ERASMUS UNIVERSITEIT ROTTERDAM

The Impact of Structure and Growth Parameters on Street Networks' Efficiency

Master Thesis

Operations Research and Quantitative Logistics

Student: Marie-Sophie Graftiaux

Student number : 430465

Supervisor: Dr. Paul Bouman

Second reader : Dr. Remy Spliet

January 2021

Abstract

In this thesis we lay the foundation for simultaneous city network growth simulation and the network efficiency and vulnerability analysis. In a first time the growth of several street networks is simulated were the shape and growth parameters are varied over the different simulations to analyse their separate effect on the efficiency and vulnerability. Subsequently street hierarchy is added to the networks. Three performance measures are then introduced, a global efficiency measure using the optimal value of the traveling salesman problem, an average efficiency measure using the inverse of the shortest path and a vulnerability measure. The results show that expenses in the development of the organisation of cities only show effect in the beginning by increasing efficiency. Once a threshold is reached the effect is not increasing anymore. We also find that the more spacious cities are, the more efficient they become and the less likely they are to be vulnerable to disruptions. Finally we find that there is a trade-off between the efficiency and vulnerability in growing cities. The more a city grows the less efficient it becomes but it also becomes less vulnerable to disruptions.

Contents

1 Introduction

According to the United Nations around 55 percent of the world's population is currently said to live in urban areas. This number will only rise in the following decades and is estimated to be around 70 percent in 2050 (Publications, 2019). This means cities all around the world will keep growing. This growth comes with a lot of challenges like employment, health care, social services, public safety, waste disposal and transport (Heinke, 1997). In this thesis, we focus on the last aspect: the transportation and in particular on street networks, as street networks are the transportation medium for the urban population distributed over an area (Dingil et al., 2018). Analysis on network efficiency can be important for policy makers such that they know the repercussions growth policies might have on the efficiency of the transportation networks in their city.

This phenomenon is not recent. In Europe, most cities have expanded from small towns in the middle ages. These are said to have grown organically, while others for example in the United States were planned from the moment the first settlements took place and can be recognised by their grid shaped centres. Not only the basic structure but also the growth patterns differ per city. Some cities grow individually with every new built settlement others grow in a more planned manner with the addition of whole neighbourhoods. The formation of structural elements in the growth process of a city is typically referred to as the morphogenesis of a city.

In the past decade, a number of mathematical models for the morphogenesis of cities were proposed, along with several statistical measures of the underlying geometrical structures of cities, their morphology. Most of these models are based on the raw street networks of cities that can be represented in a graph structure. This street based approach considers segments of streets as the elementary components of a city and assumes they contain information about further characteristics. An opposite approach would be to consider build up areas (such as building and parks) as the main information carriers in those networks (Courtat et al., 2011). In this thesis we combine several models to simulate the growth of synthetic street networks with different morphological and morphogenic characteristics.

We first simulate the growth of raw street networks that bear similar characteristics to growing and grown cities. Next to the raw networks we introduce a hierarchy measurement that distinguishes between different kinds of streets (e.g. residential streets vs arterial roads). Furthermore we are interested in the effect the morphogenic and morphological parameters have on the efficiency of the networks. By singling out parameters we investigate possible relationships between certain characteristics and the design of efficient street networks. Next to that we analyse if the growth of a city influences the efficiency of its network and if there is a difference between the growth of cities with different parameters.

To our current knowledge the morphological and morphogenic traits of cities have not been evaluated in a simulation study. With this thesis we lay the ground work to combine the field of correct city structure and growth simulation and the performance evaluation of those simulated cities.

To evaluate the performance of the simulated networks we use three measures, the first two evaluate the efficiency of the networks and the last one the vulnerability of the networks towards disruptions. The first efficiency measure looks at the overall efficiency of the network by solving a Traveling Salesman Problem (TSP). To be able to compare the optimal value between different sized cities, the deviation of the solution of the TSP using the Euclidean distance between the intersections and the shortest distance is used. The second efficiency measure uses the inverse of the shortest path between all vertices in the graph to calculate an average efficiency measure. The last performance measure evaluates the vulnerability of networks towards external disruptions. Those can for example arise from traffic or natural disasters.

We find that networks that are not organised, in other words 'tree' shaped networks, where settlements are only connected to the rest of the network by a single street have a large negative impact on the overall efficiency of the network. As soon as a minimum of organisation is added this effect disappears. This suggests that expenses in the development of the organisation of the city is only necessary until a certain 'connectedness' is reached. The spacing between new settlements and the existing network has an effect on the average efficiency of the network. The larger this spacing the more efficient the network becomes and it also becomes less vulnerable to disruptions. We find that the choice between inner development and outer development of a city barely affects the overall efficiency and vulnerability. The average efficiency on the other hand is slightly affected and we observe an increase when outer development is chosen over inner one.

When we look at the growth of individual cities we discern a trade-off between efficiency and vulnerability. In general the larger a city becomes the less efficient it is (overall and on average) but it also is less vulnerable to disruptions.

In the next chapter we present a literature review on the current solution methods and algorithms used for similar types of problems. In Chapter 3 we explain how we simulate different street networks. In Chapter 4 we explain the methods used to evaluate the efficiency of street networks and the statistical analysis tools. Next in Chapter 5 we analyse the results. In Chapters 6 & 7 we formulate our conclusions and discuss the possibilities to extend the research.

2 Literature Review

This thesis can be split into three parts. First we aim to accurately simulate the growth of street networks. Next we evaluate the efficiency of these simulated networks with three different performance measures. We then analyse the possible relationship between certain growth characteristics and the performance measures of the street networks using statistical tools. In this chapter we give an overview of the relevant literature linked to these three parts.

2.1 Morphology and Morphogenesis of Cities

Most studies analysing cities and their current morphology, that is their inherent characteristics such as form, shape and structure, focus on the modelling of static networks. In this thesis we are also interested in the formation and development of networks over time, their morphogenesis. The main difference between morphology and morphogenesis is that the first one looks at the form, shape and structure of grown cities, whereas the second one looks at the formation of said structures. That said the final structure of a city is dependent on how it has been formed such that morphology is also used in morphogenesis studies. In the following we present several morphogenic models for the simulation of cities.

Cellular automata (CA) and multi agent models define the networks as a set of blocks in a grid-environment. For each cell in the grid, which represents a piece of land, a certain probability of expansion to another cell is assigned. These probabilities are based on the behavior of biological phenomenon such as forest burns in CA models and on the interactions of humans in multi agents models. By applying these techniques city growth models can be developed such as done by Semboloni et al. (2004).

Frankhauser (1998) argues that cities' structures can be described by fractal analysis. Fractal dimension can model spatial distribution and measures non-topological aspects of cities such as different types of structures. He argues that morphogenesis models should respect this morphology of cities to be correct. He also admits that this representation lacks the definition of hierarchy in a city, which is the terminology used to differentiate highways from small streets. The strength of the CA and fractal methods lies in the simulation of the dynamics of land use and population, and are better to analyse the growth of existing structures (Courtat et al., 2011).

Since Euler's famous solution of the "Bridges of Königsberg" problem, an intuitive approach to model a city is by using a graph, where streets are the edges and the vertices represent the intersections. This approach implies that the street network provides an intuitive and relevant model of a city's structure. This approach of modelling a city is more common and also has the advantage that it provides a mathematical structure that can be used to run optimisations. The following studies use a graph structure to quantitatively model cities.

The research conducted by Pousse (2020) defines road networks of cities as a subdivision of existing spaces into smaller ones creating roads that cut through spaces. This is inspired by the tessellation of objects in the nature such as the crackles appearing when clay dries up. Pousse (2020) finds that the logarithm of street lengths of cities follow a log-normal distribution. He develops a method to recreate this network of streets by the subdivision of spaces in a square. This method allows to construct a network with very similar street length and topological distances as in real networks. This model is rather a static one and lacks the evolutionary aspect of a city. Furthermore the graph representing the city correctly imitates the street lengths of a city but does not imitate the other characteristics of a city, such as the form of a city, the non perpendicular street structures, natural elements etc. It is clear from this research that an accurate model of a city should have log normally distributed street lengths and this is what we strive for in our model.

Barthélemy and Flammini (2008) introduce a morphogenesis model for cities' street networks. Their approach takes statistical characteristics of cities around the world as a basis. From there they introduce a simple city growth model to reproduce those characteristics by growing a city step by step. The biggest assumption of their model is that new centres (which represent new homes, buildings or businesses that need to be linked to the existing network) are independently distributed. Even though this assumption has been revoked by Hagget and Chorely (1969) among others, their model accurately imitates city growth and the generated cities have empirically similar characteristics to existing cities. The model is more simple than the one proposed by Courtat et al. (2011) but is still valid and allows for lots of generalisations.

In this thesis we base the morphogenesis model on Courtat (2012)'s 'potential' approach. Courtat (2012) introduces a more elaborate model for morphogenesis than Barthélemy and Flammini (2008), where he analyses in more detail certain characteristics of cities and offers a model to simulate the growth of a city using specific growth parameters. His approach is based on a graph structure of the city network, which has advantages for the efficiency analysis we conduct on the simulated cities later on. The difficulty of the model constrains it from generalisations but offers more freedom with the adjustment of growth parameters and more closely simulates real cities. The overall accuracy of the growth model comes at the costs that cities are not as realistic in certain statistical aspects such as the cities created by e.g. Pousse (2020) that have street length distributions close to the observed one in cities around the world. We elaborate on Courtat (2012) morphogenesis model further in the next chapter.

To broaden the scope of the simple road networks as created by Courtat (2012) we also introduce some literature that investigates the dynamic flows and vulnerability of networks. Levinson and Yerra (2006) introduce an interesting model that integrates the traditional travel demand model with an agent-based revenue, cost and investment model. On a predefined network the authors propose to allocate uniformly and randomly distributed initial speeds and land use. Their model then calculates the link flows using a shortest path algorithm. They also add a cost model by collecting a toll at every passage, if the revenue exceeds the costs, maintenance can be applied on a link if this is not the case the link erodes and becomes slower. The process is repeated until an equilibrium is found or if there is no equilibrium. This model gives the opportunity to experiment with the network dynamics but the authors comment that the model is not appropriate for actual predictions on real networks. As our cities are simulated and real dynamics can not be observed this model can nevertheless be used to introduce dynamics into the model.

Yin and Xu (2010) propose a network efficiency model to define the structural vulnerability of the road network. Their model can effectively identify the critical components of a road network without analysing traffic data and dynamic flows. This research can be used to identify if a network is robust to disruptions and can therefore be used as a performance measure to analyse networks.
The above mentioned two works are used in this thesis to refine the created networks based on Courtat (2012) by adding flow dynamics and vulnerability to the network. The goal of these measures is to add another dimension of realism to the networks.

2.2 Efficiency analysis

In this section we discuss some literature on the efficiency analysis of the generated street networks. This can be achieved by solving a relevant optimisation problem on the graph structure of the network. A classic choice when thinking of transportation related optimisation problems is the Traveling Salesman Problem.

The Travelling Salesman Problem (TSP) is a combinatorial problem where a salesman has to visit n cities and return to the starting city while minimising the traveling costs made (which can be defined as the total travel distance). The Hamiltonian cycle problem, which is the problem of finding a cycle in a graph that visits each vertex once, has been proven to be $\mathcal{NP}$-hard by Karp (1972). This implies the $\mathcal{NP}$-hardness of the TSP, which means that there does not exist (to our current knowledge) a polynomial time algorithm to solve these instances. This problem can be slightly adapted to our generated networks if we define the n cities as the n intersections and the connections between the cities as the streets between two intersections in our model. As we do not consider one way streets we can define the distance from one intersection to another as the same in both ways. We therefore consider the symmetrical TSP.

The TSP is a widely applied problem that has been studied for several years. As the problem is $\mathcal{NP}$-hard exact solution approaches for large problem instances can take a very long time to solve. Naive solution approaches such as enumeration only work for instances for up until 10 cities (Ozden et al., 2017). A few exact algorithms have been developed including a branch-and-bound approach to attempt to reduce the computation times. Based on Applegate et al. (2006) branch-and-cut techniques to solve the symmetric TSP to optimality, the `concorde` code has been developed to solve instances to optimality. This code is the current exact method of choice for solving large instances as it was able to solve a 85,900 cities problem in 2013 (Ozden et al., 2017).

