P. (1995). Real-world applications of Bayesian networks. Communications of ACM, 38, 24–26.

Heinen, E., & Mattioli, G. (2019). Multimodality and CO2 emissions: A relationship moderated by distance. Transportation Research Part D: Transport and Environment, 75(September), 179–196. https://doi.org/10.1016/j.trd.2019.08.022

Hietanen, S. (2014). ' Mobility as a Service ' – the new transport model ? 12(2), 2–4.

Hoffman, S. D., & Duncan, G. J. (1988). A Comparison of Choice-Based Multinomial and Nested Logit Models : The Family Structure and Welfare Use Decisions of Divorced or Separated Women A Comparison of Choice-Based Multinomial and Nested Logit Models The Family Structure and Welfare. 23(4), 550–562.

Hoogendoorn-Lanser, s, & Nes, R. van. (2005). HOME-ACTIVITY APPROACH TO MULTI-MODAL TRAVEL CHOICE MODELLING. Transport & Planning, Technische Universiteit Delft, 9, 1–29. http://www.ncbi.nlm.nih.gov/pubmed/810049%0Ahttp://doi.wiley.com/10.1002/anie.197505391%0Ahttp://www.sciencedirect.com/science/article/pii/B9780857090409500205%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/21918515%0Ahttp://www.cabi.org/cabebooks/ebook/20083217094

Hu, H., Xu, J., Shen, Q., Shi, F., & Chen, Y. (2018). Travel mode choices in small cities of China : A case study of Changting. Transportation Research Part D, 59(February), 361–374. https://doi.org/10.1016/j.trd.2018.01.013

Jakobs, K., & Mitchel, S. (2020). OpenStreetMap quality assessment using unsupervised machine learning methods. Ottawa, Canada: John Wiley & Sons Ltd.

Javid, M. A., Okamura, T., Nakamura, F., Tanaka, S., & Wang, R. (2016). People ' s Behavioral Intentions Towards Public Transport in Lahore : Role of Situational Constraints , Mobility Restrictions and Incentives. 20, 401–410. https://doi.org/10.1007/s12205-015-1123-4

Jittrapirom, P., Caiati, V., Feneri, A., Ebrahimigharehbaghi, S., & Alonso-, M. J. (2022). PDF hosted at the Radboud Repository of the Radboud University Nijmegen Mobility as a Service : A Critical Review of Definitions , Assessments of Schemes , and Key Challenges. https://doi.org/10.17645/up.v2i2.931

Jong, D. A. de;, & Steen, M. J. van der. (2001). Ketenmobiliteit in beeld, wereldse successen.

Jong, G. C. De. (2008). The Driving Factors of Passenger Transport. 8(8).

Jonuschat, H., Stephan, K., & Schelewsky, M. (2015). Understanding multimodal and intermodal mobility. Transport and Sustainability, 7(April 2020), 149–176. https://doi.org/10.1108/S2044-994120150000007018

Kahn, C. E., Roberts, L. M., Shaffer, K. A., & Haddawy, P. (1997). CONSTRUCTION OF A BAYESIAN NETWORK FOR In 1995 , an estimated 183 , 400 women in the United States were newly diagnosed with breast cancer , and 46 , 240 died of the disease [l]. Screening mammography effectively detects early breast cancers and can inc. 1, 19–29.

Keijer, M. J. N., & Rietveld, P. (2007). How do people get to the railway station ? The dutch experience HOW DO PEOPLE GET TO THE RAILWAY STATION ? THE DUTCH EXPERIENCE. 1060. https://doi.org/10.1080/03081060008717650

Kemperman, A., & Timmermans, H. (2014a). Environmental Correlates of Active Travel Behavior of Children. Environment and Behavior, 46(5), 583–608. https://doi.org/10.1177/0013916512466662

Kemperman, A., & Timmermans, H. (2014b). Landscape and Urban Planning Green spaces in the direct living environment and social contacts of the aging population. Landscape and Urban Planning, 129, 44–54. https://doi.org/10.1016/j.landurbplan.2014.05.003

Kemperman, A., Van Den Berg, P., Weijs-Perrée, M., & Uijtdewillegen, K. (2019). Loneliness of older adults: Social network and the living environment. International Journal of Environmental Research and Public Health, 16(3). https://doi.org/10.3390/ijerph16030406

KiM. (2017). Verkenning alternatieve inwintechnieken voor verplaatsingsgegevens Kennisinstituut voor Mobiliteitsbeleid | KiM.

KiM. (2022). Het wijdverbreide autobezit in Nederland.

Kim, S., Ulfarsson, G. F., & Hennessy, J. T. (2007). Analysis of light rail rider travel behavior : Impacts of individual , built environment , and crime characteristics on transit access. 41, 511–522. https://doi.org/10.1016/j.tra.2006.11.001

Kitchin, R. (2014). The data revolution. Big data, Open data, Data infrastructures & Their consequences. Sage.

