Masters in Informatics Engineering
Dissertation
Final Report

Demand Modeling for Responsive Transport systems using digital footprints

Paulo Alexandre Costa da Silva
pacsilva@student.dei.uc.pt

Advisor:
Prof. Carlos Lisboa Bento
Co-Advisor:
Prof. Rui Gomes

Date: July 1, 2014

FCTUC DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Abstract

Regular public road transportation usually uses fixed routes and schedules, which may impair the mobility of potential users in rural areas and certain periods of the day in urban areas due to low and unpredictable demand.

Demand Responsive Transport (DRT) tries to address these problems and improve community welfare, by using flexible routes and frequencies that vary with the actual observed demand, possibly adopting route solutions in real-time. These systems try to fill in the gap between traditional transportation service and the individual transportation. DRT services must try to be sustainable, make an efficient use of vehicles and respond to users mobility needs.

Given the expected low number of requests, the application of traditional demand modeling methods is not adequate: usually, demand modeling for DRT requires a higher resolution zoning when compared to traditional transportation systems. We will need to analyze user's short term land use patterns, with a careful analysis of available data, coming from collaborative platforms like Twitter and Foursquare.

Keywords

[DRT, Demand Modeling, Discrete Choice, Mobility, Social Networks]

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Over the last decade the number of Demand Responsive Transport (DRT) services has grown since their introduction in the 60's on the United Sates, where they are known as "paratransit" (Muscat, 2012). The traditional form of DRT (e.g. services for the disabled, elderly, and where demand is too low for other public transports (Bakker, 1999) (Black, 1995)) should be distinguished from the open for-all initially investigated in (T. Sihvola, 2010). DRT's attractiveness derives from their flexibility, ability to meet demand in low or unpredicted demand areas where traditional transportation prices are high (K. Zografos, 2008), and the potential to reduce gas emissions (Julie Prud'homme, 2011) and traffic congestion.

DRT's services can operate "on demand", picking up people and dropping them of according to their needs. When referring to the non-fixed route of DRT's, it doesn't mean that there aren't any stop points. These stop maybe terminals, fixed points along the route (high demand points), and non-predefined stop points (usually the doorstep of a user). In (G. Ambrosino, 2003), the authors conclude that in order to have a well-organized DRT service, specific procedures and organizational rules must be followed to meet objectives efficiently.

There are multiple models for DRT's that focus in the optimization of routes, order and pick-up strategies but seldom consider future demand (S. Ichoua, 2006). In the field of dynamic vehicle routing and dispatching, its important to exploit information about future events in order to provide services that meet demand efficiently (providing a quality service to users) and make DRT's more viable financially (Gomes, 2013). Initial models for travel demand, focused on predicting demand for future years, to estimate the required amount of transportation supply. That prediction was carried out for long-term multiple socio-economic scenarios, alternative transportation systems and land-use configurations (Bhat, 1999). Demand modeling for flexible transportation has been using these models, which are better suited for fixed transportation, as the analysis is done in the long-term when it comes to demand. Anticipating demand by studying users short-term land use, can improve demand modeling and, ultimately overall efficiency and flexibility of the service. We take advantage of social studies and their proven effect on travel, to do so (Chapin, 1974).

The theory of utility maximization, usually through discrete choice models, is often used to study the individual decision-making. We will use such a model, namely the Multinomial Logit Model (MNL), with a social component for utility and characteristics, both derived from Social Network Analyses, where a network is constructed, linking the nodes (decision makers) that have social influence over one another, and the strength of that influence. The individual characteristics are also drawn from the profile of the entity and will be used in the MNL model.

The data for the model is to be collected from Delft, since we profit from the relationship between our University of Coimbra and the Delft University of Technology, which provided us with 80 individuals GPS traces over the course of four days.

We try to address the problem of efficiency and cost when transportation services operate in lower demand areas or times, by estimating and pinpointing the demand. That is done trough the analyses of social networks, as opposed to previous analyses consisting of surveys and phone call data. The data collect trough social networks will then be used in choice modeling, whose results can be used for transportation planning.

In State of the Art, we take a look at several methods and models used over the years to model travel behavior and estimate demand as well an overview of Social Network Analysis. The methodology for this work is presented in chapter 3, and in 4 some conclusions and future work plans are laid out.

Chapter 2
State of the Art

In this chapter we introduce some research done in the field of Demand Responsive Transportation, concepts to take into account when designing a DRT service, the importance of mobility patterns and information communication technologies as well as the models used for social network analyses and discrete choice modeling.

2.1. Service planning

A taxi service is a direct individualized or collective organized answer to the need for transport. But taxis have low capacity and a rather high fare, which seems to be the main reason to look for new types of transportation. Also public transports are facing the need to rationalize the organization. A high number of different lines serves most origins and destinations in the region, however the organizational costs are less and less covered by fare incomes and origins and destinations have also increased enormously, due to people increasing need for mobility. The typical answer to this decrease in income is to limit the frequency of services, resulting again in a lower number of customers.

In this context DRT Services may be designed according to a combination of concepts, to reduce the operational costs and to give customers a transport offer with a higher flexibility to meet their needs (G. Ambrosino, 2003):

- Route and time concepts
- The booking concepts
- The general intermodal integration concepts
- The vehicle allocation concepts

2.1.1 Route and Time

When it comes to route and time concepts, the design of the DRT service must consider the level of flexibility envisaged. In conventional public transportation schedules are defined well ahead, or in the taxis case not at all. For intermediate transport a wide range of different concepts are possible (G. Ambrosino, 2003). To get a better view off possible ways of organizing services, some concepts are presented with the following type of stops.

- Stops with fixed time, always served.
- Stops with a pre-defined passing time that is only served if requested.
- Stops that are only served once requested.
- Stops indicated by users, anywhere within the area of service.

The combination of this types of stops, leads to different service scheme. In the first scenario (Figure 1) mandatory stops are used with non-mandatory stops, in a sequential manner. This could lead to an idle period on the last stop, if none of the non-mandatory stops are requested.

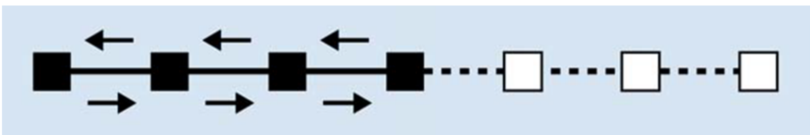


Figure 1 Scenario 1, (G. Ambrosino, 2003)

Figure 2 demonstrates a schedule with fixed stops and predefined passing times. Intertwined in those stops, are stops by request. The deviation will normally take more time than the direct route. For this reason a balance needs to be achieved between the deviations and the feasible time margins on the fixed timetable for the fixed stop of the basic route.

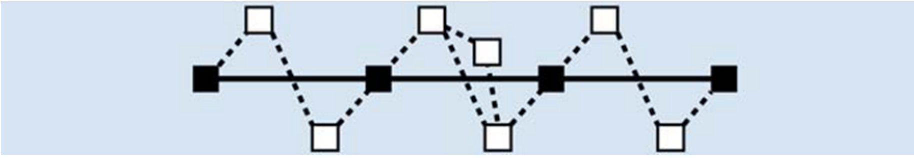


Figure 2 Scenario 2, (G. Ambrosino, 2003)

A possible third scenario (Figure 3), has one or more mandatories stops at the ends of the corridor, but allow various points with a non-fixed time to be requested between them. Statistical data must be used, so that the number of requests between the ends don't hinder the arrival time on mandatory stops.

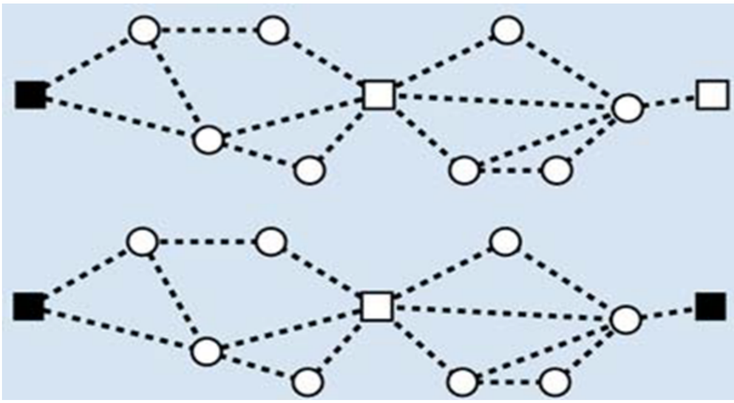


Figure 3 Scenario 3, (G. Ambrosino, 2003)

In Figure 4, the served points in the area can be any point (e.g. address of a house, a particular building etc.) instead of fixed stops. If specific user groups are to be served extra attention is required for the time needed at the stop to allow customers to enter or to leave the vehicle (e.g. with a wheelchair).

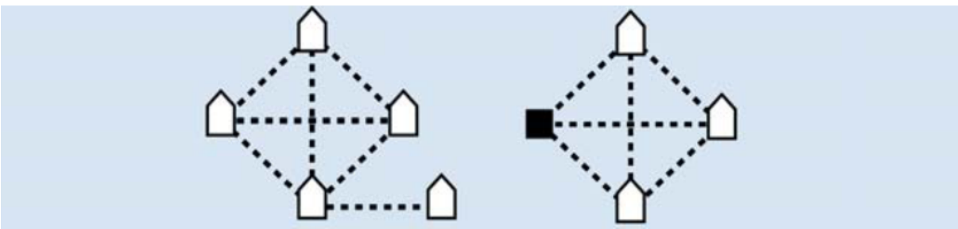


Figure 4 Scenario 4, (G. Ambrosino, 2003)

2.1.2 Booking

A crucial element for DRT Services is the booking of the trip the customer (G. Ambrosino, 2003). In booking three phases are distinguished. 1) Request for trip, 2) A proposal from the service operator, and 3) Booking confirmation by the costumer. In the first phase costumers supply both addresses for departure and destination, as well the number of persons and the departure or arrival time. Once this gets to the service operator a preliminary trip with a wide time window is presented. The costumer then confirms this, and after, receives the booking confirmation with more details.

2.1.3 Intermodale integration

A DRT service can have different roles in the general public transport offer (G. Ambrosino, 2003):

- Stand-alone: Operating mostly in rural environments, it offers service in low-density areas, without any time or spatial relation with other services.
- Feeder: Feeds service to another transport service, which completes the rest of the trip. The service area is limited with the center being the connection to the other service.
- Multiple roles: It's a bit of a combination of the two extremes services presented before. They feed the whole region, while maintaining connections to other important community services and facilities to other destinations.