As we mentioned before the TSP aims to minimise the travelling costs in the tour of the salesman. The definition of these costs can for example be the Euclidean distance between two cities. In our network the vertices are not cities but intersections and the distance between two intersections is not best defined by the Euclidean distance. There are two possible realistic measurement options for the distance between two intersections. The first and most straight forward one is the shortest path. Next to that, Courtat (2012) introduces the topological distance measure, this measurement can be seen as the simplest path. It minimises the amount of turns to reach the destination while still trying to minimise the distance. This topological distance is more realistic as it mimics the behavior of human movement through streets.

As a distance measure for the TSP we need to compute all the pairs of shortest paths in the networks. This problem is known in the literature as the All Pair Shortest Path Problem (APSPP) and has been studied by several researchers over the past years. The complexity still remains open but is largely researched for dense graphs Chan (2010). For a sparse graph the best known algorithm is the repeated application of Dijkstra's algorithm resulting in a complexity of $O(n^2\log(n) + mn)$, where n is the number of vertices and m the number of edges. For arbitrary dense graphs the Floyd-Warshall algorithm can be used and runs in $O(n^3)$ (Chan, 2010). In more recent literature several more complex algorithms have been proposed especially for dense graphs that break the $O(n^3)$ runtime and classify as low as $\frac{n^3}{2^{\Omega(\log n)^{1/2}}}$ according to Williams (2014).

2.3 Statistical Evaluation

The goal of this thesis is to analyse the possible relationship between several growth and shape parameters of cities and the efficiency of the street network. There are several methods in the statistical learning field to evaluate this relationship. To be able to infer a relationship between the characteristics and the efficiency we need to be able to model what kind of efficiency certain characteristics imply. A quite straight forward analysis can be done using a simple multiple linear regression model (Heij et al., 2004). If the relationship between the variables seems to be more complex more flexible models can be used.

3 City Morphogenesis Simulation Process

In this chapter we elaborate further on the data generation process. There are several ways to translate street networks into geometrical forms. In this thesis we choose for a graph representation of the street network as it is a largely used, intuitive form of representation and a preferred form for the performance analysis as described in Chapter 4. Other ways include the use of a cellular system (Semboloni et al., 2004) or fractals (Frankhauser, 1998) to model street networks.

There are two possible ways to retrieve graph structures of cities. First from existing cities and secondly through simulation. Publicly available street data from `openstreetmap.org` makes it possible to retrieve a graph structure for every city in the world. As each city is unique and follows a lot of potentially different morphogenesis and morphological parameters it is difficult to compare single parameters across cities. In this thesis we therefore choose to generate cities the second way through a simulation. This allows us to control and compare different parameter settings with each other and facilitates the ensuing statistical analysis.

To simulate the street networks of a cities the most accurately possible, we implement the framework and model introduced by Courtat (2012), that models and simulates the morphogenesis of cities. This model is based on the raw street network of cities that are represented in a graph structure. This approach considers that the network contains information about further characteristics of cities and can therefore be used to simulate the growth of a city. Courtat (2012) analyses several French cities and draws statistical conclusions about methods and parameters, that can be used to reproduce the growth of these cities synthetically in a simulation. We use the results of this research and extrapolate them to grow different kinds of cities by tuning parameters and therefore quantify the inherent structures of cities.

In a first time we elaborate the above mentioned graph structure that is used to simulate cities. A weighted graph $G = (V, E)$ is a structure compromising a set of vertices V and a set of edges E. Each edge (v_1, v_2) is a link that connects vertices v_1 and v_2 and that has a specific weight w assigned to it. In the street network the vertices represent intersections and the edges represent street segments with no intersection. The degree of the vertex represents the type of intersection at hand. For example if the degree is 3 the vertex represents a T-junction where a street meets another but does not cross it. At first the weight on each edge represents its length in metres, later we introduce a link speed to add street hierarchy to the network. By this we refer to street or parts of streets that are larger or in better condition and thus allow for a better flow.

Next we define the street network of a city as the result of several operations over time. These operations are defined by the building of new so called *centres*. Those represent new homes, businesses, etc., that are attached to the network by the building of streets (or street segments). Furthermore we assume the conservation of existing structures. That means that once a street is built it stays part of the street network. Lastly the network is subject to limitations as to its expansion, size etc. These limitations come forward from the main characteristics of cities.

We define four main characteristics of a city: its planning, construction, organisation and sprawling. First the *planning* of a city can either be organic or centralised. In an organic environment the city grows with each new centre. Under different conditions in centralised cities an authority or different ruling instance plans the coherent and simultaneous construction of the street network. An example of a planned city is the city of Brasilia, that has been designed and constructed to be the capital of Brazil, it is planned to look like a plane. Next, the ability to add new streets to the network is called the *construction* factor of a city. The feasibility to add new centres or streets to the network might be encumbered by natural structures in and around the cities such as rivers, lakes or mountains. A cities *organisation* can reach from a random set of settlements to a highly organised structured street network. The typical American city with parallel streets is an example of a highly structured street network, a slum displays the contrary example, as there settlements are linked randomly through the network in a 'tree'-like fashion. Lastly a city's *sprawling* factor is the amount of compromise it makes between its inner development and its outer growth. The higher the factor the more the city is going to expand outside its boundaries. The lower it is the more it is going to allow for new settlements inside its borders. Dense neighbourhoods such as city centres often displays a low sprawling factor. Industrial rings display the contrary. This factor often is not a constant for the whole city, but differs for different stages in its development.

In the simulation the assumption is made that the cities planning is organic, as this is in line with how most cities are growing. centres are thus added individually and are subsequently connected with the network. In a first step, the new centre is added to the network such that its position minimises the *potential*. As this potential is difficult to minimise we use a Monte-Carlo simulation. We randomly draw Q new possible centres in a radius v around the current network, where we choose Q to be $Q = 3N$. N represents the desired size of the network in number of centres present in the network. From those Q centres we choose to accept the one that minimises the overall potential function $\mathcal{P}$ (see Section 3.1). The radius v of the city can be adjusted during the simulation and represents the maximum diameter in which a new centre can be drawn from. Once a new centre is chosen it is connected to the rest of the network. These connections are determined with the organisation and construction parameters of a city (see Section 3.2. In the following two sections we describe how the best location for a new centre is decided with the help of a potential function and how the new centre is connected to the network.

3.1 Potentials

The allocation of new centres in a city has been studied by Courtat (2012) who has developed an approach using a *potential*. This potential determines the probability of the acceptance of a new centres to the city. Potential fields are commonly used in physics to describe the magnitude of magnetic fields for instance. They generally represent an induced region around objects. In our approach this translates to a region induced around the existing network, where new centres are rejected. The larger the street in the network the larger its field. In this way, a long street induces a larger field than a small street. Which seems logical as longer streets are usually also larger due to more traffic lanes for example. The field is defined in such a way that the preferred point is at its minimum. As the function is difficult to minimise we apply a Monte-Carlo simulation to determine the best new centre.

Courtat (2012) defines the following function inspired by the Lennard-Jones (Lennard-Jones, 1924) potential:

$$P(r) = \alpha \times \left(\left(\frac{\beta}{r}\right)^{2\eta} - \left(\frac{\beta}{r}\right)^{\eta} \right) \tag{1}$$

The function P calculates the value of the potential induced by a point of the network given its distance r to the new centre. To understand the parameters better we introduce three measurements. The rejection radius $r_0 = \beta$ around which a new centre is forbidden, the optimal distance $r^* = \beta \times 2^{\frac{1}{\eta}}$ of a new centre with respect to the existing network and the minimal value of the potential function $P^* = -\alpha \times \left(\frac{2^{\eta}-1}{4^{\eta}}\right)$. The rejection radius r_0 is constant and represents houses and other structural elements around streets, it is dependant on the parameter β. The optimal radius r^* depends on the type of city structure that is wished for, i.e. a city with a larger optimal radius allows for larger spaces between the streets. This can be because of larger houses or building complexes or simply more green space between the structures. We call the parameter η that influences the size of the optimal radius the *sprawling parameter*. The optimal value of the potential P^* determines the priority the new centre has in being placed inside the optimal radius. Therefore a larger value favours centres that are the closest to this optimal position. This measurement is dependent on the parameters η and α.

In the left panel of Figure 1 the potential field of the point $x = (0, 0)$ is represented in form of a heat map. The rejection radius r_0 is the yellow circle with radius 2. The optimal radius for new centres r^*, is the circle with the darkest points around x with radius $2\sqrt{2}$. Finally the minimal potential value is determined by $P^* = -\alpha \times \left(\frac{2^\eta - 1}{4^\eta}\right)$ and takes on value 5. The function P is defined for the values $[\beta, \infty)$ as for values inside the rejection radius r_0 the allocation of a new centre is forbidden. For other values of r, P is set to $+\infty$. We use the definition of r_0, r^* and P^* to determine the values of the parameters α, β and η in the simulation.

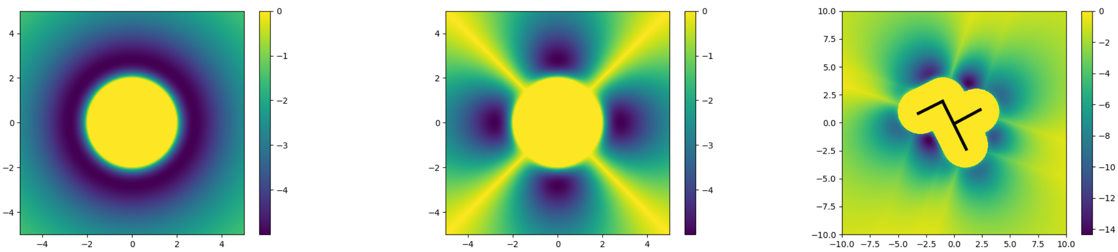


Figure 1: Potential field of a single point without (left) and with (middle) the angular effect. Potential field of a small network (right). The darkest values represent the optimal settling points for new centres. The following parameters are used $\alpha = 20,\ \beta = 2,\ \eta = 2$.

To introduce an extra dimension around the streets, the function P is enhanced with an angular effect. This means, that the optimal position of a new centre is not only dependant on its distance to the network, but also on the angle it makes with the existing structures. The consequence of this added dimension is that settlements, which make an angle of $\pi/4$ with the current network, are less likely to be selected compared to the ones aligning with the existing network.

For each generated new centre x we calculate the unitary potential with angular effect $\mathcal{P}_h(x)$ for all edges (street segments) h in the existing network (see Equation (2)). This unitary potential consists of two parts. The first one is the potential function $P(d(x, h^*))$ that can be found in Equation (1). The radius $d(x, h^*)$ represents the Euclidean distance from x to h^*, where h^* is the orthogonal projection of x on the edge h. If the orthogonal projection is not on edge h, h^* corresponds to the closest existing vertex to x. The second part of this equation is the correction for the angular effect explained in the paragraph above, where $\angle(x, h)$ represents the angle the edge (x, h^*) makes with edge h.

$$\mathcal{P}_h(x) = P(d(x, h^*)) \times |\cos(2|\angle(x, h)|)| \quad (2)$$

The overall potential $\boldsymbol{\mathcal{P}}(x)$ of a new centre vis-à-vis the whole network is the sum of all unitary potentials $\boldsymbol{\mathcal{P}}(x) = \sum_{h \in E} \mathcal{P}_h(x)$. In the middle and right panel of Figure 1 the potential of a single point and a whole network are shown respectively.

For all the potential new centres drawn in the Monte-Carlo simulation the new centre that is accepted to the network is the centre x with the lowest potential value $\boldsymbol{\mathcal{P}}(x)$. In the next section we elaborate on how this centre is connected to the existing network.

