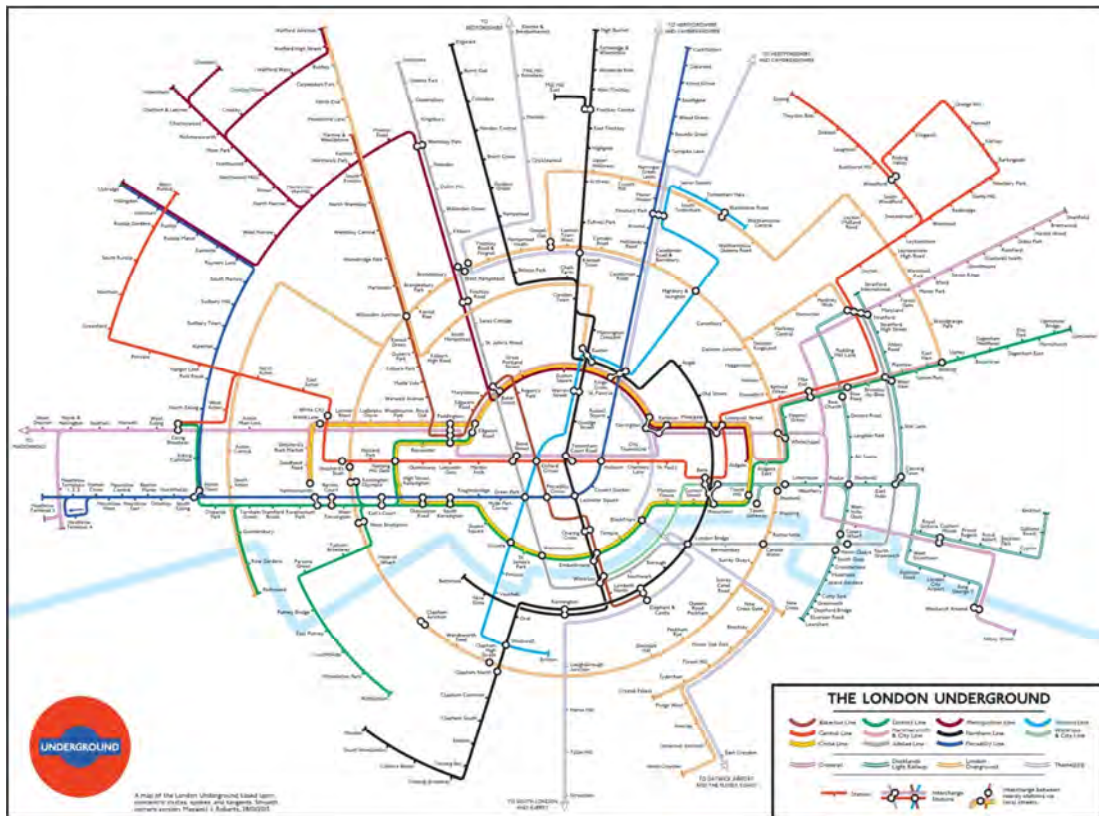


Design of robust metro networks



Research Paper Business Analytics

J.S. de Graauw

May 11, 2014

Supervisor

Prof. dr. R.D. van der Mei



VU Amsterdam

Faculty of Science

Business Analytics

De Boelelaan 1081a

1081 HV Amsterdam

Preface

As a student of the master Business Analytics at VU University Amsterdam, I needed to write a research paper. In the search for a suitable subject for my research paper, I came into contact with robustness of networks, a subject which became more and more important for the development and improvement of networks. At VU University the course 'Performance analysis of communication networks' is taught by Rob van der Mei and although I didn't attend this course, I asked him to guide me on this research paper based on his experience of network analysis. Based on his input, the existing literature and my interest, this paper has been established. I would like to thank Rob for his input and support.

Abstract

Metro networks are used by millions of people each day, all expecting a well performing network under all possible conditions. The ability of a network to continue performing well when it is subject to failures or attacks is called *robustness* (Ellens, & Kooij, 2011). Research to improve the robustness of a transport network often focuses on adding, moving or deleting certain links (Holme, Kim, Yoon, & Han, 2002). Because metro networks are a collection of transportation lines, those methodologies cannot be applied easily. In addition, these networks are linking different areas in a particular city and not specific locations. Due to the lack of information on the design of robust metro networks and the increasing degree to what new metro networks are built, the main question considered in this paper is:

How to design a metro network that maximizes robustness?

A method is given which can find the metro network with the highest robustness by three steps: Firstly the modeling part of translating reality to an applicable model. Second is the calculation of robustness. Finally the algorithm of finding the best additional metro line under certain constraints. The method gives a simple and applicable way to model areas and measures robustness differences even on graphs which are almost similar. The method is persistently capable of finding always the metro network with the highest robustness under a specific cost function.

Although this method is perfectly useful for finding a highly robust network, the complexity is increasing exponentially in relation to the number of nodes. A search for a heuristic to reduce complexity can be found in the betweenness heuristic or by methods to maximize travel time or the robustness measure. Although this heuristics are gives an approximate good design, the assumptions are very strict and worsen the applicability of the model.

Because of the detailed way of modeling reality and the possibility of the robustness measure to distinguish almost equal graphs, the method can be used best in the original method with manual offering metro networks, which can be tested by the robustness model. Offering a large number of possible models will give the model with the highest robustness inside the group of offering networks, with certainty. This method is been used in a case study based on the metro network of Amsterdam.

Index

Contents

- Preface..... 2
- Abstract..... 3
- Index 4
- Chapter 1 - Introduction..... 6
 - 1.1 Urban rail transit (metro) 7
 - 1.2 Definitions of robustness..... 8
 - 1.3 Network notation and terms 8
 - 1.4 Outline 9
- Chapter 2 - Model of a metro network10
 - 2.1 Mathematical model of the urban area10
 - 2.2 Mathematical model of the metro network.....13
 - 2.3 Paths of travelers15
 - 2.4 Continuation of this model.....16
- Chapter 3 - Measures to compare networks17
 - 3.1 Robustness measures17
 - 3.1.1 Literature review.....17
 - 3.1.2 Effective resistance18
 - 3.1.3 Network robustness index (NRI)21
 - 3.2 Travel time measure21
 - 3.2.1 Shortest path in graph **F**22
 - 3.2.2 Shortest path in graph **N**.....22
 - 3.3 Cost function.....24
- Chapter 4 - One metro line network on a greenfield26
 - 4.1 Approach and methodology26
 - 4.2 Algorithm of finding all metro lines26
 - 4.3 The 4-nodes area.....27
 - 4.3.1 The empty 4-nodes area.....28
 - 4.3.2 Placing 1 metro line in a 4-nodes area by maximizing the robustness29
 - 4.3.3 Placing 1 metro line in a 4-nodes area by maximizing robustness and a boundary at costs.....31
 - 4.4 The 9-nodes area.....32
 - 4.3.1 The empty 9-nodes area.....32

4.3.3 Placing 1 metro line in a 4-nodes area by maximizing robustness and a boundary at costs.....	33
Chapter 5 - Limitations and approaches of finding a robust metro line.....	37
5.1 Estimation of the number of edges and nodes	37
5.1.1 Dependence of costs and robustness.....	37
5.1.2 Simplify costs of one line and access the number of edges and nodes.....	38
5.2 Approach by edge betweenness to find a reasonably good design	39
5.2.1 Approach based on edge betweenness.....	39
5.2.2 Error measurement.....	40
5.2.3 Limitations	40
5.2.4. Adding weights of importance to the model	41
5.3 Approach by node betweenness to find a reasonably good design	41
5.2.1 Approach based on the node betweenness.....	41
5.2.2 Error measurement.....	42
5.2.3 Limitations	42
5.3.4. Adding weights of importance to the model	42
5.4 Approach by travel times to find a reasonably good design.....	43
5.5 Approach by maximizing effective resistance.....	43
5.5 Summary of models	43
Chapter 6 - Method in practice	44
6.1 Casus	44
6.1.1. Area of Amsterdam.....	44
6.1.2 The current metro system.....	45
6.1.3 Additional metro line	46
6.2 Approach of finding an additional metro line.....	46
6.3 Metro line with the highest robustness	46
Chapter 7 - Conclusion.....	48
Bibliography	49

Chapter 1

Introduction

Urban rail transit (metro) networks have become an important factor of the accessibility of a city. Metro networks are used by millions of people each day, all expecting a well performing network under all possible conditions. The ability of a network to continue performing well when it is subject to failures or attacks is called *robustness* (Ellens, & Kooij, 2011). In recent years, many new metro networks have been developed and existing networks have been expanded. Through this expansion, metro networks are increasingly important transport connections in a city. A disturbance in a metro network can therefore lead to large delays when travelers are unable to reach their destination. Therefore, there is an increasing demand for knowledge about robust metro network structures. Research to improve the robustness of a transport network often focuses on adding, moving or deleting certain links (Holme, Kim, Yoon, & Han, 2002). Because metro networks are a collection of transportation lines, those methodologies cannot be applied easily. In addition, these networks link different areas in a particular city and not specific locations. However, there is little information about applying robustness indicators to metro networks. Due to the lack of information on the design of robust metro networks and the increasing extent to what new metro networks are built, the main question considered in this paper is:

- ***How to design a metro network that maximizes robustness?***

To answer this main question, this research question is separated into five sub-questions:

- How to model a metro network to perform calculations on the representations of the networks?
- Which metrics are suitable to quantify robustness, costs and travel time of a metro network?
- How to design a robust metro network on a greenfield?
- What is the most effective investment to improve robustness of a metro network?
- How to design a fast yet accurate heuristic for improving robustness of metro networks?

This chapter will continue to discuss the theoretical background needed to understand the used notations and concepts. The first subsection will elaborate on the properties of an urban rail transit network. The second subsection will discuss the term *robustness* in more detail, followed with a section where more insight in the used notation and relevant terms are given.

1.1 Urban rail transit (metro)

In this paper, we focus on urban train transit systems, also known as metro systems. We define urban transit systems as metro systems with exclusive right-of-a-way, regardless of being underground, at grade or elevated. In those systems people move independently through the network using transportation vehicles which move on the rail tracks. Those networks have some specific properties:

- Developments of new network links are high investments and time consuming projects.
- Transportation vehicles are used to move on the trails.
- The transportation vehicles are not capable to depart the trails in case of any circumstance.
- All transportation is subject to a line schedule and doesn't move randomly through the transportation network.
- Users of the network move independently through the network and are capable to transfer between lines in a particular network.

In Figure 1 an example of a metro line in Mexico-City is shown. The underground network in Mexico City consists of 11 lines. Line 5 is the yellow and runs from Politecnico to Pantitlan with 12 stops. Figure 1 shows a small part of this line, from La Raza to Aragon. The stations La Raza and Consulado are transfer stations. Here it is possible here to switch to another subway line. The metro vehicles on this line can be seen in Figure 2. This part of the metro network can be seen as an example for many metro networks worldwide.



Fig 1. Example of metro line (Mexico-City)



Fig 2. Metro (Mexico-City)

1.2 Definitions of robustness

All networks are subject to disturbances. An important concept of (re)designing networks is to design networks so as to improve the ability of a network to absorb disturbances by offering alternative or high capacity routes. The intention of this design criterion is to continue performing well in case of disturbances, also named as the robustness of a network.

Robustness is the extent to which, under pre-specified circumstances, a network is able to maintain the function for which it was originally designed (Snelder, van Zuylen, & Immers, 2012)

Robustness is the ability of network to continue performing well when it is subject to failures or attacks (Ellens, & Kooij, 2011)

Network designers will pursue to design their networks such that they are highly robust, but highly robust networks are extremely limited by the investment to make. To realize a certain improvement in robustness and the urge to fall back on other parts of the network or to increase the capacity on existing parts (if possible).

1.3 Network notation and terms

The metro networks are described as graph G in context of graph theory. Graph G denoted as $G = (V, E)$ with V the vertices that are connected by the edges E . The number of vertices $|V| = n$, and the number of edges $|E| = m$. In some parts of this paper the names nodes (vertices) and links (edges) are used. This ambiguity is caused by the different names used in network theory (informatics) and graph theory (mathematics).

Path - A *path* in the graph is a sequence of edges which connect a sequence of nodes. In this paper only finite paths with two terminal nodes are captured: the *start node* s and the *end node*.

Shortest path – Path from the start node s to the end node e with the shortest distance.

Complete graph – A graph where all vertices are directly connected to each other.

Connected graph – A graph where all vertices are connected to each other by a path.

Unconnected graph – A graph for which at least one pair of vertices lacks a path.

Vertex degree - The degree of a graph vertex v of a graph G is the number of graph edges which touch v .

1.4 Outline

The remainder of this paper is organized according to the sub-questions listed in section 1.1. Each chapter refers to one of the sub-questions. Chapter 2 starts with the mathematical model of the metro network and area. Chapter 3 describes the measures used to determine robustness, costs and travel time. For those measures is the meaning, formula and the interpretation of the outcomes are discussed. Chapter 4 follows with the application of those measures for some predetermined areas. Chapter 5 describes the search of an accurate heuristic to decrease complexity. In Chapter 6 a case study in which the model is being used is developed.

Chapter 2

Model of a metro network

“How to model a metro network to perform calculations on the representations of the networks?”

2.1 Mathematical model of the urban area

Urban rail transit systems as the metro are commonly used in large cities or agglomerations. In this paper those urban areas are considered as a graph inspired on the layout of a grid. A grid can be applied to a particular area to divide the city in different smaller areas. The grid can then be translated into a graph G with the intersections of the grid as vertices and the lines as edges. An example of a grid spread out on a particular area is shown in Figure 3. The corresponding graph G is shown in Figure 4.