2.1.4 Vehicle allocation

Another important concept, is the way vehicles are allocated to each service:

- Fixed vehicle allocation: Service defined with only one vehicle available. The vehicle characteristics, determine to a large extent the type of DRT service (e.g. large doors and ramp, will tell that the service provided is for the elderly and disabled). If service is full or time windows don't allow deviations, passengers will have to choose an earlier or later service.
- Extendable vehicle allocation: If the operator doesn't want to loose passengers, the use of extra vehicles can be foreseen with certain limits (e.g. cooperation with a taxi company).
- Dynamic allocation: There is a pool of vehicles at the disposal of the DRT service. These vehicles range in size and types, offering flexibility to the service. Vehicles are allocated to the services taking into account optimization aspects and the specific requirements of the requests (e.g. accessibility for disabled).

2.2 Mobility Patterns

Identifying Urban Mobility Patterns has been a topic of continuous research in transportation planning and behavior modeling. The rapid development of information and communication technologies (ICTs) has open new possibilities for such research. A good example in transportation planning was the use of mobile phone data release by Orange for a development competition to redraw bus routes in the Ivory Coast's (Wakefield, 2013). Location-based check-in services are also a way to understand mobility as in (Samiul Hasan, 2013) where activity patterns are aggregated by categories determining the purpose. The individual activity patterns are characterized by "finding the timing distribution of visiting different places depending on activity category" and the frequency of visiting.

We will focus more on what's being done in transportation planning, namely in Demand Responsive Transport's. The next points refer to the models being used to estimate travel demand.

2.3 Aggregate and Disaggregate Models

Aggregate models where one of the first to appear to structure travel demand They are simple mathematic models, such as gravity models or an entropy model that quantified travel as a function by zones. The number of trips generated from a zone was considered to be proportional to the population in the zone, while the number of trips attracted to a zone was considered to be proportional to the number of sources of attraction in the zone.

As models began to evolve with time, aggregated models were gradually replaced by disaggregated models. The fundamental difference is that disaggregated models take into account the effects of individual socio-demographics on travel related choices. However in practice disaggregate trip-based models are sometimes implemented in an aggregate manner with aggregate zonal social-demographic data. These models don't take into account the linkages between trips. A trip from home to work and work to home are both classified as home-based work trips. This limitation opened the way for Tour-based models (Sivakumar, 2007). Human behavior is extremely complex, so much that its possible impossible to understand completely what lies behind a single individual's decision. Although it's possible to draw inferences from the patterns of choice that groups of people make. Introductions to the field of discrete choice models can be found in (Dagsvik, 2000) and (Moshe E. Ben-Akiva, 1985). The latter of those focuses in particular on how the theory is applied to travel demand. The framework for a discrete choice model can be defined by a set of general assumptions:

- Decision-maker: defining the decision-making entity and its characteristics.
- Alternatives: determining the options available to the decision-maker.
- Attributes: Measuring the benefits and costs of an alternative to the decision-maker
- Decision rule: Describes the process used by the decision-maker to chose an alternative.

2.4 Approaches for travel demand modeling

2.4.1 Trip-based

The traditional four-step demand modeling is the most used demand analysis methodology. The widespread use of the four-step model (FSM) does not imply its superior efficacy, but that it is simply the most economical option, with respect both to data requirements and simplicity of operation. In this model the influence of activity characteristics decreases, and that of trip characteristics increases. A basic FSM is defined by four sequential stages: trip generation, trip distribution, modal split, route assignment as shown in Figure 5.

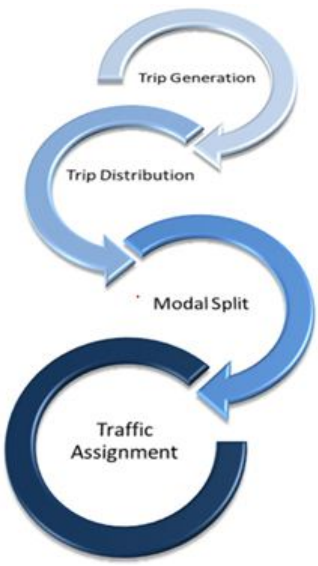


Figure 5 4-Step Model

The demand modeling process is aggregate and trip-based with limited analysis of travel behavior "Four step models are not "behavioral in nature" rather they rely on statistical correlations between demographics and traffic patterns" (VTM, 2009).

Trip generation is defined by trip production, which contains household characteristics and socio-economic factors, and trip attractions, describing land-use, household and employment by category (e.g. Industrial, commercial, Services), with multiple regression models being used to predict attractions. Trip distribution is estimated with gravity and multinomial logit models. In mode choice, the proportion between each origin and destination that uses a particular transportation mode, generally employ multinomial logit mode choice analysis (e.g. Figure 6 represents the Logit and Figure 7 the Nested Logit methods in the modal split phase). The difference being that the Nested Logit allows us to capture correlations between alternatives, hence the subdivision of the alternatives. Both models are presented in section 2.5.

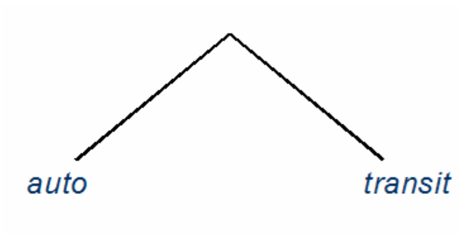


Figure 6 Logit

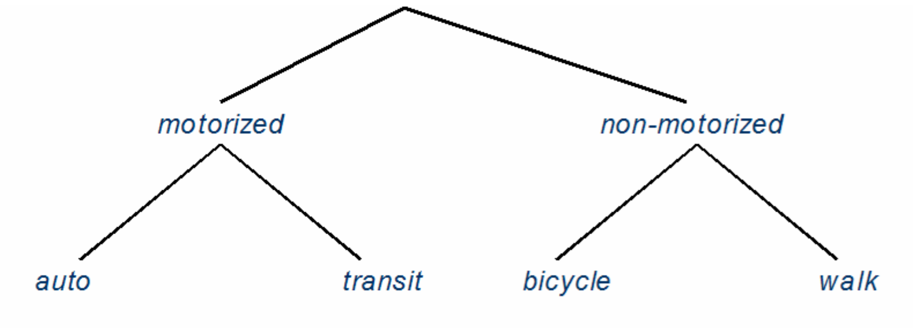


Figure 7 Nested Logit

Route assignment is typically based on deterministic (e.g. Shortest Path, Minimum Generalized Cost) or stochastic (Discrete choice) algorithms. Time-of-day choices and peak spreading are not considered in basic four step models. Instead, the 24-hour demand matrix is converted to several time-of-day matrices (e.g. AM peak, PM peak, mid-day, and other off-peak) based on observed demand shares in different time periods.

2.4.2 Tour-based

The Tour-based Four-Step modeling approach (which is an advance to the traditional Four-step), is made of multiple trips starting and ending at important points, such as home or work, thus resolving the problem of return trips being treated independently. The data requirement is the same as the Trip-based models. Tour-based models still neglect linkage between trips. For instance, if a person goes from work to home and stops midway in a grocery store, tour-based models would view this as a home-based trip and a non-home based trip. The Tour-based model includes time in its mode-destination-time period choice models in more advanced model systems (combined tour distribution and mode choice step).

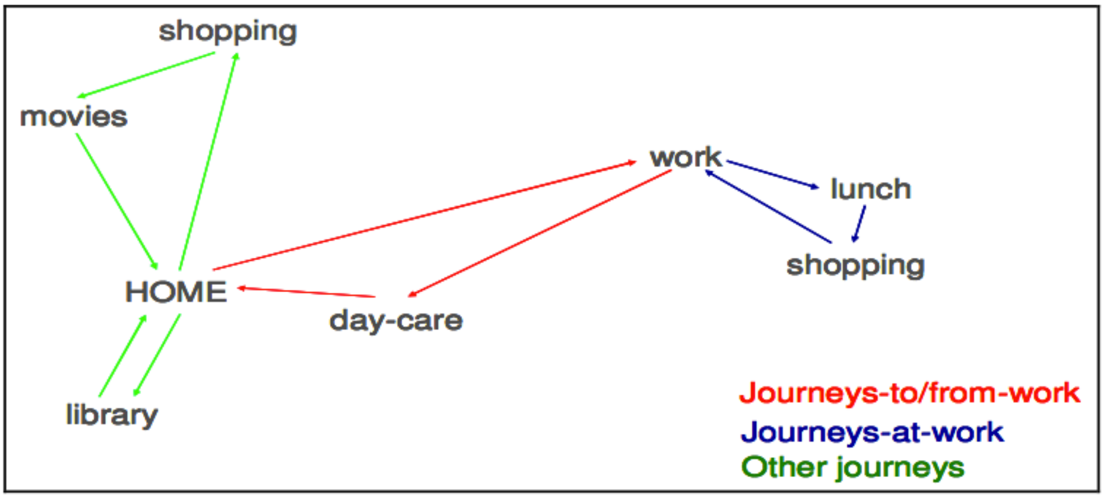


Figure 8 "Tours"

2.4.3 Activity

One of the emerging models over the last decade is the Activity-Based model first introduced in (Pendyala, 1989). This model differs from the FSM by viewing travel as a demand derived from the need to pursue activities and focuses on “activity participation behavior” (716, 2012). Travel patterns are organized within activity-based models as sets of related trips known as “Tours” (e.g. in Figure we can see that tours are now connected by activities). The tours are interdependent, with the activity pattern replacing trip and tour generation steps of trip and tour-based models.

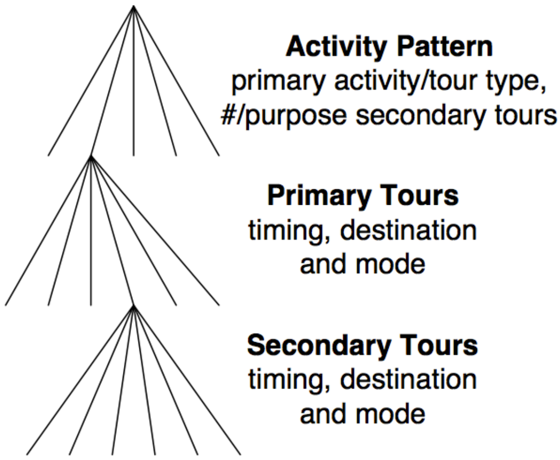


Figure 9 Model Structure

This behavior basis include various factors to determine the why, how, when, and where of performed activities and resulting travel such as socioeconomic factors. In contrast to the FSM which uses discrete trips as their standard travel unit the tour based approach compose the sets of activity-based travel analyses. Activity based models will situate tours based on possibilities derived from socioeconomics, land use and network characteristics.

