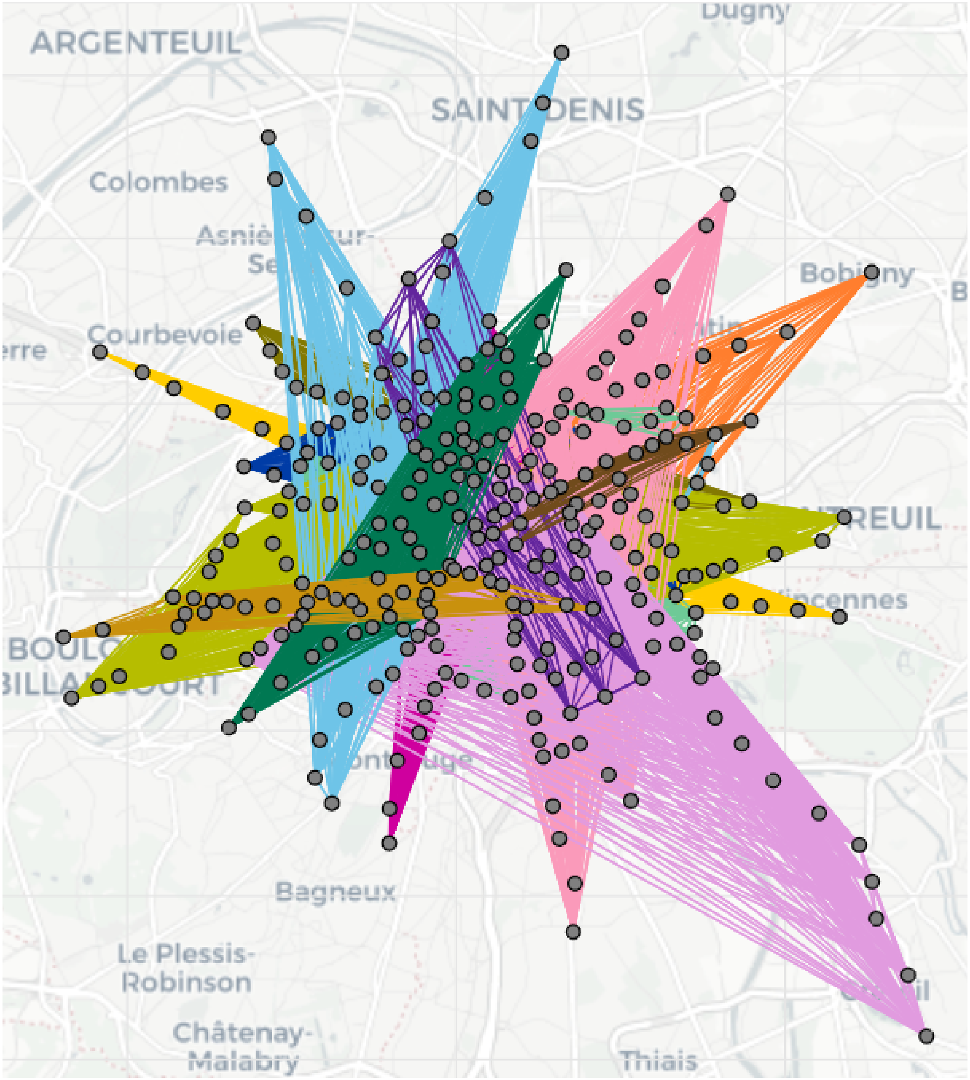
Including service information in a topological comparison of metro networks worldwide



A comparison of 51 metro networks worldwide using GTFS static data

Sam Vijlbrief

INCLUDING SERVICE INFORMATION IN A TOPOLOGICAL COMPARISON OF METRO NETWORKS WORLDWIDE

A COMPARISON OF 51 METRO NETWORKS WORLDWIDE USING GTFS STATIC DATA

by

Sam VIJLBRIEF

to obtain the degree of Master of Science in
Transport, Infrastructure, & Logistics
at Delft University of Technology
to be defended publicly on Friday October 28th, 2022

Thesis committee members:

Dr. O. (Oded) Cats,	chair
Dr. ir. P.K. (Panchamy) Krishnakumari,	daily supervisor
Dr. ir. S. (Sander) van Cranenburgh,	supervisor
Dr. ir. R. (Renzo) Massobrio,	supervisor

An electronic version of this thesis will be available at
http://repository.tudelft.nl/.

Contents

Abstract **vii**

Preface **ix**

1 Introduction **1**
- 1.1 Scientific context . . . 1
- 1.2 Problem description . . . 2
- 1.3 Solution direction . . . 2
- 1.4 Research questions . . . 3
- 1.5 Document structure . . . 3

2 Literature Study **5**
- 2.1 Scoping . . . 5
- 2.2 Bibliometric search strategy . . . 6
- 2.3 Related work . . . 6
  - 2.3.1 PT network topology . . . 7
  - 2.3.2 Service information . . . 10
  - 2.3.3 PT comparison studies . . . 11
  - 2.3.4 Summary table . . . 12
- 2.4 Research gap . . . 13
  - 2.4.1 Service information as a quantitative measure . . . 13
  - 2.4.2 Number of networks included . . . 14
  - 2.4.3 Unexplored networks and regional differences . . . 14

3 Methodology **17**
- 3.1 Network selection . . . 17
  - 3.1.1 Definition of metro . . . 17
  - 3.1.2 Overview of potential networks . . . 18
- 3.2 PT network representation . . . 19
  - 3.2.1 Representation options . . . 19
  - 3.2.2 L-space . . . 20
  - 3.2.3 P-space . . . 22
- 3.3 Primary metric . . . 24
  - 3.3.1 Accessibility . . . 24
  - 3.3.2 Metric definition . . . 25
- 3.4 Network indicators . . . 26
  - 3.4.1 Number of stations . . . 26
  - 3.4.2 Network length . . . 27
  - 3.4.3 Number of lines . . . 27
  - 3.4.4 Direct station distance . . . 28

3.5 City indicators . . . 29
3.5.1 Country/Region . . . 29
3.5.2 Population . . . 29
3.5.3 GDP . . . 29

4 Implementation 31
4.1 Networks included . . . 31
4.1.1 Data format . . . 31
4.1.2 Data retrieval . . . 32
4.1.3 Final set of networks . . . 33
4.2 Data pipeline . . . 34
4.3 Implementation of L- and P-space . . . 36
4.3.1 L-space . . . 37
4.3.2 P-space . . . 41
4.4 Shortest path calculation . . . 44
4.4.1 Attribute coefficient values. . . . 44
4.4.2 Implementation . . . 45
4.5 Additional data . . . 47
4.5.1 Additional network information . . . 47
4.5.2 Population . . . 47
4.5.3 GDP . . . 47

5 Results 49
5.1 Primary indicator . . . 49
5.1.1 Representative value . . . 49
5.1.2 Exceptions . . . 51
5.2 Correlations between the indicators . . . 53
5.2.1 Spearman . . . 53
5.2.2 Pearson . . . 54
5.3 Simple linear regression. . . . 55
5.3.1 Outliers . . . 56
5.3.2 Regions . . . 59
5.4 Multiple linear regression . . . 60
5.4.1 Full model . . . 60
5.4.2 Improved model . . . 61
5.4.3 Conclusion. . . . 64
5.5 Further comparisons . . . 65
5.5.1 Number of lines . . . 65
5.5.2 Population . . . 66
5.5.3 GDP . . . 68
5.6 Travel time calculation benchmark analysis. . . . 69
5.6.1 Correlation between the methods . . . 69
5.6.2 Direct comparison between the methods . . . 70
5.6.3 Indirect comparison between the methods . . . 72
5.6.4 Conclusion. . . . 74

6 Conclusion **75**
6.1 Key findings . . . 75
6.2 Planning and policy implications . . . 77
6.3 Limitations . . . 77
6.4 Recommendations for further research . . . 78

A Scientific paper **85**

B Median travel time maps **99**

C General network information **101**

D Network data **103**

E Histograms **105**

F Processing **109**

G Additional data sources **119**

Abstract

Public transport (PT) plays a vital role in commuting billions of travellers in cities all over the world, providing a mode that is both sustainable and accessible. Metro networks are especially apt at this considering their high-capacity and high-speed operation in urban environments. Comparing different metro networks to one another is a suitable manner for transport planners to gain insights into the characteristics of their networks and which areas of improvement exist. In the field of network science, metro networks have been studied extensively in recent decades. While this provided many new insights in the field of network science, the practical relevance for the field of transport science often remained limited. This limited relevance is primarily caused by the lack of realism of the network representations used, not incorporating the actual operation and service that the network provides. As such, this study proposes a comprehensive comparison of metro networks worldwide including service information. This comparison study includes service characteristics in the form of the total travel time indicator for shortest path calculations, which is a combination of the in-vehicle time, waiting time and number of transfers. The median of this total travel time is taken for each network and compared to that of other networks. This metric in turn is contrasted with network- and city-related characteristics in order to explore relations between these factors and to explain the patterns discovered. From this analysis, it is revealed that the travel time increases with network size. The indicator that is especially apt at explaining the differences in total travel time between networks is the number of stations combined with the average direct station distance. The total travel time methodology applied in this study shows significantly different results to other commonly used methods that rely only on in-vehicle time or hops to calculate shortest path travel times. The waiting time turns out to be the main contributor to these significant differences. Future studies can expand on this by considering other network science indicators and looking further into local indicators. In addition, the methodology could be expanded with more detailed transfer information and other PT modes.

Preface

This thesis is my final project to attain the Master Transport, Infrastructure & Logistics (TIL). When I started my Bachelor Computer Science at TU Delft in 2016, I did not expect to finish my studies with a project about metro networks. Similarly, when starting my TIL Master in 2019, wanting do something less technical and more policy-related, I did not expect my final project to involve this much programming. In the end, it seems like this thesis is the perfect match between my love for modeling and passion for public transport. I must admit, that while I think the concept of my thesis is fairly straightforward and simple to explain, until the end I struggled with explaining what exactly my work involved and why the results would be useful. I feel that finally after eight months of hard work, I have the answers to those questions, which are extensively described in this report.

Throughout this process, it was my supervisors who helped guide me to these answers. Naturally, my first thanks go out to my committee chair, Oded Cats, who I approached with this topic in the first place. While I had a vague idea of what I wanted to do for my thesis, Oded helped solidify this into a clear topic and direction. Even though he was technically my chair and thus did not involve himself directly with my thesis except for some specific moments, he helped me like a daily supervisor. His huge amount of knowledge around the topic and eagerness to help me succeed, were a great help during my progress meetings. Secondly, I'd like to thank my daily supervisor Panchamy Krishnakumari. Especially at the beginning of the project when I was still searching for the right direction, our frequent meetings helped push me in the right direction and she was always open to help me if I was stuck. Thirdly, I'd like to thank my supervisor from TPM, Sander van Cranenburgh. As this is a TIL thesis, it had to incorporate both aspects from policy as well as civil engineering. Sander helped me approach my results from the policy perspective by continually asking the question "Who benefits from your research?". My fourth supervisor, Renzo Massobrio, joined my committee later when we were working together on the code. I can think of no better example of "last but certainly not least" than Renzo as my supervisor. I lack the words to describe the amount of support I have felt from Renzo during this project. While he joined initially since we were working on the same code base, it is the frequent discussions and detailed feedback on my work that really helped me forward and kept me motivated. My sincerest gratitude to my supervisors for all the help you have given me!

