Geographical Information Management and Applications

THESIS

The Traveling Employees Problem

Author:
Laurens BAKKER BSc

Supervisor:
Aldo BERGSMA MSc

Responsible professor:
Ron VAN LAMMEREN Dr MSc

June 5, 2019

GEOGRAPHIC INFORMATION MANAGEMENT AND APPLICATIONS

Abstract

For obtaining the degree of Master of Science

The Traveling Employees Problem

by Laurens BAKKER BSc

Student employment agencies face the daily challenge of making sure that their employees reach the scheduled work location on time with as little travel time as possible, all the while trying to keep the financial cost low. This problem could be called the Travelling Employees Problem (TEP). This research has two objectives, first to formulate a conceptual model; it should solve the TEP to support the efficiency of expert planners. Second, this model will be implemented in the form of a computer script. This implementation should be sufficiently fast in terms of computational time. Based on these objectives four questions can be posed. *What concepts could be used to solve the travelling employees problem?* This question is answered by reviewing literature about the space time prism, clustering, and the value of commuting time. *How can these concepts be used to form a model that solves the travelling employees problem?* This question is answered by aligning the found concepts to form a heuristic model. *How does this model perform?* This question is answered by performing a case study on historic data from the student employment agency LINQ in Amsterdam, The Netherlands. *How do experts judge the implemented model?* Three experts are asked to first compete against the model, and later to judge the validity of model output. It is concluded that the model is able to support expert planners, but not (yet) replace them.

Acknowledgements

I am thankful to my supervisor at LINQ. I have been given the liberty in my research on the one hand, and my work related activities on the other. But first and foremost there was no doubt when I asked for two months off when the opportunity to go abroad arose rather abruptly. The question of how I would go, not if I should go was the only matter. Thus my personal development was placed above all else giving me the opportunity of a lifetime. For that I am thankful.

I am also thankful to my girlfriend, who supported me after long, sometimes grumpy days. It is the little things in life that matter most, and she knows it.

My supervisors from the University have always remained positive and kept challenging me. This motivated me to incorporate feedback and improve my research rather than making me feel beat down after a feedback session. A property that should be standard for teachers, yet seldom is.

I learned that the reality is not always as neat and clear as is science. The models we read about are applied in an environment that is not ideal, which means that real compromises are to be made. This was both true for my research as the work I did next to it.

Modelling is something I never truly questioned. Although we are told at university that models are a simplification of reality, it never deemed to me that some things might actually be 'lost in translation', so to speak.

Contents

1 Introduction

1.1 Problem identification

Student employment agencies face the daily challenge of making sure that their employees reach the scheduled work location on time with as little travel time as possible, all the while trying to keep the financial cost low. Such an agency mostly employs students that do not have a permanent side job, but students that only require work occasionally. Such an agency offers the students work on a more flexible basis compared to a permanent side job at a single employer. An agency can have multiple clients that require a variable amount of work throughout the year. An example could be a festival that requires bartenders, medical first responders, security officers, and building crews for the limited duration of the festival. Another is the case of conference centres in which the amount of work varies throughout the year. Thus the students and the agency generally have a flexible supply/demand relation instead of a fixed amount of hours per month.

Student commuting trips take place when the agreement is reached about a student going to work at a certain client at a certain time and location. The commuting questions posed above now becomes relevant. Especially since students in the Netherlands are generally limited in their modes of transport to walking, cycling, or taking public transport. Furthermore it is of importance to all three parties that the commuting trip is as efficient, in terms of time and money, as possible. For the employment agency it is also relevant to have an exact measure for expected travel cost, as this could be used on the strategic level. Failing in optimising such a trip will cost money to one of the, if not all, parties. Additionally having an unnecessarily long trip will all but improve the performance of the student at work; it has been found that longer commutes are associated with increased perceived and physical stress (Evans and Wener 2006).

Current employment agency student trip planning takes place mostly manually with the support of some digital planning support system. In general the process of planning involves evaluating whether or not students could walk, cycle, or take public transport to the working location, or that a car should be arranged. These different types of transport are also called modalities, or modes (of transport). It is also evaluated who of the available students should be assigned to the job. If a car is arranged, *'who should drive?'*, *'where should the car depart from?'*, *'what route should the car take?'*, and *'is it better to use one full car with a longer route, or two half full cars with a shorter* *route each?'*. Routes are in turn dependent on the road or public transport network. Thus implying that this is a network problem. Currently questions about the planning are mostly answered by common sense and approximate calculations.

1.2 Problem definition

What is the problem faced by the student employment agencies? In logistical planning their goal is twofold. First, they want to keep their students and employers happy. Second, they want to spend as little money on commuting as possible. This is done by doing two things. First, they should assign the students to the job that are in the most favourable position to travel. Second, they should plan this single trip as efficiently as possible. The efficiency of the trips the students make are constrained by several factors that can be categorised in one of groups; employee (student), employer, or mediator related. The mediator being the student employment agency. These factors can also be interdependent.

1. Employee (student)
   (a) *Home locations.* Given a group of employees (1...n); they each start their trip to the working location from different locations. In this research it is assumed that they initiate their trip from home.
   (b) *Drivers licence.* Students might, or might not have a drivers license. Especially in The Netherlands it can not be assumed that every student has a drivers license because the bike and public transport provide, arguably, sufficient means for travelling.
   (c) *Availability.* A single student might not always be available to go and work. Some might prefer working in the weekends, others in the evenings or on a specific day during the week.
2. Mediator
   (a) *Changing travel cost throughout time.* Travelling at different times and dates entails changing cost for equal trips. The moment at which travelling takes place is dependent on the time the shift starts and thus the below mentioned arrival time. An employee travelling by car from home to work will experience different travel time and financial cost for travelling in- or outside rush hour. The same goes for travelling by public transport. In the Netherlands students can have one of two types of discount cards with varying discount rates throughout the day, week, and year. It can therefore be said that the transport network as a whole has time dependable cost. And although this might be seen as an external variable, careful planning will allow the student employment agency to change the cost of the network, e.g. through carpooling.

(b) *Travel time versus financial cost.* A firm does not always financially compensate travel time or direct travel cost such as public transport fees or fuel consumption. Yet those cost are often made. If all students were to walk to work this would be the cheapest in terms of money, but this situation seems undesirable because faster methods of transport exist. Hence a balance between the travel time and travel cost needs to be found. This scale is to be balanced by the mediator. Throughout the remainder of this research the term cost refers to both financial as well as temporal cost.

(c) *Modality.* There are several travel modes available for the employees to travel to their work. In this research the modalities will be limited to car, public transport, bike, and walking. It is up to the mediator to select the 'best' modality for the situation at hand. These modalities are also constraining the trip back home. In other words, cars cannot be created out of thin air for the home bound trip. Also, it is particularly relevant when working hours are late in the night so that public transport connections are non existent.

(d) *Carpooling.* Given that at least one of the employees travels by car, it should be evaluated if other employees should join or not to increase overall efficiency, not to forget the reduction of the environmental footprint. The car has a maximum capacity. Common cars have five seats, including the drivers seat. These cars can theoretically depart from almost any location, although intuitively it is more practical to have the driver pick up her/his colleagues.

3. Employer

   (a) *Required amount of employees.* A certain employer might require a certain amount of employees for a certain job. In the case of the aforementioned festival, tens, if not a few hundreds, of employees are needed to facilitate the festival.

   (b) *Work location.* Throughout the day multiple employees can work at multiple locations, but generally one employee will work at one location during one day. One location where the employee performs work is the work location.

   (c) *Arrival time.* The time at which the employees are to arrive at the working location will be before the actual work starts.

More formally this can be stated as *'Given the above employee, mediator, and employer constraining factors, what is the least costly combination of trips to perform a forward and backward trip at different times?'*. Let us call this problem the **T**ravelling **E**mployees **P**roblem (TEP), with a slight wink to the classical Travelling Salesman Problem (Rez 1832).

1.3 Objectives and research question

This research has two objectives. First a conceptual model will be formulated. It should solve the TEP to support the efficiency of expert planners. Second, this model will be implemented in the form of a computer script. This implementation should be sufficiently fast in terms of computational time. Based on these objectives the following questions can be posed.

1. What concepts could be used to solve the travelling employees problem?
2. How can these concepts be used to form a model that solves the travelling employees problem?
3. How does this model perform?
4. How do experts judge the implemented model?

The first question is of a theoretical nature and will be answered through a literature review in chapter 2 Review. The second question will be answered by combining some, or all, concepts from the review in a model in chapter 3 Model formulation. The third question will be answered by applying and assessing the model through a case study in chapter 4 Application and assessment. The fourth question will be answered by letting experts compete with and validate the model output from the case study. This will be done in chapter 5 Expert Validation. Finally the findings of each of the previous chapters will be discussed against the broader literature which will result in a conclusion in chapter 6 Discussion and Conclusion.

2 Review

In seven sections this chapter will familiarise the reader with concepts that are relevant to the TEP. The goal is to be able to formulate a model in chapter 3 starting on page 23 based on the concepts explained here.

The structure of this chapter will be based on the constraints as discussed in the introduction. Of these constraints the employee (student) and employer constraints will in almost all cases be fixed, meaning that a student employment agency will have a given set of students that are available to work on the one hand, and a given location and time and required amount of students on the other. They are to mediate in matching this supply and demand. Hence the employment agency can most likely not alter any of these factors during the mediation process. The mediator constraints however, are up to the mediator to determine/influence.

As this research is hardly the first to investigate a problem such as the TEP, the first section 2.1 starting op page 6 will give a short historic overview and subsequently explain the basics of graph theory; an approach that can be used to solve problems like the TEP. It will also explain why the most intuitive and perfect method to solving these kind of problems, which is simply trying all possibilities and selecting the 'best', is not feasible in the real world.

The first mediator constraint *Changing travel cost throughout time* calls for expanding the in section 2.1 explained graph theory by allowing the graph to change throughout time. This will be done by the concept of the time expanded network in section 2.2 on page 9.

The second mediator constraint *travel time versus financial cost* calls for a quantification of both travel time and financial cost. Otherwise the two cannot be compared but philosophically. Hence section 2.3 will expand upon valuing travel time.

Then the constraints *modality* and *carpooling* remain. Both the *modality* choice and choice for *carpooling* are currently based on the experience, wit, and intuition of the planners. There are however, theoretical concepts that can be used to form a basis for heuristic methods. For example, it seems counterintuitive to assign students that live relatively far apart to carpool in the same car. Hence section 2.4 starting on page 13 will discuss theories that can be used to quantify the intuitive meaning of 'near' and 'far' to make them more tangible and quantifiable. Once it has been established how near or far two points are from each other, discrete modality choices can be made. section 2.5 starting on page 15 will elaborate on space-time related concepts that can be used to make these choices. And last the concept of clustering will be discussed in section 2.6 starting on page 18 as this can provide a basis for making carpooling choices.

Finally, section 2.7 starting on page 22 will draw a conclusion about the reviewed literature, thereby answering the first research question.

2.1 Travelling Salesman and Köningsberg

2.1.1 Köningsberg and graph theory

In two subsections the Köningsberg and travelling salesman problem will be discussed. These serve as a basis for explaining graph theory and exhaustive search respectively.

The problem of the seven bridges of Köningsberg is a mathematical problem set in former Prussia. Figure 2.1 shows the city of Köningsberg with its seven bridges. The problem to be solved is to plan a walking route that crosses each of the bridges exactly once. This is also called the Euler walk, because Euler (1736) has first investigated it. He found that this is not possible in the situation of figure 2.1. Euler concluded so by creating an abstract representation of the islands and the bridges, which can be seen in figure 2.2.

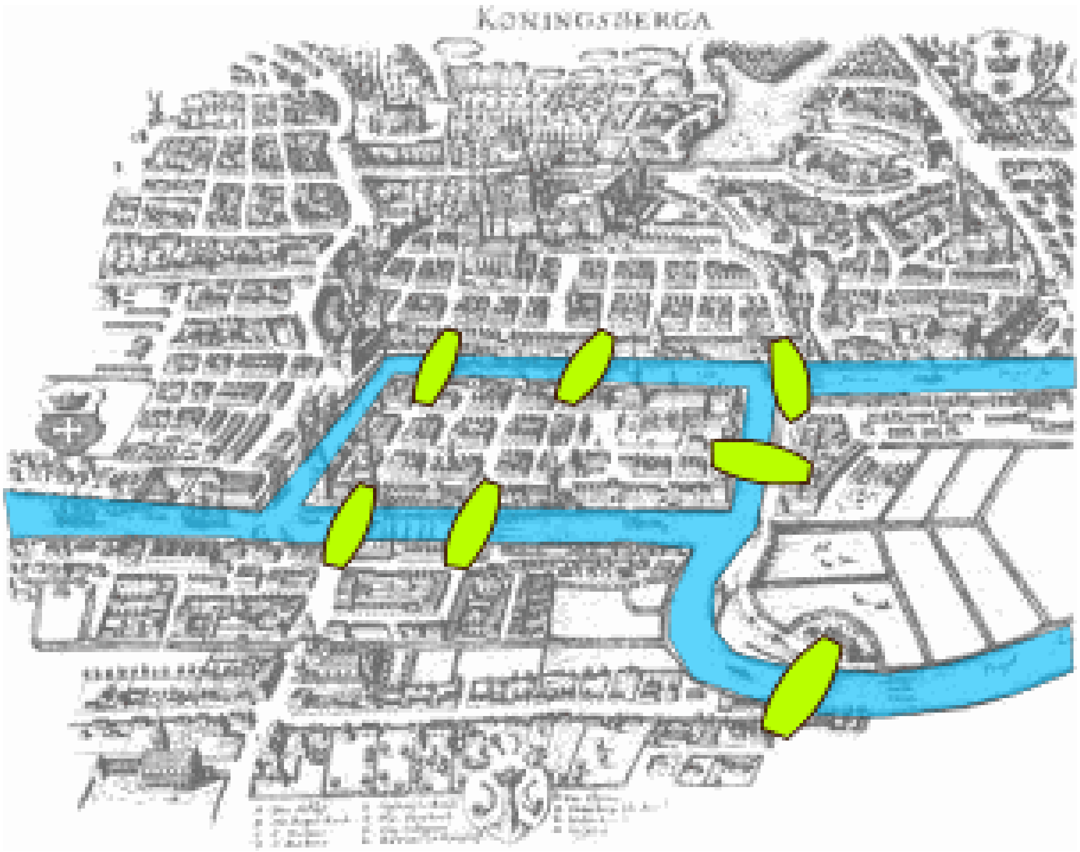


FIGURE 2.1: Map of Köningsberg in Euler's time showing the actual layout of the seven bridges (Giuşcă 2005).

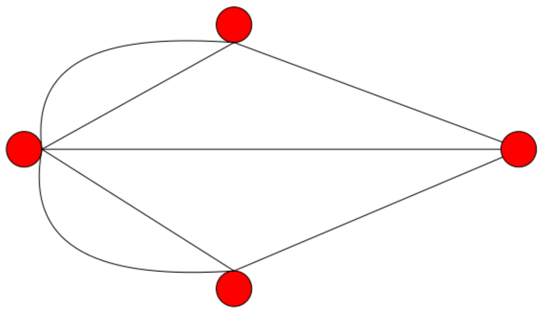


FIGURE 2.2: Köningsberg graph after Giuşcă (2005).

This representation is called a graph. In a more formal sense this is written as $G = \{V, E\}$, or a set of vertices and edges (H. Cormen T. et al. 2009, p. 589; Goodrich, Tamassia, and Goldwasser 2014, p. 789). The edges (black lines) represent the bridges that connect the islands which are the vertices (red dots). Two vertices are said to be adjacent when they are connected by an edge (T. H. Cormen et al. 2009; Goodrich, Tamassia, and Goldwasser 2014). For a road network it can be said that the edges represent the roads and that the network vertices where more than three edges meet represent the crossroads. Only when there are zero or two vertices that have an odd number of edges the Euler walk is possible. For further reading about graph theory Pinter (2010) provides a introductory starting point.

Shortest path finding on a network graph can be done by an algorithm. The Dijkstra 1959 and A* (Hart, Nilsson, and Raphael 1968) algorithms are two of the most prominent that can be used to find the shortest route on a graph. Algorithm 1 describes the Dijkstra algorithm.

Algorithm 1 Dijkstra's algorithm, after Dijkstra (1959)

visited = {empty set}
unvisited = {empty set}
candidate = {empty set}

Start
put all vertices V in *unvisted*
set all vertices to 'infinity'
Put starting vertex v_s in *visited*
$v_i = v_s$

Loop
(1) For all vertices v_j connected to v_i
evaluate the travel time cost tt to v_j

(2) If v_j is already in *candidates*
check if tt is lower than already stored cost.
Yes? Replace stored tt with tt from step 1
and continue to (4)

(3) Put v_j in *candidates* with $cost = tt$

(4) If travel time from v_i to all neighbouring
vertices is known, put v_i in *visited*

(5) Jump to candidate v_j with lowest cost
and repeat from (1)with $v_i = v_j$

The A* algorithm further builds on the Dijkstra algorithm by adding a measure of how close a vertex is to the final destination. Whereas the Dijkstra algorithm also searches in the 'wrong' direction, the A* does so less. Thus requiring less calculations, but also potentially missing out on an optimal solution. Dijkstra (1959), Hart, Nilsson, and Raphael (1968) provide starting points for further reading.

2.1.2 Travelling Salesman Problem and exhaustive search

The Travelling Salesman Problem, or TSP for short, was first introduced in a German manual (Rez 1832) and is formulated in the lines of *'Given a number of locations a salesman has to visit, what is the shortest path that visits all locations and starts and ends at the office?'*. This problem can be solved with the help of a graph. For more detailed information about the specific case of the TSP Bodin et al. (1983), Lawler et al. (2007), or Laporte (1992) can be read.

Exhaustive search is a method that can be used to solve the TSP. It entails trying all possible inputs and calculate the corresponding outputs. This is also called brute-forcing. Since all possibilities are, by definition, evaluated, the solution is guaranteed to be perfect. The problem of brute-forcing however is that it becomes time consuming when the amount of possible inputs increases. The time needed to solve a problem is called the time complexity. The time complexity of brute forcing is factorial, denoted as n!. This means that if the input is 10 (locations), the time needed to solve the problem scales with 10*9*...*2*1, also noted as 10!. Solving the TSP quickly becomes practically impossible on a human time scale. Thus heuristic solutions are needed to balance the time needed with solution quality.

2.2 Time dependable network graph

Time dependent network graphs differ from regular network graphs in that they change throughout time. Strictly speaking this means that both the adjacency and the traversal cost could change. Traversal cost are the cost that are associated with travelling over part of the network. The cost could be expressed as distance travelled, but also in time needed or the amount of shade along the route. In case of a road network this would mean that the time to travel between two points differs depending on the departure time. This is experienced daily when commuting during rush-hour; commuting usually takes longer than outside rush-hour. Time dependable adjacency is out of scope of this research because it is unlikely that the road network will frequently change within a single commuting trip. The concept of the **T**ime **E**xpanded **N**etwork (TEN) was mentioned by Ford Jr and Fulkerson (1958) as a way to cope with time dependent graphs. If $G = \{V, E\}$ is a graph, then $G^T = \{V^T, E^T\}$ is a time extended graph. Thus for any unique moment in time there is a unique graph. This can practically be done by storing the network multiple times, each copy for a fixed time window. Another possibility is to store the time dependable cost of an edge as a function of time.

2.3 Value of commuting time

This section is structured by first discussing **the value of time**. Second, the fact that **the value of time changes** is recognised. Then the concept **travel time ratio** is discussed.

The value of time is a question that comes with the travelling employees problem, as it is necessary to know how a faster route should be valued in comparison to slower route. A starting point is that the value of a travellers time is equal to the wage of that traveller (Beesley 1965). It also needs to be recognised that any persons time is finite and cannot be sold or bought to in-/decrease someones 'stockpile', thus making it a scarce good (Lugano et al. 2018). This provides an intuitive explanation on why a company CEO has his personal driver and lower wage employees do not. The underlying assumption is that when the CEO is travelling she/he is using her/his time unproductive. Since time is scarce this legitimises the hiring of a private driver. In recent years however, it has become more possible to use one's travel time productively due to the improved information technologies and the nature of knowledge work (Lyons and Urry 2005; DeSerpa et al. 1973; Hoinville and Berthoud 1970). This (partly) contradicts the classical viewpoint that travel time is useless time. The nature of the work in the student employment agencies is mostly not knowledge work and therefore the applicability of useful travel time in this context is limited. Therefore a more classical approach seems reasonable. More useful in the context of a classical (monetary) (cost-benefit) analysis is equation 2.1 after Hess, Bierlaire, and Polak (2005).

$$\beta_{tt}/\beta_{tc} \tag{2.1}$$

In which β_{tt} and β_{tc} are coefficients for respectively travel time and travel cost. In other words, it is a linear relation between travel time and cost, or just *'how much money are you willing to pay for a minute less travel time'* is assumed. Determining the values of these coefficients can be done, as proposed above, by taking a person's wage.

