MSc thesis
Geomatics for the built environment

Assignment of walking trips to pedestrian network in the context of the 4-steps travel demand model

A macroscale approach of walking

Ioanna Tsakalakidou

TU Delft

ASSIGNMENT OF WALKING TRIPS TO PEDESTRIAN NETWORK IN THE CONTEXT OF THE 4-STEPS TRAVEL DEMAND MODEL

A MACROSCALE APPROACH OF WALKING

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science in Geomatics for the Built Environment

by

Ioanna Tsakalakidou

July 2019

Ioanna Tsakalakidou: *Assignment of walking trips to pedestrian network in the context of the 4-steps travel demand model* (2019)

Supervisors:	Dr.ir. Pirouz Nourian
	Dr.ir. Goncalo Correia
Co-reader:	Dr.ir. Martijn Meijers

ABSTRACT

In the transportation planning process, the Four Step Model (FSM) is used to define the needs and requirements of the transportation system within a city or a region. Despite its wide use, the model is focused on vehicular trips and fails to represent the demand of walking activity. The limited work that has been performed to address the misrepresentation of walking in the context of the FSM, still does not address the last step of the four step model, the assignment of the walking trips to the network. Stemming from this literature gap, and the need to enhance the role of walking activity within the transportation modeling, the main question of this thesis is to develop a method for the assignment of the walking trips to the street network.

This research suggest a methodology to model large continuous space and walking trips on it. Thus, the focus is on modeling walking space taking into account the pedestrian scale for the model. To implement that, the size of the spatial analysis zones of the FSM is adjusted to walking scale. Next, continuous walking space is defined by the BGT spatial dataset that offers information for the land covers. In order to model continuous walking space two space discretization techniques are compared; first, Constrained Delaunay Triangulation (CDT) is applied on the walking area polygon, second, a regular quadrilateral grid is overlaid with the walking area. Walking activity is modeled on the network that is formed with the dual graph of each discretization method. In a subsequent step, the trip injection is performed with zone centroids and connectors which were created with the point clustering algorithm DBSCAN having as input points the graph nodes. Finally, A* algorithm is used for deterministic an all-or-nothing network assignment.

The results of the applied methodology show that addressing walking within the FSM requires a finer grained approach due to the more localized scale of walking activity. Walking space needs to be defined as a continuous surface. In order to model such surface at the scale of the FSM, a discretization method is required to handle continuous space at large scale. Comparing the result of the two discretization techniques, polygon triangulation and regular grid, the triangulation dual graph results to the 'skeleton' of the polygon while the regular grid dual graph offers a more dense and homogeneous polygon representation. Additionally, in the trip injection process, the nodes of the regular grid perform better than the triangulation graph nodes. Overall, throughout the whole implementation, the performance of using a regular grid, outweighs the polygon triangulation as discretization method.

ACKNOWLEDGEMENTS

I was able to conduct this research with the invaluable support of some people. First, I would like to thank my first mentor Pirouz Nourian for the long conversations we had during these months and the stimulation for acquiring new knowledge and employ different approaches. Additionally, I would like to thank my second mentor Goncalo Correia who with his concrete knowledge and opinion helped me overcome obstacles I ran into. I also express a great gratitude to both of my mentors who assisted my aspiration to bring together different disciplines and to use the knowledge from Geomatics for another field, this of transport planning. A great thanks goes to Martijn Meijers who offered his knowledge when I encountered programming obstacles.

Finally, a great thanks is devoted to my friends Angeliki and Elissavet who were by my side during the whole time and to Thanasis who was always willing to offer his knowledge. Also to Maria, Alexandra, Konstantinos and Nikolas, who remotely supported me during my graduation. Last, my biggest gratitude goes to my family that is always by my side.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Algorithms

ACRONYMS

FSM Four Step Model 2
CA Cellular Automata 13
ABM Agent Based Model 13
BGT Basic registration Large-scale Topography 6
WFS Web Feature Service 29
CBS Centraal Bureau voor de Statistiek 28
PDOK Publieke Dienstverlening op de Kaart 28
WMS Web Map Service 29
OD Origin and Destination 5
DBSCAN Density Based Spatial Clustering of Applications with Noise 57
TAZ Transport Analysis Zone 3
TAZs Transport Analysis Zones 10
PAZ Pedestrian Analysis Zone 6
PAZs Pedestrian Analysis Zones 6
PIE Pedestrian Index of Environment 3
ABM Agent Based Model 13
DT Delaunay Triangulation 21
CDT Constrained Delaunay Triangulation 43
GIS Geographical Information System 2
VD Voronoi Diagram 19

1 INTRODUCTION

Walking activity within the built environment has drawn the attention of transportation planners, urban designers and other urban disciplines since many years. The value of walking as a mode of transport has been recognized in many aspects of urban living. It is reported that walking is beneficial for improving public health, air quality and the overall quality of life and shifting towards a more walkable city can yield health, social and economic benefits (Litman and Steele, 2017).
Knowing where people walk, how often, and under which circumstances provides important information to the urban and transportation planning process. More specifically, knowledge around walking activity offers insights for the transportation system of the area and the needs and requirements for pedestrian infrastructure. Additionally, the information about pedestrian activity reveals where people prefer to walk and what are the mostly 'used' parts of a city. Overall, the awareness around walking can contribute to a planning process that results in a more sustainable and livable city.
Walking is affected by many different factors such as the personal characteristics (economic status, vehicle occupancy, etc.), demographics (age, working status, etc.) and the distance between origin and destination point. Additionally, the components of the built environment like the street network, the land uses, the buildings and the overall city structure are important for the generation of pedestrian trips (see A.1 for more detail).
In the late years, with the shift towards more sustainable transportation, the interest around walking increased. This gave rise to a wide part of research regarding walking activity. In the transportation field, various methods have been employed for pedestrian demand forecasting like comparison studies, sketch plans, discrete choice models, regional travel models and pedestrian demand models, an overview of which is presented by Schwartz et al. (1999). Also a big part of research that focuses on the identification of pedestrian movement patterns and behavior within space has been developed Helbing and Molnar (1995); Helbing et al. (2001, 2002). In parallel, researchers have started to investigate how pedestrian activity is related to the surrounding built environment and how the last affects physical activity (Handy et al., 2002; Saelens et al., 2003; Stead et al., 2000; Litman and Steele, 2017).
The importance of walking is quite evident in the planning process, but as it turns out it is not easily quantified and thus it tends to be undervalued. Indeed, in the existing models the focus is shifted towards motorized travel and the available tools

for non-motorized trips (walking and cycling) present much less sophistication and detail than those developed for motorized travel (TRB, 2007; Clifton et al., 2013; Kuzmyak et al., 2014). This lack of representation of walking within transport modeling is due to the fact that for many years there was a lack of accurate and detailed information of walking behavior in the travel surveys (Clifton et al., 2013). Additionally, measures of the built environment regarding the densities, the land use and the transportation system were difficult to obtain due to the demand on spatial data (Clifton et al., 2013). Yet, even though the increase of Geographical Information System (GIS) has contributed to overcome these barriers, the existing models are still inefficient for treating walking. The absence of walking in travel demand forecasting has resulted to large scale design and planning (wide roads, low densities, parking infrastructure) that not only fails to support pedestrian activity but also acts as an inhibiting factor for choosing walking over other modes of transport. (Litman and Steele, 2017).

Among the existing travel demand models the most dominant and most commonly used is the Four Step Model (FSM) and is applied at a regional and urban level. Throughout the four steps of the model, the trip generation, trip distribution, modal split and network assignment (Figure 1.1), the future demand and performance of a transportation system is estimated by measuring the number of trips generated and attracted to and from the spatial zones within an area. In a subsequent step the total number of trips are distributed to the modes of transport. Within the four step model, walking is represented as one of the travel modes but the estimates are based on vehicular parameters, making the tool inaccurate for pedestrian trip prediction.

It could be argued that there are more suitable tools to represent pedestrian activity rather than the four step model which even though considers pedestrian trips its main purpose is the vehicular demand forecasting. Though, the four step models are used as a decision making tool for policy formulation and comparison of alternative interventions at large scale and their results are decisive for the distribution of resources across the needs of the transportation system (McNally, 2007). For this reason it is important that the transport analysis techniques frame in a comprehensive manner all the travel modes and the need for pedestrian infrastructure within a region is efficiently represented in the process of travel demand forecasting.

Consequently, developing a model that supports walking, is important for improving the knowledge on how pedestrians will respond to new facilities and infrastructure and to compare planning scenarios. Considering the value of walking in contemporary cities and the need to increase the focus of transportation planning on pedestrian activity, this thesis aims to develop a method to support the decision making process for pedestrian design and planning.

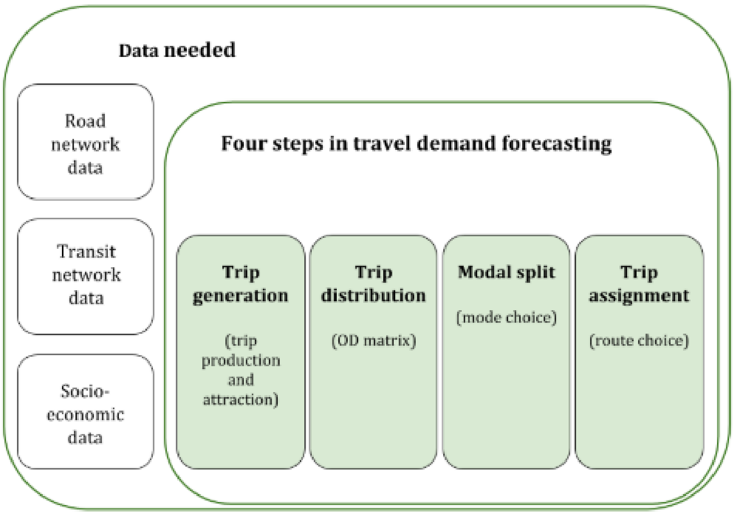


Figure 1.1: Components of the four step model (Woldeamanuel, 2016)

1.1 PROBLEM STATEMENT

The FSM is the main tool used to analyze the performance of a transportation system within a region and the definition of the needs in terms of facilities and infrastructure for the different modes of transport. Following the general tendency of the transportation analysis tools, the current design of the four step model reflects a mainly vehicle-oriented character and fails to support the aspects of walking. Despite its wide use, the model fails to represent efficiently walking activity (Clifton et al., 2016). More specifically, the large size of the Transport Analysis Zone (TAZ) tend to average out measures for pedestrians and to ignore internal trips (Clifton et al., 2016). Finally, the FSM does not reflect the evident effect of the built environment on the walking activity (Clifton et al., 2016).

The vehicle oriented character of the four step model is not widely acknowledged. Though, Clifton et al. (2016) recognize the limitations of the model and introduce a framework that enhances the existing four step models in order to better represent pedestrian activity. The advancements of their research include the introduction of finer-grained spatial zones and the use of a walk mode split model instead of the modal split step. Finally, a Pedestrian Index of Environment (PIE) is suggested so that the aspects of the built environment that affect walking are considered in the model. The research of Clifton et al. (2016) adds important value to the existing practices of travel demand forecasting for pedestrians. Yet, the model does not address the last step of pedestrian trip assignment to the network. Reviewing the limited literature regarding the assignment of walking trips to the network, a general lack of literature on the topic is recognized.

The misrepresentation of walking in the FSM results in a general lack of organized knowledge around walking activity in cities. A systematic modeling of walking would offer information on where people walk, how often, what time of the day and would reveal the most used and congested areas as well as the least used by pedestrians parts of the city. Consequently, introducing smaller scale in the model to study walking, would make the model more sensitive to the variables that affect walking. The output of such models, can be used for better pedestrian planning, improve safety and generally, would result to a better design and planning of the economic activities, the needs for transportation infrastructure and the overall city development.

Currently, the assignment of trips to the network is performed only for vehicular trips and expresses the path choice of the network users. More specifically, the trips are loaded to the network and the result reflects the spatial distribution of the trips to the network and the extent to which the transportation infrastructure is used. The network assignment of vehicular trips is addressed with multiple ways. The existing models range from macro-scale models that consider aggregate of vehicles into time periods , to micro-scale that model individual vehicles in very small time frame and also consider behavioral characteristics of the drivers (Bosa, 2017; Flügel et al., 2014). The road and transit network are represented by a graph consisting of links for each road segment and nodes for the road intersections. The selected route for each vehicle is represented on the network by a set of links or nodes.

It is obvious that the method for assignment of vehicular trips to the road network is already advanced and have specific steps. However, a process for assigning walking trips to the network has not been defined. Thus, the need to develop an approach to model walking space and trips within the context of the FSM arises.

Main purpose for implementing this step is to enhance the role of walking within the FSM and to fill the current literature gap on the assignment of walking trips to the network. In order to perform this, the goal of the thesis is to recognize the differences of walking compared to vehicular trips and to develop a method for modeling walking space and walking trips at the spatial scale of the four step model.

1.2 RESEARCH QUESTION, OBJECTIVES AND SCOPE

Considering the current limitations of the four step model to represent walking and the gaps on the existing knowledge, this thesis extends the knowledge on how to enhance the role of walking in the travel demand models. Especially, the focus will be on the last of the four steps, the assignment of the trips to the road network.

To perform this step, the first objective of this research is to specify the scale of the model. Walking is a smaller scale activity compared to vehicular trips thus, a different scale needs to be defined. Another objective is to define walking area. In vehicular modeling the road network is represented with a graph that consists of links and nodes that correspond to road parts and road intersections. On the other hand, walking is not performed with the same way and walking area may not be defined and represented with a network. The term "network" will be used from now on as a convention to denote the walking area representation, so that the term is aligned with the terminology of the existing FSM. To address these, the main research question of the thesis will be:

How to perform assignment of walking trips to the network to for the representation of pedestrian activity within the 4-steps travel demand model?

In order to answer the main question the following sub-questions need to be defined:

1. *What are the differences of pedestrian modeling comparing to vehicular modeling?*
2. *How is walking space modeled and how is pedestrian network defined?*

To clarify the scope of research the following subjects are relevant but fall out of the scope Figure 1.2:

1. The FSM is usually applied at regional and city level. Considering this, the spatial extent of the model should not be defined at microscale but at macroscale. This implies that no individual trips will be modeled but instead they will be considered at an aggregate level.
2. The step of assigning trips to the network a route choice model can be used to define the criteria of path selection. These can be parameters of the built environment, network characteristics, or personal preferences of the user. In this thesis aim is to model walking space and walking trips in space. Modeling pedestrian route choice is out of the thesis scope. This means that factors that affect network attributes, personal choice or behavioral characteristics of individuals will not be considered in the model. The route choice will be performed with a simplistic all-or-nothing approach.
3. The FSM consists of many sub-models in each step and each of these steps needs to be enhanced to better represent walking. Nevertheless, it is not aim of the thesis to address the preceding of the 'network assignment' steps. Thus, since input for the network assignment is the Origin and Destination (OD) matrix which is the output of the second step (trip distribution), the matrix will be created with random trip numbers based on the distances between the spatial analysis zones.
4. The developed model is implemented in the municipality of Delft. The application does not aspire to give any interpretations about the walking infrastructure of Delft, nor to estimate the future walking demand of the municipality. Doing so, requires to consider numerous aspects of the built environment

along with individual decision parameters. Since this is a first attempt to assign walking trips to the network, these aspects that are indeed important for modeling are left aside. Consequently, Delft is only selected as an input spatial dataset and features of the local walking infrastructure, the land uses and overall built environment characteristics are not introduced in the model.

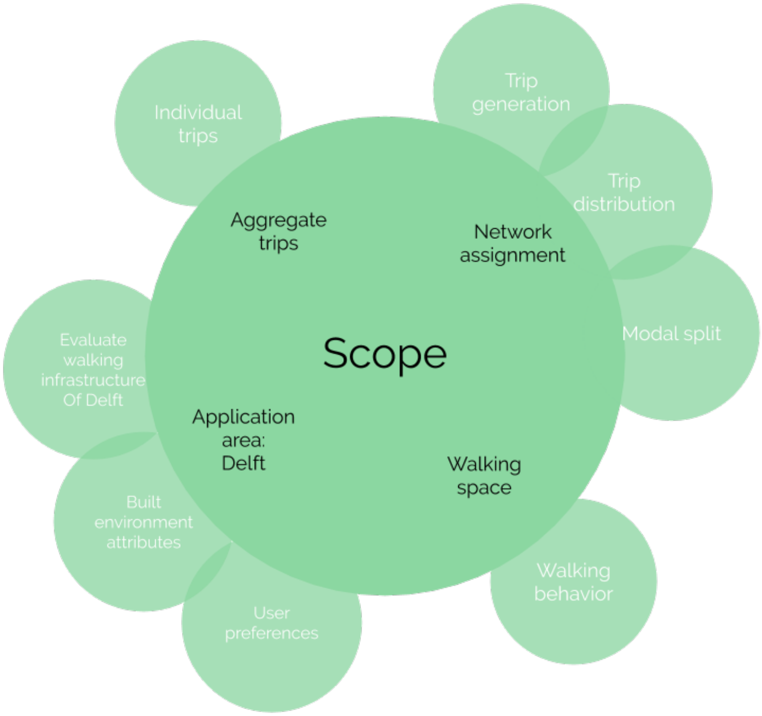


Figure 1.2: Scope of research and relevant topics

1.3 METHODOLOGY

The trip assignment to the network involves many individual steps (Figure 1.3). First, the required input datasets need to be specified. Such datasets are: the Pedestrian Analysis Zone (PAZ) which are geographic subdivisions of the area, the OD matrix that contains the number of trips starting and ending to each spatial unit and the spatial dataset that contains the geometric definition of the walking area (*walking_area* polygon). For the Pedestrian Analysis Zones (PAZs) a regular grid of $80x80m$ is used, a distance that corresponds to one minute of walking. The use of a regular grid to represent the PAZs has been introduced by Clifton et al. (2016). The OD matrix is filled with random numbers of trips based on the distance between each zone pair. For the walking area, the Basic registration Large-scale Topography (BGT) dataset is used to define the land covers that form the walking area.

Second, for the continuous surface modeling, two discretization techniques are implemented and compared; irregular triangulation network and regular quadrangular network. From each of the two discretization methods results a graph on which

walking is modeled. Triangulation is constructed with the use of *Python* library *Tri*. The regular network is implemented with a custom made algorithm.
Third, for the injection of trips to the network the zone centroids and connectors are created. In this process, the trips are allocated from the zone level (zone centroids) to the network nodes through the zone connectors. As zone centroid, the centroid of the walking area polygon within each zone is used. Next, the zone connectors are links that have as starting point the zone centroid and as ending point a graph node. These graph nodes are selected with a proposed approach based on graph nodes clustering, in order to recognize different regions of the walking area polygon. The point clustering is performed using the DBSCAN algorithm. A point from each cluster is connected with the zone centroid in order to form the zone connectors. Finally, the route choice for the trips that correspond to each origin and destination pair is performed with an all-or-nothing approach using the A* shortest path algorithm.

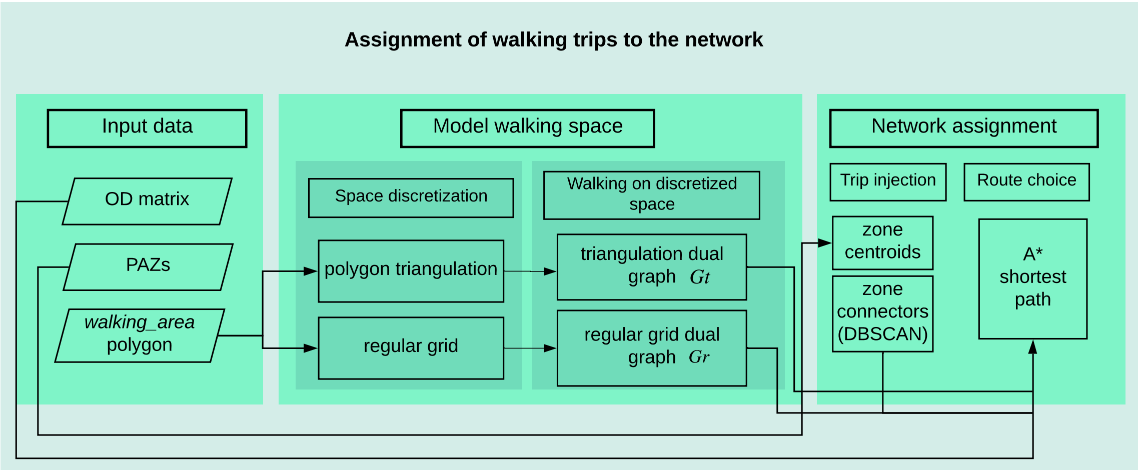


Figure 1.3: Workflow of followed methodology

1.4 THESIS OUTLINE

The thesis is structured in seven chapters in total. Chapter 2 includes all the literature review regarding transportation modeling and pedestrian modeling. In Chapter 3 the methods of modeling walking space and movement in such space are reviewed. Chapter 4 contains the description and pre-processing of all the input data, as well as the tools used. Chapter 5 is devoted to the implementation of space discretization, the allocation of the trips to the network and the final route choice. In Chapter 6 the results of the suggested procedure are presented and evaluated. Finally, in Chapter 7 the answers to the research question and sub-questions are provided and future work is suggested.

2 WALKING IN TRAVEL DEMAND MODELING

2.1 IMPORTANCE AND CHARACTERISTICS OF WALKING

Walking is the most common and simple form of human activity. People make walking trips in many parts of daily activities as main mode of transportation or as part of trip with another mode. Walking activity has reported benefits in various aspects of urban life like health, air quality, economy and the general quality of life. Thus, the last decades, there has been a turn of land planning and urban design policies towards the restriction of car usage in order to reduce urban sprawl, congestion, oil dependence, and climate change (Ewing and Cervero, 2010). Prioritizing walking in policies is crucial in order to reverse the effects of previous car-oriented policies that promoted the dependence of auto usage, reduced the access of short trip destinations and in general, have increased the overall vehicular trips (Frank, 2000).

Despite the importance of walking, it tends to be underestimated by transport planners and engineers. The rapid increase of car usage in the twentieth century directed all the transportation policies towards planning for automobile and public transport infrastructure, ignoring the non motorized travels. The are several reasons why walking was previously ignored (Olszewski, 2007). First, it is difficult to quantify and measure because pedestrian trips because unlike vehicular trips they are not always clearly defined. Second, walking infrastructure is considered low cost comparing to other transport modes and thus, it does not attract interest of large investments. Another reason is the difference of scales as walking is a more localised activity which can not be considered in parallel with the regional scale of vehicular trips. Finally, walking has a self-organising nature and there is a belief that it "will take care of itself". This means that it can take place even when the required infrastructure is not provided (sidewalks, paths, crossings, etc.).

Underestimation of walking has resulted to the shift of resources towards vehicular infrastructure (roads and parking), design for land use and facilities that reflects an orientation to large scale facilities (wide roads, low densities) and finally, leads to reduced pedestrian management and safety practices (Litman and Steele, 2017).

2.2 TRANSPORTATION DEMAND MODELING

Travel in theory is approached as the need of people to participate in activities (McNally, 2007). For the evaluation of the performance of a transportation system within a region and the estimation of the amount, the type and the distribution of the trips that are generated transportation models are utilized. Travel models aim to assess different alternatives in the transportation planning process which may include capital investments, policies land use planning and other aspects related to the transportation system (Castiglione et al., 2015). During the years of transport modeling various types of models have been developed like sketch plan models, strategic models, trip based and activity based models.

The primary travel demand model is the four step model which was developed in the 1950s and is the most widely used travel demand forecasting tool (Meyer and Miller, 1984). It is applied regional and sub-regional level and estimates the trips generated in the area for each travel mode. It is based on a discrete zoning system and the data are aggregated through zones, trips or activities. The measurements are based on a large amount of data: household demographic, socioeconomic attributes, employment totals for transportation analysis zones, measures of accessibility to employment etc. (Clifton et al., 2013).

More specifically, the model is structured with four sequential steps, the *trip generation*, *trip distribution*, *mode split* and the *network assignment* (Figure 1.1), each of which is represented with a sub-model. In the first step, the number of trips generated and attracted on each spatial unit and by all modes is measured. The measurements at this stage can be activity based or trip based. In the second step, *trip distribution*, the trips measured in the previous step are distributed to match the attractions and productions for every zone pair. Here, an impedance function is used to represent the travel behavior between each origin and destination pair. The third step of the model, *mode split*, is employed to measure how the trips are divided into the possible transportation modes (car, bus, bike, walking). Finally, in the last of the four steps, the *network assignment* the resulted trips from the previous steps are divided and assigned to the links of the network.

Before applying the model steps, some preliminary requirements and considerations are essential (Woldeamanuel, 2016). First, the modeled system includes all transport modes within a large area and needs to be segmented into smaller entities. For instance trips can be segmented based on the purpose into work based or non-work based. Another travel segmentation that can be applied is time based, splitting the trips into time periods (rush hour and non-rush hour) or into days of the week like weekdays and weekends.

Another prerequisite for the model is to divide the study area into smaller geographic sub-areas the Transport Analysis Zones (TAZs). The TAZs are used to approximate the location of origin and destination of each trip (Woldeamanuel, 2016). The definition of the transport analysis zone size should be performed in a way such that each zone reflects homogeneous travel behavior (Jafari et al., 2015). The zone size varies depending on the kind of application. It can start from the extent of a building block to larger spatial units like a neighborhood or a town (Sheffi, 1985). Aggregating the trips with the use of zonal system is necessary even in the case where the exact locations of a trip origin and destination points are known in order to avoid data availability issues (household data, travel survey data) and reduce implementation time (Qian and Zhang, 2012; Jafari et al., 2015).

The last prerequisite is the data. Socio-econonomic data like population, vehicle occupancy and employment are used to estimate how the trips are generated and distributed across the zones. Additionally, data that represent the network are needed. The final output of the model is the number of forecasted trips assigned to the network which expresses the demand of the transportation system in the application area (Meyer and Miller, 1984). The results of the model are used to forecast the impact of major transportation infrastructure decisions (McNally and Rindt, 2008). Generally, even though the the four step model has been the primary tool for demand forecasting for many years, and during these years advancements were made in the submodels, it is not the state-of-the-art. Its wide use is not because it has been the most suitable tool but for long time the only available (McNally, 2007). Yet, even though the activity based models provide better representation of the travel behavior the four step model is still dominating.

2.3 MODELING WALKING DEMAND

2.3.1 Walking in transport models

Walking as a mode of transport comprise a big part of human daily activity (Cole et al., 2006). Despite that, the need for modeling walking activity is more recent comparing to vehicular trips. Walking in transport planning, is modeled from the aspect of forecasting walking trips. Nevertheless, it is not efficiently represented by the existing models and the demand analysis is directed towards vehicular trips.

Including walking in transport models make the prediction of the travel demand more accurate for decision making and selection of policies (Rodrıǵguez and Joo, 2004). Additionally, the unbalanced representation of all the travel modes, reflects an unrealistic image for resource allocation, needs for access and mobility improvement and the needs of the transportation system in general (Rodrıǵguez and Joo, 2004).

There are many reasons why pedestrian activity is poorly or not considered at all in the travel demand models. Firstly, the spatial scale of motorized and non motorized trips is inherently different (large analysis zones for motorized trips versus small analysis zones for walking). So far, transport modeling is based on the use of large scale transportation analysis zones and high functional level of street classification (Clifton et al., 2016). Using the large zones applied for cars to estimate pedestrian trips, aggregates out aspects of walking since the small trips that are performed within a zone are neglected. Secondly, the need of different spatial scales requires additional travel surveys at a smaller spatial extent. This denotes a high demand of more detailed and accurate data on a disaggregate level. Another limitation of the four step model to represent walking lies in the fact that through the steps of the model walking trips are estimated based on vehicular parameters. This issue is evident in the trip distribution step in which the impedance of traveling between two zones is used to measure the trips between two zones. Even though the impedance is based on metrics like distance, travel time or generalized cost that differentiate among the different modes of transport this variation is not considered in the model. Finally, there are many theories that emphasize the influence of built environment on travel behavior and the selection of travel mode (Handy et al., 2002; Stead et al., 2000; Litman and Steele, 2017). Through the existing research there are many theoretical and empirical interpretations of how the structure, the road network, the land use system of an area interact with the pedestrians. Consequently, these parameters of the built environment need to be considered by the model and to be translated into practical solutions.

2.3.2 Overview of pedestrian modeling techniques

In the field of transportation, traffic models focus on the amount, type and distribution of vehicular and non vehicular - pedestrian- trips. These are statistical models that exploit existing data on pedestrians and are based on assumptions in order to make predictions for future conditions where no data are available (Raford and Ragland, 2005). Such models refer to all the modes of transport (metro, bike, car, bus, etc.).