Finally, I have to thank all the others that supported me throughout this long and sometimes arduous process. I would like to especially thank the friends that I studied with together, who I could spar with when I was stuck or to just go for a walk to clear my mind: Laura, Eva, Marijn, Salomon, Schelte, Nao and Arjan. Thanks everybody!

Sam Vijlbrief
Delft, October 2022

1

Introduction

Public transport (PT) plays a vital role in commuting billions of travellers in cities all over the world. Considering the rapid threat of climate change, improving sustainability and equity are crucial factors to keep cities habitable. PT plays a crucial role in providing transport that is both sustainable as well as accessible. A variety of different modes to transport volumes of people throughout cities exist, such as bus, tram, metro and train. In an urban environment, metro networks have a variety of factors that make them especially attractive. In terms of operation, a metro is considerably high-capacity and high-speed compared to most other PT modes. In terms of infrastructure, their (primarily) underground nature makes their location highly flexible while also having a minimal impact on the above-ground urban infrastructure. Considering these advantages and the rapid advance of technology to construct metro networks, their popularity has increased greatly over the past decade. In the last decade, almost sixty new systems have opened, nearly a third of the total number of metro networks worldwide. On the other hand, existing networks also frequently expand, seeing as the total amount of metro network infrastructure has increased by 25%, or 3,300km in total, in the past three years alone (UITP, 2022).

The subsequent parts of this introductory chapter are structured as follows. Firstly, the scientific context is provided in Section 1.1. The problem is further described in Section 1.2 along with the proposed solution direction in Section 1.3. The accompanying research questions are provided in Section 1.4. Finally, this chapter is concluded in Section 1.5 with a description of the further structure of this document.

1.1. Scientific context

Metro networks have also received much attention in literature in the recent years. The field of complex network science has been using public transport systems as a field of application since the early 2000s (Latora and Marchiori, 2002; Sienkiewicz and Hołyst, 2005). PT systems as a complex network have been extensively explored around the year 2010 with the works of Derrible and Kennedy (Derrible and Kennedy, 2009; Derrible and

Kennedy, 2010a; Derrible and Kennedy, 2010b; Derrible and Kennedy, 2011; Derrible, 2012). These researchers noted how suitable PT networks were to investigate various concepts from complex network theory. While this provided many new insights in the field of network science, the practical relevance for the field of transport science often remained limited. This limited relevance is primarily caused by the lack of realism of the network representations used. Frequently, these studies used only simple hops or in-vehicle travel time to calculate travellers' paths through the network. In doing so, service concepts such as transfer possibilities and waiting time are completely ignored. Evidently, this can lead to a misrepresentation of networks.

PT scientists recognized this gap and tried to integrate concepts from transport science into network analyses. This led to a large variety of examples such as: creating a weighted graph with passenger flows (Xu et al., 2016); investigating the relationship between network topology and ridership (Ingvardson & Nielsen, 2018); or integrating network science and accessibility analysis (Luo et al., 2019). These studies, albeit limited in number and scope, provided more insights into the relationship between the structure of PT networks and their actual service and usage patterns, providing information for policy makers on the performance of their networks.

1.2. Problem description

With this field of combining network infrastructure and service information upcoming, various aspects are still left unexplored. For example, while studies as the aforementioned do contribute new methodologies in the field, these have not been extensively used to compare different networks yet. It is vital for PT planners to understand the performance of their network on these aspects compared to other networks, in order to learn where room for improvement exists within their networks. In addition, even when comparison studies are performed, the set of networks used is usually similar: a set of a few dozen, large-sized metro networks. Logically, using only the same few dozen networks means a significant portion of the 190 metro networks worldwide (UITP, 2022) is currently not studied. Additionally, as described before, the number of networks has greatly increased in recent years, meaning many recent networks and expansions are also still unexplored. Considering previous studies primarily use large-sized networks, the effects on small- and mid-sized networks are also unclear. It can thus be concluded that a comprehensive comparison of a large set of varied networks is currently missing.

1.3. Solution direction

To fill these gaps, this study proposes a comprehensive, topological comparison study of metro networks worldwide that includes service information. This study includes service characteristics mathematically in the topological representations of the networks in order to provide a more realistic representation of these networks. An indicator is used that incorporates service information to directly compare networks to each other. This metric in turn is contrasted with network- and city-related characteristics in order to explore relations between these factors and to explain the patterns discovered. Specific attention is also paid towards outliers and regional differences in order to discover what potential patterns arise.

1.4. RESEARCH QUESTIONS

Based on the research gaps as described above, the question this study aims to answer can best be formulated as follows:

How can service information be included into a topological comparison of metro networks worldwide?

This research question is supported by the following sub-questions that provide more detail on the possible methodology and results:

- What indicator can best be used to account for service information?
- What type of data pipeline should be created to turn service information data into comparable metrics?
- Which regional differences and outliers are identified when examining networks' total travel time and their components?
- Can (a combination) of network and city factors be used to meaningfully explain the travel time of a network?
- Does including service information provide significantly different results to commonly used methods?

1.5. DOCUMENT STRUCTURE

The rest of this report is structured as follows: Firstly, the current state of the literature is reviewed in Chapter 2. In Chapter 3, the methodology for answering the research questions is further detailed. Chapter 4 details the modelling implementation for this study. The results are reported in Chapter 5. Finally, the conclusions and recommendations of this study are detailed in Chapter 6.

2

Literature Study

In this chapter, an extensive literature study is performed in order to identify the current state-of-the-art of PT network topology research. The aim is to identify what research has already been performed and what gap still exists in the literature, especially in terms of including service information. Firstly, a brief overview of the field of study is provided in Section 2.1. Secondly, the strategy applied for finding related work is further detailed in Section 2.2. Thirdly, the related works for each topic are outlined in Section 2.3. Finally, this chapter is concluded with the identification of the research gaps in Section 2.4.

2.1. Scoping

In order to perform a directed search within the literature, first the field of study needs to be narrowed down. The main topic of this study (and consequently of this literature review) is "including service information in a topological comparison of metro networks worldwide". This title also provides a natural direction for the topics of the literature review: service information for metro networks and topological comparisons of metro networks. For this literature study, the topic is expanded slightly to include all types of PT, not merely metro networks. In addition to that, there are some studies that do focus on the comparison of PT networks but that do not clearly fall into either of these two categories. Therefore, a third topic of interest for this literature study is "general PT comparison studies". A such, the following three topics are researched:

- **PT network topological studies** - This topic is focused on all studies that research PT networks from a topological perspective. These primarily consist of empirical and statistical studies into different graph theory concepts such as structure, hierarchy and topology. Works in this category specifically focus on the topology without including service information.
- **PT service information studies** - This topic is about analyzing PT networks based on their service characteristics. This includes studies into the accessibility or equity of networks as well as into the transfers, waiting times and frequency of the

service. These studies generally perform this analysis from a topological perspective, but some exceptions do exist.

- **General PT comparison studies** - The focus of this topic primarily lies in the actual comparison between different PT networks. Literature in this area can be quite diverse in their focus and also can include grey literature. Literature in this section explicitly does not incorporate network topology.

The various works identified for each of these categories are presented in Section 2.3.

2.2. BIBLIOMETRIC SEARCH STRATEGY

In this section, the search strategy applied for this literature study is briefly described. The aim of this literature review is to find the knowledge gap within the field of topological metro network comparisons incorporating service information. The general topic and corresponding subtopics were already described extensively in Section 2.1. As such, the literature review is performed with these topics in mind. The following methods are applied to find relevant works:

- **Search in academic databases** - Different search terms are applied to academic databases. The primary database used for this review is Scopus. The following search terms are applied:
  - "public transport" and "network" and "topology" (139 results on Scopus)
  - "public transport" and "network" and "comparison" (308 results on Scopus)
  - "public transport" and "service frequency" or "service information" (140 results on Scopus)

  Based on these search terms, a large variety of works is found, as is indicated above. Naturally, considering the overlap in the search terms, there is also some overlap in the works discovered by the three search terms. These works are then filtered based on their relevance for this study.

- **Recommendations by coordinators** - The coordinators of this project have provided several relevant studies as a baseline. These are also included in this literature study.

- **Forward and backward snowballing** - The techniques of both forward and backward snowballing are applied to the most relevant works. This is used especially on works that have a close relation with the aim of this study.

After all of these steps, 33 related works were retained which will be described in more detail in Section 2.3.

2.3. RELATED WORK

In this section, the related work is discussed. This section is split up into the three corresponding subtopics as described in Section 2.1.

2.3.1. PT NETWORK TOPOLOGY

The first of three topics to discuss, PT network topology, is arguably the most significant and comprehensive one. This significance becomes clear from the fact that it has been an interest of study for several decades and is continually expanding with new findings and applications. The network topology of PT networks has been of interest for transport scientists and network scientists alike, giving the field a broad variety of studies from different directions. This section largely follows a chronological order as that best describes the changes and development in the field. The most relevant studies are explained in more detail and findings from them that are relevant for this study are briefly highlighted.

EARLY WORKS

One of the earliest examples of literature on network infrastructure is by Musso and Vuchic (1988). In this study, the authors aim to identify the most important geometric characteristics of metro networks. The work is surprisingly extensive for its time providing a set of five different categories by which to quantitatively describe networks. These categories are: network size and form; network topology; relationship between network and city; quantity and quality of offered service; and service use. While the latter two categories (focusing on service characteristics) are mentioned in the study, they are only provided as a guideline and not actually used for comparison in the study itself. In terms of scope, the study includes ten of the most significant metro networks of that time. The study also focuses quite heavily on defining metro lines and systems by their geometry, placing them into specific categories. This study can be regarded as the foundation on which many future studies expanded in the field of network structure analysis.

Studies continuing and expanding on this work arose especially in the 2000s, when different studies were performed on various locations and PT modes around the world. One of the earliest examples is the work by Latora and Marchiori (2002). In their work, the authors aim to bridge the gap between the theoretical paradigms of complex network science and real complex networks. The Boston subway was chosen as the network of application. While this work was one of the first seeking a connection between PT networks and complex network science, the focus was still very much on network science itself. A continuation of this work came three years later in the highly cited study by Sienkiewicz and Hołyst (2005) analyzing bus systems in Poland. This study focuses heavily on the application of complex network science concepts on a real-world network. As such, the results themselves are simply a definition of these networks according to various topological concepts (such as degree distribution, path length properties or betweenness). Unlike Musso and Vuchic (1988) in their work, the authors do not focus on the transport characteristics of these bus networks at all, but simply treat them as complex networks.

A very similar analysis was performed on the Chinese rail network (Li & Cai, 2007) which confirmed that railway network to be scale-free. Majima et al. (2007) instead took a more transport-related approach and performed a complex network analysis to evaluate the potential of a waterbus in the Tokyo area when combined with other modes. Researchers from Korea performed a similar study applied to the subway system of Seoul in 2008 (Lee et al., 2008). Their study focused on the analysis of statistical and topological properties and also included passenger flow.

CHARACTERIZING PT NETWORKS USING NETWORK SCIENCE INDICATORS

Arguably the most groundbreaking work in the field of PT network topology came up around the year 2010 with the works of Von Ferber et al. and Derrible & Kennedy. These authors took the field to a new level with extensive studies on various network aspects of public transport networks worldwide. One of the earlier works is by Von Ferber et al. (2009). The goal of their work is twofold: define public transport networks based on statistical properties; and build a model that can create a network that would reproduce these properties based on a few simple rules. These statistical properties were identified for fourteen public transport networks. The study identified some unexpected similarities and differences between networks and created a model aimed to capture those rules.