The value of time changes. Other authors have suggested that the the relationship between travel time and value is not linear, but a logit function with parameters that change depending on the total distance of the trip (Johansson, Johan Klaesson, and Olsson 2002; 2003). Further constraints are that the general willingness to commute is finite; commuting longer than 45 minutes is undesirable (Johansson, Julian Klaesson, and Olsson 2002). The ideal commuting time can be argued to be 16 minutes (Johansson, Julian Klaesson, and Olsson 2002). However, if the commuting takes place less often the willingness to commute longer increases (Johansson, Julian Klaesson, and Olsson 2002).

It is also true that one person does not always value travel time equally, but the value is dependent on factors such as weather (Beesley 1965), modality (Wachs 1976), and uncertainty due to traffic jams or delays (Fosgerau and Engelson 2011; Carrion and D. Levinson 2012). And rather than expressing the value in monetary terms, the value can be expressed subjectively. Travel time can be experienced longer or shorter depending on the state of mind of the traveller, e.g. when she/he falls asleep in the train (Watts 2008). Horowitz (1978; 1981) also investigated the subjectivity of travel time, but he recognises that a subjective measure is less useful in the context of a cost benefit analysis.

Another limitation of the linear relation between time and value is the assumption that the value of travel time is independent of the time that is spent at the destination. This assumption is false. Travel time and the duration of the activity at the destination are positively correlated (Hamed and Mannering 1993; Kitamura 1990; Kitamura, Chen, and Narayanan 1998; D. M. Levinson 1999). This also makes sense intuitively, as this would mean that a student is willing to travel further when the time spent at work is longer. Surprisingly, completely eliminating commuting is often not desirable as commuting also has positive effects such as socialising or shopping at/near the working destination, listening to music or making phone calls while travelling, and the intrinsic enjoyment of travelling itself (Redmond and Mokhtarian 2001).

Travel time ratio is a concept that provides an alternative to the limitations of the linear time value relation. The concept of travel-time ratio is formulated by Dijst and Vidakovic (2000) as in equation 2.2.

$$\tau = \frac{T_t}{T_t + T_a} \tag{2.2}$$

The travel time ratio τ is the ratio between the time travelled T_t and the time spent at an activity T_a plus the travel time T_t, thus accounting for the interdependence of travel time and activity time. The T_t is the total time spent travelling, so it includes both the forward and backward trip. In plain language this means that we are willing to travel longer if we stay at the destination longer. The ratio τ is between zero and one. A ratio of one means there is only travel and no activity, a ratio of 0.5 would mean the time travelled equals the time of the activity.

After studying the results of the Dutch National Travel Survey from 1998 Schwanen and Dijst (2002) found that the inverse polynomial from equation 2.3 best describes the relationship between work activity duration W (in hours) and travel time ratio.

$$\tau = \frac{1}{0.036W^3 - 0.635W^2 + 3.997W} \tag{2.3}$$

It should be said that Schwanen and Dijst (2002) mention that their model has an R^2 of 0.083, which means eight percent of the variation in their data is explained by the model. They argue this is caused by the high variation in travel time ratios. They do state however, that their model from equation 2.3 shows similarities with the mean travel time ratio from their data, see also table 2.1. Thereby they illustrate the validity of their model.

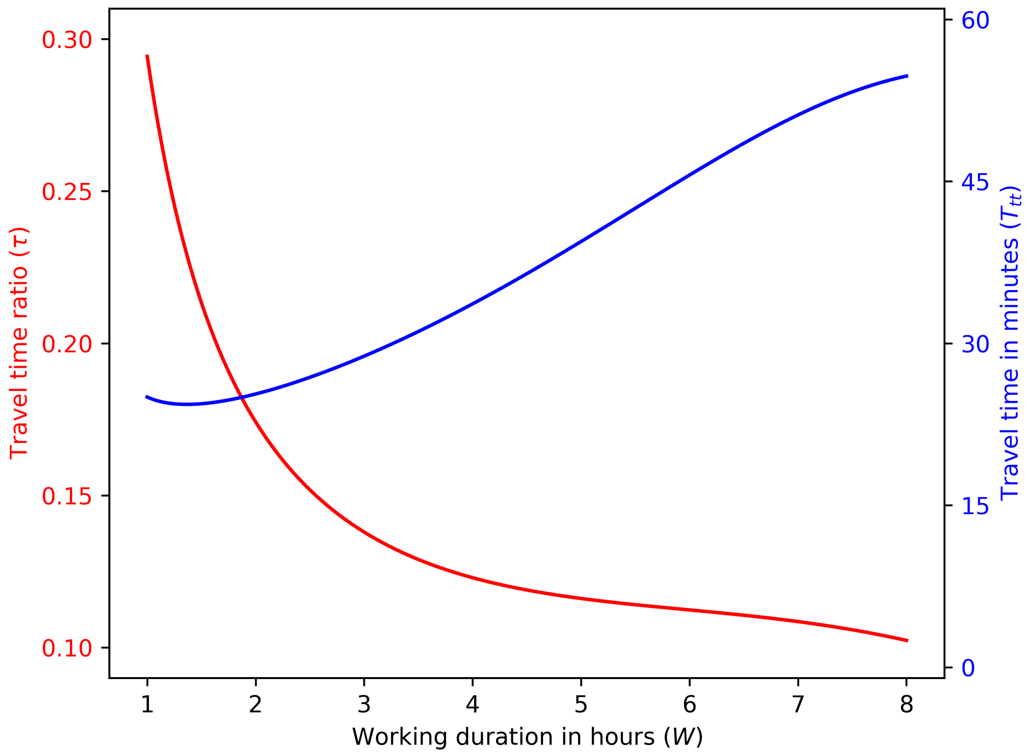


FIGURE 2.3: In red the relationship between the travel time ratio and the working duration according to Schwanen and Dijst (2002), in blue the derived relationship between travel time and working duration.

Figure 2.3 shows that the relative willingness to travel decreases as the amount of work increases, but the absolute travel time increases. Figure 2.3 is based on a nation wide data set representative of the entire working population, not students in particular. Schwanen and Dijst (2002) distinguish between households with a single part-time worker, arguably more similar to students, and other groups. Notably the mean and standard deviation of the travel time ratio are higher for single part-time worker households compared to the total average, meaning that this group is, on average, willing to travel further.

Travel time ratio	**Mean**	**Std.dev.**
Nation wide	0.105	0.078
One part-time worker	0.129	0.092

TABLE 2.1: Country wide travel time ratio and for part-time workers, from Schwanen and Dijst (2002), for the Netherlands

With this knowledge it can be argued that by adding the difference between the two groups $(0.129 - 0.105 = 0.024)$ to equation 2.3 results in a travel time ratio versus working duration regression that approximates the behaviour of students. This produces equation 2.4.

$$\tau = \frac{1}{0.036W^3 - 0.635W^2 + 3.997W} + 0.024 \tag{2.4}$$

If equation 2.2 and 2.4 are combined and it is assumed that a students travel time budget is at maximum equal to the travel time as would result from equation 2.4, equation 2.5 can be derived for a one way trip.

$$B_{tt} = \frac{1}{2} * \frac{-W^4 + 17.6389W^3 - 111.028W^2 - 1157.41W}{-40.6667W^3 + 717.315W^2 - 4515.13W + 1157.41} \tag{2.5}$$

With the knowledge of this subsection it can be concluded that students are willing to travel longer and thus farther for longer shifts. This relationship between travel time and travel distance will be more thoroughly discussed in the next subsection.

2.4 Closeness measures

Measuring how close one student lives another can be done by one of multiple measures. An intuitive measure for closeness is the straight line distance, also called the Euclidean, Pythagorean, or as-the-crow-flies distance (M.-M. Deza and E. Deza 2006), which can be defined as

$$d = \sqrt{(x^2 + y^2)} \tag{2.6}$$

where d is the distance between the two points, and x and y are the distance between the two points in the x and y dimension respectively (M.-M. Deza and E. Deza 2006). The collection of distances from each student to each other student can then be represented in an $N \times N$ matrix, with N being the amount of students. Other less intuitive distance measures exist, such as the flower-shop distance, which is the distance from point A to point B via the nearest flower shop (M. M. Deza and E. Deza 2009). More useful in the context of this research is the quickest path measure, also defined as *"It is the time needed for the quickest path (i.e, a path minimising the travel duration) between them when using [...] the network"* (M. M. Deza and E. Deza 2009, p. 328). Although these measures seem similar, their difference is visible in figure 2.4.

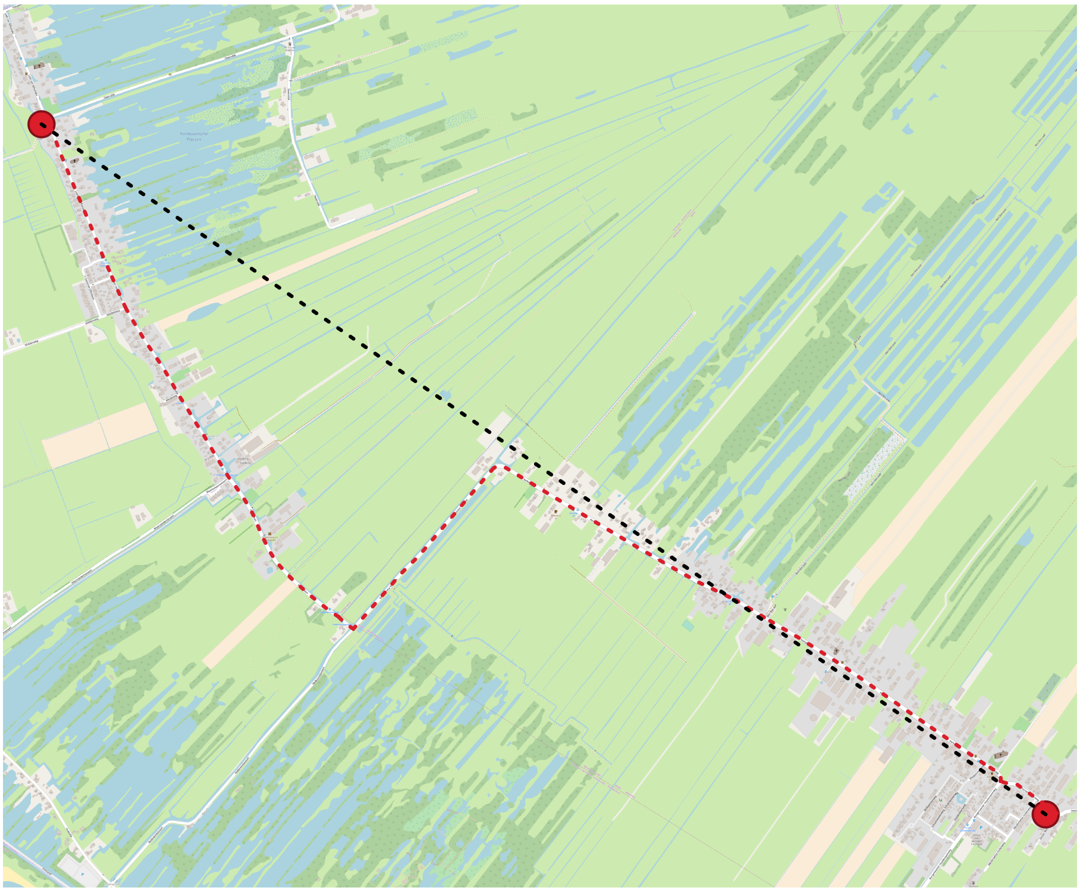


FIGURE 2.4: The Euclidean distance (black line) and quickest path distance (red line) between two points in Tienhoven, the Netherlands. The Euclidean distance is 3948 meter, while the quickest path distance is 57 minutes, based on a distance of 4793 meter and travel speed of 5 kilometre per hour.

Similar to the quickest path measure is the shortest path measure; instead of minimising travel time, travel distance is minimised. These two can be, but are not necessarily, equal. When walking the shortest path often is equal to the quickest path. When travelling by car the quickest path often is over the highways.

When the Euclidean and shortest path distance are almost equal, the shortest path is relatively straight. If they differ by a higher factor the shortest path is more non-straight. This concept is called straightness, defined as *"the efficiency between two nodes i and j is equal to the ratio of the Euclidean distance and the shortest distance between them"* (Rui and Ban 2014, p.1429).

How distance relates to time is studied in the field called time-geography. One of the main concepts in this field is the space time prism. This will be the subject of the next subsection.

2.5 Space time prism

This section discussed that **people are constrained by time and space.** After that it will discuss the time-geographic concept of **service area**.

People are constrained by time and space. If you wake up five minutes before you need to be at work 15 kilometres away, you will most likely not be on time because you simply cannot travel 15 kilometres in five minutes. A space time prism *"represents potential mobility in space with respect to time" (Miller 2017)*. In other words, it is a representation of what area can potentially be visited by someone given certain time and modality constraints. This concept was first developed by Hägerstraand (1970). Figure 2.5 shows a space time prism.

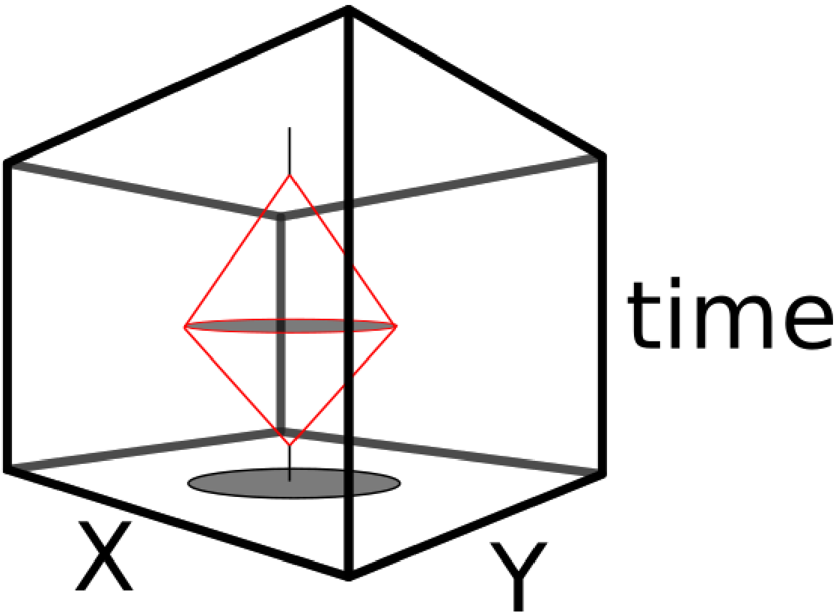


FIGURE 2.5: A space time prism in red.

The volume of the prism is called the **p**otential **p**ath **s**pace (PPS), and represents the all possible locations and times (Tong, Zhou, and Miller 2015). A 2D projection of the PPS is called the **p**otential **p**ath **a**rea (PPA) which can be used to determine whether or not visiting a certain point is possible given time and modality constraints (Tong, Zhou, and Miller 2015). Or as formulated by Patterson and Farber (2015) in their comprehensive and structured overview of the history of PPAs, *"PPAs refer to the spatial extent of where individuals can participate in activities subject to time and other (e.g. modal availability) constraints"*. Time constraints are often expressed as a time budget, meaning that a person has a finite budget of time that can be expended on travel and activities. So the time budget of a day would be 24 hours. Figure 2.6 shows the PPA for a given travel time budget T and two points i and j and the shortest path beteween them. Any point falling outside the PPA cannot be visited with the given time and modality constraints.

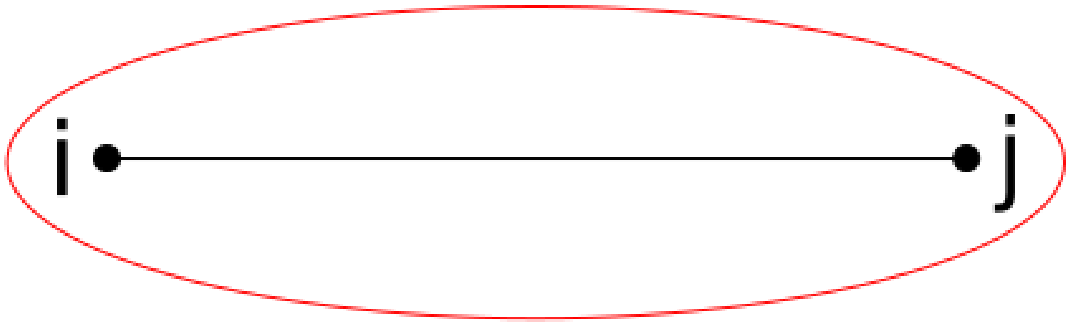


FIGURE 2.6: Schematic representation of the potential path area in red.

More formally it can be evaluated if any point k is inside the PPA when travelling from i to j for a given travel time budget T and a shortest path SP_{ij} by the following logic. The PPA is the set of points k such that the travel time from i to k plus the travel time form j to k is smaller than the travel time budget T. Or formally as in equation 2.7.

$$PPA = \{k | t_{ik} + t_{jk} < T\} \tag{2.7}$$

When the amount of points that are to be visited increases, the logic becomes as is shown in figure 2.7 and equation 2.8. The entire PPA is the sum of all coloured PPAs. They are calculated by taking the shortest path between all nodes except one, and calculating the PPA for that line segment. This is then repeated for each line segment. Note that the order in which the points are to be visited is known beforehand.

Research about maximising the PPA through determining the order of visiting the nodes seems to be scarce. It is out of scope of this research, but might provide valuable methods for carpooling and/or accessibility problems and should therefore be investigated.

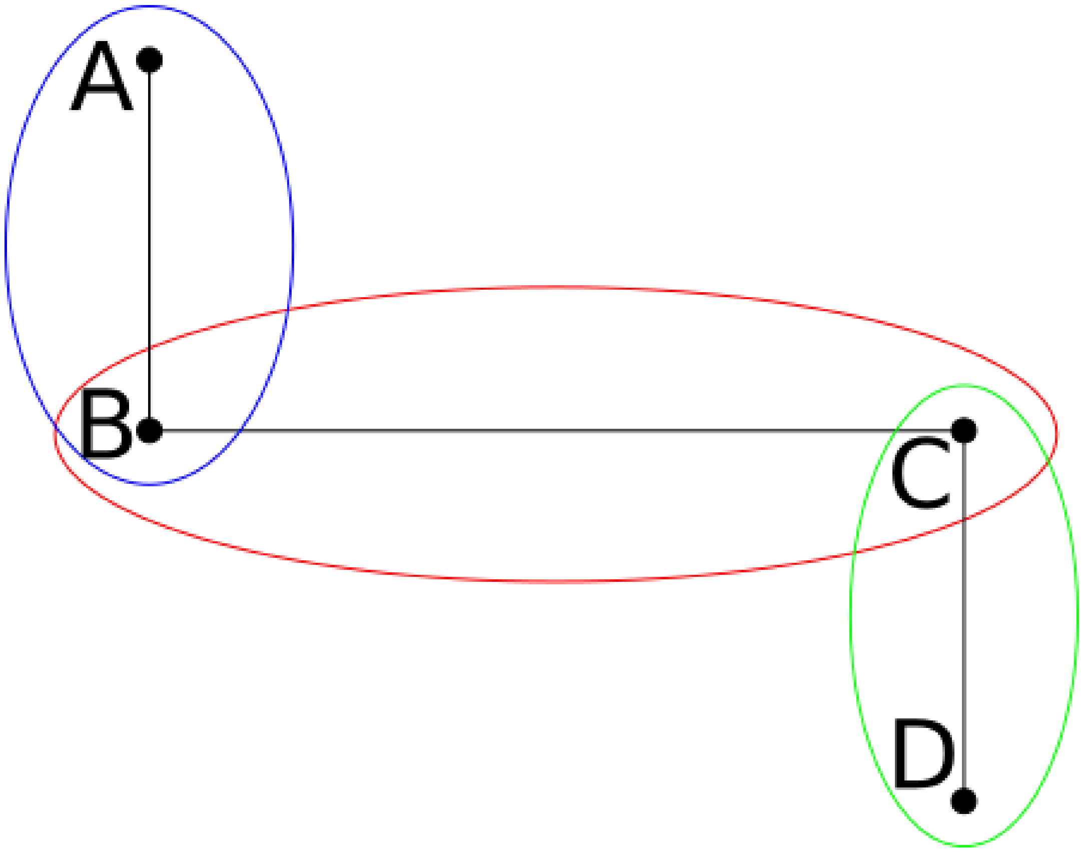


FIGURE 2.7: Schematic representation of the potential path area for multiple points. The colours match those of equation 2.8.

$$
\begin{aligned}
PPA = \\
&\{k|t_{Ak} + t_{Ak} < T - SP_{BC} - SP_{CD}\} \cup \\
&\{k|t_{Bk} + t_{Ck} < T - SP_{AB} - SP_{CD}\} \cup \\
&\{k|t_{Ck} + t_{Dk} < T - SP_{AB} - SP_{BC}\}
\end{aligned}
\tag{2.8}
$$

Service area is a concept related to the PPA. Formally *"The service area [...] is a subset of the [...] network"* (Huisman and By 2009, p. 421). Say that a maximum travel time of 30 minutes to a vertex is given. In that case the **s**ervice **a**rea (SA) of 30 minutes is that area from which it is possible to reach the vertex within 30 minutes. This area is calculated in all directions over the network from the vertex for 30 minutes. When travelling over the highway the covered distance will be larger compared to travelling through a busy city centre. Hence the service area is not necessarily a circle around the work location, but depends on the properties of the network. This is more clearly illustrated in figure 2.8.