Besides the travel models, there is a big portion of literature dedicated only to pedestrian movements. Part of research on pedestrians is focused on pedestrian flows and crowd dynamics and at a disaggregate level the behavior of pedestrians is modeled. Raford and Ragland (2005) based on the difference in scale of application, required inputs and expected outcomes, classify pedestrian models into sketch plan models, network analysis models and microsimulation models. Sketch plan models are applied at regional level for pedestrian volume estimation and are based on population statistics. In a more detailed scale, network analysis models, which are mostly a variation of the four step modeling approach, estimate pedestrian volumes on street level covering the neighborhood scale. Microsimulation models narrow down to a street or group of streets, intersections and open spaces, or even enclosed spaces like train and metro stations and focus on individual pedestrian movement based on behavioral characteristics (Raford and Ragland, 2005).

A different categorization of models for pedestrian movement is performed by Olszewski (2007) who distinguishes the classical models from the simulation models. Examples of classical models are the regression models and spatial interaction models. Simulation models differentiate from the aforementioned by introducing behavioral aspects to simulate movement instead of simplified theoretical assumptions (Olszewski, 2007). In this category numerous types of models have been developed with the most dominant types being the Cellular Automata (CA), Agent Based Model (ABM) and models based on ordinary differential equations which are mostly inspired by Newtonian mechanics and their logic is based in the consideration that attractive and repulsive forces are applied on the pedestrians.

2.4 PREVIOUS WORK ON WALKING IN THE FOUR STEP MODEL

The four step model has some limitations regarding the balance of mode representation as there is a mismatch between spatial requirements of each mode. The model has been mainly a model that focuses on vehicular trips and lacks parameters around walking due to the coarse spatial scale. Lately, Clifton et al. (2016) introduce a framework for representing the pedestrian activity in the four step model.

In their methodology they introduce finer grained analysis zones, the PAZs to better suit walking activity Figure 2.2. Additionally, they change the conventional sequence of the model's steps by applying a walk mode split model after the trip generation Figure 2.1. With the walk mode choice model the subset of walking trips is identified and separated bu the other modes. The vehicular trips are aggregated to the TAZ level in order to follow the trip distribution, the mode choice and the network assignment steps. Finally, they develop the Pedestrian Index of Environment (PIE) to associate the built environment issues with walking. PIE is introduced in the walk mode split step and uses the following six measures of the environment: activity (population and employment) density, transit access, block size, sidewalk extent, comfortable facilities, and urban living infrastructure.

The suggested framework for pedestrian activity of Clifton et al. (2016) and the conventional FSM can be applied in parallel:

The first step, *trip generation*, is applied at the PAZ level for all modes of transport. Before the second step, the number of walking trips is estimated with the walk mode split model introduced.

The second step, *trip distribution* or *destination choice*, is applied at the PAZ for walking and at the TAZ level for the other modes.

The third step, *mode choice*, is only used for the non walking trips and is applied at the TAZ level.

The last step, *network assignment*, is not addressed by the proposed framework.

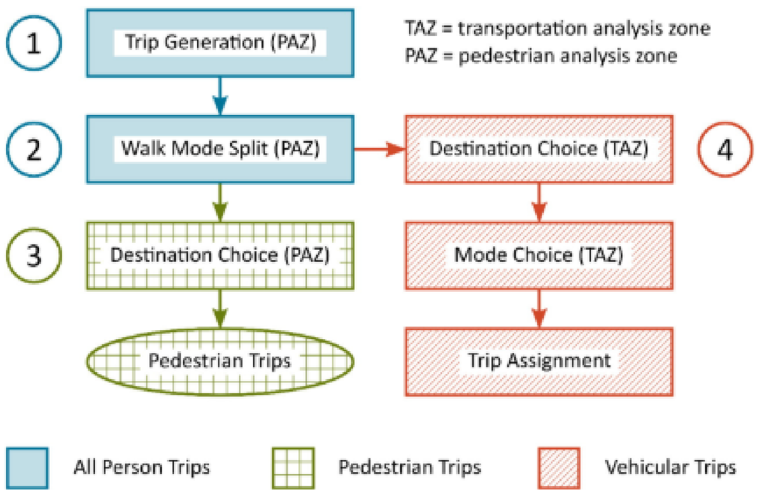


Figure 2.1: Conceptual diagram of proposed pedestrian modeling framework. Image source : Clifton et al. (2016), pg. 113

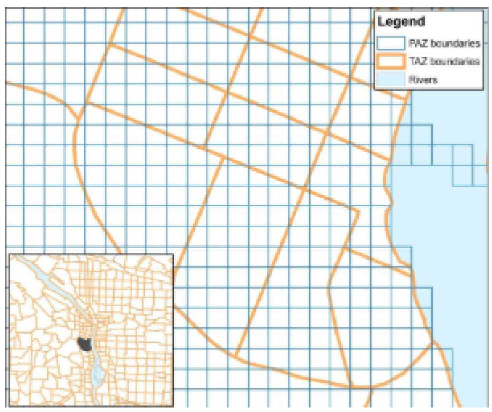


Figure 2.2: PAZs represented by a regular grid with a cell of 80x80 meters. Image source : Clifton et al. (2013), pg. 12

The enrichment of the four step model with walking parameters is an important progress for the pedestrian representation in the transport modeling. Nevertheless, while Clifton et al. (2016) introduce a methodology that involves three of the four steps of the model the final step of the model in which the trips are assigned to the network is not addressed. In this last step, the difference in spatial scale between motorized and non-motorized trips denotes the need to consider two different scales in the model. Assigning walking trips to the network requires a more detailed network representation than the network of vehicular trips and local characteristics of the network need to be considered.

2.5 NETWORK ASSIGNMENT IN TRAFFIC DEMAND MODELS

2.5.1 Network assignment of vehicular trips

Network assignment or route choice is the last step of travel demand models. This last step is not always integrated in the whole modeling process. The assignment of the trips to the network handles the allocation of the total number of trips performed in the area to the existing road and transit network. Overall, assignment can be performed with two basic approaches; either with the deterministic all-or-nothing method which is based on least-cost paths or with stochastic assignment based on probabilities of using a path (Schwartz et al., 1999).

All-or-nothing approach is very simplistic as it assumes the absence of congestion on the network and no differences in users' preferences and perception (Ortuzar and Willumsen, 1996). Not considering congestion practically means that the costs on the network are stable. Taking into consideration same user preferences and perception implies that for each pair of origin and destination points on the network all the trips will be performed on the exact same route with no deviations(Ortuzar and Willumsen, 1996).

In stochastic network assignment the trips are distributed on the network based on the probability of each route to be chosen over alternative routes. The probability of selecting a specific route is based the path utility which can be specified either by random utility models or by discrete choice models. Both random utility and discrete choice models are disaggregate behavioral models that focus on the route choice of individuals and reflect personal attributes and characteristics (Antonini et al., 2006).

All-or-nothing approach simplifies the real conditions of traffic since no congestion is assumed and consequently users only take the shortest path without searching for alternatives. On the other hand, stochastic network assignment is more realistic since congestion is considered along with attributes of the route and the user preferences. Yet, considering these parameters increases the time of collecting the required information and makes the model implementation more complicated.

Considering the difference in the resolution of the model, network assignment can be performed with simulations at a macro, meso or micro level. Macrosimulation models which perform static assignment on a continuum of vehicles, are deterministic models based on the relationships of speed and density of traffic on the network (Flügel et al., 2014). At these models the trips are aggregated into time-periods by zone pairs and not as individual vehicles (Bosa, 2017). The microscopic models are stochastic and are based on simulation of individual vehicles in very small time frames. The main difference is the consideration of driver behavior attributes and aspects of infrastructure design. In these models there is the possibility of data extraction at continuous temporal dimension. On the micro-level, network capacity is considered and micro-phenomena such as queues and bottlenecks are observed (Bosa, 2017)(Bosa, 2017). In an intermediate scale, the mesoscopic models employ a dynamic network assignment approach for a group of vehicles. They can be either stochastic or deterministic and merge the route choice capability of the macroscopic level with behavioral detail used in the microscopic level (Bosa, 2017).

2.5.2 Injection of trips to the network

The network assignment step, takes as input the OD matrix for each mode of transport which is produced in the second and third step of the model. At this point, the OD matrix provides information about which zone a trip starts from and to which zone the trip ends. Thus, the trips need to be allocated from the zone level to the network level, a process called trip injection. More specifically, the trips that correspond to each zone pair must be assigned to nodes of the network within the origin and the destination zone. The role of the *transport analysis zones*, the *zone centroids* and the *zone connectors* in the trip injection process is further explained below.

Transport analysis zones: The size of the TAZ affects the number of the assigned trips to the network. Having large TAZs reduces the number of intrazonal trips. This is a reason why the usually large size of the analysis zones used for vehicular trips is inappropriate for pedestrian trips. According to a research conducted in US for walking as a mode of transport, the results indicate that about 60% of people would not walk more than 1.6 km or 20 minutes for transportation purpose. Considering this, a larger than this distance size of zone eliminates the internal trips. Additionally, the presented interzonal trips are not realistic due to the large distance implied. Despite the advantage of the finer grained information and the detail that a small TAZ's size offers, more detailed data is required (surveys, population data, census data, spatial data) and more time to collect this data is needed.

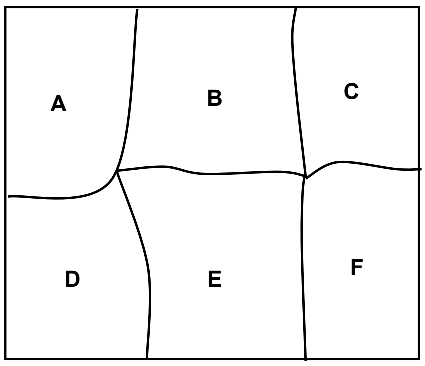
Zone Centroids: Nodes used to represent TAZs of origin and destination. The zone centroid can represent either the geographic center of the zone or the center of activities while in some cases it is manually located (WATS, 2008). For instance, when choosing to locate the centroid according to the activities of the area, if the activities are evenly spread throughout the zone, the centroid is placed in the center, while if the west side of the zone is an activity hub the centroid is located to the west side (WATS, 2008). Despite the different possible criteria of locating the centroid, the final location of the point has a small impact on the trip distribution because the actual effect of the point location is only a change in the distance of the centroid connectors (WATS, 2008).

Centroid connectors: Connectors are used to represent a path from a centroid to the network. When assigning trips to the highway or transit network, connectors are usually part of the local street network that are connected to the the higher hierarchy

road network. Thus, connectors should follow the overall road network structure and not cross other types of functions like buildings, water areas etc. Connectors' role is very important for the result of the trip assignment to the network, due to the fact that it is a cost-free alternative based on the local street network and may attract more trips (Jafari et al., 2015). Additionally, the number of centroid connectors affect the load of specific parts of the network. Particularly, fewer connectors may lead to concentrate trips at specific links of the network (Jafari et al., 2015). For large size TAZs 1-3 connectors are suggested to be used (WATS, 2008).
The trip injection process can become more clear through the example case in Figure 2.3 and Figure 2.4. The origin and destination matrix for the example area contains the number of trips for each pair of zones. From zone A to zone B 12 trips will be channeled to the graph starting from the centroid of zone A and ending to the centroid of zone B. In the same way, 17 trips will start from the centroid of zone B to end up to the centroid of zone A.

O \ D	A	B	C	D	E	F
A	0	12	32	4	25	8
B	17	0	11	22	13	16
C	34	25	0	17	20	22
D	56	34	21	0	14	9
E	14	10	19	32	0	4
F	11	5	3	1	12	0

(a) OD matrix



(b) TAZS

Figure 2.3: Example of OD matrix and TAZs

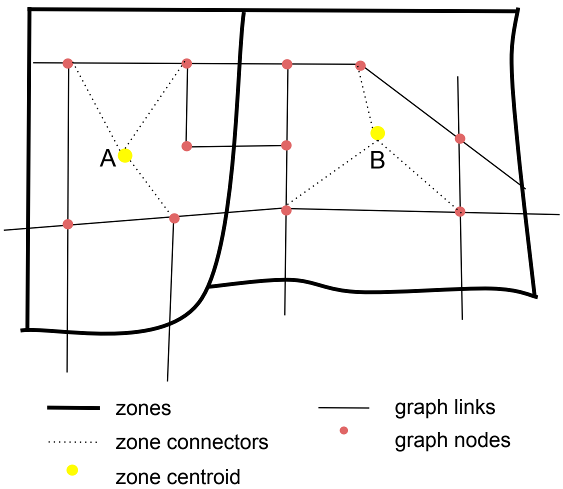


Figure 2.4: Use of zone centroids and zone connectors to inject the trips of the OD matrix to the road network in the example area

2.5.3 Network assignment of walking trips

While there is a large number of techniques to address this step of the model for vehicular trips, there is a gap in the literature for assigning walking trips to the network. To better understand the requirements of this step, it may be helpful to consider two sub-models that constitute the whole network assignment step. First, is the route choice model in which aspects that define the selection of a route instead of another are introduced. Such aspects may be parameters of the built environment, network characteristics and personal user preferences. The second sub-model is the space model, that describes the space on which walking takes place. In this subsection an overview of existing techniques for the route choice is presented. Techniques for modeling walking space will be presented in Chapter 3.

Pedestrian route choice has been studied out of the context of the four step model and mostly performed at microscale based on individual behavior. Borgers and Timmermans (1986a) suggest a route choice model that predicts the effect of retail stores on pedestrian behavior providing results for the profitability of shopping streets. Hoogendoorn and Bovy (2004), develop a theory of pedestrian behavior in which individuals are considered to schedule their activities, the activity areas and the selected paths route characteristics simultaneously in order to maximize the utility of walking activity. Antonini et al. (2006) present a pedestrian discrete choice model for walking behavior in which local interactions of individuals with other pedestrians are modeled.

Pedestrian route choice has not been applied within the four step model and at macroscale. Performing route choice for walking at large scale requires data that represent accurately the pedestrian network, and the physical features that affect walking trips (Liu and Andersson, 2004). These data requirements may have been prohibiting in the past due to insufficient datasets of pedestrian network, but currently this problem is resolved with the use of GIS (Clifton et al., 2013).

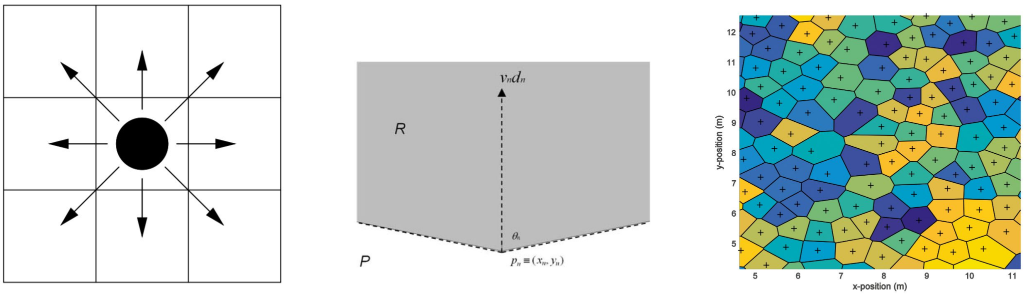
Nevertheless, previous work on pedestrian network assignment in the context and scale of the four step model has not been performed. Singleton and Clifton (2013) mention that among the Metropolitan Planning Organizations of US, no one perform pedestrian network analysis in their travel demand surveys. It is important to mention that there is an apparent absence of relevant research and application in Europe. Even though there is relevant work on route choice, models are applied at microscale leaving the macroscale unexplored.

3 MODELING WALKING SPACE AND PEDESTRIAN MOVEMENT

3.1 SPACE REPRESENTATION TECHNIQUES IN PEDESTRIAN SIMULATION

In the previous chapter a luck of modeling walking activity at large scale was recognized. This implies the absence of techniques for modeling walking space at macroscale. Despite that, there is a significant amount of research applied on modeling walking space at microscale. The need for modeling of walking space has emerged along with the need for pedestrian behavior simulation. Behavior of pedestrians has been studied for the prediction of pedestrian movements in cases like emergency situation, evacuations, crowd management, but also in occasional situations like traffic management, urban planning etc. Overall, two main approaches consider pedestrian movement; pedestrians seen as aggregate and flows of crowds are analogous to fluids; and pedestrians studied as individuals at a microscopic level (Antonini et al., 2006).

The method for representing continuous space varies in such models. For instance, in the social force model of Helbing and Molnar (1995) space is continuous and movement is modeled with a change in the location x(t). In cellular automata models physical space is discretized with a grid of cells which is overlapped to the walking area (Schadschneider, 2001) (Figure 3.1, (a)). Antonini et al. (2006) introduce an individual based discretization method to represent physical space based on pre-known position of each individual (Figure 3.1, (b)). Hoogendoorn et al. (2017) study pedestrian dynamics and use Voronoi Diagram (VD) to discretize continuous space. In this approach space is discretized based on the location of each pedestrian (Figure 3.1, (c)). Another approach is used by Borgers and Timmermans (1986b), in which space is represented with a network.



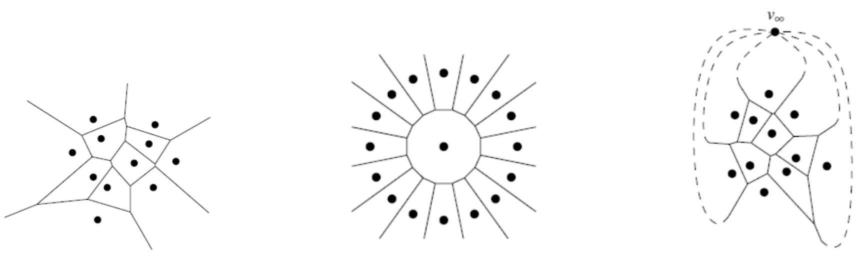
(a) (b) (c)

Figure 3.1: Space modeling techniques. (a) Cellular automata model. Space is discretized with cells. Source: Schadschneider (2001), (b) Space that corresponds to each individual based on direction and visual field. Source: Antonini et al. (2006). (c) Voronoi diagram of pedestrians. Each cell corresponds to a single pedestrian. Source: Hoogendoorn et al. (2017)

3.2 HOW TO MODEL CONTINUOUS SPACE

Walking trips take place towards all possible directions in the 2D surface thus, walking space is a continuous 2-dimensional surface. In reality walking space is part of a terrain that is described as a 2-dimensional surface in the 3-dimensional space. Considering that walking trips are performed on a continuous surface S which represents the walking area, practically means that the origin and destination points p_o, p_d belong to an infinite set of alternative locations within S. Consequently, modeling the 2-dimensional walking area requires discretization of the continuous space.

One way to do so is with the Voronoi diagram in which an area is subdivided into regions based on distance to a set of predefined points P, called sites. The sites may correspond to a point on a surface or may represent a polygon area. The VD results in N regions on the plane, one for each point in P such that the region of a site $p \in P$ contains all points in the plane for which p is the closest site. The regions of the subdivisions are the Voronoi cells (Figure 3.2) (De Berg et al., 2008).



(a) (b) (c)

Figure 3.2: Voronoi subdivisions.Source: De Berg et al. (2008)

Another way to divide continuous space into smaller continuous entities is to triangulate it. Based on a polygon, triangulation is the decomposition of the polygon into triangles by a set of non-intersecting diagonals drawn by the polygon vertices (Figure 3.3a). Additionally, triangulation can be applied on a finite set of points P, in which the triangulation of P is a planar subdivision of space into triangles whose vertices are the points of P (Figure 3.3b) (De Berg et al., 2008).

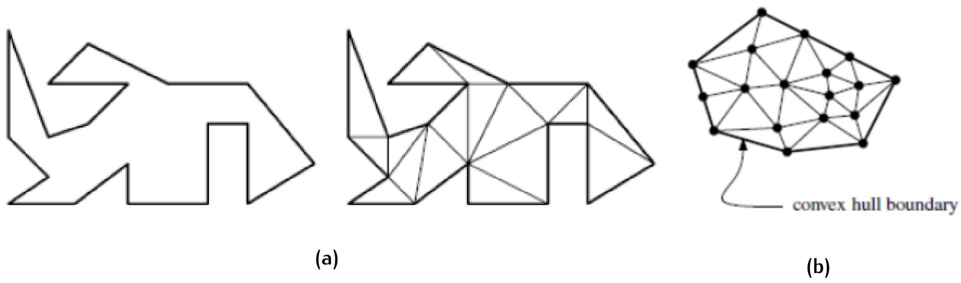


Figure 3.3: (a) Polygon Triangulation. (b) Triangulation of set of points. Source: De Berg et al. (2008)

A special case of triangulation of a set of points is the Delaunay Triangulation (DT). The Delaunay graph is the dual graph of the VD. This means that for every face in the Delaunay graph there is a vertex of the Voronoi diagram. The Delaunay triangulation is obtained by adding straight-line edges to the Delaunay graph (Figure 3.4) (De Berg et al., 2008).

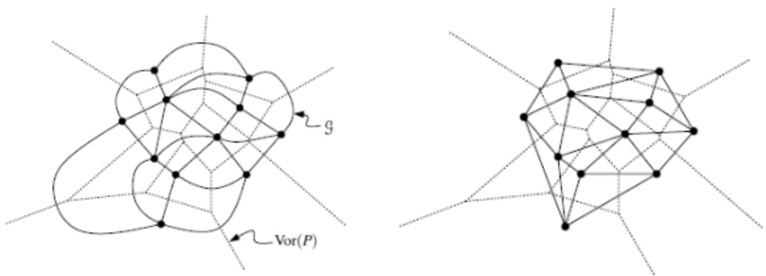


Figure 3.4: Left: Voronoi dual graph. Right: Delaunay graph. Source: De Berg et al. (2008)

Another approach for the representation of free space is by obtaining the trapezoidal map. It is used in the case where in free space exist obstacles or holes. The trapezoidal map of a set of non-intersecting line segments (that define polygon obstacle or hole), is constructed by drawing a bounding box that includes the line segments and draw two vertical extensions from every segment vertex. One vertical extension towards the top of the bounding box until it hits another line segment or the bounding box itself and a second vertical extension towards the bottom of the bounding box. The last step is to remove the trapezoids inside the obstacles or holes (Figure 3.5) (De Berg et al., 2008).

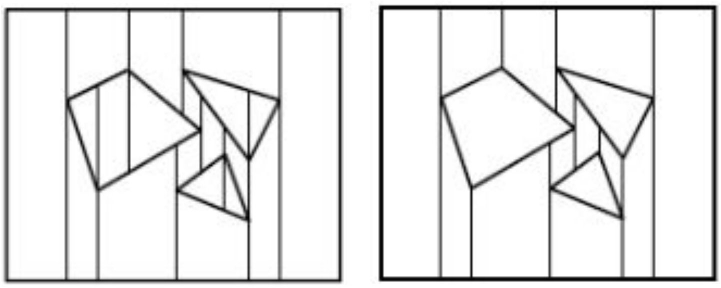


Figure 3.5: Trapezoidal map. Source: De Berg et al. (2008)

Finally, continuous space can be partitioned with a uniform or a non-uniform mesh. In such method, the space is divided into small regions or elements which are usually triangles or quadrilaterals. The difference between a mesh consisting of triangles and a triangulation is that the vertices of the mesh triangles are not required to belong to the input set of points. Extra points can be added called Steiner points. Steiner points of a triangulation are points that are not part of the initial triangulation edges and are employed to make the triangulation more uniform and dense. More information on the definition of these points is provided by Hwang et al. (1995). The resulting triangulation is called Steiner triangulation (Figure 3.6) (De Berg et al., 2008). To reduce the computation time in the applications non-uniform meshes are used.

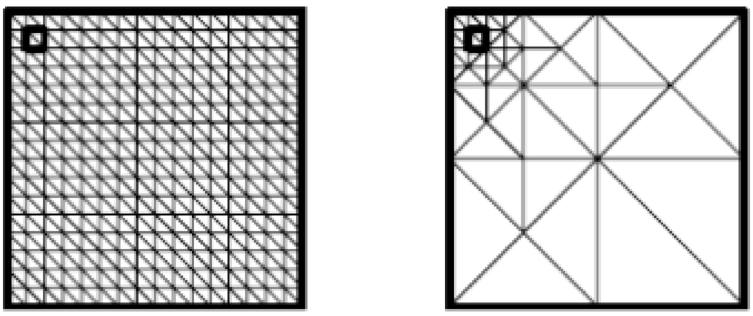


Figure 3.6: Left: Steiner points on a uniform mesh. Right: Steiner points on a non-uniform mesh. Source: De Berg et al. (2008)

3.3 MOVING IN CONTINUOUS SPACE

Having defined possible methods to discretize continuous space, it is time to elaborate on how to move in such space. The movement in a continuous surface is connected with the methods of discretization of continuous space. The existing techniques that simulate moving or walking are based on the space discretization methods described in Section 3.2.

More specifically, in robot motion planning, when designing autonomous robots it is important that the robot moves in space without having to say how to do it, and that it moves in such way that collisions are avoided. A simplification of this motion planning problem is to consider a point-robot with no dimensions. In order to find the path that the point robot will move on in free space the trapezoidal map of the previous section can be used. Based on this the road map is constructed (Figure 3.7) which is represented as a graph embedded in the free space with obstacles. The road map graph is based on finding the center of each trapezoid plane and the nodes on the midpoint of each vertical extension. An arc is created between two nodes if and only if one node is the center of the trapezoid and the other node lies on the boundary of the same trapezoid. The arcs are embedded in the plane as straight line segments (De Berg et al., 2008).

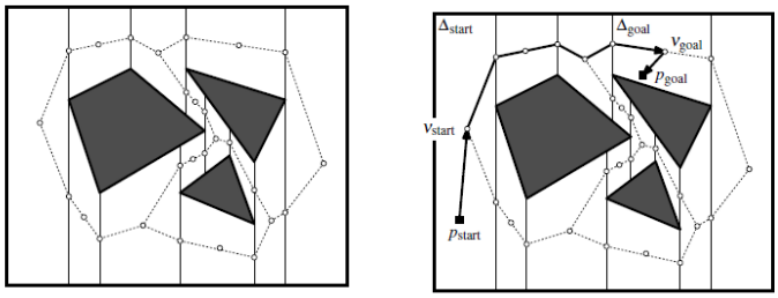


Figure 3.7: Road map of space with obstacles. Source: De Berg et al. (2008)

In case that shortest path from any starting point to an end point is needed obviously the road map is not sufficient in providing it in the free space. Any shortest path between a starting point Pa and an ending point Pb in a free space with obstacles is a polygonal path that passes through the vertices of the objects (Figure 3.8).

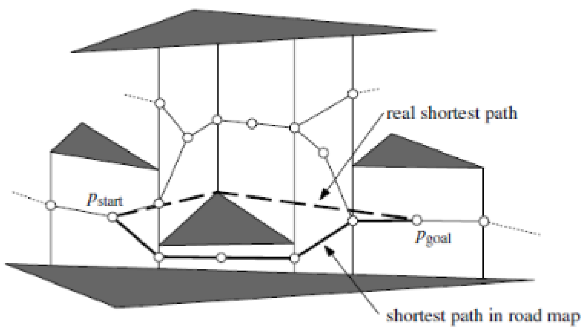


Figure 3.8: Shortest path on road map. Source: De Berg et al. (2008)

The shortest path on the road map is not the same with the real shortest path. The real shortest path is called visibility graph (Figure 3.9). More specifically, the visibility graph of a set of obstacles S is a graph whose nodes belong to S and there is an arc between two vertices v and w if they can see each other, which means that the line segment vw does not intersect with the interior of any obstacle in S. Two vertices that can see each other are considered mutually visible and the line segment that connects them is a visibility edge.

Finally, it is possible to represent moving in a continuous surface that is partitioned with a triangulation network or a mesh, by "walking" over the line segments of the triangulation or the mesh, or over the line segments of their corresponding dual graphs.

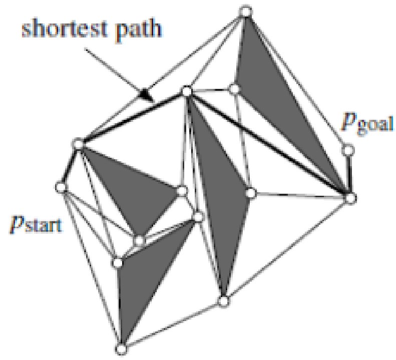


Figure 3.9: Visibility graph. Source: De Berg et al. (2008)

3.4 EVALUATING THE SPACE DISCRETIZATION TECHNIQUES FOR MODELING WALKING SPACE

Among the aforementioned techniques for discretizing continuous space and modeling walking, two general cases can be recognized. One is to have as input a set of points and the other is based on an polygon input. Point based discretization requires a predefined location of the input set of points that may represent walkers, activities or points of interest. Having a polygon input, requires that information for the area limits is provided and represented with the polygon boundary.