A work from the same year albeit with a slightly different focus came from Derrible and Kennedy (2009). In their work, the authors analyze nineteen metro networks worldwide, showing a relationship between network design and ridership. Network design was modeled using three indicators: coverage, directness and connectivity. The study shows a strong correlation between these aspects and the ridership of the metro networks, suggesting the effect of network design on ridership is significant. While this study does not specifically consider service characteristics, it does consider more than just the infrastructure. Both directness and connectivity relate to the routes and number of transfers (which technically are choices by the operator) and the ridership is included in the analysis. The study was expanded on by the authors in the following years with three new studies: Derrible and Kennedy (2010a), Derrible and Kennedy (2010b) and Derrible (2012)

Derrible and Kennedy (2010a) similarly to Derrible and Kennedy (2009) try to characterize metro networks using indicators. This work specifically focuses on adapting graph theory concepts into well defined public transport-specific applications and use this method to characterize networks. The three dimensions used in this study are: state, form and structure. These dimensions in turn consist of several indicators each. While the study does contain 33 different metro networks, these are merely used as examples of the three concepts and are not compared and contrasted themselves.

In a similar work from the same year (Derrible & Kennedy, 2010b), the authors aim to address topological networks in complex network science using the same data set of 33 networks. The interest here lies in the complexity of the networks and the effects thereof on the robustness. The networks in this study are not actually compared to each other nor does the study consider service characteristics.

The third of the works by Derrible on network science in metro networks specifically focuses on network centrality (Derrible, 2012). An alternative graph representation using only terminal and transfer nodes was used to compare 28 metro networks on their betweenness centrality.

While thematically similar, the work by Derrible and Kennedy (2011) takes a different approach and instead reviews the existing literature on applying graph theory and network science to the field of transit network design. At the time, this application was still fairly new and unexplored but has since seen a huge increase in popularity. The study provides a useful overview of the indicators found in other studies that could be applied to public transport networks. A similar review was created by Lin and Ban (2013). The main difference is that this work by Lin and Ban (2013) reviewed all transport networks, including for example aviation and maritime transportation networks as well.

The work by Von Ferber et al. (2009) and the consequent works by Derrible and Kennedy kick-started the increase in popularity of public transport network analysis. The works following this, can be generally split into two categories: complex network theory and PT network analysis. Whereas the complex network theory works merely use PT networks as a field of application, the PT network analysis works aim to use network theory to learn more about PT networks themselves. The developments in the two fields are described in the corresponding subsections below.

COMPLEX NETWORK THEORY

Part of this increase in popularity can be attributed to the nature of public transport networks. The aforementioned studies showed that PT networks (and especially metro networks) are very well-defined and frequently show useful network characteristics. PT networks are for example usually planar by nature and compared to some other real-life networks (such as social networks) have a relatively small size and can thus be visually inspected. In addition, these networks seem to frequently show the aforementioned small-world or scale-free properties in a real-world situation. This has led to many network science researchers to investigate these and other properties in networks in their local areas.

One of the earlier works in this era is by Zhang et al. (2013). In their work, the authors investigate the topological characteristics of urban rail transit networks to discover any universal patterns in them. They do this based on a variety of aspects such as betweenness, degree, shortest path and network failures. Xu et al. (2013) performed similar research for the PT networks in 330 Chinese cities.

Instead of applying known concepts to new regions, some authors instead aim to develop new methods to apply to complex network analysis in the PT application. Dimitrov and Ceder (2016) create a new methodology for examining topological characteristics of PT networks. In addition, the authors also focused on the data extraction process and solving operational tasks in practice.

Shanmukhappa et al. (2018) similarly apply a new method to the bus networks of London, Hong Kong and Bengaluru. Their method uses a supernode representation and end-to-end travel delay to evaluate the topological efficiency of these networks.

Wei et al. (2019) examine the Chinese high-speed rail network to investigate its network structure. Using metrics similar to Derrible and Kennedy (2009), they discovered that the network shows a hierarchical structure as opposed to the commonly-found scale-free structure.

While PT networks are often treated as unimodal networks, attempts have been made to create integrated networks as well. An example of this is by Hong et al. (2019) in which they create such a network for the metropolis of Seoul.

PT NETWORK ANALYSIS

The interest in PT networks from a complex network perspective is not just of interest for complex network scientists, but for transport scientists as well. Transport scientists use the concepts from network science to compare networks to each other or to better understand specific networks.

A good example of this approach is the work by Haznagy et al. (2015) where the urban PT systems of five Hungarian cities are analyzed. In their work, they find interesting similarities even in cities with varying morphologies. An example of this is the existence of a few high-degree nodes and an abundance of low-degree nodes. Their discoveries of this independence within Hungary are supported by different researchers in other regions such as China (Xu et al., 2013) and Poland (Sienkiewicz & Hołyst, 2005) suggesting that these findings apply globally.

Wu et al. (2017) compare six of the world's biggest metro networks using a new centrality measure called "node occupying probability". The study finds that these networks perform better under random attack than targeted attack and that some variation exists between these networks on the performance of these measures.

Shanmukhappa et al. (2019) similarly to Derrible and Kennedy (2011) and Lin and Ban (2013) review the current developments in PT network analysis in 2019. Based on the literature, they drew a variety of conclusions on the state-of-the-art in the field. Especially relevant for this study were their conclusions on metro network studies. The main conclusion was that undirected and weighted graphs are best suited for the nature of metro networks. In addition, they also recommended performing a more in-depth analysis of PT networks to get proper insights, as a simple topological analysis of the graph simplification of a network does not provide many practical insights.

2.3.2. SERVICE INFORMATION

The second topic of interest is that of service information. This field has been upcoming for the past few years where the network theory, as described in Section 2.3.1, is combined with the actual service or operation of a PT network. The studies in this field are quite varied in their nature and approach as the topic can be addressed from many different sides. The field can be roughly divided up into two categories: operation/ridership and accessibility. As such, studies divided into those categories are discussed in the corresponding sections below.

OPERATION/RIDERSHIP

One approach of how to combine network topology and service information is by considering the operation or ridership of a metro system. Xu et al. (2016) incorporate trip data in order to create a weighted passenger flow network of the Beijing subway system. Through this new methodology, the researchers were able to identify the spatial mobility patterns in the urban area, providing useful insights for policy makers. This methodology can be extended to be applied to other cities and modes as well. Saidi et al. (2017) similarly use generalized passenger travel costs combined with network theory to compare the urban rail transit systems of six large cities.

Amini et al. (2016) take a slightly different approach. Instead, they investigate which type of urban network structure is most suitable for a city, depending on its traffic conditions. While not specifically applied to the area of PT, the authors do attempt to bridge the gap between network theory and real-life by applying the origin-demand data of Mashhad, Iran, to different possible network structures. Their study reveals that the most suitable network structure depends on the amount of traffic in the city, indicating a relationship between network structure and traffic volumes.

Whereas the approach of the aforementioned studies is to integrate service information or demand data into a network formulation, Ingvardson and Nielsen (2018) approach the problem from the opposite direction. Instead, they look how network topology (among other factors) influences the ridership of public transport systems. Their study finds that extensive rail network coverage (among which metro was counted as well) is positively correlated with ridership. This correlation seemed largest for metro, because of its high passenger-carrying capacity. They also discover a strong relationship between urban density and metro network coverage, as a dense metropolitan area requires a high-capacity transport mode and vice-versa. Their study does not find significant influence of network topology indicators on ridership. Having significant transfer possibilities does however increase the mobility of the system, leading to a higher attractiveness of the entire system.

ACCESSIBILITY

Another approach is to consider how the structural topology and service of a metro system relate to its accessibility. A first example of this is by Luo et al. (2019) who integrate accessibility and network science, using a network science-based approach. They formulate an accessibility indicator of travel impedance based on the generalized travel cost (GTC), a combination of in-vehicle travel time, waiting time and the number of transfers. Their work compares eight tram networks using this newly defined indicator. The analysis reveals that, when including these service properties, there is a higher spatial disparity in PT accessibility. In addition, their work shows that larger networks exhibit a larger average travel impedance.

A work in the following year by the same authors takes this integration a step further by evaluating whether passenger flow distribution can be estimated by looking solely at the network properties of a PT system (Luo et al., 2020). Their work reveals that some network indicators can indeed be used to estimate passenger flow with reasonable accuracy. This is especially the case for indicators using weighted graphs and those related to the space-of-service, indicating the higher level of realism of those representations.

Jin et al. (2017) take a more comprehensive approach and consider the accessibility of the whole high-speed rail (HSR) network of East Asia, comparing multiple countries. Their study considers how the structure of the network improves accessibility to different hot-spots. They suggest that further development of an integrated East Asia HSR system could reduce the travel time between the major cities but admit the impacts are lower than national developments and that there are many hurdles to realize this.

2.3.3. PT COMPARISON STUDIES

The final topic of interest is comparison studies which do not explicitly incorporate network topology. In this section, a more in-depth analysis is performed on the existing PT comparison studies. This specifically involves the studies which compare aspects differently than network topology and service information as those are already extensively described in their corresponding sections. As comparison studies such as these are also quite prevalent in grey literature, that category is also included in this analysis.

The first study to consider is a report by McKinsey (2021). In this report, the authors compare the urban transportation system of 25 global cities and focus on what makes them successful on specific aspects. The number of metrics in the report is quite large, but some examples include affordability, availability, efficiency and convenience. A key element of this report is that these metrics are not only judged using objective metrics but also subjectively through user perceptions. In addition, the results are compared to earlier studies in order to evaluate the growth or decline of networks in certain aspects. This provides a picture from a very different perspective than the studies as performed in Sections 2.3.1 and 2.3.2 which are purely focused on the actual network itself and service and less on the user experience.

Taecharungroj (2022) takes an entirely different approach to comparing PT networks. In their work, they compare 127 urban rail transit networks based on traveller reviews via TripAdvisor. The dimensions used for comparison were not predetermined but are instead inferred using machine learning. This means the comparison is executed based on the traveller's perspective and desires, gaining insights that might otherwise be lost. The study identifies which dimensions are most relevant for which networks, giving operators insights for where to improve.

In the study by Weckström and Mladenović (2020), the authors compare the public transport policy of 24 mid-sized cities in Nordic Europe. Their comparison is quite extensive, including the development trajectory, quantity and structure of services as well as planning objectives and measures. The comparison is performed based on five newly developed performance measures. Their study finds that these cities follow fairly similar planning principles, explained by their proximity and cultural similarity. The actual performance of these networks does vary however. Several potential explanations are provided for this varying performance, but no conclusive answer can be given.

Bastidas-Zelaya (2021) seeks to understand the planning and development of metros in Latin America and to find out what makes the planning and development of those metro systems unique compared to other systems around the world. There are four main features they discover to be unique for Latin American metro systems: a higher relative demand; more financial sustainability; metro projects are all relatively new; and finally, hardly any public-private partnerships.

2.3.4. SUMMARY TABLE

All related works are presented in Table 2.1. This summary table provides an overview of all the included related works along with some key information about each of those works. The following columns are present:

- **Networks** - This column simply describes how many networks are considered in each work.
- **Topological** - This column provides information about whether the work takes a topological approach or not.
- **Comparison** - This column describes whether a related work compares different networks or not.
- **Service** - This column informs whether the work includes service information. This could be in the form of for example demand data or scheduling data.