3.2 Construction and Organisation

Once a new centre x is accepted into the network it needs to be connected to the network. In real cities this translates to the construction of a street or street segment connecting this centre to the network. In our simulation a new centre takes the form of a vertex in the graph, a connection is an edge. We first define the visible set of neighbours VN. These are vertices in the graph that can be reached by x without crossing another edge on the way. To this set we add add the set of (visible) orthogonal projections of x on the edges already present in the network. Linking x to all the vertices in the set VN might allow for connections that are too close to each other or that are crossing each other, and that are thus not realistic city features. Therefore we define the relative neighbourhood RN, as the feasible subset of all the connections in VN. To create a set of feasible links we use Delaunay's triangulation. In a graph setting this triangulation of the space works as follows. The vertices p_i, p_j and p_k are three different vertices in the set of vertices P. The triangle $p_i p_j p_k$ is drawn if and only if there are no other vertices of P inside the circumcircle of $p_i p_j p_k$. This translates to the acceptance of the edge between two vertices p_i and p_j if and only if there is no vertex in the graph closer to either p_i or p_j. In our case we calculate the Delaunay triangulation for the vertices in VN. Only the vertices that are directly connected to x in this triangulation are part of the set RN.

The n possible connections in RN are then ordered $(s_1, ..., s_n)$, such that s_1 represents the shortest connection from x to any vertex in RN and s_n the furthest. The amount n' of accepted connections follows a Binomial distribution with $n' \sim \mathcal{B}(\omega, n-1) + 1$. By adding one we make sure that the shortest connection s_1 is always accepted. Connections $s_2, ..., s_{n'}$ are then also added. The *organisation parameter* $\omega \in [0, 1]$ defines how many connections are allowed. The more this parameter is close to zero the more tree like the city appears favouring dead end streets that are linked to the network. On the other hand the closer it is to one the more connected the network becomes which implies a higher organisation factor. To illustrate the difference we refer to Figure 2, where three simulations are depicted with varying organisation factors.

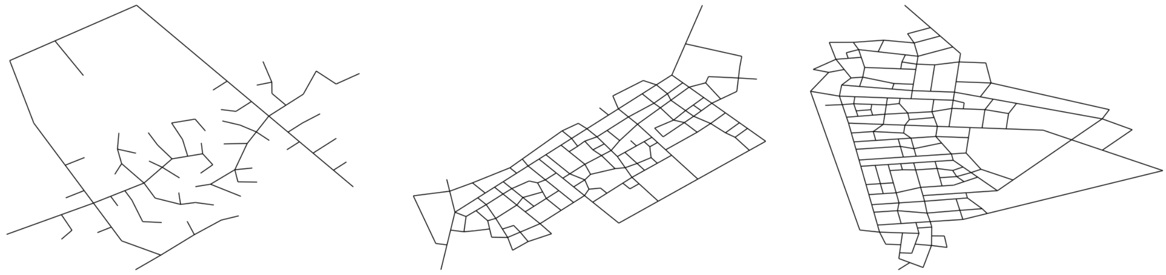


Figure 2: Three cities with varying organisation. From left to right the organisation factor ω is 0, 0.5 and 1. All other parameters are kept constant across the three simulations.

3.3 Sprawling

We now look at one of the last main characteristic of cities that we have not discussed yet, which is the city's sprawling. We distinguish three measures with respect to sprawling: the *sprawling* (η), the *sprawling factor* (ζ) and the *sprawling frequency* (ζ_f). The first one has been introduced in Section 3.1 and plays a role in the potential function. The lower η the larger the potential field and thus the more spacious the city becomes. The second and third one are used to model different growth patterns. A city with a low sprawling factor continues to develop in the city centre which means that new centres appear inside the city or near its borders. On the contrary a city with high sprawling allows for the development of centres outside the city centre as well. During the simulation with a probability of ζ_f the rejection radius is expanded by the factor ζ. This changes the potential field and allows for outside development. In real cities this is observed in the form of industrial or suburban neighbourhoods located on the outskirt of the city, which over time get gentrified and join the city's centre. To illustrate this in Figures 3 & 4 we depict several simulations where the parameters ζ and η are individually varied over several simulations. Usually during the growth of a city the three introduced measurements might vary across time. A larger city might not sprawl as much in the end as in the beginning of its growth. Nevertheless to keep the amount of simulations tractable we did not vary the parameters during the growth of a city.

Figure 3: Three cities with varying sprawling factor ζ. From left to right the factor ζ is 1, 5 and 10. All other parameters are kept constant across the three simulations.

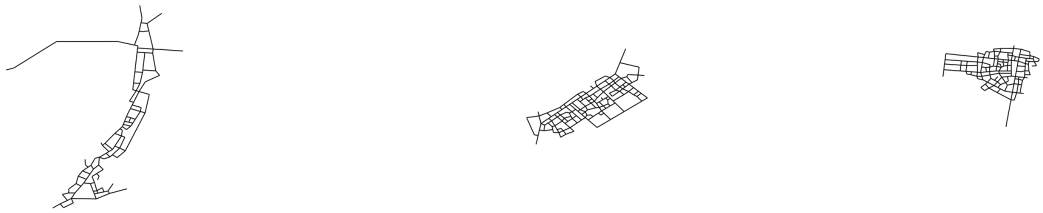


Figure 4: Three cities with varying sprawling parameter η. From left to right the parameter η is 0.3, 0.6 and 0.9. All other parameters are kept constant across the three simulations.

3.4 Refinements

To enhance the realism of the simulation we also consider the following refinements or construction elements of cities. To forbid extremely long and unrealistic streets the parameter l_{max} represents the maximum street length. Furthermore as the chance of a new settlement being exactly on an existing street is quite small any new point or orthogonal projection that is $c \cdot r_0$ away from an existing segment in the network is incorporated into the existing street or connected to the nearest intersection instead of adding a very small connection to it. A realistic value for c is 0.3.

Finally some cities are built around structuring natural elements such as a river. To simulate this we start the simulation with an existing network (i.e. four or five points that form a geometrical structure). This brings some randomness in the positioning of the street network and contributes to the fact that each city has a unique form and unique natural constraints. Next to that Courtat (2012) has found that by introducing structural elements the shape of cities differs from a circle to a more realistic city like polygon.

3.5 City Growth Simulation

In this section we discuss the parameters variation and other choices made for the morphogenesis simulation of cities. The process is described in details in the previous sections. We summarised the different parameters with a short description in Table 1. To keep the amount of simulations feasible we decide to only vary certain parameters. Only the parameters marked with an asterisk (*) are varied over the different simulations. We fix the values for the resting parameters based on small simulation runs and values found to be accurate by Courtat (2012). The expansion radius should always be set to the radius where the entirety of the potential fields exceeds a certain value to allow for a more accurate Monte-Carlo simulation. In our case we set the radius to 400m beyond the furthest vertex and adjust it in every simulation.

Parameter	Description	Value
N^*	amount of centres	[10, 20,..., 210, 220]
v	maximum expansion radius	varies
β	parameter in the rejection radius r_0	10
α	parameter in the optimal potential value V^*	5
ω^*	organisation factor	[0, 0.25, 0.5, 0.75, 1]
η^*	sprawling parameter used in the optimal settlement radius r^*	[0.3, 0.6, 0.9, 1.2]
ζ^*	sprawling factor	[1, 5, 10]
ζ_f	sprawling frequency	0.1
l_{max}	maximum street segment length	400
c	street separation factor	0.3

Table 1: Parameter Settings for the Simulation

3.6 Hierarchy

Street hierarchy is a term used in urban planning terminology to describe different types of streets. Until now we did not distinguish a simple one way street from a four lane boulevard. In this section two methods are presented to introduce this extra dimension of street hierarchy into our simulation.

3.6.1 Closeness

The *closeness* measurement is a centrality measure on street networks that evaluates the topological distance of a street towards the rest of the network. The higher the closeness measure the more central this street is. This means that it can be reached from all other streets in the network by using the least amount of streets. Streets that are more accessible are also more used by the population to move from A to B and therefore those streets are usually larger and in a better condition allowing for more flow or/and a higher travelling speed. We should take note at this point that until now we have worked with street segments, that are represented as a single edge in the graph structure. For the closeness measurement we are interested in the relationship between entire streets, where streets are usually formed by more than one street segment. In cities, usually several street segments with the same name form a street. We therefore in a first time explain how we combine street segments to form streets. We then elaborate how we use the ensuing street network to determine the closeness of each street segment.

In a first time we create street elements on the network, by combining several street segments, which we from now on refer to as *ways*. To create those we use an algorithm developed by Lagesse (2015) that can be used to combine street segments into ways in our simulated networks. This is done by combining adjacent street segments following a set of rules. For each vertex in the graph, which represents an intersection, a set L_{adj} of all adjacent edges or street segments is made. For each pair of edges in the set L_{adj} the intersection angle is calculated. The pair of edges that forms the largest angle is combined to form a new way or is added to an existing way. The pair is then deleted of L_{adj} and the intersection angles of the pairs left in L_{adj} are calculated again. If the largest angle of all remaining pairs of edges is lower than a threshold angle of θ_{min} the edges are separately added to the list of ways as they form the end of a way or a way consisting of a single street segment. To illustrate in Figure 5 a network is depicted where each way is attributed another colour.

Once the ways are calculated we make a new graph, representing the street network. In this graph each vertex corresponds to a way and the edges connect ways that are linked by an intersection in the network. We use this simplified graph to calculate the closeness measure.

The closeness measure assesses how close a street is towards the rest of the network in a topological way. The topological distance can be seen as the easiest path between two points instead of the shortest one. A more extensive definition is given in the next chapter. The closeness measure of way i is defined as : $closeness_i = \frac{1}{\sum_{j \in N} topo_{ij}}$, where $topo_{ij}$ is the amount of ways on the path to way j when starting in way i. This means that the larger the closeness measurement for way i, the easier it is to reach way i from any point in the network making it an important structure. In the original graph each street segment or edge gets the closeness measure of the way it belongs to. In Figure 6 a network is depicted with the closeness measurement on its ways.

By making the assumption that streets, that play a more central role in the network are also more important in size and/or allow for a faster flow, we define a new weight measure for the

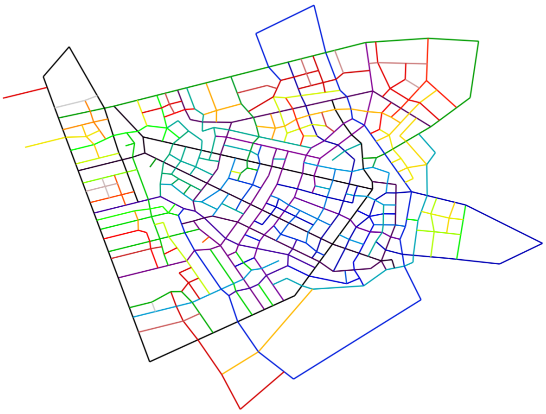


Figure 5: Network depicted with the way object. Street segments corresponding to the same way have the same colour.

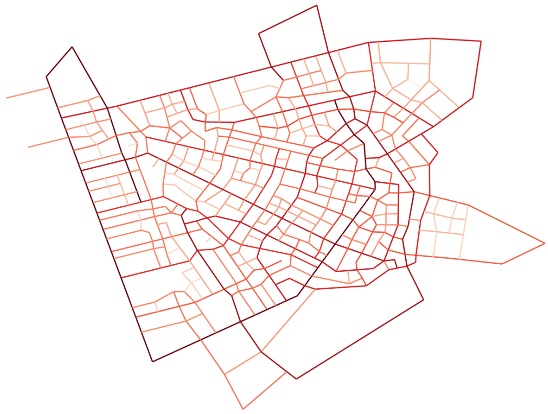


Figure 6: Closeness measurement of a network, the darker a way the more central it is in the network and, the higher its hierarchy.

street segments (arcs) in the network. Street segments are now not only characterised by their length, but also by their importance in the network. If the segment has a larger centrality measure we assume a faster link speed and therefore adjust its weight accordingly. Assuming that the maximum speed on a link is 50km/h and the slowest 10km/h, the link's weights, which were only characterised by length at first, are left equal for the most central streets and are multiplied by at most 5 in case they are the least central.

3.6.2 Self-Organisation Model

In this section we define a second model that can be used to introduce link hierarchy into a network. This model is based on the model introduced by Levinson and Yerra (2006). They find that hierarchies are not only the product of design choices but also happen with the free use of a network. We use this research to first create a demand model that creates flow in the network. Once demand is present in the network a cost, revenue and investment model determines which edges are used most.This is done in terms of speed. The higher the speed on an edge the more it is used and the more the model invests further in it. The model is iterated until an equilibrium is found or it is clear that there will be no equilibrium. An outline of the self-organisation model for street hierarchy can be found in Algorithm 1.