Koh, P. P., & Wong, Y. D. (2016). Influence of Socio-Demography and Operating Streetscape on Last-Mile Mode Choice. 19(2), 38–54.

Krygsman, S., & Dijst, M. (2001). Multimodal trips in the Netherlands microlevel individual attributes and residential context. Transportation Research Record, 1753, 11–19. https://doi.org/10.3141/1753-02

Kuhnimhof, T., Chlond, B., & Von Der Ruhren, S. (2006). Users of transport modes and multimodal travel behavior: Steps toward understanding travelers' options and choices. Transportation Research Record, 1985, 40–48. https://doi.org/10.3141/1985-05

Kujala, P., Kallio, J., & Niemelä, M. (2019). Income Inequality, Poverty, and Fear of Crime in Europe. https://doi.org/10.1177/1069397118799048

Lauritzen, S. L. (1995). The EM algorithm for graphical association models with missing data. Computational Statistics & Data Analysis, 19, 191–201.

Lee, D., Derrible, S., & Pereira, F. C. (2018). Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling. Transportation Research Record, 2672(49), 101–112. https://doi.org/10.1177/0361198118796971

Lee, S., & Abbott, P. A. (2003). Bayesian networks for knowledge discovery in large datasets : basics for nurse researchers. 36, 389–399. https://doi.org/10.1016/j.jbi.2003.09.022

Li, M., Song, G., Cheng, Y., & Yu, L. (2015). Identification of Prior Factors Influencing the Mode Choice of Short Distance Travel. 2015.

Limtanakool, N., Dijst, M., & Schwanen, T. (2006). The influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium- and longer-distance trips. 14, 327–341. https://doi.org/10.1016/j.jtrangeo.2005.06.004

Liu, C., Ph, D., Erdogan, S., Ph, D., Ma, T., & Ducca, F. W. (2016). How to Increase Rail Ridership in Maryland : Direct Ridership Models for Policy Guidance. 142(4), 1–10. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000340

Loukopoulos, P., (2007). *A classification of travel demand management measures. In: G¨arling, T., Steg, L. (Eds.), Threats From Car Traffic to the Quality of Urban Life: Problems, Causes, and Solutions.* Elsevier, Amsterdam, pp. 275–292

Lu, X., Liu, T., & Huang, H. (2015). Pricing and mode choice based on nested logit model with trip-chain costs. 44, 76–88. https://doi.org/10.1016/j.tranpol.2015.06.014

Ma, T. (2015). Bayesian networks for multimodal mode choice behavior modelling: a case study for the cross border workers of Luxembourg. Transportation Research Procedia, 10(July), 870–880. https://doi.org/10.1016/j.trpro.2015.09.040

Nes, R. Van. (2002). Design of multimodal transport networks A hierarchical approach.

Neapolitan, R.E. (1990). *Probabilistic Reasoning in Expert Systems: Theory and Algorithms,* John Wiley & Sons, New York.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Francisco, CA: Morgan Kaufman.

Przybylowski, A., Stelmak, S., & Suchanek, M. (2021). Mobility behaviour in view of the impact of the COVID-19 pandemic-public transport users in gdansk case study. Sustainability (Switzerland), 13(1), 1–12. https://doi.org/10.3390/su13010364

Racca, D. P., & Ratledge, E. C. (2004). Project Report for " Factors That Affect and / or Can Alter Mode Choice " prepared for Delaware Transportation Institute and The State of Delaware Department of Transportation by College of Human Services , Education , and Public Policy University of Dela. March.

Redman, L., Friman, M., Gärling, T., & Hartig, T. (2013). Quality attributes of public transport that attract car users: A research review. Transport Policy, 25, 119–127. https://doi.org/10.1016/j.tranpol.2012.11.005

RideShark (n.d.). Multimodal trip matching. Received from: https://www.rideshark.com/category/transport-demand-management/page/8/

Rietveld, P. (2000a). Non-motorised modes in transport systems : a multimodal chain perspective for The Netherlands. 5, 31–36.

Rietveld, P. (2000b). The accessibility of railway stations : the role of the bicycle in The Netherlands. 5, 2–6.

Rietveld, P., & Steg, L. (2000). Behoeften , mogelijkheden en gedragskeuzen met betrekking tot het verplaatsingsgedrag : een multidisciplinair perspectief . Conceptueel model verplaatsingsgedrag.

Scheiner, J. (2010). Interrelations between travel mode choice and trip distance : trends in Germany 1976 – 2002. Journal of Transport Geography, 18(1), 75–84. https://doi.org/10.1016/j.jtrangeo.2009.01.001

Schlossberg, M. (2013). PREDICTING TRANSIT RIDERSHIP AT THE STOP LEVEL: THE ROLE OF SERVICE AND URBAN FORM.