The visibility graph based on the trapezoidal map as a discretization method requires that there are obstacles or objects to be avoided at a known location. Additionally, the obstacles must be homogeneously spread into the area so that the whole area is approximated with the visibility graph. This means that in a case where there are few or no obstacles this method can not be applied or the space representation will be coarse (Figure 3.10).

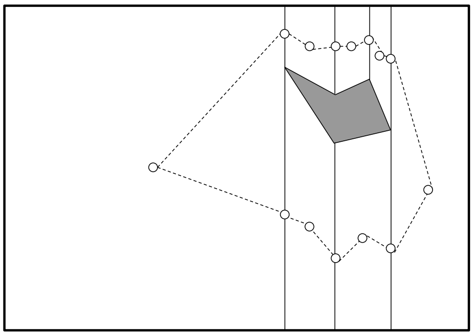


Figure 3.10: Trapezoidal map and road map on a polygon with one obstacle

The polygon triangulation technique for space discretization is merely based on the vertices of the polygon borders. In practice, this means that the triangulation vertices will lie on the polygon border and not in the interior of the polygon. Representing walking on the triangulation edges implies that part of the "walking" will be performed on the border. This can be avoided by considering the dual graph of the triangulation, which will ensure that moving only happens on the polygon interior.

Space discretization with a mesh either regular or irregular uses additional points inside the polygon resulting to a more dense space partitioning. Similarly, with the triangulation "walking" can be performed either on the edges of the mesh or on the edges of the dual graph of a mesh in order to avoid polygon borders.

The dual graphs of both polygon triangulation and triangular mesh are presented in Figure 3.11. Generally, the dual graph is not geometrically defined, thus, the triangle centroids are usually used to construct it. The triangulation dual graph G_t mostly represents the skeleton of the polygon while in the more dense discretization of the mesh, the dual graph G_m is a denser representation of the polygon area. Apparently, G_t is less costly in terms of storage and computation time comparing to G_m. Nevertheless, the scale of the application and the required detail in space representation are also important determinants in selecting which method to apply.

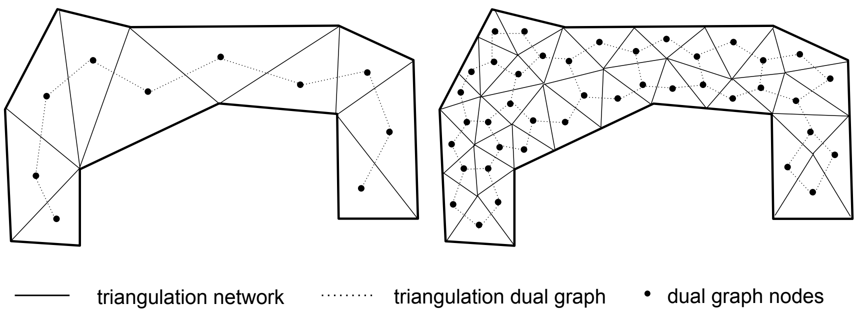


Figure 3.11: Left: Polygon triangulation and its dual graph. Right: Polygon triangular irregular mesh and its dual graph.

4 DATA PREPARATION AND TOOLS FOR THE NETWORK ASSIGNMENT

4.1 DEFINITION OF THE WALKING SPACE AND SCALE FOR THE APPLICATION

The application of modeling walking space and distributing trips will be based on spatial data. Before selecting which and what type of the dataset will contain the spatial definition of walking space, it is important to specify what constitutes a walkable space. Generally, the four step model is used to estimate the demand of a transportation system within a city or a region. Urban and regional scale is useful for large distance trips, and reflects the travel demand within and throughout the areas that are included. Walking is a more local activity and traveling at a municipality level or regional level does not take place on foot. This automatically denotes the need for a smaller spatial extent of the application area.

Apart from the size of the area the level of detail needs to be specified. In the four step model, the step of trip assignment to the network can be performed either on macro-level or on micro-level. At the macro-level the trips aggregated usually by spatial units, by time periods or by purpose of the trip. The micro-level simulations mostly focus on individual behavior. Considering the two different scales within the travel demand forecasting models, macroscale modeling provides information about the load of walking trips on the network of the whole area. On the other hand, microscale modeling could be applied to investigate localized effects of the infrastructure on walking. Within the scope of this thesis the assignment of pedestrian trips to the network will be performed at macro-level without introducing any individual behavioral characteristics. This comes to alignment with the main goal of this research which is to model walking space based on spatial data and not walking behavior which has already been widely researched.

Walking network unlike vehicular network is not explicitly defined. Car or transit network is represented by line segments on the local, collector and arterial roads. Also the junctions of the streets are represented with nodes. On the contrary, walking area unlike streets is not hierarchically classified according to function and capabilities of the area. Additionally, physical turns on the walking do not correspond to turn on the walking area as happens with the road network. Movement of pedestrians is performed freely towards all directions and does not take place on a straight line. So the space that corresponds to walking trips can not be presented as linear.

Based on these observations, the definition of the walking space for the model is a two-fold task. Firstly, it is important to describe the walkable area. This requires a semantical categorization of space through which the walkable and non-walkable areas are distinguished. Walkable areas are the public spaces like squares, pavements, pedestrian streets, stairs, bridges etc. Non walkable areas are the roads, the buildings, water surface, agricultural areas and so forth. Secondly, the geometric form of space needs to be specified. Both the semantic categorization and the geometric form are provided through the spatial datasets for the land use. The definition of walking space will be further elaborated in following section 4.2.

4.2 DATA REQUIREMENTS AND PRE-PROCESSING FOR WALKING AREA

The application of assigning walking trips to the network will be based on spatial data and more specifically on existing datasets that are open and free. Implementation area for the assignment of walking trips to the pedestrian walking space will be the city of Delft in the Netherlands. The exact borders of the study area follow the statistical subdivisions of municipalities into districts and neighborhoods (wijken and buurten). The municipality of Delft consists of 13 districts and 90 neighborhoods. Application area will be the central district (binnenstad) which includes 10 neighborhoods.
The spatial datasets are acquired from Publieke Dienstverlening op de Kaart (PDOK) which is the main provider of open free spatial in the Netherlands (https://www.pdok.nl/datasets). The borders of the administrative units for the whole country are available through PDOK website from the Centraal Bureau voor de Statistiek (CBS), the Dutch statistics organization. The name of the dataset used is 'CBS Wijken en Buurten 2018', from which the borders of the municipality of Delft are extracted (Figure 4.1).

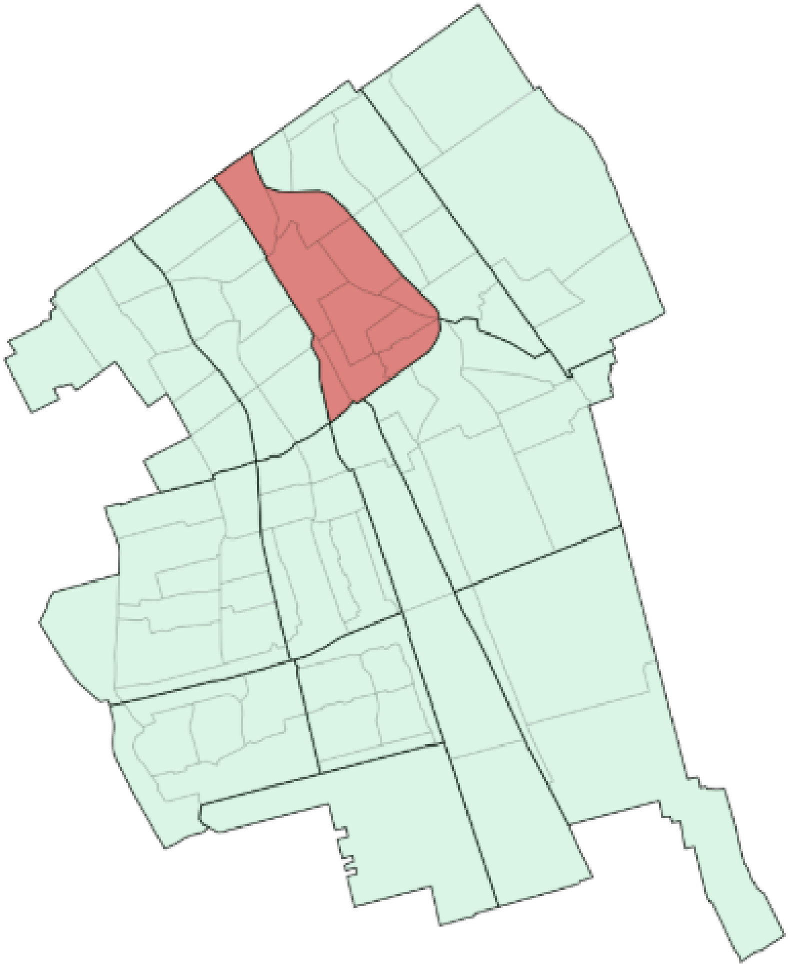


Figure 4.1: Area selection: Central district of Delft

The required information concerning the walkable areas in the city of Delft is provided through the BGT dataset. BGT (Basisregistratie Grootschalige Topografie) which stands for Basic Registration of Large-Scale Topography is a large-scale digital map for the whole country that provides the exact location of physical objects such as buildings, roads, water, railway lines and agricultural sites. The dataset is provided for downloading in gml format or through Web Feature Service (WFS) and Web Map Service (WMS) servers. The area coverage on the website is splitted into tiles that contain vector spatial features for various land objects. The features provided through the tiles do not follow the administrative borders. In order to keep the features contained in the municipality of Delft, the Delft administrative units were overlaid with the BGT dataset. The detailed classification of the physical object type is provided through the information layers shown in table Table 4.1. The existence of a detailed dataset with the classification of land cover is important for the semantic selection of walking space. The layer of interest is the 'wegdeel'='road part' through which detailed information about the road infrastructure is provided. The selected land covers are: 'footpath', 'footpath on stairs', 'pedestrian area' and 'local road' (Table 4.2). The later category is selected to ensure that crossings are included since the pedestrian crossings are not signified in the dataset. Generally, walking can decline from the pavement or a square and take place also in other land covers, like the bike lanes, the woonerfs or even in the private space among buildings. Nevertheless, the aim of the model is to focus on the use of walking infrastructure and consequently only such land covers are selected to form the final walking area.

Wegdeel (road part)
Waterdeel (water part)
Tunneldeel (tunnel part)
Overbruggingsdeel (Bridging Part)
Onbegroeidterreindeel (Uncultivated area Undeveloped Area)
Ondersteunendwegdeel (Auxiliary Traffic Area)
Functioneelgebied (functional area)
Begroeidterreindeel (overgrown terrain)

Table 4.1: BGT dataset classification

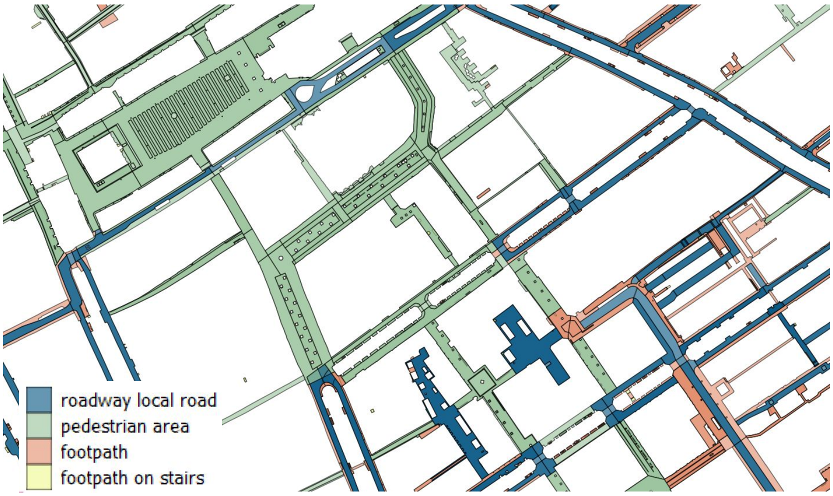
Wegdeel (road part)	parkeervlak	parking area
	voetpad	footpath
	rijbaan lokale weg	roadway local road
	voetpad op trap'	footpath on stairs
	transitie	transition
	rijbaan regionale weg	roadway regional road
	fietspad	bicycle path
	inrit	driveway
	OV-baan	OV track
	voetgangersgebied	pedestrian area
	woonerf	woonerf
	spoorbaan	railroad track
	rijbaan autoweg	roadway motorway
	ruiterpad	equestrian path
	overweg	Crossing, level crossing
	baan voor vliegverkeer	runway for air traffic
	rijbaan autosnelweg	roadway motorway

Table 4.2: Functions contained in Wegdeel dataset

Next, it is important to specify the form of these walkable areas. Geographic data follow specifications of Open GIS consortium (OGC). The objects in a geographic dataset are spatial objects or features that represent an object in reality and are defined as geometric sets. According to SFS (2010), spatial objects can be collections of Points, LineStrings and Polygons that are defined in 0, 1 and 2-dimensional coordinate space respectively. The BGT dataset represents the terrain functions with polygon features providing the geometric definition of walking space which will be 2-dimensional polygon geometric object (Figure 4.2). With a closer view on the walkable area (Figure 4.3a) it can be observed that the dataset consists of several polygon features. Each feature corresponds to a different object on the terrain represented by a polygon.

For the final walking area dataset the individual polygons are merged into one large polygon area. This polygon, named as walking_area will be the core representation of walking space for the trip assignment in this thesis (Figure 4.3b). The data processing and implementation steps are described in detail in Section B.1

Figure 4.2: Walkable areas in Delft center. *Visualised in QGIS*.



(a) Land uses that form the walking area

(b) Walking area uses merged

Figure 4.3: Definition of walking area polygon

4.3 SOFTWARE SPECIFICATION AND TOOLS

Throughout the implementation process multiple tools are employed for the analysis, computation and visualization including *Python*, *PostgreSQL,pgRouting* and *QGIS*.

Python (version 2.0 and 3.0) is a programming language that can be used to create processing tools. External libraries were used for multiple purposes. For the polygon triangulation the library *Tri* created by Martijn Meijers was used. Tri is developed to create the Delaunay triangulation and the Voronoi diagram of an input set of points. Comparing to other available libraries for triangulation, Tri is suitable when the under triangulation polygon is provided through a geographic dataset. The advantage of *Tri* is that it takes as input polygon shapefiles and the polygon edges can be used as constraints for the triangulation so that the polygon edges are preserved in the resulting triangulation. Another important advantage of this library is that the output triangulation contains also the adjacency information of the triangles. *Tri* is developed for *Python* 2.0. In Table 4.3 all the used libraries and the application domain is listed in alphabetical order.

Library Name	Python Version	Use
Csv	3.0	Import and export csv files
Fiona	2.0 / 3.0	Read and write shapefiles
Geopandas	3.0	Read, write and analyze geospatial data
Math	3.0	Provides access to mathematical functions
NetworkX	3.0	Creation and manipulation of networks
Numpy	3.0	Container of generic data
Pandas	3.0	Dataframe structure for manipulation and analysis
Pyshp	2.0 / 3.0	Reading and writing shapefiles
Psycopgr	3.0	Wrapper of pgRouting for route analysis
Psycopg2	3.0	PostgreSQL database adapter for Python
scikit learn	3.0	Identification of points clusters
Shapely	2.0 / 3.0	Manipulation and analysis of geometric objects
Time	2.0 / 3.0	Measure query execution time
Tri	2.0	Constrained Delaunay triangulation on the polygon

Table 4.3: Python libraries used

PostgreSQL is an open source object-relational database system. The spatial extension *postGIS* is a prerequisite for spatial functions with *PostgreSQL*.
pgRouting is an open source extension for *PostgreSQL* for geospatial routing functionality. It was used through the *Python* wrapper *Psycopgr* to compute the shortest path between every pair of origin and destination points. The algorithm used is the A* algorithm based on the distance.
QGIS is an open source, geographic information system tool used for result visualization.

4.4 TRAVEL DATA

4.4.1 Definitions for input data

The four step model is a data demanding model based on spatial and socio-economic data. Usually the data regarding the origin and destination of the trips per transport analysis zone are derived from already performed surveys (household travel surveys and travel activity diaries). The distribution of the number of trips to each

origin and destination pair of zones is implemented in the second step of the model. In this step, the trips are assigned to each pair of zones (i, j) using a gravity model that considers the size of each zone and the impedance (distance, travel time, generalized cost) of traveling between them. The output of this step is the OD matrix.
OD matrix represents the travel demand within the study area. More specifically, it contains the number of trips for each pair of origin i and destination j zone. It is a square matrix with dimensions equal to the number of zones in the area such that if N= number of zones, OD = (NxN) matrix. The step in which the OD matrix is produced is not within the scope of this research. Thus, the matrix will be created and filled with random trip numbers.
The definition of the matrix is based on the zonal system which divides the study area into smaller spatial units, the TAZ. As already mentioned, transport analysis zones are geographic subdivisions of a region into zones used to aggregate household and travel data as well as trips that start and end to each zone.
Thus, the application of network assignment for walking trips denotes the need for pedestrian scale zone size. The selected zone size follows the method of Clifton et al. (2016), in which they suggest as zones, a set of regular grid with cell size 80m which corresponds to the distance of 1 minute walking-what they call Pedestrian Analysis zones (PAZ) (Figure 4.4). The total number of PAZ in the study area is 276 zones, resulting to a 276x276 origin and destination matrix.

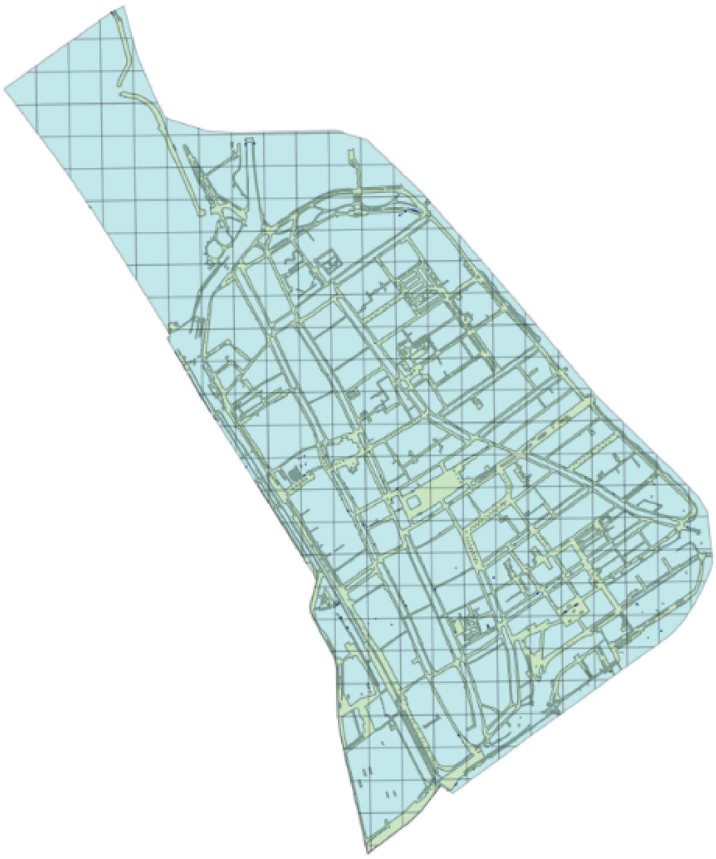
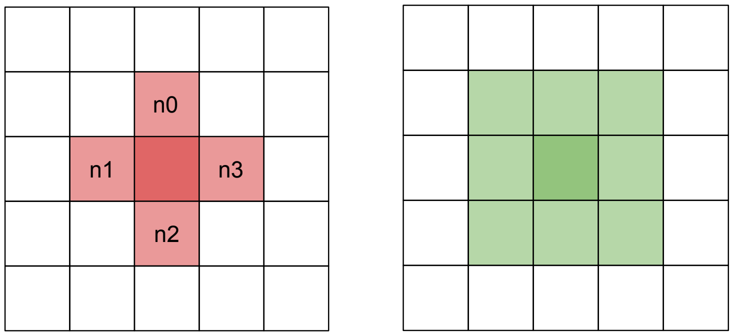


Figure 4.4: Pedestrian Analysis Zones (PAZ)- grid 80x80m

4.4.2 Creation of zones grid

The pedestrian analysis zones were constructed in Python 3 using the module Geopandas. The process of the implementation of the zones is based on the use of spatial data and consists of three main parts. First, the $80x80m$ grid was built in the rectangular bounding box of the municipality of Delft dataset Algorithm 4.1. The adjacencies of each cell were stored using the Von Neumann 4-neighborhood (Figure 4.5a) (Weisstein, 1666). The numbering of id definition is visualized in Figure 4.7. Second step is to intersect the grid with the borders of Delft municipality polygon. Finally, the removed cells, must be also removed from the list of the stored adjacencies.



(a) Von Neumann 4-neighborhood **(b)** Moore 8-neighborhood

Figure 4.5: Regular grid cell adjacencies

The final output for the grid is a Geopandas GeoDataframe in which for each cell defined by the column id, the column geometry, the u,v coordinates and the four neighbor ids are specified by the columns geometry, row, col, n0_id, n1_id, n2_id, n3_id respectively. Where no adjacent cell exists, value -1 is stored (Figure 4.6).

	geometry	id	col	row	n0_id	n1_id	n2_id	n3_id
0	POLYGON ((81699.99997400001 449899.999974, 817...	1	0	0	-1	-1	2	95
1	POLYGON ((81699.99997400001 449819.999974, 817...	2	0	1	1	-1	3	96
2	POLYGON ((81699.99997400001 449739.999974, 817...	3	0	2	2	-1	4	97
3	POLYGON ((81699.99997400001 449659.999974, 817...	4	0	3	3	-1	5	98
4	POLYGON ((81699.99997400001 449579.999974, 817...	5	0	4	4	-1	6	99
5	POLYGON ((81699.99997400001 449499.999974, 817...	6	0	5	5	-1	7	100
6	POLYGON ((81699.99997400001 449419.999974, 817...	7	0	6	6	-1	8	101
7	POLYGON ((81699.99997400001 449339.999974, 817...	8	0	7	7	-1	9	102
8	POLYGON ((81699.99997400001 449259.999974, 817...	9	0	8	8	-1	10	103
9	POLYGON ((81699.99997400001 449179.999974, 817...	10	0	9	9	-1	11	104

Figure 4.6: Geodataframe containing the adjacencies stored

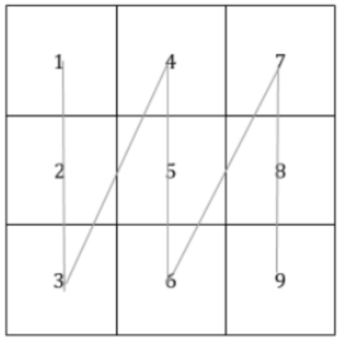


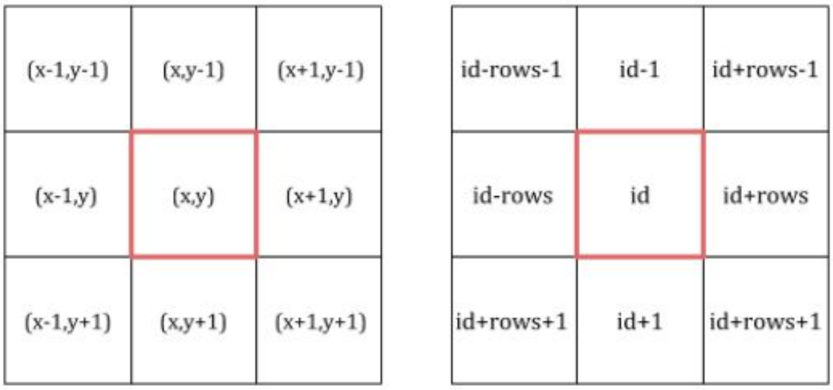
Figure 4.7: Cell id numbering

Algorithm 4.1: Creation of zones grid

```
Result: geodataframe (table): grid_cells.
Parameters: width equal = 80, height = 80;
xmin,ymin,xmax,ymax = total bounds of Delft center;
number_of_columns = (xmax-xmin)/width;
number_of_rows = (yxmax-ymin)/height;
Xleftorigin = xmin;
Xrightorigin = xmin + width;
Ytoporigin = ymax;
Ybottomorigin = ymax- height;
cell_geometry = [];
cell_ids = [];
cell_uv_coordinates = [];
neighbor1_id = [];
neighbor2_id = [];
neighbor3_id = [];
neighbor4_id = [];
id_counter = 0;
for column in [0,number_of_columns] do
    Ytop = Ytoporigin;
    Ybottom = Ybottomorigin;
    for row in [0,number_of_rows] do
        Polygon = [(Xleftorigin, Ytop), (Xrightorigin, Ytop), (Xrightorigin,
          Ybottom), (Xleftorigin, Ybottom)];
        Ytop = Ytop-height;
        Ybottom = Ybottom-height;
    end
    Xleftorigin = Xleftorigin + width;
    Xrightorigin = Xrightorigin + width;
end
```

4.4.3 Generation of Origin and Destination matrix

The creation of the synthetic data representing the number of trips for the OD matrix is implemented in Python 3 using the modules *Numpy* and *random*. The main logic for processing in the step is that the size of the whole study area does not denote walkable distances since walking trips do not take place in the whole municipality but only at smaller spatial extent. Thus, we define for each zone z, a neighborhood Nz of surrounding zones at a specific radius, so that having as origin the zone z, the destination zones are restricted only in the Nz. From now on in this section the words zone and cell will be used interchangeably denoting the pedestrian analysis zone. The neighborhood definition is based on the cells' adjacencies stored during the zone creation process which was described in Section 4.4.2.
The neighborhood Nz of each cell z is defined by using as measuring unit the cell of the zones' grid, which size corresponds to $80x80m$ or 1 minute of walking. This method is based on the stored adjacencies of each cell. In literature, the suggested method to access the adjacent cells of a seed cell is by the u, v coordinates of each cell (Figure 4.8a). Another similar approach introduced here is to use the ids instead (Figure 4.8b). The approach based on the cell id, has the advantage of using one integer value instead of a tuple of two integers, making the storage, manipulation, and access of the cell more efficient.



(a) (u, v) defined adjacencies **(b)** *id* defined adjacencies

Figure 4.8: Methods of accessing adjacent cells

The size of the neighborhood is defined as being proportional to the average walking distance of a person. Based on a research conducted in US about walking as a mode of transport, we use the maximum prefered by pedestrians 20 minutes of walking as presented by Watson et al. (2015). Consequently, the neighborhood of each cell is formed by all the cells that are up untill the $20th$ cell towards the left, right, top and bottom of cone z. This walking time is measured from the zone's border and not from the center which means that the time spent for walking within this cell is not considered for the 20 minutes of walking. The final neighborhood for each cell will be at maximum $21x21$ cells.