The equations from the PPA can more or less be translated directly to fit the SA. Any point k is within the service area of an origin o when the travel time between o and k is smaller than the travel time budget T.

$$SA = \{k | t_{ok} < T\} \tag{2.9}$$

FIGURE 2.8: An example of a service area (ESRI 2014).

2.6 Clustering

Clustering is a term that, at least intuitively, describes a spatial concentration of 'something'. More formally we have to distinguish between clustering and kernel density. Cluster is a term used to describe "clusters of similar or dissimilar neighbouring values for a given attribute" (Chun and Griffith 2013, p. 92), also called hot and cold spots. Kernel density has to do with a concentration of points.

Kernel density estimation is a method that is used to estimate the density of kernels, or points, by plotting a smooth surface over each point. This can be seen in figure 2.9. This could be used to find high concentrations of students among a group of available students. The search radius, or bandwidth, can be varied. This in turn results in a differentiated map as can be seen in figure 2.10. Research into how the bandwidth should be determined is plenty, see for example Sheather and Jones (1991), Turlach (1993), Hall et al. (1991), or Jones, Marron, and Sheather (1996).

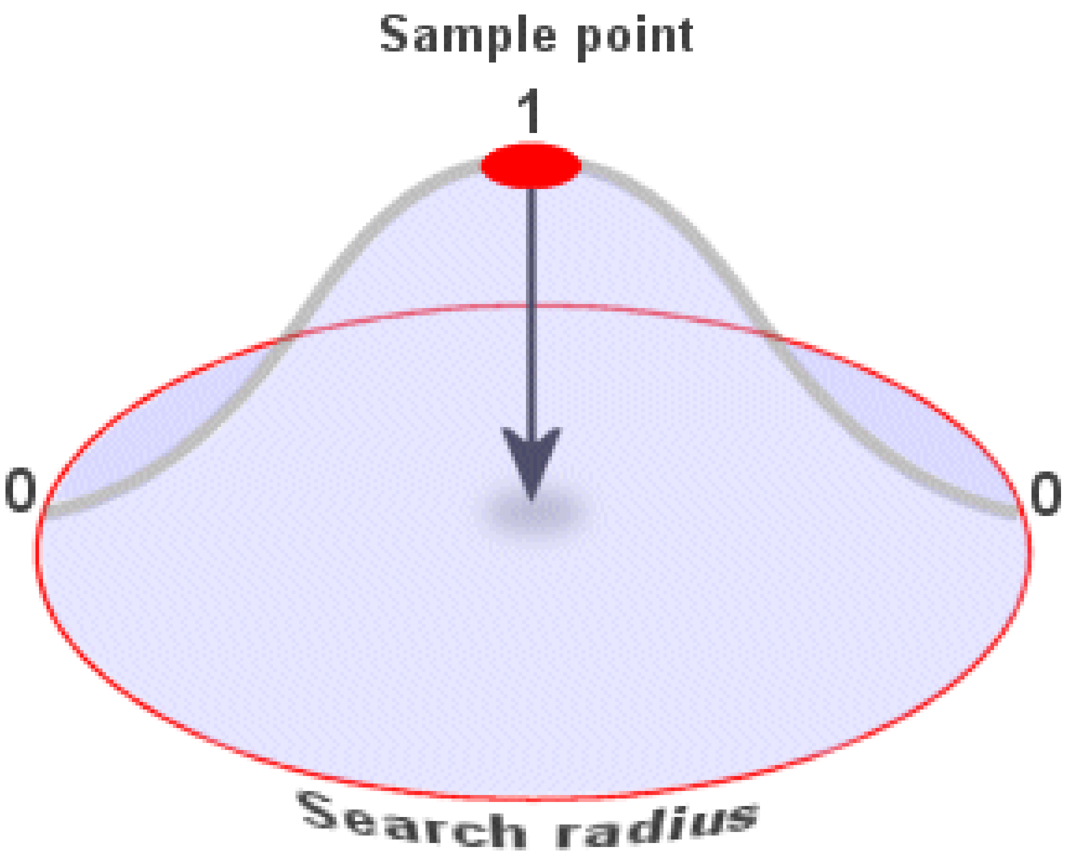


FIGURE 2.9: Kernel density calculation method (*Kernel density calculations* 2015)

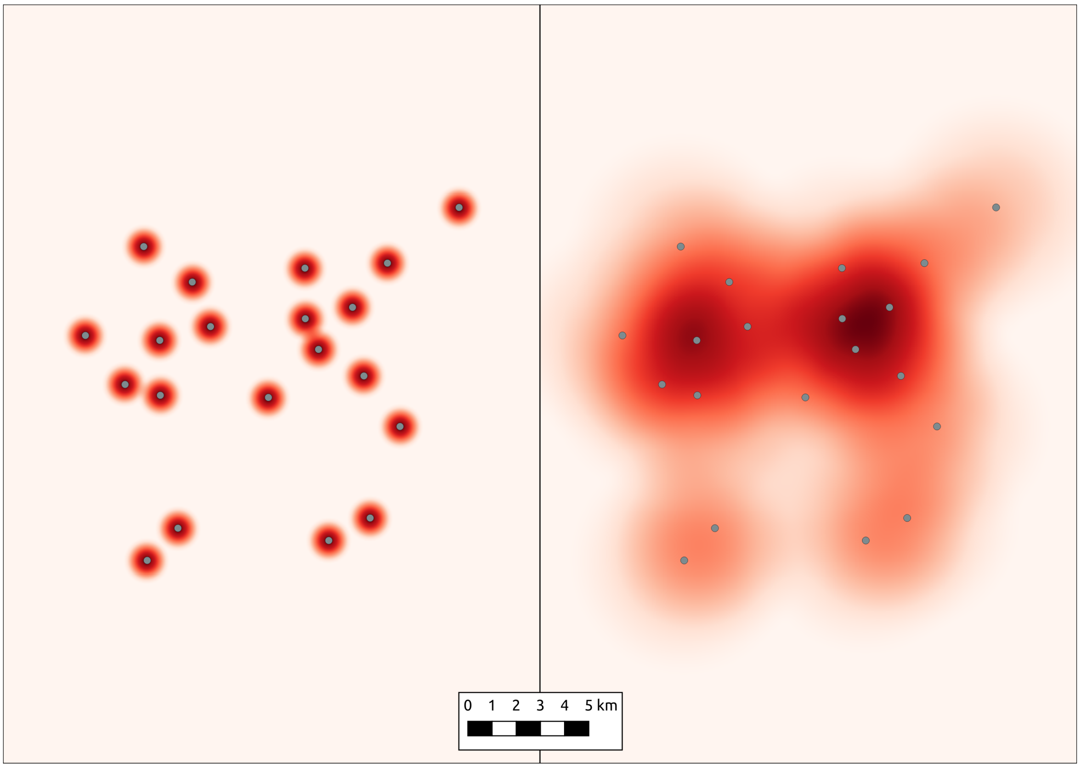


FIGURE 2.10: Comparison of two kernel density estimations with a search radius of one kilometre (left) and five kilometres (right). Both are based on the same location data (dots).

Nearest neighbours is a method that is also related to clusters. In our case can be used to determine what students should be in which car. For filling a car with four students and a driver simply the four students that live nearest to the driver could be selected. Note that any of the previously discussed (section 2.4) distance metrics can be used to determine the four nearest students. The other way around is also possible. A student gets assigned to the car that is nearest to him/her. This can be seen in figure 2.11.

FIGURE 2.11: Student is assigned to car 2 because it is closer than car 1.

K-means builds further on the principle nearest neighbour. MacKay (2003) define the k-means algorithm as *"The K-means algorithm is an algorithm for putting N data points [...] into K clusters. Each cluster is parameterized by a vector* $\mathbf{m}^k$ *called its mean"* (MacKay 2003, p. 285). In other words, it is an algorithm to divide a total of N data points into a total of K groups, or clusters, based on their position in space. The algorithm consists of three steps. These are explained in algorithm 2 and figure 2.12.

Algorithm 2 K-means after MacKay (2003)

Initialization step:
Set K means $\{\mathbf{m}^{(k)}\}$ to random values

Assignment step:
Each data point n is assigned to the nearest mean

Update step:
The K means $\{\mathbf{m}^{(k)}\}$ are adjusted to the mean of all data points n_m that are assigned to that particular mean $\mathbf{m}$

Repeat the assignment and update step until no data points switch between means

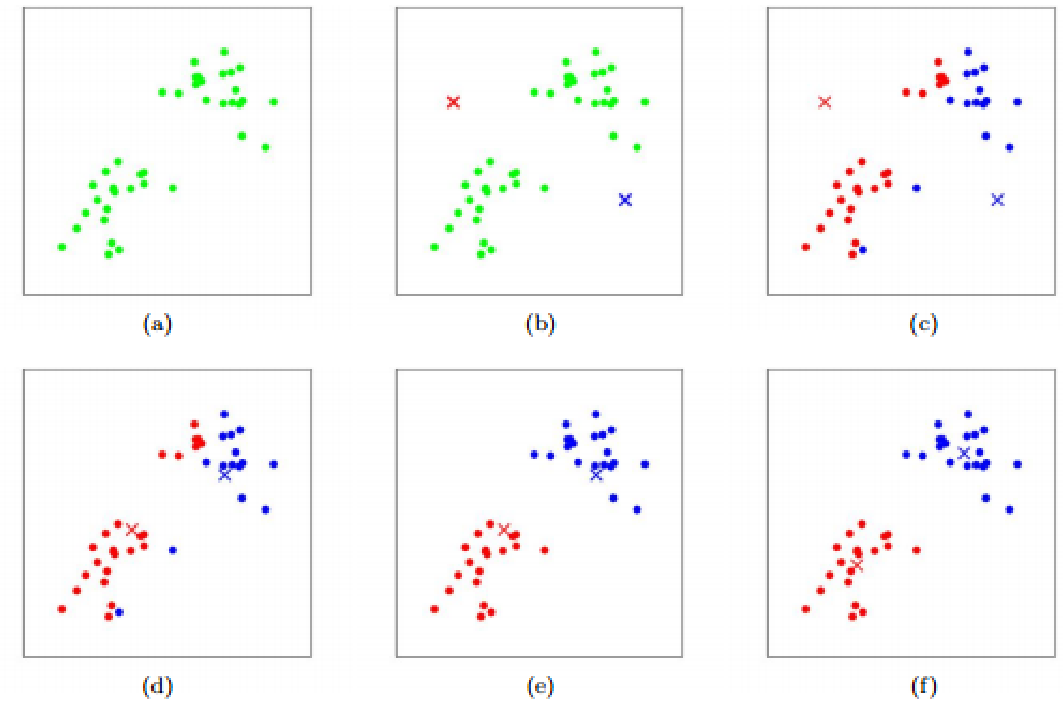


FIGURE 2.12: Adopted from Piech (2012). (a) All data points. (b) $k = 2$ random means in red and blue are created. (c) Each data point is assigned to, or coloured, to its nearest mean. (d) The means are recalculated as the average of all data points in that cluster. (e) The colouring is done again. (f) The means are recalculated.

Initialisation of the means in the above example has been done at random. For other methods of initialisation and further reading see Hamerly and Elkan (2002).

2.7 Conclusion

This chapter aimed to answer the research question *What concepts could be used to solve the travelling employees problem?*. In the previous sections a brief history of the TSP and Köningsberg, time dependable network graphs, the value of commuting time, closeness measures, space time prisms and clustering were given. Along the way a gap in literature was identified; research about *PPA* maximisation through changing the order of trip waypoints seems to be scarce or non-existent.

It was found that *changing travel cost throughout time* can be handled by the concept of a TEN. Balancing *travel time versus financial cost* can be done by quantifying commuting time and relating it to the duration of the work via the travel time ratio. *Modality* choices can be made by using discrete spatial decision boundaries such as the service area or space time prism. And finally the process of choosing whether or not *carpooling* is efficient in a given spatial arrangement can be supported by clustering methods such as the kernel density estimation or k-means.

3 Model formulation

In this chapter it will be explained how the second research question *"How can these concepts be used to form a model that solves the travelling employees problem?"* can be answered by taking the concepts of the previous chapter and aligning them to form a model.

The model should solve the problem that has been stated in section 1.2: *"First, they should assign the students to the job that are in the most favourable position to travel. Second, they should plan this single trip as efficiently as possible"*. These two problems will be split in the model in section 3.1 Selecting the students and 3.2 Determine the most efficient way of travel accordingly. Section 3.4 will conclude this chapter by summing up the findings.

3.1 Selecting the students

Selecting a subset Assigning the students that are in the most favourable position to travel entails that a selection be made from among all the available students. In other words, a subset is selected. But what then is a favourable position to travel? Intuitively this comes down to selecting students that live 'near' over students that live 'far' from the working location in terms of travel time. But how near or far? Travelling 1.5 hours for a shift of 30 minutes will be judged as 'too far' by most. On the other end is an 8 hour shift at 5 minutes walking from your home location.

Willingness to travel Given that we know how long a shift is, how long can we expect a student to travel? Or, what is the travel time budget of a student? The in subsection 2.3 Value of commuting time formulated equation 2.4 ($\tau = \frac{1}{0.036W^3 - 0.635W^2 + 3.997W} + 0.024$) provides a relationship between the duration of an activity and the travel time (ratio). In our case the activity duration, or shift length, can be assumed to be known beforehand by the employment agency This also means that for each shift a travel time ratio, and thus travel time, can be calculated via equation 2.5 ($B_{tt} = \frac{1}{2} * \frac{-W^4 + 17.6389W^3 - 111.028W^2 - 1157.41W}{-40.6667W^3 + 717.315W^2 - 4515.13W + 1157.41}$).

Selecting students based on the B_{tt} This travel time budget B_{tt} is dependent on shift length and hence equal for each student for a given shift. To see whether it is possible for a student to reach the working location without expanding the travel time budget B_{tt} at least two algorithmic approaches can be used.

- Calculate the SA based on the B_{tt} and select students inside the SA
- For each student calculate the actual travel time and compare to see if ht is lower than the B_{tt}

The first approach entails an SA calculation followed by a spatial query. The second involves a single travel time calculation per student. Depending on the amount of available students one method might be preferred over the other. In general it can be assumed that a per student travel time calculation is more complex, and thus slower, than a per student spatial query. On the other hand, calculating an SA might be slower than a single, or a few, route calculations. Therefore it can be assumed that the first method is perferred for larger groups of students, whereas the second method is preffered for smaller groups of students. The exact size of 'larger' and 'smaller' is to be determined empirically and will depend on software and hardware.

A disadvantage of the SA based algorithm is that a single SA is used, thus representing a single point in time. However, the SA changes as traffic conditions change throughout time. Different students should start their commute at different points in time, e.g. student 1 leaves at 08:30 and student 2 leaves at 08:45. Both times have different SA's, so the SA at what point time should be used to select the students?

Because of this limitation and the fact that open source API's are more readily available for routing than for SA calculations the choice is made to use the per person travel time calculation method in the model.

More formally,

- The set of students $\mathbf{st}$ that are selected from all available students $\mathbf{st_a}$
- consist of the individual students $\mathbf{p}$ such that
- their work bound travel time $\mathbf{T_{t,wb,p}}$ between the home location $\mathbf{HL_p}$ and work location $\mathbf{WL}$ is smaller than the travel time budget $\mathbf{B_{tt}}$. The travel time $\mathbf{T_{t,wb,p}}$ is a function of
  - modality $\mathbf{m}$
  - the work location $\mathbf{WL}$
  - the home location $\mathbf{HL_p}$
  - the traffic conditions, i.e. the state of the $\mathbf{TEN}$ which is a function of
    * for the work bound trip: the time the shift starts $\mathbf{t_{start}}$
    * for the home bound trip: the time the shift ends $\mathbf{t_{end}}$

So the entire selection logic then becomes as in equation 3.1

$$\begin{aligned}
st = \{p| \\
&T_{t,wb,p}(m, WL, B_{TT}, G^t(t_{start})) <= B_{tt} \cap \\
&T_{t,hb,p}(m, WL, B_{TT}, G^t(t_{end})) <= B_{tt} \cap \\
&p \in st_a\}
\end{aligned} \tag{3.1}$$

This logic is represented in algorithm 3. On line three the travel time budget is calculated based on the travel time ratio as presented in equation 2.5 on page 13. On line six and seven the travel time for both the work bound and home bound trip is calculated. Then in line ten through twelve it is checked whether this travel time is shorter than the travel time budget. This information is then stored as an attribute of the student on lines eleven and thirteen. If either, or both, of the trips are not possible the student is not selected. This is (not) done in lines fourteen through seventeen. After all, that student does not have the possibility to complete the round trip. If there are more students available and able to travel than needed, the closest, based on Euclidean distance, are selected. The reason that the Euclidean distance is used is because it is assumed that this is a proxy for the travel time of each modality. The algorithm then returns the set of selected students.

Algorithm 3 Selecting students based on the $T_{t,wb/hb,p}$

1: **function** SELECT SUBSET OF STUDENTS($st_a, L_{shift}, t_{start}, G^t, WL, M$)
2: $\quad st = \{\}$ ▷ Start with empty selection
3: $\quad B_{tt} = \frac{1}{2} * \frac{-W^4+17.6389W^3-111.028W^2-1157.41W}{-40.6667W^3+717.315W^2-4515.13W+1157.41}$ ▷ W in hours
4: $\quad$ **for** *each modality* $m \in M$ **do**
5: $\qquad$ **for** *Each student* $p \in st_a$ **do**
6: $\qquad\quad T_{t,wb,p}(m, WL, B_{TT}, G^t(t_{start}))$
7: $\qquad\quad T_{t,hb,p}(m, WL, B_{TT}, G^t(t_{end}))$
8: $\quad$ **for** *each student* $p \in st_a$ **do**
9: $\qquad$ **for** *each modality* $m \in M$ **do**
10: $\qquad\quad$ **if** $T_{t,wb,p} <= B_{tt}$ **then** ▷ Can p reach WL with m?
11: $\qquad\qquad$ *add m to possible work bound modalities* $M_{p,wb}$
12: $\qquad\quad$ **if** $T_{t,hb,p} <= B_{tt}$ **then**
13: $\qquad\qquad$ *add m to possible home bound modalities* $M_{p,hb}$
14: $\qquad$ **if** $M_{p,wb} \not\eqsim \emptyset$ & $M_{p,wb} \not\eqsim \emptyset$ **then**
15: $\qquad\quad$ *Add p to st*
16: $\qquad$ **else**
17: $\qquad\quad$ *Do not select student*
18: $\quad$ **if** $|st| > n$ **then** ▷ n is the number of students required.
19: $\qquad$ remove students that have the largest Euclidean distance from the work location so that $|st| = n$

$\quad$ **return** st

Car minimum travel time threshold δ Considering that a car is faster than walking, even on short distances, it will likely be the case that for most students the travel time by car $T_{t,wb/hb,car}$ is smaller than the travel time budget B_{tt}. This is not desirable as initiating transport by car comes with fixed cost for the employment agency that has to organise the transport and ensure that the car keys are at the right place (student) at the right time. To deal with this the car should not be used for trips shorter than a given threshold δ, e.g. 10 minutes. This threshold can be dealt with by checking if the travel time $T_{t,wb/hb,car}$ is smaller than the travel time budget B_{tt} and larger than the threshold δ

Strengths and weaknesses can be identified now that algorithm 3 is formulated.

First the travel time budget B_{tt} is used as a measure for whether or not a student is willing to travel a certain distance. Since the duration of the shift is included in the calculation of B_{tt} its real world applicability is increased. However, the empirical study of Schwanen and Dijst 2002 the calculation is based on a model with an R^2 of 0.083. Therefore further, more recent, empirical testing could improve the model.

The travel time budget B_{tt} is used as a binary border in selecting students. A student is in or out, even if it is just by one second. Even if this student might greatly improve the carpooling possibilities for the whole group.

The travel time that is tested against the travel time budget is calculated on a time dependent network. Therefore real world applicability is improved. The practical limits of rush hour or public transport not available in the middle of the night are dealt with by the model.

The time complexity of algorithm 3 is dependent on the chosen method for calculating the travel time of any given modality and can be said to equal $n * m * 2*$ times the time complexity of the route finding algorithm, with n being the amount of students, m being for each modality that is included in the model scope, and 2 being for one home and one work bound trip.