Algorithm 1: Self-Organisation Model

Initialisation:

1. Demand Model: Assign amount of generated and attracted trips to vertices in network
2. Set initial speeds on arcs
3. Initialize time $t = 0$

while *equilibrium is not found or* $t < 100$ **do**

Flow Model:

1. Determine cost in terms of time of commuting on edges given current speed (v_a^t)
2. Given these costs, determine flow for each pair of vertices in the network

Cost Model:

Determine the usage costs C_a^t based on the flow on arcs.

Revenue Model:

Determine the revenue E_a^t generated from the flow on arcs.

Investment Model:

Determine the new speeds v_a^{t+1} for each arc.

if *The new speeds differ less than 0.05 from the previous speeds* **then**

Equilibrium found

$t + 1$

Demand Model

Until now we have introduced a framework to build a street network without taking demand and flows into consideration. In this section we introduce flows into the model. We only consider three basic cases. The first one being an 'American' type of city, where the residential areas are situated in the outskirts of the city and the business in the centre of the city. In this case the demand generating vertices are situated on the outskirt whereas the demand attracting vertices are situated in the centre. Next, we consider a second case where we presume the opposite type of city, where we assume the residential areas to be mainly located inside the city and the industry more on the outskirts. In the last model we assume equal distribution of demand attraction and generation over the whole city.

In the two first above mentioned cases we have to discern the city centre from the outskirts. The best way to do this is by clustering the intersections into two groups: those belonging to the city centre and the ones belonging to the outskirt. We use a spectral clustering technique to cluster the vertices into two groups. This is done by solving a graph partitioning problem using eigenanalysis. In Figure 7 this partitioning of the vertices is depicted on two networks. We remark that the spectral clustering algorithm has more difficulty with the partitioning of the right city than with the left. As the partitioning of a city in neighbourhoods is a very subjective question, in the framework of synthetic cities we have to make the assumption that this method simulates the reality as close as possible. As the partitioning is used to create demand it is not crucial for it to be exact therefore we accept this error probability in the demand creation model.

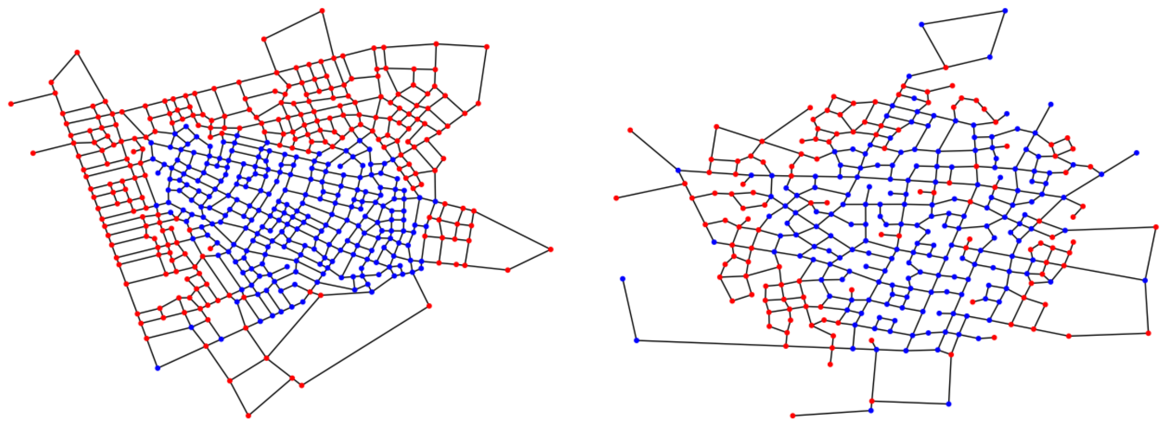


Figure 7: Two network partitioning cases. The blue intersections represent the city centre and the red ones the city outskirts. We remark that the partitioning is better in the left panel than in the right one.

Once the vertices are partitioned. Each vertex i is assigned an amount p_i of *generated trips* and an amount q_i of *attracted trips*. If vertex i is situated in a demand attracting area p_i is a uniformly distributed random variable between 0 and 20. If vertex i is situated in a demand generating area p_i is a uniformly distributed variable between 80 and 100. Similarly q_i is a random variable that is respectively uniformly distributed on (80-100) and (0-20). The maximum flow (100) is a trivially chosen. The exact amount of flow does not matter, the difference between demand attracting and demand producing vertices is what is important such that the flow can be simulated correctly. We choose for a 1:5 ratio.

Flow Model

In this section we calculate the *flow* f_{rs}^t between any two vertices r and s in the graph. To do this we first need to define the cost of commuting between two vertices. In this model the cost of commuting between two points at time t is assumed to consist of two parts: travel time converted into monetary value and toll. We define d_a^t the *cost of commuting* on edge a at time t:

$$d_a^t = \frac{l_a}{v_a^t} + \rho_0 \cdot (l_a)^{\rho_1} \cdot (v_a^t)^{\rho_3}, \tag{3}$$

where l_a and v_a^t are respectively the *length* and *speed* of edge a at time t and ρ_0, ρ_1 and ρ_3 are coefficients explained in Table 2.

The more general travelling cost d_{rs}^t between two vertices r and s is the sum of all the edge travelling costs incurred by the edges on the shortest path between r and s. The shortest path is calculated using Dijkstra's algorithm and as weights the costs d_a^t. This algorithm is explained in more details in Chapter 4. The flow f_{rs}^t between r and s at time t is dependant on the travelling cost d_{rs}^t and the generated (p_r) and attracted (q_s) trips at vertices r and s respectively.

$$f_{rs}^t = p_r \frac{q_s \cdot h(d_{rs}^t)}{\sum_{s \in N} q_s \cdot h(d_{rs}^t)}, \tag{4}$$

where the function h is a negative exponential model with parameter w: $h(d) = e^{-w \cdot d}$.

Cost, Revenue and Investment Model

Based on the flow between different vertices the cost and revenue and the ensuing reinvestment can be determined for each edge in the graph.

The *revenue* E_a^t of each edge a at time t is calculated by multiplying the flow and the toll. Where the 'toll' is dependant on the *speed* at time t (v_a) and *length* (l_a) of edge a. This means that the flow and revenue are positively linked.

$$E_a^t = (\rho_0 \cdot (l_a)^{\rho_1} \cdot (v_a^t)^{\rho_3}) \cdot (\psi \cdot f_a^t), \tag{5}$$

where ψ is a parameter that scales the flow annually and ρ_0, ρ_0 and ρ_0 are the coefficients of Table 2.

The *cost* C_a^t of edge a at time t is the cost needed to keep the edge in the present condition given its flow, speed and length.

$$C_a^t = \mu \cdot (l_a)^{\alpha_1} \cdot (f_a^t)^{\alpha_2} \cdot (v_a^t)^{\alpha_3}, \tag{6}$$

where μ is the annual unit cost of maintenance for a arc. The coefficients α_1, α_2 and α_3 are coefficients from Table 2.

Finally based on the cost and revenue of each edge in the investment model, the speed of the edges is updated, as shown in Equation (7). This new speed is then used in the flow model again in iteration $t+1$. The whole process is iterated until an equilibrium is reached or it is clear that and equilibrium will not be reached.

$$v_a^{t+1} = v_a^t \left(\frac{E_a^t}{C_a^t} \right)^{\beta} \tag{7}$$

In Figure 8 we depict the hierarchy of the network for the three partitioning cases using the self-organisation model as explained above. The darker the street segment the more important it is and the higher its hierarchy in the city. In the left panel we observe the hierarchy of the streets if every vertex generates an equal amount of trips as it produces. In that case the hierarchy seems quite equally distributed among street segments. In general redundant connection streets seem to be less important than others. In the middle panel we depict the hierarchy of streets in the scenario where the demand is generated in the city centre and attracted in the city outskirts. Here we see that the street segments that connect the city centre with the outskirt have a higher hierarchy and that in general the streets in the centre have a higher hierarchy measure than the ones on the outskirt. In the right panel the the hierarchy of streets in the opposite scenario is depicted. We see that if the situation is inverted the opposite effect on the hierarchy is observed.

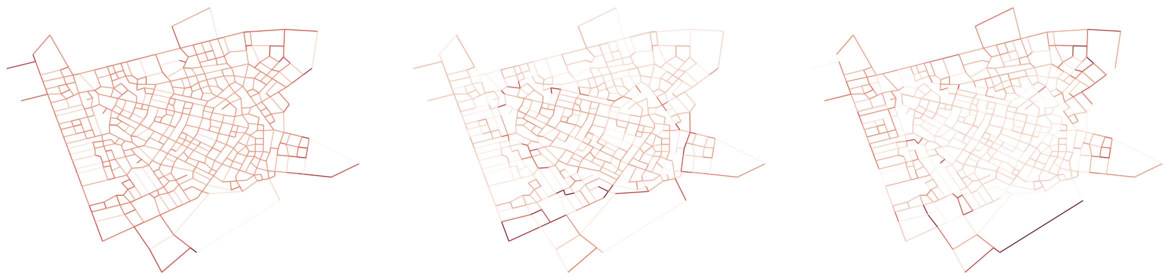


Figure 8: Networks with the hierarchy measurement in three scenarios. The darker the street segment the larger its hierarchy in the city. In the left panel the demand is equally distributed. In the middle panel the demand generating vertices are located in the city centre. In the left panel the opposite demand is used where demand generating vertices are located in the city outskirts instead.

Variable	Description	Value
v_a^0	Initial speed (integer)	1
w	Coefficient in flow model	0.01
ρ_0	Coefficient in revenue model	1.0
ρ_1	Length power in revenue model	1.0
ρ_3	Speed power in revenue model	0.2
τ	Tax rate in revenue model	1.0
ψ	Revenue model parameter	365
μ	Unit cost in cost model	365
α_1	Length power in cost model	1.0
α_2	Flow power in cost model	0.75
α_3	Speed power in cost model	0.75
β	Coefficient in investment model	1.0

Table 2: Variable descriptions and values used in the self-organisation model

4 Methodology

In this chapter we discuss the different methods that are used to evaluate the possible effect certain morphogenesis and morphological characteristics of cities have on the efficiency of their street network. First we introduce the Traveling Salesman Problem (TSP) as well as the different ways to measure distances between intersections in the network. We then introduce the three performance measures used in the efficiency analyses of the simulated networks: The Overall Efficiency Measure (OEM) using the TSP, the Average Efficiency Measure (AEM) using the shortest path and a vulnerability measure evaluating network vulnerabilities to perturbances. Finally we introduce a multiple linear regression model used to evaluate certain relationships statistically.

4.1 Travelling Salesman Problem

The Travelling Salesman Problem (TSP) is one of the most studied combinatorial optimisation problem where a salesman needs to make a tour visiting n cities only once and return to the first city while minimising traveling costs. We define $G = (V, E)$ as a graph where V is the set of vertices with cardinality n and E the set of edges with cardinality m. Furthermore we define a cost matrix C where c_{ij} represents the cost made to go from vertex i to j. Using integer programming the TSP can be formulated as follows:

$$\min \quad \sum_{i \in V} \sum_{j \in V} c_{ij} x_{ij} \tag{8}$$

$$\text{s.t.} \quad \sum_{j \in V} x_{ij} = 1 \qquad \forall i \in V \tag{9}$$

$$\sum_{i \in V} x_{ij} = 1 \qquad \forall j \in V \tag{10}$$

$$\sum_{i \in S} \sum_{j \in S} x_{ij} \leq |S| - 1 \qquad \forall S \subset V, S \neq \emptyset \tag{11}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i, j \in V \tag{12}$$

where the variable x_{ij} takes on value 1 if there is an edge between vertex i and j and 0 otherwise. In this integer programming the objective function (8) is to minimise the costs of the tour. The first (9) and second (10) constraint ensure that there is at most one arrival and one departure from each city. This ensures a single visit to each city. The last constraint (11) makes sure that there are no sub-tours and that the tour spans all cities in one single tour.

To solve this problem to optimality Applegate et al. (2006) propose a solver code `concorde`. This solver uses branch and bound and problem specific branch and cut techniques (Ozden et al., 2017) and is currently the exact method of choice to solve this type of problems.