Senbil, M., Kitamura, Æ. R., & Mohamad, Æ. J. (2009). Residential location, vehicle ownership and travel in Asia : a comparative analysis of Kei-Han-Shin and Kuala Lumpur metropolitan areas. 325–350. https://doi.org/10.1007/s11116-009-9195-y

Steg, L. (2003). Can Public Transport Compete With the Private Car? IATSS Research, 27(2), 27–35. https://doi.org/10.1016/s0386-1112(14)60141-2

Tonne, A. C., Beevers, S., Armstrong, B., Kelly, F., Wilkinson, P., Tonne, C., Beevers, S., Armstrong, B., Kelly, F., & Wilkinson, P. (2008). socioeconomic inequalities Linked references are available on JSTOR for this article : Air pollution and mortality benefits of the London Congestion Charge : spatial and socioeconomic inequalities. 65(9), 620–627. https://doi.org/10.1136/oem.2007.036533

Tyrinopoulos, Y., & Antoniou, C. (2013). Factors affecting modal choice in urban mobility. 27–39. https://doi.org/10.1007/s12544-012-0088-3

Ulloa, L., Lehoux-Lebacque, V., & Roulland, F. (2018). Trip planning within a multimodal urban mobility. IET Intelligent Transport Systems, 12(2), 87–92. https://doi.org/10.1049/iet-its.2016.0265

Van Exel, N. J. A., & Rietveld, P. (2009). Could you also have made this trip by another mode? An investigation of perceived travel possibilities of car and train travellers on the main travel corridors to the city of Amsterdam, The Netherlands. Transportation Research Part A: Policy and Practice, 43(4), 374–385. https://doi.org/10.1016/j.tra.2008.11.004

Verhoeven, M. (2010). Modelling life trajectories and mode choice using Bayesian belief networks (Vol. 1, Issue 2010). https://doi.org/10.6100/IR667904

Wang, F., & Ross, C. L. (2018). Machine Learning Travel Mode Choices: Comparing the Performance of an Extreme Gradient Boosting Model with a Multinomial Logit Model. https://doi.org/10.1177/0361198118773556

Wen, C., Wang, W., & Fu, C. (2012). Latent class nested logit model for analyzing high-speed rail access mode choice. Transportation Research Part E, 48(2), 545–554. https://doi.org/10.1016/j.tre.2011.09.002

WHO, W. H. O. (2020). Situation Report: World Health Organization. Coronavirus Disease 2019 (COVID-19), 2019(February), 11. https://apps.who.int/iris/bitstream/handle/10665/331475/nCoVsitrep11Mar2020-eng.pdf?sequence=1&isAllowed=y%0Ahttps://pandemic.internationalsos.com/2019-ncov/ncov-travel-restrictions-flight-operations-and-screening%0Ahttps://www.who.int/docs/default-source

Yan, X., Levine, J., & Marans, R. (2019). The effectiveness of parking policies to reduce parking demand pressure and car use. Transport Policy, 73(September 2018), 41–50. https://doi.org/10.1016/j.tranpol.2018.10.009