The logic of the algorithm for the construction of the neighborhood Nz of a cell z, is based on incrementally finding up to 20 vertical neighbors to both directions (above and below), and for each one of them, incrementally find up to 20 horizontal neighbors to both directions (left and right). The way that the algorithm traverses the cells is visualized in Figure 4.9. The pseudocode if the neighborhood creation is provided in Algorithm 4.2. The data structure that stores the neighbor cells for each z, is a dictionary with key the idz and values a list with the neighbors ids such that:

$$neighbor\_dict = \{idz : [n_1, n_2, n_3, \ldots, n_k]\}$$

Algorithm 4.2: Neighborhood creation

```
Parameters: size = 20;
cell_dictionary = {cell_id: [u, v]};
weight_dict = {};
neighborhood_dict = {};
for cell in grid_cells do
    cell_neighbors = [];
    cell_order = [];
    for index in range [0, size + 1] do
        if if index equal to 0 then
            for i in range [1, size + 1] do
                cell_neighbors append right_cell_id ;
                cell_neighbors append left_cell_id ;
                right_cell_order append i;
                left_cell_order append i;
            end
        else
            cell_neighbors append top_cell_id ;
            cell_neighbors append bottom_cell_id ;
            top_cell_order append i;
            bottom_cell_order append i;
        end
    end
    neighborhood_dict key = cell;
    neighborhood_dict values = cell_neighbors;
    weight_dict key = cell;
    neighborhood_dict values = cell_order;
end
```

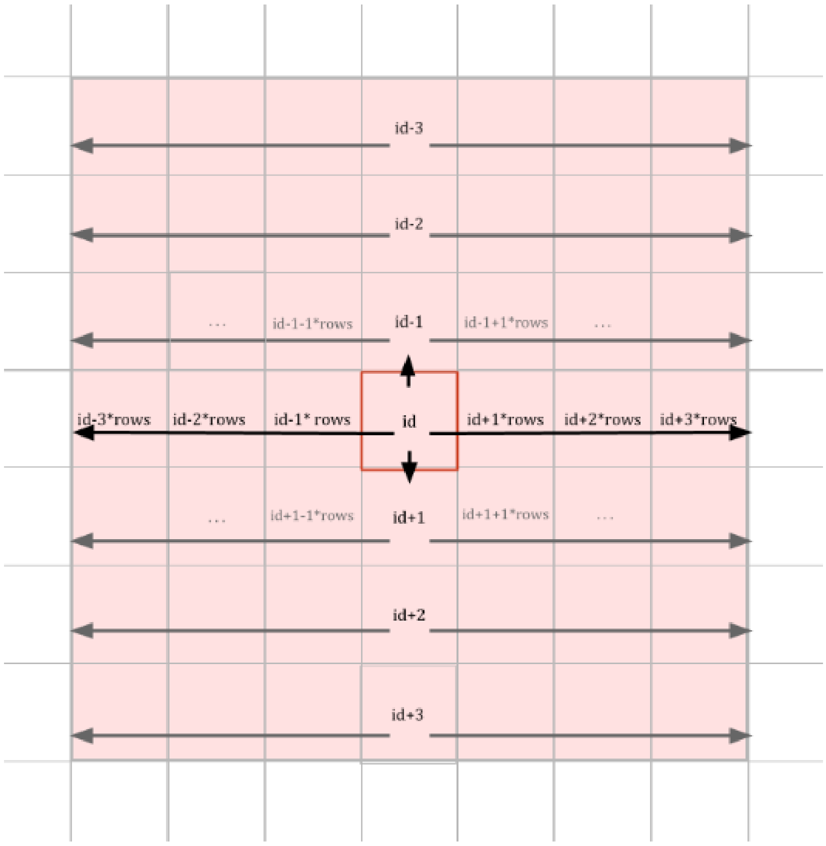


Figure 4.9: Visualization of traversing the adjacent cells to construct the neighborhood.

Having the neighborhood for each zone, the OD matrix is created with the Python module *numpy* based on two data structures. The first is the 1-dimensional list of ids which contains all the 276 ids of the zones sorted in ascending manner. The second is a zero square matrix with dimensions $276x276$. The two structures are connected with index correspondence such that the 10th element of the list refers to the $[10, 10]$ element of the zero array.

The OD matrix is generated by assigning random integers in the range of $[0, 100]$ to the columns of each row whose indices correspond to the ids list index of the neighbor cell Algorithm 4.3. Besides that, for each neighbor cell of the seed cell, a weight is assigned based on order of contiguity. This means that moving away from the seed, the first set of adjacent cells based on Moore's neighborhood (Figure 4.5b) are $1st$ order neighbors, the second set of cells contains $2nd$ order neighbors and so forth (Figure 4.10). The order of contingency affects the number of trips assigned to each cell with its neighbors, such that for each pair of the seed cell and a neighbor n, the contingency order of n with the seed is divided with the random generated integer. This produces more trips for the pairs that are closer together and less trips for the zone pairs that are farther away. Additionally, the zones that are out of the neighborhood rage of z are assigned zero trips.

Algorithm 4.3: OD matrix with random trips

Result: 276*x*276 OD sparse matrix

OD = zero matrix;
for *cell in cells_neighborhood_dict* **do**
    OD_cell = neighbor_order/random_number [range: 0-100];
end

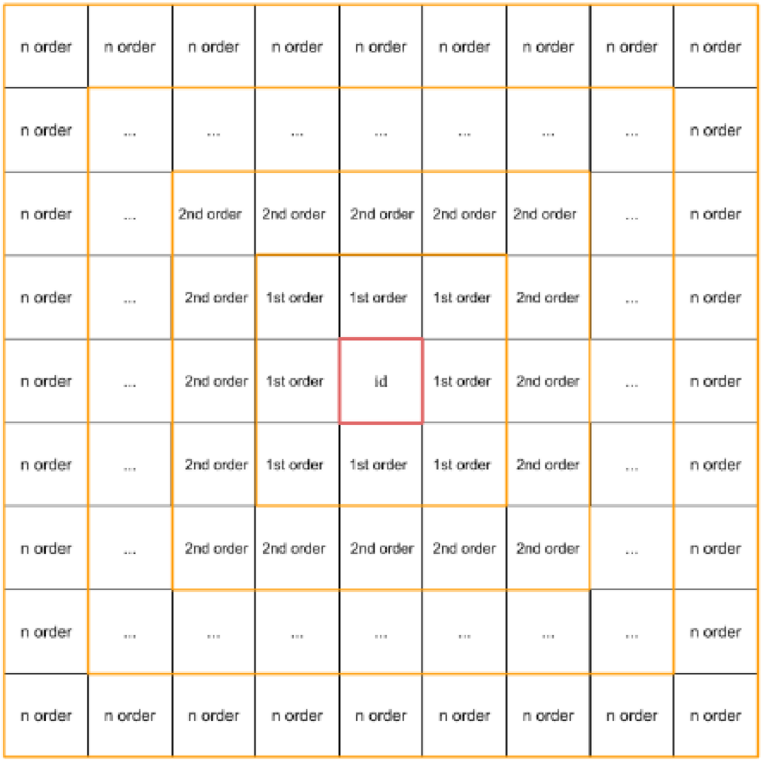


Figure 4.10: Order of neighboring cells

5 IMPLEMENTATION OF NETWORK ASSIGNMENT

5.1 SELECTING METHOD FOR MODELING WALKING SPACE

In this thesis we aim to develop a methodology for modeling walking trips in continuous space within an area at the spatial extent of a municipality. In Chapter 3 a literature review on available methods to model continuous space were presented. For choosing a methodology of modeling continuous space, there are some criteria that need to be considered. Firstly, the FSM is usually applied at a large scale as it also considers vehicular trips that are performed on bigger distances than walking trips. This means that not only the size of the implementation area is big also the spatial dataset covering the whole extent of the area is big in terms of storage. As an outcome, large datasets used as the inputs and outputs require more computation time especially when using algorithms with high complexity. Complexity of a procedure or an algorithm gets higher when for applying the same process for a bigger set of data, the time increases in a non-linear way.

Secondly, the coarseness of representation will be considered. As mentioned in Section 3.4 the various techniques presented offer different representation density of the polygon. The triangulation dual graph and the road map represent the skeleton of the polygon, while the mesh offer a denser space representation.

Thirdly, an automated procedure that can be applied to other cases and areas is essential for the method developed. Automation means applying existing algorithms made available through tools and libraries which would be complex enough to be customarily developed under the scope of a thesis. Treating continuous 2-dimensional space at large scale, requires discretization into smaller 2-dimensional entities. Modeling a small continuous physical space without discretization can be implemented mathematically using a function to describe the change of a person's location. The boundaries of the potential space could be defined through the domain space of the function. Nevertheless, such an approach is not applicable for a very large walking area that is defined at a municipality or regional level and discretization is required.

Considering the performance of method in terms of space representation, complexity and the automation of the procedure, we will compare two possible methods for treating continuous space, polygon triangulation Section 5.2 and regular grid representation Section 5.3. The polygon triangulation is constructed with automated way. For a dense representation, ideally a mesh would be constructed. Nevertheless, the available tools for mesh construction are not compatible with geographic spatial datasets like the *walking_area* polygon shapefile. Consequently, a regular grid will be customarily constructed and overlaid with the walking area polygon. Each of the two methods includes the progression of different approach and particular steps and process. Details on the two methods are provided in the following sections.

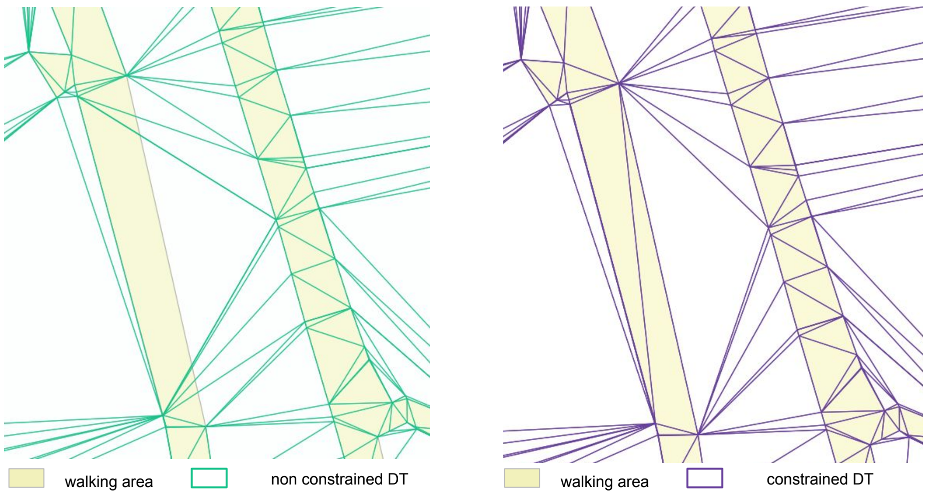
5.2 SPACE DISCRETIZATION WITH POLYGON TRIANGULATION

5.2.1 Implementing polygon triangulation

The triangulation of the *walking_area* polygon is based on the polygon edges and vertices. As a result, the internal boundaries need to be removed and instead keep only the geometry that forms the shape of the area. To achieve this the polygon features are merged into a single polygon. The output merged polygon as presented in Section 4.2, is the input for the polygon triangulation.
Merging a set of features is the combination of multiple features of the same layer into one feature. This process applied on the whole *walking_area* of Delft, results to a single polygon that is complex due to the many holes and concavities it has. Additionally, even though merging results into a representation of the area that seems to be correct, the geometric definition of the polygon is not valid. According to SFS (2010), a valid polygon among others, is defined by the outer border with an exterior ring (LinearRing) and if holes exist, with interior boundaries. Merging process results in wrong definition of these boundaries making impossible to handle the polygon's boundary and holes.
Despite the lack of information about whether the polygon vertices and edges are part of the exterior or interior boundaries, the information that the vertices and edges hold can be still utilized. Moreover, spatial overlay functions such as intersect, contain, within etc. can be successfully performed meaning that there is still functionality even with an invalid boundary definition.
There are multiple alternative python libraries that implement polygon triangulation. For instance, *Triangle* from Jonathan Shewchuk is a powerful library that constructs Delaunay Triangulation. Through this library the possibility to construct either constrained or constrained conformed triangulation is provided. Nevertheless, Triangle is not designed and aligned with geographic datasets. Consequently handling the polygon of the *walking_area* with complex geometry and invalid definition makes this Python library improper. Another available tool is the *scipy.spatial* library, using the module *Delaunay*, that performs polygon triangulation given a set

of input points (polygon vertices). Yet, the resulting triangulation does not consider the edges of the given polygon.

Most suitable for our purpose is the Python library *Tri*, that is built to construct the Delaunay triangulation of a polygon having as input geographic data. Using as input the merged polygon in a .shp format a Constrained Delaunay Triangulation (CDT) is created, having as input the polygon vertices and as constraints the polygon edges. The advantage of being able to introduce the polygon edges as constraints to the triangulation process is that it ensures that the all the polygon edges will be part of the triangulation edges. The difference of a constrained Delaunay Triangulation and a non constrained Delaunay Triangulation becomes more clear in Figure 5.1.



(a) Non constrained Delaunay triangulation created with *scipy.Delaunay* in Python

(b) Constrained Delaunay triangulation created with tri.Delaunay in Python

Figure 5.1: Methods of accessing adjacent cells

The result of applying CDT in the area is presented in Figure 5.2. The next step is to remove the triangles that are not inside the polygon. This requires spatial overlaying of the triangulation dataset with the walk area polygon. To reduce the implementation time and the computation load considering the large dataset, instead of overlaying the two polygon geometries (triangulation and walk area polygon), the centroid of each triangle is overlayed with the walk area polygon. A workflow of the implementation of these steps is provided in Section B.2. The final triangulated walking area is shown in Figure 5.3.

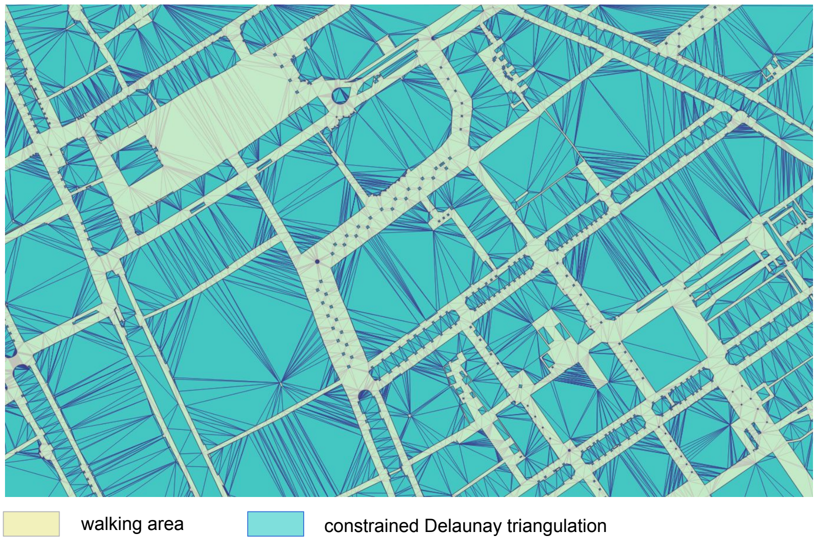


Figure 5.2: Constrained Delaunay triangulation in application area

Figure 5.3: Walking space discretization with polygon triangulation

5.2.2 Path planning in the triangulated polygon

After discretizing the continuous surface with triangles we need to define the way to move from one point of the surface to another. Traditionally in traffic assignment, trips are assigned between a pair of origin and destination points based on the shortest path. Further on this we will elaborate in Section 5.4.3. Consequently, when moving on discretized continuous space it is important to define the possible paths.

Moving on a triangulated surface can be implemented either by 'walking' on triangles' edges or by moving from one triangle to another. The first option requires that the path between the starting and ending point will be composed by triangles' vertices. It also implies that since the borders of the walking polygon are also part of the triangulation edges, part of a route or possibly a whole route will be on the polygon border. To avoid this, the walking routes are represented by the dual graph of the triangulation. The dual graph is topologically defined by representing each triangle face with a vertex and connect the vertices of triangles with an edge when two faces share a common edge (Figure 5.4). Therefore, the topological relations will be the same with the triangulation so that each vertex that represents a triangle will have the same neighboring vertices with the corresponding triangle neighbors. The geometric construction of the dual graph requires the specification of the location of a point that represents each triangle. Such a point that corresponds to each triangle and can be easily defined is the centroid of a triangle. The resulting the dual graph is presented in Figure 5.5. The algorithm for implementing the graph is presented in Algorithm 5.1.

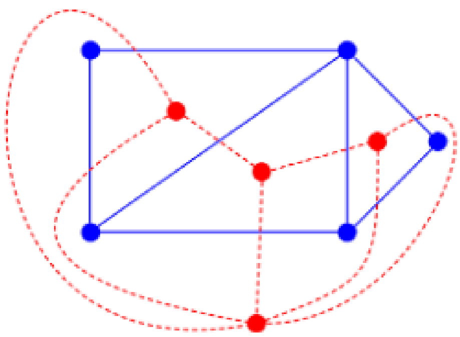


Figure 5.4: Dual graph (source: wikipedia)

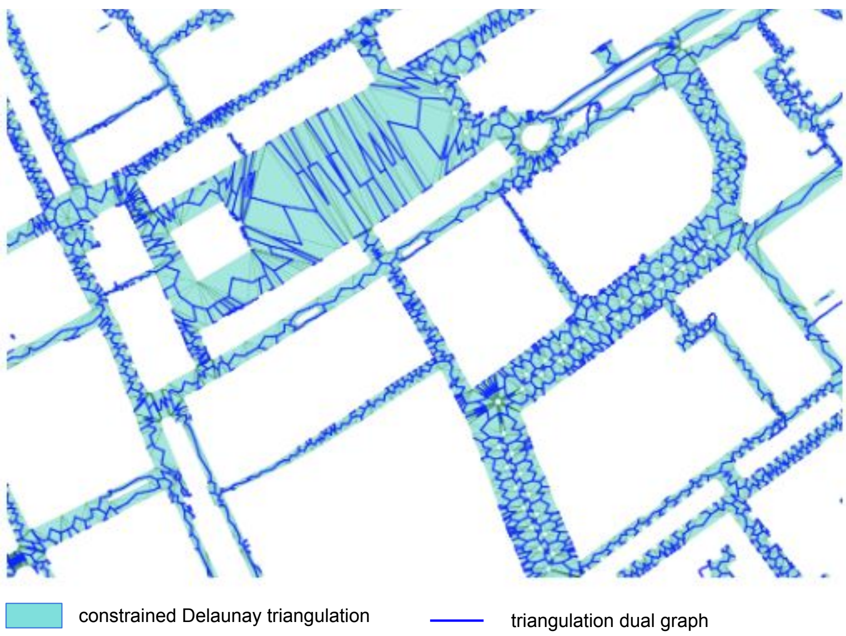


Figure 5.5: Dual graph of the triangulated walking area polygon

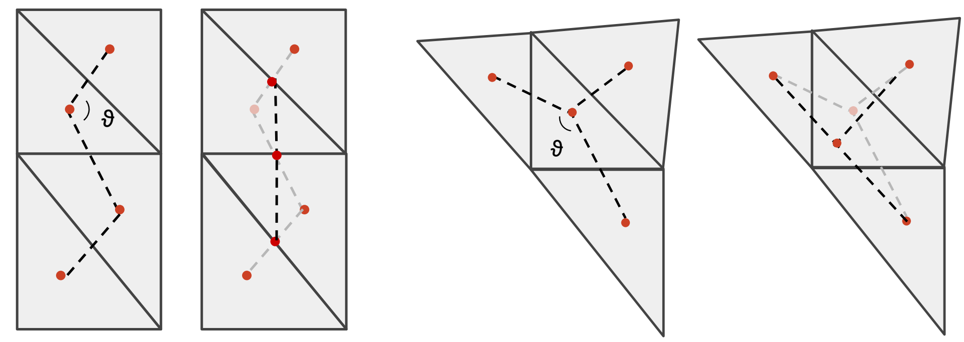
Algorithm 5.1: Dual graph of the triangulation

```
Result: lines LIST of line segments with start/end coordinates
triangle_dict = {triangle_id : [neighbor1_id, neighbor2_id, neighbor3_id]};
centroid_dict = {triangle_id : [x1, y1]};
lines = [];
for triangle in triangle_dict do
    point_0 = (x, y) #get triangle_id coordinates from centroid_dict;
    point_1 = (x, y) #get neighbor1_id coordinates from centroid_dict;
    point_2 = (x, y) #get neighbor2_id coordinates from centroid_dict;
    point_3 = (x, y) #get neighbor3_id coordinates from centroid_dict;
    line = linestring [(point_0, point_1)];
    lines.append(line);
    line = linestring [(point_0, point_2)];
    lines.append(line);
    line = linestring [(point_0, point_3)];
    lines.append(line);
end
```

The form of the dual graph of the triangulation is affected by the non-uniformity of the triangulation. The triangles with highly acute angles, combined with non-homogenous triangle shapes and densities, result in a dual graph that is not organically spread throughout the hole area. Another issue of the dual graph is the zig-zagged form which affects the correctness of the estimated path length. To reduce the impact of the zig-zags a 'relaxation' or smoothing of the line segments is performed. Some extreme cases of very acute angles and dense zig-zags are observed when the triangles have reversed orientation.

The graph smoothing is based on parsing all the points of the graph one-by one and distinguishing two cases: the points of the graph that have two neighbors (case 1) and the ones with three neighbors (case 2) according to the triangulation adjacency information. For case 1, the angle ϑ that is formed by the point P and the two neighbors n1, n2 is calculated. Aiming to eliminate the acute angles, if $\vartheta < 90°$ then P is relocated to the midpoint of n1, n2. Having eliminated the acute angles, the whole graph is 'relaxed' by relocating each point P, n1, n2 to the middle of the vertical to their direction distance (Figure 5.6a). The implementation code of graph relaxation is presented in Algorithm 5.2.

For case 2, the points with three neighbors (Figure 5.6b), the three angles $\vartheta 1, \vartheta 2, \vartheta 3$ are measured. If there is $\vartheta < 90°$ then P is relocated to the middle of the points forming this angle, else, P is relocated to the midpoint of the points that form the biggest ϑ.



(a) case 1: Triangle with 2 neighborss **(b)** case 2: Triangle with 3 neighbors

Figure 5.6: Dual graph smoothing

The above process produces a smoother graph that is not representing a route with a straight line but with a more relaxed and smooth set of line segments (Figure 5.7).

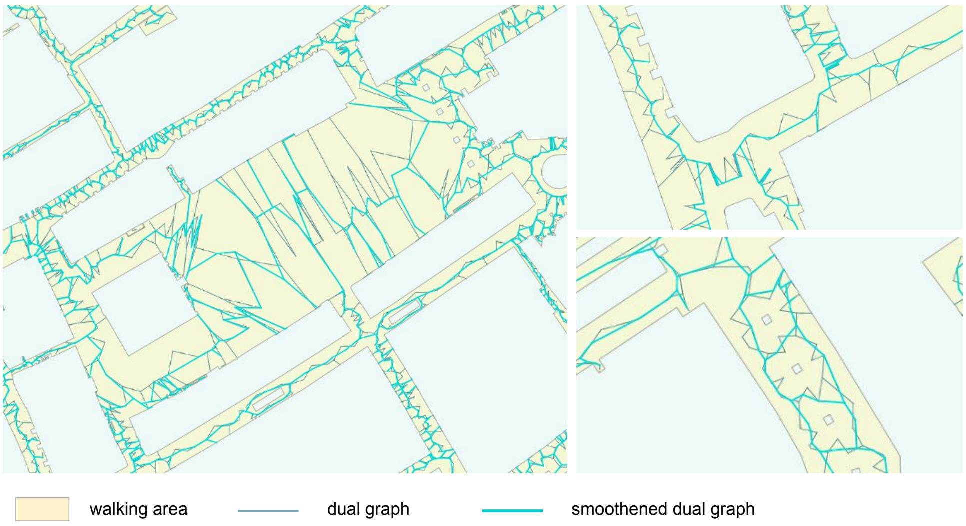


Figure 5.7: Dual graph before and after smoothing

Within the walking space, two major types areas are recognised; the long and narrow paths that correspond to the sidewalks and the wide open areas such as large open spaces and squares. The relaxed graph follows quite well the form of long and narrow areas but in the large open spaces the graph does not follow the area form. This denotes the need for a more dense representation of large open areas. Technically, triangulation is implemented based on points that lie on the polygon borders. Due to this fact, the large regions of a polygon are covered with bigger triangles. This also explains the non-uniform density and shape of the triangles. A more dense space tessellation requires additional points inside the polygon. Such a discretization can be achieved with the use of a mesh and is implemented in the following section.

Algorithm 5.2: Dual graph relaxation

```
Result: LIST of linestrings geometries
triangle_dict = {triangle_id : [neighbor1_id, neighbor2_id, neighbor3_id]};
centroid_dict = {triangle_id : [x1, y1]};
for triangle in triangle_dict do
    number_of_neighbors = len(dictionary values);
    if number_of_neighbors equal to 3 then
        p0 = (x, y) #get coordinates from centroid_dict;
        p1 = (x, y) #get coordinates from centroid_dict;
        p2 = (x, y) #get coordinates from centroid_dict;
        p3 = (x, y) #get coordinates from centroid_dict;
        ang1 = angle (p1, p2, p3);
        ang2 = angle (p1, p2, p4);
        ang3 = angle (p1, p3, p4);
        if min(ang1, ang2, ang3)<= 75 then
            if ang1 <= ang2 and ang1 <= ang3 then
                p0 = midpoint(p2, p3)
            if ang2 <= ang1 and ang2 <= ang3 then
                p0 = midpoint(p2, p4)
            if ang3 <= ang1 and ang3 <= ang2 then
                p0 = midpoint(p3, p4)
        else
            if ang1 >= ang2 and ang1 >= ang3 then
                p2 = midpoint(p1, p2);
                p3 = midpoint(p1, p3);
                p1 = midpoint(p2, p3);
            if ang2 >= ang1 and ang2 >= ang3 then
                p2 = midpoint(p1, p2);
                p4 = midpoint(p1, p4);
                p1 = midpoint(p2, p4);
            if ang3 >= ang1andang3 >= ang2 then
                p3 = midpoint(p1, p3);
                p4 = midpoint(p1, p4);
                p1 = midpoint(p3, p4);
        end
    if number_of_neighbors = 2 then
        p0 = (x, y) #get coordinates from centroid_dict;
        p1 = (x, y) #get coordinates from centroid_dict;
        p2 = (x, y) #get coordinates from centroid_dict;
        p2 = midpoint(p1, p2);
        p3 = midpoint(p1, p3);
        p1 = midpoint(p2, p3);
end
```

5.3 SPACE DISCRETIZATION WITH REGULAR MESH

5.3.1 Implementing the regular mesh

Building a mesh within a polygon can be implemented either from the ground up or with the use of existing tools and libraries. Developing an algorithm that constructs a mesh from the very beginning, is a process that demands a lot of programming skills and plenty of time to implement. On the other side, libraries that automatically construct a triangular mesh within a polygon are not aligned with geographic datasets. Besides that, the invalid polygon definition, makes the use of existing tools for mesh generation impossible to use. Therefore, the need for a method that overcomes the issues of the polygon size and definition arise.
Before constructing the grid, the geometry of the cells should be defined. Usually triangles or quadrilaterals are used. The algorithm used constructs a regular quadrilateral grid; regular grid ensures a uniform density over the whole area; the quadrilateral cells are offered for easier adjacency building and uniform shape. The construction of the mesh that overlays with the walking_area polygon is a process that is independent from existing tools for mesh generation. The algorithm that constructs the regular grid is based on the same logic of the algorithm of zone creation with the only difference being that of a smaller cell size. Additionally, it is a robust method that works for whatever shape and size of polygon.
Selecting the cell size of the grid was based on the desired resolution and on the polygon particularities (Figure 5.8). The resolution is defined by the extent of the expected detail in the representation of the polygon characteristics, and the desired density of the mesh. The polygon particularities refer to the concavities, the very narrow areas and the very small holes. The optimal cell size would be big enough to respect all this individualities of the polygon shape.

Figure 5.8: Particularities in the polygon geometry. Some long and narrow areas and holes.

In practice, this would result in a very small cell size, up to 1m, reducing the level of abstraction of the representation and increasing dramatically the computation time. Consequently, after testing various cell sizes, the selected cell size of the mesh is 5m which provides enough resolution to represent walking space at macroscale.
For the grid construction, one major requirement is to store the adjacencies of the cells. The algorithm for the grid creation and the adjacency storing is the same with the algorithm used to construct the origin and destination zones as described in Section 4.4.2. The difference is in the size of the cell which is much smaller.
The first step is to construct the grid of the $5 \times 5m$ over the bounding box of the *walking_area* polygon. The next step is to remove those grid cells that do not intersect with the walking area. To perform this spatial overlay function which is computationally demanding processes, especially when dealing with large spatial datasets, spatial indexing was used to reduce drastically the implementation time.
The form of the final regular mesh is presented in Figure 5.9. The selected cell size results in eliminating some details of the polygon such as small holes, long and narrow strips of walking space and makes areas that are non-connected to be connected Figure 5.10. In terms of walking in the polygon, this implies that even if in reality there is not a path connecting a place with another, in the mesh these areas seem to be connected. Despite this loss of accuracy, it is still efficient for representing walking trips at a resolution that corresponds to the macroscale.