4.2 Cost Matrix

The cost matrix C needed to solve the TSP as described from (8) to (12) is in a first time defined in terms of distances between all vertices in the graph G. At a later point we introduce hierarchy into the simulated cities and multiply certain distances with factors to introduce link speed. There are several possible ways to define the distance between two vertices. The most straight forward distance measure is the Euclidean distance, which represents the distance as the crow flies between to intersections, this distance measure can be used as a benchmark as it is the shortest distance between two intersections. The shortest distance between two intersections using the existing network is the *shortest path* between two vertices. We refer to the problem of finding all pairs of shortest paths in the graph as the All Pairs Shortest Path Problem (APSPP). A final (more realistic) measurement for the distance between two intersection is what Courtat (2012) defines as the *topological distance*. This is the 'simplest' path between two intersection in terms of times the path needs to take turns to reach the destination. This concept is illustrated in Figure 9.

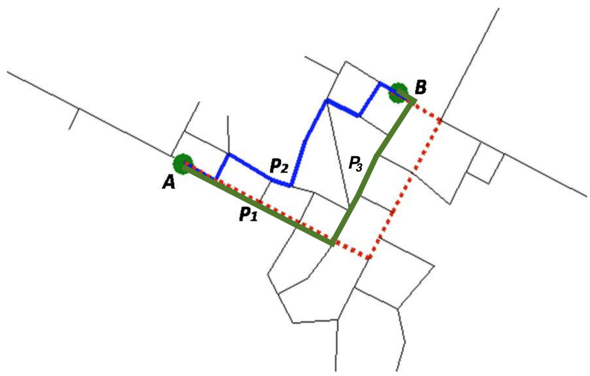


Figure 9: The points A and B represent two points in a graph. The blue line (P_2) represents the shortest path from A to B, the red dotted line (P_1) and the green line (P_3) represents a simplest path and the shortest simplest path. Modified picture from Courtat (2012).

All Pairs Shortest Path Problem

Several algorithms can be used to solve the APSPP. In our case we want to solve this problem on a sparse graph with positive weights (as street lengths are always positive), which means that the repeated application of Dijkstra's shortest path algorithm is the best known solution to solve this type of problem.

The density of a graph $G = (V, E)$ is defined as $D = \frac{2|E|}{|V|(|V|-1)}$. If D is close to one the graph G is called dense. The contrary leads to a sparse graph. In sparse graphs the repeated application of Dijkstra's algorithm on all pair of edges is the quickest way of solving the APSP problem as the time complexity of this algorithm is dependant on the amount of edges $O(n^2\log(n) + mn)$.

To illustrate we take another solution method for the APSP: the Floyd-Warshall algorithm that has a complexity of $O(n^3)$. This algorithm is faster for all graphs where,

$$m > n^2 - n\log(n). \tag{13}$$

Because in our case we do not allow for edges to cross we have a planar graph. Using Euler's formula for planar graphs we know the number of edges m in a graph is bounded by the following equation: $m < 3n - 6$. As the inequality in (13) never holds we can conclude that the planar graphs that are generated are sparse and that Dijkstra's algorithm performs better than Floyd-Warshall at all times.

Algorithm 2: Dijkstra's Algorithm

Input: Graph $G = (V, E)$, origin vertex u, destination vertex v
Initialise:
- Q a set of unprocessed vertices
- N_i the set of all neighbour vertices of $i \in V$
- d_i shortest known distance from u to $i \in V$
- $parent_i$ the parent vertex of $i \in V$
- $length(i, j)$ edge length between neighbouring vertices i and j

```
for i ∈ V:
    d_i ← ∞
    parent_i ← undefined
    add i to Q
  set d_u ← 0
while Q is non empty
  select i with min d_i in Q
  remove i from Q
  for j ∈ N_i
    d_new ← d_i + length(i, j)
    if d_new < d_j
      d_j = d_new
      parent_j = i
```

Return: d_i, $parent_i$ $\forall i \in V$

In Algorithm 2 the pseudo code for Dijkstra's algorithm is given. This algorithm calculates the shortest path between and origin vertex u and a destination vertex v. In this algorithm we start by processing the origin vertex u. The distance to this vertex is marked as 0 and the distance d to all other vertices in the graph is set to positive infinity. In each iteration the next closest unprocessed vertex is chosen as the current vertex , starting with u in the first iteration. First the distance d of all adjacent vertices to this unprocessed vertex is calculated. The distance is defined as the distance d to the current vertex plus the distance from this vertex to the adjacent one. If this value is less than the previously recorded distance, the distance d is updated. Once the destination vertex v has been visited the shortest path can be retrieved by backtracking each vertex' parent starting at vertex v. Where the parent vertex is the last vertex that updated the distance before the vertex has been processed.

Topological Distance

The most plausible choice for humans to move from A to B is probably the simplest path. Courtat (2012) defines this simplest path as the topological distance between two intersections. This corresponds to the path between to points that minimises the amount of turns first and the distance second. The length of the shortest path with the least amount of turns is the path defined as the topological distance between two points. This distance measure is computationally expensive compared to the shortest path as all paths need to be generated and evaluated on the amount of turns and the path length.

4.3 Performance Measures

In this section, we introduce three performance measures to evaluate the efficiency of the generated cities in the simulation. The first two evaluate the efficiency the last one the vulnerability of a network. The first efficiency measure uses the optimal value of the TSP. The optimal value represents the cost of a tour visiting all intersections in the network once, where the costs incurred on each edge are defined in a cost matrix. We define a network to be more efficient if a cheaper tour is possible through the city network, as it means that it is easier to reach all the intersections. The second performance measure evaluates the efficiency of the network based on the inverse of the shortest distance between intersections in the network. By taking the average of all the shortest distances this measure differentiates itself by providing an average efficiency measurement of the network. The last measure that we use is the vulnerability of the network to disruption. Vulnerability is measured by the effect of discontinuities in the network on the efficiency measure. Those disruptions might be due to a wide range of events ranging from daily traffic accidents to natural disasters incapacitating parts of the network.

4.3.1 Overall Efficiency Measure

We assess the overall efficiency of the network by using the optimal value of the TSP. This measure looks at the accessibility of the whole network. To be able to compare it across growing and differently sized networks, we are constrained to scale the optimal value with a benchmark. Therefore we solve the TSP for each city two times with different cost matrices as input. The first cost matrix uses the Euclidean distance as a distance measure and the second one the shortest path. The optimal value using the Euclidean distance (O_e^*) is the benchmark value. The Overall Efficiency Measure OEM is the deviation from the optimal value, using the shortest path (O_s^*), towards the benchmark value.

$$OEM = \frac{O_e^* - O_s^*}{O_s^*} \tag{14}$$

4.3.2 Average Efficiency Measure

We now introduce the second efficiency measure. Instead of looking at the accessibility of the whole network, this measure looks at the average shortest distance between two intersection. According to Yin and Xu (2010) the network efficiency is inversely related to the shortest distance between vertices. We introduce the Average Efficiency Measure

$$AEM = \frac{1}{N(N-1)} \sum_{i \neq j} e_{ij}, \tag{15}$$

where e_{ij} is defined as the inverse of the shortest distance $e_{ij} = \frac{1}{d_{ij}}$. When no path exists between vertices i and j, e_{ij} is set to 0. Due to its inverse relationship, the larger the AEM, the less efficient a city is.

4.3.3 Vulnerability

Yin and Xu (2010) propose a structural vulnerability model to assess the vulnerability of networks. The efficiency measure AEM is the base of the model for the structural vulnerability of a network. The vulnerability of the street network towards a given intersection (or vertex) n is calculated as follows:

$$V_n = \frac{AEM - AEM_n}{AEM}, \tag{16}$$

where AEM_n is the efficiency of the network as described in Equation (15), when vertex n is removed from the network. This measurements captures the dependence of the network to this one intersection. Since $AEM \geq AEM_n$ we know that $0 \leq V_n \leq 1$. The average vulnerability of the whole network is calculated as follows:

$$V = \frac{1}{N} \sum_{n \in N} V_n. \tag{17}$$

4.4 Multiple Linear Regression

Assuming the easiest relationship between the different parameters and the performance measures we run a linear regression model to find the average effects the parameter variations have on the dependent variables. The linear model is given by the following equation:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k + \epsilon, \tag{18}$$

where we define y as the dependant variable. The parameter vector β is estimated in this regression such that the sum of squared errors are minimised. The constant parameter is β_1. The vector of explanatory variables is defined as x, x_1 is defined to take on value 1. The unobserved residuals are taken into account in ϵ. We use this model to see if and what kind of combined effect the parameters have on the performance measures. For instance to analyse the joint effect the four parameters η, ζ, N and ω have on OEM. We run the regression where $y = OEM$, and the vector x corresponds to $[1, \eta, \zeta, N, \omega]$. The vector of coefficients that minimizes the sum of squares β tells us the average effect each coefficient has on the OEM. With the help of a t-test we can determine the significance of each coefficient. In the linear regression model several assumptions are made to be able to guarantee that β is the best linear unbiased estimator of the dependant variable. First of all as mentioned before we assume a linear relationship between the dependent variable and the vector of variables x. Secondly we assume that x is fixed and the parameters in the vector β are constant. Next the random disturbances ϵ should be randomly distributed with mean zero and and equal variance. Finally all pairs of disturbances should be uncorrelated.

5 Results

In this chapter we evaluate which different morphogenesis and morphology factors have an effect on the efficiency of a network. To come to our conclusions we simulated the growth of different cities by varying their growth and structure parameters. In a first time we have simulated 60 cities. For each city its growth is saved every 10 iterations with a maximum of 220 iterations. Each iteration corresponds to a time unit where a new center is added and connected to the network. We then enhanced the simulated street networks with the introduction of street hierarchy. Subsequently we calculate performance measures on the grown and growing cities to compare and evaluate their efficiency and vulnerability over time and between each other.

5.1 Hierarchy

In Section 3.6 two models are proposed to introduce hierarchy into the model. The more elaborate model is the self-organisation model. As described in 3.6.2 this model has to iterate several times and is thus quite expensive in computation times. Trials on small instances show that it is not feasible (in the scope of this thesis) to calculate it for all stages of the growing cities, we therefore choose to add hierarchy using this model on the 60 largest instances of cities. The other model makes use of the closeness measurement and is described in Section 3.6.1. This measurement is easily calculated even on large instances and therefore provides a very cheap alternative to the more complex self-organisation model. To visualize the difference in Figure 10 the OEM is shown for the 60 largest cities for three types of street hierarchisation. In the first case no hierarchy is used in the second one hierarchy is calculated with the closeness measurement and in the third one with the self-organisation model.

It is difficult to analyse the correctness of both hierarchy measurements as hierarchy is not only dependent on geographical factors and usage but also on results of political and town planning actions. Even though those are partly incorporated in the growth model for the street network by Courtat (2012) there is not a clear trend in all cities. The two models aspire to approach reality a bit more but do not intend to mimic it perfectly.