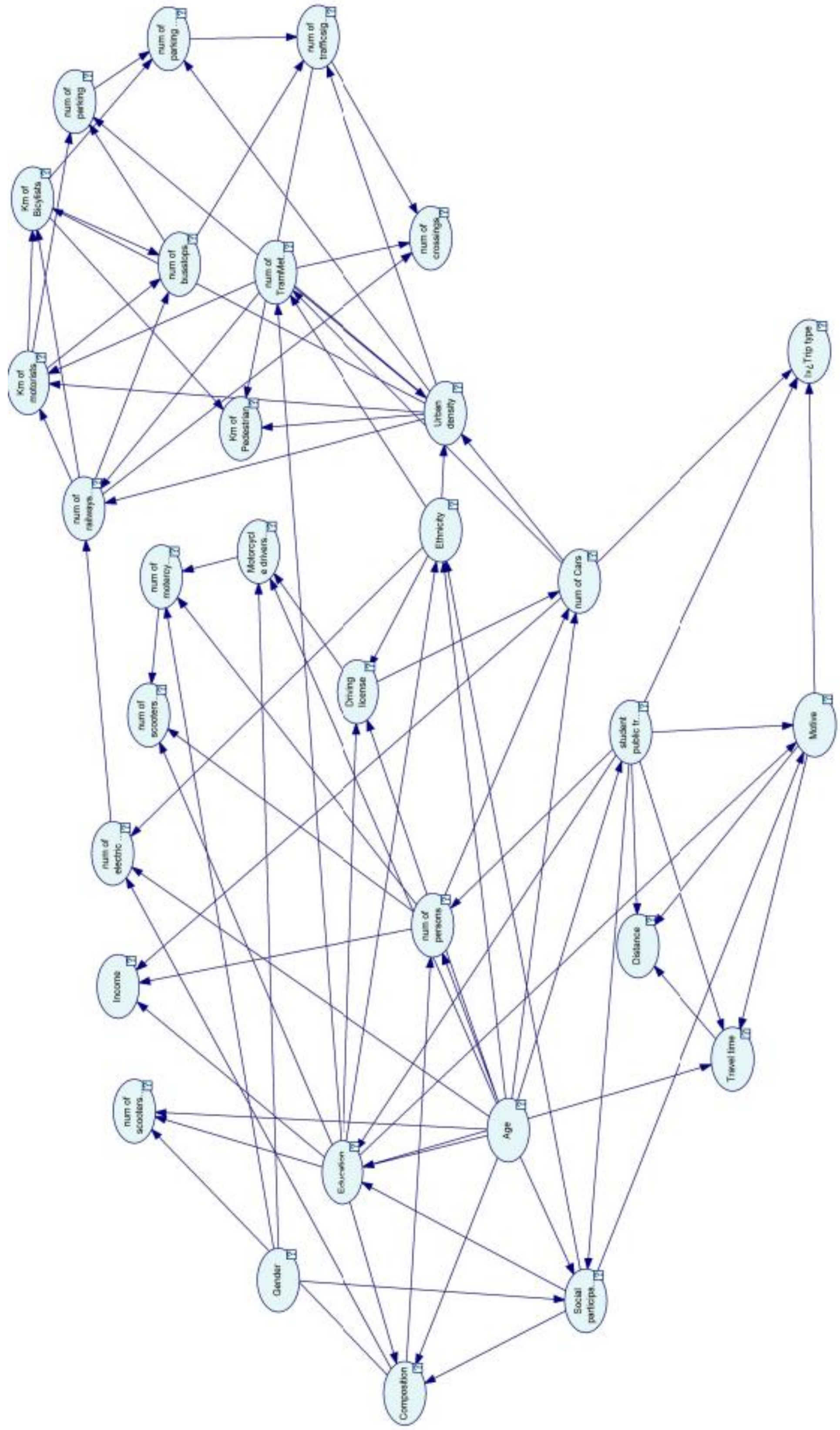
Yankaya, U. (2010). MODELING THE EFFECTS OF LAND USE CHARACTERISTICS ON MODE CHOICE FOR HOME – BASED WORK TRIPS : THE CASE OF ISTANBUL Uğur YANKAYA İ zmir Institute of Technology December , 2010 MODELING THE EFFECTS OF LAND USE CHARACTERISTICS ON MODE CHOICE FOR HOME – BASED.

Zhu, Z., Chen, X., & Xiong, C. (2018). A mixed Bayesian network for two-dimensional decision modeling of departure time and mode choice. Transportation, 45(5), 1499–1522. https://doi.org/10.1007/s11116-017-9770-6

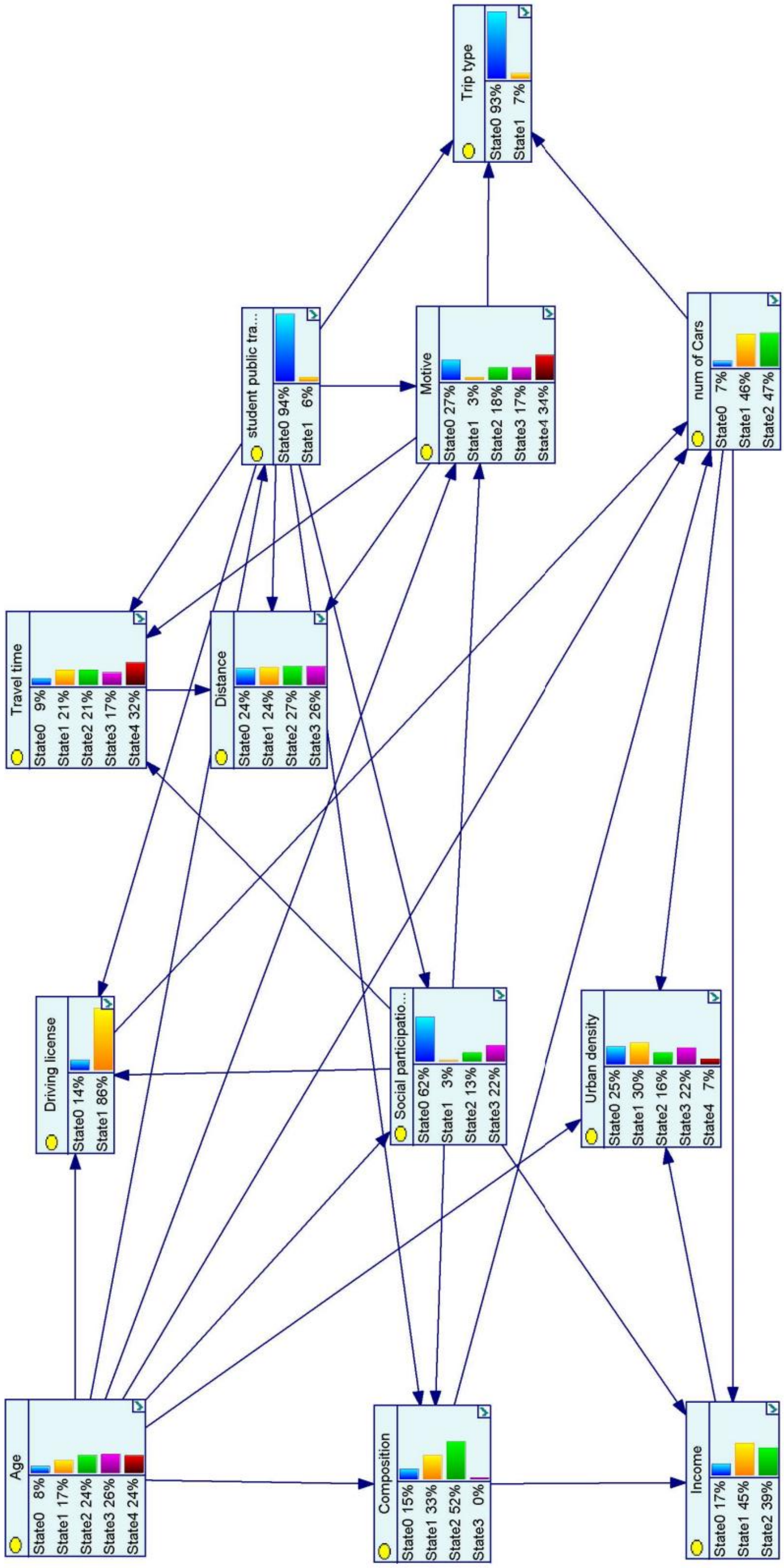
Zumkeller, D., Manz, W., Last, J., & Clond, B. (2005). Die intermodale Vernetzung von Personenverkehrsmitteln unter Berücksichtigung der Nutzerbedürfnisse (INVERMO). Schlussbericht, http://verkehrspanel.ifv.unikarlsruhe.de/download/INVERMO_Abschluss_Praesentationen/INVERMO_Schlussbericht_Executive_Summary.pdf