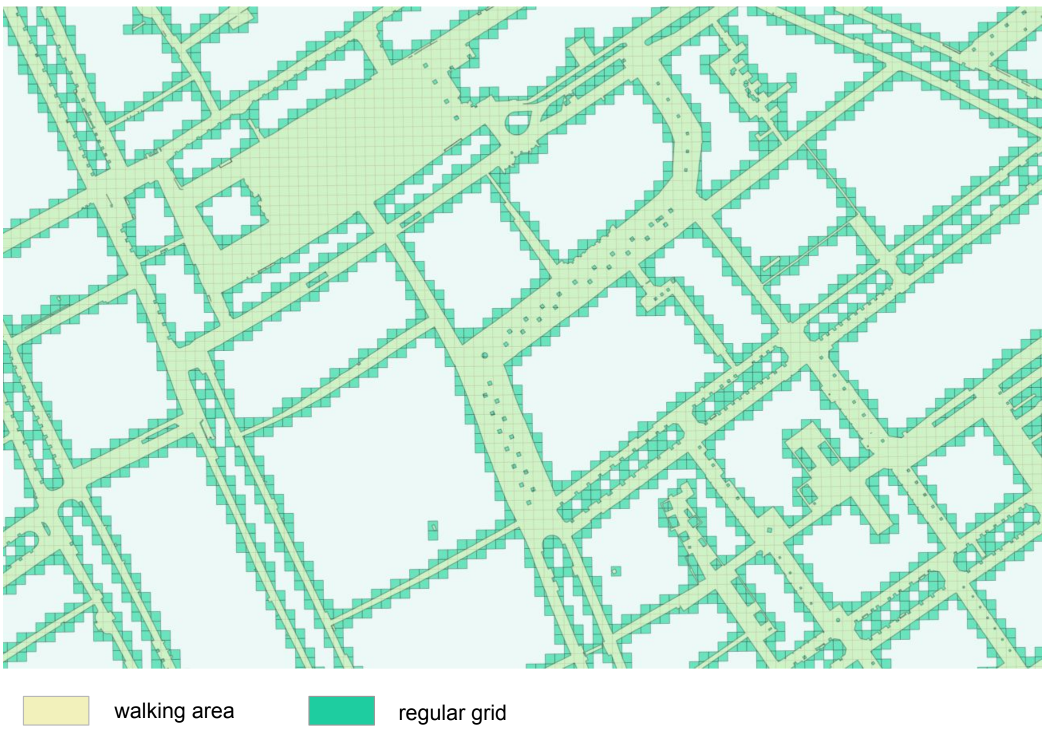


Figure 5.9: Space discretization with regular mesh of $5x5m$

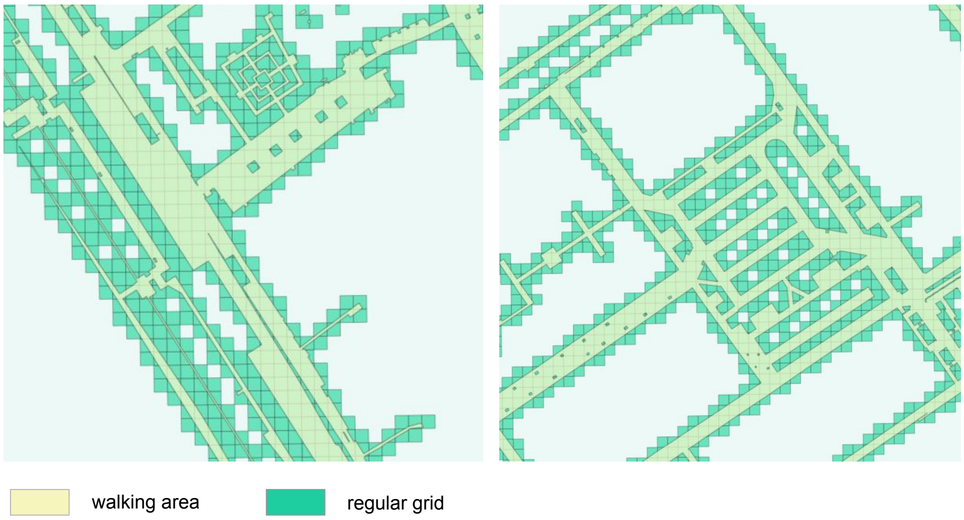


Figure 5.10: Example of mesh eliminated information of the polygon shape

5.3.2 Path planning on the regular mesh

Walking on the discretized space is going to be modeled with the dual graph, same as described with the triangulation method. The difference of constructing the dual graph of a regular quadrangular mesh compared to the triangular is that each quadrangle has at most 4 neighbors instead of three Algorithm 5.3. The resulted dual graph is a grid graph in which all the line segments of the graph have the same length (Figure 5.11).

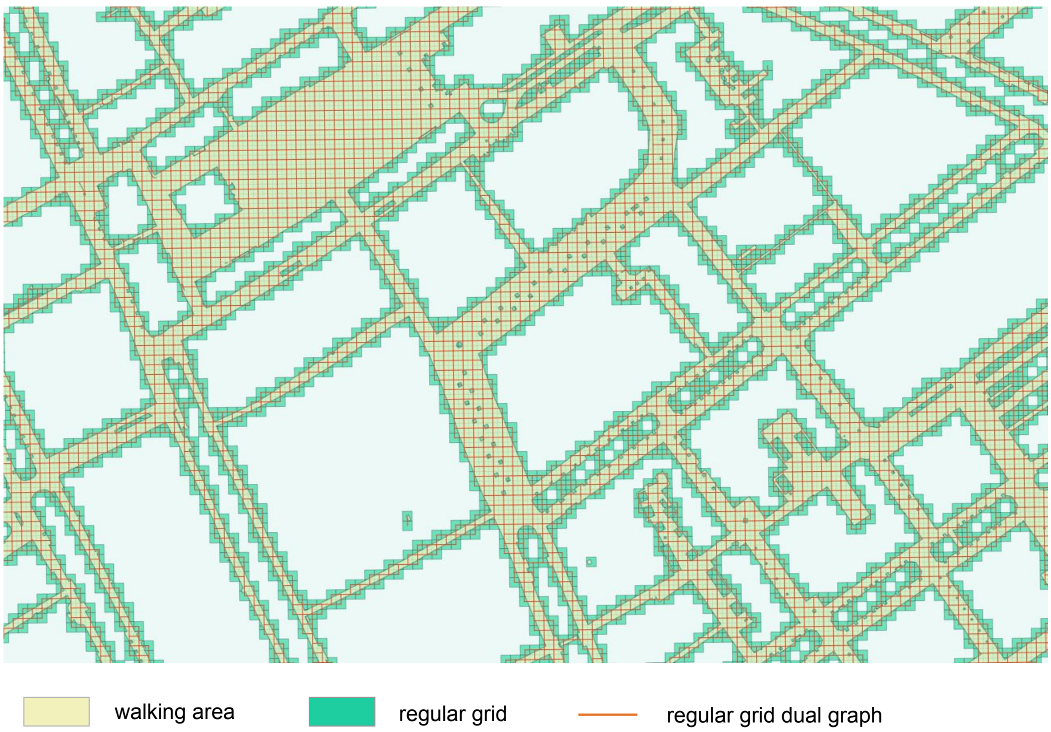


Figure 5.11: Dual graph of the regular quadrangular grid

Algorithm 5.3: Regular grid dual graph

```
Result: lines LIST of line segments with start/end coordinates
cell_dict = {cell_id : [neighbor1_id, neighbor2_id, neighbor3_id,
  neighbor4_id]} #dictionary of cells;
centroid_dict = {cell_id : [x1, y1]} #dictionary of cell centroids ;
lines = [];
for cell in cell_dict do
    p0 = (x, y) #get coordinates from centroid_dict;
    p1 = (x, y) #get coordinates from centroid_dict;
    p2 = (x, y) #get coordinates from centroid_dict;
    p3 = (x, y) #get coordinates from centroid_dict;
    p4 = (x, y) #get coordinates from centroid_dict;
    line = linestring [(point_0, point_1)];
    lines.append(line);
    line = linestring [(point_0, point_2)];
    lines.append(line);
    line = linestring [(point_0, point_3)];
    lines.append(line);
    line = linestring [(point_0, point_4)];
    lines.append(line);
end
```

The graph follows the form and structure of the regular mesh. Resulting from the fact that the regular grid does not preserve the polygon borders, some areas that are unconnected in reality, have a connection in the graph. Another drawback of using this method for implementing a mesh, is that the quadrilateral form of the grid results in multiple changes in direction which affect the final distances. These are factors that depend on the cell size and the desired spatial resolution.

5.4 NETWORK ASSIGNMENT

5.4.1 Network preparation

The network analysis was performed on the smoothened dual graph of the triangulation and on the regular grid graph of the mesh. Both graphs are undirected that means that moving on each graph segment is performed in both directions. Additionally, both graphs are simple since there are no more than one edge for each pair of neighboring vertices and no loops (Gibbons, 1985).

Before applying the shortest path algorithm, the graph was imported in pgAdmin with the spatial extension pgRouting where some preprocessing was required on the graph table (see Figure B.3 for steps in detail). The first step required is to built the routing topology using *pgr_createTopology* based on input geometry information. This procedure ensures that for any given edge of the graph, the ends of that edge are connected to a unique node which is also node of other edges of the network. This operation generates two tables: The first table is the vertices table that contains the id and the geometry information of the graph nodes. The second table is the same edge table updated with information about the source and target nodes of the edge referencing the vertex id from the first table.

The next step is to check the topology created for possible errors in the graph with *pgr_analyzeGraph*. Such error could be that line segments cross each other without having a node on the intersection point. Despite the fact that the graph was created in a way that such errors do not exist, this step was applied to ensure the validity of the graph.

Another action needed before obtaining the final graph, is to remove those parts of the graph that are disconnected from the main graph and consist of few number of links. These are created due to the existing islands on the walking area polygon and need to be removed so that trips will not start or end from them. To implement this, the *Python* library *NetworkX* was used (see Section B.3). The graph components with less that 15 connected links are removed (Figure 5.12). This number of connected components is selected to ensure that all the unconnected links of graph are eliminated. In the process of removing the disconnected components, 487 nodes and 1.016 links were deleted from the triangulation dual graph, while 91 nodes and 84 links were removed from the regular grid dual graph.

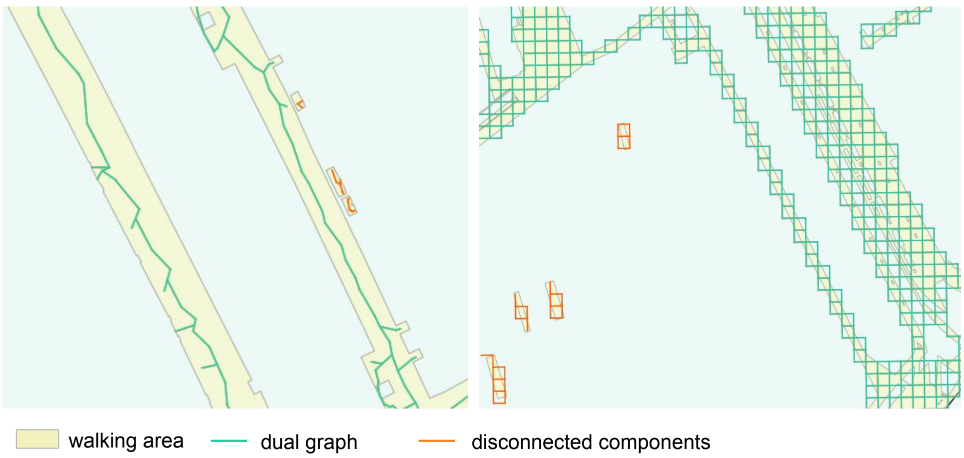


Figure 5.12: Disconnected components of graphs.

The final topologically clean graphs are both simple, undirected graphs. The dual graph of the triangulation covers the polygon with 52.605 nodes and 52.092 links while the graph of the grid contains 22.260 nodes and 36.099 links. The links and nodes of the triangulation dual and the regular grid dual are visualized in Figure 5.13, Figure 5.14 respectively.

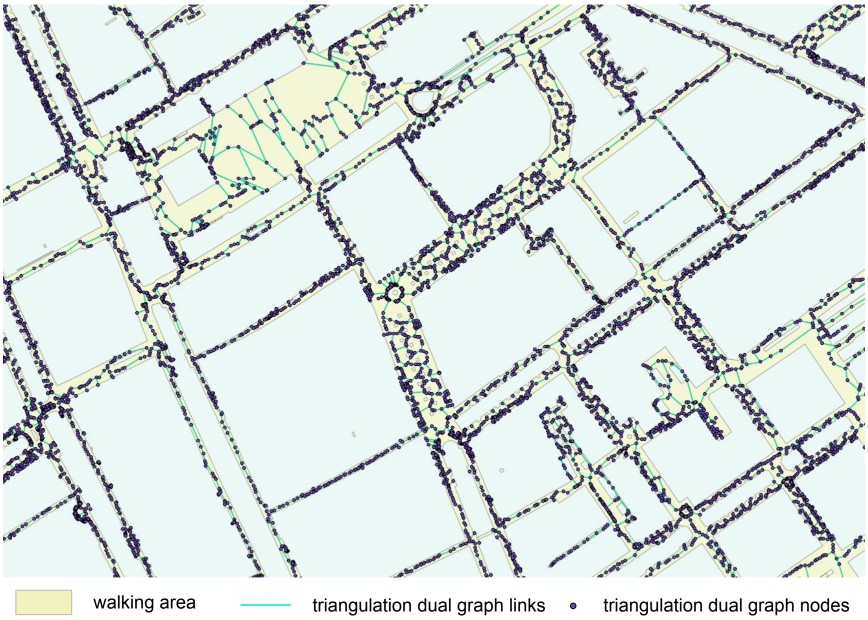


Figure 5.13: Links and nodes of the final triangulation dual graph. In total: 53.092 links/ 52.605 nodes

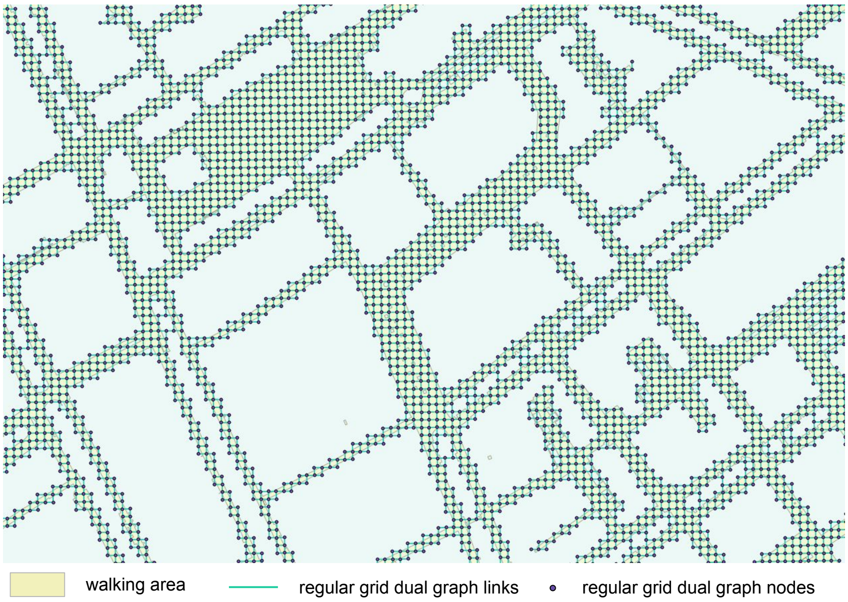


Figure 5.14: Links and nodes of the final regular grid dual graph. In total: 36.099 links/ 22.260 nodes

Before obtaining the shortest paths, the cost on each link of the graph is represented by the distance. The link distance is the more simple cost to consider. Instead the cost could express the utility of each link based on built environment attributes like the surrounding land uses, the shops, the bus stations, etc. Nevertheless, this analysis is out of the scope of this research. The length of each link is computed using the *postGIS* function *ST_LengthSpheroid* that calculates the length of a geometry considering the spheroid of the defined coordinate system. In this case the equatorial

radius and the flattening of EPSG 28992, the projected coordinate system for the Netherlands, are considered for the length calculation.
Having built the topology with *pgRouting*, the shortest paths for each origin and destination pair are implemented with the *pgRouting* wrapper for *Python psycopgr*.

5.4.2 Injection of trips to the network

The assignment of the trips to the network is a process in which, the information for the trips provided through the origin and destination matrix need to be extracted from zone level to individual locations of the network (graph nodes) for each origin and destination pair. The procedure followed for the vehicular network assignment usually involves locating the zone centroid and some zone connectors to link the trip ends (origin point and destination point) with the existing network. The same method will be used for the allocation of walking trips to the network. More specifically, the trips have as origin point the centroid of each origin zone and as destination the centroid of each destination zone. The centroids are connected to the network through the connectors, with the use of DBSCAN algorithm which is used to recognize polygon regions form the graph points. The steps of implementing the trip injection are presented in Figure 5.15

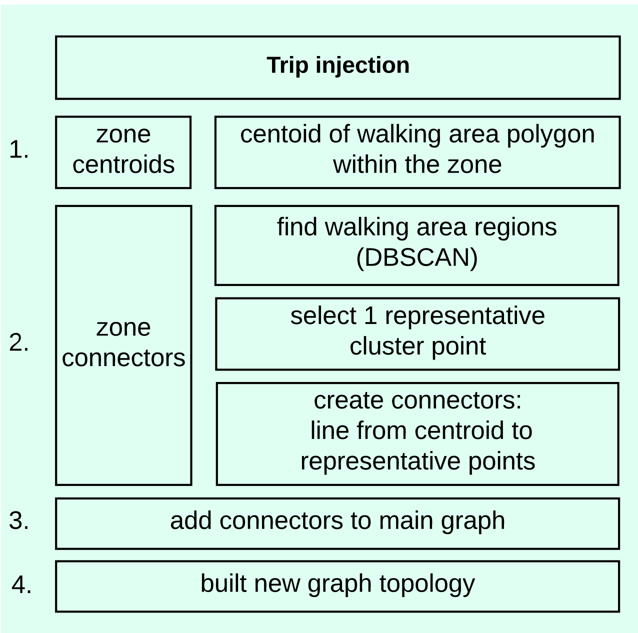


Figure 5.15: Trip injection steps

Injecting walking trips to the pedestrian network differs from the vehicular approach. First, the pedestrian analysis zones with size $80x80m$, are much smaller than the transport analysis zones therefore, they cover a relatively small part of the network. Additionally, the graph representing the walking network is dense implying that more than one graph links correspond to a section of sidewalk or a public square. Consequently, the selection of the starting and ending node of a trip within a zone does not affect the trip distribution dramatically since the graph nodes are few meters apart.

As injection point for each zone the centroid of the walking area is used. This is not located at the center of the regular grid but is the centroid of the intersection of each cell with the *walking_area* polygon. For some zones the centroid is not located inside the *walking_area* polygon (Figure 5.16). Next, some connector points must be specified, through which the trips will be channeled to the network. The connections of the zone centroid to the connection points will be considered as part of the graph with 0 weight assigned. There is not an exact method on how to select these connectors nor is it defined how many should they be. Thus, a method based on recognizing regions of the walking area is suggested.

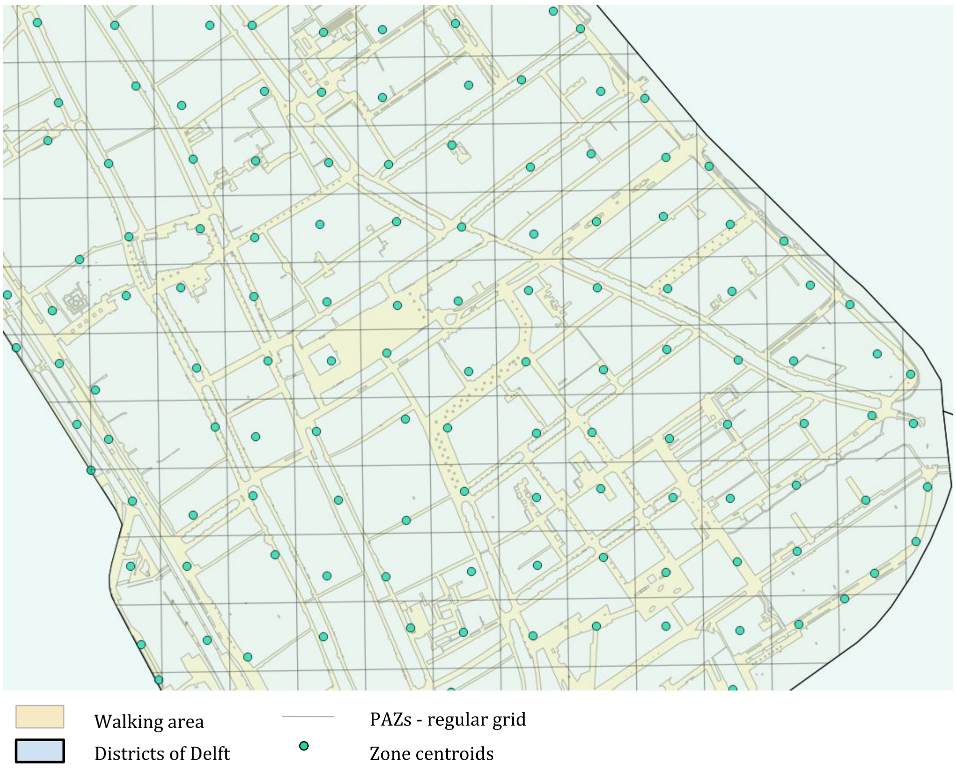


Figure 5.16: Zone centroids

More specifically, focusing only on the spatial form of the walking area, the proposed method for selecting the graph nodes that connect with the zone centroid is based on identifying the different possible regions of walking area. Different regions are defined as parts of the walking area polygon that are not not continuous, not continuously having the same direction (parts of a turn on the polygon), not connected (i.e. parallel, pavements), or sections that have different width. Aiming to identify such different regions within a zone, the clustering algorithm Density Based Spatial Clustering of Applications with Noise (DBSCAN) is used.

DBSCAN is a clustering algorithm developed by Ester et al. (1996) to identify clusters of arbitrary shape. What this algorithm does is, having as input a set of points in space, to group the points that are closer together by considering a neighborhood for each point. Also points that are separate from the main dataset are classified as outliers. More specifically, a minimum density is estimated, getting as input a threshold for the number of neighbors *min_samples* within radius ε which is defined with arbitrary distance measure (Schubert et al., 2017). Points that are more than the *min_samples* within the radius ε become the core points. The points with the radius ε from a core point belong to the same cluster with the core point (Schubert et al., 2017).

Using as input the graph nodes, DBSCAN distinguishes groups of points that form a group or cluster based on their distances. In practice, for each zone, the graph points that are contained, pass through the DBSCAN algorithm, which returns the output clusters of points. From each cluster one random point is selected. The set of random selected cluster points for each zone, are the final points connected with the zone centroid forming the zone connectors. Through these all the trips are assigned to the network based on the OD matrix. DBSCAN is implemented in *Python* through the library *scikit-learn*. The preparation of input and output data structures are described in Section B.4.1

Graph point clustering is implemented both on the triangulation graph and on the regular grid graph. The result of region recognition in each zone by graph point clustering is highly affected by the different form of the graph in each case. The points from the triangulation dual graph are clustered into more and smaller groups while the points from the regular grid are less segregated and thus, grouped into fewer clusters. This is an effect of the distance between points. In the first case of the triangulation points, the distance among points is variant while the distance between the grid points is fixed to 5 meters. Some examples of the result of point clustering is presented in Figure 5.17 - Figure 5.23.

Figure 5.17: Different polygon regions within a zone

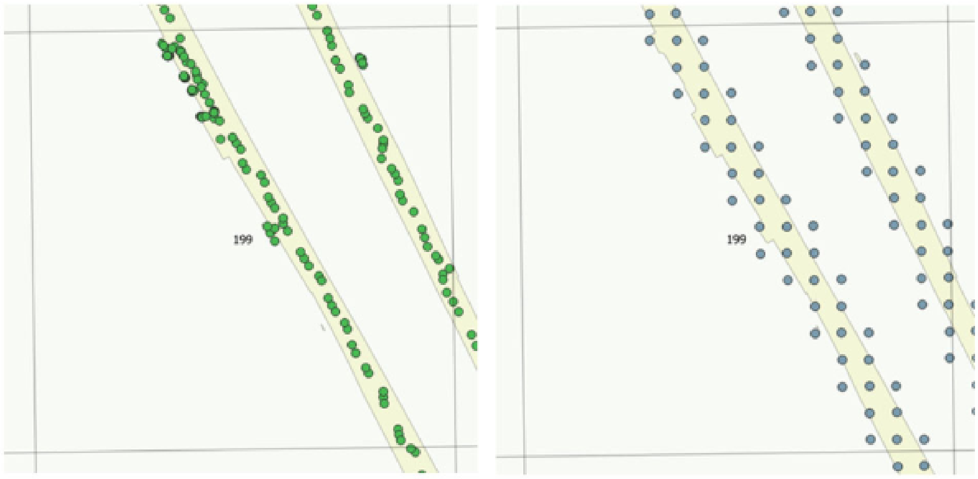


Figure 5.18: Zone: 199 Left: points of triangulation dual graph. Right: points of regular grid dual graph.

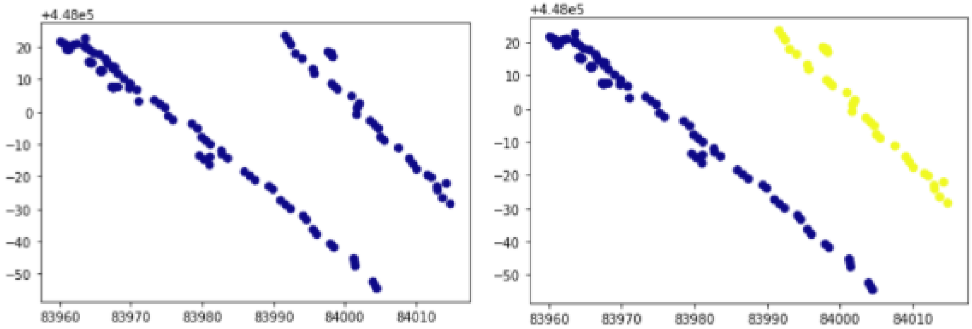


Figure 5.19: Left: points of triangulation method unclustered. Right: points clustered into two different clusters. Zone:199, Parameters:eps=5, min_samples=3, metric='euclidean', algorithm='brute').

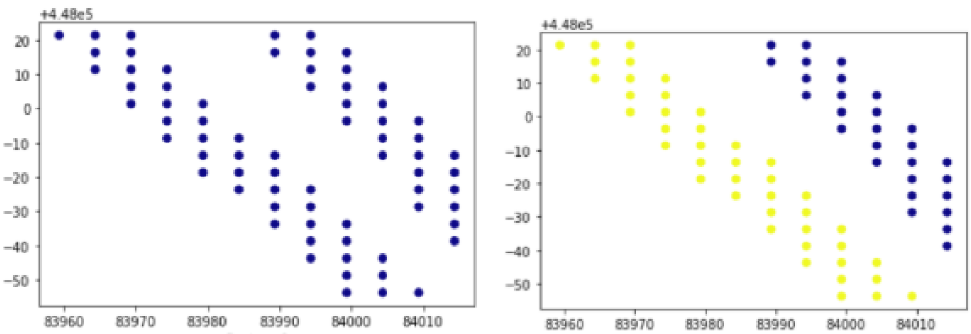


Figure 5.20: Left: points of regular grid unclustered. Right: points clustered into two different clusters. Zone:199, Parameters:eps=5, min_samples=3, metric='euclidean', algorithm='brute').

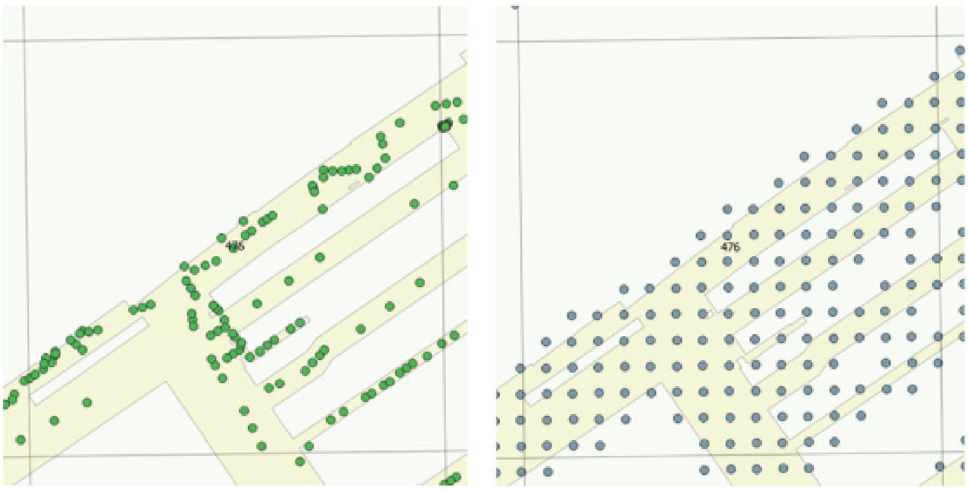


Figure 5.21: Zone:476 Left: points of triangulation dual graph. Right: points of regular grid dual graph.