In Figure 10 we can observe that in general the closeness and self-organisation model follow similar trends. Due to the randomness in the creation of demand in the self-organisation model the OEM using the self-organisation model is more volatile than the one using closeness but we see the same general trend. Even though the self-organisation model for the simulation of network hierarchy has its advantages, we for now assume that the closeness measurement as a

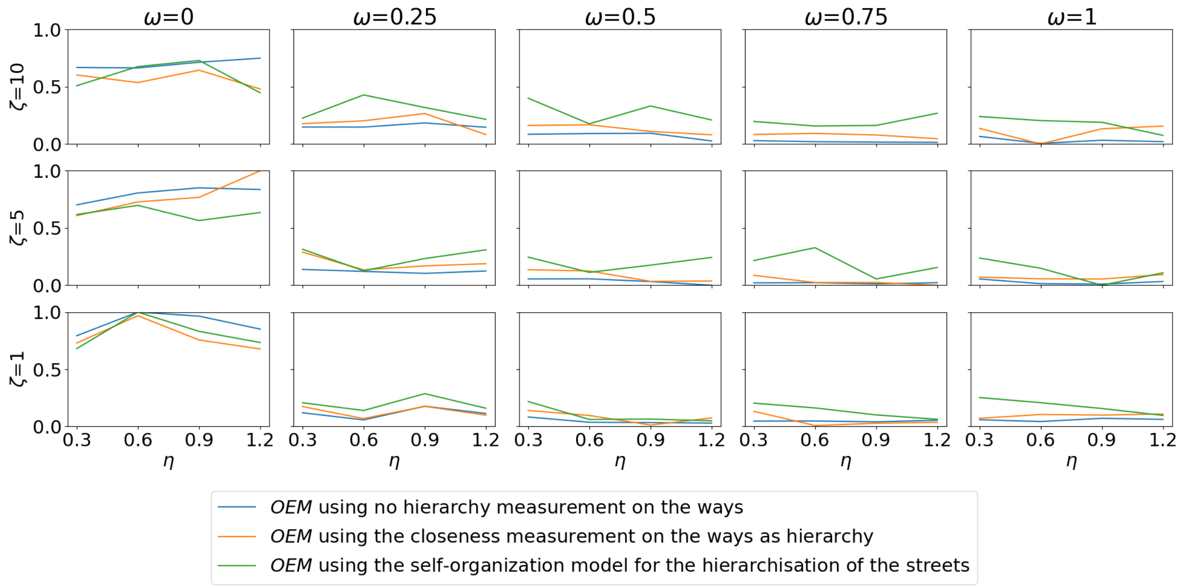
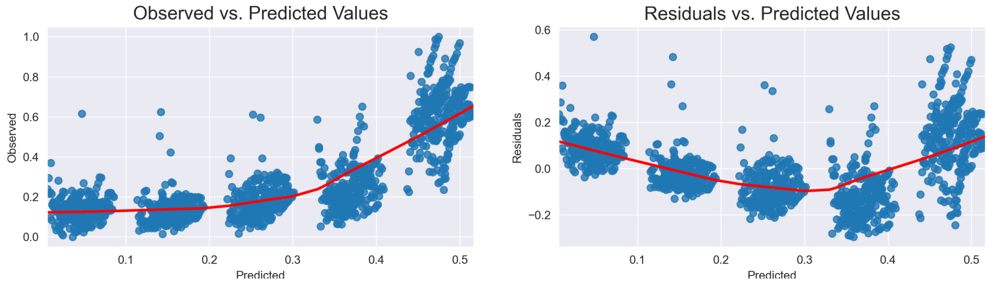


Figure 10: Overall Efficiency Measure for cities with different values of the sprawling parameter η, sprawling factor ζ and organisation factor ω for three different hierarchy measurements

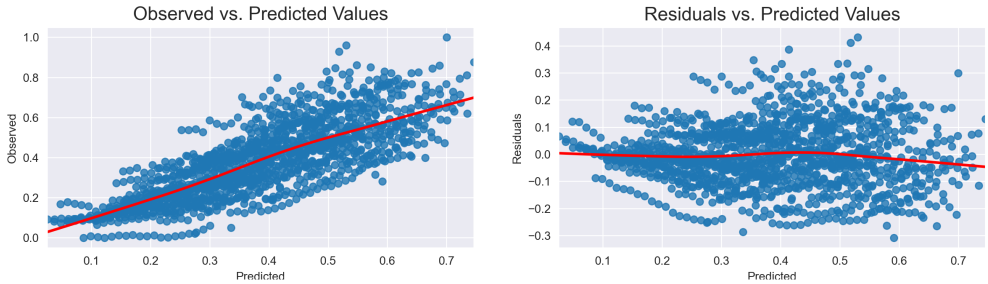
basis for network hierarchy is a computational cheap alternative, that yet gives promising results. In the remainder of this thesis we use the closeness measurement as the hierarchy measure of the street network in terms of link speeds as explained in Section 3.6.1.

5.2 Efficiency Analysis

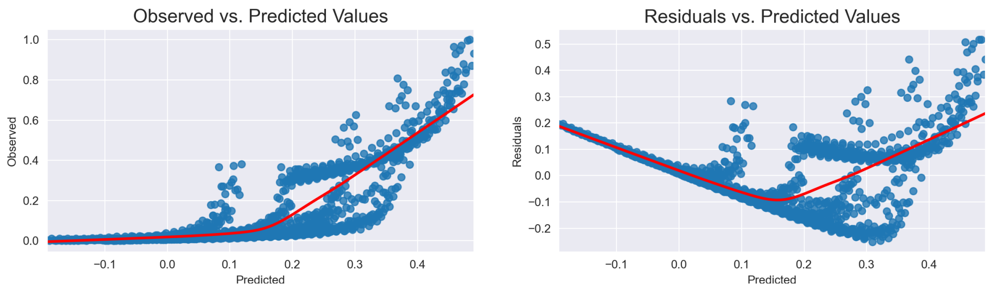
In this section we evaluate which morphological and morphogenic parameters influence the performance measures of the simulated street networks. As can be consulted with more detail in Table 1 the organisation factor ω, the sprawling factor ζ, the sprawling parameter η and the size (N) of the simulated cities are varied across each simulation run. This results in the simulation of 60 cities whose growth is measured every 10th iteration starting at N=20 until N=220. In total we calculate the three performance measures for 1260 simulated networks. Figures 13, 14 and 15 depict the three performance measures on each simulated network in the form of a heat map. For each simulated city the color represents the value of the performance measure, where the blue color represents a more favorable outcome than red (i.e the network is more efficient or less vulnerable). In Table 3 the results of three multiple linear regression with robust standard errors are depicted. Each regression is used to explain average effects and combined significance that the parameters have on the performance measures. We first discuss the assumptions made in the multiple linear regression and then go on with a full analysis of our results.



(a) This figure shows the observed values for the OEM against the predicted ones in the left panel and the observed residuals of the regression against the predicted ones in the right panel. In both cases the desired outcome would be that the observations lie on a straight line, this would confirm that OEM can be modeled by a linear relationship. We can visually assess that this is far from being the case.



(b) This figure shows the observed values for the AEM against the predicted ones in the left panel and the observed residuals of the regression against the predicted ones in the right panel. In both cases the desired outcome would be that the observations lie on a straight line, this would confirm that AEM can be modeled by a linear relationship. Even though the line through the data seems straight the data points themselves are not gathered around the desired line. This also indicates that the relationship is probably not linear.



(c) This figure shows the observed values for the VM against the predicted ones in the left panel and the observed residuals of the regression against the predicted ones in the right panel. In both cases the desired outcome would be that the observations lie on a straight line, this would confirm that VM can be modeled by a linear relationship. We can visually assess that this is far from being the case.

Figure 11: Six plots to visually assess the linearity of the relationship between the three depend variables and the data.

Regression

The estimators of the three linear regressions are only the best linear unbiased estimators if the assumptions made in Section 4.4 hold. We will now therefore first analyse the regressions before we can go on with the discussion of the results. In a first time we assumed that the relationship between the data and the dependent variable is linear. In Figure 11 we can see that this might not be the case. As we only are interested in the trend of the data and the joint significance of parameters we will for now continue to use this simplified model for our data. The assumption of a fixed vector of variables and a constant vector of estimators is met. Next, we look at the unobserved residuals (ϵ in (18)). The assumption that their mean is equal to zero is met for the three regressions. The Durbin-Watson test that detects autocorrelation is less than two for all three regressions and thus suggests that the residuals are positively correlated. Furthermore, in Figure 12 we can see that the residuals are probably heteroskedastic. The absence of a constant variance of the residuals should be shown by a straight line. The two statistical tests for detecting homoskedasticity Breusch-Pagan (p=0.000) and Goldfeld-Quandt (p=0.000) are also both rejected in all three regressions. Serial correlation and heteroskedasticity do not however influence the unbiasedness of the estimators. In the case of both heteroskedasticity and serial correlation the estimator just underestimates the standard errors of the coefficients and the t-stat tends to exaggerate the significance of these coefficients. To solve this we introduce Heteroskedastic and Autocorrelation Consistent (HAC) standard errors. For the right interpretation of our results we included HAC standard errors into the three regressions in Table 3 and thus made our standard errors robust. A lag of $T^{\frac{1}{4}}$, where T is the sample size, is chosen in accordance with Newey and West (1986). Those results will be used to explain certain trends we find in the data as explained in more details next.

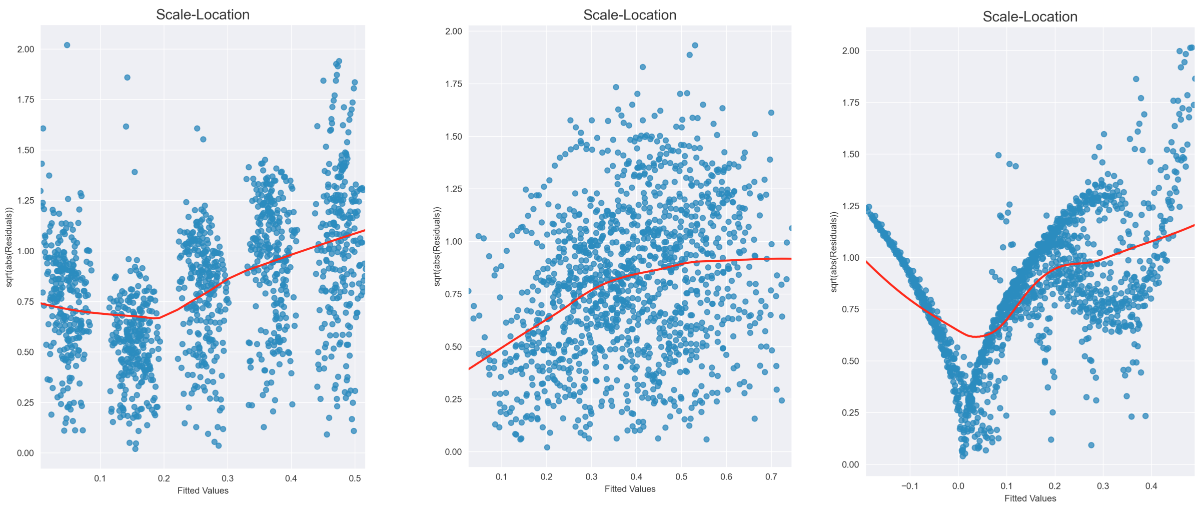


Figure 12: Figure showing the standardised residuals against the fitted values for the three dependent values OEM in the left panel, AEM in the middle panel and VM in the right panel. The residuals should have a constant variance and thus a straight line should go through the data. We see that in the three panels this is not the case suggesting heteroskedasticity.

Result Analysis

In a first time we look at the Overall Efficiency Measure OEM. This performance measure takes the deviation of the optimal value of the TSP using the shortest distance cost matrix towards the benchmark value using the Euclidean distance cost matrix. By taking the deviation we can compare results for different sized cities as well as different development stages of a same city.

A few observations stand out. Fist of all there is a big difference in the OEM between the cities with organisation factor $\omega = 0$ and $\omega > 0$. As soon as cities leave the 'tree' like form that is achieved when $\omega = 0$ their efficiency as measured by the OEM increases considerably. For cities that have $\omega > 0.25$ there does not seem to be a significant difference anymore. Together with the other parameters the sprawling parameter η and factor ζ do not have a joint significant effect on the OEM. If we look at inner city growth the positive trend does not seem to be statistically significant. Visually we can only confirm this effect for cities with a low value for $\omega = 0$. In Figure 13 a representation of the OEM coefficient is given for the 60 simulated cities at 21 stages during their development. All the results of the regression can be found in the first panel of Table 3.

Next we look at the Average Efficiency Measure AEM as introduced in Section 4.3.2. This performance measure uses the inverse of the shortest distance as an efficiency measure for the whole network. We find that the sprawling parameter η has the largest effect on the AEM of a city. When η increases by 0.3 the AEM also increases by 10.5% on average. This means that the street network becomes more efficient when the spacing between new centres and the existing network increase. The inverse effect can be observed, with a significant lower impact, when looking at the sprawling factor ζ. When the factor increases by 5 the network becomes 6% less efficient on average. In other words this refers to the fact that the more a city chooses for outside development the less efficient it becomes. Lastly when looking at the growth of each city we observe a negative effect on the efficiency measure. For each 10 new centres ($N = 10$), the AEM decreases by 1.2%. This means that the larger the city becomes on average in terms of new centres, the less efficient it becomes according to the AEM. In Figure 14 a representation of the AEM coefficient is given for the 60 simulated cities at 21 stages during their development. All the results of the regression can be found in the second panel of Table 3.