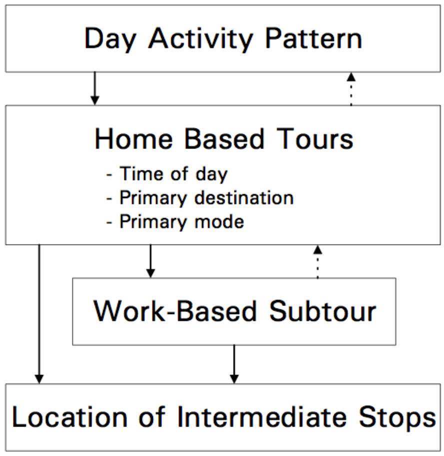


Figure 10 Activity-Based Model, Portland

2.4.4 Overview

In Figure 11 its possible to observe the connection between activities to tours to complete the schedule, to the semi-linkage in tours and to the non-existence of linkage between trips.

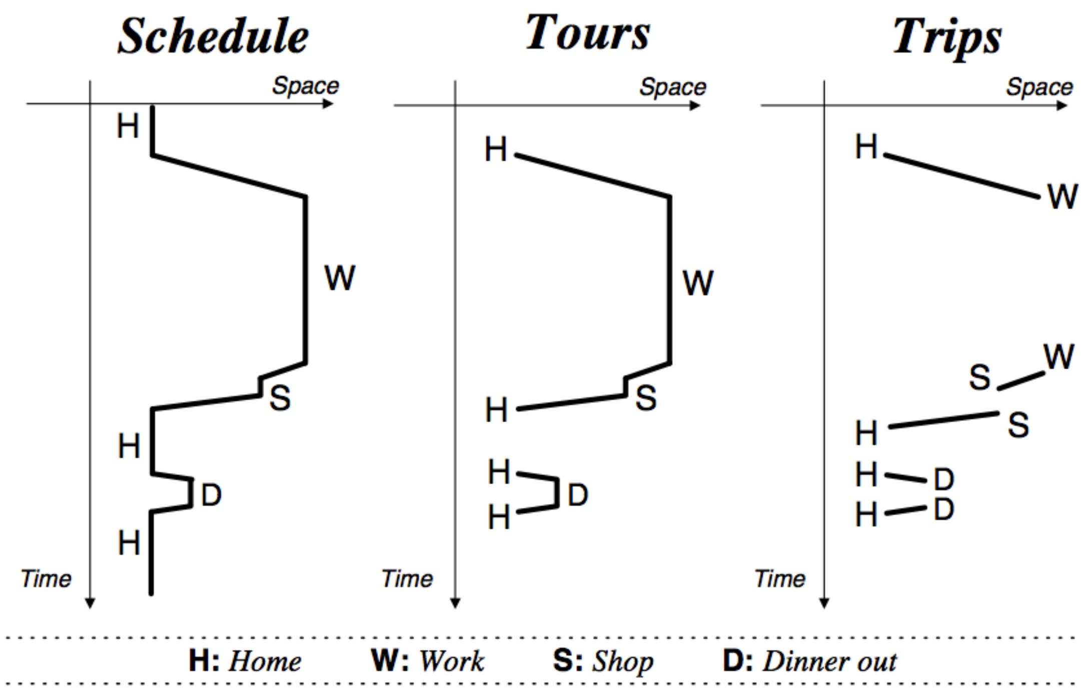


Figure 11 Travel behavior, Ben-Akiva

Table 1 highlights some aspects about the linkage approach, both positive and negative. More specifically the effect of the non-existence, partial, full linkage between trips.

Table 1 Models highlights

Approach	Highlights
Trip-Based	• Person trips as the unit of analyses • Temporal aggregation • Behavior modeled in earlier steps unaffected by choices modeled in later steps • Neglects linkage between trips
Tour-Based	• Explicitly chains trips in tours • Doesn't integrate well the time dimension • One of the benefits of estimating tours rather than trips is that coordinated decisions within a household may be modeled comprehensively based on a wider set of influential factors • Neglects linkage between trips • Same data requirements • Not behaviorally realistic
Activity-Based	• Travel demand is derived from demand for activities • People face time and space constraints that limit their activity schedule choice • Tours are independent • Occurs dynamically with influence from past and anticipated future events • The larger choice set is a weakness

2.5 Modeling methods

In this section we present the methods for the models previously presented. First the aggregated gravity method and then the discrete choice methods also referred as disaggregated, meaning that the decision-maker is assumed to be an individual.

2.5.1 Gravity

Proposed in (Tinbergen, 1962) to explain international bilateral trade, was named after Newton universal law of gravity by its similarities. In gravity model, we start from assumptions about trip making behavior and the way it is influenced by external factors (e.g. Trip-production and trip-attractions in the Four-step model). An important aspect of the use of gravity models is their calibration, that is the task of fixing their parameters so that the base year travel pattern is well represented by the model (NPTEL, 2006).

2.5.2 Random Utility theory

Most frequently discrete choice models implement a fixed coefficient utility function that is linear in the parameters. But human behavior isn't linear, as we well know, so to better describe it, random coefficients were used in utility functions giving birth to the Random Utility theory.

Random Utility theory is based on the hypothesis that every individual is a rational decision-maker, maximizing utility relative to his or her choices (Cascetta, 2009). Considering a discrete set $S = \{S_1, \ldots, S_n\}$, this is called the choice set. The utility that individual n associates with alternative i in the choice set S_n is given by:

$$U_{in} = V_{in} + \Sigma_{in}$$

Where V_{in} is the deterministic part of the utility, and Σ_{in} is the random term, capturing uncertainty. The alternative with the largest total utility is chosen (Jonas Anderson, 2010). Based on the theory of utility maximization and conditional on assumptions placed on the random error component, several formulations such as the multinomial probit or logit models may be derived.

2.5.2.1 Multinomial Logit Model

The Logit family as spawned from the initial Binary Logit Model, becoming widely used for travel demand analysis, because of its tractability, even if imposes restrictions on the covariance structure. The Multinomial Logit Model is derived from the assumption that the error terms of the utility functions are independent and identically Gumbel distributed (Ben-Akiva, 1999). Assuming that individual utility deviations from mean utility in a homogenous segment are statistically independent for different alternatives and have a probability distribution, the MNL model can be derived.

The probability that a given individual *n* chooses alternative *I* within the choice set *C* is given by:

$$P(i|C_n) = \frac{e^{uV_{in}}}{\sum_{j\, C\, c_n} e^{uV_{jn}}}$$

where the probability of *I* is divided by the sum of the utility of all the others alternatives.

2.5.2.2 Nested Logit Model

The Nested Logit Model, first proposed in (Ben-Akiva, 1973) and (Ben-Akiva, 1974). is an extension of the Multinomial Logit Model designed to capture some correlation between alternatives. It is based on the partitioning of the choice set C_n into *M* nests C_{mn} such that

$$C_n = \bigcup_{m=1}^{m} C_{mn}$$

The utility function is composed of a term specific to the alternative and a term associated with the nest. If *I* is an alternative from C_{mn} we have

$$U_{in} = V_{in} + \mathcal{E}_{in} + V_{c_{cn}} + \mathcal{E}_{c_{cn}}$$

The error terms $\mathcal{E}_{in}$ and $\mathcal{E}_{c_{cn}}$ are supposed to be independent. As in the Multinomial Logit Model, the error terms $\mathcal{E}_{in}$ are assumed to be independent and identically Gumbel distributed, with scale parameter that can be different for each nest. The Nested Logit Model is designed to capture choice problems where alternatives within each nest are correlated. No correlation across nests can be captured by the Nested Logit Model. When alternatives cannot be partitioned into well separated nests to reflect their correlation, the Nested Logit Model is not appropriate.

2.5.2.3 Multinomial Probit Model

One problem with the multinomial logit methods was the Independence of Irrelevant Alternatives (IIA) assumption imposed by them, as can be illustrated by the red-bus-blue-bus example: If a commuter chooses from a car and a red bus with equably probabilities, 0.5, and then add a third mode, the blue bus. Assuming that the commuters wouldn't care for the color of the bus, we should have $\text{prob(car)} = 0.5$, $\text{prob(redBus)} = 0.25$ and prob(blueBus) =0.25. When in reality we end up with $\text{prob(x)} = 0.33$ for all tree transportation modes. The IIA leads to a failure by not taking into account that the blue and red buses are very similar, and are "perfect substitutes" (Wooldridge, 2002). One alternative to break down the IIA assumption therefor consists in allowing correlation between alternative. The Probit model assumes that the error terms of the utility are normally distributed. The model captures explicitly the correlation among all alternatives.

2.5.3 Overview

Table 2 presents the principal highlights of the methods, presenting strengths and weaknesses.

Table 2 Methodologies highlights

Type	Method	Highlights
Aggregate	Gravity and Entropy Maximization models	Initial models used to estimate travel demand. Have a high explanatory power and the data is easily available. This models don't consider the individual, instead the hole population
Disaggregate	Multinomial Logit	Assumes that individual utility deviations from mean utility in a homogenous segment are statistically independent for different alternatives and have a probability distribution. Limitation explained with red/blue bus paradox
	Nested Logit Multinomial Probit	To overcome the paradox mention above, the Nested Logit and the Multinomial Probit methods, allow for correlation for different alternatives. Due to the complexion of the Probit model the Logit has been more popular, because of its tractability, but imposes restrictions on the covariance structure.

2.6 Social Network Analysis

With the evolution of technologies, and the number of devices/applications that use location-based services (LBS), a new opportunity arises to study travel behavior. Before LBS the studies focused on phone records to understand the mobility of the user, extrapolating the location from the cell towers. Location-based social networks are growing in popularity, so it makes sense to conduct a Social Network Analysis (SNA) to explain the motivation for, and characteristics of, travel behavior. Although we can infer friendship in both cell phones (identifying the caller or receiver) and social networks (friends), the later is more sporadic, but easier to access. So we will focus on SNA. If the members of a household influence travel, then friends can also be of influence. The SNA has two major components, actors (persons, groups or organizations) who interact within each other, and relationships. The main objective of SNA is to analyze the link between people and their dynamics. Seems evident that there is a relationship between Information and Communications Technologies (ICT) and travel (de Graaff, 2007) (Kenyon, 2006) (Kwan, 2007) (Lyons, 2009).