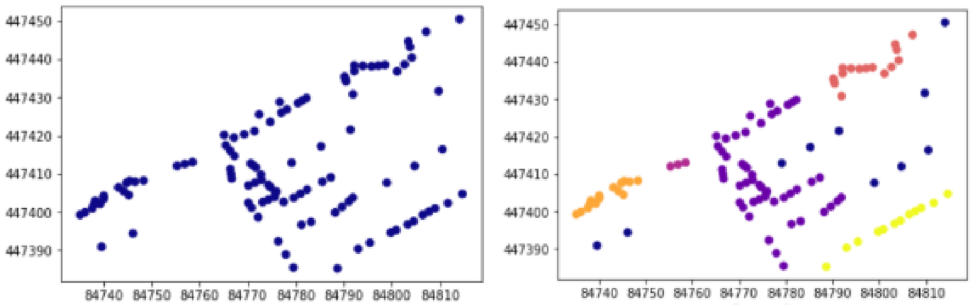


Figure 5.22: Left: points of triangulation method unclustered. Right: points clustered into two different clusters. Zone:476, Parameters:eps=5, min_samples=3, metric='euclidean', algorithm='brute').

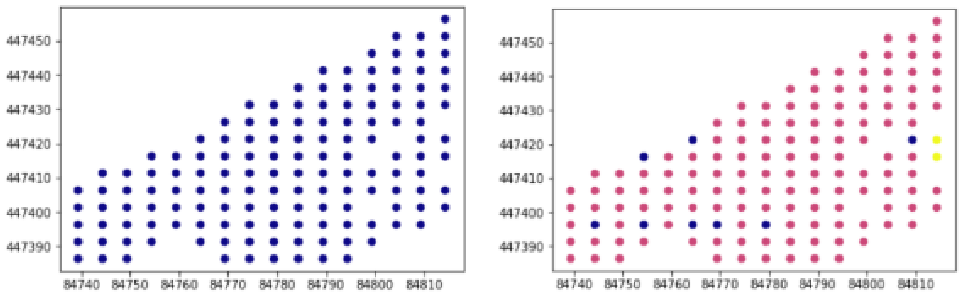


Figure 5.23: Left: points of regular grid unclustered. Right: points clustered. Zone:476, Parameters:eps=5, min_samples=3, metric='euclidean', algorithm='brute').

From the produced clusters in each zone, one representative point is randomly selected and connected with the zone centroid in order to form the zone connectors. The number of resulted clusters and therefore the number of connectors is 1084 for the triangulation dual graph, and 761 for the regular grid dual graph (Figure 5.24, Figure 5.25).

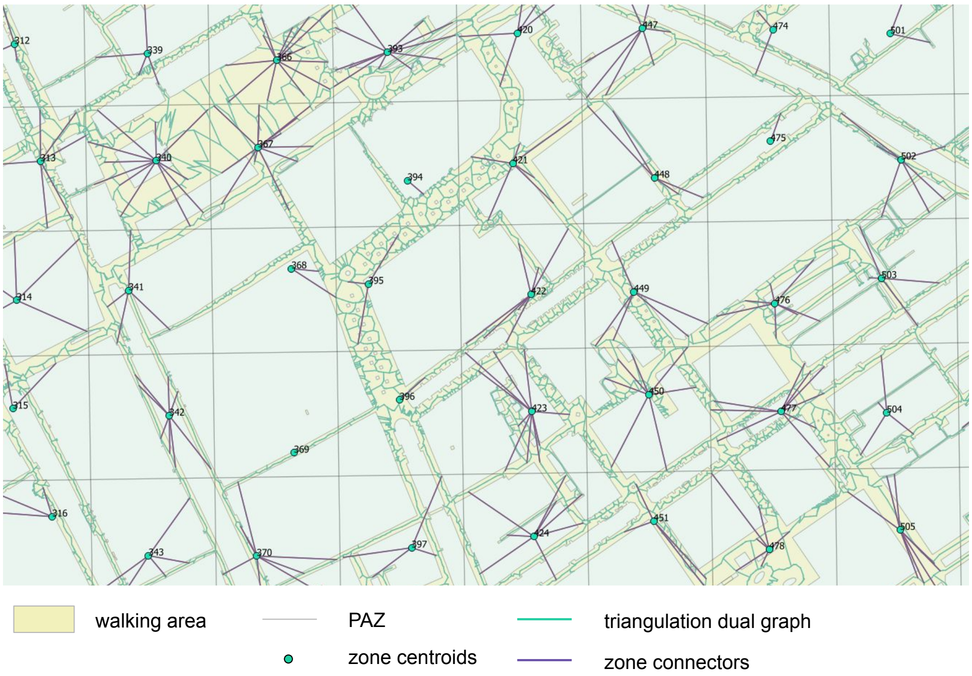


Figure 5.24: Zone connectors for triangulation dual graph created using DBSCAN to select graph points

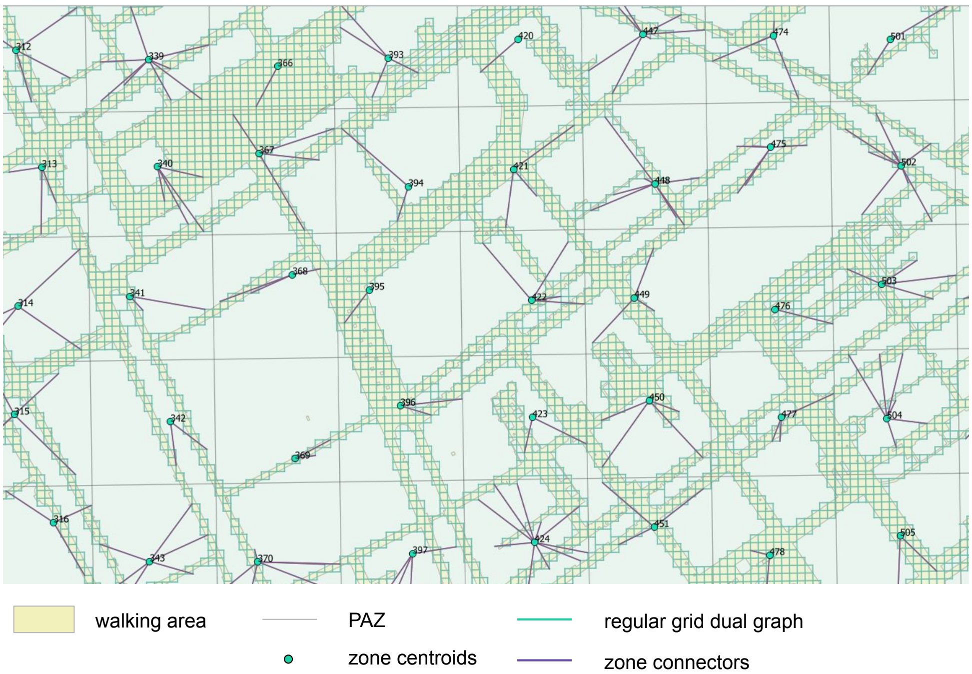


Figure 5.25: Zone connectors for regular grid dual graph created using DBSCAN to select graph points

5.4.3 Route choice

At this point the trips are projected from zone level to the nodes of the graph. The next step is to define how the route choice is performed for each trip. The route selection can be implemented either with a simple all-or-nothing trip assignment or with the use of stochastic model. The first approach considers that there is no network congestion and that there are no other parameters affecting the route choice, thus, all the trips with the same origin and destination are performed on the same route. The stochastic route choice assumes that some trips deviate from the shortest path due to individual user behavior, network characteristics and congestion on the network.

Using a stochastic method provides a more realistic result for the network assignment. Nevertheless, implementing a stochastic route choice model requires to consider behavioral information and built environment data, which is out of the thesis scope. Additionally, the trip data used are random trip numbers. Applying a stochastic network assignment without having trip data based on travel surveys and real population data, does not provide any valuable information regarding how the walking infrastructure is used. Consequently, since this thesis does not intent to provide any interpretation for the demand of walking in the study area, a simple all-or-nothing approach is implemented. The route choice is performed in order to visualize the result of the trips on the two different networks and to evaluate the effect of the connectors on the distribution of the trips.

The assignment of walking trips to the network is performed in *pgRouting* through the Python wrapper *psycopgr* on the network with the added connectors. The shortest path algorithm selected is the A* algorithm which is a more efficient version of Dijkstra algorithm for one-to-one path selection.

Before applying the shortest path algorithm on the origin and destination pairs, the following preliminary steps need to be performed. First, the connectors of the two graphs are appended to the existing dual graph network. Second, the network topology needs to be created again and assign zero cost to the connectors. Third, the centroid id needs to be passed to the equivalent graph node. After applying A*, the connectors are removed from the graph. This step is needed because in some zones, the centroid is located outside of the walking area and the formed connectors intersect with the walking area polygon and with the existing network. The preliminary steps required before applying the A* shortest path algorithm on the graph are described in detail in Section B.5.

The final assignment on the two networks after performing A* is presented in Figure 5.26 and Figure 5.27. The origin and destination for the two graphs is the same, since all the trips start from the zone centroid and are channeled through the network via the zone connectors. The zone connectors are different for each graph and were created after applying clustering on the graph points. The number of trips for each OD pair is generated with a random integer generator. On the triangulation network 1.077.703 random trips correspond to 67.979 OD pairs while on the regular grid network 1.0777.593 trips were assigned to 68.042 OD pairs.

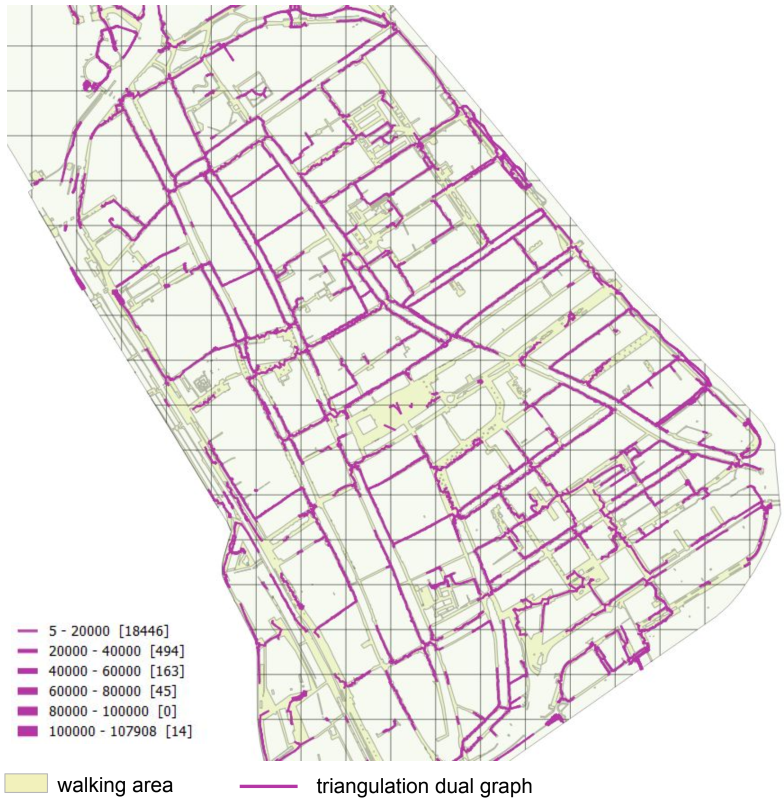


Figure 5.26: Assignment of trips on the network; triangulation dual graph

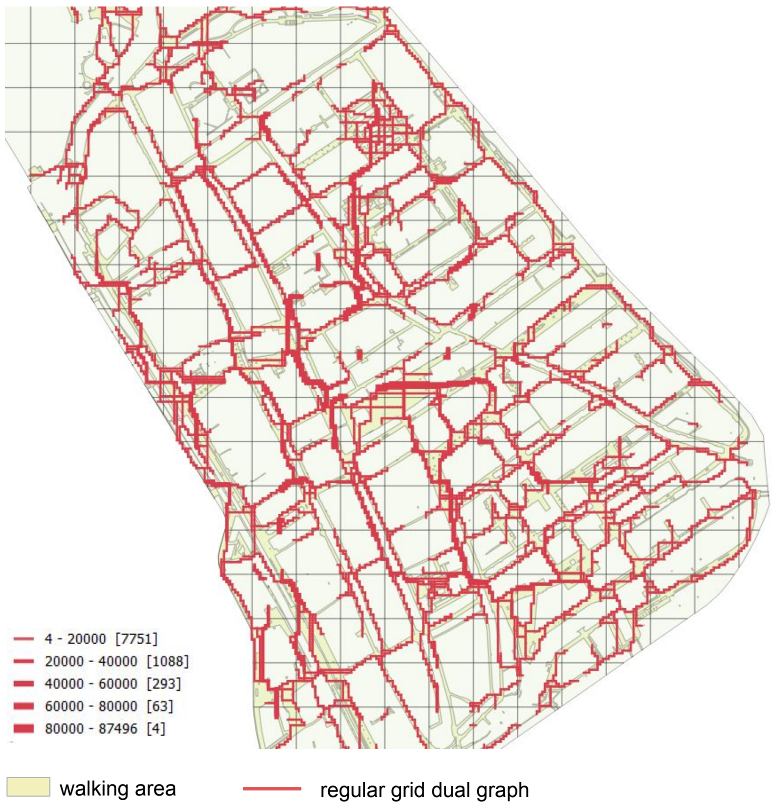


Figure 5.27: Assignment of trips on the network; regular grid dual graph

6 RESULTS AND ANALYSIS

6.1 MODELING WALKING SPACE METHOD

6.1.1 Using discretization to model continuous space

Walking space has been modeled for various applications and in different types of models. The existing techniques for modeling continuous walking space are devoted to microscale models where the walkable space is not very big nor does it have a complex geometry.
The nature of the transport planning or urban planning applications requires modeling walking space at large scale and the walking area boundaries need to be defined. For that purpose, the use of geospatial data is needed. The discretization of continuous surface is inevitable when dealing with such the large geospatial datasets that enclose big geographic areas such as districts, cities, regions, etc.
Selecting the proper space discretization method depends on how the walking area is defined. For instance, when there are mostly linear parts, and no wide open spaces, space representation does not need to be necessarily dense and the polygon skeleton provided by the dual graph of a polygon triangulation can be adequate. On the contrary, when there exist large open spaces, like public squares, a more coarse space representation is required such that offered by a mesh or a grid.

6.1.2 Evaluation of triangulation as discretization method

To implement the polygon triangulation, Python library *Tri* was used. This library, developed by Martijn Meijers, is the only existing tool that is designed to take as input geospatial datasets. This tool is developed to implement Delaunay Triangulation (DT) of a given set of points, and also takes as input polygon edges as constraints to construct Constrained Delaunay Triangulation (CDT). *Tri* performs well for a polygon of the application area size. Nevertheless, in an effort to triangulate the walking area polygon within the whole municipality, the process was not successful since some areas of the polygon were not triangulated.
Regarding the resulting triangulation of the polygon, a CDT was constructed based on the edges and vertices that lie only on the polygons borders. Consequently, the produced triangular network of the walking area polygon results to a non-uniform

space tessellation in terms of shape and density. Regarding the shape, it is described as quite anisotropic as there exist triangles with very acute angles. An ideally isotropic triangulation would contain triangles close to equilateral (Botsch et al., 2010). Regarding the density, the triangulation is non-uniform as the elements appear to be more dense on the curved borders of the polygon and less dense along the long straight polygon edges. This lack of uniformity is an expected outcome considering the polygon complexity and the large number of holes, as described in Section 5.2.1. Additionally, considering that it is a method that is based on points that only lie on the triangulated polygon border such non-uniformity is rational.
The big variation in triangles' shape and size also affects the dual graph that is produced. More specifically, the graph is highly zig-zagged, with uneven line segments and in-homogeneous density. The zig-zagged form is improved with the graph 'relaxation' which is implemented in Section 5.2.2.
Yet the problem of inhomogeneous density is not resolved. Overall in the walking area, two cases of walking space are recognised; the long and narrow polygon regions that correspond to the sidewalks and the wide open areas such as large open spaces and public squares. The graph follows quite well the form of long and narrow areas. This means that considering the long and narrow parts of the walking area, the relaxed graph follows the linear form of the long and narrow polygon. On the other side, the dual graph of the polygon triangulation resembles the polygon skeleton. When using the graph of the triangulation to represent wide open areas, walking space is misrepresented as the network is coarse and many parts of the walking area are not represented by the graph. This means that modeling walking in such cases, requires a more refined representation. This is a limitation of using triangulation method for this purpose. A more dense space tessellation would require points inside the polygon. This can be implemented with a regular or an irregular mesh.

6.1.3 Evaluation of regular grid as space discretization method

The need for a more dense space representation, denotes the urge for discretization technique that uses not only points of the polygon border, but also points of the polygon interior. This is achieved with a mesh construction. Implementing a mesh in a polygon, is a demanding procedure in terms of programming. The existing tools that offer the possibility to construct a mesh in a polygon, are not compatible with geospatial datasets. Consequently, the mesh construction could not be applied for the *walkin_area* polygon.
In order to achieve a dense space tessellation, a regular grid was constructed instead of a mesh. The construction of regular grid for space discretization has the benefit of being independent of existing tools and libraries. Additionally, the algorithm it is based on, provides the flexibility of adjusting the spatial resolution of the grid. The

benefit of using this algorithm is that it can be applied for any type of geospatial datasets and is easily transferable through different programming languages.
Concerning the result of this discretization method, main characteristic is that the grid cells cover the whole area homogeneously, with same size elements and uniform density. Especially in the large open spaces of the polygon, where the movement takes place towards all directions it is important to have a dense graph.
On the other side, overlaying a regular quadrilateral mesh over the polygon results to a space representation in which the polygon geometry is described with a level of abstraction. The smallest the size of the mesh size, the more precise the representation of the walking area. Generally, constructing a mesh that does not 'respect' the polygon boundaries introduces some issues of how accurate the reality is represented. For instance, the polygon holes that are smaller than the cell resolution are eliminated. Additionally, disconnected areas appear to be connected and the graph provides a path where connections do not exist in reality. Finally, the long narrow patches of the polygon on which movement can be considered linear, is zig-zagged affecting the computation of shortest paths.

6.2 TRIP INJECTION

6.2.1 Evaluation of the use of DBSCAN for the trip injection

The trip injection is performed with the use of centroids and connectors for each PAZ. Since there is no rule for the creation of connectors, the point clustering algorithm DBSCAN, is introduced. More specifically, the points of the graph are used to recognize the form of the polygon regions and to divide them into clusters, each one representing a different region on the polygon. A random point from each cluster is selected as representative point of each cluster and is connected with the zone centroid to form the zone connectors. The use of point clustering algorithm to recognize polygon regions is an approach that has not been applied before for the trip injection. Thus, there are some benefits and some drawbacks. In the next paragraphs, the method will be evaluated regarding the ease to use, the suitability over other clustering algorithms and the performance on the two datasets.
DBSCAN can be applied with the machine learning library *scikit-learn* in *Python*. The process is automated and can be easily applied and reproduced.
The advantage of DBSCAN algorithm over other point clustering algorithms (K-means, spectral clustering, Ward) is that it does not require to pass in the parameters a predefined number of clusters. In this case of point clustering, the number of different polygon regions within each zone is unknown and varies for each zone, thus, the number of clusters is unknown. There exist also few other clustering algorithms in which the number of clusters is not predefined. Namely, the affinity propagation,

and the mean shift algorithm. Nevertheless, DBSCAN provides better clustering result. Additionally, it is the fastest clustering algorithm. Thus, this method is not affected by the network size. For a detailed comparison of the algorithms see Section B.4.2.

The output number of clusters depends on the parameters given to the algorithm. The exact parameters that are passed to the DBSCAN are the maximum distance ε between points, the minimum neighborhood size $min\_samples$, the distance metric and the type of algorithm (for algorithm options provided: auto, ball_tree, kd_tree, brute). The two first parameters are those that affect mostly the number of identified clusters. After testing different combinations of parameters, $\varepsilon = 5$ and $min\_samples = 3$, resulted to better clustering result for both graphs.

The resulted clusters are affected by the input set of points and how they are spread throughout the walking area. This becomes more clear, when seeing the points of the graph without the walking area polygon. If the graph points are able to point out the polygon shape, then the clustering process is more probable to result in recognizing polygon areas. Inversely, when the graph points are unevenly scattered over the polygon, then the actual form of the polygon is not effectively represented by the points, thus, the clusters may not represent different polygon areas. Taking a look to the points of the two graphs, (Figure 6.1, Figure 6.2) it becomes obvious that the triangulation graph points are non-uniformly spread over the polygon. On the other hand, the regular grid nodes are equidistant, have homogeneous density and 'describe' the polygon shape better.

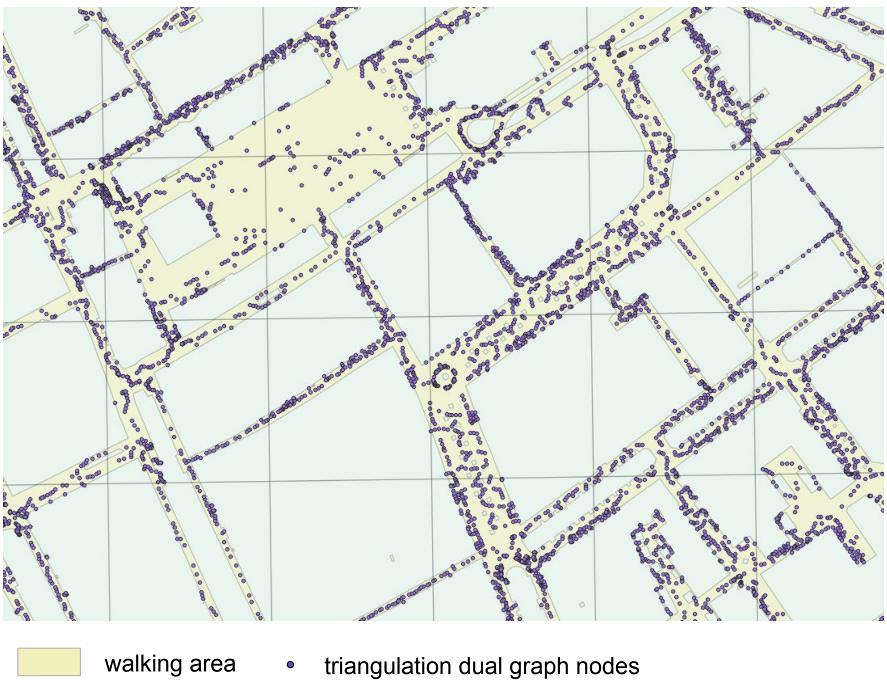


Figure 6.1: Triangulation dual graph nodes

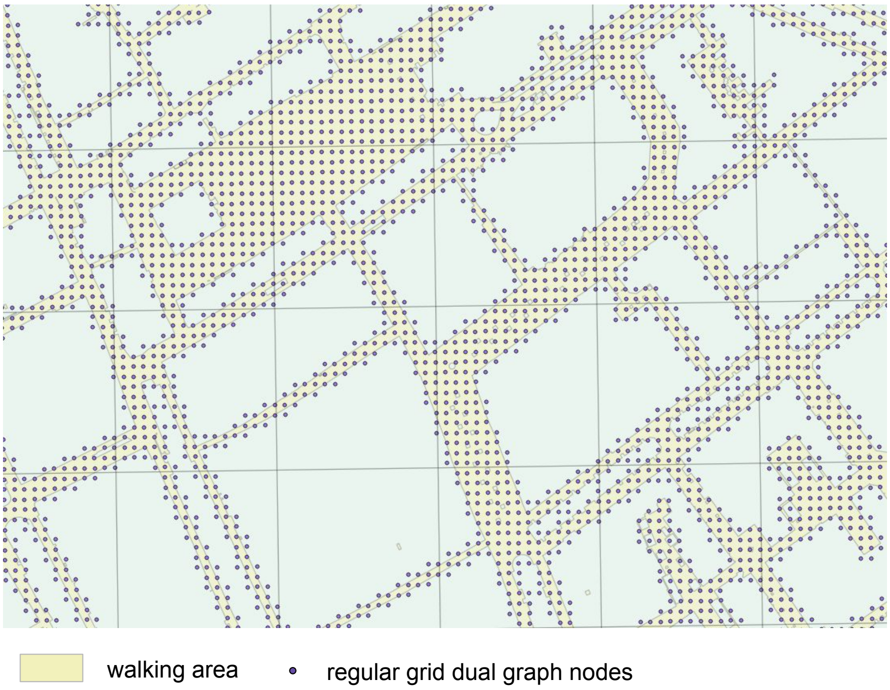


Figure 6.2: Regular grid dual graph nodes

The identification of point clusters is highly affected by the parameters set to the clustering algorithm like the maximum distance between two points or the minimum samples of a point neighborhood. For example, recognizing groups of points that are closely packed together depends on their distances. Points with variation of the in between distances, like the triangulation points, are more likely to be grouped into more clusters. This is because the triangulation of the polygon follows the borders and the geometry of the polygon and smaller areas are preserved. On the other hand, with the regular grid the polygon borders are not retained and unconnected areas are presented as connected while at the same time very small holes and discontinuities are ignored. Consequently, points that are more uniformly spread across the area like the regular grid points, are less possible to be grouped into more than one cluster. As an extent, the different regions of a polygon are more easily recognized and clustered when the graph points are not evenly spread on the polygon. At the same time, when the graph points are not homogeneously spread, it is possible that resulting clusters do not correspond to different regions on the polygon.

A drawback of implementing this algorithm is that the set of parameters' values is not the same for all the zones. Fox example, a combination of values set for ε and $min\_samples$ that result in good cluster identification for one zone, may not recognize the existing polygon regions in another zone.

The connectors are created by selecting one random point from each cluster. This is because the DBSCAN algorithm is designed in such way that there is not a representative point for each cluster. Selecting a random point has the drawback that for two neighboring clusters it is possible that the randomly selected points are close to each other. Nevertheless, since the walking area within each zone is relatively small and the graph nodes are closely located to each other, selecting a random cluster point does not change dramatically the final route choice.

6.2.2 Possibilities offered with point clustering

Using clustering algorithm to identify different cases of the walking area is a promising technique for the model. At this point, the proposing method only considers the form of the polygon area. Enhancement to this would be to introduce additional criteria for selecting which point within each cluster will be the representative point through which trips will be injected. Although it is out of the scope of this research, information concerning the land use, the functions as well as other built environment measures around the walking area could be added to the point selection criteria. For instance, a commercial street may attract more trips and a residential street may generate more trips. But the extent to which a specific number of commercial uses affects the attraction/generation of walking trips is unknown. Since these parameters are not quantified, they can not be introduced as criteria for the trip injection yet.

6.3 EVALUATION OF ROUTE CHOICE MODEL

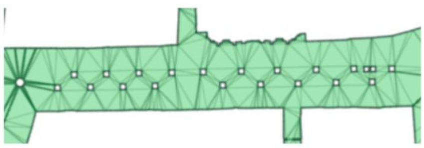
Applying an all-or-nothing deterministic route choice model is a simplistic method for trip assignment. Yet, as already mentioned, in Section 5.4.3, the purpose of the thesis is not to evaluate the use of walking infrastructure in the study area. Since no real travel data are used, no results can be extracted from the route assignment. The route choice step was implemented to provide a more clear visualization of the walking trips in the walking area, and to evaluate the two graphs that result from the selected space discretization methods.

6.4 COMPARISON OF THE TWO DISCRETIZATION METHODS THROUGH THE WHOLE PROCESS

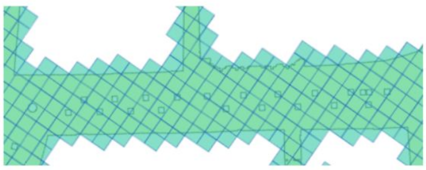
The discretization of continuous walking space is implemented with two techniques; a CDT is implemented to triangulate the polygon and a regular quadrangular grid is overlaid with the walking area. These two methods are different in implementation and space representation. The differences regarding the implementation, the polygon representation, the performance of each graph nodes in the DBSCAN point clustering algorithm and their suitability for the final the trip assignment will be thoroughly analyzed below.