3.2 Determine the most efficient way of travel

Possible modalities Now that the students that are going to work have been selected based on whether they can reach the work location without expending their B_{tt}, it has to be determined what modality each student should take. There are four modalities. In other words, the set of possible modalities M contains the elements m, which are $walk$, $cycle$, public transport pt, and car, as in equation 3.2

$$M = \{walk, cycle, pt, car\} \tag{3.2}$$

So that means that theoretically four modalities might be used. It is likely that if you can reach a location by foot within the time budget, you can also reach this location by bike. It might also be possible to reach a location both by bike and car, but not by public transport because of a bad connection. There are four modalities, so theoretically there are $4^2 = 16$ combinations of possible modalities for a one way trip. Each of the possibilities is explained in table 3.1.

one element:	
$\{walk\}$	Possible when e.g. a footbridge across a canal forces cyclists, public transport, and cars to take a longer route via another bridge.
$\{cycle\}$	Possible when walking is too far, but public transport and car take too long. This might be in a congested city with poor public transport.
$\{pt\}$	Possible when location is too far away to walk or cycle, and car connectivity is poor due to e.g. rush hour, but public transport connectivity is good.
$\{car\}$	Possible when in rural areas with distances too long to walk or cycle, but with poor public transport connectivity.
two elements:	
$\{walk, cycle\}$	Possible for city areas with congestion and poor public transport.
$\{walk, pt\}$	Possible in congested city with a large non-cycle area between origin and destination.
$\{cycle, pt\}$	Possible in city with congestion, good public transport, and a distance too far to walk.
$\{walk, car\}$	It is possible when there is poor public transport, and a large non-cycling area which takes to long to cycle around. With the faster car however, you are able to reach the destination within the travel time budget.
$\{cycle, car\}$	In rural areas with poor public transport but a distance too large to walk.
$\{pt, car\}$	Long distances too far to cycle or walk.
three elements:	
$\{walk, cycle, pt\}$	Distance that can be walked or cycled, public transport is well organised, but the car is congested.
$\{walk, cycle, car\}$	Rural areas with poor public transport but a distance that can be walked or cycled, or crossed by car.
$\{walk, pt, car\}$	Possible when a non-cycling area is between origin and destination that can be circumvented by public transport or by car.
$\{cycle, pt, car\}$	Distances that are too far to walk but well reachable by bike, car, and public transport.
four elements:	
$\{walk, cycle, pt, car\}$	Origin and destination are near and have good connection for all modalities.

TABLE 3.1: Each home location HL_p can be located in the SA of a certain modality m. This table shows all possible combinations.

Thus there are fifteen combinations of possible modalities; the sixteenth possibility is the empty set. These possibilities are for a one way trip. Let us call the possible modalities for the work bound trip M_{wb}, which is a subset of (or equal to) M from equation 3.2 on page 26. A similar convention will be adhered to for the home bound trip M_{hb}.

Since we also have to travel back home, this trip also knows 15 possibilities. So theoretically there are fifteen possibilities for the work bound, and fifteen possibilities for the home bound trip; the total amount of possibilities therefore equals $15 * 15 = 225$.

From these possibilities it should be decided what modality a student should take. When choosing a modality it has to be respected that it is generally not desirable to leave bikes and cars at the work location, and that bikes and cars cannot be created out of thin air at the work location. In other words, if a student cycles to work, he/she has to cycle back home on the same bike. If a car is taken to work, it also has to be taken back home. Thus the only modalities allowing for asymmetry within a single round trip are public transport and walking. Hence the round trip can be completed in one of six ways:

1. Both work and home bound trip are done by walking
2. Both work and home bound trip are done by cycling
3. Both work and home bound trip are done by public transport
4. Both work and home bound trip are done by car
5. The work bound trip is done by walking and the home bound trip is done by public transport
6. The work bound trip is done by public transport and the home bound trip is done by walking

In some cases multiple modalities for the round trip will be possible. In that case, it has to be evaluated which of the available modalities is quickest for the student. Table 3.2 on page 30 shows for all 255 combinations of modalities that can be used. Let us call the contents of each individual cell the modalities that are possible for the round trip, or M_{rt}.

Home Bound \ Work Bound	Walk	Cycle	PT	Car	Walk, Cycle	Walk, PT	Cycle, PT	Walk, Car	Cycle, Car	PT, Car	Walk, Cycle, PT	Walk, Cycle, Car	Walk, PT, Car	Cycle, PT, Car	Walk, Cycle, PT, Car
Walk	walk	x	PT > Walk	x	Walk	(Walk), (PT > Walk)	PT > Walk	Walk	x	PT > Walk	Walk	Walk	Walk	x	Walk
Cycle	x	Cycle	x	x	Cycle	x	Cycle	x	Cycle	x	Cycle	Cycle	x	Cycle	Cycle
PT	Walk > PT	x	PT	x	Walk, PT	(Walk > PT), (PT)	PT	Walk > PT	x	PT	PT	Walk > PT	(PT), (Walk > PT)	PT	PT
Car	x	x	x	Car	x	x	x	Car	Car	Car	x	Car	Car	Car	Car
Walk, Cycle	Walk	Cycle	PT > Walk	x	(Walk), (Cycle)	(Walk), (PT > Walk)	PT > Walk	walk	Cycle	PT > Walk	(Walk), (Cycle)	(Walk), (Cycle)	(Walk), (PT > Walk)	(Cycle), (PT > Cycle)	(Walk), (Cycle), (PT > Walk)
Walk, PT	(walk), (Walk > PT)	x	(PT), (PT > Walk)	x	(Walk), (Walk > PT)	(Walk), (PT), (Walk > PT), (PT > Walk)	(PT > Walk), (PT)	(walk), (walk > PT)	x	PT	(Walk), (PT), (Walk > PT), (PT > Walk)	(Walk), (Walk > PT)	(Walk), (PT), (Walk > PT), (PT > Walk)	(PT), (PT > Walk)	(Walk), (PT), (Walk > PT), (PT > Walk)
Cycle, PT	Walk > PT	Cycle	PT	x	(Walk > PT), Cycle	(Walk > PT), (PT)	(Cycle), (PT)	Walk > PT	Cycle	PT	(Cycle), (PT)	(Cycle), (Walk > PT)	(PT), (Walk > PT)	(Cycle), (PT)	(Cycle), (PT), (Walk > PT)
Walk, Car	Walk	x	PT > Walk	Car	Walk	(Walk), (PT > walk)	PT > Walk	(walk), (car)	Car	(PT > Walk), (Car)	(walk), (PT > Walk)	(Walk), (Car)	(Walk), (Car), (PT > Walk)	(Car), (PT > Car)	(Walk), (Car), (PT > Walk)
Cycle, Car	x	Cycle	x	Car	Cycle	x	Cycle	car	(cycle), (car)	Car	Cycle	(Cycle), (Car)	Car	(Cycle), (Car)	(Cycle), (Car)
PT, Car	Walk > PT	x	PT	Car	Walk > PT	(Walk > PT), (PT)	PT	(walk > PT), (car)	Car	(PT), (Car)	(PT), (Walk > PT)	(Car), (Walk > PT)	(PT), (Car), (Walk > PT)	(PT), (Car)	(PT), (Car), (Walk > PT)
Walk, Cycle, PT	Walk	Cycle	PT > Walk	x	(Walk), (Cycle), (Walk > PT)	(Walk), (Walk > PT), (PT)	(Cycle), (PT)	(walk), (walk > PT)	Cycle	(PT), (PT > Walk)	(Walk), (Cycle), (PT), (walk > PT), (PT > Walk)	(Walk), (Cycle), (Walk > PT)	(Walk), (PT), (Walk > PT), (Pt > Walk)	(Cycle), (PT), (PT > Walk)	(Walk), (Cycle), (PT), (Walk > PT), (PT > Walk)
Walk, Cycle, Car	Walk	Cycle	PT > Walk	Car	(Walk), (Cycle), (Walk > PT)	(Walk), (PT > Walk)	(Cycle), (PT > Walk)	(walk), (car)	(cycle), (car)	(Car), (PT > Walk)	(Walk), (Cycle), (PT > Walk)	(walk), (cycle), (Car)	(Walk), (Car), (PT > Walk)	(Cycle), (Car), (PT > Walk)	(Walk), (Cycle), (Car), (PT > Walk)
Walk, PT, Car	Walk	x	PT > Walk	Car	(Walk), (Walk > PT)	(Walk), (PT), (Walk > PT), (PT > Walk)	(PT), (PT > Walk)	(walk), (Walk > PT), (Car)	Car	(PT), (Car)	(Walk), (PT), (Walk > PT)	(Walk), (Car), (walk > PT)	(Walk), (PT), (Car), (Walk > PT), (PT > Walk)	(PT), (Car), (PT > Walk)	(Walk), (PT), (Car), (Walk > PT), (PT > Walk)
Cycle, PT, Car	x	Cycle	PT	Car	(Walk > PT), (Cycle)	(Walk > PT), (PT)	(Cycle), (PT)	(walk > PT), (car)	(cycle), (car)	(PT), (Car)	(Cycle), (PT), (Walk > PT)	(Cycle), (Car), (Walk > PT)	(PT), (Car), (Walk > PT)	(Cycle), (PT), (Car)	(Cycle), (PT), (Car), (Walk > PT)
Walk, Cycle, PT, Car	Walk	Cycle	PT > Walk	Car	(Walk), (Walk > PT), (Cycle)	(Walk), (PT), (Walk > PT), (PT > Walk)	(Cycle), (PT)	(walk), (car), (walk > PT)	(cycle), (car)	(PT), (Car)	(Walk), (Cycle), (PT), (walk > PT), (PT > Walk)	(Walk), (Cycle), (Car), (Walk > PT)	(Walk), (PT), (Car), (Walk > PT), (PT > Walk)	(Cycle), (PT), (Car), (PT > Walk)	(Cycle), (PT), (Car), (Walk > PT), (PT >

TABLE 3.2: Modality choice based on the modalities available for the work (per column) and the home bound trip (per row). When a cell is red, no round trip can be made. Green cells have one modality that can satisfy both the work and home bound trip. Yellow cells have multiple modalities can be used to complete the round trip. Each option that is to be evaluated is given in brackets, different options separated by comma's. If the modality is different for the work and home bound trip this is denoted with "work bound modality > home bound modality".

There is a distinction to be made between when a student must use, or can use a modality. Respectively green and yellow in table 3.2. In the case of car travel this is relevant because this limits the possible combinations that have to be explored for students travelling together by car. Say that three students may not, three students must, and two students may use the car. That can be written as

- {0,0,0,1,1,1,2,2}
  - 0 = a student may not use the car, but may use other modalities
  - 1 = a student must use the car, it is the only modality available to the student
  - 2 = a student may use the car, but may also use other modalities

This is an attribute of a student, just as his/her home location. So if st is the set of students p, each p now has several attributes:

- HL_p: the home location of the student
- DL_p: whether or not the student has a drivers licence
- CR_p: whether a student may not, must, or may use the car

With this information algorithm 4, on page 33, can be formulated. The algorithm starts by creating a pool of students on line two. This pool contains the students that require a route to be planned for them. Once it is known what modality and route a student will take, the student is removed from the pool.

For any student that does not have the possibility to travel by car, formally $car \notin M_{p,rt}$ the planning is relatively straight forward because the non-car modalities are not interdependent on the modalities of other students. For these students the routes per modality are calculated and the fastest is chosen. This is done in lines two through seven.

In line nine through twelve it is checked whether there are any students left in the pool. If there students left, the algorithm should continue, otherwise it is done.

On line fourteen the loop is started that empties the rest of the pool. The variable i on line fifteen keeps track of the amount of loops that have been performed. On line sixteen the amount of cars needed if all students that could or must, would take the car gets calculated by dividing the size of the pool $|pool|$ by the capacity per car. Thereby assuming a homogeneous fleet of cars.

On line seventeen the k-means clustering is performed as explained in section 2.6. Then on line eighteen through twenty-six it determined if a cluster has a driver. If it has, the shortest route for that cluster is calculated. It is assumed that the car departs from the cluster centroid, and that all students cycle from their home to the cluster centroid. The total time for the car trip, including the cycling to/from the centroid, is than compared to their other modalities.

If the other modality is faster the student is assigned that modality and he/she is removed from the pool. If for each student in a cluster the car is the fastest option, all members are assigned the car and the entire cluster is removed from the pool.

Lines twenty-eight through thirty-one determine if the loop should be run again. First, if the pool is empty the algorithm is done. Second, if any students have been removed form the pool since the last iteration a new iteration is started.

Finally, line thirty-two is initialised if no students were removed from the pool since the last iteration. In that case all remaining students are assigned modality available to them but the car. If they can exclusively travel by car they cannot be planned by the model. Thus the algorithm leaves room for situations in which an optimal route for a student cannot be planned.

Algorithm 4 Generate minimal travel time routes for all students

```
function GENERATE ROUTES(st, WL)
    pool = st
    for each student p ∈ pool do
        if CR_p = 0 then
            for each modality m ∈ M_p do
                function CALCULATE FASTEST PATH(m, HL_p, WL, G^t)
            Select fastest modality and remove p from pool

    if |pool| = 0 then
        DONE                                         ▷ All routes are calculated
    else
        CONTINUE                                     ▷ Pool not empty

    STARTLOOP
    i = iteration
    k = RoundUp(|pool|/CarCapacity)                  ▷ Amount of cars needed
    Perform K-means clustering(pool, k)
    for clusters c ∈ k do
        if c contains DL then     ▷ Does cluster have p with drivers license?
            CalculateShortestRoute(c)
            if fastest modality for any p_{CR=2} ∈ c != car then
                Set modality of p to fastest modality        ▷ Not car
                Remove p from pool
            else                                     ▷ This cluster is optimal
                Set modality for students p in this cluster c to car
                remove p ∈ c from pool

    if |pool| = 0 then
        DONE
    else if |pool_i| != |pool_{i-1}| then     ▷ |pool| changed, but not empty
        GOTO Startloop
    else                             ▷ Only clusters without drivers remain
        for c ∈ k do
            for p ∈ c do
                if CR_p = 2 then
                    Remove car from M_{p,rt}
                    Choose between non-car modalities
        Remaining students in
        pool have CR_p = 1 and do not
        fit in a cluster with driver. END
```

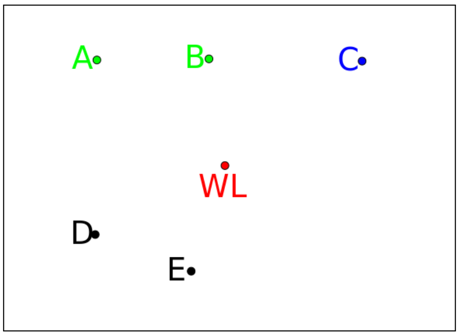
Visually this is as in figure 3.1. Depending on the spatial distribution and attributes of the students the exact steps might differ. It can be seen that a set of students $\{A, B, C, D, E\}$ is going to work at the work location WL.

Students A and B must travel by car. Both have a drivers license. Student C may travel by car and public transport. Student D may travel by public transport and bicycle. Student E can walk or cycle. This is seen in figure 3.1a.

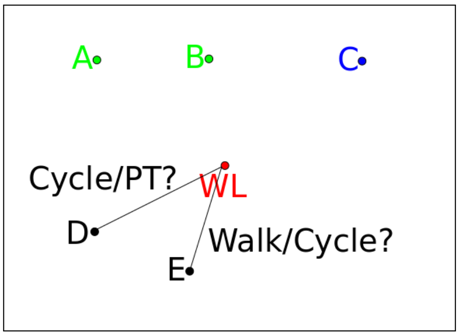
Figure 3.1b shows that it is first evaluated for all students that cannot go by car what their fastest route is. Then figure 3.1c shows that the students D and E are assigned public transport and cycling respectively. They are then removed from the pool, visualized by the color grey.

Then figure 3.1d shows that for the remaining students k-means clustering is performed. In this case $k = 1$. In figure 3.1e it is evaluated for each cluster if it is faster to take the car, or for each individual to go by any of their other modalities. In this case A and B have no option but the car, but for C it is also evaluated whether public transport is quicker. Since for C the public transport is faster, C is assigned public transport and removed from the pool in figure 3.1f.

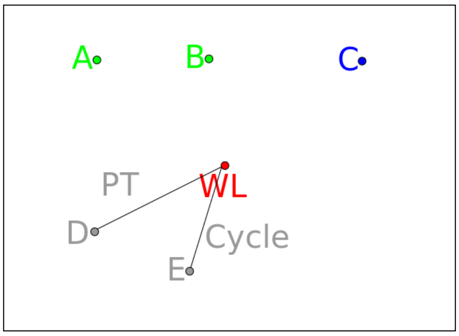
Now the algorithm starts the loop again with the remaining students. This is seen in figure 3.1g. Finally the pool is empty in figure 3.1h. All students are assigned a modality.



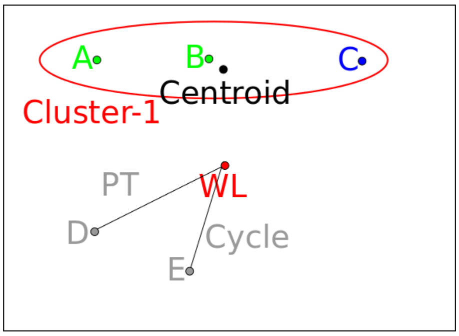
(A) (Initial state) A and B can only travel by car. C may travel by car or by public transport. D and E cannot travel by car



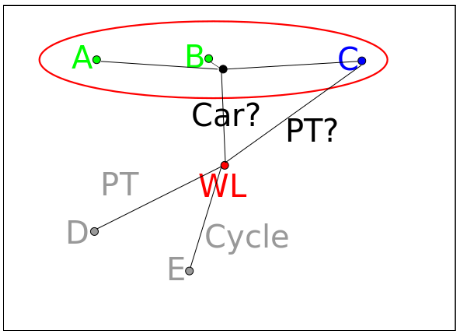
(B) (Lines 1-6) First evaluate all students that do not have *car* in their M_{rt}



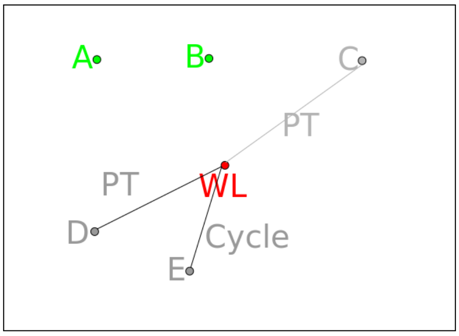
(C) (Lines 7-12) For D and E set the modality to the fastest



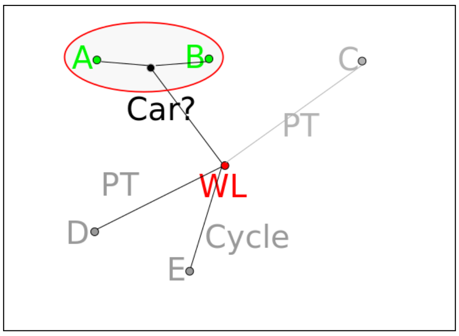
(D) (Lines 13-17) For the remaining students calculate k clusters and their centroids



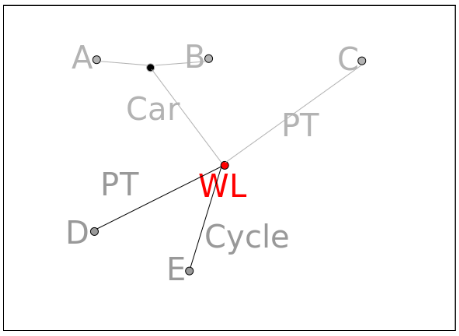
(E) (Lines 18-22) Evaluate the modalities for each of the students in the cluster.



(F) (line 23 and 28-31) For student C the public transport is faster than the car, so C is assigned to the public transport.



(G) (Lines 17-20) Again perform k-means clustering and evaluate the *car*.



(H) (Lines 24-26) Set the *car* for A and B. No students left in the pool, so the algorithm is done.

FIGURE 3.1: Algorithm 4 explained visually

Route finding has been mentioned on line twenty of algorithm 4, of which the basics have been explained in section 2.1.1. In the algorithm the shortest path has to be found for a cluster c. On other words, for a maximum of five students of which at least one has a drivers license an optimal route by car has to be formulated. In general two tactics can be chosen for this purpose. One is to let everyone gather at a single pick-up point from where the car then departs. Another is to have the car drive past every individual student's home.

The latter has the disadvantage that the first person to start travelling, which is by definition the driver, will have to travel furthest of everyone. This means the students would have an incentive to not want to be the assigned driver. An advantage of the former method is that all students can start travelling about at the same moment to the pick-up location. This is more fair to the entire group.