Fig. 3: Grid on plan (Bedum, Netherlands)

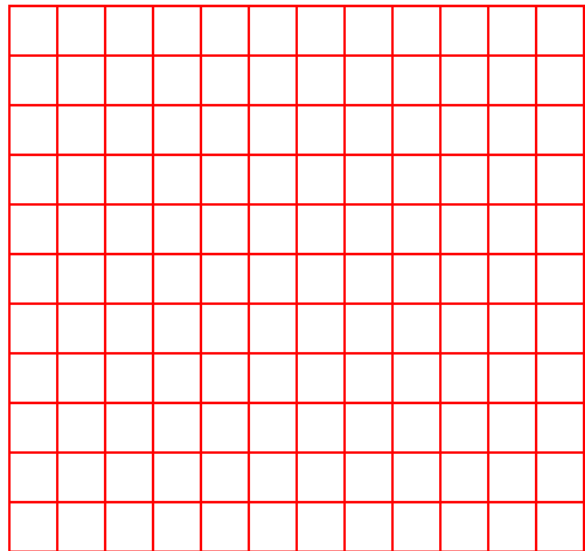


Fig. 4: Corresponding graph G of Fig. 3

In this paper graph G is denoted as $G = (V, E)$ with V the vertices and E the edges linking the vertices. The number of vertices $|V| = n$ and the number of edges $|E| = m$. In some parts of this paper the names nodes (vertices) and links (edges) are used. Figure 4 shows the grid-graph of the plan of Bedum. The grid is dividing the area of Bedum in 144 smaller areas. For this graph the number of vertices $n = 169$ and the number of edges $m = 288$.

The connected graph

Graph G includes all parts of a particular area and connects all possible vertices. This initial graph is the same for all equivalent areas and represents all possible links in an area. Those links can be added to metro lines which collectively represents the metro network. If a specific metro network is analyzed by the links added in the network, this leads to an unconnected graph if we disregard the other links. To avoid this situation, the possibility of

'walking from vertex to vertex' is added to the model, to ensure the connectivity of all vertices. To accomplish this situation the following assumptions and notations are made:

- The initial grid-graph G represents an extensive area by placing a grid over it.
- Subgraph $F \in G$ contains all vertices from graph G , which are possible to reach by foot and all links in between them which are walkable.
- Subgraph $M \in G$ contains all vertices which are possible to adopt in a metro network and are adopted in graph F , and the links in between them which are possible to adopt in a metro line.

Graph F denoted as $F = (V_F, E_F)$ with V_F the vertices and E_F the edges linking the vertices. The number of vertices $|V_F| = n_F$ and the number of edges $|E_F| = m_F$. The graph M is denoted as $M = (V_M, E_M)$, with $|V_M| = n_M$ and $|E_M| = m_M$.

With those assumptions all nodes in a particular area are reachable with or without a metro line connecting them. The nodes in G which are not adopted in F or M are not relevant to adopt in the model. The initial situation of each model is graph $\{F, M\}$. This assumption has also the advantage of a clear zero point to measure the travel times and robustness from: an area without an additional metro line, equal to graph F . The following example shows for the area of Bedum the graphs G, F and M .

The municipality of Bedum will better connect the city through a metro network. Suppose we want to analyze the area of Bedum further, the following graphs are determined:



Fig. 5: Urban area

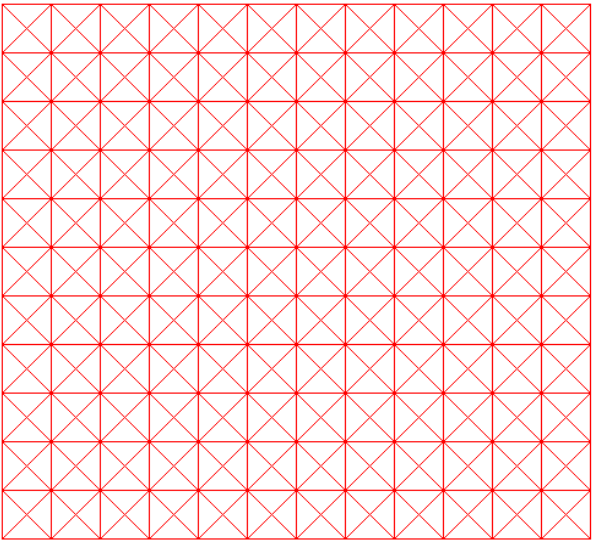


Fig. 6: Grid-graph G of area

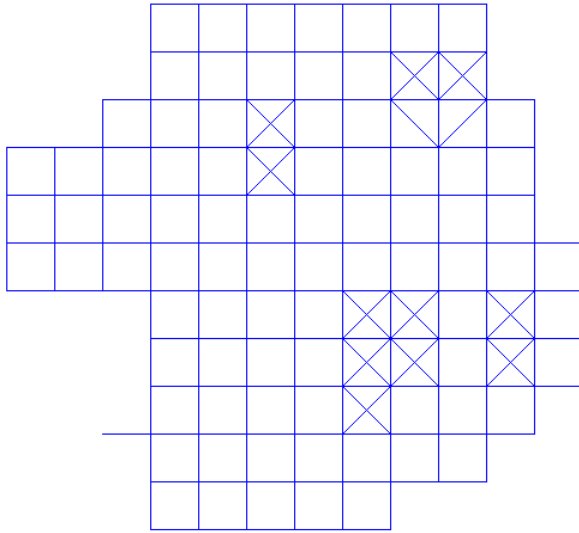


Fig. 7: Graph F of the urban area Bedum

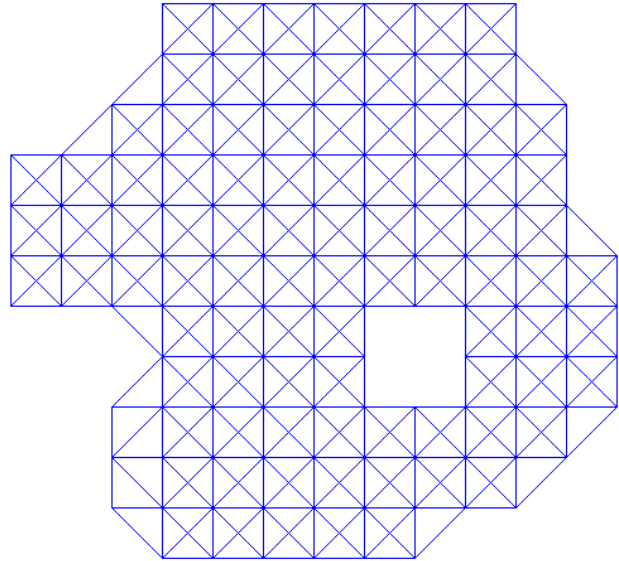


Fig. 8: Graph M of the urban area Bedum

Figures 6 till 8 show the graphs G , F and M for the area of Bedum. Graph F is equivalent to the streets in Bedum and graph M shows the possibilities in building a metro line in Bedum. Noteworthy is the empty area in graph M because of (for example) limitations in building a metro through this area. This example shows that by the properties of the graph F and M , the model for each environment can be specified.

Reachability of neighbors

The graph G represents a two-dimensional area, in which not every link between a pair of nodes is sensible. In a practical way of thinking: a link between two neighbor nodes can be made easily for low costs, but a link between two extreme nodes (far apart nodes) would be much more expensive and difficult. And in addition, many intermediate nodes will be beaten and the distance to be covered becomes larger. In order to reduce complexity the variable k is introduced. The variable k is the radius in which neighbors can be reached with a single edge, where the distance between two nodes is kept 1. The case $k < 1$ stands for a model where no node can be reached from another node because the scope is too narrow. $k = 1$ refers to a model with a scope of 1 and includes all links between horizontal and vertical neighbors (Figure 3). $k = \sqrt{2}$ refers to the model with all horizontal, vertical and diagonal neighbors (Figure 6). Also other values of $0 \leq k \leq \text{graph radius}$ are possible.

Besides the value of k the reachability of neighbors also depends on the links adopted in F and M . If links with a ratio of $k = \sqrt{2}$ are not adopted in F and M but valid through graph G , the reachability of the model is in fact smaller than $\sqrt{2}$, although k is set $\sqrt{2}$. Figures 2 and 6 show a grid-graph with $k = 1$ and $k = \sqrt{2}$ respectively.

2.2 Mathematical model of the metro network

The graph M shows all possible metro links which can be adopted in the model. A metro network N is a sub graph of M with $N = (S, P)$ with S the vertices referred to herein as stations and P the edges referred to herein as metro parts. The number of stations is defined as $|S| = s$, the number of metro parts as $|P| = p$. A metro network N consists of one or more metro lines. Every line is denoted as L_i , with $i = 1, \dots, l$ and $|L| = l$. Therefore, a metro network is always a collection of metro lines: $N = \{L_1, \dots, L_l\}$. Every line in the model has a different collection of nodes and edges; $L_i \neq L_j$ holds. N is a connected graph.

A metro network consists of p edges and s vertices. The set of vertices s can be divided into: (1) transfer stations s_t , (2) monotonic stations s_m and (3) end stations s_e . Transfer stations are stations where passengers have the possibility to transfer to another metro line, end stations are positioned at the end of a line, but if an end station is also a transfer station the name 'transfer station' is used. All other intermediate stations are monotonic stations:

$$s - (s_e + s_t) = s_m.$$

To get a better understanding of the modeling of the metro networks, it will be explained on the basis of an example.

The municipality of Bedum has developed a metro line plan. The specification of this plan is shown in Figures 9 and 10.

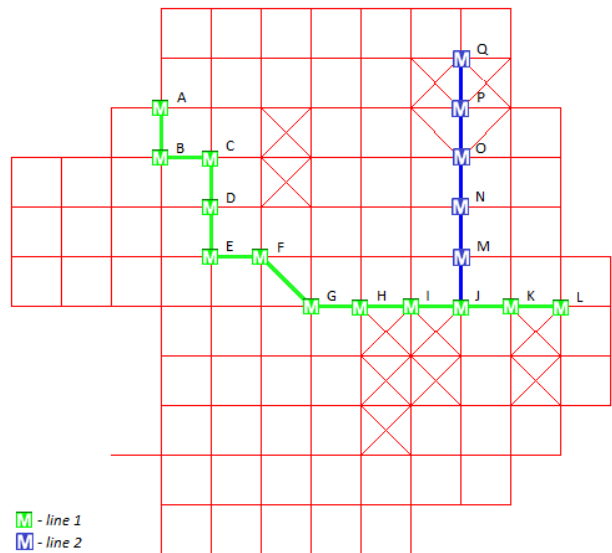
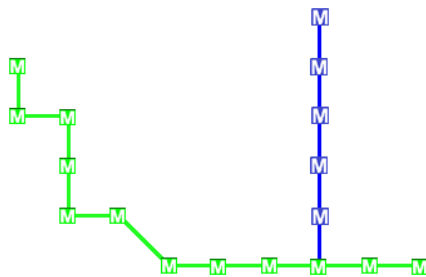


Fig. 9: Graph N specified as the metro network

Fig. 10: Graph N and F together

Figure 9 shows the metro network N consisting of two lines. The network consists of $s = 17$, $p = 16$. The stations can be divided in $s_t = 1$ (J), $s_e = 3$ (A, L, Q) and $s_m = 13$. Figure 10 shows the final model of all connections: graphs F and N together.

Metro line characteristics

A metro line needs to meet some characteristics to be a valid metro line in the system. The basic principle is that every subset of M is a candidate for N , but some subsets will only be practical when carried out by multiple metro lines. The definition of a metro line will be underpinned by the following assumptions:

- A metro line i is a sub graph L_i of M and contains all links adopted in the metro line i and the thereby connected nodes.
- A metro line i follows a finite *path* in graph M .
 - A finite path is a sequence of edges which connect a sequence of nodes with two terminal nodes are captured: the start node s and the end node e (Fig. 11), and it is possible to have $s = e$ (ring trail, Fig. 12)
 - A metro line visits nodes once, except when $s = e$ and the node revisited is s and the line doesn't visit another node after revisiting s (Fig. 13).
 - Metro lines don't cross, cross over edges added to them already (Fig. 14).

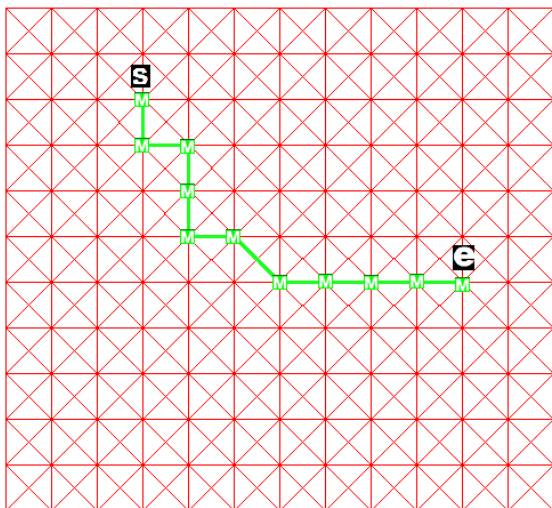


Fig. 11: Valid metro line with $s \neq e$

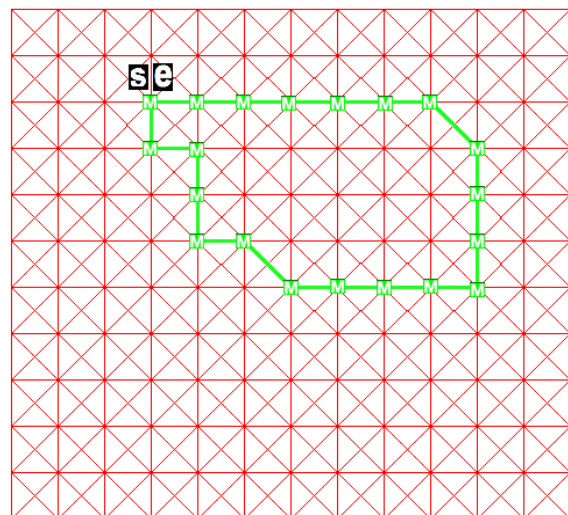


Fig. 12: Valid metro line with $s = e$

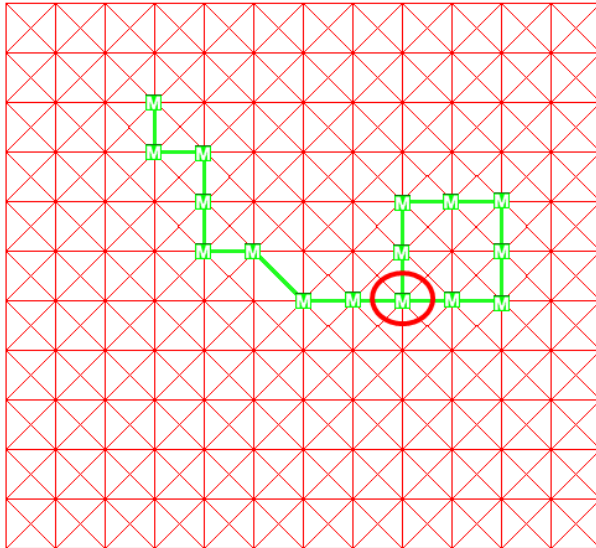


Fig. 13: Invalid metro line, line revisit node with $s \neq e$.