Table 2.1: A summary table of all related works including some additional information

Paper	Networks	Topology	Comparison	Service
Musso and Vuchic, 1988	10	yes	yes	no
Latora and Marchiori, 2002	1	yes	no	no
Sienkiewicz and Hołyst, 2005	22	yes	yes	no
Li and Cai, 2007	1	yes	no	no
Majima et al., 2007	6	yes	yes	no
Lee et al., 2008	1	yes	no	no
Von Ferber et al., 2009	14	yes	yes	no
Derrible and Kennedy, 2009	19	yes	yes	yes
Derrible and Kennedy, 2010a	33	yes	no	no
Derrible and Kennedy, 2010b	33	yes	no	no
Derrible and Kennedy, 2011	-	yes	no	no
Derrible, 2012	28	yes	yes	no
Zhang et al., 2013	30	yes	yes	no
Lin and Ban, 2013	-	yes	no	no
Xu et al., 2013	330	yes	yes	no
Haznagy et al., 2015	5	yes	yes	no
Amini et al., 2016	1	yes	no	yes
Dimitrov and Ceder, 2016	1	yes	no	no
Xu et al., 2016	1	yes	no	yes
Jin et al., 2017	4	yes	yes	yes
Saidi et al., 2017	6	yes	yes	yes
Wu et al., 2017	6	yes	yes	no
Ingvardson and Nielsen, 2018	48	yes	yes	yes
Shanmukhappa et al., 2018	3	yes	yes	no
Hong et al., 2019	1	yes	no	no
Luo et al., 2019	8	yes	yes	yes
Wei et al., 2019	6	yes	no	no
Shanmukhappa et al., 2019	-	yes	no	no
Luo et al., 2020	2	yes	yes	yes
Weckström and Mladenović, 2020	24	no	yes	yes
Bastidas-Zelaya, 2021	20	no	yes	yes
McKinsey, 2021	25	no	yes	yes
Taecharungroj, 2022	127	no	yes	yes

2.4. RESEARCH GAP

Based on the related works described in Section 2.3, several gaps in the literature can be identified. These gaps are further detailed in the corresponding subsections below.

2.4.1. SERVICE INFORMATION AS A QUANTITATIVE MEASURE

The first and arguably biggest gap in literature is the lack of using service as a quantitative measure in comparisons between metro networks. Incorporating service information quantitatively would make it much easier to directly compare networks to each other and gain information about their performance. It can also provide more scientific evidence for findings from qualitative studies as well as a basis for decision making. While recent studies do aim to fill this gap, the amount of literature in this area is still comparatively low. In addition, the focus of these studies is varied and none concretely focus on comparing networks using service information as a measure for accessibility (with the sole exception of (Luo et al., 2019)). Including service information quantitatively consequently also requires a lot of data and a pipeline to process it, which therefore can also be considered as gaps in the literature.

2.4.2. Number of networks included

Nearly all of the related works use either one or multiple networks for application. From Table 2.1, it becomes clear that the exact number of networks included differs significantly per study. These studies can roughly be divided into three categories in terms of included networks:

1. **One network/no comparison** - A portion of the explored literature does not actually compare networks but instead uses one (or a few) networks to apply a newly developed methodology on.
2. **Small comparison (<12 networks)** - A second group of studies does perform a comparison study of multiple networks but restricts this to only a handful of networks. Usually these studies do not aim to create a complete image of all possible networks but instead use this comparison to demonstrate a new methodology and compare its applicability on a variety of cases.
3. **Middle-sized comparison (~12-30) networks** - This group of studies features a larger group of networks that are compared. Frequently, the intent of these studies is to extract patterns and groupings from these networks in order to get an image of the overall state of networks.

With some exceptions — (Taecharungroj, 2022) (127 networks), (Ingvardson & Nielsen, 2018) (48 networks) and (Xu et al., 2013) (330 networks) — all explored studies fall into one of these categories. It can be noticed that these comparisons are far from exhaustive, considering that there are around 190 metro networks in total (UITP, 2022). As such, there is a clear gap in the literature of metro comparison studies with 30+ networks.

2.4.3. Unexplored networks and regional differences

Lastly, in addition to the number of networks lacking in literature, the actual networks that are explored is also an area for improvement. Nearly all studies mentioned in this review perform one of either two comparisons: a comparison on the largest and most significant metro networks (Musso and Vuchic, 1988; Von Ferber et al., 2009, Derrible, 2012); or a comparison of studies in one particular area (Xu et al., 2013; Weckström and Mladenović, 2020; Bastidas-Zelaya, 2021). The most notable exception to this is Taecharungroj (2022) who analyzes 127 networks worldwide of varying sizes. However, because of the reliance on TripAdvisor data, this study has a very different focus than most others and does not in fact consider the topology nor service information of these networks. This focus from these studies on either large networks or specific networks, means there are many networks that have not been explored at all or only in a regional context. In addition, including many networks from different regions means it is also possible to evaluate regional differences. Researching these regional differences is another gap this study aims to fill.

3

METHODOLOGY

In this chapter, the methodology applied for this study is described. Firstly, the manner in which networks are selected is briefly explained in Section 3.1. Secondly, the network representations used for this study are described in Section 3.2. Thirdly, the primary metric used for comparison is described in Section 3.3. Lastly, the secondary metrics based on network- and city-related aspects are explained in Sections 3.4 and 3.5 respectively.

3.1. NETWORK SELECTION

The first step in the process is selecting the metro networks to include in this study. In this section, the overview of potential networks to include is described. Firstly, the definition of what constitutes a metro network is explained in more detail in Section 3.1.1. Following this definition, a potential set of networks to include in this study follows, which is briefly illustrated in Section 3.1.2.

3.1.1. DEFINITION OF METRO

Metro, subway, heavy rail and (mass) rapid rail transit are all definitions for a specific type of mass urban rail public transport. As becomes clear from the title, for this study the moniker "metro" is used to indicate networks of this type. The exact definition differs depending on the exact source, but there a few general rules that nearly all sources adhere to, as also described by the International Association of Public Transport (UITP, 2022):

- Located in an urban area
- Exclusive right-of-way
- Grade-separated through either tunnels or elevation
- High frequency and capacity vehicles

3

For this study, these four rules are taken as the main determinant to include networks. While networks can technically only be described as a metro if they meet all four of these criteria, some networks exist that do not fully meet all of the criteria but are still classified as metro. This especially refers to the fourth requirement of "high frequency and capacity" which is not a binary requirement nor does it have strict limits defined. This definition of metro therefore excludes tram and light-rail on the one hand and suburban and commuter rail on the other hand. While this excludes some cities that rely heavily on these types of transportation as their primary urban PT method, this does ensure the comparison between different networks stays pure, as the metro networks included should be fairly similar from a systematic point of view.

3.1.2. OVERVIEW OF POTENTIAL NETWORKS

Following the definition as described in Section 3.1.1 as closely as possible, an initial selection of potential networks can be created. This selection is based on varying sources listing metro networks around the world, such as UITP; and network planners and operators. The aim of this initial list is not to be completely exhaustive, but to include nearly all significant networks that can be classified as metro according to the aforementioned definition. The total list is provided as a world map in Figure 3.1. All cities with a metro network are indicated in red on this map. Some cities have multiple systems that could be identified as separate metro systems (such as the Tokyo Metro and Toei Subway in Tokyo, Japan). Depending on the exact composition and transferability between these systems, it is determined on a case-by-case basis whether these systems are considered as one integrated system. The desire is to have only one (integrated) system per city in order to keep the comparisons with other cities clear.

As described in Section 2.4.2, there is a literature gap both in terms of the number of networks included as well as the specific networks that are included in related works. Therefore, the aim of this study is to include as many networks in the comparison as possible, in order to get the most complete view of the world's metro networks. In the ideal situation all networks present in Figure 3.1 are included in the final set used for analysis. Naturally, this is hard to actually realize considering data limitations. The final number of networks that is actually included in this study is described further on in Section 4.1.

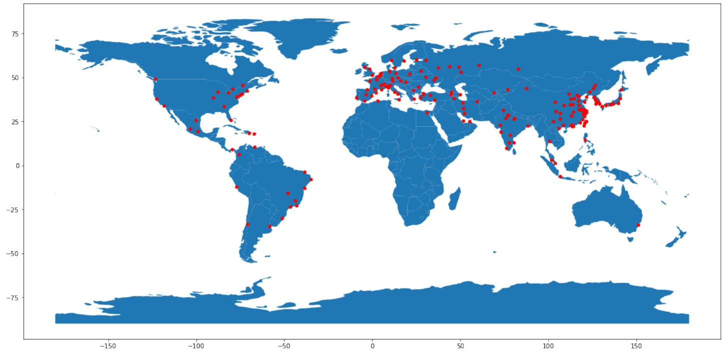


Figure 3.1: All 190 cities in the world with a metro network

3.2. PT NETWORK REPRESENTATION

The focus of this study is to create a comparison of metro networks that also includes the service information quantitatively. Similar to other studies, this is done by representing the metro networks topologically using concepts from graph theory. From these graph representations, metrics can be computed that can be used to compare the networks. In this section, the chosen representations are extensively described.

3.2.1. REPRESENTATION OPTIONS

When representing PT networks using graph theory, there are generally four options for the type of representation used. How to translate a metro map into these four options, is displayed graphically in Figure 3.2.

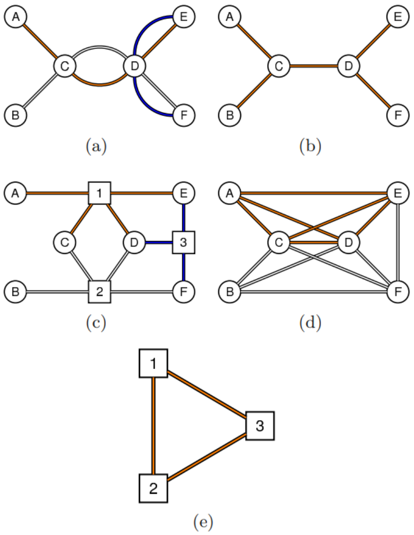


Figure 3.2: (a) A simple metro map (b) L-space (c) B-space (d) P-space (e) C-space (Von Ferber et al., 2009)

These four spaces are four different ways of displaying the topology of the metro network, both in terms of its infrastructure as well as its service. Whereas the L-space provides only information on the infrastructure of the system, B-, C- and P-space provide information on the service of the system. As both the infrastructure and service are relevant for this study, it is sensible to use two representations: one for infrastructure and one for service. As L-space is the only representation describing infrastructure, it warrants no further explanation that it is used as the infrastructure representation. For service, the options are threefold. C-space is the least attractive option as it only provides information on which lines are connected, losing vital information about exactly which stations are on each line. B- and P-space both include this information and are thus equally rich in information in that aspect. B-space includes some extra nodes detailing the exact line names/numbers connecting the different stations. As this extra information is not directly relevant for this study, P-space is used instead.

In order to understand exactly what information is captured in these representations and how this is formulated mathematically, further explanations are provided in the corresponding Subsections 3.2.2 and 3.2.3 below.

Before defining each representation separately, a general mathematical definition of a graph is provided. A graph G is represented mathematically by $G = (V, E, w)$. In this graph, V is a set of vertices or nodes while E is a set of pairs of nodes (i, j), also known as edges. Edge (i, j) represents the two nodes i and j joined by this edge. Edges can also have an edge weight function w that maps each edge $(i, j) \in E$ to a real-valued weight $w(i, j)$. In addition to the main weight, other information might be captured using additional labels $\mathrm{l}(i, j)$.

3.2.2. L-SPACE

As mentioned in the introduction of this section, the L-space or alternatively space-of-infrastructure is a topological representation of the infrastructure of the network. In this representation, stations are represented as nodes while the tracks/tunnels between them are represented as edges. This is the common way to represent L-space where the information about the different metro lines/routes is lost and only the infrastructure remains. Within this basic concept of L-space, there are two factors that can differ per implementation: edge weight and edge direction. These concepts are explained in more detail below.