The pick-up point should be equally far from each student in terms of travel time. Considering the computational complexity the Euclidean geographic mean of the students in the cluster can be taken as an approximation. The route from the pick-up point to the work location can then be calculated by the Dijkstra, A*, or any other shortest path finding algorithm.

3.3 Discussion

Equation 2.5 is based on the research by Schwanen and Dijst (2002). This research is into historical data and hence equation 2.5 is also based on the average of this historical data. So data that describes the average travel time is used to predict what the maximum travel time is ought to be. It seems hence that equation 2.5 is conservative for estimating the maximum travel time budget for a student. The empirical description of B_{tt} for a student should be further fine-tuned through quantitative research.

The model is implicitly individualistic in nature. In case student A with a drivers license could either go by public transport (30 minutes) or by car (35 minutes), and in the latter case other students could carpool and thus save a total of 100 person minutes, the model would still decide that student A goes by public transport, thus losing 95 person minutes for the group. On the other hand, a more utilitarian or altruistic model might decide that student A has to travel 95 minutes, or more, extra to save 100 person minutes in the whole. Whatever is preferable is ultimately a matter of ethical opinion. On a higher level this same model is utilitarian, as the choice has been made to minimise travel time, instead of travel cost in euro's. On the other hand, student that live closer are preferred over students that live further, even if only by a second. The personal situation of the students is not included, nor is their working history. These are ethical matters, and should be taken into account when the model is applied.

A homogeneous fleet of cars is assumed. This can be an invalid assumption in the case of an heterogeneous fleet that would render the model unusable.

Only students that have been selected are used in the analysis. Students that are not selected but could potentially greatly reduce the travel time of the group are not 'allowed to enter' the equation.

The model assumes that the students start travelling from their home location. In practice this is a too narrow scope. Students might start travelling from the University, or from their parents house where they are staying in the weekend. Although the home location is conceptually easily replaced by the 'travel start' location, this would require this information to be known beforehand, something that was not possible in the case study.

It is possible for a student to not be able to travel according to the model. When a student can exclusively travel by car and does not have a drivers license, the situation can occur where all other drivers are assigned to a cluster or some other modality, leaving the first unable to travel to work.

The time complexity is ultimately dependent on the fastest path and k-means algorithm. Yet both should be solvable on a reasonable time scale considering that planning over a 100 students is rare and the road network and its complexity are finite. The loop used to find the solution in the worst case assigns a minimum of one student per iteration, thus also remaining workable.

3.4 Conclusion

This chapter aimed to answer the question *"How can these concepts be used to form a model that solves the travelling employees problem?"* A model was formulated that takes the in chapter 1 explained employee (*home location, drivers license, and availability*) and employer (*required amount of employees, work location, and arrival time*) constraints as input. With these input a two step process is then executed, namely performing a preselection and then planning an optimal set of routes and modalities for this selection. This two steps are a formalised process to handle the mediator constraints as explained in chapter 1.

The algorithms as formulated in algorithm 3 and 4 provide a model that is able to solve the TEP with the concepts discussed in chapter 2. As discussed at the end of section 3.1 and 3.2 several ethical and practical facts should be considered before the actual application of the model. In line with contemporary ethics the model should not be applied without human supervision and understanding of the conceptual steps, and their strengths, weaknesses, and pitfalls.

4 Application and assessment

The case study will help in answering the question *How does this conceptual model perform in a case study?* Section 4.1 will discuss background information of LINQ, the firm where the caste study is performed. Section 4.2 will discuss how the model will be implemented practically. Section 4.3 will present the results that the implemented model has given in the case study. Section 4.5 will discuss the results and the meaning of those results for the model validity. Conclusions will be drawn in section 4.6.

4.1 Context

General information about LINQ. It has two offices with the main office in Amsterdam, the Netherlands. Students are employed among others in security, logistics, office, and hospitality related jobs. Over 1700 students were responsible for roughly 170 thousand man hours in 2017. The office supporting the students is run by around 25 full- and part time-employees. They are among others concerned with matching students to jobs, and planning the commuting trips. Thus they are a central authority for planning the trips.

The author has been working at LINQ since 2013. Hence the travelling employees problem had already been identified, but never investigated in a scientific setting. The research in its entirety will be performed at the main office of LINQ in Amsterdam within a nine month period between September '18 and May '19.

The data that LINQ has made available for this research includes data about the employees such as living location and work history. Additionally the work locations are known.

Figure 4.1 shows a per municipality count of unique employees that have worked in 2016 and 2017. It shows that the count is highest in Amsterdam and Utrecht, 583 and 775 respectively. After that Rotterdam and Groningen with 94 and 80 respectively.

The cars that are available at LINQ are not the students own cars. In the Netherlands it is not customary for students to have their own car. Hence the cars that are described in the case study are those of LINQ. This means that any cluster centroid/pick-up point is the location where the driver should bring the car after picking it up in consultation with LINQ's planners.

Available students As stated in the problem definition (1.2, p.2) the 'available students' are those that have indicated that they are available to be scheduled for a job. In other word, they are available to go and work, but do not necessarily have to be assigned to a job.

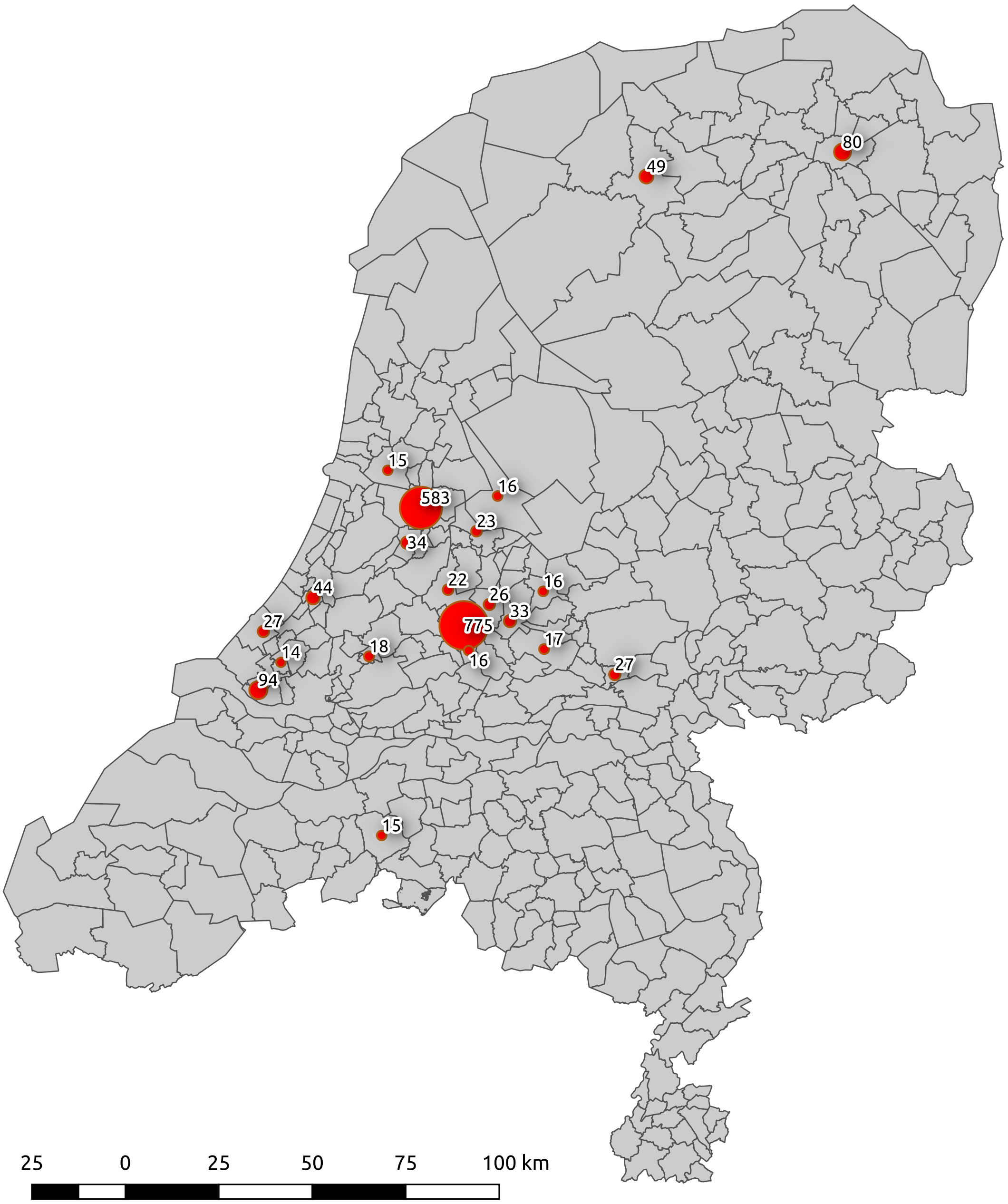


FIGURE 4.1: Overview of the per municipality count of unique employees that have worked in 2016 and 2017, only the municipalities with more than ten employees are shown.

The data about work locations is embedded in a 'locations' database. The locations include all locations where some activity has taken place. These locations consist of mostly work locations, but includes also non-working locations, for example where a company barbecue was given. The current data structure does not allow for filtering without per case manual processing. The location data is automatically geocoded and stored with latitude and longitude information. Two thirds of the 3548 locations have been successfully geocoded. Figure 4.2 gives an overview of the work locations. This is mostly in the areas of Utrecht, Amsterdam, and Rotterdam.

Thus between figure 4.1 and 4.2 it can be seen that the working locations are spread more throughout the country than the students. The working locations are mostly limited to municipalities with, or next to, relatively large cities.

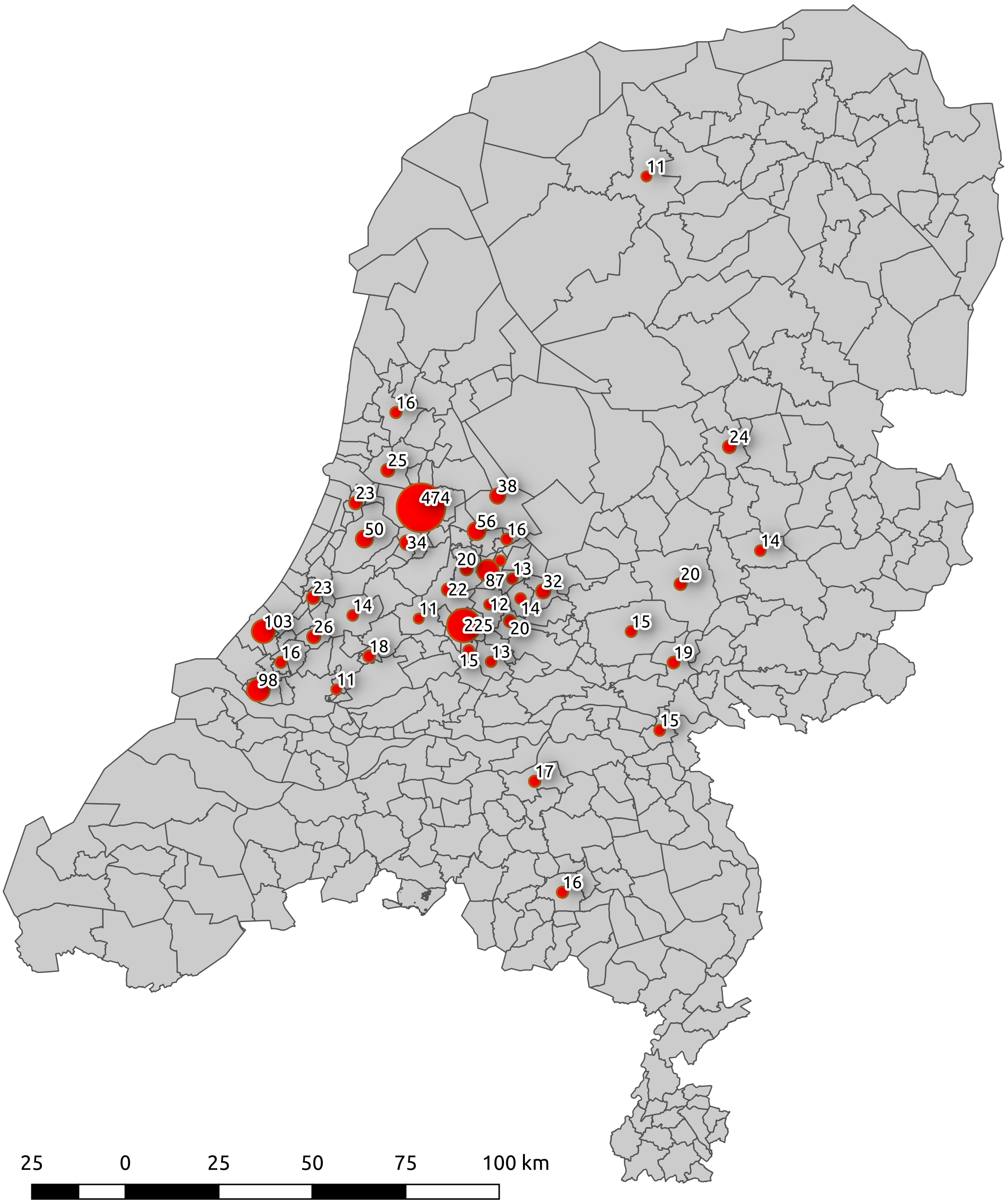


FIGURE 4.2: Overview of the per municipality work locations, only municipalities with more than 10 locations are shown.

The scientific quality is influenced by the cooperation with LINQ. It will have a positive impact on scientific quality because:

1. A dedicated office environment is available to the author.
2. The proximity to the real world travelling employees problem will increase the overall understanding of the problem. The availability of uncodified knowledge in close proximity will enhance the overall quality.
3. The office environment is prone to serendipity through informal discussion about the subject.
4. The research progress will be presented to the office staff on a hand full of occasions, effectively exposing it to experts and peers providing feedback.

On the other hand it is no secret that scientific and commercial interests do not always align. Therefore practical measures were taken to safeguard the scientific quality of this research; it was formalised in the contract between the author and LINQ that the supervisors of the University have a binding vote about research content.

4.2 Model implementation

The acronym TESS can be used for the implemented model. The implemented model will be able to solve the travelling employees problem. Therefore it will from now on be referred to as **T**ravelling **E**mployees **S**upport **S**ystem, or shortly TESS. TESS thus refers to the implemented conceptual model; the higher level conceptual model itself will be referred to as 'conceptual model'.

TESS has been implemented in the form of multiple Python scripts. The routing functionality for walking, cycling, and going by car is built by calling the Google Maps API. The public transport routing is done by calling the 9292OV API, which covers all the public transport throughout the Netherlands. The k-means clustering is done via the scikit-learn module of Python. The Euclidean distance calculation is done in the Rijksdriehoekstelsel coordinate system. The part of the code that is not concerned with TESS's integration in LINQ's systems will be attached to or available with this thesis. The details of the workings of the code are not within the scope of this research, mainly because the model conceptually works as explained in chapter 3. Also, the code is over 1500+ lines; most of it contains technical details such as how the API's are called, how errors are handled, how logs are kept, files are written, how dates and times are handled, coordinate systems converted, and how API responses are parsed.

In the implementation of the model it was assumed that the departure time for the work bound trip equals the start time of the work. Ideally the departure time would be calculated from the start time of the shift and the duration of the work bound trip. This is however a circular problem, as the duration of the trip depends on the departure time. This problem is solvable, but does not lie within the scope of this research, and more practically not within the scope of the Google maps API.

Three scenario's have been used to test TESS. In each of these scenario's the employee (amount, location, and (not) having a drivers license) and employer (time, location, and required amount of students) constraints as explained in chapter 1 have been varied.

In each scenario a random subset of students has been selected from the historical location data from LINQ. Also a work location was randomly selected. Randomly selecting a subset ensures that spatial patterns present in the historical data remain intact. The students were randomly given drivers licenses. In the historical data of LINQ the ratio of drivers license possession cannot be determined reliably, as it is not known of over two thirds of the students whether they have a license or not. Therefore the national ratio of 0.5 in the age group of 16-25 years is used (CBS 2019; CBS 2018). The amount of available students was varied between five and twenty.

Since it is practically impossible to test TESS in all possible scenario's, let alone qualitatively discuss them, three scenario's will be run. In all scenario's the δ as discussed in section 3.1 will be set to a value of 10 minutes. This value has been determined by discussing it with planning experts from LINQ. All dates are on 'normal' days, i.e. not holidays or other days with special circumstances with a significant impact on travel time, such as sports events.

Scenario:	Start time work	End time work	Duration	Day of week	Date	Required students	Available students
A	08:00	16:00	08:00	Monday	15-04-2019	1	5
B	12:00	18:00	06:00	Wednesday	10-04-2019	6	15
C	15:00	03:00	12:00	Saturday	13-04-2019	11	20

TABLE 4.1: Three scenario's and their respective attributes.

The choice to not keep all but one variable constant between scenario's allows for the widest variation of scenario's to be able to monitor how TESS performs on coping with *changing travel cost throughout time*, balancing *travel time versus travel cost*, making *modality* choices, and deciding whether or not carpooling should take place.

The trade off of this approach is that effects of individual variables are not easily discerned. Changing only one of the variables at a time, as is customary in quantitative research, would mitigate this problem, but then an impractical amount of scenario's needs to be run. Let alone testing all possible scenario's; here the same problem is faces as with the exhaustive search as discussed in subsection 2.1.2 (p.8).

4.3 Case study results

4.3.1 Scenario A

Scenario A: a single student is needed, where five are available. The shift is from 08:00-16:00 (8 hours) local time on Monday 15th April 2019.

Input To determine what students are available for this scenario, five students were randomly selected from LINQ's historical data. This resulted in five students dispersed across the region Amsterdam, Utrecht, Rotterdam, and Zuiderwoude, see figure 4.3. Table 4.2 shows the student's coordinates, and whether or not they have a drivers license. On the right side their travel times for a one way trip have been calculated by TESS. These are the travel times that are used by TESS to select a subset of students from among the available students. The travel time is shown in seconds. For walking and cycling trip symmetry is assumed.

ID	City	X	Y	Has drivers license	Duration seconds Walk One way	Cycle One way	Public Transport Work bound	Home bound	Car Work bound	Home bound
0	Utrecht	137886	455969	No	31321	8518	4560	3600	2796	2994
1	Zuiderwolde	130898	493824	Yes	10162	2853	4320	3660	1664	1590
2	Amsterdam	116668	484595	Yes	4555	1462	2340	2100	968	1196
3	Rotterdam	92192	437988	Yes	48457	13625	6060	3720	3432	3563
4	Utrecht	136152	453498	No	31901	8795	4380	3660	2641	2769

TABLE 4.2: Students used as input for scenario A, with their respective attributes. Locations are in the Rijksdriehoekstelsel coordinates (EPSG:28992). Also shown are their one way trip travel times per modality. This is the travel time that is compared to the B_{tt} by TESS.

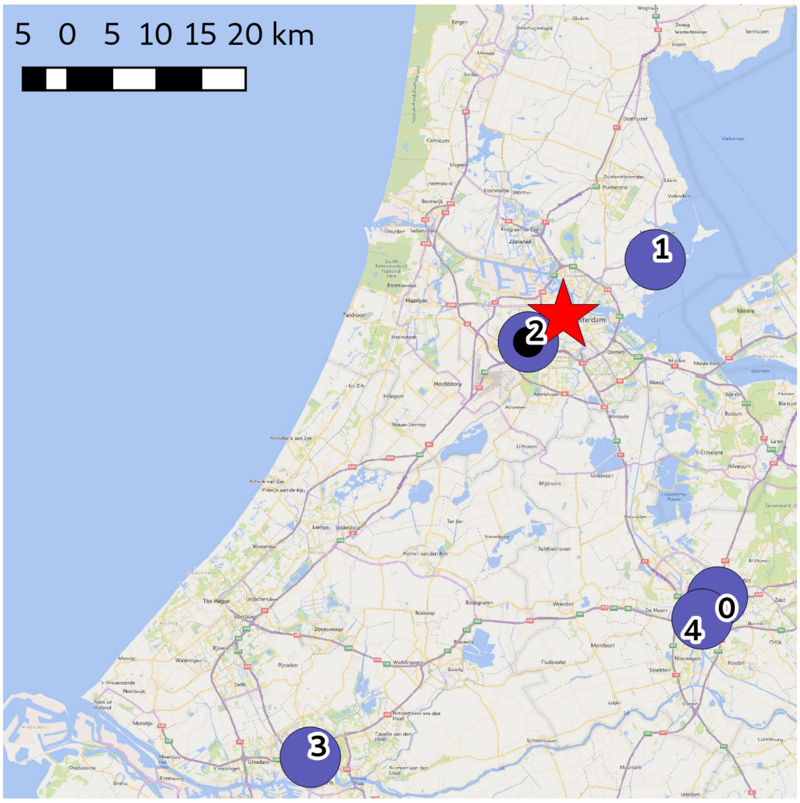


FIGURE 4.3: Results for scenario A in the Randstad area, the Netherlands. The work location (red star) in Amsterdam, available students (blue circles), and selected student (black dot).