2.6.1 Periodic & Social Mobility Model

In (Eunjoon Cho, 2011), the authors aim to understand human motion and dynamics by answering tree questions: Where do we move? How often do we move? And how do social ties interact with movement? To answer these questions, data from Gowalla, BrightKite and cell phone records were used. The model was build on the assumption that the majority of movement occurs between latent states, such as home and work. The user movement is anchored by these locations and some in between as they commute in between them.

- First a spatial model is made for every user. This model represents the latent states of each user, modeled with Gaussian distribution. Inferring the user's check-ins to identify which state (home, work) they came from.
- Secondly, a temporal model is made to describe movement between the locations. The probability distribution is the mixture of "home" and "work" of a user at a given time, governed by the temporal model.
- Finally, the model is extended to include social network-driven mobility. The probability of user X checking in at a place P, depends on how long a friend Y check-in on P and the distance between the two friends.

With the training of the Periodic & Social Mobility Model (PSMM) completed, both spatio-temporal and social components are used to calculate the check-in probability in a certain place at a certain time.

2.6.2 Radiation Model

The study by (Alexey Tarasov, 2013) aims to improve the PSMM by removing the Gaussians in the spatial model, using the radiation model instead and to predict user location based on information about previous check-ins and social ties. To study the intensity of the flow T_{ij} between the locations i and j for the distance R_{ij} with populations m and n respectably, the radiation model is described as follows.

$$T_{ij} = T_i * \frac{m * n}{(m + S_{ij}) * (m + n + S_{ij})}$$

Where T_i corresponds to the population leaving that location, and S_{ij} total population placed within the circle of radius R_{ij}, not including i and j. Using this model to express the probability of a individual moving from i to j, we have:

$$P_{ij} = \frac{m * n}{(m + S_{ij}) * (m + n + S_{ij})}$$

Were m and n now represent venues capacities rather than population. It works on the assumption that higher capacity venues will be more attractive.

2.6.3 Social interactions on discrete choice

In discrete choice models the effect of social dimensions was first formalized for both the binomial (Brock W. A. and Durlauf, 2001) and the multinomial (Durlauf B. W., 2002) cases. In the general method, an agents utility is formed by both private and social component. The private component corresponds to the decision-maker's characteristics and his neighborhoods characteristics. The social component represents the strength of social utility and the percentage of others in the neighborhood selecting the same alternative in the choice set (Tim J. Ryley and Alberto M Zanni and Tim, 2011).

$$U_j^n = h_j^n + Jp_j + \varepsilon_j^n$$

Where h_j^n represents the deterministic private part of utility (decision maker's and neighborhood characteristics), J the social strength in utility, p_j the probability of j selecting the same alternative, denotes the decision maker's expectation and ε_j^n the random private utility (Tim, 2013).

The study of social influence in the decision-making has been investigated by (Carrasco, 2008) and (Axhausen, 2008). (Axhausen, 2008) connects travel with social networks, arguing that daily life revolves around family, colleagues, friends and shopping. The study in (Paez, 2007), an additional element refers to the conformation to social norms, implying that decision-makers are more likely to choose a particular alternative if more peers have already chosen the same alternative.

Chapter 3
Methodology

The data used in this work comes from Twitter, Instagram, Foursquare and GPS traces collected by eighty individuals for a period of four days. We started collecting tweets in the beginning of January using the Twitter Stream API with a radius of fifteen kilometers from the center of Delft, Holland, with the following information: Tweet_ID (enables user identification), timestamp and coordinates. For Instagram, the data collection started latter and focused on the media shared by its users.

Using Twitter's and Instagram REST API we will create a list of friends for each user, considering as friends the users that mutually follow each other. Having this information available will enable us to start out SNA.

Using Foursquare API its possible to access the Venues database, so that we can extract the check-ins count, category and location. With location and category, its possible to cross-reference with the coordinates from tweets, so that we know witch locations are frequently visit by a certain user and the its type.

Later in this chapter we present the results obtained from the MNL model and their use with the aid of a simulator.

The phases of this are presented in the Gantt diagram below. With the delay in the previously planned Gantt being shown by the blue part of the bar.

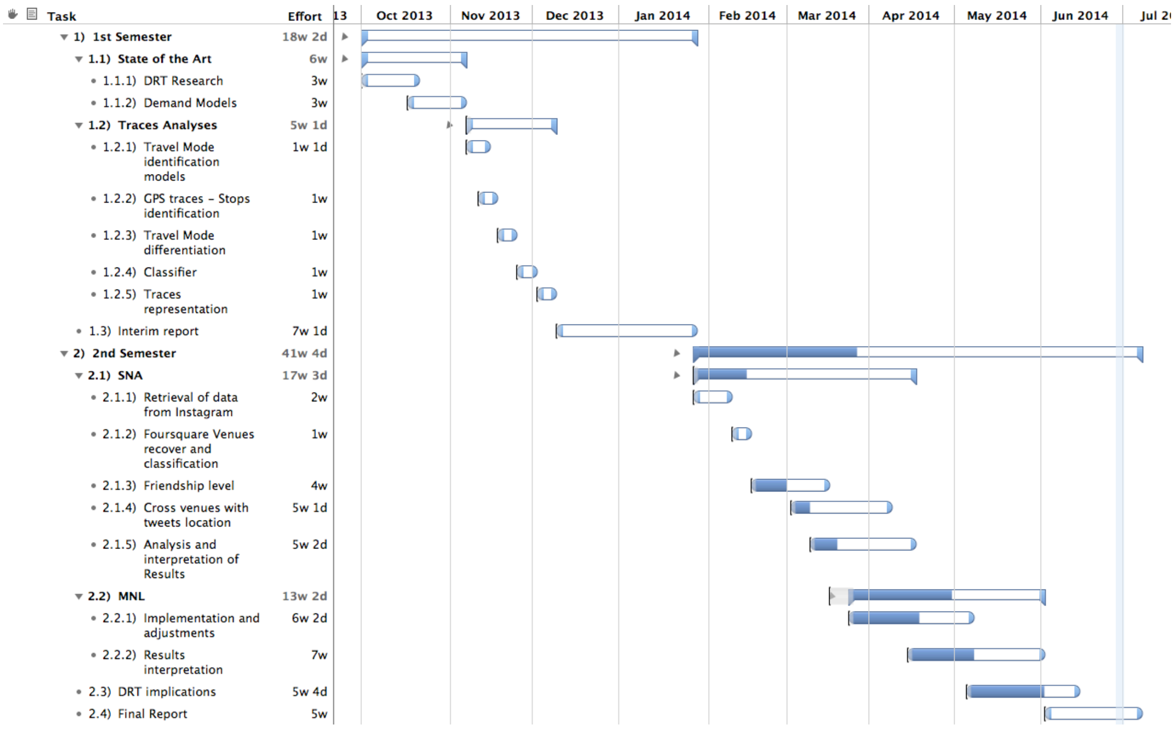


Figure 12 Gantt diagram

3.1 Data Preparation

3.1.1 Data gathering

For gathering data from social networks, Twitter Stream and Instagram API were used. The APIs work on a callback basis, so it was necessary to create a server, so that when new updates where available, the server would be notified. The callback on Twitter was made when a new tweet was posted and the Instagram callback when new media was shared. In the Twitter Stream the callback comes with the new tweet, as opposed to Instagram, that only notifies that new media is available. To work around this, when a callback from Instagram arrived, we make a subsequent request for the last data in the region, and filter old data from the response.

The data obtained is cleared of personal values as to ensure privacy, this values are mostly the user name, since age, gender and home location are not available. The data saved to the SQL database was:

- User identification - Enable new tweets identification and tractability.
- Geo-location - To know the real world location of the user.
- Timestamp - The post date of the image or tweet.
- Tweet – The tweet message was saved, although it was not used.

To get the geo-located points of interest, we used the FourSquare API. We made hundreds of requests to get the fifty most popular venues, within a radius of thirty meters of a given center. From those calls we obtained thousands of venues, with their identification, geo-location, category and total number of check-ins made. The total data amounts to 489 distinct categories from 37506 different venues.

Table 3 represents the amount of data collected and the collection period. FourSquare doesn´t have a start and stop time, as it was run a single time, contraire to Twitter and Instagram subscriptions. Figure 13, shows roughly the subscription zone for Instagram. The area for Instagram was rather small due to API limitations.

Table 3 Collection amount and time.

API	Start time	Stop time	Total data
Twiiter	28-12-2013	06-05-2014	513.146
Instagram	11-03-2014	27-05-2014	15.594
FourSquare	-	-	37.506