EDGE WEIGHT

When transforming a metro network into an L-space representation, it is possible to add a main weight to the edges in order to represent extra information. Below, the different possibilities for these edge weights are described along with what information this weight describes.

- **Unweighted** - By far the most commonly used option is to not assign any weight to the edges at all (Sienkiewicz & Hołyst, 2005) (Von Ferber et al., 2009) (Derrible & Kennedy, 2009). In this implementation, each edge simply represents a connection between two stations without any additional information about the length or duration of this connection.

- **Station-to-station distance** - A way to provide more information on the edges is to use the distance between stations. This simply describes the distance covered by the metro network between two stations or in other words the length of an edge. In the local sense, this weight provides information about how far apart certain stations are compared to others. Globally, this weight can provide information about the average interstation distance in the network or the total length of the network. For this study, the station-to-station distance is defined as $\mathrm{l}(i, j)$.

- **Average station-to-station in-vehicle travel time** - An alternative that is partly dependent on the aforementioned distance, is the in-vehicle travel time between stations (Luo et al., 2019). This weight is a combination of the distance, the type of vehicles used (i.e., as they generally determine the maximum speed of operation) as well as the infrastructural intricacies of the network (e.g., sharp bends, speed limits because of external factors). As such, this weight provides more detailed insights into both the infrastructure of the network as well as its operation as opposed to the distance. Mathematically, $\text{ivt}(i, j)$ represents the in-vehicle time of edge (i, j).

From these three options, the in-vehicle travel time is used as the main edge weight for this study. It provides a more detailed picture of both the infrastructure and operation as opposed to a distance-based or unweighted graph. It also has great potential for further integration with other service aspects when combined with P-space. In addition, the station-to-station distance is used as a secondary label for calculating certain indicators such as the total network length.

Edge direction

Firstly, it is important to establish that all L-space representations are so-called "simple" graphs. This means that each edge can only exist once. In addition, self-loops are also not permitted in this representation. In graph theory, edges can be one of two types: undirected or directed. An undirected graph, or simply "graph", is a graph in which the edges do not have an orientation. As this is an undirected graph, each node pair can only have one edge between them and thus also $(i, j) = (j, i)$. In a directed graph on the other hand, edges have an orientation, and as such for any edge $(i, j) \neq (j, i)$. This means between every pair of nodes, two edges can exist: one for each direction.

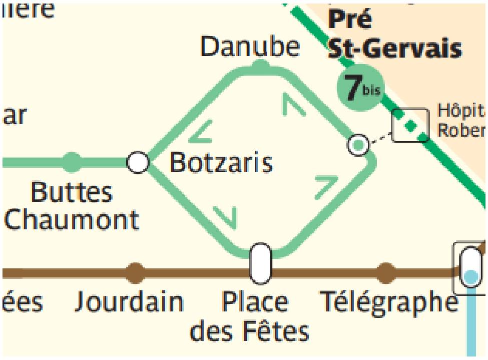


Figure 3.3: An example of directed metro lines in the Paris network[1]

As L-space graphs are commonly represented using unweighted edges (as described in the previous section), usually there is little reason to use directed edges. When using unweighted edges, one undirected edge between two nodes is simply enough to represent the potential to move between these two stations. This is a valid choice for metro networks as vehicles nearly always travel in both directions between stations. There are however exceptions to this as can be seen in Figure 3.3. On this turquoise line (7bis) in the Paris metro network, vehicles travel only from Botzaris to Place des Fêtes to Pré St-Gervais to Danube and back to Botzaris, without travelling in the other direction. While this is a very uncommon feature in metro networks, it is prevalent enough to account

[1] Based on: https://www.ratp.fr/sites/default/files/plans-lignes/Plans-essentiels/Plan-Metro.1653922257.pdf

for it in the L-space implementation. In addition to accounting for these exceptions, it is also sensible to use directed edges when using weighted edges, as is the case for this study. This is because the exact weight might differ depending on the travel direction. The travel time in one direction can for example be higher as the vehicle has to travel up a slope while it travels downwards in the opposite direction. In order to properly represent these potential differences, directed edges are used in the L-space implementation of this study.

3

CONCLUSION
In conclusion, the L-space consists of a set of nodes representing stations and a set of directed edges representing the rail infrastructure between the stations. The edges are weighted with the in-vehicle travel time between the two stations for that direction. L-space can thus be defined as a graph $\mathrm{L} = (V, E, w)$ where V = a set of nodes, $E \subseteq \{(i,j) \mid (i,j) \in V^2 \wedge i \neq j\}$ with main edge weight $w : (i,j) \rightarrow \mathrm{ivt}(i,j)$ and an additional label $\mathrm{l}(i,j)$.

3.2.3. P-SPACE
Whereas L-space describes the physical, infrastructural state of the network, the information about the service (e.g., where the lines are, which stations can be transferred at) is lost. The P-space, or space-of-service, is the chosen space to represent the actual service that is run on the infrastructure described by the L-space. In this representation, similarly to L-space, the stations are represented as nodes. An edge between two stations represents the fact that those two stations share a line and thus have a direct means of travel between them (i.e., without a transfer). For P-space, the same types of decisions have to be made as for L-space in terms of the edge weight and direction. These choices are briefly described below.

EDGE WEIGHT
Similarly to L-space, P-space has the option to use either weighted or unweighted edges. Depending on the exact weight chosen, P-space can provide different types of information about the level-of-service in the network. These different options are described below.

- **Unweighted** - The most common representation, similar to L-space, is to have unweighted edges (Sienkiewicz & Hołyst, 2005)(Von Ferber et al., 2009) (Xu et al., 2013). In this representation, each movement (with the exception of the first) from one node to the other simply represents a transfer. As such, this representation can provide information about the number of transfers necessary for any trip through the network.

- **Vehicle capacity** - Another option is adding the vehicle capacity for a certain line to the edges as a weight. With this representation, along with information about transfers, information about the potential number of travellers making this trip would be known.

- **Frequency** - A different weighted option is to add the line frequency as a weight to the edges. This would thus provide information about how often in a certain time period vehicles travel across this line. For this study, the frequency is arguably much more interesting than capacity as it provides more insights into the schedule made by the planner. The frequency itself is however difficult to combine with the in-vehicle travel time that is used as a weight for L-space.

- **Average waiting time** - This weight turns the frequency of a line into an average waiting time instead (Luo et al., 2019). This is done through the following formula:

$$w : (i, j) \rightarrow \mathrm{wt}(i, j) = \frac{60}{(veh_{tot} \div p)} \div 2$$

  In this formula veh_{tot} represents the total number of vehicles travelling on the line in a set time period p. This could for example be the total number of vehicles in a full day of 24 hours. This number of vehicles is then divided by the period to get the vehicles per hour. Dividing 60 by the vehicles per hour, gives the maximum waiting time in minutes. This maximum waiting time is in turn divided by 2 in order to get the average waiting time.

  To illustrate exactly how this works, let us consider a metro line with a frequency of six vehicles per hour. For this, the assumption is made that these vehicles are evenly distributed throughout the hour. While this might not always be the case, this is very common for metros and public transport in general. Six vehicles per hour is the equivalent of one every ten minutes. As such, the maximum waiting time for a vehicle is ten minutes (i.e., when a traveller arrives just as the previous vehicle is leaving). This is however the absolute worst case scenario which is not entirely realistic. Another assumption made here, is that travellers arrive evenly distributed, or in other words do not consider the scheduling of the metro. This is a valid assumption for metros considering their high-frequency scheduling is in fact one of the core features of a metro system (UITP, 2022). This high-frequency schedule means there is generally no need for travellers to consider the timetable, as vehicles will always arrive within a reasonable time. Considering that travellers thus arrive randomly, they arrive on average halfway through the maximum waiting time. In this case that would give an average waiting time of five minutes.

Using this average waiting time as an edge weight means the P-space provides information on both the number of transfers needed as well as the average time travellers have to wait for each leg of their trip. As such, the P-space now provides a lot of information on the service and scheduling of the network. Combining this waiting time with the in-vehicle time from L-space, provides a fairly comprehensive image of the travel time that travellers actually experience. It is for this reason that the average waiting time is used as the main P-space weight in this study. In addition, the route/line to which each edge belongs is also added as a secondary label. In this way, the different routes are identifiable in P-space. Mathematically, the route of each edge is represented by $\mathrm{r}(i, j)$.

EDGE DIRECTION

In addition to the edge weight, the decision must also be made whether to use directed edges or not. Considering that P-space, similar to L-space, uses edge weights that might differ per travel direction, it is sensible to also use directed edges in P-space. An example of how the average waiting time might be different per travel direction is the possibility of extra express trains that only travel in one direction thus lowering the average waiting time for that direction.

CONCLUSION

In conclusion, the P-space contains the same nodes as L-space, representing each station. Edges exist between each station that is connected by the same line. The weight of these edges represents the average waiting time on that line between the two stations. P-space can thus be defined as a graph $P = (V, E, w)$ where V = a set of nodes, $E \subseteq \{(i,j) \big| (i,j) \in V^2 \wedge i \neq j\}$ and $w : (i,j) \rightarrow \mathrm{wt}(i,j)$. An additional label $\mathrm{r}(i,j)$ exists to represent the route of each edge.

3.3. PRIMARY METRIC

Based on the L- and P-space representations as described in Section 3.2, a variety of metrics can be created. These metrics are a descriptor of different network attributes including both topological and operational properties of the networks. In this section, the primary metric used for comparison is described in more detail. This metric is the main factor which is used to evaluate the performance of networks. Networks can thus be directly compared based on their performance on this indicator. To help explain why certain networks score better than others, secondary attributes are used for comparison. These are further described in Sections 3.4 and 3.5 hereafter.

As described in Chapter 1, the aim is to create a primary metric that includes service information quantitatively. In addition, considering the overall goal of accessible PT, it is also important to relate this factor to the concept of accessibility. Firstly, this concept is explained in a little more detail in Section 3.3.1. Afterwards, the metric is further explained and defined in Section 3.3.2.

3.3.1. ACCESSIBILITY

In order to be able to properly compare networks, a topic on which to compare them must be chosen. Based on the gaps identified in Section 2.4, accessibility is chosen as the most suitable dimension for comparison. As accessibility tends to be a term with varied meanings, it is important to narrow down the exact meaning that is applied in this study. In transport planning, the concept generally refers to a measure of the ease of reaching certain destinations. In the specific case of a metro network, this concept can apply to multiple different facets of the network. Examples are the proximity of inhabitants to the nearest stations, the coverage of the network compared to the size of the city or how quickly points in the city can be accessed with the metro network. As this study takes a topological approach focusing on the network and its service, it is sensible to define accessibility in that sense. As such, the exact location of stations, their reachability in the urban environment and similar factors are not considered. As these topics have been extensively studied already (e.g., Acampa et al., 2019; Igualada, 2015), it is sensible to focus on the network itself instead. For this study, accessibility is focused on how quickly travellers can travel through the network. In this way, transport planners can learn how their network performs compared to others in terms of service accessibility. Based on the earlier introduced network representations, a metric for defining this accessibility is introduced below.