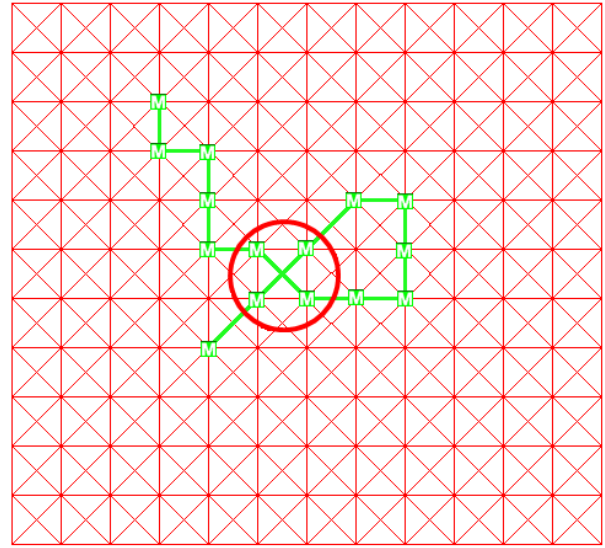


Fig. 14: Invalid metro line, cross over edge

2.3 Paths of travelers

People are moving through the transportation network by a path from the start node s to the end node e , with s, e arbitrary. The sequence of nodes taken to get from s to e depends on the offered transportation lines and availability of those lines in combination with the available footpaths. Regardless of the network, each individual person strives for the path with the shortest travel time d_{se} . The two paths s to e and e to s are considered the same. This leads to $\frac{n_F * (n_F - 1)}{2}$ paths to be considered in the model.

In all models the travelers have the choice to travel either by metro or by foot. The main difference is the speed of moving through the network. This speed difference is expressed by α : the ratio between time to travel over one link by metro and by foot. The time to travel on 1 straight link by metro is taken 1. To travel an oblique link, the time is calculated by the theorem of Pythagoras (see Figure 15).

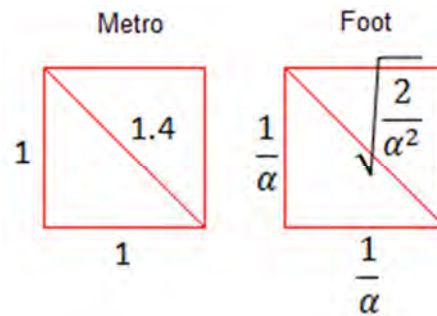


Fig. 15: Time to travel by foot and metro

$$\alpha = \frac{1}{\text{travel time by foot 1 straight link}}$$

To provide more insight about the travel paths, the following example is given:

Figure 16 shows the grid-graph of Bedum. In all the models the time to move on one link with a metro is kept 1. A traveler will move from *A* to *B* and chooses for the shortest path to travel (black arrows). This case is based on the fact that the metro moves faster than a pedestrian ($\alpha < 1$) and the time to wait for a metro is excluded.

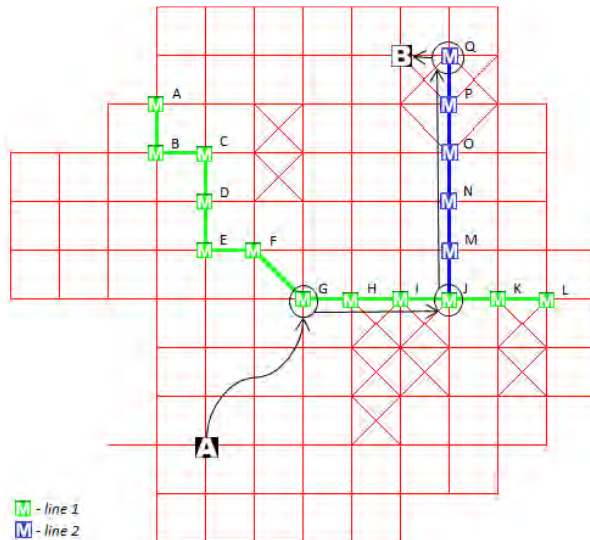


Fig. 16: Path of traveler from *A* to *B*

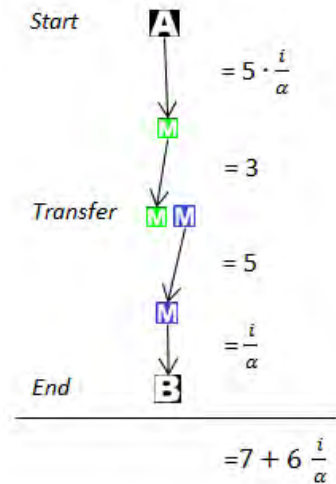


Fig. 17: Path of traveler based on Figure 16

When the metro is run over 7 times faster $\alpha = \frac{1}{7}$. The time to travel from *A* to *B* is $7 + 6 \frac{1}{1/7} = 49$ (see Figure 17). When no metro is available the time to travel from *A* to *B* is $7 \cdot 11 = 77$. The reduction by the metro network is 40%. This model however is based on the assumption of no waiting times. Suppose the metro inter arrival times are uniformly distributed with a mean time of 10 and transitioning from the green to the blue line is a direct connection and takes no extra time will result in an average waiting time of 5. The time to travel from *A* to *B* will be $49 + 5 = 54$. The reduction by the metro network comparing the network without a metro line will be 30%.

2.4 Continuation of this model

This chapter describes the modelling process of an area with a metro network. Through the example of the area around Bedum the theoretical assumptions will be clarified. In the next chapters this model will be used to gain insight in robust metro network designs corresponding to the area. Chapter 3 will follow with robustness, travel time and costs measures which can be applied on this model. Chapter 4 will continue to use the model in practice.

Chapter 3

Measures to compare networks

“Which metrics are suitable to quantify robustness, costs and travel time of a metro network?”

To assess different urban rail transit networks, three measures are used: the robustness R , the travel time T and the costs C . In this chapter the mathematical approach and definition of those measures are given.

3.1 Robustness measures

This paragraph discusses the robustness measures suitable to determine the robustness of rail transit networks.

3.1.1 Literature review

The topic of robustness measures is relatively new, but recently became, because of the extreme increase in demand for knowledge of network architecture in general, led by the increase in computer networks over the last years.

No robustness measure especially for rail transit networks has been defined in the literature. The approach of Derrible (2010) comes closest: he defines a *robustness indicator*, the ratio between transfer stations and monotonic stations (see section 2.2 for explanation of definitions). But this indicator grows linearly with the size of the network, is not valid for round passes with zero transfer stations, and is not capable to distinguish almost equal networks. Although this measure is applied to underground networks, this indicator is not taken into account for further research.

Research in the field of transportation networks brings the research of Scott, Novak, Aultman-Hall, & Guo (2006). In this research a *robustness measure* is defined, from the perspective of the field of transportation network architecture. They disclose that the standard measure, the V/C ratio (volume to capacity ratio), is referred to be restrictive in analyzing a transportation network as a whole. They define a new measure the Network Robustness Index to identify the robustness of a link in comparison to the whole network. This robustness indicator works fine as long every node in the network is accessible via at least two edges (vertex connectivity of minimal 2). This measure is applicable to the graph F and N together and is a good indicator of the dependence of a link to the environment. This measure could be useful for determining the robustness of metro network. This measure is

taken into account for further research. The definition of the Network Robustness Indicator is defined in section 3.2.

Ellens & Kooij (2011) evaluate classical graph measures and spectral graph measures which are intuitively relevant for evaluating the robustness of a network. The analysis of fourteen measures has shown that all measures are able to place some small example graphs in the same order of robustness as we would do intuitively, but not all measures are able to distinguish the given sample graphs. From this research one measure is useful for the stated research question in this paper: the effective resistance. This measure is applicable for every connected graph with vertex degree ≥ 1 , which makes the effective resistance a perfect measure for both graph G , M and F . This measure is described more detailed in the research of Ellens et al. (2011). In this research this measure is taken as robustness measure for all graphs.

Based on this literature review, the following robustness measure is explained in detail:

- **Effective resistance** - Spectral graph measure, specially proposed for robustness measurement based on the Laplacian spectrum
- **Network robustness index** – Index measure proposed from the transportation network design field.

3.1.2 Effective resistance

The effective resistance is a measure in the range of ‘spectral graph measures’. This measure is inspired by the resistor theory from physics. The graph is seen as an electrical circuit, where two vertices are connected and the electricity find its way through the network. Every edge (i, j) corresponds to a resistor of r_{ij} ohm. The resistance of the whole system can be calculated by the series and parallel manipulations as known in physics.

Classical effective resistance

Two edges corresponding to resistors with resistance $r_1 = 1$ and $r_2 = 1$ Ohm. If those edges are series connected, the effective resistance is $r_1 + r_2 = 1 + 1 = 2$ Ohm. If those edges are parallel connected, the effective resistance is $\frac{r_1 r_2}{r_1 + r_2} = \frac{1}{2}$ Ohm. The effective graph resistance is the sum of all effective resistances over all pairs of vertices i, j :

$$R = \sum_{1 \leq i < j \leq n} R_{ij}$$

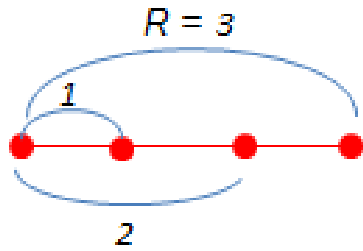


Fig. 18: Series connected part of a graph

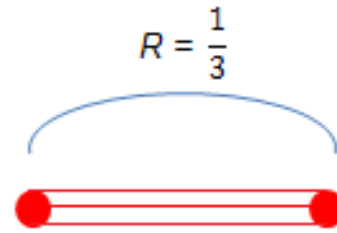


Fig. 19: Parallel connected part of a graph

Figure 18 and 19 show the effect of the effective resistance applied to graphs. If robustness is in effect, the resistance should be as low as possible. Figure 19 shows that with more alternative routes the resistance is lower and Figure 18 shows that while the connected paths are longer, the resistance increases. These two main effects of the effective resistance are crucial for a robust network.

Effective resistance by Laplacian eigenvalues

An additional advantage of the resistance metric is its relation to the Laplacian eigenvalues (Klein & Randic, 1993). This method is, especially when calculating the effective resistance computational, an advantage above the classical series and parallel manipulations. To calculate the Laplacian eigenvalues of the graph, some calculations need to be done. We start with three matrices: adjacent matrix A , degree matrix D and the Laplacian matrix Q .

The adjacency matrix A is a matrix, with a 1 or 0 at the position (i, j) according to whether i and j are adjacent $A(i, j) = 1$ or not $A(i, j) = 0$. The degree matrix is a diagonal matrix D with at position (i, i) the corresponding vertex degree of i . The Laplacian matrix Q is defined as the difference between the degree matrix and the adjacent matrix: $Q = D - A$.

By the characteristics of the Laplacian matrix, the eigenvalues of the Laplacian are real, non-negative and the smallest is 0. We can order the eigenvalues such that $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Taking only the eigenvalues ≥ 0 , leads to the following function which links the eigenvalues to the effective resistance:

$$R = n \sum_{i=2}^n \frac{1}{\lambda_i}$$

For the full proof of this formula, see the work of Klein and Randic (1993) or the work about effective graph resistance by Ellens et al. (2011).

Ellens (2011) proved that the effective graph resistance strictly decreases when an edge is added to the network. Although the work doesn't specify the findings for large graphs, it is

plausible that the same outcomes can be found by applying the effective graph resistance on large graphs.

This measure expresses the resistance of the network, and the higher the resistance, the lower to the robustness: maximizing robustness, leads to minimizing the resistance.

Effective resistance with weights

The effective resistance is not a measure that can easily incorporating weights. This measure looks at the total effective resistance in which all the possible alternative routes between all points are examined and not to specific routes in general. Yet there is a possibility to circumvent this restriction, by dividing an area into regions of priority. The edges in the high priority areas get a higher priority in the form of a higher residence on those edges. Therefore a kind of priority between different areas can be incorporated. The priority weights are applied to graph F . The following example shows the effect on prioritize different parts of a graph.