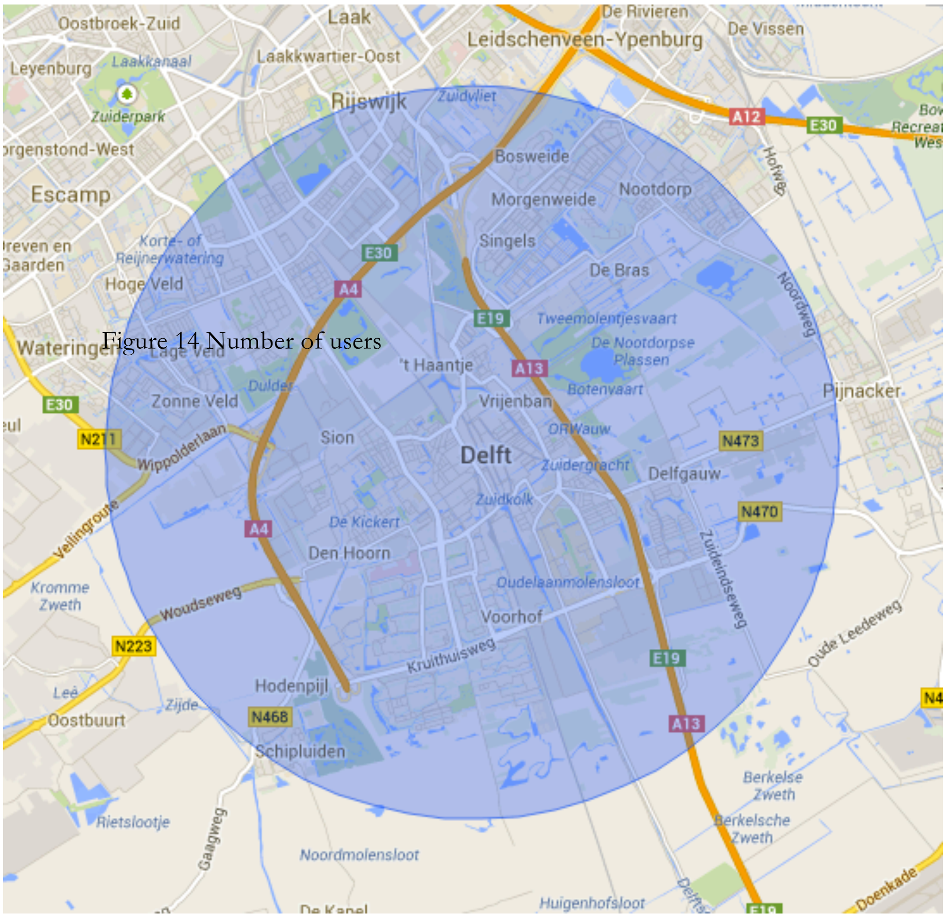


Figure 14 Number of users

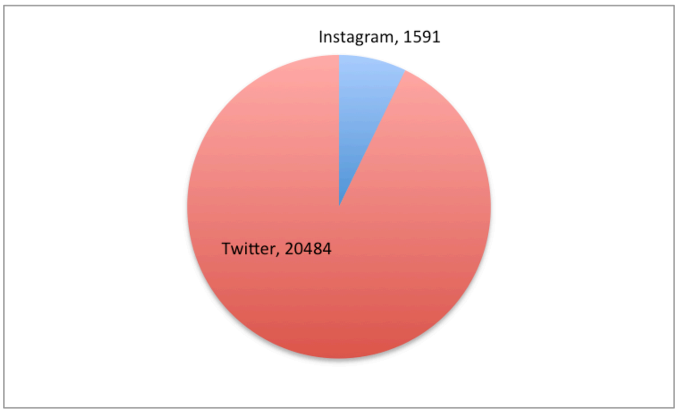
Figure 13 Instagram subscription

The data gathered was from Holland, with focus on Delft. The subscription zone for Instagram, compasses a radius of 5 kilometers from the city center. For Twitter, a square was created in the stream that covers a bigger area, so that we can get the surrounding data from Delft.

3.1.2 Match data

Once the data collection period was finished, the next step was to add information to the user's posts. To do so, we calculated the distance from every post to all venues, and considered posts within a radius of 10 meters from every the venue center, to be the user location. From the total amount of Twitter data gathered we were able to geo-tag 44.335 posts locations.

3.1.3 Data statistics



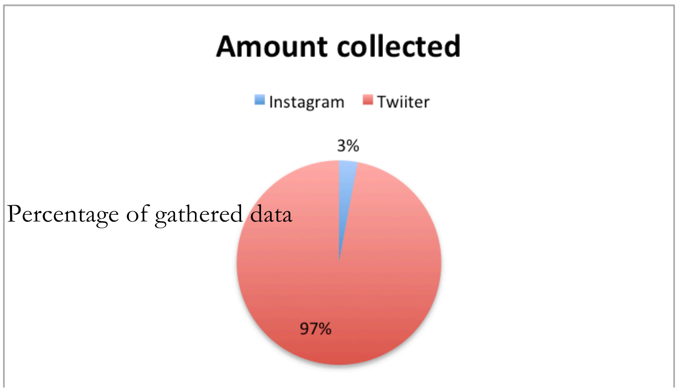


Figure 15 Percentage of gathered data

The above figures describe first amount of data capture and the second, the number of users that generated the data. At this point we discarded the Instagram data due to limited data and limited number of individual user's posts.

3.2 SNA analyses

As was said previously, we consider as friends social network users from the gathered post's that follow each other. To get the friendship, we had to use the user unique identification from the post, and request the user's that the user follows and that follow him. This process took quite some time, due to the quantity of user's and the limitation imposed by Twitter over the number of requests in time (rate limit), of 15 requests per 15 minutes.

3.2.1 Friendship

After getting the following and followers from the 20.484 users, we got a sum of 176.583 friendships. That data was then further analyzed, in order to see if a user that follows another user, is followed by him, and this resulted in a small reduction of friendships. Also we are interested in friends that reside inside our area of analyses, so the only significant friendships considered where the ones between users that posted around Delft. This further reduced the data to 35.457 friendships.

3.2.2 Important locations

For the data to be analyzed with the Discrete choice model, we needed to know the user home and work location, since we are interested only in the user movement patterns before and after work. Home and work where the starting points, to which the distance to points of interest were measured.

To get this central location, a cluster algorithm was used, namely DBScan. From the algorithm we obtained at least two clusters from the posts. Users who didn't have enough data to generate the clusters where then discarded. The main clusters where analyzed with the assumption that most of the tweets occur at the home and work locations, and were then divided in those two locations by the mean hour in each cluster.

Since the cluster corresponds to a number of points in a certain area, the centrality of the post's was used, and assigned to each user as their home or work location, as shown in Figure 16, where the green dots compose the cluster and the red dot the central point.

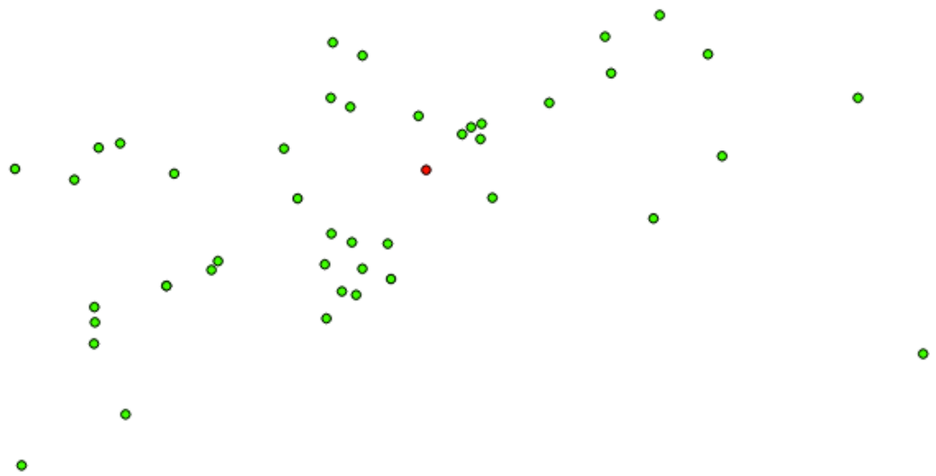


Figure 16 Cluster

3.2.3 Ties strength

As proposed previously, the Multinomial Logit model, hast to take into account the strength between users in its analyses.

Friendship between users has different strength's. To get and measure different ties strength, the tie mutuality, propinquity, mutual friends and multiplexity were used.

- Mutuality: The number of times, two users frequented the same location.
- Propinquity: Is tendency for users to have more ties with geographically close others. It takes the value of the distance between users central locations.
- Mutual friends: The number of friends that two users have in common.
- Multiplexity: With a Default value of 1 and a value of 2 if the two friends reside or work close to one another.

The final value is calculated and normalized between 0 and 1.

3.2.4 Pruned data

After the analyses of the posts, friendship, user area, we further discard users whose tweets information don´t have a location associated, since there's no information to be provided to the model. Table 4 gives an idea of the discarded data.

Table 4 data reduction

Type	Initial	After
Users	20.484	8.595
Friendships	176.583	35.457
Posts	513.146	40.149
Venues	37.506	6.749

As we can see, the initial data was quite reduced, and is further trimmed after we remove unwanted locations.

3.3 Demand Modeling

The R statistics system is a free to use tool that works for just about any statistical procedure, since its likely that someone has already written an R package to handle them. Such example of a package is the mlogit package (Written by Yves Croissant) for estimation of discrete choice methods. The following formula was used for our work,

Mlogit (choice~distance+friendship+attractiveness, CS)

where choice is the variable that indicates the choice made for each individual among the alternatives and the distance, friendship and attractiveness being the alternative specific variables with generic coefficients from the choice set *CS*.

3.3.1 Choice Set division

In the data set with the posts and venues associated, which we will now call choice set, there is a lot of data that doesn't interest us, some because they don´t provide useful information (e.g. Event Space, Town, City, Tunnel, River, Trail, Parking, etc.) or because they are work or residential places (e.g. School, University, Neighborhood, Hostel, Housing, etc.), in this cases the demand patterns are well established and can be met by traditional transportation, not requiring the route flexibility or timetable characteristic of the DRT service. To get around this, the data containing specific categories was erased of the choice set. The 6.749 venues resulted in 5.464.

The resulting venues will be our alternatives in the Discrete Choice analyses. As the number of alternatives is quite big, we grouped those venues in 6 main categories, represented in table 5 and Figure 17.

Table 5 Categories

Name	Categories
Food	Sandwich, Snack, BBQ, Pizza, Breakfast, Steakhouse,…
Bar	Bar, pub, nightlife, club,...
Shop	Convenience, Store, Shop, Market, Store,…
Entertainment	Theater, Zoo, Concert, Gym, Museum,…
Travel	Tram, Bus, Boat,…
Appointment	Office, Bank, Doctor, Voting, Hospital,…

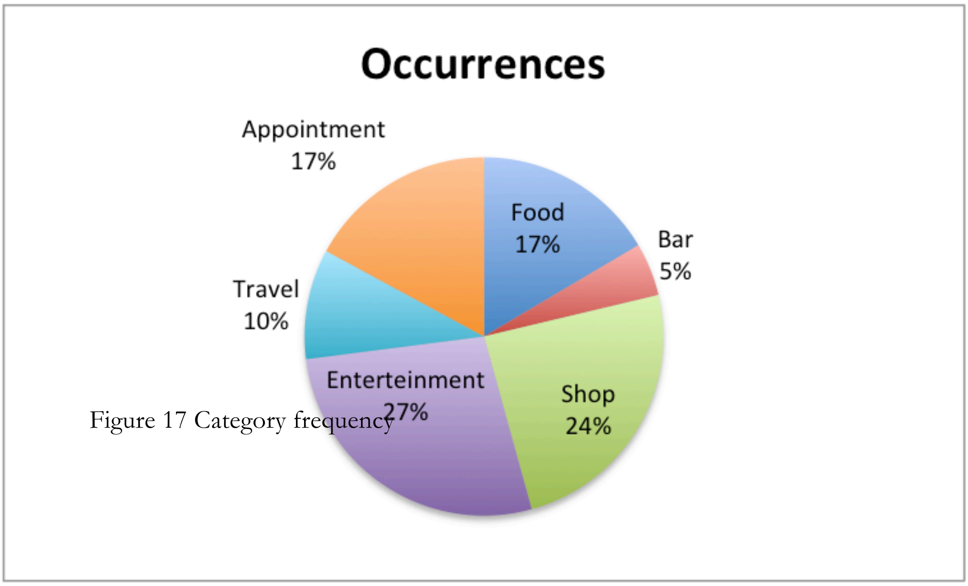


Figure 17 Category frequency

After grouping the categories into 6 main categories, we could use this 6 as our number of different alternatives for the Multinomial Logit model. But then we would only get results concerning each of those 6 alternatives, which are quite generic. Since we want to use the model to predict probabilities of destination choice for DRT services, whose service requires a higher resolution we will then generate data for all the venues and use those categories only to filter unnecessary data.