Appendix

A. Learned network structure 1- trip type



B. Final learned network 1 with CPTs - Trip type



C. Probabilities of trip characteristics – Network 1 – Trip type

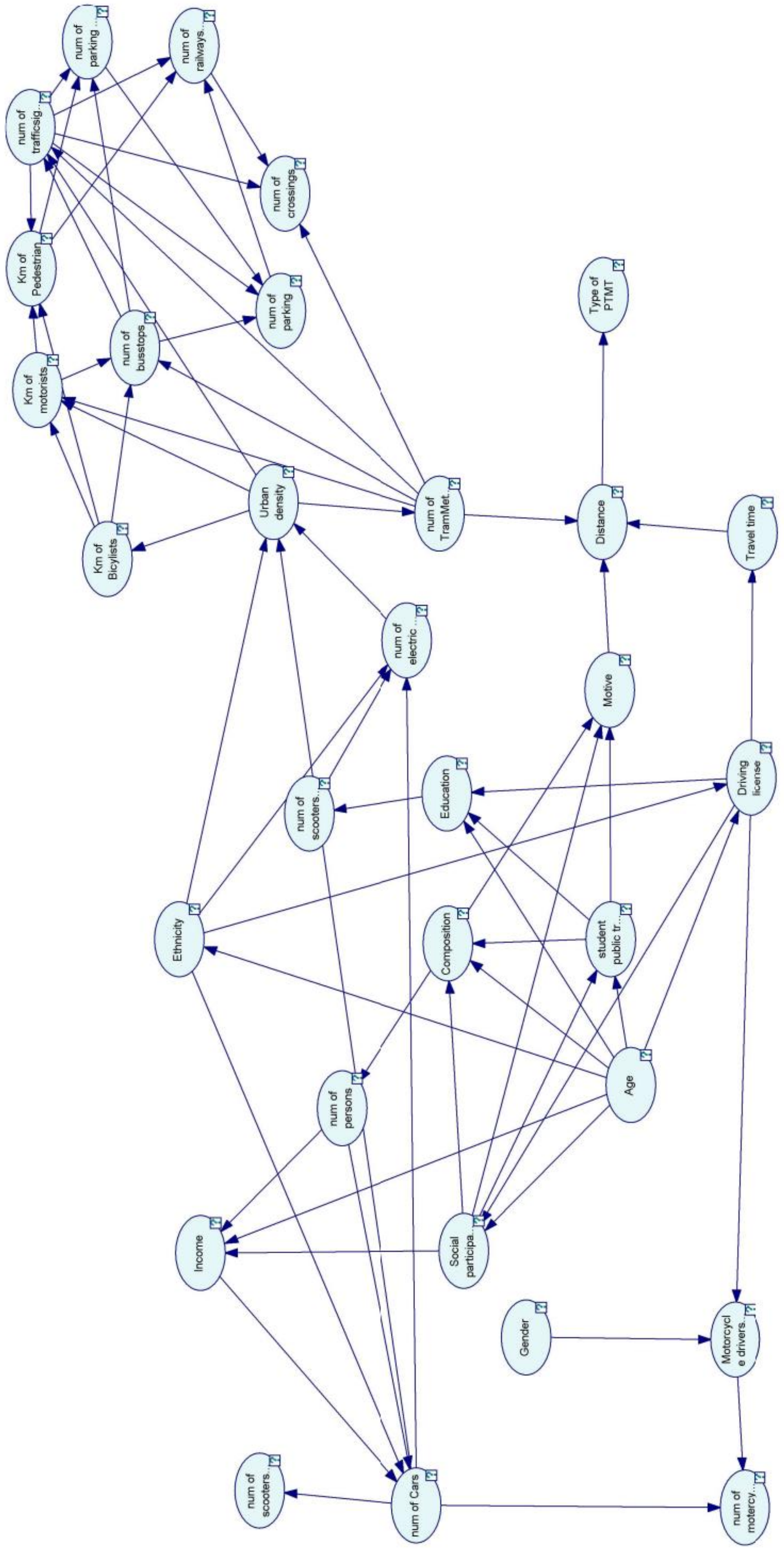
Table 20: Updated probabilities hard evidence travel time