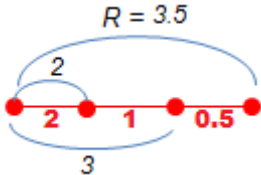


Fig. 20: Series connected part of a graph with priority weights

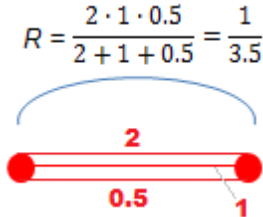


Fig. 21: Parallel connected part of a graph with weights

Figures 20 and 21 show the same examples of the graphs in Figures 18 and 19 but with priority weights. One edge is divided by 2 to reduce priority, a second edge is multiplied by 2 for increasing priority and the third edge is kept equal. This process leads to three edges with weights 2, 1 and 0.5. For series connected graphs the effective residence stays equal if the weights are summing up to the same value (Fig. 20). For parallel connected graphs this is impossible, because the ratio between summing and multiplying values should be equal: a two variable problem with always one solution (for positive values, see Fig. 21). Consequently models with different priority weights on F are difficult to compare.

To incorporate the connections in F as more difficult than connections in M the same method is used. A connection in F is α times higher (section 2.3) in resistance than the connections in M . In practice leads this method to a resistance of a connection in F with two weights: the priority weights and the weight α .

3.1.3 Network robustness index (NRI)

A study (Scott, Novak, Aultman-Hall, Guo, 2006) from the field of transportation network design provides a measure to determine the value of an individual edge within the overall transportation system. This measure is based on the flow (volume of traffic) and travel time on each link relative to the complete network and takes the spatial relationships and rerouting possibilities into account.

Let d_{ij} be the shortest path from vertices i to j when the whole system is available. The sum of all lengths of the shortest paths is the cost value c . This expression is called the Wiener index in graph theory:

$$c = \sum_{i=1}^n \sum_{j=i+1}^n d_{ij}.$$

The same cost value can be calculated within the model when a particular edge a is removed, with d_{ij}^a the shortest path in the network without edge a .

$$c_a = \sum_{i=1}^n \sum_{j=i+1}^n d_{ij}^a.$$

Finally, the network robustness index can be computed for link a . This index relies on the addition of link a to the whole working network:

$$NRI_a = \frac{c}{c_a}.$$

The average can be taken as network robustness index for the entire network. A disadvantage of this network is that without a particular link, the system still has to be fully connected:

$$NRI = \frac{1}{m} \sum_{a=1}^m NRI_a.$$

3.2 Travel time measure

The travel time measure is the average time of an arbitrary path through the system. This corresponds to the average shortest path in a network equivalent to the average distance belongs to the set of 'classical graph measures'.

The *average distance* \bar{d} of a graph is the average path distance over all pairs of vertices, with d_{ij} the shortest path between i and j (see section 1.3.1) and w_{ij} the weight of the path between i and j . This weight can be used to take different paths from i to j with a specific heaviness in calculation.

$$T = \bar{d} = \frac{1}{2W} \sum_{i=1}^n \sum_{j=j+1}^n d_{ij} (w_i + w_j), \quad W = \sum_{k=1}^n w_k$$

This measure is based on the shortest path in a graph. The shortest path in a graph is solvable with the classical *Dijkstra's Shortest Path algorithm*. This algorithm holds for graph F , but is not applicable when the graphs F and N are considered together. Graph N consists of lines with a specified transfer time. The shortest path is equal to the shortest travel time, where transfers increase the total travel time. The following two subparagraphs show how to deal with the shortest path in all models: F and $\{F, N\}$.

3.2.1 Shortest path in graph F

In the grid graph F , the shortest path is solvable by the ordinary Dijkstra's Shortest Path algorithm from s (start node) to e (end node). This path will also be the shortest path in graph $\{F, N\}$ if there is a direct connection in N with one line available who's imitating the shortest path given by the algorithm in F . The reverse of this statement is generally not true: if there is not a direct line available in N whose imitating the shortest path by the algorithm in F , it still can be the shortest path in $\{F, N\}$ but this is not commonly given. Appendix I describes the ordinary Dijkstra's Shortest Path algorithm.

3.2.2 Shortest path in graph N

In graph N the shortest path is solvable with the Dijkstra's Shortest Path algorithm with an extension. The extension takes a transfer time f in addition to determine the shortest path. The following expression is used to determine the transfer time form line l_i to l_j at station s :

$$f_{l_i l_j}(s) = \frac{IATime l_j}{2} + w_{l_i l_j}(s)$$

In this formula, the inter arrival times of metro vehicles of an arbitrary line are uniformly distributed and $w_{l_i l_j}(s)$ is an approximation of the extra time to move through a particular station s to transfer from line i to line j . In this paper we assume $w_{l_i l_j}(s) = 0$ and $f_{l_i l_j} = \frac{IATime l_j}{2}$, unless otherwise stated.

To keep the shortest path algorithm for graph N greedy, the assumption that $f_{l_i l_j}$ is constant for all i and j is kept. Based on this assumption the algorithm can be described as follow:

Step 0: This algorithm is based on graph N as collection of l lines.

Step 1: The line-based algorithm makes use of a transition matrix T_{l_i} where all direct known distances for one line are summarized. No calculation is needed to fill this initial matrix of every line l_w .

$$T_{l_w} = \begin{matrix} & N_1 & N_2 & N_3 & \dots & N_n \\ \begin{matrix} N_1 \\ N_2 \\ N_3 \\ \vdots \\ N_n \end{matrix} & \begin{bmatrix} 0 & t_{12} & t_{13} & \dots & t_{1n} \\ t_{21} & 0 & t_{23} & \dots & t_{2n} \\ t_{31} & t_{32} & 0 & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & t_{n3} & \dots & 0 \end{bmatrix} \end{matrix}$$

Step 2: This matrix undergoes a recursive process till a steady system is found by the following property:

$$t_{ij} = \min(t_{ik} + t_{kj}, \text{for all nodes } k)$$

Through this property all possible routes (from node i to j via k) through the network are considered and if a shorter path is found, the transition time is replaced by the shorter variant. The process will continue till equilibrium is reached.

Step 3: Steps 1 and 2 are performed for all lines in N .

Step 4: After the recursive processes of all lines, two lines (or network parts) are combined by the following property:

$$t_{ij} = \min(T_{l_w}(t_{ik}) + T_{l_v}(t_{kj}) + f, T_{l_v}(t_{ik}) + T_{l_w}(t_{kj}) + f, T_{l_v}(t_{ij}), T_{l_w}(t_{ij}), \text{for all nodes } k).$$

Step 5: The two network parts v and w are now a new part of the network summarized in a new transition matrix with shortest paths. This obtained matrix is now ready to be merged with another part of the network.

3.2.3 Shortest path in graph $\{F, N\}$

The shortest paths in graph $\{F, N\}$ is a combination of the shortest path in graph F and N . The matrix of the shortest paths of $\{F, N\}$ is generated by the following formula:

$$t_{ij} = \min(F(t_{ik}) + N(t_{kj}), N(t_{ik}) + F(t_{kj}), N(t_{ij}), F(t_{ij}) \text{ for all nodes } k).$$

This formula combines all possibilities between walking or taking the metro in the graph.

3.3 Cost function

The cost function puts a constraint on the design of networks. The reason to use a cost function is the fact that every design of a metro network is subject to a budget. The development of an extreme comprehensive network is highly robust, but will never be implemented because of the high investment.

The cost function is defined as follows:

$$C = K + c_p \sum_{i=1}^p p_i k_i + c_m s_m + c_e s_e + c_t \sum_{j=1}^{s_t} s_j(l) + \sum_{k=1}^r r_k(l)$$

The variables are:

- fixed cost K
- the cost of placing a metro part with length 1 c_p
- total distance of all metro parts $\sum_{i=1}^k p_i k_i$, with p_i the number of metro parts from the k_i th layer and k_i de ratio value of the i th layer (the distance of a straight grid line is kept 1)
- costs for building a monotonic station c_m
- number of monotonic stations s_m
- costs for building an end station c_e
- number of end stations s_e
- costs for building a transfer station c_t
- s_t the number of transfer stations
- $s_j(l)$ the number of lines passing transfer station j
- the number of underground or leveled crossings r
- $r_k(l)$ is the number of lines crossings at cross r_k .

The definition of the different types of stations can be found in section 2.2. The following example shows the calculation of costs for a metro network. The municipality of Bedum considers the metro network of Figure 22 and is determining the costs by the costs function.

- The fixed costs are equal to K .
- The costs for the metro parts are equal to $c_p \sum_{i=1}^p (p_i k_i) = c_p (18k_1 + 5k_{\sqrt{2}}) = c_p (18 + 5\sqrt{2})$, based on the number of metro parts in the radius of $k = 1$ (18) and the number of metro parts in the radius of $k = \sqrt{2}$ (5).
- The costs for the transfer stations are $c_t \sum_{j=1}^{s_t} (s_j(l)) = c_t (2 + 2) = 4c_t$, based on two transfer stations, both connecting two lines.

- The costs for the crossings are $\sum_{k=1}^r (r_k(l)) = 2$. In the metro network is one cross present, where the yellow and green lines are crossing underground or leveled. This cross concerns two lines which lead to total costs of 2.
- The costs for building the end stations $c_e s_e = 3c_e$, because three end stations are kept in the model.
- The costs for building monotonic stations $c_m s_m = c_m (s - (s_e + s_t)) = 20c_m$. The monotonic stations are all stations which are not a transfer of end station (section 2.2).

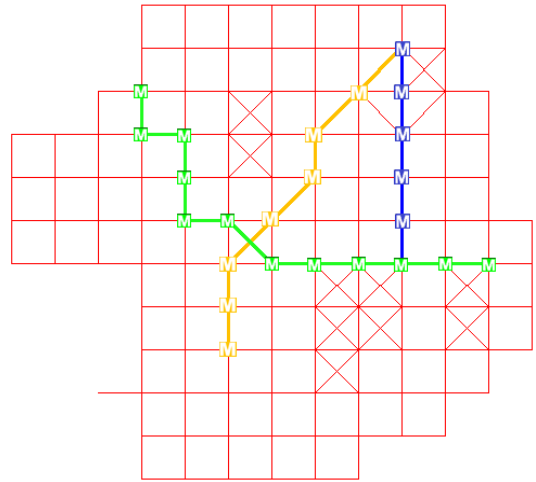


Fig. 22: Example of metro network in Bedum shown as graph F and N .

This enumeration leads to an overall cost C of $K + c_p(18 + 5\sqrt{2}) + 4c_t + 2 + 3c_e + 20c_m$.

Chapter 4

One metro line network on a greenfield

“How to design a robust metro network on a greenfield and what is the most effective investment to improve robustness of a metro network?”

This chapter analyzes how a metro line should be placed on a greenfield area for maximum robustness. For a graph N , with $N = L_1$ (one metro line included) is searched, to minimize the travel time, maximize robustness and the feasibility under a limited budget.

4.1 Approach and methodology

To find an answer to the question which metro line can be adopted best, all possible networks N as sub graph of M are considered in respect to the graph F . In this methodology all possible networks should be searched first and then be measured by the measures determined in chapter 3. In this way it is always possible to find the optimal network with respect to robustness subject to a cost function. However, this method has a high complexity because the number of possible networks N is increases exponentially when the number of nodes is increased.

To handle this approach, the model is divided into three smaller problems: finding all possible metro networks with one line, obtaining all measures for the resulting networks and analyzing the obtained data and models.

4.2 Algorithm of finding all metro lines

The first sub-problem as defined in section 4.1 is finding all possible metro networks $N = L_1$ in M . This problem is equivalent to finding all possible metro lines in M what is equal to finding all paths ($s = e \wedge s \neq e$). Almost all algorithms can find all paths with a complexity of at least $O(n^4 \cdot n!)$ but possibly worse. The extreme complexity makes it difficult to analyze all possibilities in N for large graphs. This chapter therefore focuses on small graphs to obtain insight in robust designs, to develop a generic approach (chapter 5).

For this research an algorithm based on a recursive birth-death search is developed and implemented in the statistical software package R with the use of the external R-package *igraph*. The algorithm is given next:

Algorithm 1 Get All Paths

```
1. Procedure getPaths(G)
2.   for (i = 1 to (m of G))
3.     T <- delete.edges(G,i)
4.     paths <- findPaths(T,get.edgelist(G)[i,1],get.edgelist(G)[i,2]_
       ,paths)
5.     paths <- findPaths(T,get.edgelist(G)[i,2],get.edgelist(G)[i,1]_
       ,paths)
6.   End for
```

Algorithm 2 Finding All Paths

```
1. Procedure findPaths(G,s,e,a, paths)
2.   if has.no.neighbors(G,e) then
3.     paths <- paths + found combination s-a-e
4.     return(paths)
5.   else
6.     sp <- shortest.paths(G, e)
7.     T <- G
8.     T <- delete.vertices(T, e)
9.     For (every node i in sp with length 1)
10.      if (i == s && noLineCross(s,a+e,s))
11.        paths <- paths + found combination s-a-e-s
12.      ElseIf (noLineCross(s,a+e,i))
13.        paths <- findPaths(G,s,i,a+e,paths)
14.      End if
15.    End for
16.    paths <- paths + found combination s-a-e
17.    return(paths)
18.  End if
```

4.3 The 4-nodes area

The 4-nodes area is used to gain insight in the behavior of the robustness measures and the designs corresponding to the 4-nodes area.

In the 4-nodes model assumptions of parameters described in chapter 2 and 3 are made.

The speed of walking compared to traveling by metro α is kept $\frac{1}{7}$ (see section 2.3). This value is based on the average speed of walking and metro journey times in multiple cities. In the

cost function the costs of a metro part with length 1 is kept 1 and the cost for a station $c_m, c_t, c_e = 2$.

In this model homogeneous and heterogeneous areas investigated. In a homogeneous area all nodes are equally important while in heterogeneous areas the nodes are distinguished on importance by weights.