There are two main choices when dealing with individual data for discrete choice methods.

- *Reveled preferences data* which means that the data are observed choices of individuals.
- *Stated preferences data* in this case the individual faces a virtual situation of choice.

Our data corresponds to the reveled preferences data, since we know that user **u** was at a specific location **l.** Also we have repeated observation for each user, and in the MNL there's only one alternative chosen by each individual. To make the most of the gathered data, the choice set is then divided by hours with the finite number of alternatives varying for each hour and main category as shown in Table 6, as well as the number of observations.

Table 6 Choice set division.

Hour	*Observations*	*Alternatives*
0	14	4
1	9	3
2	7	2
3	3	1
4	0	0
5	26	4
6	37	10
7	65	14

8	117	26
9	98	21
10	117	30
11	205	58
12	172	38
13	155	36
14	164	39
15	144	36
16	138	36
17	184	45
18	188	38
19	140	34
20	88	20
21	91	24
22	24	8
23	32	8

Data sets for multinomial logit estimation deals with some individual, that make a choice of one alternative among a set of several alternatives. Alternatives with only 2 or less individuals were removed.

For each dataset we have the number of alternatives selected in each hour, which will be our finite set of alternatives for each individual. As we can see in the above table the number of alternatives and observations vary a lot trough the hours. These alternatives correspond to the individuals choice of alternative for a certain hour. Since we divided it by hours, the number of alternatives will vary and the same can be said for the observations, since we need to select one observed choice of alternative from the data set.

3.3.2 Variables considered

When working with multinomial logit models, 3 types of variables must be taken into account:

- alternative specific variables x_{ij} with a generic coefficient β,
- individual specific variables z_i with an alternative specific coefficient γ_j,
- alternative specific variables w_{ij} with an alternative specific coefficient δ_j.

The variables used for the data-frame where:

- distance – the venue distance to the user central point,
- check-ins – the total number of check-ins in each alternative for each user,
- friendship – the sum of the individual friendship for each alternative,

- choice – the alternative selection.

Since, we can't directly extract user information from the social network (e.g. age, gender) our data doesn´t contain individual specific variables, and the alternative specific variables have a generic coefficient, since we consider that the number of check-ins, distance and friendship have the same value for all alternatives (e.g. 1 euro on a train is the same as 1 euro in a taxi, but the same cannot be said for time, which would be an alternative specific coefficient). Choice takes values of "yes" and "no", if the alternative was chosen or not by the user.

Chapter 4
Results

To estimate this data, it is not enough to give the attributes of the chosen alternative, but on all alternatives. For example, its not sufficient to determine the distance, attractiveness and friendship of the choice made by the individual, its also necessary to know those attributes if other alternative had been chosen by the same individual.

4.1 MNL

We will present the results and estimation parameter for one choice set, namely the one representing the choices made at hour 21, which has 24 alternatives and 91 observations. Table 7 corresponds to the frequency of the alternatives chosen, while Table 8 and 9 presents the results.

Table 7 Alternative frequency

Venue	*Frequency*	*Probabilities*
Stadion Feijenoord	0.032967	0.04490835
EkoPlaza	0.065934	0.03791752
Station Den Haag HS	0.054945	0.05494505
La Mer	0.032967	0.03303908
Diner Company	0.032967	0.03777430
LantarenVenster	0.032967	0.04320755
Station Rotterdam Centraal	0.153846	0.12336128
BIRD	0.043956	0.03794951
Maassilo	0.043956	0.03530117
Emma	0.021978	0.02478593
Station Den Haag Centraal	0.054945	0.05070866
Lucent Danstheater	0.032967	0.02519137
Zaal 3	0.032967	0.03212339
De Banier	0.054945	0.09073518
Randstadrail javalaan	0.032967	0.02411359
Kot Treinpersoneel	0.032967	0.02666024
Spuimarkt	0.032967	0.02543518
Doerak	0.032967	0.04017886
Paard van Troje	0.032967	0.03294463

Ahoy Rotterdam	0.043956	0.03654922
Restaurant Meram	0.043956	0.03907711
Live Tv Show	0.021978	0.04634078
Stadskwekerij den haag	0.010989	0.03433138
Oudedijk 166 A2	0.032967	0.04312563

The alternative frequency is the percentage that an alternative was chosen in the choice set and the probability is the predicted probabilities for all the alternatives by the model. Comparing the two values for more than one dataset, we can see that the discrepancy between them is small, meaning that the alternatives are being well predicted.

Table 8 Coefficients Intercepts

Intercepts	**Estimate**	**Std. error**	**t-value**	**Pr(>| t |)**
Emma	0.372905	0.994081	0.3751	0.7076
Kot trein.	0.704053	0.900677	0.7817	0.4344
Doerak	0.126631	0.964026	0.1314	0.8955
Ahoy Rott.	0.510281	0.803442	0.6351	0.5254
Maassilo	0.252294	0.850796	0.2965	0.7668
EkoPlaza	1.091659	0.814275	1.3407	0.1800
Spuimarkt	0.754647	0.902787	0.8359	0.4032
S. Den Haag	0.999597	0.825177	1.2114	0.2258
Diner C.	0.237901	0.965728	0.2463	0.8054
Stat. Rott.	0.746720	0.723048	1.0327	0.3017
La Mer	0.318648	0.848980	0.3753	0.7074
Rest. Meram	0.426253	0.810011	0.5262	0.5987
Stadskwekerij	-1.837684	1.830452	-1.0040	0.3154
Lucent	0.765183	0.901061	0.8492	0.3958
Zaal	0.344013	0.850835	0.4043	0.6860
LantarenVenster	0.074625	0.852844	0.0875	0.9303
Randstadrail	0.809245	0.901150	0.8980	0.3692
De Banier	-1.721859	1.343160	-1.2819	0.1999
Den Haag Cent-	0.608758	0.875576	0.6953	0.4869
BIRD	0.476525	0.799113	0.5963	0.5510

Paard	0.507464	0.924042	0.5492	0.5829
Tv	-1.218937	1.394182	-0.8743	0.3820
S. Feijenoord	-0.074893	0.887996	-0.0843	0.9328

Table 9 Coefficients variables

Variables	**Estimate**	**Std. error**	**t-value**	**Pr(>| t |)**
distance	-0.107414	0.022597	-4.7534	2.000e-06 ***
friendship	0.094441	0.081570	1.1578	0.2469
attractiveness	0.342170	0.061907	5.5272	3.254e-08 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

As we can see, the result from the model gives the estimation results: the estimated coefficients, their estimated standard error, the t-statistics, the probability, under the null hypothesis that the true value of the coefficient is zero, of observing a t-value greater than the computed one; and finally a graphical indication (stars) of the significance level of the coefficient.

As expected the distance as a negative value, meaning that as the distance increases (and the other distances to the alternatives remain the same) the probability of that alternative being chosen falls.

The estimated model will give us the predictions for all alternatives for all individuals, making it useful for transportation analyses.

4.2 Evaluation

As we previously showed (4.4.1) our model makes good predictions. The model also generates for each individual the probability for all alternatives, even when there are no indications on the data set that individual **i** chose alternative **j**. Since our variables are significant and the predictions are good we could say that the prediction for unobserved individual choices are also good – the predicted probabilities for each alternative are the sum of the predicted probability of all individuals for an alternative.

Since we have some observations for the same user, and only one is passed as the choice to the choice set, we could see if the probabilities predicted match those observations. The only problem with this approach, is that the number of alternatives in the choice set is far superior to the individual observations of alternatives. To get around this, we use the sum of the probabilities predicted for each alternative and compare them to the observed choices of all individuals (Table 10).

Table 10 evaluation

Venue	*Avg. Probabilities*	*Observations frequency*
Stadion Feijenoord	0.04490835	0.0229007
EkoPlaza	0.03791752	0.0458015
Station Den Haag HS	0.05494505	0.0381679

La Mer	0.03303908	0.0229007
Diner Company	0.03777430	0.0458015
LantarenVenster	0.04320755	0.0229007
Station Rotterdam Centraal	0.12336128	0.1374045
BIRD	0.03794951	0.0305343
Maassilo	0.03530117	0.0229007
Emma	0.02478593	0.0229007
Station Den Haag Centraal	0.05070866	0.0839694
Lucent Danstheater	0.02519137	0.0229007
Zaal 3	0.03212339	0.0229007
De Banier	0.09073518	0.0839694
Randstadrail javalaan	0.02411359	0.0381679
Kot Treinpersoneel	0.02666024	0.0305343
Spuimarkt	0.02543518	0.0229007
Doerak	0.04017886	0.0687022
Paard van Troje	0.03294463	0.0305343
Ahoy Rotterdam	0.03654922	0.0305343
Restaurant Meram	0.03907711	0.0381679
Live Tv Show	0.04634078	0.0534351
Stadskwekerij den haag	0.03433138	0.0305343
Oudedijk 166 A2	0.04312563	0.0305343

The data presented in the table above, is from the data set corresponding to hour 21. As we can see the probabilities have a closer match in some alternatives than others, this can be explained by the individual alternative frequency observed, as the observation data doesn't have the same number of observations per individual. Other data sets also present close similarity between the sum of the alternatives predictions per individual and the frequency of individuals observations.

Chapter 5
Simulation

Simulation and analyses tools, constitute a powerful mean to evaluate system capacities and have been used intensely in transportation literature and others.