Trip type	No evidence	Evidence > 30 min travel time	Evidence 31-60 min travel time	Evidence 61-90 min travel time	Evidence 91- 120 min travel time	Evidence > 120 min travel time
Unimodal car	0.935	0.944	0.941	0.939	0.933	0.927
Multimodal	0.065	0.056	0.059	0.061	0.067	0.073

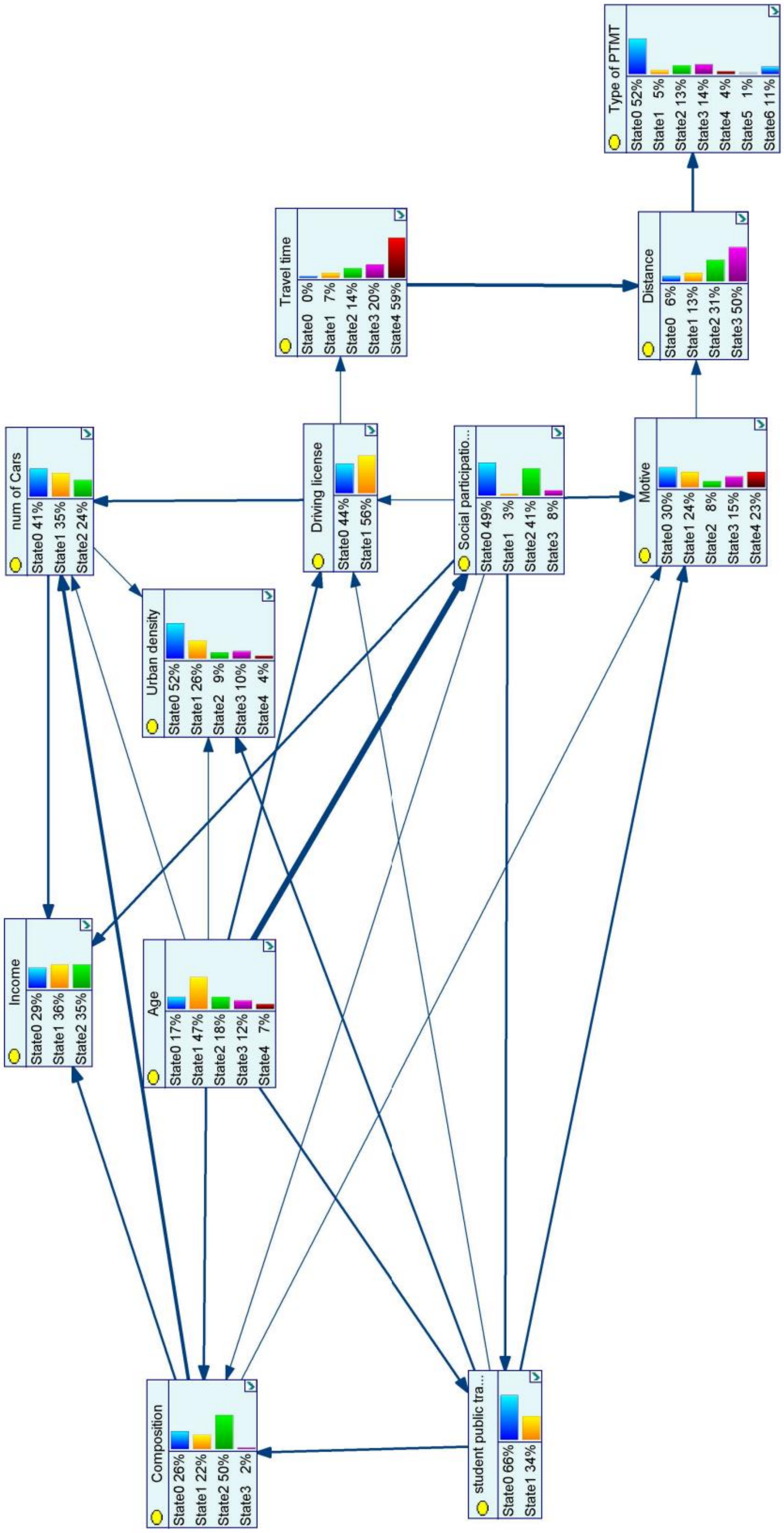
Table 21: Updated probabilities hard evidence distance