4.3.1 The empty 4-nodes area

The empty 4-nodes area is represented as a graph F of 4 nodes with $k = 1$ and $k = \sqrt{2}$ (Figures 23 and 24). In these models $\frac{n(n-1)}{2} = 6$ unique combinations of start- and end node can be found. The 4-node model with $k = 1$ knows a robustness R of 35 and $T = 9\frac{1}{3}$. The 4-nodes model with $k = \sqrt{2}$ knows a robustness of R of $23\frac{1}{3}$ and $T = 7\frac{14}{15}$. The robustness of the area with $k = \sqrt{2}$ is higher than the robustness of the area with $k = 1$ because this area knows more alternative paths between the nodes s and e . The average travel time is also lower because of the more direct paths between the nodes. Costs are 0, because no metro network is built.

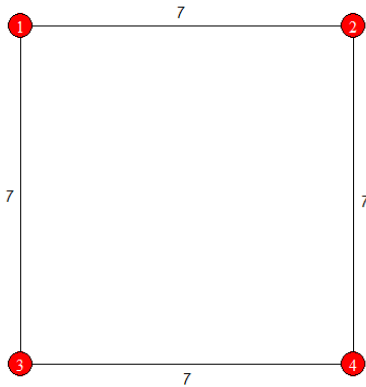


Fig. 23: homogeneous greenfield area of a 4-nodes model with $k = 1$

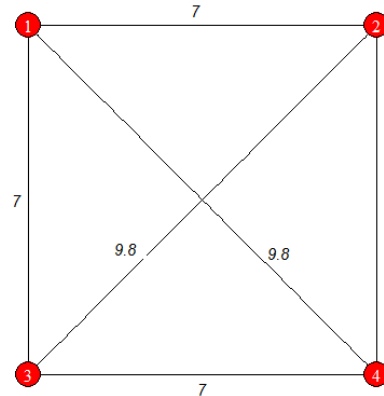


Fig. 24: homogeneous area of a 4-nodes model with $k = \sqrt{2}$

Figures 23 and 24 are both show a homogeneous area. This means that no priority is given to specific nodes or edges and all nodes and edges are equal. Figures 25 and 26 show a heterogeneous 4-nodes area with different priorities. Section 3.1.2 shows a different approach for determining robustness of a heterogeneous. This priority weights can't be specified on nodes, but can specify individual edges and areas. Figure 23 and 24 show an area consists of a city center and two suburbs. The priority inside the city center is kept 2, the priority between city center and suburb is kept 1 and the priority between the suburbs 0.5. These measures are included in graph F .

The model with $k = 1$ knows a robustness R of 35.78 and the weighted travel time $T_w = 9.15$ and the model with $k = \sqrt{2}$ a robustness R of = 23.46 and the weighted travel time $T_w = 7.89$. The robustness of an empty asymmetric model is higher than an empty symmetric model. This phenomenon is described in section 3.1.2.

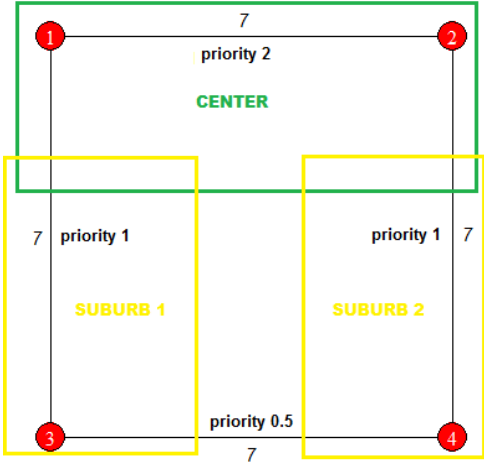


Fig. 25: Heterogenous greenfield area of a 4-nodes model with $k = 1$

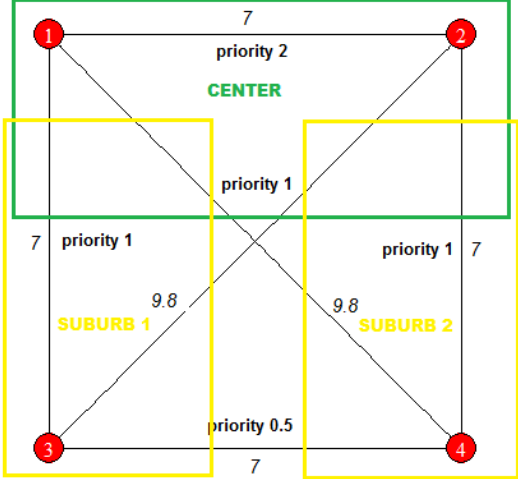


Fig. 26: Heterogenous greenfield area of a 4-nodes model with $k = \sqrt{2}$

4.3.2 Placing 1 metro line in a 4-nodes area by maximizing the robustness

The 4-nodes model with $k = 1$ has 13 ways to place 1 metro line, the 4-nodes model with $k = \sqrt{2}$ has 31 ways of placing 1 metro line in the system. In the homogeneous areas the diversity between the different possible graphs for N is lower because all symmetric versions of N are generating the same values. With an unlimited amount of money, the best design is in all cases the same (see Fig. 27, 28). Instinctively, it is not surprising that the design corresponding to graph N leads to the highest robustness. It is the only design where four edges are allowed and all nodes are connected by a subway line. In addition, the design is known as a bypass, what makes each node accessible from two sides.

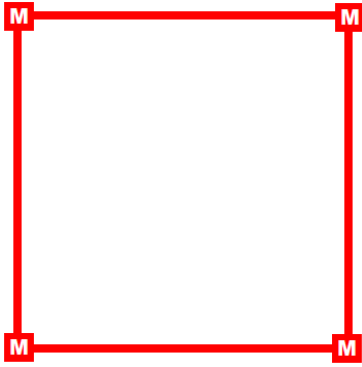


Fig. 27: Metro network in 4-nodes model with highest robustness for hetero- and homogeneous area, $k = 1$

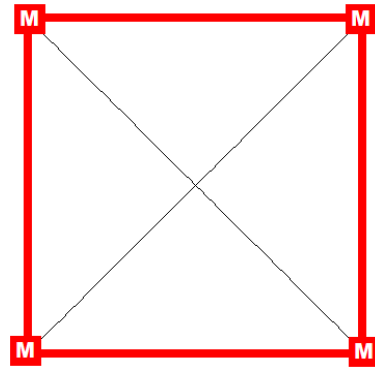


Fig. 28: Metro network in 4-nodes model with highest robustness for hetero- and homogeneous area, $k = \sqrt{2}$

The robustness of all designs is described in the following table:

	Fig. 23	Fig. 25	Fig. 24	Fig. 26
Asym	Sym	Asym	Sym	Asym
k -value of F	$k = 1$	$k = 1$	$k = \sqrt{2}$	$k = \sqrt{2}$
k -value of M	$k = 1$	$k = 1$	$k = \sqrt{2}$	$k = \sqrt{2}$
C - empty N	0	0	0	0
T_w - empty N	9.333	9.150	7.933	7.892
R - empty N	35	35.78	23.333	23.46
C	17	17	17	17
T_w	1.333	1.308	1.333	1.308
$T_w, t_f = 5$	6.333	6.308	6.333	6.308
R	4.375	4.319	4.091	4.042
Improvement R	87,50%	87,93%	82,47%	82,77%
Improvement T_w	85,72%	85,70%	83,20%	83,43%
Improvement $T_w, t_f = 5$	32,14%	31,06%	20,17%	20,07%

Table 1: Measurement results of the 4-nodes model (Fig. 23 to 26)

Based on the results in table 1 some statements can be made. Firstly, it is important for the improvement in robustness which value of k is used to construct for F . The less the area is connected through F , the more the addition of a subway line in robustness and equivalent time travel is. Secondly, all the models describe the same graph N as the most robust

design. This is mainly because of the few possible ways to place a line in a 4-nodes model and therefore perhaps not useful to draw a conclusion. Finally, the improvement in time traveled with transfer time t_f is significantly lower than without the transfer time. A transfer time of 5 refers to a continuous schedule with every 10 time units a metro. A metro schedule with every 15 time units a metro (expectation of 7.5) will never beat the time to walk in the models and the metro line will never be used. The value of transfer time is a high impact parameter.

4.3.3 Placing 1 metro line in a 4-nodes area by maximizing robustness and a boundary at costs.

The 4-nodes model with $k = 1$ has 13 ways to place 1 metro line, the 4-nodes model with $k = \sqrt{2}$ has 31 ways of placing 1 metro line in the system.

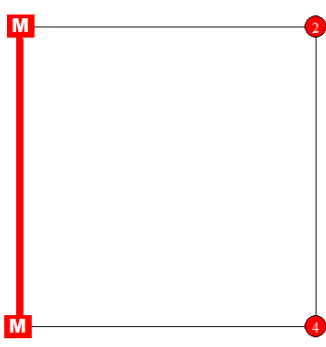


Fig. 29: Symmetric model
 Sym: $R = 25.200$
 Asym: $R = 24.983$
 $C = 10$

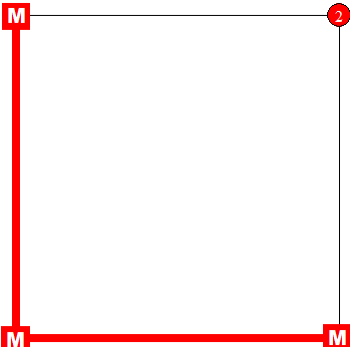


Fig. 30: Symmetric model
 Sym: $R = 14.924$
 Asym: $R = 18.416$
 $C = 13$

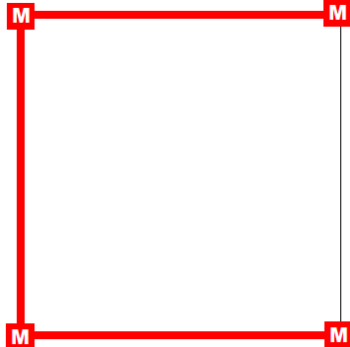


Fig. 31: Symmetric model
 Sym: $R = 7.159$
 Asym: $R = 7.076$
 $C = 16$

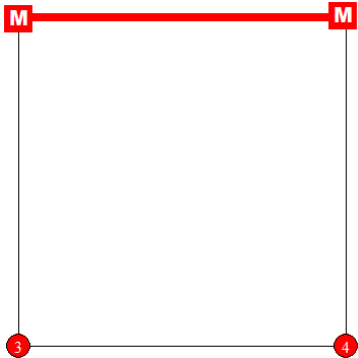


Fig. 32: Asymmetric model
 Sym: $R = 25.200$
 Asym: $R = 21.443$
 $C = 10$

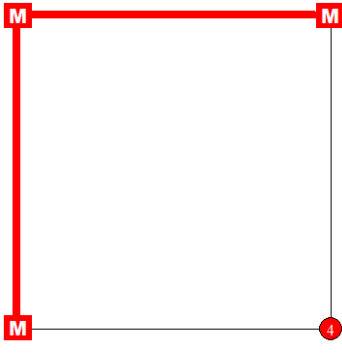


Fig. 33: Asymmetric model
 Sym: $R = 14.924$
 Asym: $R = 11.562$
 $C = 13$

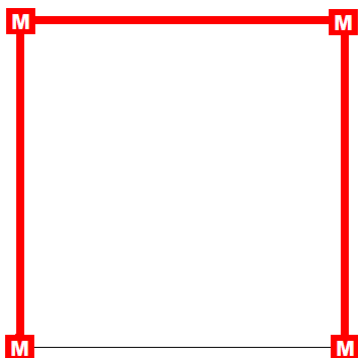


Fig. 34: Asymmetric model
 Sym: $R = 7.159$
 Asym: $R = 6.373$
 $C = 16$

Figure 29 till 34 show different models with the robustness values and costs. Figure 29, 30 and 31 are based on a symmetric model, Figure 32, 33 and 34 on the asymmetric model.

Through the comparison of the models the influence of the priority weights is visible. For the symmetric designs the position of the line is not important, only the structure. The asymmetric models are far greater in versatility and are looking for a particular model which provides the high priority parts of a subway line first.

Through this approach and methodology, the travel time, costs and robustness are calculated for every possible metro line in the model. This generates a data set in which for every threshold in costs, a metro line with the highest robustness can be obtained. The 9-nodes model in the following section shows more diversity and insights.

4.4 The 9-nodes area

This chapter is to develop some insight about the behavior of the robustness measures and the designs corresponding to the 9-nodes area. In this model some assumptions on parameters of chapter 2 and 3 are made. The speed of walking comparing to traveling by metro is kept $\frac{1}{7} = \alpha$. This value is based on the average speed of walking and metro journey times. In the cost function the costs of a metro part with length 1, is kept 1 and the costs for a station $c_m, c_t, c_e = 2$.

4.3.1 The empty 9-nodes area

The 9-nodes area is represented as a graph F of 9 nodes with $k = 1$ and $k = \sqrt{2}$ (Figure 34 and 35). In these models $\frac{n(n-1)}{2} = 36$ unique combinations of start- and end node can be found. The 9-node model with $k = 1$ knows a robustness R of 241.5 and $T = 14$. The 9-nodes model with $k = \sqrt{2}$ knows a robustness R of 145.02 and $T = 11\frac{2}{3}$. The robustness of the area with $k = \sqrt{2}$ is higher than the robustness of the area with $k = 1$ because this area knows more alternative paths between the nodes s and e . The average travel time is also lower because of the more direct paths between the nodes. Costs are 0, because no metro network is built.