3.3.2. METRIC DEFINITION

A common way to define accessibility in PT networks is with the shortest path travel time through the network. Considering the topological approach that this study takes, this is a sensible direction to follow here as well. The travel time in a graph can be defined based on, for example, the number of nodes crossed in an unweighted graph or the total distance in a weighted graph. Neither of these methods explicitly incorporate the service of the metro network however, which is a desired characteristic for the research. Based on the L- and P-space representations used in this study, it is possible to create a composite travel time consisting of the in-vehicle travel time, waiting time and number of transfers. This can be mathematically defined as described in Equation 3.1 below.

$$tt(i,j) = ivt^{\mathrm{L}}(i,j) + \alpha * wt^{\mathrm{P}}(i,j) + \beta * tf^{\mathrm{P}}(i,j) \tag{3.1}$$

$tt(i,j)$ - The total (shortest path) travel time between node i and j (min)
$ivt^{\mathrm{L}}(i,j)$ - The total in-vehicle travel time between node i and j from L-space (min)
$wt^{\mathrm{P}}(i,j)$ - The total waiting time between node i and j from P-space (min)
$tf^{\mathrm{P}}(i,j)$ - The number of transfers needed in the shortest path between node i and j from P-space (-)
α - A positive integer, constant penalty per minute of waiting time (min/min)
β - A positive integer, constant penalty per transfer (min/transfer)

The shortest path between two arbitrary nodes i and j is thus the path with the lowest total travel time which is a combination of the factors $ivt^L(i,j)$, $wt^P(i,j)$ and $tf^P(i,j)$. Considering the values for these three factors have to come from two different representations, the calculation of the shortest path is non-trivial. Calculating the shortest path for each of these factors separately might result in different paths, meaning the values cannot be combined. Instead, the shortest path calculation is done firstly in L-space based on the in-vehicle time, after which the corresponding waiting time and number of transfers for that path are retrieved from P-space. Naturally, the shortest path in L-space based on in-vehicle time might not be the shortest path in terms of total travel time. As such, the k-shortest paths based on in-vehicle time in L-space are considered. The corresponding waiting time and transfer information are then retrieved from P-space for all of these paths. The addition of these three components (multiplied by the respective penalty factors) then determines which path has the shortest total travel time and is thus considered the shortest path. The exact implementation of this along with the values for k and the attribute coefficient values α and β are described in Section 4.4.

Mathematically, the shortest path between two arbitrary nodes i and j can be defined as follows. Firstly, the definition of a path in a graph must be provided. A path in a directed graph is a sequence of nodes $N = \{v_1, v_2, ..., v_n\} \in V \times V \times ... \times V$ such that there is a directed edge (v_x, v_{x+1}) from each node v_x to its successive node v_{x+1} where $1 \le i < n$. This is thus a path from node v_1 to node v_n with length $|N| = n-1$.

Given a directed graph G, a real-valued weight function $w : E \rightarrow \mathbb{R}$, the shortest path between two arbitrary nodes i and j is the path $SP = (v_1, v_2, ..., v_n)$ (where $v_1 = i$ and $v_n = j$) that over all possible n minimizes the sum $\sum_{x=1}^{n-1} w(v_x, v_{x+1})$. This general notation can be adopted for the calculation of the shortest path calculation in L-space as follows:

$$ivt^{\mathrm{L}}(i,j) = \min \sum_{x=1}^{n-1} \mathrm{ivt}(v_x, v_{x+1}) \tag{3.2}$$

This calculation provides both a shortest path SP in L-space between node i and j as well as the corresponding in-vehicle time $ivt^L(i,j)$. The waiting time and number of transfers corresponding to this path SP can then be gained from P-space to determine the value of the total travel time. This is done through the following formulae.

$$wt^{\mathrm{P}}(i,j) = \mathrm{wt}(v_1, v_2) + \sum_{x=2}^{n-2} \begin{cases} \mathrm{wt}(v_{x+1}, v_{x+2}), & \text{if } \mathrm{r}(v_x, v_{x+1}) \neq \mathrm{r}(v_{x+1}, v_{x+2}) \\ 0, & \text{if } \mathrm{r}(v_x, v_{x+1}) = \mathrm{r}(v_{x+1}, v_{x+2}) \end{cases} \tag{3.3}$$

$$tf^{\mathrm{P}}(i,j) = \sum_{x=1}^{n-2} \begin{cases} 1, & \text{if } \mathrm{r}(v_x, v_{x+1}) \neq \mathrm{r}(v_{x+1}, v_{x+2}) \\ 0, & \text{if } \mathrm{r}(v_x, v_{x+1}) = \mathrm{r}(v_{x+1}, v_{x+2}) \end{cases} \tag{3.4}$$

The total waiting time for a shortest path SP is the summation of the waiting time of the initial edge and the waiting time of each edge that is on a new route. In this way, each time a transfer is made, the waiting time for the new line is added to the total waiting time. The number of transfers is calculated in a similar manner but instead for each edge on a new route, the sum is increased by 1 to indicate a transfer has been made.

3.4. NETWORK INDICATORS

The primary indicator is used as a performance indicator to compare networks. In order to relate this performance to other factors, four network-related indicators are defined. These indicators are all related to the size of the network. While this means there is some natural overlap between them, it is interesting to note the differences in their descriptive power based on the results from this study. In the respective subsections below the indicators are described along with an explanation of why they are included.

3.4.1. NUMBER OF STATIONS

The first network-related indicator included is the number of stations. This indicator simply describes the total number of stations in the metro network. The number of stations is a manner to define the size of the network. The main reason for including this indicator is to see whether a relationship exists between the travel time (as described by the primary indicator) and the network size. It is hypothesized that a positive correlation between these factors exists considering the intuitive notion of increasing travel times in larger networks and related studies (e.g., (Luo et al., 2019)) confirming this pattern. This indicator is calculated simply as the total number of nodes in L-space (P-space would be equally possible as the nodes are the same in both graphs). Mathematically, this indicator can be defined as per Equation 3.5

$$ns^{\mathrm{L}} = |V(\mathrm{L})| \tag{3.5}$$

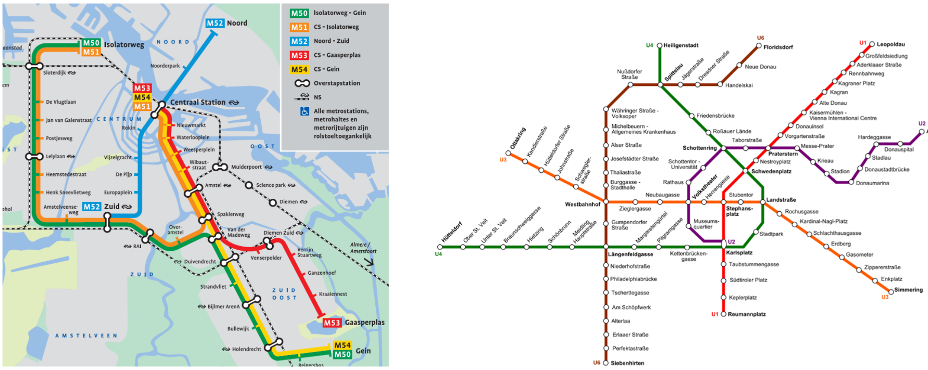
3.4.2. NETWORK LENGTH

Another network-related indicator describing network size is the total network length. While also a descriptor of size, like the number of stations, this indicator also incorporates the coverage of the network better. This is because an increase in network length implicitly means covering more area. This also relates it further to the concept of accessibility, as network coverage has a direct relationship with how accessible the network is. As this indicator is fairly similar to the number of stations, the hypothesized pattern is the same for this indicator. Naturally, some individual differences between networks can be expected considering the slight differences with the number of stations. The calculation of this indicator is the simple addition of the length of all edges in L-space divided by two. This division by two is required considering each edge exists twice by the directed nature of the chosen representation. This will alter the results slightly for networks that do actually have directed edges (such as the Paris example mentioned in Section 3.2). Considering the infrequency with which this occurs, this is deemed an acceptable simplification. The total network length is defined mathematically as per Equation 3.6.

$$l^{\mathrm{L}} = \frac{\sum_{(i,j)\in E(\mathrm{L})} \mathrm{l}(i,j)}{2} \tag{3.6}$$

3.4.3. NUMBER OF LINES

While the number of lines might seem similar to the previous two indicators, it in fact describes a much more complex process than that. The number of lines is namely much closer related to the service of a network, as a line is in fact a service run on a piece of infrastructure. Planners have some freedom of choice in defining what exactly constitutes a line. Comparing this indicator to the travel time can reveal whether having fewer or more lines increases the travel time. Naturally, the expectation, similar to the other two network indicators, is that an increased number of lines leads to an increased travel time. Adding an extra line generally leads to more stations and a longer network length which in turn lead to an increased travel time. Based on intuition, some caveats can be made about this expected similarity however. The freedom of a transport planner to determine what exactly constitutes a line means that the results can in fact be quite different. Figure 3.4 illustrates this difference quite clearly. While both the Amsterdam and Vienna network have five lines, what exactly constitutes a line in each network is very different. Whereas Amsterdam has multiple parallel lines, Vienna exclusively has non-parallel, crossing lines. This in turns leads to a variation between them in terms of both the number of stations and the network length. To be specific: Amsterdam has a network length of 39 km and 39 stations, while Vienna has a length of 80km and 98 stations. As such, it becomes clear that while these networks have the same number of lines, they are in fact very differently sized with Vienna being more than twice as large in both aspects.



(a) Amsterdam (b) Vienna

Figure 3.4: Two five-line networks with a different geometry[2]

The number of lines can be defined mathematically as per Equation 3.7.

$$nl^{\mathrm{P}} = \left| \{ \, r(i,j) \mid \forall (i,j) \in E(\mathrm{P}) \, \} \right| \tag{3.7}$$

3.4.4. DIRECT STATION DISTANCE

The final network indicator is the average direct station distance. More specifically, the direct (i.e. geodesic) distance between all station pairs is calculated, of which the average value is then taken for comparison. This is classified as a network indicator considering it mostly relates to the locations of the stations in the network, which can be seen as a network choice. On the other hand, the location of stations is also strongly related to the lay-out and population hotspots of the city. As such, it is also in part a city-related indicator. The hypothesis is that this indicator is positively correlated with the travel time, considering they are both indicators of the distance between stations, whether measured in direct distance or travel time. Nevertheless, this indicator also contains information about the directness of the network, or in other words, whether the metro network covers distance between stations in a straight line. Considering that this directness differs per network, as also shown by Derrible and Kennedy (2010a), these differences are also expected to be discovered in the comparison with travel time. This indicator is calculated by taking the direct distance between each pair of stations (defined using the function d). Similarly to the number of stations, this indicator is calculated in L-space but can be calculated using P-space as well. The average direct station distance is defined mathematically as per Equation 3.8.

$$d^{\mathrm{L}} = \frac{\sum_{i \in V(\mathrm{L})} \sum_{j \in V(\mathrm{L})} d(i,j)}{|V(\mathrm{L})|^2 - |V(\mathrm{L})|} \tag{3.8}$$

[2] (a) retrieved from: https://www.gvb.nl/sites/default/files/metrokaart.pdf, (b) retrieved from: https://www.introducingvienna.com/metro

3.5. CITY INDICATORS

In addition to indicators related to the network itself, a second set of indicators is included related to the city in which the network resides. The goal of these indicators is to discover if a relationship exists between the characteristics of the city and the metro network operating in it. The three indicators are discussed in the corresponding subsections below.

3.5.1. COUNTRY/REGION

The first city-related indicator is that of the city's country and region. Considering the number of related works that focus on the same sets of networks or specific regions, it is sensible to consider regional differences as a relevant factor for this study. As such, both the country and region of each network is added as a data point. The country indicator is fairly self-explanatory, but in terms of region a little extra explanation must be provided. For this study, the considered regions are: Africa, North-America, South-America, Asia and Europe. This is similar to all inhabited continents with the exception of Oceania which is included with Asia considering the extremely low number of networks in Oceania. The regions are not used as a distinct indicator for comparison but are instead considered in all comparisons as a secondary feature.