The use of simulation software, became essential to service planning. The simulation can comprise requests booking, vehicle utilization, route definition, stops and schedule, allowing the validation of optimized services that provide DRTs a quick demand response and flexibility in terms of time and client location and at the same time minimizing costs., improving overall efficiency.

To show the usefulness of the analyses made in chapter 6, we feed the probabilities predicted in our model to a decision support system, developed by (Gomes 2013).

5.1 Simulation tool

The simulation tool used, resulted from the work done in creating a system to support decisions, developed by (Gomes 2013) in his PhD work. The decision support system is based on the model of Dynamic Vehicle Routing (DVR), and integrates optimization and simulation for operational costs reduction and maximization of service quality. The system presents some specific characteristics such as:

- Vehicle capacity
- Requests time window
- Stops, can be any pre-defined point along the route, for passenger retrieval or drop off.

The system allows booking and real time requests. The later may need new route and schedule calculations. The development of this tool was based in the random event modulation by stochastic models for generating requests in space and time. The time between consecutive requests is modeled with a negative exponential distribution. The origin and destination locations are defined by the Origins-Destinations matrix of the service area. The events associated to the user service follow a Poisson distribution, applied when generating transport requests, request cancelation and no show of users. Events related with the vehicle, like stop arrival or malfunctions during service are also modulated with Poisson distributions.

5.1.1 Logic structure

The "DecisionSupportSystem" (DSS) enables two interfaces, one dedicated booking requests (IRequest) and another dedicated to real time events (IEvents). IRequest allows the client applications to make new transportation requests and cancelations. IEvents allows, for example, to external systems to insert event for route re-planning . The vehicles can be equipped with hardware that sends events like vehicle breakdown, service pause and no show to the DSS. Figure 18, represents the different layers presented in the system logic structure.

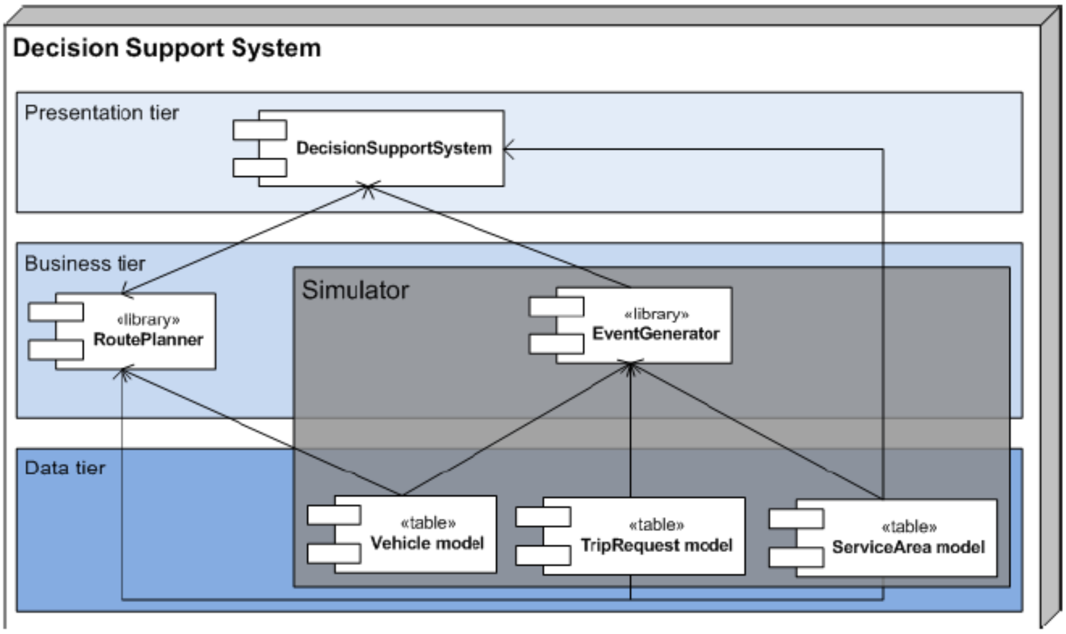


Figure 18 Logic system structure

5.1.2 Functionality and Parameters

In terms of service vehicles, the simulation allows for homogeneous fleet and recourse to sub-contracted vehicles, with different capacities, acquisition and fuel cost. The private fleet and the sub-contracted fleet can be defined by determining the following parameters:

- Capacity
- Vehicle acquisition cost
- Fuel cost per kilometer
- Commercial velocity

The simulator user determines the rate of request generation, rate of cancelations and no shows. The Poisson distribution determines the intervals between requests, cancelation and no shows generation.

The user can define the service operation. The expected travel time is estimated trough mobility studies in the area.

The simulator defines the network trough nodes, that represent the stops and edges that represent the connection between stops. The flow between nodes is estimated according to the generation and attraction probabilities for each zone/stop introduced by the user.

5.2 Simulation results

Since the purpose of this work was to model demand for responsive transportation, we will not vary parameters in the simulator, with the intention of planning or optimizing the service. We use the simulator only to provide some perspective to the work done by the discrete choice modeling. To that effect, the probabilities predicted by the model are feed to the simulator in the form of two vectors, one containing the probabilities of all the alternatives in the choice set, which will be the destination vector, and the vector containing all the individuals will be the origins vector.

The model results used in the simulator are from the data set corresponding to the hour 21, with the limitation of considering only venues that where the individuals choices in more than 5 distinct occasions, this produced 4 locations and 22 different individuals occurrences. The idea was just to have a simple example - since the purpose of this work was to model demand for responsive transportation, and not plan a DRT – and so we imposed this limitation on the data set. There were also some limitations on the simulation tool on the size of the test instances given the precision we wanted in terms of trip distribution.

The probabilities predicted from the MNL were used to form the destination vector and the individual locations form up the origins vector. For the depot of the service we used the geographic center of Delft. Some noteworthy variables in the simulator configuration are: the number of vehicles in the fleet (5), their commercial speed, window of operation (1 hour and 30 minutes), number of requests (200), operation costs (vehicle, gas) and the network that was created with the venues and individuals location as well as the OD matrix.

The parameter called "degree of dynamism" refers to the percentage of request made in real time (i.g, during the service operating hours). We left it at 50% meaning that the number of initial request when the service starts is 100 and that 100 more will arrive during the service.

For public transportation, Delft offers the options of train, tram and bus. The city has two train stations (Delft Zuid and Delft Center), two tram lines, number 1 (Scheveningen - Delft Tanthof) and 19 (Leidsehage - Delft Tanthof) and local and regional buses (Figure 19).

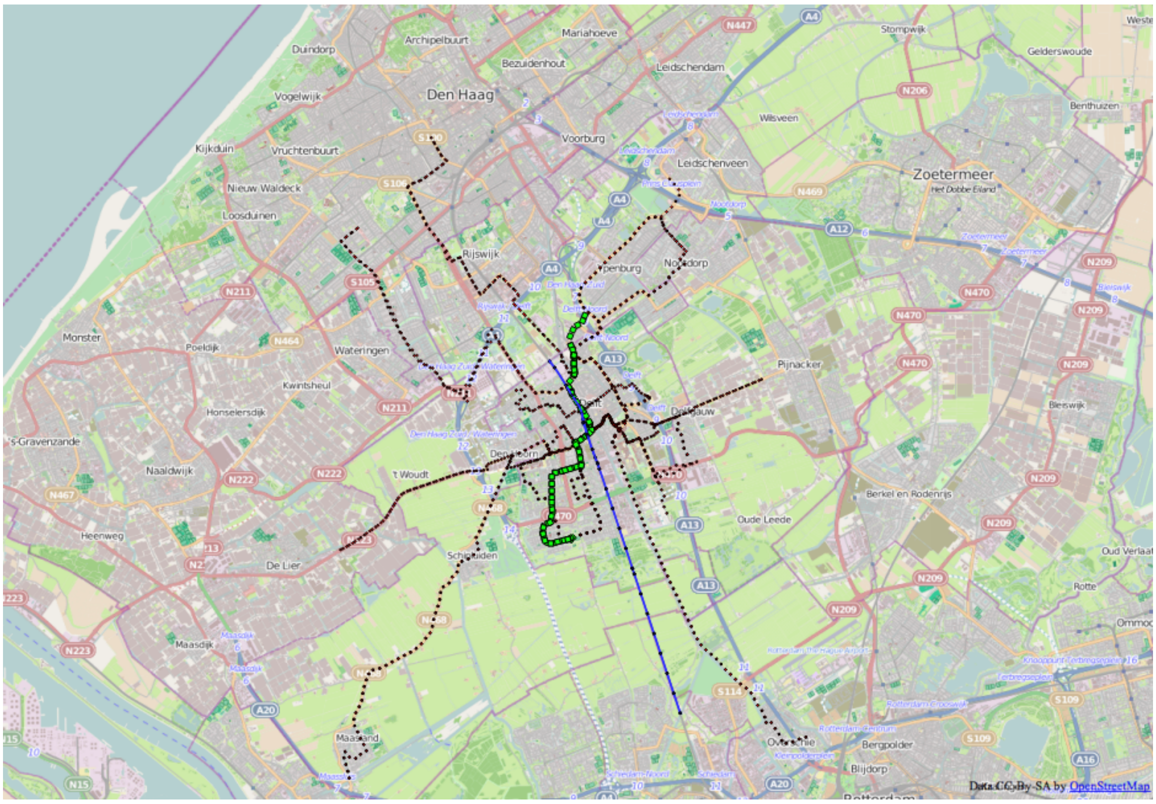


Figure 19 Red, Bus and Railway lines

The figure represents Delft public transport routes. The red squares are the bus service, the green are the trams and the blue the railway.

The traditional public transport operates on a fixed route basis. Services being identified with a fixed or timetabled operating pattern. As we can see from the five routes calculated by de simulator in Figure 20 in response to the demand, the flexibility offered by a DRT (booking, real-time and door-to-door) service is more adequate to meet that demand compared to the traditional transportation route and schedule.

Figure 20 Route 1 and 2

Figure 21 is the representation of the 5 (blue, red, yellow, orange and violet) routes calculated by the simulator and the public transportation system.