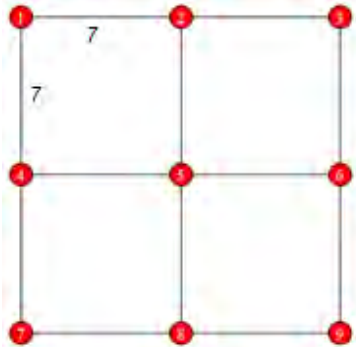


Fig. 34: Homogeneous greenfield area of a 9-nodes model with $k = 1$

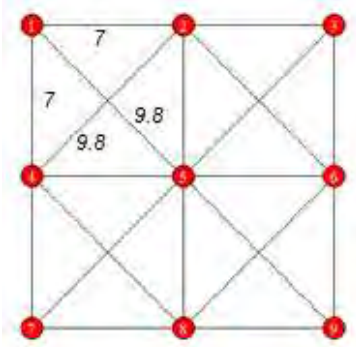


Fig. 35: Homogeneous area of a 9-nodes model with $k = \sqrt{2}$

Figure 34 and 35 both show a homogeneous area. Figure 36 and 37 show a heterogeneous 9-nodes area with different priorities. Section 3.1.2 showed a different approach for robustness on a heterogeneous area. This priority weights can't be specified on nodes, but can specify individual edges and areas. Figure 36 and 37 show an area existing of a city with a city center, two suburbs and an inhabited countryside. The priority inside the city center is kept 2, the priority of the other parts of the city 1.5, the suburbs are priority 2 and the countryside knows a priority of 0.5. All the priorities on the links are average priority values of the two nodes connected by the link. These measures are included in graph *F*.

The empty model gives different robustness measures. The model with $k = 1$ knows a robustness R of 241.5 and the weighted travel time $T_w = 13.84$ and the model with $k = \sqrt{2}$ a robustness R of = 145.02 and the weighted travel time $T_w = 11.53$. The robustness of an empty asymptotic model is higher than an empty symmetric model. This phenomenon is described in section 3.1.2.

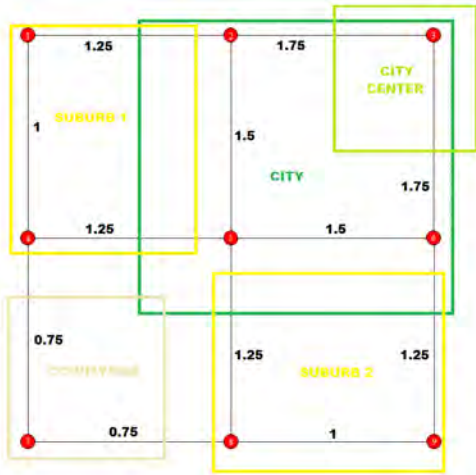


Fig. 36: Heterogeneous greenfield area of a 9-nodes model with $k = 1$, with priority weights

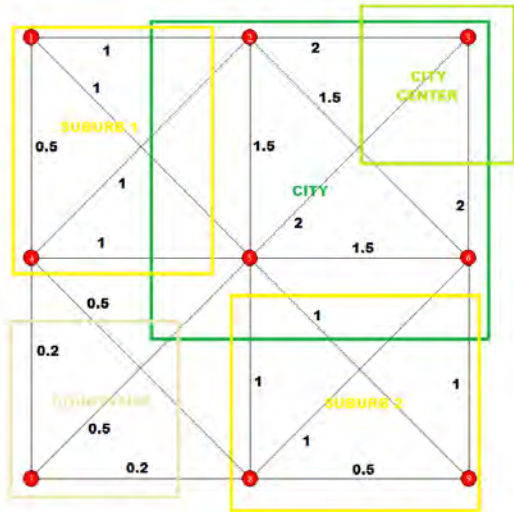


Fig. 37: Heterogeneous greenfield area of a 4-nodes model with $k = \sqrt{2}$

4.3.3 Placing 1 metro line in a 4-nodes area by maximizing robustness and a boundary at costs

The 9-nodes model with $k = 1$ has 337 ways to place 1 metro line, the 9-nodes model with $k = \sqrt{2}$ has 3.786 ways of placing 1 metro line in the system. The following pictures (Fig. 38 to 45) show the designs of metro networks with the highest and second highest robustness. The longest metro line has a length of 8 ($k = 1$) or 9 ($k = \sqrt{2}$).

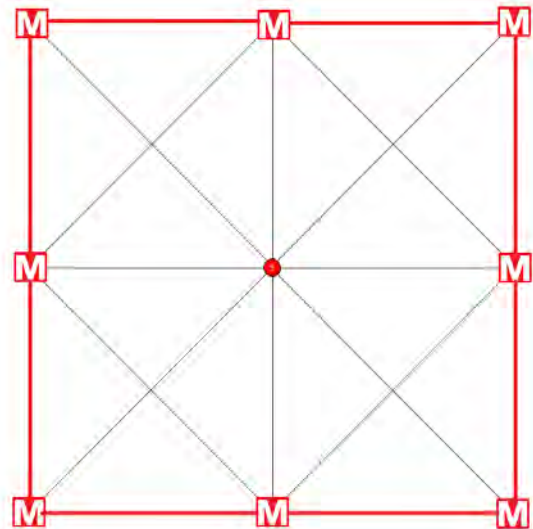
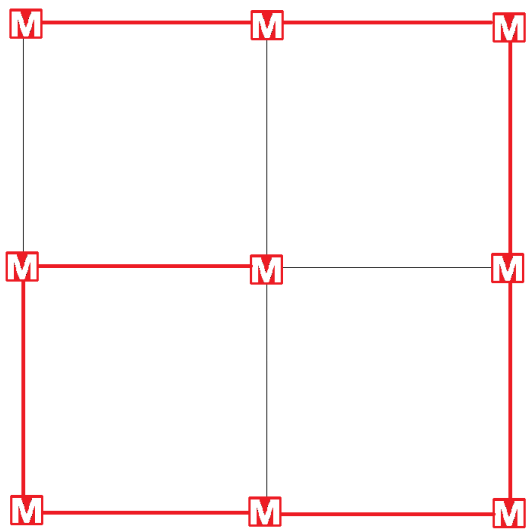
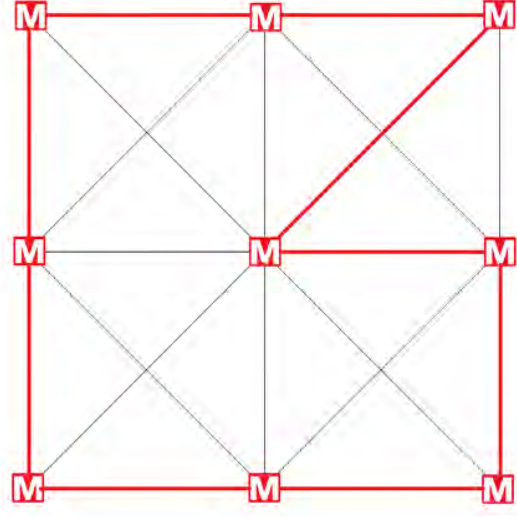
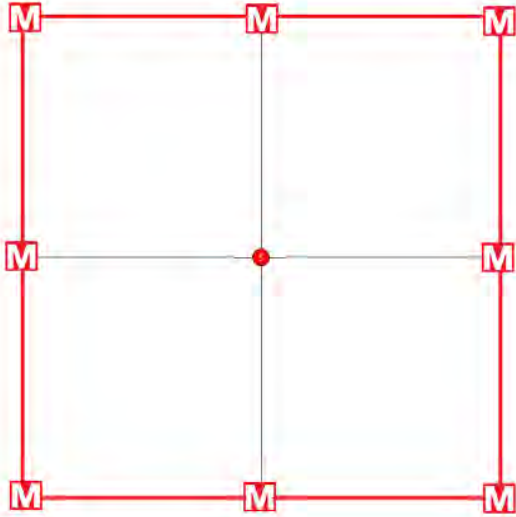


Fig 38, 39: Metro network in 9-nodes model with highest and second highest robustness for homogeneous area, $k = 1$

Fig 40, 41: Metro network in 9-nodes model with highest and second highest robustness for homogeneous area, $k = \sqrt{2}$

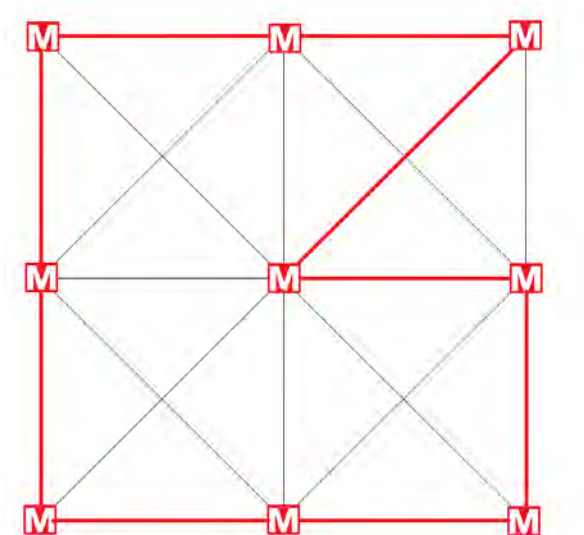
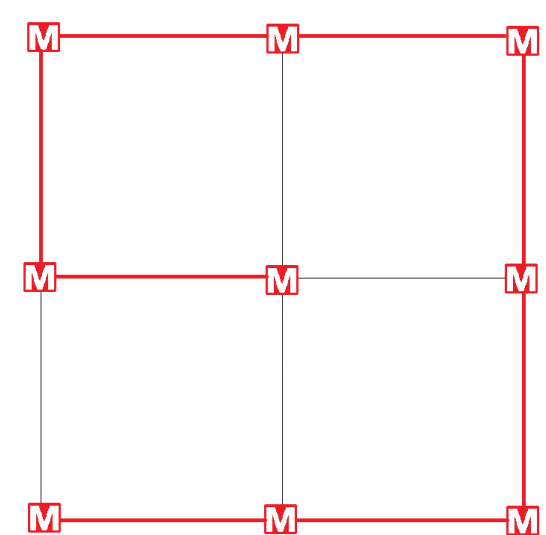
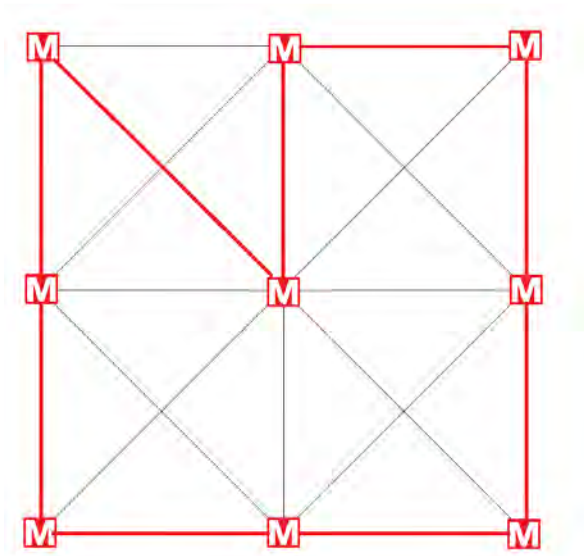
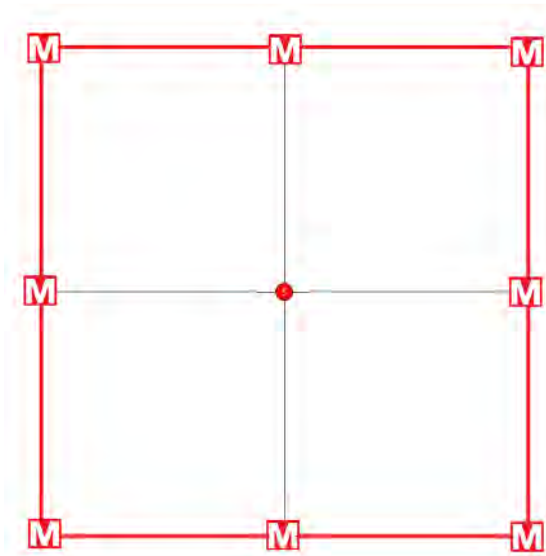


Fig 42, 43: Metro network in 9-nodes model with highest and second highest robustness for heterogeneous area, $k = 1$

Fig 44, 45: Metro network in 9-nodes model with highest and second highest robustness for heterogeneous area, $k = \sqrt{2}$

Those figures show some interesting results. (1) For all models, a metro network constructed as bypass results in a higher robustness than a straight metro line, as long as networks over approximately the same size are compared. (2) The robustness is also significantly higher when the number of nodes and edges are increased.

	Fig. 38	Fig. 40	Fig. 42	Fig. 44
Asym	Sym	Sym	Asym	Asym
k -value of F	$k = 1$	$k = \sqrt{2}$	$k = 1$	$k = \sqrt{2}$
k -value of M	$k = 1$	$k = \sqrt{2}$	$k = 1$	$k = \sqrt{2}$
C - empty N	0	0	0	0
T_w - empty N	14	11.67	13.84	11.53
R - empty N	241.5	145.02	241.5	145.02
C	29	32	29	32
T_w	3.44	2.5	3.56	2.58
$T_w, t_f = 5$	8.44	7.5	8.56	7.57
R	52.71	37.86	58.90	40.32
Improvement R	78.17%	73.89%	75.61%	72.20%
Improvement T_w	75.43%	78.58%	74.28%	78.32%
Improvement $T_w, t_f = 5$	39.71%	35.73%	38.15%	34.95%

Table 2: Measurement results of the 4-nodes model (Fig. 34 to 36)

In table 2 some results are valid which are also mentioned for the 4-nodes model. In this model it is also important for the improvement in robustness which value of k is used to construct for F . The less the area is connected through F , the more the addition of a subway line in robustness and equivalent time travel is. Furthermore, the improvement in time travel with transfer time t_f is significant lower than without the transfer time. A transfer time of 5 means a continue schedule with every 10 time units a metro. A metro schedule with a metro every 15 time units a metro (expectation of 7.5) will never beat the time to walk in the models and the metro line will never be used. The value of transfer time is a high impact parameter.

Chapter 5

Limitations and approaches of finding a robust metro line

“How to design a fast yet accurate heuristic for improving robustness of metro networks?”

In all models so far all possible networks are searched and the robustness is calculated on the basis of the models. Hence it is always possible to find the optimal network with regard to robustness subject to a cost function. However, this method has such high complexity that for large graphs this method is useless. This chapter looks for a general methodology, which has a lower complexity, but gives an approximately good network.