Considering the automation of the process, the polygon triangulation can be performed in an automated way with the use of various existing Python libraries. The library *Tri* is suitable when geospatial data are used as input. Nevertheless, this library did not perform well when applied on a very large polygon. Instead, the regular grid method, is based on a very simple algorithm to create a grid that overlays with the polygon and performs well for any size of polygon.

Regarding the resulted polygon representation, the triangulation consists of non-homogenous size and shape of elements but offers a precise representation of the complex area of the polygon (Figure 6.3). More specifically, it reproduces the exact polygon geometry, respecting the concativities and holes. On the other side, regular grid offers a consistent shape and size. The drawback of this method is the poor representation of complex polygons since it does not 'respect' holes and concativities and does not follow the borders. These drawbacks of using regular grid can be downsized with the possibility to easily reduce the cell size.



(a) Constrained Delaunay Triangualtion

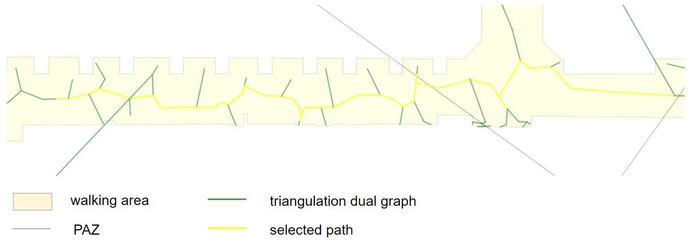


(b) Regular grid

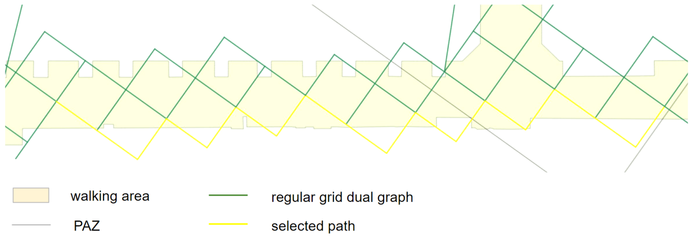
Figure 6.3: Comparison of polygon triangulation and regular grid

Another important thing to consider is the length of a path on the graph. Measuring the length of a linear road segment on both graphs the total length on the regular grid graph is measured equal to 83.30m, the same segment on the triangulation dual graph is computed equal to 75.25m while the actual straight line length is equal to 64.54m (Figure 6.4)[1]. This indicates that the actual length of the shortest paths is falsely bigger in both cases with the bigger deviation from the actual length being on the graph of the regular grid. However, the shortest path is computed between the routes on the same graph which means that since all the paths appear to be longer than the reality, this does not affect the selection of a route over another.

1 the measured selected paths are parts of the produced shortest paths after running the A* algorithm.



(a) Route length on the triangulation dual graph

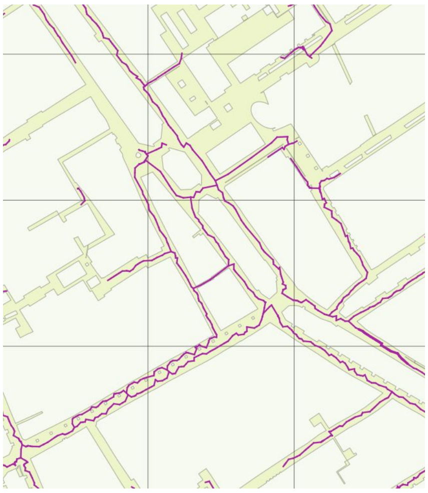


(b) Route length on the regular grid graph

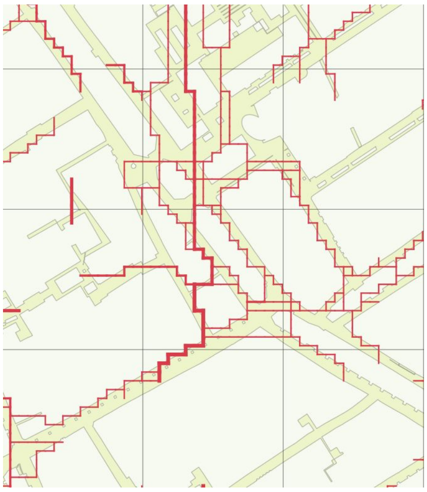
Figure 6.4: Comparison of route lengths

Seeing the size of the produced network on which the assignment of trips is performed, the triangulation dual graph, consists of significantly more links and nodes (53.092 links, 52.605 nodes) compared to the regular grid dual graph (36.099 links, 22.260 nodes). The implementation time of the A* algorithm that was used to perform the final route choice is affected by the number of edges and vertices (see Section B.4.2 for algorithm complexity). Considering this together with the more dense representation of the walking area polygon, the regular grid is more efficient. Considering the performance of DBSCAN when using the nodes of each graph for region recognition, the triangulation dual graph nodes are non-uniformly spread across the area, resulting in classes that do not correspond to a distinct polygon region. This also means that more connectors are produced. The regular grid dual graph nodes, are more suitable for representing the polygon shape since they are homogeneously distributed over the walking area polygon.

Furthermore, the size of the network has also an impact on the produced routes. More specifically, in the process of assigning trips to the two graphs, 1.077.703 trips where assigned on 19.162 links of the triangulation dual graph, while 1.077.593 trips where assigned on 1.191 links of the regular grid dual graph. This indicates that in the two cases, almost the same number of trips was allocated to a more than double size of links (the case of triangulation dual graph), affecting the load of the links and the final result of trip assignment (Figure 6.5). At this point no results can be extracted about which of the two methods offers a better network representation for the trip assignment. Comparing the resulted routes in both cases with actual pedestrian routes (gps data) could be a valuable validation technique for the two methods.



(a)

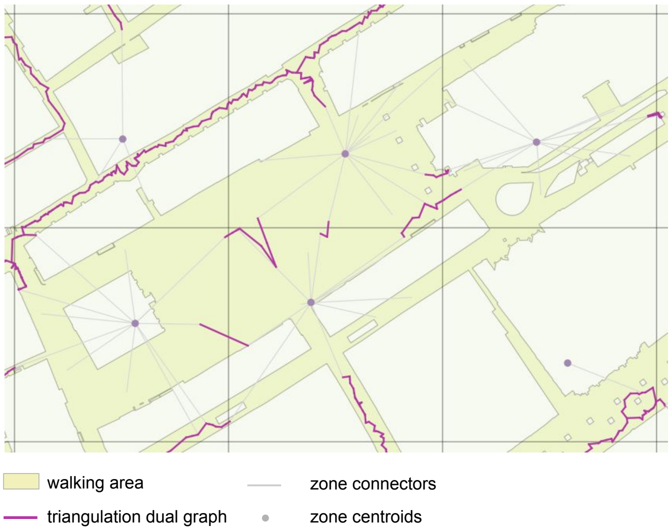


(b)

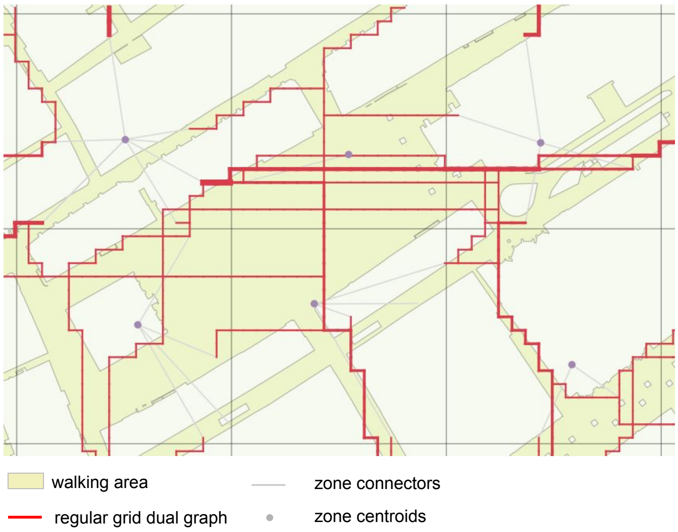
Figure 6.5: Effect of network size on the final trip assignment

Finally, comparing the two methods with regard to the allocation of the trips to the network, two things need to be evaluated; the effect of the connectors and the effect of the graph density to the final trip assignment. Considering the connectors, it is observed that in some zones the trips originate from nodes close to the zone boundary resulting to having zones with no trips in the center Figure 6.6a. This effect could be mitigated if instead of selecting a random point from each cluster, the closest to the centroid point was selected. The effect of having trips 'pushed' towards the zone border is affected by the network density. In particular, it is more dominant when using the triangulation dual graph which offers a coarse representation of wide, open areas. This effect is eliminated with the dense representation that the regular grid offers Figure 6.6b.

Overall, each method has some advantages and disadvantages. Nevertheless, it seems that the disadvantage of the regular grid to precisely represent the polygon form, can be regulated by reducing the cell size of the grid. Thus, it can be concluded that the uniform shape and size of the cells and the homogeneous density are more suitable for large space discretization. Additionally, the grid network provides a more dense walking area representation. The extra length on the routes added by dual graph of the regular grid, is a phenomenon present over the whole network and thus, it does not affects only specific routes. Furthermore, the dual graph nodes are better input for the polygon regions recognition through the DB-SCAN algorithm since they evenly represent the polygon geometry. Finally, the intrazonal trips are better distributed throughout the network within the zone with the dense network of the regular grid dual graph.



(a) Trip assignment on triangulation dual graph



(b) Trip assignment on triangulation dual graph

Figure 6.6: Effect of connectors and graph density on the trip assignment

7 CONCLUSIONS AND FUTURE WORK

7.1 RESEARCH QUESTIONS

The conclusions of the thesis will be presented in alignment with the research questions as posed in the introduction, starting from the sub questions.

1. *What are the differences of pedestrian modeling comparing to vehicular modeling?*
   The main difference between vehicular trips and walking trips is that modeling walking introduces a difference in scale. The four step model estimates the travel demand at regional and sub-regional level. The whole area is divided into large zones the Transport Analysis Zones (TAZs), which are the core spatial unit through the four steps. The zone size depends on the type and extent of the application and can start from building block size up to the size of a neighborhood or a city. The size of the zone is important determinant for the number of trips performed within the total area and when large zones are used the internal trips are eliminated.
   Large spatial analysis zones are inefficient for representation of walking because walking is not performed on large distances, and the use of large zone size results to the elimination of the intrazonal trips. Thus, the need for finer resolution of transport analysis zones arises to achieve a better representation of walking trips.
   Using a regular grid of $80x80$ meters for the Pedestrian Analysis Zones (PAZs) refines the spatial resolution of the model and less trips are eliminated. Additional benefit of using same size of zones is that they constitute a unit of measure since 80 meters correspond to 1 minute of walking. Additionally, having uniform zone size is advantageous when there is a need to project census data and travel surveys to smaller from large administrative units to smaller subdivisions.
   Another distinction between pedestrian modeling and vehicular modeling is the representation of network. Road network is classified according to its function and parameters like width, direction, turns and network capacity are considered. It is modeled and specified by a graph in which the physical junctions, crossings, turns of the road network are represented in the graph. Walking areas are classified according to their functions as pedestrian crossings, public

squares and pavements and in urban planning the term pedestrian network implies that there is connectivity among the walking infrastructure spaces. In transport modeling the term has not been defined before. Actually in this research the term walking network has been used by convention to align with the terminology of the Four Step Model (FSM).

Answering this sub question had the purpose of acknowledging the differences and challenges that modeling walking introduce. Despite that, after constructing the walking network from the dual graph of the two discretization methods, some more conclusions can be extracted regarding the differences of road network and walking network. More specifically, the form of walking network does not follow the physical form of the walking area. This means that a change in direction on the graph does not necessarily entail a turn on the physical space and vice versa. Additionally, junctions and crossings in reality are not present in the graph of walking.

2. *How is walking space modeled and how is pedestrian network defined?*

   Unlike cars, walking is an activity that is not spatially restricted. Vehicular trips take place on a clearly defined road or transit network on specific direction and with limitations on how to move or turn. Vehicular trips are modeled on a road network representation consisting of links and nodes that correspond to road segments and crossings respectively. On the other side, walking is not restricted and is performed freely towards all possible directions. Walking is not one-directional nor is the walking space. Representing walking linearly implies that the possibility of walking towards many directions is not considered. Therefore, the need for a continuous representation of space arises.

   In order to model continuous walking space at large scale, it is important to define what is the walking space and how it is delimited. In the model, walking space is defined with the use of spatial dataset that has distinct categories of land covers and the pedestrian areas are demarcated. Through this dataset, the boundaries of the continuous walking area are also set.

   The large scale of the application area, reveals that modeling of continuous space requires some discretization into smaller continuous spatial units. Continuous space can also be modeled as a continuous surface with the use of mathematical functions but the size and form of the walking area are prohibiting to do so. More specifically, the walking area polygon, is large in size and with particularities in geometry like concave areas and holes. This denotes the need to treat continuous walking surface as a set of smaller continuous subdivisions.

   Space discretization is performed in two ways. The first method is the polygon triangulation and the second is discretization with the use of a regular grid. Polygon triangulation is implemented with the use of *Python* library *Tri* which constructs a Constrained Delaunay Tiangulation (CDT) for the walking area polygon. The produced outcome is a set of triangles non-uniform in size and density. More specifically, the set of triangles is composed by small triangles in long and narrow areas and much larger on wide continuous areas, offering a coarse discretization result. This is resolved with the regular grid construction.

The regular grid is implemented with a custom algorithm which offers the possibility to change the cell size. The selected cell size of $5x5m$ offers sufficient generalization of the polygon geometry for this macroscale application. The homogeneous shape and size of cell and at the same time the uniform density of elements outweigh the limitations of the polygon triangulation. On the other hand, this method does not respect the geometry of the walking area polygon and in some cases the topological relations change (unconnected areas on the polygon are connected in the grid). Nevertheless, this granularity is eliminated at the macroscale.

Regarding the definition of walking network, in both methods walking is performed on the dual graph of the triangulation and the regular grid respectively. The dual graph of the triangulation follows the properties of the triangulation. That means that the non-homogeneous size and shape of the triangles affect the form of the dual graph which ends to be highly zig-zagged and provides coarse representation of large open areas. The zig-zagged effect is smoothened by applying a relaxation algorithm based on the angle between neighbor points. Yet, the triangulation dual graph is still ineffective for wide areas since it offers a 'skeleton' representation of the polygon and not a dense reproduction of the polygon shape.

The walking area network produced with the regular grid is a simplification of the polygon geometry and even though it follows the geometry and form of the polygon it does not 'respect' it precisely. As an outcome, some paths that do not correspond to the reality appear in the walking graph. Still these false connections are very small (at the size of the grid cell) and have a minor affect on the route length.

A benefit of using the regular grid dual graph lies in the density of the network. The regular grid dual graph covers uniformly the walking area and represents the wide spaces that the triangulation dual graph fails to do. Though, in the long and narrow polygon areas, the graph adds more turns and changes in direction.

How to perform assignment of walking trips to the network for better representation of pedestrian activity within the 4-steps travel demand model.

For the assignment of walking trips to the network, this thesis develops a procedure to model walking space and walking trips on the walking network. The main concepts introduced are the finer grained scale of spatial analysis zones needed, the walking network which is provided through the continuous walking area surface, and the role of the polygon form to recognize different regions on the walking area. The steps of the suggested workflow are presented in Figure 7.1.

Walking activity is not performed on long distances and requires smaller zones in order to consider the intrazonal trips. Thus, the size of the transport analysis zones is redefined and a regular grid of cell size $80x80m$ is used. This grid size proposed by Clifton et al. (2016), corresponds to 1 minute of walking and can be easily used as a unit of measure for trips.

Walking area and walking network have not been defined before in transport planning. In this thesis, walking space is defined as a continuous surface and provided with the use of spatial data. The data is a polygon dataset that contains information for the land covers on the terrain. Modeling walking on continuous walking space at large scale, requires the discretization of the walking area polygon. Performing discretization with the use of a regular quadrangular grid, a dense and uniform reproduction of continuous walking surface is provided. In extend, walking network which is defined by the dual graph of the grid, is a dense and uniform network.

Finally, the assignment of the walking trips to the network is based on the way that the walking network graph nodes represent the walking area polygon. The trips are injected from the OD pedestrian analysis zones to the graph points with the use of DBSCAN and the route choice is performed with a simple an all-or-nothing approach.

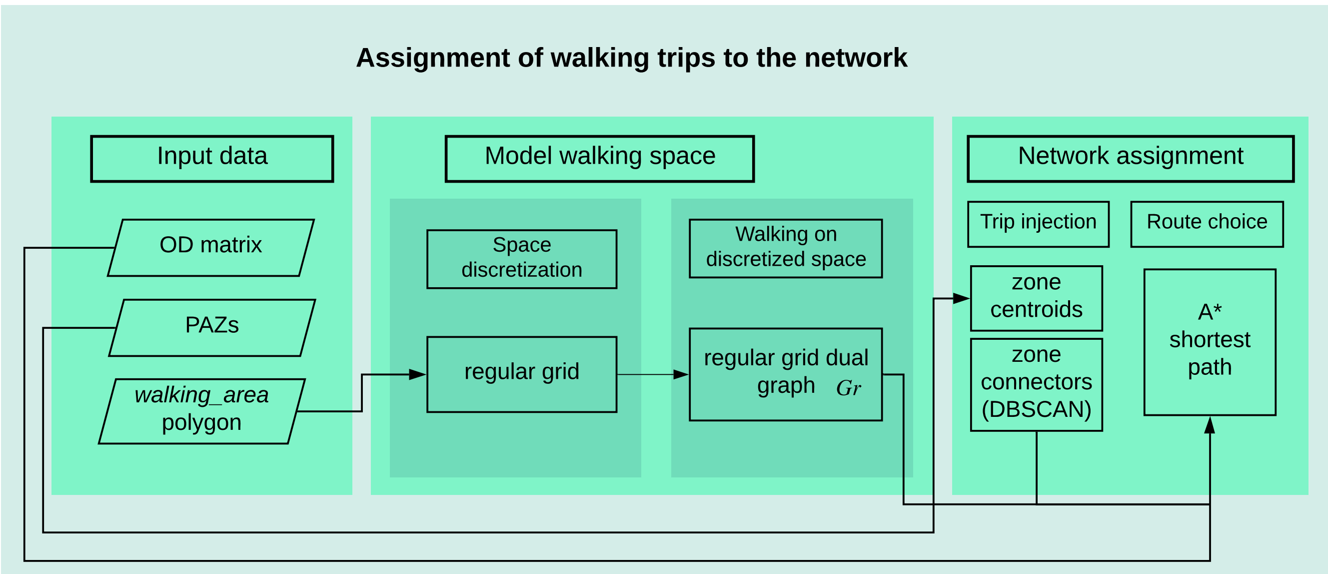


Figure 7.1: Suggested workflow for network assignment

7.2 REFLECTION AND DISCUSSION

Overall, this thesis attempted to approach a problem of transportation modeling that has not been a priority to resolve. This explains why there exists limited literature on modeling walking within the Four Step Model and in parallel why modeling walking space at large scale has not been performed. Despite this lack of interest, the FSM still remains the main tool to estimate the needs of a transportation system within an area. Thus, walking should not be ignored or misrepresented among the others mode of transport. Additionally, developing a tool to assist pedestrian modeling at large scale, can contribute to sustainable urban and transport planning process.

The proposed methodology, aims to enhance the role of walking within the FSM and is should be compatible with the existing steps. Indeed, the suggested procedure is aligned with the existing steps with the only difference being that after the third step of the model the vehicular trips and walking trips are treated with different procedures and at different scales Figure 7.2.

Regarding the developed methodology, using a smaller size for the PAZ, is more suitable for the scale of walking activity. Though, the availability of data at this granular level is something that needs to be considered. Usually, travel surveys and questionnaires used to get the estimates for the trips, are performed within administrative units and mostly at larger spatial entities. Therefore, real data at grid zone level, in most cases will not be available. This can be resolved by projecting the existing data to the smaller PAZs with various existing demographic and statistic techniques.

Regarding the space modeling method, the suggested technique for polygon discretization, is a regular grid. The advantage of this method is that it provides a dense polygon representation and uniform size and shape of the elements. Still the regular grid has the drawback of not respecting the geometry of the walking area polygon. This issue could be resolved with a triangular mesh instead of the regular grid. However, constructing a mesh is not as simple as constructing a regular grid. It is also possible that a remeshing would be needed in order to achieve uniform size and density of the triangular elements. These highlight the need for an automated process for mesh creation, which is currently not available for large geographic datasets. The increasing applications that are based on meshes can make such a tool available in the future.

One novelty of this thesis is that it introduces a point clustering method to create the connectors for the trip injection. Despite the fact that the DBSCAN algorithm was evaluated as the most suitable among the other point clustering algorithms, there are still some cases that it does not distinguish some polygon regions. Reducing the size of the grid cells could offer better clustering results.

The final route choice was performed with a deterministic all-or-nothing method. At a further step, where real travel data are used, and there is availability of built environment and behavioral data, a stochastic route choice model will be more realistic. Yet, in the context of this research a stochastic network assignment would offer no additional information for the use of walking infrastructure.

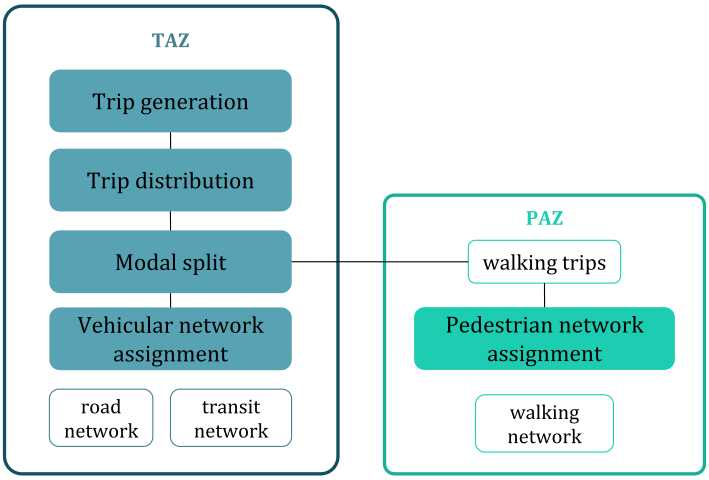


Figure 7.2: Proposed procedure in FSM

7.3 CONTRIBUTION

This thesis emphasizes the role of walking activity within the four step model. By addressing the assignment of walking trips to the network this work broadens the limited literature devoted to pedestrian activity within the FSM.

The main contribution of this work is that it suggests a method to model continuous walking space at large scale. Through this, also a network representation is provided to modeling walking at macroscale. This is not the first time that walking space is modeled. In fact, there exists a wide variety of simulation models for walking. The difference is that while these models refer to the micro-scale, the thesis examines the limitations that underlie in the use of large scale datasets. Another contribution regards the method to inject the trips to the network. The suggestion of a point clustering algorithm to create the connectors, offers the possibility to connect the polygon form with the allocation of the trips to the network. For example, through the point clustering, the connectors can be located on different part of the walking infrastructure like different sidewalk sides.

Finally, the combination of different disciplines like transportation modeling and geomatics points out the additional possibilities of bringing together two scientific fields.

7.4 FUTURE WORK

There are several recommendations to extend the possibilities of this work in the future. These are outcome of the time limitation that comes along with the scope of a thesis, and limitations that have to do with the available tools and data.

- Reproject travel data from TAZ to PAZ level.
  The source of the trip data are travel surveys and questionnaires which offer information at the TAZ level. Acquiring data at smaller spatial units is more a more fine grained and demanding procedure. Thus, producing the inputs for the model at PAZ level is still something that needs to be considered. This could be implemented by using existing data available for larger spatial units and project them to the pedestrian alanysis zone level.

- Use built environment data and measures for trip injection and selection of clusters.
  Walking is affected by the surrounding built environment. Including measures of the built environment will make the estimates of walking more realistic. For instance, built environment data could be used to estimate the trips for the origin and destination pairs. Another part of the model that considering built environment data could be for the route choice. More specifically, the number existing commercial uses, the network characteristics or the existence of points of interest could be introduced in a utility function to measure the probability of selecting a route over another.

- Harmonize the scale of walking with the large scale of the FSM.
  The demand of walking trips is estimated along with the demand for other modes of transport at a regional level. Measures and estimates from the whole area are projected to small spatial units, the TAZs. It became clear that treating vehicular trips and walking trips requires working in two scales, large scale for vehicles and smaller scale (but not microscale) for pedestrians. Studying pedestrian trips at smaller geographic units, may require to re-project the scale of walking to the large scale of the whole FSM. In order to harmonize the two scales, a process to bring together the pedestrian networks of different subareas' is needed.

- Validate the model.
  The performance of such a model can be validated with the use of real trip data. Even though usually travel data are not available at the level of the spatial analysis zones, there is an abundance of data produced by pedestrians cell phones and location sharing. Having real information on where people walk, can improve the definition of walking space and network in the model.

BIBLIOGRAPHY

Antonini, G., Bierlaire, M., and Weber, M. (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, 40(8):667–687.

Borgers, A. and Timmermans, H. (1986a). City centre entry points, store location patterns and pedestrian route choice behaviour: A microlevel simulation model. *Socio-economic planning sciences*, 20(1):25–31.

Borgers, A. and Timmermans, H. (1986b). A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. *Geographical analysis*, 18(2):115–128.

Bosa, P. G. (2017). Dynamic assignment models and their application in the portland metro region.

Botsch, M., Kobbelt, L., Pauly, M., Alliez, P., and Lévy, B. (2010). *Polygon mesh processing*. AK Peters/CRC Press.

Castiglione, J., Bradley, M., and Gliebe, J. (2015). *Activity-based travel demand models: A primer*. Number SHRP 2 Report S2-C46-RR-1.

Clifton, K. J., Singleton, P. A., Muhs, C. D., and Schneider, R. J. (2016). Representing pedestrian activity in travel demand models: Framework and application. *Journal of transport geography*, 52:111–122.

Clifton, K. J., Singleton, P. A., Muhs, C. D., Schneider, R. J., and Lagerwey, P. (2013). Improving the representation of the pedestrian environment in travel demand models, phase i.

Cole, R., Leslie, E., Bauman, A., Donald, M., and Owen, N. (2006). Socio-demographic variations in walking for transport and for recreation or exercise among adult australians. *Journal of physical activity and health*, 3(2):164–178.

De Berg, M., Cheong, O., Van Kreveld, M., and Overmars, M. (2008). Computational geometry: Introduction. *Computational Geometry: Algorithms and Applications*, pages 1–17.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231.

Ewing, R. and Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American planning association*, 76(3):265–294.

Flügel, S., Flötteröd, G., Kwong, C. K., and Steinsland, C. (2014). Evaluation of methods for calculating traffic assignment and travel times in congested urban areas with strategic transport models. *TØI report*, 1358:2014.

Frank, L. D. (2000). Land use and transportation interaction: implications on public health and quality of life. *Journal of Planning Education and Research*, 20(1):6–22.

Gibbons, A. (1985). *Algorithmic graph theory*. Cambridge university press.

Handy, S., Cao, X., and Mokhtarian, P. L. (2006). Self-selection in the relationship between the built environment and walking: Empirical evidence from northern california. *Journal of the American Planning Association*, 72(1):55–74.

Handy, S. L., Boarnet, M. G., Ewing, R., and Killingsworth, R. E. (2002). How the built environment affects physical activity: views from urban planning. *American journal of preventive medicine*, 23(2):64–73.

Helbing, D., Farkas, I. J., Molnar, P., and Vicsek, T. (2002). Simulation of pedestrian crowds in normal and evacuation situations. *Pedestrian and evacuation dynamics*, 21(2):21–58.

Helbing, D. and Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282.

Helbing, D., Molnár, P., Farkas, I. J., and Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and planning B: planning and design*, 28(3):361–383.

Hoogendoorn, S. P. and Bovy, P. H. (2004). Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*, 38(2):169–190.

Hoogendoorn, S. P., Daamen, W., Knoop, V. L., Steenbakkers, J., and Sarvi, M. (2017). Macroscopic fundamental diagram for pedestrian networks: theory and applications. *Transportation research procedia*, 23:480–496.

Hwang, F. K., Richards, D. S., Winter, P., and Widmayer, P. (1995). The steiner tree problem, annals of discrete mathematics, volume 53. *ZOR-Methods and Models of Operations Research*, 41(3):382.

Jafari, E., Gemar, M. D., Juri, N. R., and Duthie, J. (2015). An investigation of centroid connector placement for advanced traffic assignment models with added network detail. *Transportation Research Record: Journal of the Transportation Research Board*, 2498:19–26.