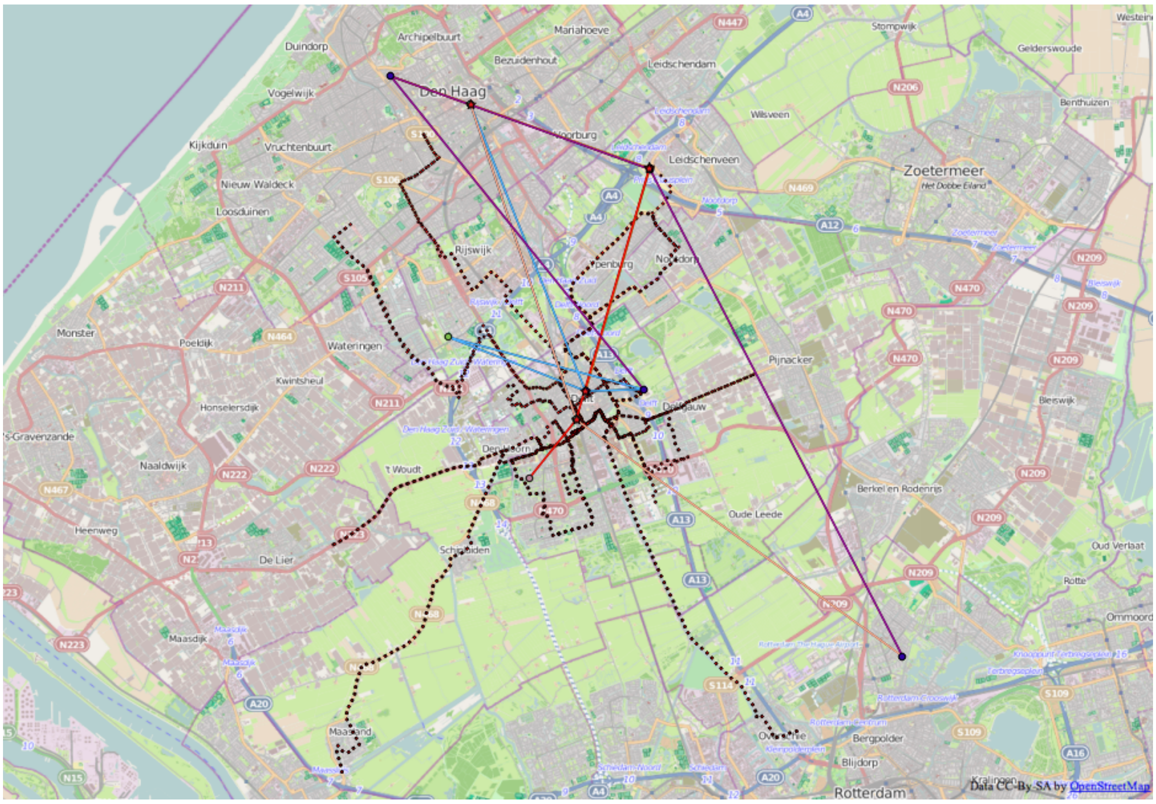


Figure 21 Route 1,2,3,4 and 5

Chapter 6
Conclusions

To have a DRT service more viable from a financial point, its important to explore ways to predict future demand. With the rapid growth in Information and Communication Technologies, new doors to explore demand are opened, as is the case with social networks. That gives way for behavior modeling by way of Social Network Analysis. Since travel is behavioral in nature and that friends can impact our travel decisions, its important to explore such relation.

The data gathered from Twitter and Instagram, the later being discarded and the former presenting a significant number of non-important data, such as work locations or non-identified locations through foursquare, as well as the number of observations for the same venue being reduced, presents a challenge when using the model. Also the frequency of quality posts (with identified locations) for each user, makes it difficult to generate a pattern for each user.

The model predictions were reasonable good when tested against the user observed choices. The results from the MNL model show meaningful relationships between distance and attractiveness for all the different alternatives, with the variable *distance* being the most significant, meaning that longer distances almost always reduce the attractiveness of a destination, all else being equal. The same can be said for the attractiveness variable, but the friendship variable doesn't have the same impact to the individual when choosing an alternative.

The value of the analyses done becomes apparent when worked on with the simulator, where we can observe the impact that the demand predicted has in terms of costs, number of vehicles and route planning of the service. Also we can see that this type of demand its not adequate to the traditional transportation, since it was focused outside the well established home-work and work-home patterns.

Since the analyses of the social network done in this work, doesn't produce individual characteristics like age, gender and socio-economic, it would be interesting for future work to include data mining algorithms to extract some of those values from tweets, join this with socio-demographic factors and add features specific to each venue, to better understand the motivation behind the choice made. With this our model would have individual specific variables was well as alternative specific without generalized coefficients, and we could also test with other logit models to relax the IIA property of the MNL presented.

Bibliography

1 Muscat, Adrian. Sustainable Mobility in Malta. *Smart City* (November 2012).

2 Bakker, P. Large scale demand responsive transit systems - a local suburban transport solution for the next millennium. *European Transport Conference, Cambridge, UK* (1999).

3 Black, A. Urban mass transportation planning (1995).

4 T. Sihvola, L. Ha¨me, and R. Sulonen. Passenger-pooling and trip-combining potential of high-density demand-responsive transport. *TRB 89th Annual Meeting* (2010).

5 K. Zografos, et al. A Methodological Approach for Developing and Assessing Business Models for Flexible Transport Systems. *Transportation*, 35 (2008), 777-795.

6 Julie Prud'homme, Didier Josselin, Jagannath Aryal. Quantitative Analysis of Pollutant Emissions in the Context of Demand Responsive Transport. *Computational Science and Its Applications - ICCSA 2011* (2011), 439-453.

7 G. Ambrosino, J.D. Nelson, M. Romanazzo. Demand Responsive Transport Services: Towards the Flexible Mobility Agency (2003).

8 S. Ichoua, M. Gendreau, and J. Potvin. Exploiting knowledge about future demands for real-time vehicle dispatching. *Transportation Science, 40(2)* (May 2006), 211-225.

9 Gomes, R. Dynamic vehicle routing for demand responsive transportation systems. *PhD Thesis, FEUP* (2013).

10 Bhat, C.R., and Koppelman, F.S. A Retrospective and Prospective Survey of Time-Use Research. *Transportation, 26(2)* , 119-139.

11 Chapin, F.S. Human Activity Patterns in the City: Things People do in Time and in Space New York (1974).

12 Wakefield, Jane. *Mobile phone data redraws bus routes in Africa*. BBC News, 2013.

13 Samiul Hasan, Xianyuan Zhan, Satish V. Ukkusuri. Understanding Urban Human Activity and Mobility Patterns Using Large-scale Location-based Data from Online Social Media (August 2013).

14 Sivakumar, Aruna. A Synthesis of Transport Modelling Methodologies (September 2007).

15 Dagsvik, John K. Multinomial Choice and Selectivity. *Research Department of Statistics Norway* (2000).

16 Moshe E. Ben-Akiva, Steven R. Lerman. Discrete Choice Analysis: Theory and Application to Travel Demand (1985).

17 VTM. Implementing Activity-Based models in Virginia (July 2009), 31.

18 Pendyala, Ram M., Huichun Zhao, Ryuichi Kitamura. Telecommuting and Travel Demand: An Activity-Based Impact Assessment. *Research Report UCD-ITS-RR-89-02* (1989).

19 716, NCHRP. Travel Demand Forecasting: Parameters and Techniques (March 2012), 170.

20 NPTEL. Trip Distribution (May 2006).

21 Cascetta, E. Transportation Systems Analysis. *Springer Optimization and Its Applications 29* (2009).

22 Jonas Anderson, Jan UBØE. Some aspects of random utility (January 2010).

23 Ben-Akiva, Michel Bierlaire. Discrete Choice Methods and Their Application to Short Term Travel Decisions (April 1999), 43.

24 M., Ben-Akiva. Structure of Passenger Travel Demand Models. *Ph.D dissertation* (1973).

25 M., Ben-Akiva. Note on the Specification of a logic model with utility functions that include attributes of competing alternatives. *MIT* (1974).

26 Wooldridge, Jeffrey M. Econometric Analysis of Cross Section and Panel Data (2002), 501-502.

27 de Graaff, T., Rietveld, P. Substitution between working at home and out-of- home: the role of ICT and commuting costs. *Transportation Research Part A: Policy and Practice 41* (2007), 142-160.

28 Kenyon, S. The 'accessibility diary': discussing a new methodological approach to understand the impact of Internet use upon personal travel and activity participation. *Journal of Transport Geography 14* (2006), 123-134.

29 Kwan, M.P. Mobile communications, social networks, and urban travel: Hypertext as a new metaphor for conceptualizing spatial interaction. *The Professional Geographer 59* (2007), 434-446.

30 Lyons, G. The reshaping of activities and mobility through new technologies. *ournal of Transport Geography 17* (2009), 81-82.

31 Eunjoon Cho, Seth A. Myers and Jure Leskovec. Friendship and Mobility: User Movement In Location-Based Social Networks (2011).

32 Alexey Tarasov, Felix Kling and Alexei Pozdnoukhov. Prediction of User Location Using the Radiation Model and Social Check-Ins (August 2013).

33 Brock, W. A. & Durlauf. Discrete choice with social interactions. *Review of Economic Studies*, 68 (2001), 235-260.

34 Brock, W. A. & Durlauf. A multinomial-choice model of neighborhood effects. *American Economic Review - Papers and Proceedings* , 92 (2002), 298-303.

35 Ryley, Alberto M Zanni and Tim J. Exploring the possibility of combining discrete choice modelling and social networks analysis: an application to the analysis of weather-related uncertainty in long-distance travel behaviour. *Discussion paper for presentation at the International Choice Modelling Conference* (July 2011).

36 Ryley, Alberto M Zanni and Tim J. Discrete choice and social networks: an analysis of extreme weather uncertainty in travel behaviour (March 2013).

37 Carrasco, J.A., Hogan, B., Wellman, B., Miller, E.J. Collecting social network data to study social activity-travel behavior: an egocentric approach. *Environment and Planning B: Planning and Design 35 (6)* (2008).

38 Axhausen, K.W. Social networks, mobility biographies, and travel: survey challenges. *Environment and Planning B: Planning and Design 35 (6)* (2008).

39 Paez, A., Scott, D.M. Social influence on travel behavior: a simulation example of the decision to telecommute. *Environment and Planning A 39* (2007), 647-665.

40 A. Páez, D.M. Scott. A Discrete Choice Approach to Modeling Social Influence on Individual Decision Making (2007).

41 Gulyás, Elenna R. Dugundji and László. Sociodynamic Discrete Choice Applied to Travel Demand: Multi-Agent Based Simulation and Issues in Estimation (2012).