Trip type	No evidence	Evidence 30-50 km distance	Evidence 51-100 km distance	Evidence 101-250 km distance	Evidence > 250 km distance
Unimodal car	0.935	0.951	0.943	0.930	0.918
Multimodal	0.065	0.049	0.057	0.070	0.082

D. Learned network's structure 2 – Types of PTMTs



E. Final learned network 2 with CPTs - Types of PTMTs



F. Probabilities of personal characteristics – Network 2 – Types of PTMTs

Table 22: Updated probabilities hard evidence age

Types of PTMTs	No evidence	Evidence	Evidence	Evidence	Evidence	Evidence
		Age < 21	Age 21-35	Age 36-50	Age 51-65	Age > 65
Walking + Public transport + Walking	0.523	0.534	0.520	0.520	0.519	0.530
Cycling + Public transport + Cycling	0.053	0.051	0.054	0.054	0.054	0.052
Cycling + Public transport + Walking	0.128	0.126	0.129	0.129	0.129	0.126
Walking + Public transport + Cycling	0.142	0.140	0.143	0.143	0.143	0.140
Car + Public transport + Walking	0.036	0.035	0.037	0.037	0.037	0.036
Car + Public transport + Cycling	0.011	0.010	0.011	0.011	0.011	0.011
Other	0.106	0.103	0.106	0.106	0.107	0.104

Table 23: Updated probabilities hard evidence student public transportation smartcard

Types of PTMTs	No evidence	Evidence	Evidence
		No smartcard	Yes smartcard
Walking + Public transport + Walking	0.523	0.524	0.522
Cycling + Public transport + Cycling	0.053	0.053	0.053
Cycling + Public transport + Walking	0.128	0.128	0.129
Walking + Public transport + Cycling	0.142	0.142	0.143
Car + Public transport + Walking	0.036	0.037	0.036
Car + Public transport + Cycling	0.011	0.011	0.011
Other	0.106	0.106	0.106

Table 24: Updated probabilities hard evidence driver's license

Types of PTMTs	No evidence	Evidence	Evidence
		No driver's license	Yes driver's license
Walking + Public transport + Walking	0.523	0.535	0.514
Cycling + Public transport + Cycling	0.053	0.051	0.055
Cycling + Public transport + Walking	0.128	0.126	0.130
Walking + Public transport + Cycling	0.142	0.140	0.144
Car + Public transport + Walking	0.036	0.035	0.037
Car + Public transport + Cycling	0.011	0.011	0.011
Other	0.106	0.103	0.108

Table 25: Updated probabilities hard evidence social participation

Types of PTMTs	No evidence	Evidence	Evidence	Evidence	Evidence
		Employed	Unemployed	Student	Other
Walking + Public transport + Walking	0.523	0.518	0.535	0.526	0.539
Cycling + Public transport + Cycling	0.053	0.054	0.052	0.052	0.051
Cycling + Public transport + Walking	0.128	0.129	0.125	0.128	0.124
Walking + Public transport + Cycling	0.142	0.144	0.138	0.142	0.138
Car + Public transport + Walking	0.036	0.037	0.036	0.036	0.035
Car + Public transport + Cycling	0.011	0.011	0.011	0.011	0.011
Other	0.106	0.107	0.103	0.105	0.103

G. Probabilities of household characteristics – Network 2 – Types of PTMTs

Table 26: Updated probabilities hard evidence household composition

Types of PTMTs	No evidence	Evidence Single household	Evidence Couple	Evidence Couple/ single + child(ren)	Evidence Other
Walking + Public transport + Walking	0.523	0.522	0.522	0.524	0.527
Cycling + Public transport + Cycling	0.053	0.053	0.053	0.053	0.053
Cycling + Public transport + Walking	0.128	0.128	0.128	0.128	0.127
Walking + Public transport + Cycling	0.142	0.142	0.142	0.143	0.141
Car + Public transport + Walking	0.036	0.037	0.037	0.036	0.036
Car + Public transport + Cycling	0.011	0.011	0.011	0.011	0.011
Other	0.106	0.106	0.106	0.105	0.105