Followed steps and result TESS first checks whether the students live close enough so that a trip can be made within the travel time budget of 2083 seconds, or 35 minutes, based on equation 2.5, for a one way trip. This means that the initial selection consisted of students with id's 1 and 2. Since there is only one student needed for this scenario, student 2 is selected because this student is closer (3.921m) in Euclidean space to the work location than student 1 (10.309m), as explained in section 3.1 (p.23).

Now the modality will be chosen. Student 2 is put in cluster 0, the only cluster. This cluster has at least one student with a drivers licence, so the car modality is possible. The cluster centroid equals the living location of student 2; this student is the only student in the cluster. Hence the time the student takes to cycle to the cluster centroid is 0 seconds. The drive time from the cluster mean to the work location is 968 seconds; the trip back takes 1001 seconds. Now the modalities for a round trip are:

What	Value (seconds)	Comments
B_{tt}	4166	For the round trip
Walking	9110	Outside B_{tt}
Cycling	2924	Within B_{tt}
Public transport	4440	Outside B_{tt}
Walking work bound – public transport home bound	6655	Outside B_{tt}
Public transport work bound – walking home bound	6895	Outside B_{tt}
Car	1969	within B_{tt}, higher than 2δ

TABLE 4.3: Scenario A, modalities and their duration (round trip) in seconds for student 2.

Resulting in the advice that student 2 should go by car. Note that this decision is solely based on the fact that the car has the shortest travel time, not that it is within B_{tt}. The validity of this output will be discussed more thoroughly in subsection 5.1.1 on page 62.

4.3.2 Scenario B

Scenario B: six students are needed, where fifteen are available; the shift is from 12:00-18:00 (6 hours) local time on Wednesday 10^{th} April.

Input Fifteen students from the regions Amsterdam, Utrecht, Wageningen, Nijmegen and Eindhoven were selected as input from LINQ's historical data. Their attributes can be seen in table 4.4 and figure 4.4. The working location, which too is randomly selected from the historical data is also located in Amsterdam.

ID	City	X	Y	Has drivers license	Duration seconds Walk One way	Cycle One way	Public Transport Work bound	 Home bound	Car Work bound	 Home bound
0	Amsterdam	126058	489460	Yes	17482	5452	3900	3240	1196	1184
1	Zeist	146108	456693	Yest	77064	22792	5940	6000	3109	2911
2	Amsterdam	121356	483913	No	12342	3712	1920	1980	1103	933
3	Wageningen	173931	442038	Yes	126178	35384	6660	6780	4183	4107
4	Nijmegen	188133	428319	No	162096	47414	7440	7320	4983	5051
5	Amsterdam	123534	486056	Yes	12934	4182	2940	2940	1709	1477
6	Utrecht	138004	452976	Yes	72556	19984	4500	4440	2491	2529
7	Utrecht	138713	456071	No	70934	19520	4440	5340	2512	2591
8	Utrecht	135114	456177	No	65864	18436	3540	3840	2359	2457
9	Zeist	143593	455417	No	79050	21840	5340	5160	2994	2950
10	Utrecht	135002	457308	No	63740	17712	4080	4980	2400	2477
11	Eindhoven	160836	379182	Yes	203774	56290	8040	8940	5246	5221
12	Amsterdam	123217	485224	No	13746	3986	2580	2700	1159	1236
13	Utrecht	137288	452825	No	72200	19964	4320	4020	2366	2395
14	Utrecht	137620	456472	No	68698	18974	4260	4620	2681	2764

TABLE 4.4: Students used as input for scenario B, with their respective attributes. Locations are in the Rijksdriehoekstelsel coordinates (EPSG:28992). Also shown are their one way trip travel times per modality that have been used for the preliminary selection. The selected students are shown in green.

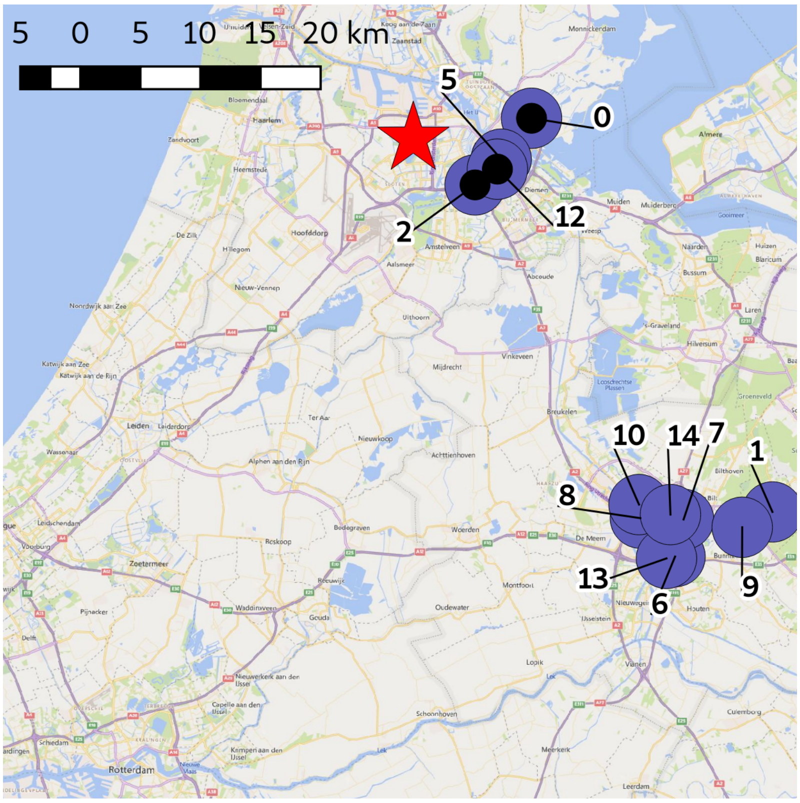


FIGURE 4.4: Results for scenario B in the Randstad area, the Netherlands. The work location (red star) in Amsterdam, available students (blue circles), and selected students (black dot). Students with ID's 3, 4, and 11 are not shown because their home locations are in Wageningen, Nijmegen, and Eindhoven respectively. Out of the scope of this figure.

Followed steps and result The travel time budget for a shift of six hours is 1705 seconds, or 29 minutes for a one way trip. Thus students 0, 2, and 12 are selected. This is less than the required six. This seems odd as student 5 seems to live closer to the working location than student 0. Yet an examination of the travel times of students 5 shows that he/she lives four seconds too far away by car to stay within B_{tt}. The home bound trip by car is within the B_{tt}.

With the three selected students the loop to determine the optimal modality is started. All three students fall within the same cluster. Within this cluster student 0 has a drivers license. Students 0, 2, and 12 would take 1040, 868, and 294 seconds respectively for a one way trip between their home and the cluster centroid. Drive time from cluster centroid to work is 1443 seconds, and from work to the cluster centroid is 1617 seconds. Thus in the first loop the following travel times for the car modality are calculated for a round trip by summing twice the cyling time and the back and forth car duration:

- student 0: 5140 seconds ($2 * 1040 + 1443 + 1617$)
- student 2: 4796 seconds ($2 * 868 + 1443 + 1617$)
- student 12: 3648 seconds ($2 * 294 + 1443 + 1617$)

If this is compared to all the other modalities available to each student both student 0 and 12 should go by car, but student 2 should go by bike. Hence student 2 is assigned to go by bike, and the loop is restarted. In the second loop the car travel times for the round trip are as follows:

- student 0: 5056 seconds
- student 12: 4308 seconds

Hence student 12 is also assigned to go by bike as this is faster. Again calling for a restart of the loop. In this third loop student 0 is assigned to go by car. So the final configuration becomes:

- student 0 goes by car and takes 2380 seconds for the round trip
- student 2 goes by bike and takes 3712 seconds for the round trip
- student 12 goes by bike and takes 3986 seconds for the round trip

This also means that the travel time budget for the round trip, which is 3410, is exceeded for students 2 and 12. This is because they are selected by TESS based on their car trip duration from their home directly to the work location. In the second part the students are combined in clusters and hence their car travel time increases by the cycle time from/to the cluster centroid and the difference between the driving time from the centroid to the working location and the driving time from their home location to work and back. Here this resulted for student 2 and 12 in a longer travel time by car than by bike, which in turn means that for both students 2 and 12 it is advised that they should go by bike. In this final step the B_{tt} is not taken into account by TESS, nor are not-selected students despite the fact that they might have faster modalities available. This will be more thoroughly discussed in section 4.5 (p.58).

4.3.3 Scenario C

Scenario C: eleven students are needed, where twenty are available, the shift is from 15:00-03:00 (12 hours) local time on Saturday-Sunday 13-14 April.

Input Twenty students and a working location were randomly selected from LINQ's historical data. Again the working location is in Amsterdam. Most students are from the region Amsterdam, Rotterdam, and Utrecht, with student 8 and 16 living in Nijmegen and Tiel respectively. This is shown in table 4.5 and figure 4.5.

ID	City	X	Y	Has drivers license	Duration seconds Walk One way	Cycle One way	Public Transport Work bound	Public Transport Home bound	Car Work bound	Car Home bound
0	Zeist	145059	455021	No	70428	18034	4800	18300	2534	2514
1	Maarssen	131444	462112	Yes	44102	11588	4800	19800	1774	1763
2	Amsterdam	120210	484478	No	2996	1022	1320	3840	344	603
3	Rotterdam	93096	436789	Yes	91430	25690	4560	12000	3507	3546
4	Rotterdam	95162	437725	Yes	87628	25218	4860	12420	3249	3321
5	Utrecht	134449	455418	No	55210	13940	3840	18360	1877	1839
6	Utrecht	136970	455400	Yes	57004	14386	3900	18960	2353	2392
7	Rotterdam	91832	437984	Yes	89352	24962	3900	10260	2973	3117
8	Nijmegen	187242	427946	No	153120	40032	6900	21300	4756	4761
9	Velserbroek	104491	494327	No	34128	9778	4140	7440	1596	1669
10	Zeist	144575	457128	No	66830	17610	4800	19320	2524	2494
11	Rotterdam	94476	441535	Yes	82216	22758	6060	13560	3125	3266
12	Utrecht	138130	456496	Yes	58206	14718	3720	18840	2223	2246
13	Maarssen	132524	460029	No	47854	12046	4380	19260	1661	1590
14	Rotterdam	93163	437425	No	90422	25252	4380	12240	3363	3468
15	Utrecht	137974	453287	Yes	60552	15300	3840	16320	2029	2025
16	Tiel	158961	433911	No	107950	29032	8940	20820	3232	3200
17	Amsterdam	122038	487143	Yes	8836	3038	1800	3780	1419	1241
18	Utrecht	136152	453498	Yes	59046	15266	4080	18660	2059	2071
19	Amsterdam	119096	485876	Yes	6396	2150	2400	4440	1060	850

TABLE 4.5: Students used as input for scenario C, with their respective attributes. Locations are in the Rijksdriehoekstelsel coordinates (EPSG:28992). Also shown are their one way trip travel times per modality that have been used for the preliminary selection. The selected students are shown in green.

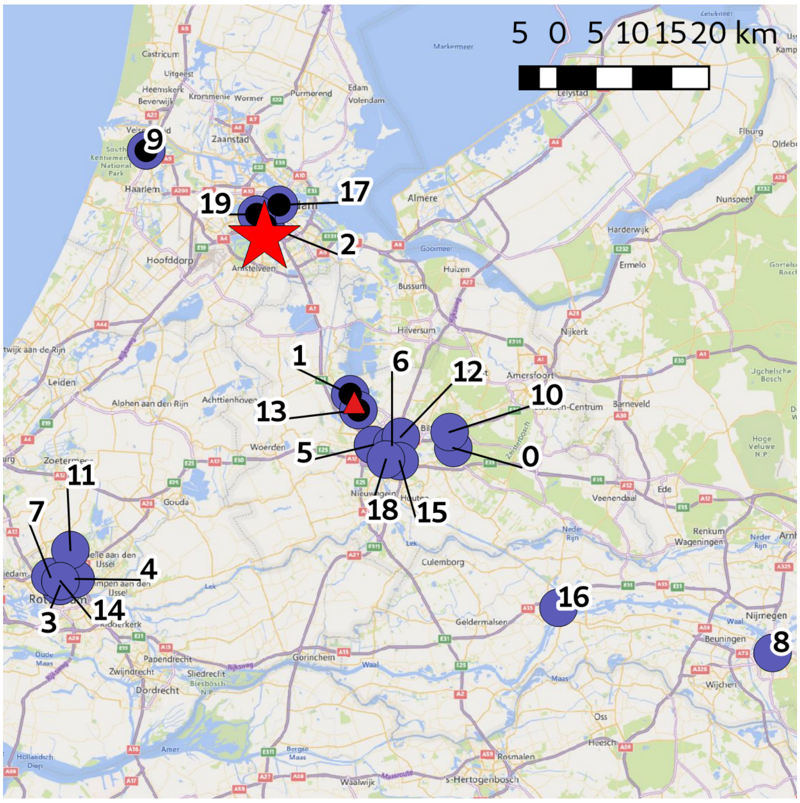


FIGURE 4.5: Results for scenario C. The work location (red star) in Amsterdam, available students (blue circles), and selected students (black dot). Also the location that the car for cluster 1 departs from (the cluster greographic mean) is shown as a red triangle.

Followed steps and result Based on a B_{tt} of 1811 seconds six students were selected, even though eleven are needed. All not-selected students cannot reach the work location, or their home location from work, within the B_{tt}. Since this shift ends at 3 o'clock at night, it is seen that all students have asymmetry between their work bound and home bound public transport duration by a factor between 1.5 and 5.

For student 2 the CR_p variable, see section 3.2 (p.26), is set to 0 because the student cannot go by car. The available modalities for this student are public transport, walking, or cycling work bound, and the car, walking, or cycling home bound. Although the car modality could be used within the B_{tt} to reach the work location, the duration is shorter than the δ of 600 seconds. Because this student has a CR_p of 0 the fastest non-car modality is chosen. In this case the student goes by bike with a one way travel time of 1022 seconds.

The remaining students (1, 9, 13, 17, and 19) are then clustered via the k-means method, resulting in a single cluster with at least one driver. The drive time from this cluster to the work location and back are 922 and 965 seconds respectively. Considering that this cluster centroid is located somewhere in the middle of Amsterdam and Utrecht, rather high cycle times for the students to reach the cluster centroid are set.

- Student 1: 4652 seconds to cluster mean, 11191 seconds total for the round trip
- Student 9: 5876 seconds to cluster mean, 13639 total
- Student 13: 4881 and 11649
- Student 17: 2318 and 6523
- Student 19: 2241 and 6369

This results in that student 9, 17, and 19 should go directly by bike, and 1 and 13 by car. Students 9, 17, and 19 are set to the biking modality and removed from the pool. For students 1 and 13 a new iteration is started, resulting in that both should go by car with the cluster centroid, where the car trip starts, in the middle between. They take 317 and 366 seconds respectively for a one way cycling trip between the cluster centroid and their home. The drive time from the cluster centroid to work and back is 1647 and 1660 seconds respectively, thus resulting in a total travel time of 3941 and 4039 respectively for the round trip.

4.4 Varying the travel time budget

Although the scientific process is portrayed as linear, it often is not. This section aims to further explore the workings of TESS after the experts have had made comments as explained in chapter 5. In short, it was commented that TESS does not always returns the needed amount of students. This is deemed a problem by the experts. It will also be discussed in chapter 5 that the travel time budget is highly variable per individual and that the approach formulated in equation 2.5 was based on highly variable data as described by Schwanen and Dijst (2002). Thus to see if this high variation can be captured in the model, scenario B will be repeated with different travel time budgets. Finally, it will also be repeated with a set of available students in Amsterdam, effectively making all students live within the B_{tt}.

The travel time budget is calculated with the work duration W in hours as input. In this section five extra scenario's will be run; D, E, F, G, and H.

- D: B_{tt} will be calculated based on $W + 0.5$
- E: B_{tt} will be calculated based on $W + 1.0$
- F: B_{tt} will be calculated based on $W + 1.5$
- G: B_{tt} will be calculated based on W as usual, but 900 seconds will be added to the result.
- H: Equal to B, but with a different set of available students

The 9292OV API does not allow for queries in the past. Scenario's D, E, F, G, and H are investigated after the date of scenario B, 10^{th} of April 2019, so the date of scenario's D, E, F, G, and H has to be changed for TESS to be able to run. The date will be changed for scenario's D, E, F, and G to Wednesday 1^{st} of May because this is also a regular Wednesday. For H it is set to the 8^{th} of May. No changes in public transport schedule were made during that time, making the scenario's equal. Scenario's D, E, F, and G will be discussed first, scenario H after that.

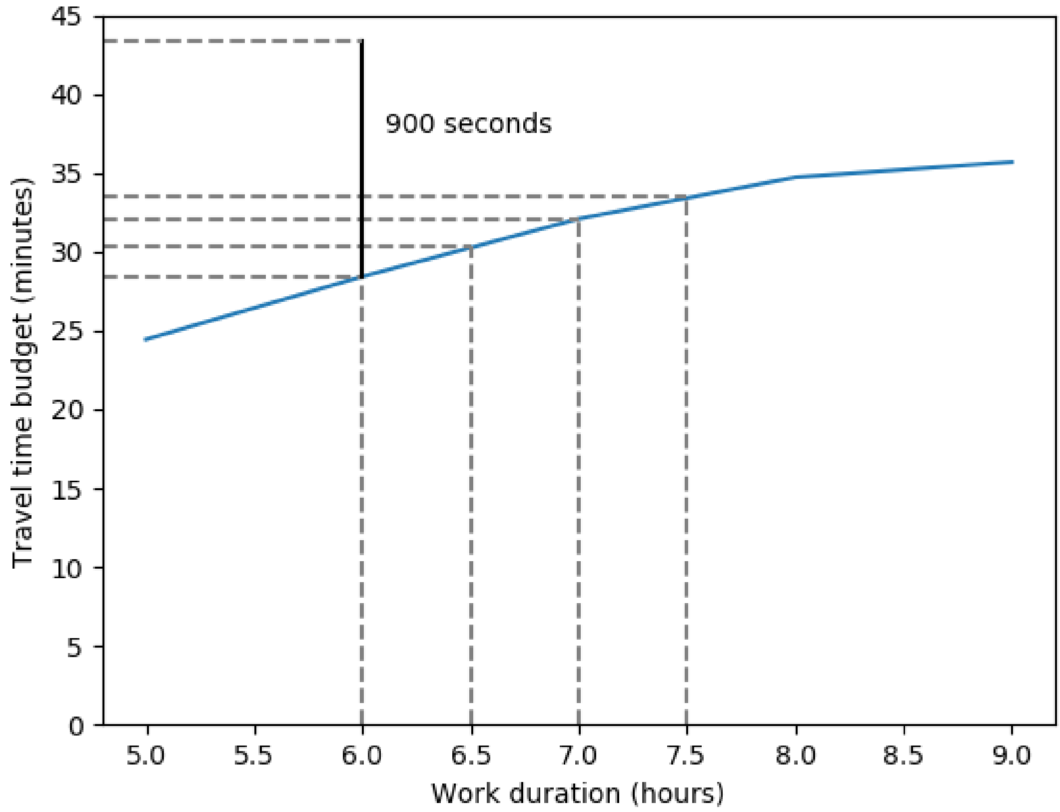
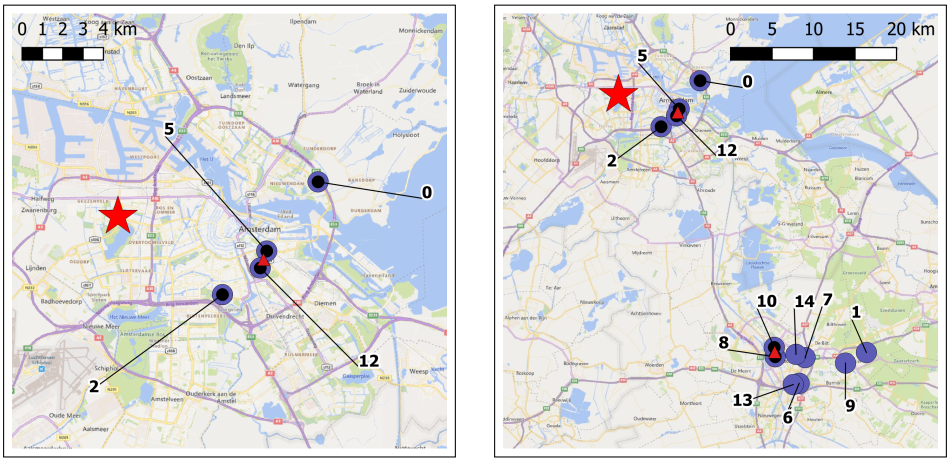


FIGURE 4.6: The relationship between work duration and travel time as by equation 2.4 (p.13) in blue. The grey lines indicate the travel time budget for scenarios D, E, F, and G.