3.5.2. POPULATION

The size of a city is usually defined by either its population or surface area. For this study, population is chosen as the indicator of a city's size, since the surface area is already indirectly included in some other indicators such as direct station distance. Population also captures information not yet directly captured by any other indicator. The hypothesis is that there is a positive relationship between the population and the travel time in the network. A larger population generally means a larger city which in turn suggests a larger network with longer travel times (as hypothesized in Section 3.4). Considering however, that the metro is not the main mode of public transport for all cities, this relationship is most likely weaker than some of the others. A city in which the bus is the main PT mode will most likely have an underdeveloped metro network. As such, it can be expected that there will also be big differences in networks' travel times for this comparison.

3.5.3. GDP

Another city-related aspect that is important to include, is its wealth. Commonly, this wealth is defined using the Gross Domestic Product (GDP), which is also what is used for this study. The reason to investigate whether a relationship exists with the travel time, is because of the potential relationship between wealth and infrastructure investments. Naturally, a metro network requires many high-cost investments in its infrastructure. A city with a higher level of wealth is most likely able to afford more of these significant investments. Therefore, a correlation, if existent, should be expected to be positive between the GDP and travel time. On the other hand, having more money to invest in the metro network can also help reduce the travel time as the network gets faster and more efficient by providing more transfer opportunities and faster vehicles. As such, it is difficult to provide a strong hypothesis for this indicator.

4

Implementation

In this chapter, the implementation of the methodology from Chapter 3 is described. Firstly, the actual networks included in this study are described in Section 4.1. Afterwards in Section 4.2, the data processing for this study is discussed. The actual manner in which the L- and P-space are implemented is detailed in Section 4.3. The implementation of the shortest path algorithm used to calculate the primary indicator is detailed in Section 4.4. Lastly, this chapter is concluded in Section 4.5 with a description of the additional data that was retrieved for this study.

4.1. Networks included

In order to retrieve the proper data for the metro networks in this study, a data format must be chosen. This data format is described in Section 4.1.1. The actual retrieval of this data for all networks is described thereafter in Section 4.1.2. Lastly, the set of networks included in this study is described in Section 4.1.3.

4.1.1. Data format

As described in Section 3.1 there are around 190 potential metro networks that can be included in this analysis. As the chosen implementation of L- and P-space requires scheduling data in order to calculate the weights, a data format must be chosen to retrieve this data. As the goal is to include as many of these 190 networks as possible, it is desirable to take a data format that is both openly available and commonly used. As such, it was decided for this study to make use of the General Transit Feed Specification (GTFS) format (Google, 2022). This format was chosen for a few reasons specifically.

Firstly, this is the most commonly used internationally accepted format for public transport data. A lot of PT operators provide data in this format meaning that the chances of the relevant data being available is larger. In addition to that, it provides a specific format for operators to use, meaning the data should be quite universal and require less adaptation when using it. Secondly, because of the commonness of this format, a lot of corresponding software exists. The existence of processing software for this format should ease the analysis process of this study significantly.

Within the GTFS format there are two types of data: static and realtime. As the name suggests, static refers to static data such as station locations and scheduling data. Real-time on the other hand shows the location of vehicles traveling through the network in realtime. As this study aims to translate scheduling data into a metric for comparison, GTFS static is the most suitable to use. This also increases the potential networks to be included, as GTFS static is a lot more commonly available than GTFS realtime.

While this format is the most commonly used international format, not all network operators actually provide this format publicly and for free. The next section provides more details on how exactly this data is retrieved for all networks.

4.1.2. DATA RETRIEVAL

As mentioned in Section 4.1.1 the data format used for this study is GTFS static. In order to acquire this data, the following public sources were considered:

- **Mobility Database** - The primary data source used for this study is the so-called Mobility Database (MobilityData, 2021). The Mobility Database is a platform by the non-profit organization MobilityData that hosts a variety of publicly available PT data sets in varying formats (among which is GTFS static). This platform has a database which should contain nearly all available GTFS files for cities around the world. The cities that have a metro network as mentioned in Section 3.1 are all searched for in this database. For those networks that have GTFS data present, this data is downloaded and stored locally for future access in the further steps of the study.

- **Manual search** - For networks that appeared to not have any GTFS data according to the Mobility Database, a second search is executed. This search consists of a manual Google search of the city/network and the term "GTFS" or "data". In this manner, a few extra databases were acquired as well.

Through these two methods, the GTFS data is retrieved for a total of 69 networks. These remaining networks are presented on a world map in Figure 4.1. The networks highlighted in red are those that actually have GTFS data available, whereas the black ones are networks that do not have GTFS data available. There are a few things to note about the networks that do and do not have data available. While the Asia-Pacific region actually has the majority of metro networks in the world, there are only a few that actually have publicly available GTFS data. Especially the Indian, Chinese and Japanese regions have hardly any networks available with the exception of Cochin, Hyderabad, Hong Kong and Kobe. The European and North-American regions on the other hand, nearly all have GTFS data available, with a few exceptions, primarily in the UK, Italy and Eastern Europe. In the Latin-American and Middle-Eastern regions the availability varies but for both regions there are a handful of available networks.

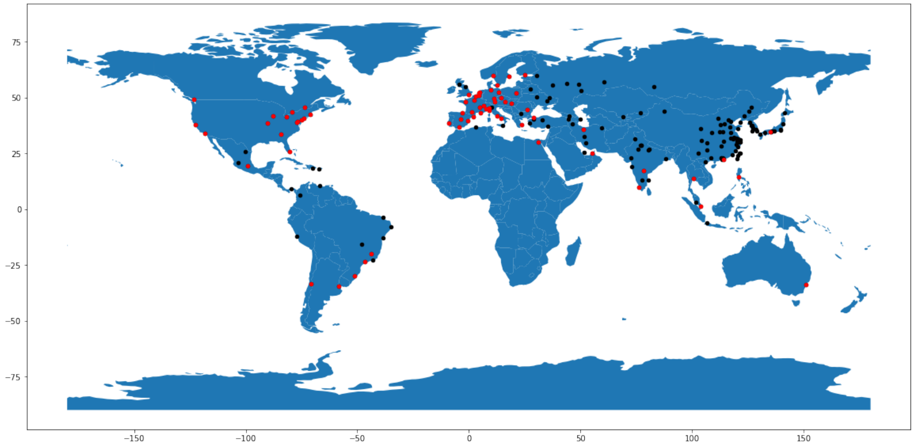


Figure 4.1: All 69 metro networks in the world with GTFS data available indicated in red

4.1.3. FINAL SET OF NETWORKS

In order to actually be able to use the networks in the analysis, the GTFS data for all of these networks needs to be verified and processed. This exact process is described further in Section 4.2. Following this process reveals a variety of reasons why some networks are excluded from the final set:

- **Data is missing** - For quite a significant set of networks, it seemed as if GTFS data existed for the metro network, while it in fact did not. In these cases, GTFS data exists for the city for other modes such as bus and tram but not for metro. For some cases, the metro line does exist in the data but is not programmed as metro, as it is in actuality a commuter rail or tram. The networks that were excluded because of missing data are: Bangkok, Belo Horizonte, Hong Kong, Lausanne, Miami, Manila, Porto Alegre, St. Louis and Sydney.

- **Data cannot be processed** - For six networks, the GTFS data does exist, but could not be processed in the pipeline. The exact reason for not being able to process differs greatly per network but is usually related to the way the GTFS files are written. The networks that were excluded for this reason are: Barcelona, São Paulo, Istanbul, Mexico City, Singapore and Tehran.

- **Data is incorrect** - For three networks, the GTFS data existed and could be processed but was not in fact correct data. Not correct in this case refers to how the programmed network does not match up with the actual network. This could mean missing stations due to construction or extra stations that are in fact unused. These networks are excluded as these mistakes in the programming will greatly affect the metric calculation and consequently yield incorrect results. The networks that were excluded for this reason are: Bucharest, Hamburg and Munich.

For networks with data that is incorrect or could not be processed, Appendix F can be referred to for a more specific case-by-case explanation of why they were excluded.

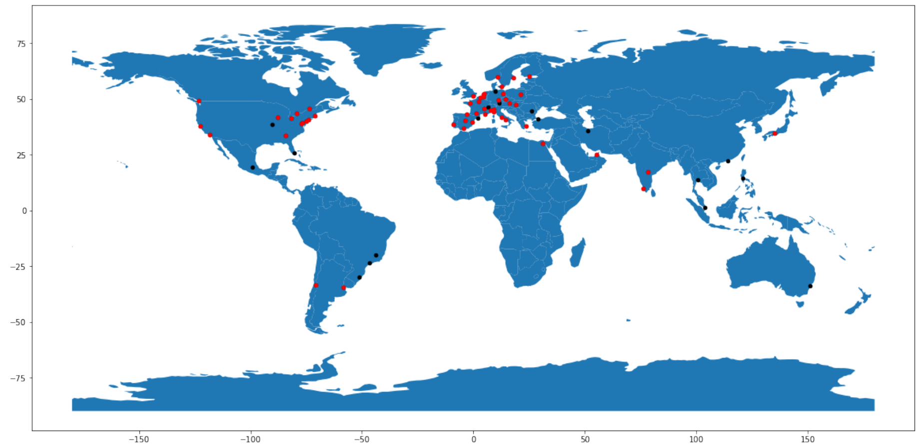


Figure 4.2: The 51 metro networks included in the analysis indicated in red

Following these exclusions, a total of 51 networks remain to be used in the final analysis. This final set of networks can be found in Figure 4.2. The final 51 networks are indicated in red, whereas the 18 removed networks are indicated in black.

Unfortunately, this final filtration step removes a lot of networks from already underrepresented regions, meaning Europe and North-America now heavily dominate the data set. More detailed information such as the number of lines, city GDP and city population are provided for the included networks in Appendix C.

4.2. DATA PIPELINE

In this section, the pipeline that was created in order to turn GTFS data into network representations is briefly described. This section only focuses on the actual processing of the pipeline, not the implementation details for the network representations. Those are instead extensively described in Section 4.3 hereafter.

The pipeline used to process the data for this study is programmed in Python (Python, 2022). A Jupyter Notebook (Jupyter, 2022) is created containing the entire pipeline to go from loading-in GTFS data to outputting metrics based on this data. The pipeline uses a wide variety of Python libraries in order to process the data, calculate metrics and output the data into the desired graph formats. In principle, this pipeline is able to process any type of (correctly programmed) GTFS metro file into the desired network representations and corresponding metrics. The three most significant libraries used are gtfspy (Kujala et al., 2018), NetworkX (Hagberg et al., 2008) and Bokeh (Bokeh Development Team, 2018). The gtfspy library is mainly used to process the GTFS data and is explained in more detail below. NetworkX is used to represent the networks as graphs. Bokeh in turn is used to visualize the NetworkX representations. The full pipeline is visualized in Figure 4.3.