Table 27: Updated probabilities hard evidence household income

Types of PTMTs	No evidence	Evidence Below average income	Evidence Average income	Evidence Above average income
Walking + Public transport + Walking	0.523	0.524	0.523	0.522
Cycling + Public transport + Cycling	0.053	0.053	0.053	0.053
Cycling + Public transport + Walking	0.128	0.128	0.128	0.128
Walking + Public transport + Cycling	0.142	0.142	0.142	0.143
Car + Public transport + Walking	0.036	0.037	0.036	0.036
Car + Public transport + Cycling	0.011	0.011	0.011	0.011
Other	0.106	0.106	0.106	0.106

Table 28: Updated probabilities hard evidence # of cars within a household

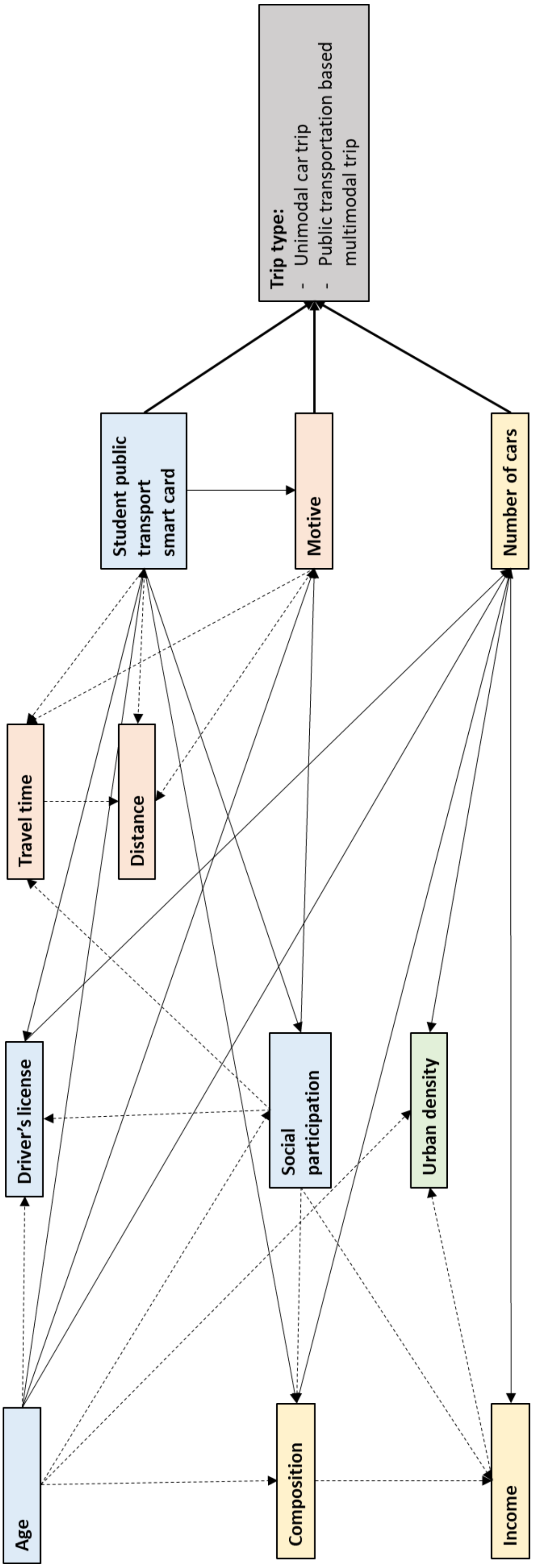
Types of PTMTs	No evidence	Evidence No cars in Household	Evidence 1 car in Household	Evidence 2 or > cars in Household
Walking + Public transport + Walking	0.523	0.524	0.523	0.523
Cycling + Public transport + Cycling	0.053	0.053	0.053	0.053
Cycling + Public transport + Walking	0.128	0.128	0.128	0.129
Walking + Public transport + Cycling	0.142	0.142	0.143	0.143
Car + Public transport + Walking	0.036	0.037	0.036	0.036
Car + Public transport + Cycling	0.011	0.011	0.011	0.011
Other	0.106	0.106	0.106	0.106

H. Probabilities of environmental characteristics

Table 29: Updated probabilities hard evidence urban density

Types of PTMTs	**No evidence**	**Evidence**	**Evidence**	**Evidence**	**Evidence**	**Evidence**
		Very highly urban	**Highly urban**	**Moderately urban**	**Little urban**	**Not urban**
Walking + Public transport + Walking	0.523	0.523	0.523	0.524	0.524	0.525
Cycling + Public transport + Cycling	0.053	0.053	0.053	0.053	0.053	0.053
Cycling + Public transport + Walking	0.128	0.128	0.128	0.128	0.128	0.128
Walking + Public transport + Cycling	0.142	0.142	0.143	0.142	0.142	0.142
Car + Public transport + Walking	0.036	0.037	0.037	0.036	0.036	0.036
Car + Public transport + Cycling	0.011	0.011	0.011	0.011	0.011	0.011
Other	0.106	0.106	0.106	0.105	0.105	0.105

I. Revised conceptual model - Trip type



J. Revised conceptual model – Types of PTMTs