Secnario's D, E, F, and G The scenario's and their corresponding B_{tt} can be seen in figure 4.6. They are as follows:

- B: $B_{tt}(6.0) = 1705$, or 28 minutes 25 seconds (the original scenario)
- D: $B_{tt}(6.5) = 1820$, or 30 minutes 20 seconds
- E: $B_{tt}(7.0) = 1925$, or 32 minutes 5 seconds
- F: $B_{tt}(7.5) = 2014$, or 33 minutes 34 seconds
- G: $B_{tt}(6.0) + 900 = 2605$, or 43 minutes 25 seconds

Scenario B is the 'original' scenario. For scenario's D, E, and F the W is increased over scenario B. Thus also increasing $B_{tt}(W)$ As can be seen in figure 4.6 The effect of increasing W increases the B_{tt} with up to five minutes for an increase in W of 1.5. It could be said that this situation simulates paying up to 1.5 hours of travel time. Scenario F is shown by the solid black vertical line which represents the 900 seconds that are added to the initially calculated B_{tt}.



(A) Scenario's D, E, and F (B) Scenario G

FIGURE 4.7: Scenario's D, E, F, and G

Figures 4.7a and 4.7b show the maps that represent the outcomes of scenario's D, E, F, and G. There was no difference in outcome between scenario's D, E, and F. Of the six students needed only four were selected by TESS. These scenario's do differ from scenario B because student 5 is now also selected. Additionally, student 5 and 12 are selected to go by car, with the pick-up point in the middle.

Paying for up to 1.5 hours of travel time seems to not be enough incentive for the students outside Amsterdam to travel to the working location. The validity of this will be more thoroughly discussed in subsection 5.1.3 (p.65). Table 4.6 shows the exact modalities and travel times for the round trip of each student. Note that in scenario G two cars are used, one for students 5 and 12, and another for students 8 and 10. It is also shown in the figure 4.7b that only scenario G returns the amount of students that are required.

Scenario	B		D, E, F		G	
	Modality	Duration	Modality	Duration	Modality	Duration
Student 0	Car	2380	Cycle	5454	Cycle	5454
Student 2	Bike	3712	Cycle	3712	Cycle	3712
Student 5	-	-	Car	3232	Car	3232
Student 12	Bike	3986	Car	3078	Car	3078
Student 8	-	-		-	Car	5391
Student 10	-	-		-	Car	5435

TABLE 4.6: Travel times and modalities for scenario's B, D, E, F, and G

Scenario H In scenario H, 15 students living in Amsterdam were randomly selected from the historical data of LINQ. The rest of the variables are equal to scenario B, with the exception of the date. That is set to the 8^{th} of May for the same reason as in scenario's D, E, F, and G.

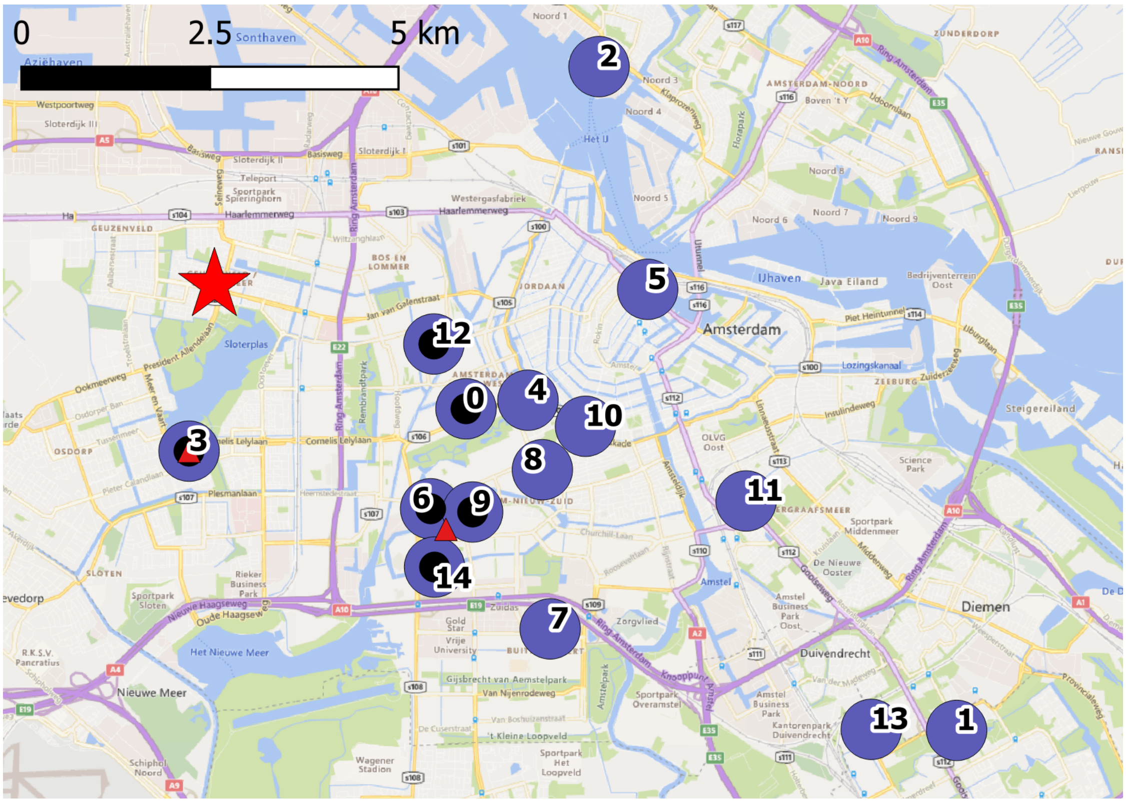


FIGURE 4.8: Scenario H, Amsterdam, The Netherlands

ID	X	Y	Has drivers license	Duration seconds Walk One way	Cycle One way	Public Transport Work bound	Public Transport Home bound	Car Work bound	Car Home bound
	119550	486115	Yes	7016	2124	1620	1620	766	860
1	126057	481847	Yes	20482	5786	2880	2760	1023	1017
2	121314	490656	No	10098	3234	2700	3120	1058	998
3	115877	485553	Yes	5108	1744	1440	2700	637	647
4	120372	486229	No	8272	2498	1980	1920	998	978
5	121959	487701	Yes	10062	3072	1980	1800	1376	1361
6	119081	484788	Yes	8286	2652	1800	2400	914	894
7	120666	483194	No	12926	3936	1860	1920	839	893
8	120563	485306	No	9372	2836	2280	2400	1116	1208
9	119638	484745	No	9264	2856	2100	2820	1018	990
10	121135	485883	No	9592	2754	2160	2100	1206	1338
11	123265	484892	Yes	14128	4186	2400	2640	1134	1107
12	119124	486975	No	5528	1708	1080	1200	666	665
13	124925	481861	No	19812	5584	2220	2400	1158	1077
14	119133	484015	No	9370	3046	1680	1680	827	803

TABLE 4.7: The input for scenario H

Table 4.7 shows that each student has at least one modality that allows the student to reach the work location within the travel time budget of 1705 seconds. Figure 4.8 shows that the selected students are the ones that live closest to the work location by their Euclidean distance. This in in line with the algorithm as explained in algorithm 4. It also shows two car clusters.

Student 6, 9, and 14 take one car. Student 3 takes another. Student 0 and 12 should take the bike, as this is faster than public transport.

The experts have commented in chapter 5 that taking the car in Amsterdam might not be the best option due to traffic and parking. The former is no argument since TESS includes traffic in the calculations. The parking argument is covered by TESS via the δ variable, which ensures the car is not selected for 'short' rides that should be cycled.

4.5 Discussion

TESS is successful in producing workable output within an acceptable time. During the testing and building of TESS, she has been run hundreds of times on an input ranging from 1 to 30 students. Each model run took no more than 30 seconds, often less, on a low end machine with 2GB RAM and 1.8 GHz dual core processor. It also means that the process of model formulation and TESS implementation have been more of an AGILE process as is customary in software development, rather than a linear process as is customary in formal science. It can be stated that TESS's output seems applicable in the real world, as will be tested in chapter 5.

TESS also successfully accounts for public transport time tables. Public transport did not seem to be an option in scenario C, as the home bound trip took much, by a factor of 1.5-5, longer than the work bound trip because the shift ends at 03:00. This formal number is acquired by TESS in less than a second via an API call. A human planner cannot acquire this same information in the same time. As will be further discussed in chapter 5 human planners rely more on gut feeling rather than exact measurements, e.g. 'Student A cannot go back by public transport because it is 3 'o clock at night' versus 'Student A would take 4203 seconds to go back by public transport at 03:14'. This seems to be an advantage that TESS has over a human planner. Also the way TESS can deal with increasingly larger groups in just seconds is an advantage.

Besides the quick run time and exact information, points of attention have been also been identified. First, as is shown by scenario's B and C, TESS might not always return the amount of students that is needed for the job; in some cases she will return less. The reason for this is that students cannot make the home and work bound trip with any modality within the B_{tt} for a one way trip. Scenario's D, E, and F showed that the B_{tt} calculation should be further investigated because it is currently based on highly variable as described by Schwanen and Dijst (2002).

Second, the δ parameter – the parameter that determines how long a car trip should minimally be before the car can be chosen as a modality – has been fixed on a value of 600 seconds for the scenario's. While informally discussing the value with experts it was noted that this value should change throughout space, as an urbanised environment allows a shorter distance to be covered by car than a rural environment in the same time. This is what happened in scenario H. A more formal discussion is found in chapter 5. In the latter case 10 minutes by car might equal an hour by bike, while in the city 10 minutes by car might equal 5 minutes by bike. This influences when a car is preferred over a bike. In the rural case a lower δ seems appropriate. The captures the fixed cost related to parking the car, getting a driver to the car, and ensuring the keys are also where the driver and the car are. These cost change throughout space, and might even change per individual. The spatial variation of δ should be further investigated to improve the model. It might also be possible to estimate the δ on a per shift basis.

Third, the B_{tt} is used in the selecting of the students, not in their modality choice in the second part of the model. Not-selected students are also not considered in this second step of the model, even though they might still improve the overall solution. This can result in that a student gets the advice to go by bike, as the car cluster is inefficient, resulting in that the travel time budget is exceeded. Also, the B_{tt} for a one way trip is assumed to be half the B_{tt} for that entire work activity. It is, however, possible that the work bound trip takes three fourths of the budget, if the home bound trip takes than one fourth. This asymmetry problem is largest at the edges of the $SA(B_{tt})$, as even a small asymmetry between the work and home bound trip could lead to not selecting a student while the total trip still is within the B_{tt}. This could be solved by checking duration of the round trip and comparing it with the full B_{tt} rather than a per one way trip logic.

Finally, for the walking and cycling modality trip symmetry is assumed. Although this seems a fair assumption for a case in The Netherlands, this assumption will be violated when TESS is deployed in a mountainous area.

4.6 Conclusion

This chapter aimed to answer the question *How does this conceptual model perform in a case study?* The implemented version of the model, TESS, showed to create logically sound although not perfect outputs in acceptable run times when it was tested with input based on historical data of the student employment agency LINQ. Therefore it can also be said that the research objective *This implementation should be sufficiently fast in terms of computational time* has been met.

In the scenario's the employee and employer variables were varied. Throughout this variation TESS showed to be capable of dealing with *changing travel cost throughout time* in a logical manner. Public transport timetables were honoured. Balancing *travel time versus financial cost* was well done by TESS. In each scenario the selected students seem to be the ones that live nearest to the work location. Although combined with the *modality* choice it did not go equally logical each time; the output that a student that lives far should go by bike as in scenario C is not logical. This flaw has to do with the fact that after the initial selection takes place, the modalities that were used for that selection are no longer honoured by TESS.

Furthermore, the *carpooling* is both logical and illogical. The carpool clusters that are made such as in scenario D, E, F, G and H seem efficient, given that the δ parameter is assumed to be optimal. However, they could be improved even further if students outside the initial selection were to be allowed to carpool. A good example would be scenario C, where student 9 could be replaced by any student living in Utrecht.

One final remark is that TESS does not always honour the inputs it is given. Not always the required amount of students that are needed by an employee are selected. This could be called a flaw of TESS, or more logically the job simply is too far away and travel time compensation is too limited.

5 Expert Validation

The goal of the expert validation is to evaluate how 'good' the model is in terms of the employee, mediator, and employer constraints as formulated in chapter 1. This is relevant as it provides a basis for further research and to gain an understanding as to what degree TESS can be applied. To reach this goal the question *How do experts judge the model?* will be answered.

The experts will be planners from LINQ. They face the travelling employees problem on a daily basis. Since the planning at LINQ is centralised, they are also the 'hub' where the opinions and knowledge of the student that are being planned accumulates.

The experts will be presented with scenario's A, B, and C as explained in chapter 4. They will have not seen the scenario's beforehand. They are also not (yet) aware of the results of TESS on that same scenario. The experts are asked how they would plan in that specific scenario. They are given a A4 paper map equal to the maps of chapter 4, with the difference that the results of TESS are not visible on their maps. They will be asked to select and plan the students in those scenario's. If asked for they will be told the employee and employment constraints, which they will either have to memorise or write down. They are asked to speak out loud what they are doing. The planning is then confirmed by the experts either writing/scribbling down the planning or verbally announced to the interviewer. The interviews take place at the lunch table (although not during lunchtime) of LINQ. The experts will have availability over pen, paper, and their mobile phone.

After the planning has been finalised the experts will be presented with the output of TESS of those same scenario's. Similarities and difference between the output of the expert and output of TESS will be discussed in which the interviewer is to refrain from any steering comments. The interviewer will, when asked, elaborate on the workings of TESS. Directly after the results of TESS have been presented the interview will become unstructured to allow the expert to discuss what he/she deems important. Thus allowing for both scenario specific as well as more general comments. No time limit has been set. All interviews, including the planning of the scenario's, took about 15-30 minutes. They stopped only when the expert was 'done', not terminated by the interviewer. The experts agreed with the results being presented anonymously in this report. During the whole of planning the scenario's and the unstructured interview the interviewer took minutes.

The structure of this section is as follows. Section 5.1 will discuss scenario's A, B, and C respectively. For readability the figures from section 4.3 (the map of each scenario) will be duplicated in each subsection. Any scenario specific comments will also be presented here. Section 5.2 will discuss experts comments that are of a more general nature. It will also discuss the state of TESS more generally. In section 5.3 conclusions will be drawn.

5.1 Scenario specific findings

5.1.1 Scenario A

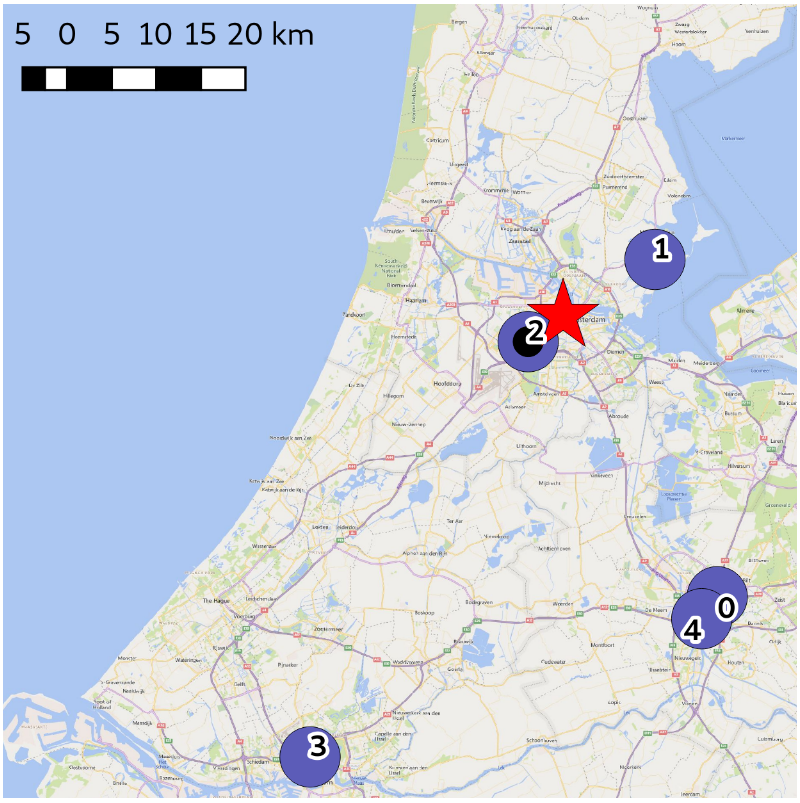


FIGURE 5.1: Duplicate of figure 4.3 (p.47). Five students are available, one is needed. The shift is eight hours long and starts at 08:00hrs local time on Monday 15th April 2019.

Expert A chose to send student 2 via public transport because he/she lives nearest. The concept of nearness was not further explained by the expert, but it is the authors interpretation that expert A used Euclidean distance. Student 0, 3, and 4 were said to live too far away. For student 1 it was commented that the accessibility with public transport was not good enough. This seemed to be based on the assumption that the rural region of student 1 means that the public transport accessibility mus thus be 'bad'.

On the output of TESS (student 2 by car) it was commented that this does not seem feasible. The reason was given that the parking cost for a single car in the city centre of Amsterdam outweigh the hourly profit of a single student. Hence such a shift would cost rather than produce income for a student employment agency.

Expert B chose to send either student 1 or 2. It was commented that student 1 is able to travel for free during the weekdays, hence he/she is preferred. But is was also asked whether or not the time and/or distance travelled were compensated for by the client to LINQ. And if that were the case travel cost made for public transport would be no selection criterion. It was chosen ad-hoc that in the scenario the same 'rules' for the compensation of travel time were valid as is the case at one of the clients of LINQ.

On the recommendation of TESS the same comments as expert A made were also made by expert B. Also it was noted that the time needed to find a parking spot would outweigh the benefit of taking the car. Further points will be discussed in section 5.2 (p.66).

Expert C first asked about what the rules were with regard to the client paying for travel expenses and travel time. First the same rule set as for experts A and B was offered, but this expert indicated not having experience with that particular rule set. Hence the choice was made to not have any compensation by the client. Consequently expert C chose to send student 2 either by bike, or by public transport with the transport fee being for the student him/herself.

The solution offered by TESS was not deemed preferable because of the limited availability of parking space, the time needed to search for this space, and the cost for parking.

5.1.2 Scenario B

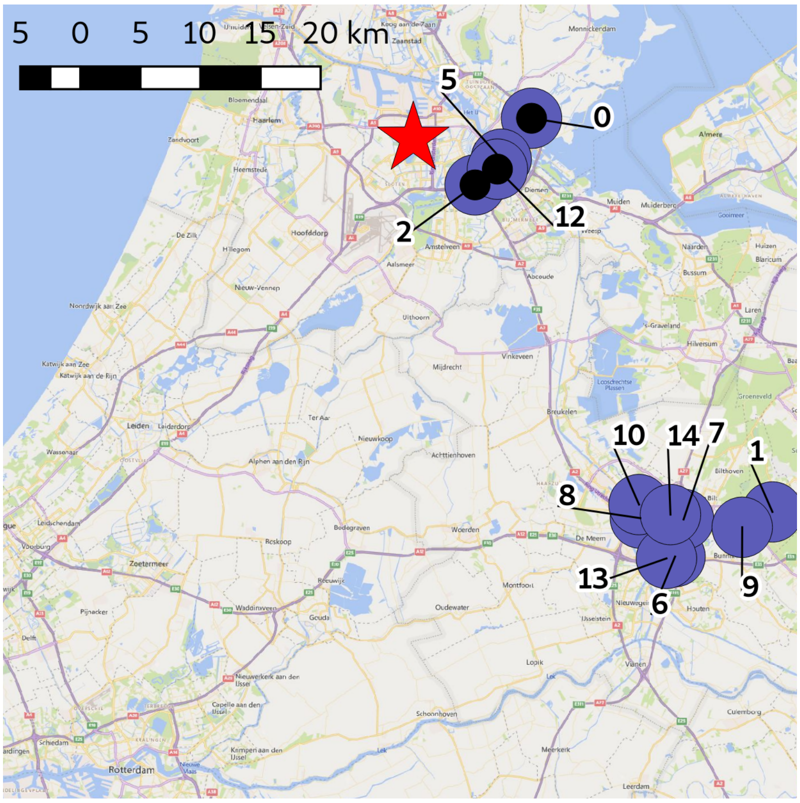


FIGURE 5.2: Duplicate of figure 4.4 (p.50). Fifteen students are available, six are needed. The shift is six hours long and starts at 12:00hrs local time on Wednesday 10th April 2019.

Expert A chose to send a car with a capacity of nine. LINQ has a non-homogeneous fleet of cars, one of which has a total capacity of nine students. TESS does not model a heterogeneous fleet of cars. The bus should depart from LINQ's office in Utrecht and carry those that are from Utrecht, namely students 6, 7, 8, 10, 13, and 14. It was also mentioned that this is more preferable than having several students travelling solo, because the perceived certainty of the whole group arriving on time increases. This will be more thoroughly discussed in section 5.2.