Kuzmyak, J. R., Walters, J., Bradley, M., and Kockelman, K. M. (2014). *Estimating bicycling and walking for planning and project development: A guidebook*. Number Project 08-78.

Lawrence, D. L. and Low, S. M. (1990). The built environment and spatial form. *Annual review of anthropology*, 19(1):453--505.

Litman, T. and Steele, R. (2017). *Land use impacts on transport*. Victoria Transport Policy Institute Canada.

Liu, X. and Andersson, C. (2004). Assessing the impact of temporal dynamics on land-use change modeling. *Computers, Environment and Urban Systems*, 28(1-2):107–124.

McNally, M. (2007). Chapter 3: The four step model. *Handbook of Transport Modeling*.

McNally, M. G. and Rindt, C. (2008). The activity-based approach.

Meyer, M. D. and Miller, E. J. (1984). Urban transportation planning: a decision-oriented approach.

Nakamura, K. (2016). The spatial relationship between pedestrian flows and street characteristics around multiple destinations. *IATSS Research*, 39(2):156–163.

Olszewski, P. (2007). *Walking as a mode of transport - a planning and policy perspective*.

Ortuzar, J. d. D. and Willumsen, L. (1996). Modelling transport. wiley, new york.

Pinjari, A. R., Pendyala, R. M., Bhat, C. R., and Waddell, P. A. (2007). Modeling residential sorting effects to understand the impact of the built environment on commute mode choice. *Transportation*, 34(5):557–573.

Qian, Z. S. and Zhang, H. (2012). On centroid connectors in static traffic assignment: Their effects on flow patterns and how to optimize their selections. *Transportation Research Part B: Methodological*, 46(10):1489–1503.

Raford, N. and Ragland, D. R. (2005). Pedestrian volume modeling for traffic safety and exposure analysis.

Rodrıǵguez, D. A. and Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D: Transport and Environment*, 9(2):151 – 173.

Saelens, B. E., Sallis, J. F., and Frank, L. D. (2003). Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of behavioral medicine*, 25(2):80–91.

Santos, G., Maoh, H., Potoglou, D., and von Brunn, T. (2013). Factors influencing modal split of commuting journeys in medium-size european cities. *Journal of Transport Geography*, 30:127–137.

Schadschneider, A. (2001). Cellular automaton approach to pedestrian dynamics-theory. *arXiv preprint cond-mat/0112117*.

Scheiner, J. (2010). Interrelations between travel mode choice and trip distance: trends in germany 1976–2002. *Journal of Transport Geography*, 18(1):75–84.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., and Xu, X. (2017). Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):19.

Schwartz, W., Porter, C., Payne, G., Suhrbier, J., Moe, P., and Wilkinson III, W. (1999). Guidebook on methods to estimate non-motorized travel: Supporting documentation. Technical report.

SFS, O. (2010). *Implementation Standard for Geographic information - Simple feature access*.

Sheffi, Y. (1985). Urban transportation networks.

Singleton, P. A. and Clifton, K. J. (2013). Pedestrians in regional travel demand forecasting models: State-of-the-practice. In *92nd Annual Meeting of the Transportation Research Board, Washington, DC*, pages 13–4857.

Stead, D., Williams, J., and Titheridge, H. (2000). Land use, transport and people: identifying the connections. *Achieving sustainable urban form*, pages 174–186.

TRB (2007). *Metropolitan Travel Forecasting: Current Practice and Future Direction–Special Report 288*. Number 288. Transportation Research Board.

WATS (2008). *Travel demand model improvements. Technical memorandum 4- Zone centroid and centroid connector placement. Washtenaw Area Transportation Study*.

Watson, K. B., Carlson, S. A., Humbert-Rico, T., Carroll, D. D., and Fulton, J. E. (2015). Walking for transportation: what do us adults think is a reasonable distance and time? *Journal of physical activity and health*, 12(6 Suppl 1):S53–S61.

Weisstein, E. W. (1666). von neumann neighborhood. http://mathworld.wolfram.com/vonNeumannNeighborhood.html.

Woldeamanuel, M. G. (2016). *Concepts in Urban Transportation Planning: The Quest for Mobility, Sustainability and Quality of Life*. McFarland.

A RELATIONSHIP BETWEEN BUILT ENVIRONMENT AND WALKING

A.1 FACTORS THAT AFFECT WALKING AS MODE CHOICE

There are numerous parameters that define how frequently people walk, whether people walk and where walking takes place. For instance, sociodemographic characteristics like income, age and gender, vehicle occupancy, among other factors, have influence on the decision of travel mode (Santos et al., 2013). Additional factors of the environment like safety of an area, time of day/year also determine the decision to walk (Saelens et al., 2003).
Besides these factors, the last two decades, a significant amount of studies have emerged that acknowledge the relation and influence of the aspects of the built environment with the travel behavior and the selection of the transport mode. In order to understand the connection of the built environment with the human activity it is important to make explicit the meaning that the term built environment conveys.
The concept of the built environment refers to the design, construction, management and use of all the infrastructure that is man-made, as long as it has relationship to human activities over time. Generally, it encompasses places and spaces that are created by people and in a more abstract view, to any physical adjustment of the natural environment through human activities (Lawrence and Low, 1990). Handy et al. (2002) defines the built environment as a constantly changing system that is comprised of *urban design*, *land uses* and *transportation system*, as well as the patterns of human activity within the physical environment. The built environment is a multidimensional system in which urban design deals with the physical arrangement, the function and appearance of the elements of a city. The term land use is referring to the distribution, the location and the density of different types of human activities (Handy et al., 2002). Examples of categorized human activities can be described as residential, commercial, recreational, and other types of land uses. Finally, the transportation system refers to the physical infrastructure related with transportation activity, such as the roads or the sidewalks (Handy et al., 2002).
Travel behavior is a human activity that is strongly affected by the multiple elements of the built environment. Especially, pedestrian trips are affected in a different manner comparing to the other motorized transport modes. Pedestrian movement is freely performed and do not strictly follow the network. Due to the relative low speeds, and short distances that a pedestrian could go through, walking is affected

by the surrounding built environment. Through the literature, the dimensions of the built environment that affect the travel behavior with a specific focus on walking activity can be categorized to the following:

Size of the city. In a research conducted by Scheiner (2010), the mode choice is studied with respect to city size, trip distance and car availability. Comparing walking propensity in large cities and small towns, the results indicate that car owners that live in cities are more inclined to walk a specific distance that the car owners that reside in towns. Generally, short trips (¡1.5 km) are more possible to be conducted with non- motorized modes in large cities than in small towns.

The structure of the area. This dimension is applied to the regional scale unlike the following dimensions which focus on local/ neighborhood scale. The regional structure reflects the form of residential development. One aspect of the structure is the centralization or decentralization which can be seen as the continuity/discontinuity of development. It is formed by the distribution of activities and transportation facilities within the region (Handy et al., 2002). A disperse developed region with low density is more car oriented while a compact region with higher densities stimulates non-motorized means of transport. Another aspect of the regional structure is the centrality. This attribute is reflected by the jobs or the extent of commercial, recreational and other important functions of an area. The centrality of a region can imply shorter distances among areas and functions, thus non-motorized means are promoted (Litman and Steele, 2017). Looking at a region through this dimension it can be characterized as monocentric if the central activities are identified in a single area or as polycentric in the case where there are multiple hubs of central activities.

Density and intensity of development. This refers to the amount of activities found in an area and can be defined by measuring demographics (i.e. employment) and/or by infrastructure measures (i.e. building area per unit of area, floor-area ratio) Frank (2000). The effect of the activity density to the generation of pedestrian trips has been extensively researched (Handy et al., 2002; Stead et al., 2000; **?**; Rodrıǵguez and Joo, 2004; Ewing and Cervero, 2010). The results show that high population density, job density and resident density reduce distances and increase the travelling on foot. Though, Rodrıǵguez and Joo (2004) in their research on the relationship between physical environment and mode choice, find that the employment density in the destination is more important factor than the residential density at the origin of a trip. The same is supported by Santos et al. (2013) as they find that population density is not a significant measure for mode choice due to the collinearity with other variables.

Mix of land uses. It is defined as the relative proximity of different land uses in a given area. Generally, the mix of land use can be measured at many levels, from a building to a street or to a neighborhood. The mix of land uses affects the physical separation of the activities and therefore the trip generation (Stead et al., 2000). The lack of land use mix namely the separation of activities generates need for trips of longer distance and therefore the motorized means of transport are promoted. By increasing the mix of land use travel distances are minimized and thus pedestrian trips are promoted (Litman and Steele, 2017).

Street network. The design of the street network and characteristics like the block size, the street width the infrastructure quality and the network form, can directly affect the travel behavior. The design of streets in a way that the speed of motorized means is maintained low, can deteriorate the use of cars and promote walking (Litman and Steele, 2017). In this dimension of built environment three characteristics of street networks are important, connectivity, accessibility and scale. The notion of *connectivity* refers to the directness and the available alternative routes of a network from one point to another. It can be also described as the existence of connection among all the transport networks (street, public transport, pedestrian and bicycle networks). Connectivity can be measured by the number of blocks or intersections per unit of area (Frank, 2000). Increased connectivity within an area, results in a more dense road network, that provides more alternative routes and with shorter distances between start and end point that which is more appealing for walking (Frank, 2000; Handy et al., 2002). *Accessibility* is described as the extent to which a network can be accessed and is mainly measured with the flows (Nakamura, 2016). Accessibility is interconnected with connectivity in a way that a poorly connected network with large streets, few vertical connections and cul-de-sacs affects the accessibility of the area. Finally, the *scale* is described by Handy et al. (2002) as human or automobile scale and refers to the tree-dimensional space that is bounded by the buildings.

Additional factors that have been reported as determinants for choosing walking as a mode of transport are the existence of pedestrian sidewalks which promotes walking, and the presence of slope on the terrain, which is observed that decreases the attractiveness of walking (Rodrıǵguez and Joo, 2004). Moreover, the existence of aesthetic qualities that refer to the attractiveness or appeal of a place increase walking trips (Handy et al., 2002).

Factor	Effect	Author
Built environment factors		
City size	Same distance: cities: + towns: -	Scheiner (2010)
Centralization (continuity of development)	+	Handy et al. (2002)
Centrality (existence of important functions)	+	Litman and Steele (2017)
Density and intensity	Population density:+ Employment density: + Residential density:+	Pinjari et al. (2007) Rodrıǵguez and Joo (2004) Handy et al. (2002) Stead et al. (2000) Handy et al. (2006) Ewing and Cervero (2010)
Mix of land use	+	Stead et al. (2000) Saelens et al. (2003) Handy et al. (2006) Scheiner (2010)
Connectivity	+	Frank (2000) Handy et al. (2002)
Accessibility	+	Handy et al. (2006) Nakamura (2016)
Other factors		
Physical activity options (bike routes, sidewalks, parks, public transit)	+	Handy et al. (2006) Rodrıǵguez and Joo (2004)
Car ownership	-	Scheiner (2010)
Safety (quiet, low crime, low traffic, street lighting)	+	Handy et al. (2006)
Attractiveness	+	Scheiner (2010) Rodrıǵguez and Joo (2004) Handy et al. (2006)

Table A.1: Factors that affect walking as a mode choice. (+ positive impact on walking / - negative impact on walking)

B DATA PRE-PROCESSING, IMPLEMENTATION DETAILS AND WORK-FLOWS

B.1 CREATE WALKING AREA

Input dataset for the walking area is a gml file provided by the PDOK website. To import the gml file in python, it was first inserted in QGIS and exported as .shp. The worflow of getting the final output file is presented in Figure B.1.

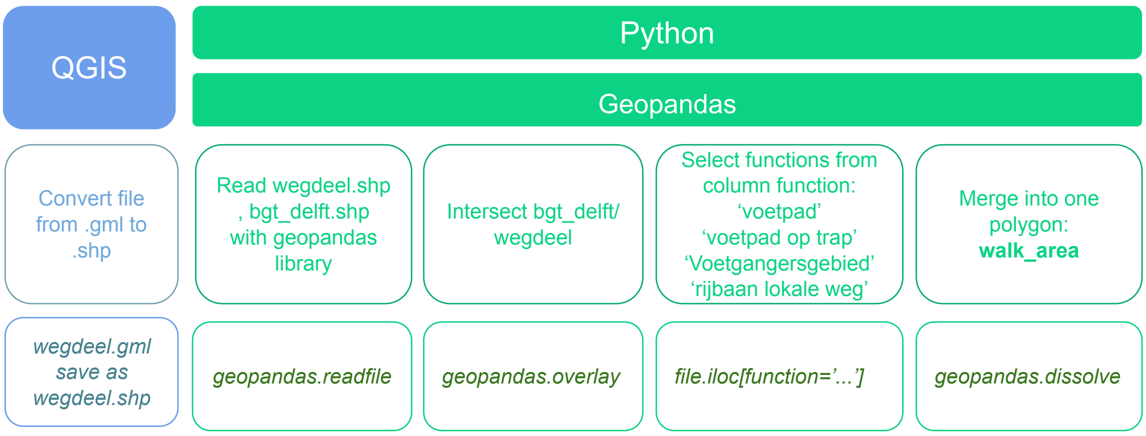


Figure B.1: Workflow of walking area creation

B.2 POLYGON TRIANGULATION

The triangulation of *walking_area* polygon is performed with the use of *Python* library *tri*. The result of this library is actually a CDT with triangles covering the whole area. Removing the triangles that are not contained in the polygon requires a spatial overlay function between two polygon datasets the *triangulation.shp* and the *walking_area.shp*. Due to the large dataset of the triangles, instead of performing intersection of polygons the centroid of each triangle is used and the spatial overlay function point within polygon is performed (Figure B.1). The *Python* library *geopandas* is used for handling all the spatial datasets and performing spatial analysis functions. Within *geopandas*, for the overlay functions, spatial indexing is used by default, resulting to important decrease of the computation time.

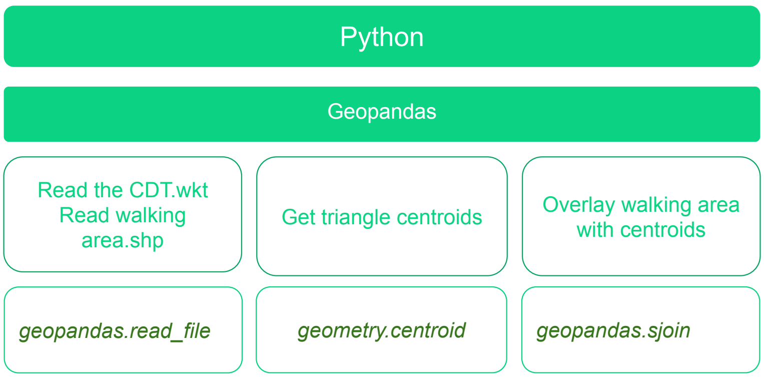


Figure B.2: Workflow of final triangulation inside walk area polygon

B.3 NETWORK PREPARATION

Network preparation

1. Import dual_graph.shp in the database/ table name:dual_graph

2.
```
alter table dual_graph
     add column source integer,
     add column target integer,
     add column cost_len double precision,
     add column cost_time double precision,
     add column rcost_len double precision,
     add column rcost_time double precision,
     add column x1 double precision,
     add column y1 double precision,
     add column x2 double precision,
     add column y2 double precision,
     add column to_cost double precision,
     add column rule text,
     add column isolated integer
```

3.
```
alter table dual_graph
rename column geom TO the_geom
```

4.
```
select pgr_createTopology('dual_graph', 0.001);
```

5.
```
select pgr_analyzeGraph ('dual_graph',0.001,
'the_geom','id','source','target');
```

6.
```
update dual_graph
     set x1 = st_x(st_startpoint(the_geom)),
         y1 = st_y(st_startpoint(the_geom)),
         x2 = st_x(st_endpoint(the_geom)),
         y2 = st_y(st_endpoint(the_geom))
```

7.
```
update dual_graph
     set cost_len =
st_length_spheroid(the_geom,'SPHEROID["WGS84",637
8137,298.25728]')
```

8.
```
update dual_graph
     set cost_len =
st_length_spheroid(the_geom,'SPHEROID["Bessel
1841",6377397.155,299.1528128]')/1000
```

Figure B.3: Create topology for the network in *postgreSQL* with *pgRouting*

After building the topology, of each graph, the unconnected components are removed with the *NetworkX* function *connected_components* in *Python*. This function uses a Depth First Search (DFS) algorithm, applied on the edge table of the graph, and returns the connected components. The implementation is presented step by step in Figure B.4.

Before implementing this process in *Python* an attempt was made with the *postgreSQL* function *pgr_connectedComponents*. Nevertheless, this is an experimental function in *pgRouting* and the resulted disconnected components were not correct.

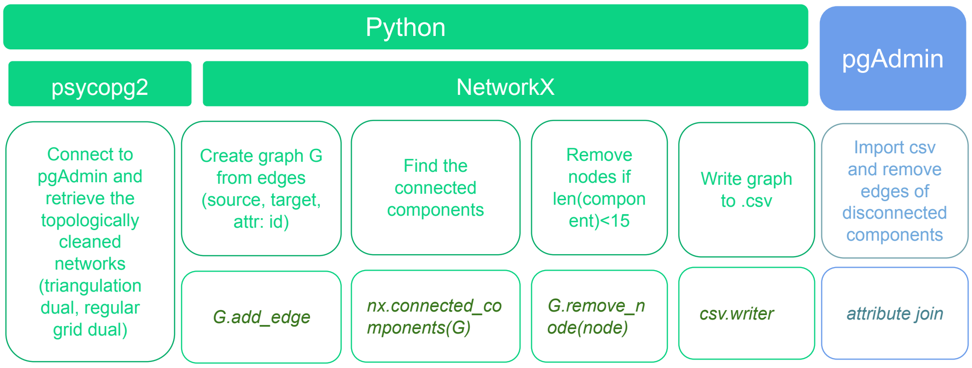


Figure B.4: Removing disconnected components full steps in *Python* and *pgAdmin*

B.4 APPLYING DBSCAN

B.4.1 Inputs and outputs of DBSCAN

Implementing DBSCAN on the graph nodes, requires some preliminary steps concerning the data processing and the preparation of data structures that will be used. Firstly, OD matrix is a sparse matrix with most of its elements being zeros. To avoid the time and storage needed for processing the zeros, the non-zero elements of the sparse OD matrix are stored in a list such that :

$$\text{OD\_ list} = [[\text{i}, \text{j}, \text{ trip \_num}]]$$

Another prerequisite is to find for each PAZ which points of the graph it contains. To avoid repetitive spatial overlay functions, the graph points dataset is overlaid with the zones, and the zone id is added as an extra column in the points data structure.
The output of this process is a dictionary data structure containing the injection points for each zone id such that:

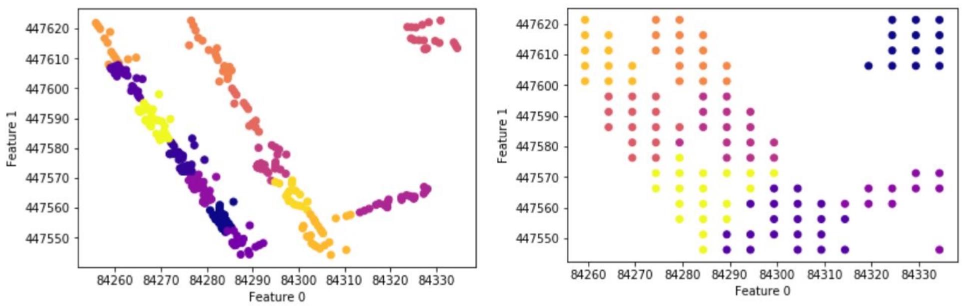
$$\text{injection\_ pts} = \{\text{zone\_id} : [\text{p1}, \text{p2}, .., \text{pn}]\}$$

B.4.2 Comparison of DBSCAN with other clustering algorithms

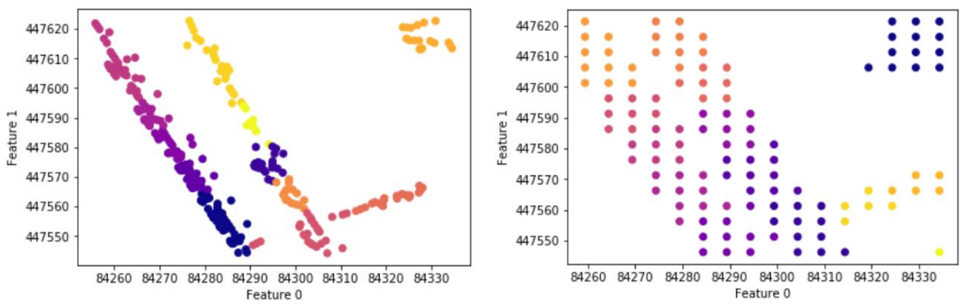
Among the various point clustering algorithms, a comparison on those that do not require as input a predefined number of clusters is performed. DBSCAN is compared with the Affinity propagation alogrithm and the mean shift algorithm Table B.1. DBSCAN outweigh the other two algorithms in terms of speed performance and clustering result. The difference on the clustering result among the three algorithms is visualized in Figure B.5.

Affinity propagation	Mean shift	DBSCAN
slow in execution	slow in execution	remarkably fast
assume globular clusters	assume globular clusters	clusters do not need to be globular
hard to set the right parameters	intuitive and meaningful parameters	hard to get parameter epsilon right
stable across runs	stability under runs can vary	stable across runs

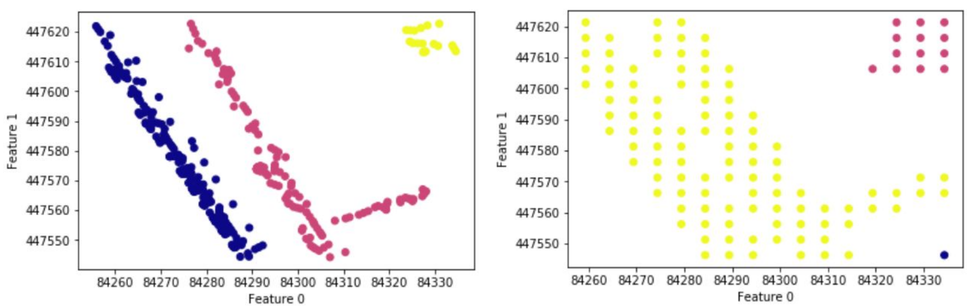
Table B.1: Differences of DBSCAN, Affinity propagation and mean shift point clustering algorithms. scource: hdbscan Clustering Library



(a) Point clustering with Affinity propagation algorithm



(b) Point clustering with Mean shift algorithm



(c) Point clustering with DBSCAN algorithm

Figure B.5: Comparison of point clustering algorithms

B.5 STEPS NEEDED TO IMPLEMENT ROUTE CHOICE WITH A*

A* algorithm for the all-or-nothing route choice was implemented in *pgRouting*. The complexity of A* implemented in *pgRouting* is $O((E+V)*\log V)$ where E is the number of graph edges and V is the number of graph vertices.
Before implementing A*, a preparation of each graph was needed. In particular, the connectors were appended to the graph edges and the network topology was created for the new graph. After creating the new graphs topology two tables are created containing the information of the graph: *graph_edges* and *graph_vertices*. These table were renewed by passing the centroid id to the graph vertices. This step is performed because the information about the trips' origin and destination is at zone (equivalent to centroid) level. At this point, the A* was implemented to the *graph_edges*. The final step is to remove the connector edges from the resulting routes. The process is analytically presented step-by-step in Figure B.6, Figure B.7, Figure B.8, Figure B.9.

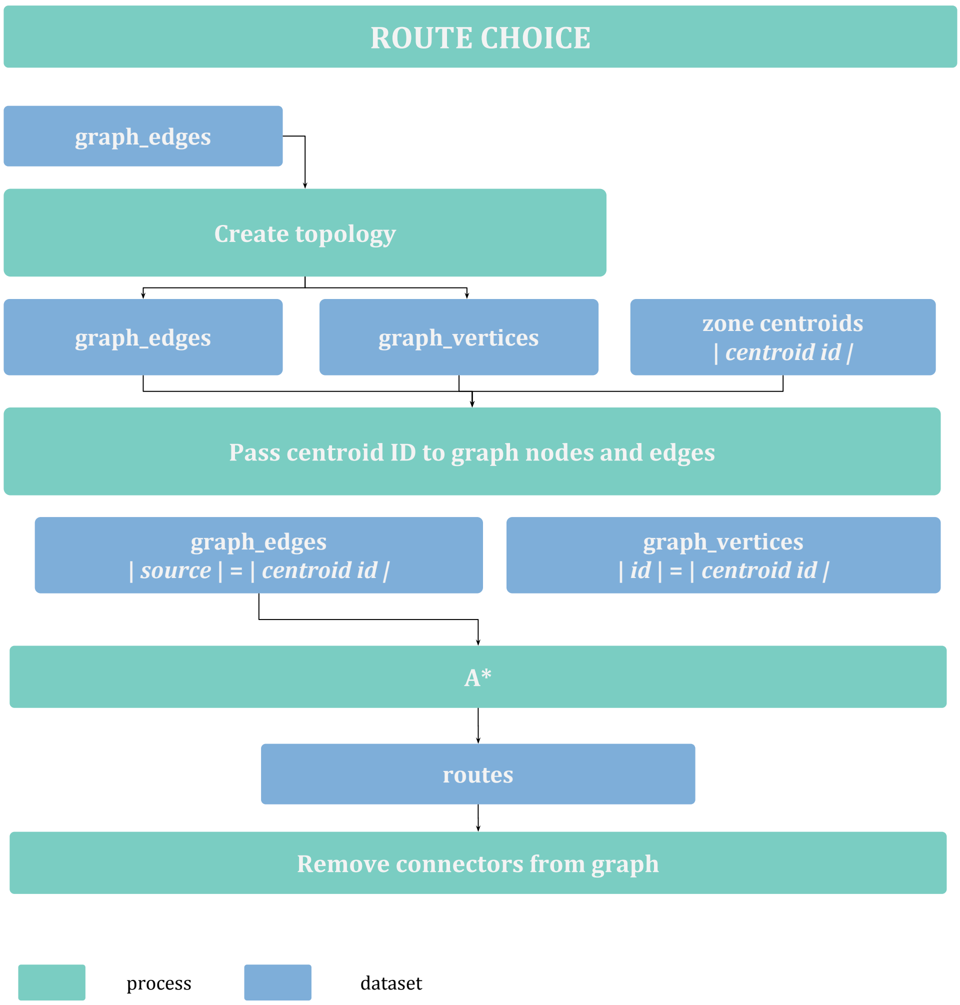


Figure B.6: Steps of route choice implementation

Pass centroid ID to graph nodes and edges

Implementation in pgAdmin

input datasets: graph_edges | graph_vertices | zone_centroids

```
create table centroids_equal_to_graph_nodes as
    (select a.id, b.id as cntr_id
    from graph_vertices a, zone_centroids b
    where ST_within(a.the_geom ,st_buffer( b.geom,0.05)))
```

```
update graph_vertices
    set id = cntr_id+30000 #add large int to avoid duplicate ids
    from con_reg_centr_id
    where graph_vertices.id=con_reg_centr_id.id
```

```
update graph_edges
    set source_zone_id = cntr_id+30000 #add large int to avoid duplicate ids
    from centroids_equal_to_graph_nodes
    where graph_edges.source=
    centroids_equal_to_graph_nodes.id
```

Figure B.7: Procedure of passing centroid id to the graph edges and graph veritces

A*

Implementation in Python with *psycopgr*

input datasets: graph_edges | od_pairs = [orig, dest, trip_num]

```
for pair in od_pairs:
    orig=pair[0]
    dest=pair[1]
    sql=( "with temp as (
        SELECT * FROM pgr_astar(
            'SELECT id, the_geom, source_zone_id as source, target, cost_len
            as cost, x1, y1, x2, y2 FROM graph_edges', %s, %s, directed :=
            false, heuristic := 1))select t.seq, t.path_seq, t.node, t.edge,
            t.cost, t.agg_cost , ST_AsText(a.the_geom) from graph_edges a
            join temp t on a.id=t.edge" ,(orig,dest))
```

Figure B.8: Applying A* algorithm on graph edges

Remove connectors from graph

Implementation in pgAdmin

input datasets: graph_edges zone_connectors

```
create table routes_no_connectors as(
    select a.*
    from graph_edges as a
    where a.id not in (
        select a.id
        from graph_edges a, connectors b
        where ST_Equals(a.the_geom , b.geom)))
```

Figure B.9: Remove connectors from graph edges

COLOPHON

This document was typeset using LaTeX. The document layout was generated using the arsclassica package by Lorenzo Pantieri, which is an adaption of the original classicthesis package from André Miede.