The start with an estimation of the number of edges and nodes a design will be shown. Subsequently, two methods are proposed to give an approximately good design.

5.1 Estimation of the number of edges and nodes

In this paragraph we estimate the number of edges and nodes of the design with the highest robustness based on the costs. This estimate is valid because, in general, the larger the network of the metro, the more robust it is. Adding a link increases the robustness, deleting a link reduces the robustness. The relation between the number of nodes/edges and the costs is discussed here first. Second, an explanation of a simplified cost function. Finally, the estimation is described.

5.1.1 Dependence of costs and robustness

All models in chapter 4 show dependence between the cost and the robustness. This dependence is instinctively plausible. The more can be invested, the larger the network would be, the better the robustness. Besides the dependence some clusters are shown because of the fact that the cost measure is discreet. Figure 46 shows the relationship for the 9-nodes model.

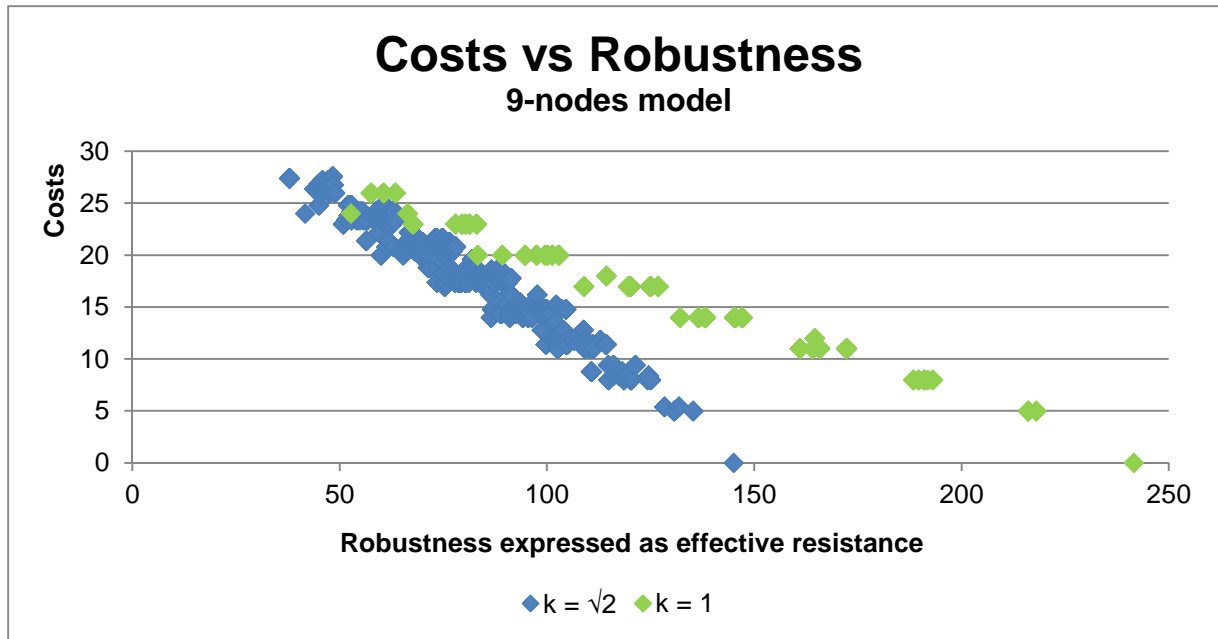


Fig 46: Plot of robustness (expressed as effective resistance) in subject to costs

This means that, given a specified available amount to invest (equals costs), a prediction of the robustness of a metro line can be given. And the improvement of robustness to a particular area can be given to compare the estimation of the robustness with the robustness of an empty model ($M = 0$).

5.1.2 Simplify costs of one line and access the number of edges and nodes

The variable cost is based on a fixed amount K , the number of stations s and metro parts p and the number of metro parts which have to be built over each other. When one metro line is built, the costs of building over each other will always be zero; otherwise it would contradict the rules of metro lines (section 2.2). And a metro system with one metro line doesn't have any transfer stations. If we take the model with $k = 1$, all metro parts have a length of 1. With these assumptions the costs variable can be simplified: model with one metro line, $k = 1$:

$$C = K + p + 2 \cdot s$$

With this formula, it is simple to determine the number of edges and nodes for the best design under a specified number of costs. There are two options for the number of stations s and metro parts p : the metro line with $s \neq e$ and the bypass metro line with $s = e$.

$$(s, p) = \begin{cases} s = \min \left[\left\lceil \frac{(C - K) + 1}{3} \right\rceil, n \right], p = s - 1 & \vee s = \min \left[\left\lceil \frac{C - K}{3} \right\rceil, n \right], p = s, \text{ when } C - K \geq 5 \\ s = 0, p = 0, & \text{ when } C - K \leq 5 \end{cases}$$

This formula consists of two parts $C - K \geq 5$ and $C - K \leq 5$: when the available amount of money (equals costs C) is less than the fixed costs K , there are not enough resources to build a metro line at all.

If the constraint $C - K \geq 5$ is valid, the formula gives two solutions in number of edges and number of nodes for the best metro line corresponding to metro line with $s = e$ and $s \neq e$.

$$s \neq e: s = \min \left[\left\lceil \frac{(C - K) + 1}{3} \right\rceil, n \right], p = s - 1$$

$$s = e: s = \min \left[\left\lceil \frac{C - K}{3} \right\rceil, n \right], p = s$$

The formula is calculating the number of pairs consisting of a node and an edge which can be paid by the amount $C - K$ for both models. The formula can give a solution in which no design can be found, because not for every combination of stations/parts a metro bypass can be found. Although, the metro line with $s \neq e$ will always give a suitable solution.

With this formula, the best network design can be found in two kinds of network structures, the metro bypass with s stations and s metro parts. And the metro line with s stations and $s - 1$ metroparts. This approach is reducing complexity of finding a suitable model to $O(|V| \log |V|)$.

5.2 Approach by edge betweenness to find a reasonably good design

In section 4.1 is a method given to estimate the number of edges and nodes of the highest robustness design. In this paragraph an approach for finding the best design is given based on the estimation of section 4.1 and the graph measure betweenness.

5.2.1 Approach based on edge betweenness

When knowing how many stations and metro parts the design needs to have for the best robustness, the question left is concerned with the place where they should be placed. Instinctively we should place the metro line to connect the 'most important parts of an area'. In graph theory a measure is designed to quantify the importance of a network link in a graph: *the edge betweenness*. The betweenness denotes the number of shortest paths passing through a particular edge. The betweenness can give information about the most 'important links' which can be underpinned by a metro line. The edge betweenness of all edges can be calculated in $O(|V||E|)$.

When determining the edge betweenness of all edges in graph G , a ranking arises between all the edges. This ranking can be used to place metro parts on important links in the network. The Algorithm of Edge Betweenness Design is given below:

Step 1: Estimate the number of metro parts p and stations s based on a fixed available investment C . Complexity: $O(1)$

Step 2: Calculate all *edge betweenness* values. Complexity: $O(|V||E|) = O(n \cdot m)$

Step 3: Take the p parts with the highest edge betweenness value. Complexity: $O(1)$

LOOP (If network arisen is not a valid metro line) Complexity: Best case: $O(1)$, worst case $O(|V||E|)$

Step 4: Take the p parts with the highest edge betweenness value if the previous design(s) are not accessible.

END LOOP

5.2.2 Error measurement

This approach gives a good approximation of a suitable design. Based on the edge betweenness design, the design can be ranked from best to worse. This same ranking can be done based on the robustness. Using ranking statistics the ranking error can be calculated:

$$e_B = \frac{\sqrt{\sum_{i=1}^{all\ designs} \left[\frac{r_{R_i} * 100}{|r_R|} - \frac{r_{B_i} * 100}{|r_B|} \right]^2}}{total\ designs}$$

With r_R the rank of the design by the effective resistance (true outcome) and r_B the rank of the design by the betweenness measure. The e_B for the 9nodes model is equal to 0.1753 rank, where the ranks are scales in a $[1, \dots, 100]$ scale.

5.2.3 Limitations

This approach is developed to get an acceptable design where high robustness is pursued. This method also has some limitations. These will be explained in this section.

1. The most important links in the underlying graph are taken into account. This means the graph M cannot adopt links, which do not exist in graph F . An assumption to overcome this problem, is taking the most important links of graph M . But in doing so, the advantage of knowledge about the current road area disappears. This can be shown by a river with two bridges in an area. Taking the most important links in graph F will lead to the links equal to the bridges. Taking the graph M , where the river can be bypassed everywhere, the algorithm will possibly take the middle line, whether there is a bridge or not.

2. Limitation (1) leads to the fact the model is only meaningful for a square area with no bulges. And those areas are precisely the areas where intuition to make the best position of a subway line is easier. The assumption to make every area square, is losing the advantage of the availability of the area without the metro line; see objections limitation (1).
3. The process of finding a valid metro line can be long, when the graph goes out of proportions and when the available investments grow.

5.2.4. Adding weights of importance to the model

The approach in this section is based on the importance of certain links and indirectly the number of shortest paths through a link. In this model the importance of particular connections on other connections cannot be taken into account easily. The shortest paths counted can multiplied by a weight value of importance, but this value should be deduced on the nodes s and e . This modified method can send the model towards a model with a higher robustness, but the limitations of section 5.2.2 will still be present.

5.3 Approach by node betweenness to find a reasonably good design

Section 5.2 is focused on a model based on the edge betweenness. The model could even focus on the node betweenness. In this case not the links, but the particular nodes can be taken into account. This model will have some advantages of the previous model, discussed in the following subsections. The node betweenness is also a classical network measure.

5.2.1 Approach based on the node betweenness

The measure node betweenness originated from almost the same conditions as the edge betweenness. In graph theory a measure is designed to quantify the importance of a network node in a graph: *the node betweenness*. The betweenness denotes the number of shortest paths passing through a particular node. The betweenness can give information about the most 'important nodes' which can be underpinned by a metro line. The node betweenness of all nodes can be calculated in $O(|V||E|)$.

When determining the node betweenness of all edges in graph G , a ranking arises between all nodes. This ranking can be used to place metro parts between important nodes in the network. The Algorithm of Node Betweenness Design is given below:

Step 1: Estimate the number of metro parts p and stations s based on a fixed available investment C . Complexity: $O(1)$

Step 2: Calculate all *node betweenness* values. Complexity: $O(|V||E|) = O(n \cdot m)$

Step 3: Take the s stations with the highest node betweenness value. Complexity: $O(1)$

Step 4: Draw a connection between the captured nodes. Complexity: $O(1)$

LOOP (If network arisen is not a valid metro line) Complexity: Best case: $O(1)$, worst case $O(|V||E|)$

Step 5: Take the s stations with the highest node betweenness value and draw a connection between the captured nodes. If the previous design(s) are not accessible.

END LOOP

5.2.2 Error measurement

This approach gives a good appropriation of a suitable design. Based on the NODE betweenness design, the design can be ranked from best to worse. This same ranking can be done based on the robustness. Using ranking statistics the ranking error can be calculated:

$$e_B = \frac{\sqrt{\sum_{i=1}^{\text{all designs}} \left[\frac{r_{R_i} * 100}{|r_R|} - \frac{r_{B_i} * 100}{|r_B|} \right]^2}}{\text{total designs}}$$

With r_R the rank of the design by the effective resistance (true outcome) and r_B the rank of the design by the betweenness measure. The e_B for the 9-nodes model equal to 0.2064 rank, where the ranks are scales in a $[1, \dots, 100]$ scale.

5.2.3 Limitations

This approach is developed to get an acceptable design where high robustness is pursued. This method also has some limitations. These will be explained in this section.

4. The limitations (1) and (2) from section 5.2.3 are still present in the model based on node betweenness. Although the results indicate that the node betweenness is less susceptible to this effect than the edge betweenness.
5. The process of finding a valid metro line can be long, when the graph goes out of proportions and when the available investments are growing. An advantage of the model by node betweenness above edge betweenness is the fact there are less nodes than edges (especially when k is high). The change to find a valid metro line by connecting nodes will be higher than by connecting edges.

5.3.4. Adding weights of importance to the model

The model can adjust weights to the model based on node betweenness easier than the model based on edge betweenness. The weights can be assigned to particular nodes. When multiplying the betweenness values of a particular node by the weights, the subdivision of nodes can be taken into account.

5.4 Approach by travel times to find a reasonably good design

In chapter 4 a remarkable finding was that the ranking of models that were found when the travel time is minimized, yielded a similar ranking. The travel time will give an approximately robust design. This has to do with the fact that the development of a subway line (adding a link) will lead to a higher robustness. Just like adding a link will reduce the travel time. Although an algorithm to find the model that minimize the travel time is of the same complexity as finding a model with the highest robustness. An approximation of such problem will lead to the same research as done for maximizing robustness.

5.5 Approach by minimizing effective resistance

W. Ellens (2011) has done much research about maximizing robustness in graphs by minimizing the effective resistance. She defines two heuristics to maximize robustness mathematically, but with both methods the error increases when the number of nodes is increasing. In the first heuristic the Laplace eigenvalues are minimized. In the second method clique chains are used to optimize a graph with a given number of nodes and edges. Finally Ellens concludes this paper gives many insights in maximizing robustness, but this paper doesn't give a clear solution, especially for large graphs.

5.6 Summary of models

In this paragraph we are trying to find an algorithm of placing one metro line in an area to maximize robustness, and starts with the development of a model, which tries every possible line in the area to compare robustness of all designs. This model knows a complexity of at least $O(n^4 \cdot n!)$ but possible worse. This complexity can be reduced by taking the betweenness of the underlying graph and find by greedy manner the path with the highest betweenness for an estimated number of edges and nodes. This reduces the problem to a complexity of at least $O(nm)$ and in worst case $O(n^2m^2)$.