We finally look at the effect the parameters have on the vulnerability of cities towards disruptions in the network. First of all we see that the effect the organisation factor ω has on the vulnerability of a city is quite large for cities that are not organised. Here similarly to the effect on the OEM the vulnerability significantly decreases when a city leaves the 'tree' like state. As soon as cities reaches a organisation of $\omega = 0.25$ the average vulnerability decreases considerably. Given that a city meets the requirement of having $\omega \geq 0.25$ it is not subject to a lot of vulnerability from its intersections anymore. The sprawling parameter η and the sprawling factor ζ do not have a very large effect nor significant combined effect on the vulnerability of a city. This means that whether the city prefers inner or outer development and bigger or smaller spaces between new settlements does not have a large impact on the vulnerability of its network. When looking at the inner growth of cities we again see a clear trend. For each 10 new centres ($N = 10$) added to the city the vulnerability decreases by 1% point. On average the larger the city becomes the less vulnerable it is to disruptions. In Figure 15 a representation of the vulnerability measurement VM is given for the 60 simulated cities at 21 stages during their development. All the results of the regression can be found in the third panel of Table 3.

	OEM			
	coefficient	st. error	t-stat	p-value
constant	0.4877	0.014	16.842	0.000
N	0.0002	$6.07 \cdot 10^{-5}$	1.648	0.099
η	-0.0442	0.011	-1.758	0.079
ω	-0.4311	0.011	-15.416	0.000
ζ	0.0004	0.001	0.199	0.843
	AEM			
	coefficient	st. error	t-stat	p-value
constant	0.3038	0.012	10.754	0.000
N	-0.0012	$5.02 \cdot 10^{-5}$	-13.957	0.000
η	0.3512	0.009	16.444	0.000
ω	0.0444	0.009	2.106	0.035
ζ	-0.0123	0.001	-6.572	0.000
	VM			
	coefficient	st. error	t-stat	p-value
constant	0.5135	0.013	17.746	0.000
N	-0.0013	$5.53 \cdot 10^{-5}$	-14.372	0.000
η	-0.0327	0.010	-1.523	0.128
ω	-0.3717	0.010	-16.393	0.000
ζ	-0.0007	0.001	-0.375	0.708

Table 3: Coefficients of three linear regressions on dependent variables OEM, AEM and VM. Standard Errors are HAC using 6 lags and without small sample correction. Significance is judged on a 95 percent confidence interval. P-values under 0.05 are considered statistically significant.

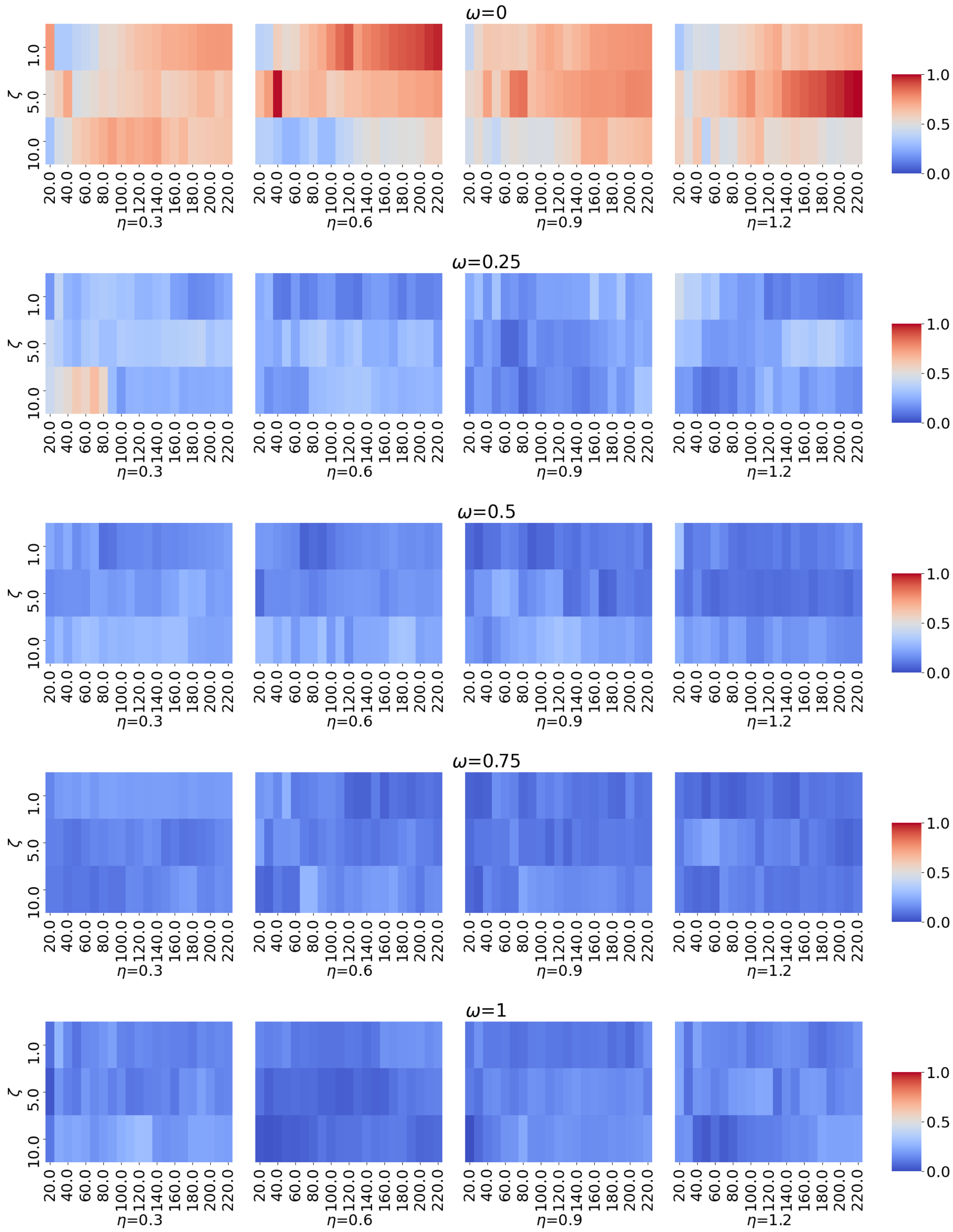


Figure 13: This figure depicts the Overall Efficiency Measure OEM for all the simulated networks. In this figure 20 heat maps are depicted. From top to bottom in each row the organisation factor ω increases by 0.25. From left to right for each column the sprawling parameter η increases by 0.3. Each heat map shows the OEM for three growing networks with constant ω and η. The three rows represent a new network with increasing sprawling factor ζ. From left to right each color strip represents the OEM for a network with N from 20 to 220. The OEM is given on a scale from zero to one. A network with an OEM tending to zero, is coloured blue. If the value tends towards one, it takes on the colour red.

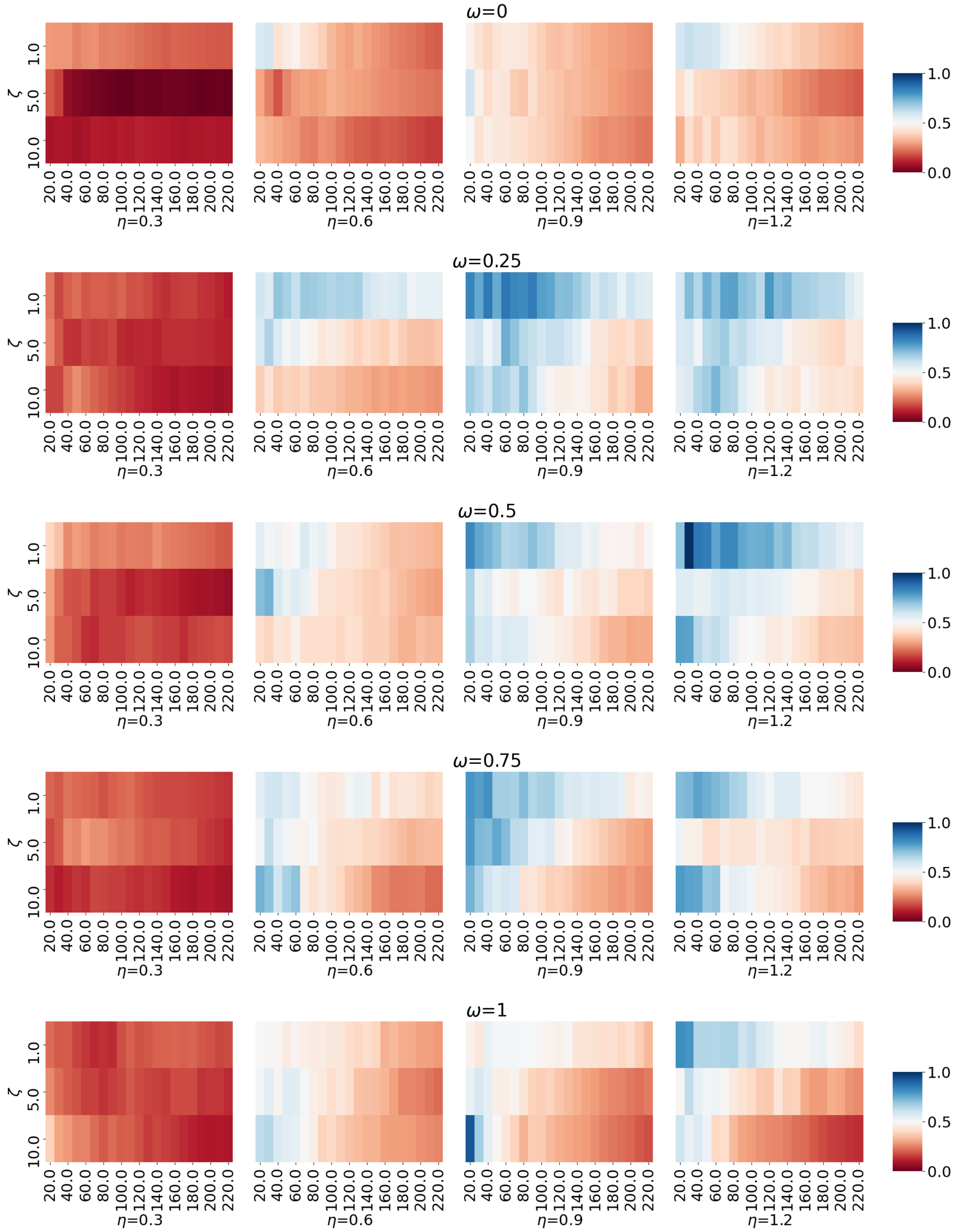


Figure 14: This figure depicts the Average Efficiency Measure AEM for all the simulated networks. In this figure 20 heat maps are depicted. From top to bottom in each row the organisation factor ω increases by 0.25. From left to right for each column the sprawling parameter η increases by 0.3. Each heat map shows the AEM for three growing networks with constant ω and η. The three rows represent a new network with increasing sprawling factor ζ. From left to right each color strip represents the AEM for a network with N from 20 to 220. The AEM is given on a scale from zero to one. A network with an AEM tending to zero, is coloured blue. If the value tends towards one, it takes on the colour red.