Expert B chose to send students 0, 2, 5, and 12 by public transport, after having looked at the type of discount card each has. They each have a card that enables the students to travel for free at the date and time of that scenario. Then the remaining two students were selected from the Utrecht area with the guideline that they need to be in possession of the same card as the above students, and that they live near the Utrecht Centraal train station.

The output of TESS, in which only three of the required six students were given an advice, was said to be waste full because it did not plan students that were available even though they could work, thus making for a loss of money.

Expert C chose to send a total of four or five students from Utrecht by car. This car will be driven by student 6. A single student from Amsterdam is also selected. He/she will go by bike or public transport, with any cost for the student him/herself.

5.1.3 Scenario C

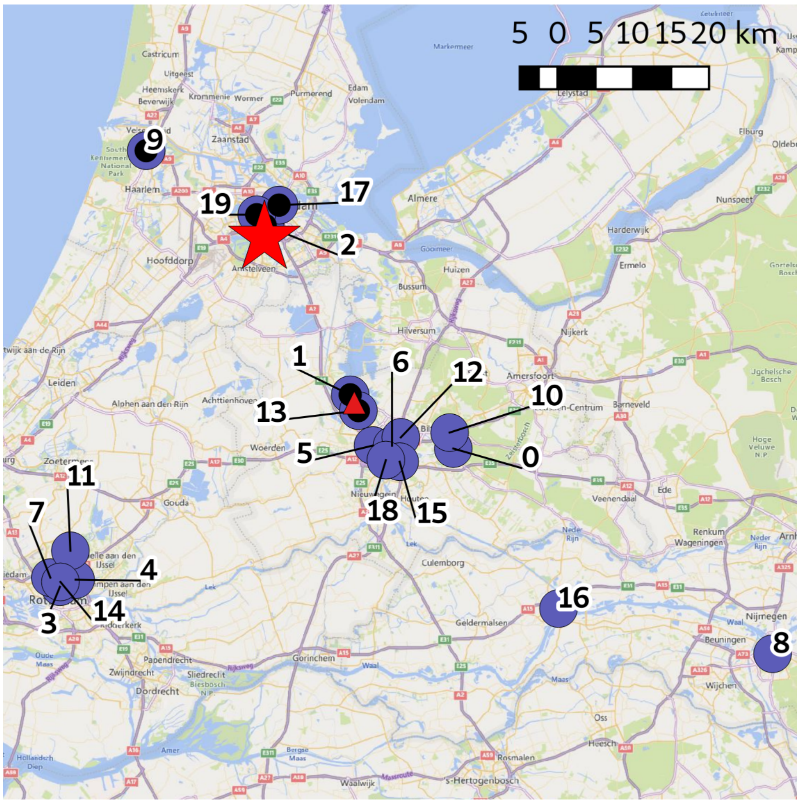


FIGURE 5.3: Duplicate of figure 4.5 (p.53). Twenty students are available, eleven are needed. The shift is twelve hours long and starts at 15:00hrs local time on Saturday 13th April 2019.

Expert A planned student 2, 17, and 19 by bike. The remaining students were planned by go by car from Utrecht. Again the car with a capacity of eight was used. Considering the late home bound trip it was uttered that students 1 and 13 would be dropped of along the way home, as they live near the Amsterdam-Utrecht route. Ultimately only seven of the remaining eight needed students live in Utrecht. For the final student no clear solution was given, as the conversation took of in the direction of the type of client, rather that the specific solution for this case. This is further discussed in section 5.2 (p.66).

On viewing the output of TESS it was commented that sending student 9 by bike for a bike ride of over 162 minutes at three o'clock at night after a 12 hour shift is not humane. TESS is less empathetic.

Expert B Planned the bus from Utrecht, with the comment that the driver would get paid an hour extra to bring each student home during the night. Students 0 and 10 were also considered, as they could take the car home for the night and bring it back the next day. This was not possible however because neither is in the possession of a drivers license.

Expert C Even though students from the area of Rotterdam are available, expert C commented that the first option always is from the Amsterdam-Utrecht area. It is the authors opinion that this is due to the experience of the expert.Expert C mainly has experience with planning in the Amsterdam, Utrecht, and Bussum area. Expert C checked the 9292OV.nl phone app to see what the options were to travel from the work location to Utrecht. There are none by public transport until 5:30 in the morning. Hence it was concluded that public transport to Utrecht for the home trip was not feasible. After this three groups were formed. Students 2, 17, and 19 were sent by bike. Students 5, 6, 13, 15, and 18 were sent by a car driven by student 6. Students 0, 10, and 12 would also go with a car driven by student 12.

On seeing the results of TESS it was commented that it was not so 'weird' that not all students from Utrecht were selected, even though this resulted in that less students that required were planned by TESS, as mentioned by the other experts. Reason being that this expert recognised the 'resistance' encountered when trying to plan students from Utrecht to go to work in Amsterdam.

5.2 General findings and Discussion

Internal validity Multiple experts asked about the compensation of travel time/distance by the client to LINQ. They indicated that for different clients different rules for travel time/cost compensation are negotiated. This had not been included a priori in the scenario's; the scenario's were designed with TESS in mind, and TESS does not take this variable into account. The decision for a certain rule set was chosen ad-hoc. The rule set for experts A and B is that of one client that pays fifteen euro travel cost per student, independent of actual made cost. For expert C the travel cost compensation was set to zero.

Variation in willingness to travel was mentioned by all experts. The different rules concerning the compensation of travel time/distance per client also results in different willingness to travel per individual for different clients. Hence ideally a model would provide different outputs for different travel time compensation rule sets ceteris paribus.

Experts also stated that some of their students do not mind travelling 1.5 hours by train as long as that student can travel with a friend. Others do mind. Overall it can be said that the willingness to travel varies greatly among individuals and shifts. This high variation between individuals is in line with results by Schwanen and Dijst (Schwanen and Dijst 2002) who found a high coefficient of variation, or relative standard deviation, of 0.743 in travel time ratio among individuals. Also, the willingness to travel might increase if some other shifts in that week are close to home. In other words, individual trips depend on other trips in the (near) future/past. This interdependence has not been modelled in TESS. Research about past and future trip interdependence also seems to be scarce.

On the other hand, the willingness to compensate travel time and distance from LINQ to the student also depends on the client. If a client is 'important' the planners are more willing to compensate a students travel expanses at the cost of LINQ, if this keep the client happy. For 'not important' clients this will be done to a lesser degree.

TESS is faster than the expert planners. Output was produced in all cases in tens of seconds, while the experts took longer as the scenario's increased in size. Notable was that expert planners seemed to employ a cluster based approach. A type of approach also employed by TESS. This also makes that for the experts planning time probably does not scale linear with the amount of students, but merely with the amount of clusters. Exact performance of the planning process remains a point for further research.

In such a research it might be found that there is a high variation among individual planners and the time of day, day of the week, and how well that planner slept that night. TESS does not get tired and hence performs consistently. She also does not quit her job, resulting in the retention of knowledge whereas with humans (uncodified) knowledge is lost through the coming and going of employees.

In theory the maximum amount of students that TESS can plan is higher than any planner can do. In theory TESS could plan for the entire country for hundreds of shifts in a matter of minutes. For a human planner this would be near-impossible.

Financial cost rather than temporal cost seemed to be the main factor that the experts used for planning. Considering that LINQ is a commercial firm this seems logical. If a student has to travel further than he/she is willing to do, he/she will not accept the shift. So rather than letting the planner do the selecting a priori, students select themselves a posteriori.

Group travel over solo travel One of the experts said that it is preferable to have a single group travelling by car with a reliable driver, who practically takes on the role of 'senior', over five students travelling solo via public transport. Also the travelling contributes to the 'team spirit' of the students. It allows new students to be mentored by the more experienced students. Something that would not happen if the new student had to travel solo. In the latter case the certainty of timely arrival at the client is lower than in the former case. This statement is based on experience, but logically the opposite might also be true. If this single bus has any trouble, the whole group comes later. Yet if every student travels solo, the chance of each of them being late seems lower as this would require nine, instead of one, delay-event. On the other hand, perhaps travelling in a group stimulates the students to come on time to not keep their colleagues waiting. Multiple mechanisms can be thought of; which is true should be investigated empirically.

The travel time budget calculation is based on empirical data which was concerned with a working duration of at most eleven hours. Using this same data for shifts with a duration of twelve hours might give unexpected and invalid results.

Planning is dynamic process that takes – and changes throughout – time. Most likely empoyee and employer related factors will change. Say that three students from the Utrecht region are available three days before the actual work takes place. A professional planner might anticipate that in the next day another two students somewhere from Utrecht will become available, and therefore reserve a car for the total of five students. Whereas TESS cannot handle uncertainty in the availability of work and students, nor can it anticipate future situations other than travel time.

5.3 Conclusion

This chapter tried to answer the question of *How do experts judge the implemented model?*

TESS ability to account for *changing travel cost throughout time* has been shown to be quick and more exact than human planners can do. While experts think in terms of 'rush hour', 'middle of the night', TESS can quantify this within seconds, even for larger groups. TESS also is able to balance *travel time versus financial cost*, something that the experts were not found to be doing at all. They simply assumed that any available student should travel to the work location. The question of 'if', rather than 'how' is not posed.

The *modality* and *carpooling* choices of TESS are not optimal yet. Going by car in the city of Amsterdam was deemed not optimal due to parking fees. Also not allowing students to carpool because they lived outside the B_{tt} did not always seem logical to the expert.

So generally it can be said that TESS is fast, exact, and constant in gathering and extracting information needed for the planning process; the actual planning itself is, for now, too complex to be exclusively handled by TESS.

6 Discussion and Conclusion

The four research questions have been answered in chapters 2 through 5. Furthermore there have been two research objectives;

- Formulate a conceptual model that can solve the TEP to support the efficiency of expert planners
- The implementation of this model should be sufficiently fast in terms of computational time

Both objectives have been met. The conceptual model as applied in TESS meets the second objective. The first objective has been met because although TESS might not give a perfect solution, it provides expert planners with a starting point for their planning.

The model, and thus TESS provide a solid basis for student employment agencies to start automating, or hybridising their transport planning. The model is not yet at a level to plan autonomously. Further fine-tuning of the model parameters should be the next step to get the model to become applicable.

This type of planning problem is highly complex. In this research and literature that is at its base simplifications of reality were made. This reductionist approach might be applicable to a limited degree on the process of planning, considering that the actual planning of students, who on their own make their work-travel-leisure decisions is highly complex. Adding more variables should thus be needed to increase the relevance of this model. Yet the question whether or not TESS will some day be able to perform the planning process completely is up to the reader and his/her philosophical stance on how man and machine will (co)exist in the future. For now a hybrid approach seems the most feasible; the exactness and fast data retrieval of the computer with the creativity and intuition of the human.

This research has been the first to formulate the travelling employees problem. In a way it is a 'new' problem, but due to the transdisciplinary nature and huge volume of related research as explained by Eksioglu, Vural, and Reisman (2009) this claim might not be justified. However it seems that the case of (student) employment agencies has not been investigated as literature on the subject is scarce or non-existent.

This research adds to the body of work that is related to graph theory and its wide applicability. Mondal and De (2017) give an overview of the application of problems that might be investigated using graph theory; here again the claim could be made that the TEP has never been investigated before, but it could also be claimed that the TEP is a problem somewhere in between a matching problem and the TSP. Although the problem as phrased here is specific to (student) employment agencies, similarities exist with other research as can already be concluded by the wide body of literature discussed in chapter 2. It also shows that these same concepts can be applied in a computer script, rather than remain mathematical and/or philosophical concepts.

But overall, this research resulted in TESS, and TESS will support student employment agencies in facing their daily challenge of planning travelling employees.

Bibliography

Beesley, Michael E (1965). "The value of time spent in travelling: some new evidence". In: *Economica* 32.126, pp. 174–185.

Bodin, L. et al. (1983). *Routing and Scheduling of Vehicles and Crews: The State of the Art*.

Carrion, Carlos and David Levinson (2012). "Value of travel time reliability: A review of current evidence". In: *Transportation research part A: policy and practice* 46.4, pp. 720–741.

CBS (2018). *Bevolking; geslacht, leeftijd en burgerlijke staat, 1 januari*. URL: https://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=7461BEV&D1=0&D2=a&D3=1-27,101-105,121-123,131&D4=l&HD=110621-1139&HDR=T,G3,G1&STB=G2 (visited on 04/11/2019).

— (2019). *Personen met een rijbewijs; rijbewijscategorie, leeftijd, regio, 1 janruari*. URL: https://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=83488ned (visited on 04/11/2019).

Chun, Yongwan and A. Griffith Daniel (2013). *Spatial Statistics Geostatistics*. SAGE, London.

Cormen Thomas, H. et al. (2009). *Introduction to Algorithms*. Cambridge, Massachusetts London, England: The MIT Press.

Cormen, Thomas H et al. (2009). *Introduction to algorithms*. MIT press.

DeSerpa, Allan C et al. (1973). "Microeconomic theory and the valuation of travel time: some clarification". In: *Regional and Urban Economics* 2.4, pp. 401–410.

Deza, Michel Marie and Elena Deza (2009). "Encyclopedia of distances". In: *Encyclopedia of Distances*. Springer, pp. 1–583.

Deza, Michel-Marie and Elena Deza (2006). *Dictionary of distances*. Elsevier.

Dijkstra, Edsger W (1959). "A note on two problems in connexion with graphs". In: *Numerische mathematik* 1.1, pp. 269–271.

Dijst, Martin and Velibor Vidakovic (2000). "Travel time ratio: the key factor of spatial reach". In: *Transportation* 27.2, pp. 179–199.

Eksioglu, Burak, Arif Volkan Vural, and Arnold Reisman (2009). "The vehicle routing problem: A taxonomic review". In: *Computers & Industrial Engineering* 57.4, pp. 1472–1483.

ESRI (2014). *ServiceAreaTask*. URL: https://developers.arcgis.com/flex/api-reference/com/esri/ags/tasks/ServiceAreaTask.html (visited on 09/13/2018).

Euler, Leonhard (1736). "Solutio problematis ad geometriam situs pertinentis (The solution of a problem relating to the geometry of position". In: *Commentarii Academiae Scientiarum Imperialis Petropolitanae*.

Evans, Gary W and Richard E Wener (2006). "Rail commuting duration and passenger stress." In: *Health Psychology* 25.3, p. 408.

Ford Jr, Lester R and Delbert Ray Fulkerson (1958). "Constructing maximal dynamic flows from static flows". In: *Operations research* 6.3, pp. 419–433.

Fosgerau, Mogens and Leonid Engelson (2011). "The value of travel time variance". In: *Transportation Research Part B: Methodological* 45.1, pp. 1–8.

Giuşcă, Bogdan (2005). *Map of Köningsberg in Euler's time showing the actual layout of the seven bridges, highlighting the river Pregel and the bridges*. URL: https://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg#/media/File:Konigsberg_bridges.png (visited on 10/04/2018).

Goodrich, Michael T, Roberto Tamassia, and Michael H Goldwasser (2014). *Data structures and algorithms in Java*. John Wiley & Sons.

Hägerstraand, Torsten (1970). "What about people in regional science?" In: *Papers in regional science* 24.1, pp. 7–24.

Hall, Peter et al. (1991). "On optimal data-based bandwidth selection in kernel density estimation". In: *Biometrika* 78.2, pp. 263–269.

Hamed, Mohammad M and Fred L Mannering (1993). "Modeling travelers' postwork activity involvement: toward a new methodology". In: *Transportation science* 27.4, pp. 381–394.

Hamerly, Greg and Charles Elkan (2002). "Alternatives to the k-means algorithm that find better clusterings". In: *Proceedings of the eleventh international conference on Information and knowledge management*. ACM, pp. 600–607.

Hart, Peter E, Nils J Nilsson, and Bertram Raphael (1968). "A formal basis for the heuristic determination of minimum cost paths". In: *IEEE transactions on Systems Science and Cybernetics* 4.2, pp. 100–107.

Hess, Stephane, Michel Bierlaire, and John W Polak (2005). "Estimation of value of travel-time savings using mixed logit models". In: *Transportation Research Part A: Policy and Practice* 39.2-3, pp. 221–236.

Hoinville, Gerald and R Berthoud (1970). "Identifying preference values: report on development work". In: *Social and Community Planning Research, London*.

Horowitz, Alan J (1978). "The subjective value of the time spent in travel". In: *Transportation Research* 12.6, pp. 385–393.

— (1981). "Subjective value of time in bus transit travel". In: *Transportation* 10.2, pp. 149–164.

Huisman, Otto and Rolf A de By (2009). *Principles of Geographic Information Systems, An introductory textbook*. ITC, Enschede, The Netherlands.

Johansson, Börje, Johan Klaesson, and Michael Olsson (2002). "On the non-linearity of the willingness to commute". In:

— (2003). "Commuters' non-linear response to time distances". In: *Journal of Geographical Systems* 5.3, pp. 315–329.

Johansson, Börje, Julian Klaesson, and Michael Olsson (2002). "Time distances and labor market integration". In: *Papers in regional science* 81.3, pp. 305–327.

Jones, Michael C, James S Marron, and Simon J Sheather (1996). "Progress in data-based bandwidth selection for kernel density estimation". In: *Computational Statistics* 11.3, pp. 337–381.

Kernel density calculations (2015). URL: http://www.geography.hunter.cuny.edu/~jochen/GTECH361/lectures/lecture11/concepts/Kernel%20density%20calculations.htm (visited on 10/08/2018).

Kitamura, Ryuichi (1990). "Trip chaining behavior by central city commuters: A causal analysis of time-space constraints". In: *Developments in dynamic and activity-based approaches to travel analysis*.

Kitamura, Ryuichi, Cynthia Chen, and Ravi Narayanan (1998). "Traveler destination choice behavior: effects of time of day, activity duration, and home location". In: *Transportation Research Record: Journal of the Transportation Research Board* 1645, pp. 76–81.

Laporte, Gilbert (1992). "The traveling salesman problem: An overview of exact and approximate algorithms". In: *European Journal of Operational Research* 59.2, pp. 231–247.

Lawler, EL et al. (2007). *The traveling salesman problem: a guided tour of combinatorial optimization*.

Levinson, David M (1999). "Space, money, life-stage, and the allocation of time". In: *Transportation* 26.2, pp. 141–171.

Lugano, Giuseppe et al. (2018). "D2. 2–Mobility and Travel Time Report". In:

Lyons, Glenn and John Urry (2005). "Travel time use in the information age". In: *Transportation Research Part A: Policy and Practice* 39.2-3, pp. 257–276.

MacKay, David JC (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Miller, Harvey J (2017). "Time geography and space-time prism". In: *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, pp. 1–19.

Mondal, Basudeb and Kajal De (2017). "An Overview Applications of Graph Theory in Real Field". In:

Patterson, Zachary and Steven Farber (2015). "Potential path areas and activity spaces in application: a review". In: *Transport Reviews* 35.6, pp. 679–700.

Piech, Chris (2012). *K Means*. URL: http://stanford.edu/~cpiech/cs221/handouts/kmeans.html (visited on 10/23/2018).

Pinter, Charles C (2010). *A book of abstract algebra*. Courier Corporation.

Redmond, Lothlorien S and Patricia L Mokhtarian (2001). "The positive utility of the commute: modeling ideal commute time and relative desired commute amount". In: *Transportation* 28.2, pp. 179–205.

Rez, A.H. (1832). *Der Handlungsreisende — wie er sein soll und was er zu thun hat, um Aufträage zu erhalten und eines glücklichen Erfolgs in seinen Geschäften gewiß zu sein — von einem alten Commis-Voyageur*.

Rui, Yikang and Yifang Ban (2014). "Exploring the relationship between street centrality and land use in Stockholm". In: *International Journal of Geographical Information Science* 28.7, pp. 1425–1438.

Schwanen, Tim and Martin Dijst (2002). "Travel-time ratios for visits to the workplace: the relationship between commuting time and work duration". In: *Transportation Research Part A: Policy and Practice* 36.7, pp. 573–592.

Sheather, Simon J and Michael C Jones (1991). "A reliable data-based bandwidth selection method for kernel density estimation". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 683–690.

Tong, Lu, Xuesong Zhou, and Harvey J Miller (2015). "Transportation network design for maximizing space–time accessibility". In: *Transportation Research Part B: Methodological* 81, pp. 555–576.

Turlach, Berwin A (1993). "Bandwidth selection in kernel density estimation: A review". In: *CORE and Institut de Statistique*. Citeseer.

Wachs, Martin (1976). "Consumer attitudes toward transit service: An interpretive review". In: *Journal of the American Institute of Planners* 42.1, pp. 96–104.

Watts, Laura (2008). "The art and craft of train travel". In: *Social & Cultural Geography* 9.6, pp. 711–726.