Chapter 6

Method in practice

This chapter shows the application of the model of the area Amsterdam. The model used is a replica of the model described in section 2 and because of the complexity of the models as described in Chapter 5, not all possibilities in the area will be considered.

6.1 Case study

Amsterdam is the capital city of the Netherlands and has currently an operating metro system with four lines. The model described in chapter 2 till 4 is used to find the answer to the question: *which additional metro line should be placed to maximize robustness?* The initial model is based on the graph F and N . The additional metro line will be a sub graph of M .

6.1.1. Area of Amsterdam

The captured area of Amsterdam is shown in Figure 47 as graph F and based on a $k = 1$ model. The area of Amsterdam is not a symmetrical area. It is divided in nine districts: *Noord*, *Zuid*, *Oost*, *West*, *Nieuw-West*, *Amstelveen*, *Diemen*, *Zuidoost* and *Centrum*. These districts are captured in the graph as different colors. To incorporate the importance of a specified link in the graph, the population density of the districts is used as weights of importance. These weights are used in both, the calculation of the robustness and the travel time.

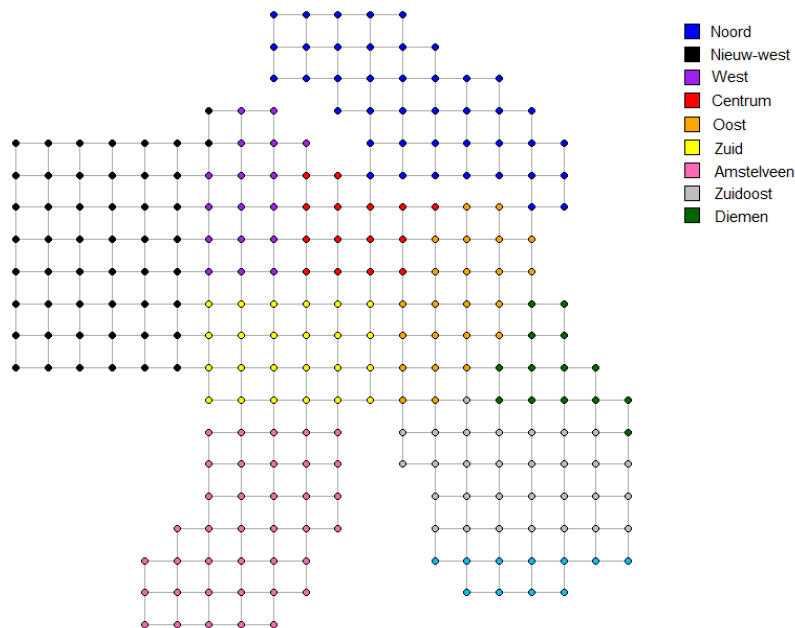


Fig. 47: Graph of the area of Amsterdam

The weights used are summarized in the table below. All weights are calculated by population density of the neighborhood divided by 1,000. This is done to keep the weights small without losing essential information. Because the Centrum area knows a low population density (1.351), this density is multiplied by 4 to incorporate the need of facilities in this area.

Neighborhood	Weight w_i	Neighborhood	Weight w_i
Centrum	5.404	Amstelveen	2.011
Noord	2.079	Zuid-Oost	4.082
West	15.477	Oost	6.917
Nieuw-West	4.819	Diemen	2.000
Zuid	8.814		

This initial area without any metro system can be taken as zero point to express the effect of existence of the metro network in Amsterdam on the robustness. Based on graph F the weighted average travel time is 78.01 time units and the robustness 22,305.34 (effective resistance). In this initial model all public transportation is hereby excluded.

6.1.2 The current metro system

The current metro system is shown in Figure 48 and the graph representation in Figure 49. The robustness of the area with the metro network is 20,671.64. This is an improvement of more than 7% to the whole area of Amsterdam.



Fig. 48: Metro network in Amsterdam

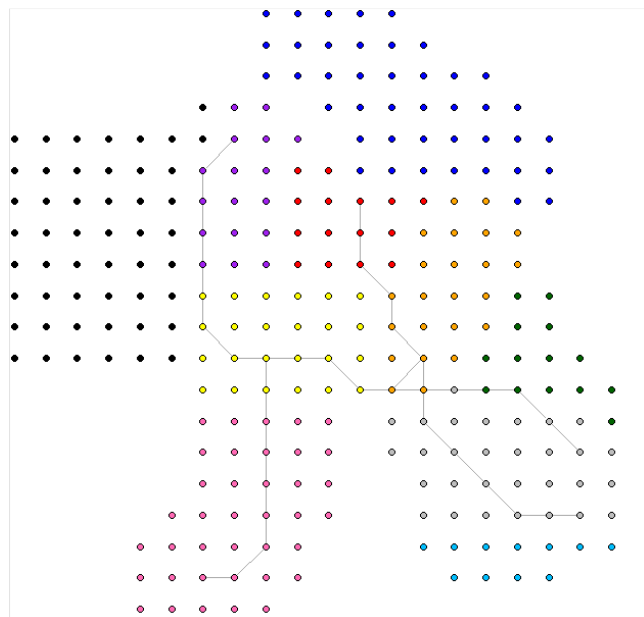


Fig. 49: Graph with metro network of Amsterdam

6.1.3 Additional metro line

In this case study the current metro network is extended by an additional metro line. The maximum ratio of connection two metro stations is $k = \sqrt{2}$. The following costs function specified:

$$15 \geq 5 + |k_1| + \sqrt{2}|k_2| + s_m + s_e + s_a$$

The variable s_a is kept as number of existing stations with an additional platform built in this case study. For this expansion a specified amount of money is available to cover the additional costs K and the development of metro parts with a maximum length of 5 and construction/expanding of 5 metro stations. Crossing over an existing metro line is disregarded.

6.2 Approach of finding an additional metro line

The process of finding all possible metro lines under the specified costs limit and obtain all robustness values of those networks is based on complexity impossible. In chapter 5 some alternative solutions are given, but all the alternative methods have restrictive assumptions. In this case study of the area of Amsterdam the view of people on the best models and the advantages of the robustness model are combined. Fifty metro lines are determined as possible additional metro line in subject to costs function. All those fifty metro lines are used as input of the model and only those models are compared and determined.

6.3 Metro line with the highest robustness

The model with the highest robustness is the model shown in Figure 50. Besides the model with the highest robustness also other similar models have been tried, but none with robustness equal to the final.

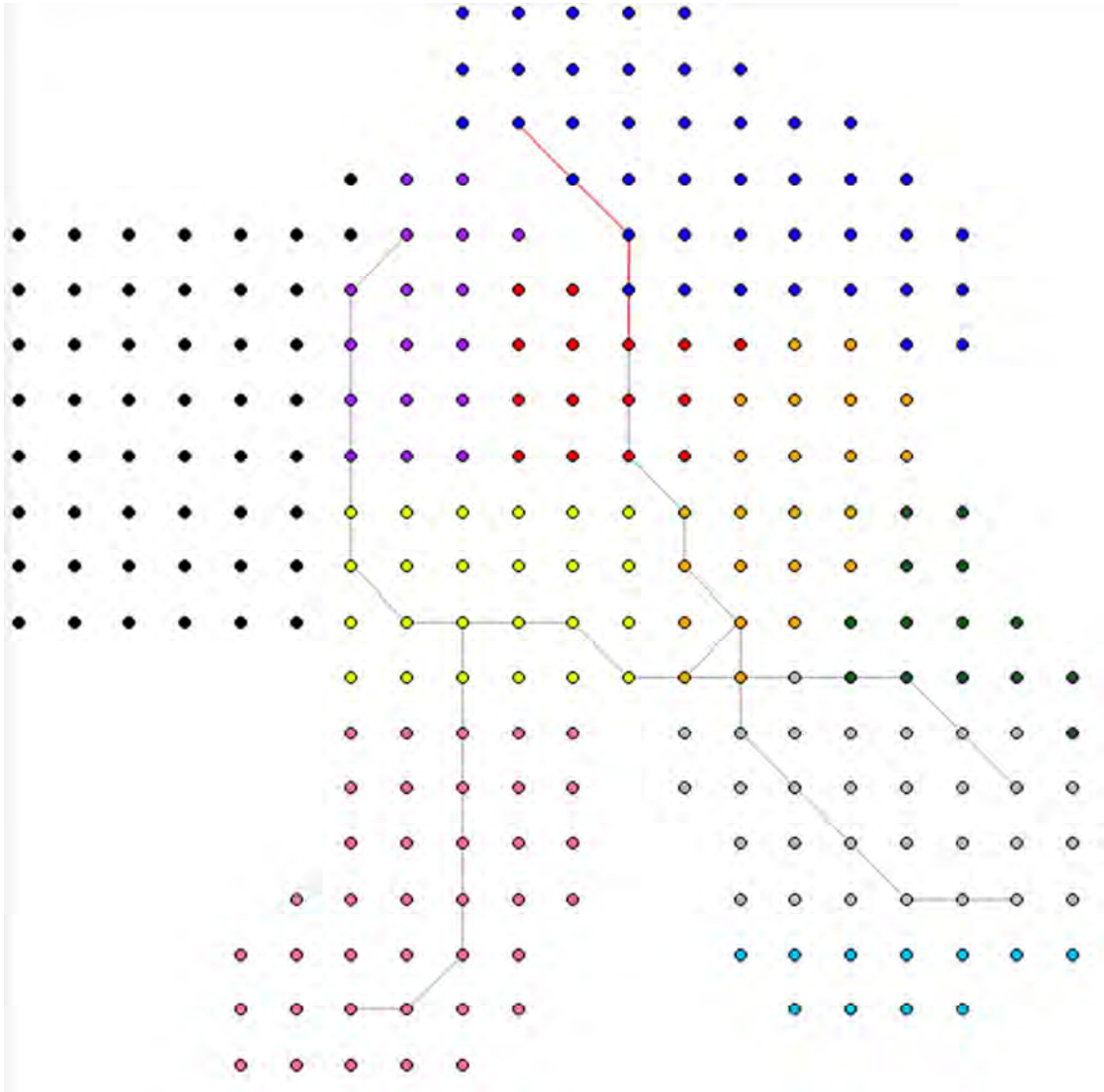


Fig. 50: Amsterdam metro network with additional metro line with the highest robustness

The fact that the system is chooses for a metro adjustment in the neighborhood *Noord* is plausible. The neighborhood *Noord* is only available by two ways: a tunnel/boat and a bridge. With the development of a metro line to the northern part the robustness will see a considerable positive improvement: almost 6% to the existing metro network.

6.4 Conclusion

The case of Amsterdam shows that the detailed way of modeling and the possibility of the robustness measure by the model can distinguish almost equal graphs. None of the 50 defined metro lines has equal robustness values. Offering a large number of possible models will give the model with the highest robustness inside the group of offering networks, with certainty. Finding the design with the highest robustness in all possible designs is impossible because of the greatness of the model.

Chapter 7

Conclusion

A method is proposed to find the metro network with the highest robustness by three steps: Firstly the modeling part of translating reality to an applicable model. Second is the calculation of robustness. And finally the algorithm of finding the best additional metro line under certain constraints.

The model of the first step is flexible and applicable to most areas in reality. Although some variables should be chosen on forehand: the connectivity of the point in the network (value of k), the distance between the different points of the grid and the position of the boundaries of the incorporated area. The effect of the choice of variables is not determined in this paper. In the examples throughout the paper these variables are chosen on practical insight. Further research can be done about the effects of those variables.

The method used has a highly detailed and reliable way of measuring robustness. This advantage can be used on different designs of metro networks. Metro networks which seem incomparable at first sight, can easily be distinguished by the model. Even in a complex network such as Amsterdam, small adjustments on metro lines can be recognized by the robustness indicator.

The given method is capable of always finding the metro network with the highest robustness under a specific costs function. Although this method is perfectly useful for finding a highly robust network, the complexity is increasing exponential with respect to to the number of nodes. A search for a heuristic to reduce complexity can be found in the betweenness heuristic or by methods to maximize travel time or the robustness measure. Although this heuristic gives an approximate good design, the assumptions are very strict and worsen the applicability of the model. Further research should be done about a better heuristic in finding computationally an acceptable result.

Because of the detailed way of modeling reality and the possibility of the robustness measure to distinguish almost equal graphs, the method can be used best in the original method with manual offering metro networks, which can be tested by the robustness model. Offering a large number of possible models will give the model with the highest robustness inside the group of offering networks, with certainty. This method has been used in a case study on the Amsterdam metro network.

Bibliography

Costa, L., Rodrigues, F., Travieso, G., & Villas Boas, P. (2007), Characterization of complex networks: A survey of measurements, In *Advances in Physics* (pp. 167-242)

Derrible, S., & Kennedy, C. (2010). The complexity and robustness of metro networks. In *Physics A* (pp. 3678-3691).

Ellens, W., & Kooij, R.E. (2011). Graph measures and network robustness (p1).

Ellens, W., Spijksma, F.M, Mieghem van, P., Jamakovic, A., & Kooij, R.E. (2011), Effective graph resistance. In *Linear Algebra and its applications 435* (pp. 2491-2506)

Holme, P., Kim, B.J., Yoon, C.N., & Han, S.K. (2002). Attack vulnerability of complex networks. *Physical Review E*, 5(65), 56109

van der Meer, E. (2012). Comparing measures of network robustness.

Scott, D.M., Novak, D.C., Aultman-Hall, L., & Guo, F. (2006). Network Robustness Index: A new method for identifying critical links and evaluating the performance of transportation networks.

Snelder, M., van Zuylen, H.J., & Immers, L.H., (2012). A framework for robustness analysis of road networks for short term variations in supply. In *Transportation Research Part A* (pp. 828-842).

Yingying, D., & Feng, L. (2013). Structural robustness of city road networks and based on community. In *Computers, Environment and urban Systems 41* (pp. 75 -87).