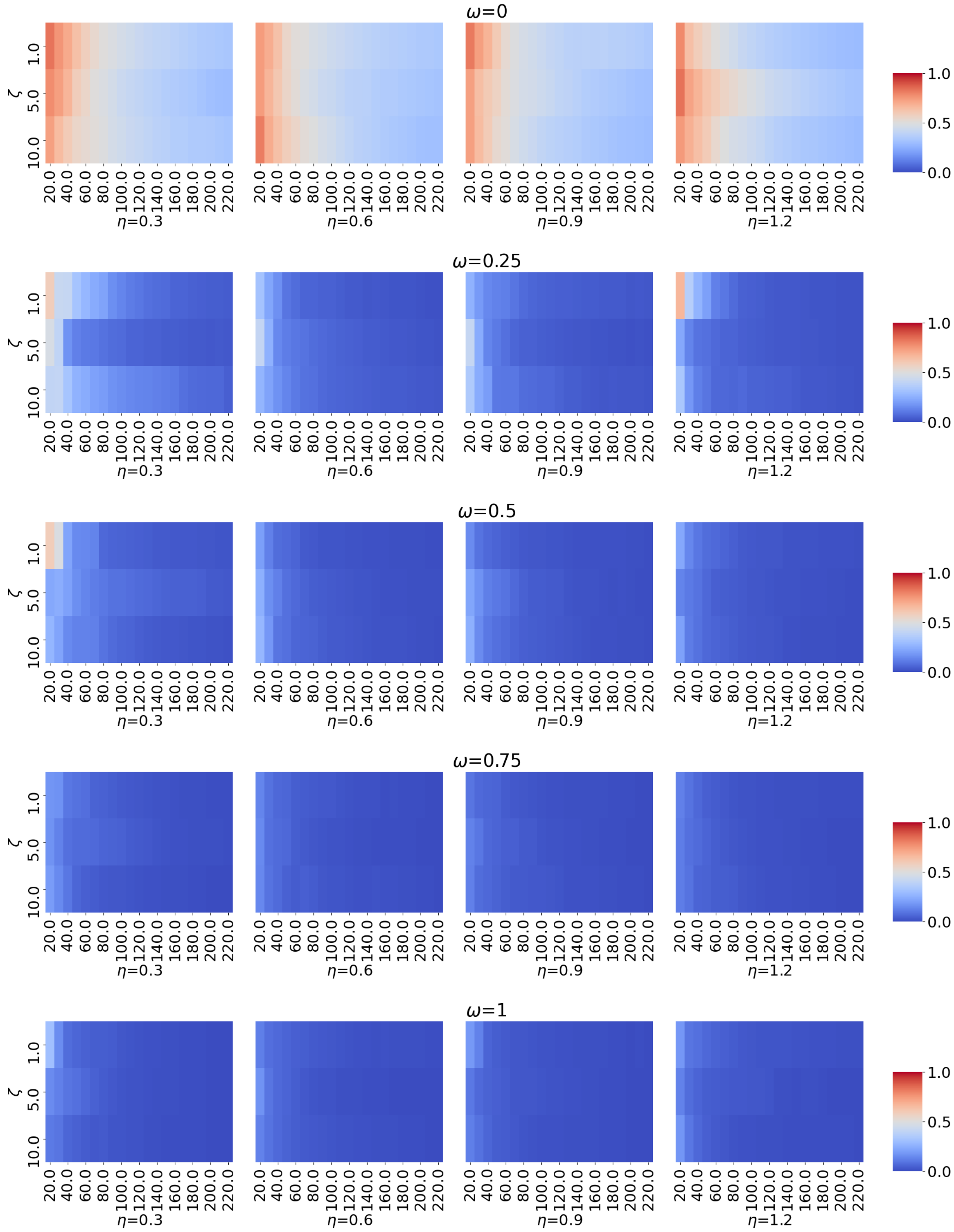


Figure 15: This figure depicts the vulnerability measure VM for all the simulated networks. In this figure 20 heat maps are depicted. From top to bottom in each row the organisation factor ω increases by 0.25. From left to right for each column the sprawling parameter η increases by 0.3. Each heat map shows the VM for three growing networks with constant ω and η. The three rows represent a new network with increasing sprawling factor ζ. From left to right each color strip represents the VM for a network with N from 20 to 220. The VM is given on a scale from zero to one. A network with an VM tending to zero, is coloured blue. If the value tends towards one, it takes on the colour red.

6 Conclusion

The growth of cities and urban networks has a big momentum in the literature as it brings with it a lot of challenges on a social, economic, political and logistical level. In this thesis we propose the simulation of growing city networks with different morphological and morphogenic parameters and the subsequent analysis of the parameters' effect on network efficiency and vulnerability.

There are different ways of representing a city mathematically. We choose for a representation of the street network using graph theory as it offers an intuitive way of representing this network while also offering the mathematical framework for further analyses. In a simulation study we simulate 60 different street networks that vary in parameters such as sprawling, organisation and size. Their size is recorded every 10-th iteration until iteration 220 is reached. These simulations follow a method introduced by Courtat (2012) and are based on a potential function that determines the optimal allocation of new centres. Those centres are then connected to the network by street segments. We also add hierarchy to our network with the help of a centrality measure. To implement this measure we cluster street segments to form ways and rank them on their centrality. More central streets have a higher hierarchy in the city, which is quantified in an increase of the crossing speed.

In a following step we analyse the efficiency and vulnerability of those simulated street networks. We propose three performance measures. The first two evaluate the efficiency of the network and the last one its vulnerability. The first measure uses the well known optimization problem of the Travelling Salesman Problem (TSP), where a path needs to be found that visits all the vertices in the network exactly once. The optimal value of this problem is used to evaluate the overall efficiency of the network. To be able to compare the values across different networks we take the deviation of the TSP using the shortest path towards the TSP using Euclidean distances between intersection. The second measure evaluate the average efficiency of the network using the inverse of the shortest path between all intersections in the network. The last performance measure calculates the vulnerability of the street network to disruptions of the intersections.

When looking at the growth of cities we find that in general the larger a city becomes, the less efficient it is. Where we have to distinguish that the average efficiency decreases more than the overall efficiency. On the other hand we find that the vulnerability of networks decrease when they are growing. Furthermore when looking at the effect the different parameters have we also see clear trends.

The organisation of a city only has a large impact on the overall efficiency and the vulnerability if it is close to 0. That means, that if the network of a city is tree like its overall efficiency decreases and its vulnerability to disruptions increases. As long as a minimum of organisation is present (i.e. it is larger than 0.25) we find that it does not have a large impact on the efficiency and vulnerability anymore.

The sprawling parameter, which refers to the spacing between new centers and the existing network, has a considerably low impact on the overall and the average efficiency. In general we can say that if the spaces between new centres added to the network get larger, the efficiency of a city slightly improves. The increase in the sprawling parameter does not seem to influence the vulnerability of a city significantly.

The sprawling factor that reflects the choice between inner development and outer expansion does not influence the overall efficiency and vulnerability significantly. A slight increase in the average efficiency is observed when more outside development is chosen over the inner development.

In conclusion we can say that in general a trade-off needs to be made when cities are growing between efficiency and vulnerability. A more vulnerable city can be subject to more congestion for example but a less efficient city is not attractive for trade for example. Furthermore as long as the city is organised to a minimum of $\omega = 0.25$ the building of connections does not seem to add much to the efficiency and vulnerability of the network. Finally we find that cities should be generous with the space between new centres and the existing network. Next to the fact that extra space can be used to have more plants or pedestrian streets, we find that it has in general a positive effect on the efficiency and vulnerability of a city.

7 Discussion and Further Research

In this thesis we analysed the effects morphological and morphogenic parameters have on the efficiency and vulnerability of street networks in cities. In this chapter we discuss the main limitations as well as suggestions for possible further research.

First of all we mostly based the simulation of the different street networks on the morphogenic and morphological parameter ranges introduced by Courtat (2012). These are mostly based on French towns that exist since the middle ages. A more extensive analysis of existing street networks could give more insight in the morphological parameters of cities and which geographical factors have influence on them. For example which parameters reflect on the location of a city (e.g. mountainous region, urban region, different state, region or country). As an example distances between and orientation of streets in mountainous regions are different than cities that expand in valleys or plains. Open source street maps from the whole world can be gathered from `openstreetmaps.org`.

A more extensive analysis on morphogenic parameters is more difficult to achieve, as gathering data from historic maps is a very cumbersome process. Once the data from ancient maps is retrieved it can be used to adjust the simulation process to reflect the historical process of city expansion more accurately. An accurate simulation model can then be extrapolated to other cities and save this time consuming process. This thesis can be used to extrapolate results found using historic maps and to use this knowledge in the simulation of the growth of other cities.

Due to time constraints we restrained the simulations to 60 different kinds of cities by varying three parameters. The variation of more parameters can give some extra insight into which characteristics influences the efficiency of the networks. In this thesis we also decided to keep the parameters fixed for the whole simulation of cities. It would be interesting to relax this assumption by simultaneously varying them. For example during the growth the sprawling factor can be set to 1 in the beginning and later be expanded to 10, thus simulating a city that in a first time focuses on inner development and only years later decides to expand outside its borders. The same can be done by varying the organisation factor and thus building different kinds of neighbourhoods in a city. Those adjustments can greatly improve the 'realness' of the simulations. To be able to statistically analyse those relationships a lot of simulation runs need to be done.

Because of the small scale of the simulation done in this thesis the statistical analysis is kept at a basic level. We have shown that a linear regression can be used to look at basic trends and joint significance. Nevertheless the relationship does not seem to be linear and thus this might not be the right model for predictions. We thus encourage further analysis on non-linear models to better predict behaviors. By creating more cities a more in depth statistical analysis can be performed such that more relationships can be inferred from the data. Next to that we encourage to look into more linear regressions. We now only ran three linear regressions, but relationships between variables when one parameter is fixed can also be studied apart.

Another extension of the proposed simulation is to extend the size of the created cities. This extension can be used to analyse the extreme growing of certain cities witnessed in the past decades and how it affects the network's efficiency. We warn that this extension might be time intensive as the algorithm is sensitive to the amount of vertices and edges in the network.

Finally we like to point out that we analyse the efficiency of a very simplified network. Our network does not for example take congestion, traffic lights, pedestrian street and one way streets into account. For the introduction of flows extensive research on traffic models in real cities should be made to be able to model it correctly into simulated cities. This research can be used to for example improve the self-organisation model proposed in this thesis.

References

Applegate, D. L., Bixby, R. E., Chvatal, V., and Cook, W. J. (2006). *The traveling salesman problem: a computational study*. Princeton university press.

Barthélemy, M. and Flammini, A. (2008). Modeling urban street patterns. *Physical Review Letters*, 100(13).

Chan, T. M. (2010). More algorithms for all-pairs shortest paths in weighted graphs. *SIAM Journal on Computing*, 39(5):2075–2089.

Courtat, T. (2012). *Walk on City Maps - Mathematical and Physical phenomenology of the City, a Geometrical approach*. Theses, Université Paris-Diderot - Paris VII.

Courtat, T., Gloaguen, C., and Douady, S. (2011). Mathematics and morphogenesis of cities: A geometrical approach. *Physical Review E*, 83(3).

Dingil, A. E., Schweizer, J., Rupi, F., and Stasiskiene, Z. (2018). Transport indicator analysis and comparison of 151 urban areas, based on open source data. *European Transport Research Review*, 10(2):58.

Frankhauser, P. (1998). The fractal approach: A new tool for the spatial analysis of urban agglomerations. *Population: An English Selection*, 10(1):205–240.

Hagget, P. and Chorely, R. J. (1969). *Network analysis in geography*. Arnold, London.

Heij, C., Heij, C., de Boer, P., Franses, P. H., Kloek, T., van Dijk, H. K., et al. (2004). *Econometric methods with applications in business and economics*. Oxford University Press.

Heinke, G. W. (1997). The challenge of urban growth and sustainable development for asian cities in the 21st century. *Environmental monitoring and assessment*, 44(1-3):155–171.

Karp, R. M. (1972). Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer.

Lagesse, C. (2015). *Lire les Lignes de la Ville*. Theses, Universite Paris Diderot-Paris VII.

Lennard-Jones, J. E. (1924). On the determination of molecular fields. *Proc. R. Soc. Lond. A*, 106(738):463–477.

Levinson, D. and Yerra, B. (2006). Self-organization of surface transportation networks. *Transportation science*, 40(2):179–188.

Newey, W. K. and West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. Technical report, National Bureau of Economic Research.

Ozden, S., Smith, A., and Gue, K. (2017). Solving large batches of traveling salesman problems with parallel and distributed computing. *Computers Operations Research*, 85:87 – 96.

Pousse, R. (2020). *Caractérisation et modélisation du réseau viaire.* Theses, Université de Paris - Laboratoire Matière et Systèmes Complexes.

Publications, U. N. (2019). *World Urbanization Prospects: The 2018 Revision.* UN.

Semboloni, F., Assfalg, J., Armeni, S., Gianassi, R., and Marsoni, F. (2004). Citydev, an interactive multi-agents urban model on the web. *Computers, Environment and Urban Systems*, 28(1):45 – 64. Geosimulation.

Williams, R. (2014). Faster all-pairs shortest paths via circuit complexity. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 664–673.

Yin, H. Y. and Xu, L. Q. (2010). Measuring the structural vulnerability of road network: A network efficiency perspective. *Journal of Shanghai Jiaotong University (Science)*, 15(6):736–742.