As mentioned before, gtfspy is the main library used for processing GTFS data and was developed by researchers at Aalto University (Kujala et al., 2018). This library is able

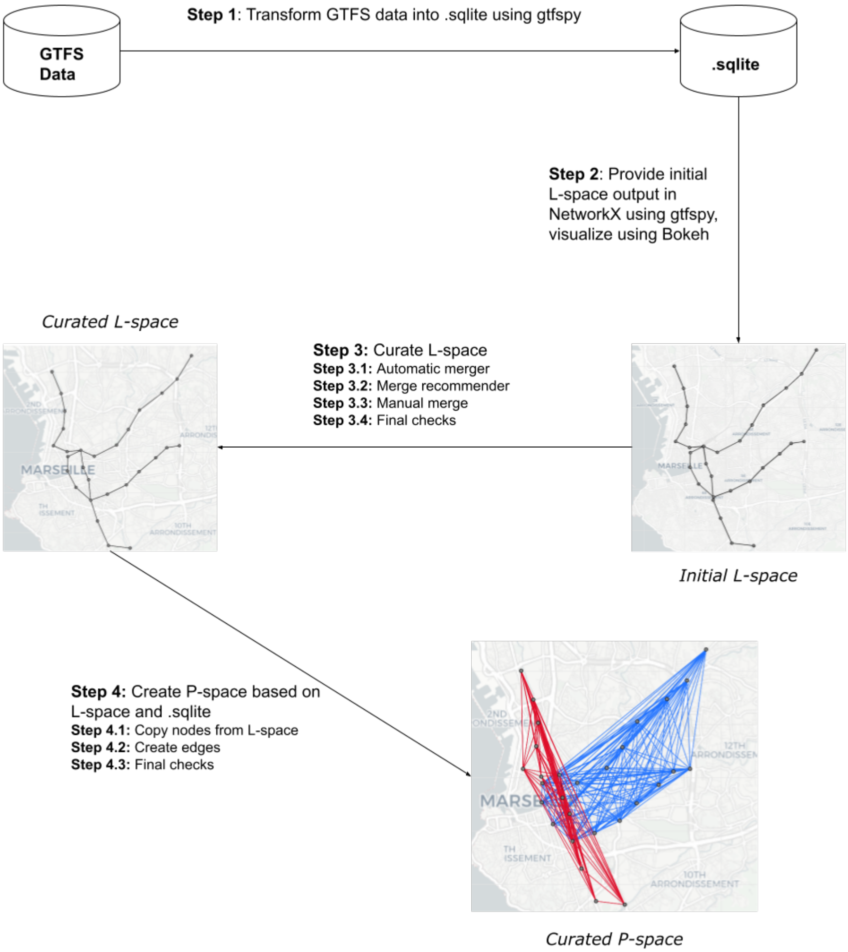


Figure 4.3: The implementation pipeline of this study

to process the GTFS files of a specific network and turn them into a more usable data format. Specifically, the GTFS files are transformed into the .sqlite format (SQLite, 2022) which is a format ideal to perform data analyses on (Step 1). This data can then be directly accessed and used. A variety of pre-made queries exist that can be used to query the data, while it is also possible to query the database with custom queries. In this way, the desired data can be withdrawn from the GTFS data sets without extraneous effort. Turning GTFS files into .sqlite files takes up a varying amount of time depending on the way in which the GTFS files were programmed. More details on this processing stage (including potential failures) for each individual network is provided in Appendix F. In addition, gtfspy provides an initial graph-based L-space output (Step 2). This initial L-space can be further processed into a curated L-space representation (Step 3) which in turn can be turned into a corresponding P-space (Step 4). These final two steps of the process are further explained in detail in Section 4.3.

4.3. IMPLEMENTATION OF L- AND P-SPACE

The two network representations as defined in Section 3.2 need to actually be implemented in order to use them for further analysis. In the following Sections 4.3.1 and 4.3.2, the L- and P-space implementations are explained in further detail. In these explanations, the steps followed from the GTFS file to the L- and P-space outputs are described sequentially. Considering the the practical application of these representations, as well as the cumbersome nature of creating them, the curated L- and P-space representations have been made publicly available at https://doi.org/10.4121/21316824 (L-space) and https://doi.org/10.4121/21316950 (P-space). In order to illustrate the implementation process of these representations, the metro network of Marseille is used as an example, which can be found in Figure 4.4. This network has two lines: M1 (blue) and M2 (red). The lines meet at two stations: Saint Charles and Castellane. These stations consequently are the only stations in the network where a transfer to another line is possible.

Figure 4.4: The metro network of Marseille[1]

[1]Retrieved from: https://upload.wikimedia.org/wikipedia/commons/4/42/M%C3%A9tro_de_Marseille.svg

4.3.1. L-SPACE

Translating a GTFS data set to an L-space representation, requires decisions on the exact translation. While gtfspy has built-in methods which do provide a graph-based output, this output is not verified and contains some implementation decisions. As such, a set of further processing steps is implemented to achieve the desired L-space representation. The first two steps are undertaken by the gtfspy library. All subsequent processing steps have been specifically implemented in this study's pipeline. The steps followed in this process are explained in more detail below.

1. **(gtfspy) GTFS data to .sqlite** - The initial step to take is the processing of GTFS data sets to .sqlite representations. This step is entirely undertaken by the gtfspy library and provides an .sqlite file as output. The duration of this step varies heavily depending on the exact implementation of the GTFS data set but generally lasts between five minutes and one hour per network.

2. **(gtfspy) .sqlite to initial graph output** - In addition to creating an .sqlite file, gtfspy also provides the ability to output an initial graph representation of the network. The output is in the form of an L-space graph representation. As mentioned before, the output is provided as a NetworkX graph by gtfspy. For this representation, both the nodes and edges are explained below. The explanation is concluded with the visualization of this graph output for the Marseille network.

   #### NODES

   In terms of nodes, the representation is fairly straightforward. Each station is simply represented by one single node. Unfortunately, the initial representation provided by gtfspy from the GTFS data is not always this straightforward. This is because some operators program the stations separately for each direction. As an example, take the second station on line M2 in Figure 4.4, named Bougainville. In the GTFS data, this station actually exists twice. This represents the two directions in which vehicles travel on this line and cross this station. Naturally, in actuality this is in fact one and the same station having both north- and southbound platforms. Nevertheless, the graph representation provided by gtfspy outputs both stations as separate stations. Considering the fact that in the desired L-space output, these stations are modeled as one, this is something that needs to be resolved. How exactly this is solved, is explained in step 3 onwards.

   #### EDGES

   The edges are provided as directed edges with three edge labels.

   The first of these labels is the average duration (i.e., in- vehicle travel time) which is calculated based on the provided GTFS scheduling data. This is calculated by gtfspy, taking all trips between two nodes and averaging out their travel time, which is then taken as the duration for that node pair. This duration corresponds to the main edge weight, in-vehicle time, as defined in Section 3.2.2.

   The second label is the number of vehicles (i.e., the frequency) travelling between the two nodes in the given time period. This time period is a specific day and time of day within the schedule of the provided GTFS data and is partly determined by gtfspy and partly by the analyst. The gtfspy library actually has a built-in method that picks a representative day from the full schedule in which at least 90% of the maximum number of trips are run. This can thus be seen as a representative day with normal operation. The analyst in turn can manually decide on a suitable time period to consider trips for. Considering that the peak hours are different per network and the aim of this study is to get a global picture of networks, a time period stretching the whole daytime was chosen. To be more precise, the time period of 05:00 until 23:59 was considered. This therefore excludes potential night schedules, considering not all networks run operations at night.

The third label is the so-called "route_I_counts". This lists the same frequencies as the number of vehicles, but then split per route (or line). This split is especially relevant for networks in which multiple lines run in parallel across the same edge. More details about this label are provided in the explanation of the P-space implementation in Section 4.3.2.

GRAPH OUTPUT

The initial graph output provided by gtfspy for the Marseille network is shown in Figure 4.5. Considering the stations in the network are provided with latitude and longitude coordinates, the graph can be plot directly onto the map of Marseille. In this visual representation it is difficult to see the fact that most nodes are in fact two (or more) nodes. The station of Castellane (the bottom crossing of the two lines) is however clearly two stations that do not overlap while they should in fact be the same station. The first step to resolving this, is the "automatic merge" step.

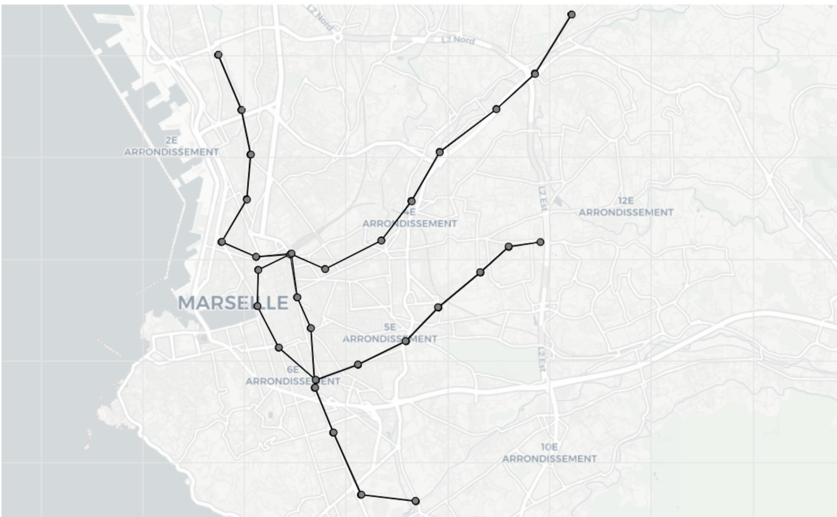


Figure 4.5: The initial L-space representation for the Marseille network

3. **Automatic merge** - In order to ensure that each station is only present once, an algorithm is created that automatically merges stations. The algorithm is fairly straightforward and has two conditions that must both be met in order to merge stations: similarity in name and proximity. In terms of name, the stations need to have the exact same name to be considered for merger. Considering the Bougainville example from Marseille, this is indeed the case. The second condition relates to the geodesic distance between the two stations. If this distance is below a certain threshold, set by the analyst, the stations are merged. For this study the threshold of 200m was used. This algorithm ensures that stations that are clearly the same station are indeed merged. In addition, it prevents accidentally merging stations that have the exact same name, but are not in fact the same station. This is a very rare case but does for example occur in the New York network.

   The merger retains the name and coordinates of the first node. Considering the proximity of these nodes (200m maximum), this is deemed as an acceptable simplification. The output from the algorithm is a list of merged nodes as well as the new graph representation.

4. **Merge recommender** - Unfortunately, this automatic merger step is not sufficient for all networks. Considering that some networks model stations with different names (e.g., Main Street - Northbound and Main Street - Southbound), these stations are not merged by the automatic merger algorithm. In order to ensure that these stations are still merged, the merge recommender algorithm is used. This algorithm works in a similar manner to the automatic merge algorithm, but instead of automatically merging stations with similar characteristics, it will instead recommend to merge those and give the analyst freedom to either agree or disagree. The merge recommender algorithm uses much looser restrictions as the analyst has the possibility to prevent incorrect merges.

   The recommender algorithm consists of two sequential stages with different conditions. The first round does not consider the similarity in name at all and is instead based solely on the proximity. The distance used for this study is 20m. The idea of this stage is that stations that are that close, simply have to be the same station, regardless of what name is given to them in the data. The second stage considers both similarity in name and distance. The similarity in name is modeled using the so-called Levenshtein distance to measure the difference between two strings. For this study, a similarity of 75% is used. In terms of spatial distance, a distance of 500m is used. Both of these values are considerably more lenient than for the automatic algorithm. As such, this captures most cases where merging is required, but also some which should in fact not be merged. The exact effectiveness of this algorithm depends strongly on the manner in which the GTFS data is programmed, but for most networks nearly all relevant merges are captured by using this algorithm.

   An example of the prompt the analyst receives is provided in Figure 4.6. Considering that all relevant stations are already merged after the automatic merge stage for Marseille, the example of Madrid is used instead. In this example, the algorithm prompts the analyst with the question of whether two stations named "Diego de

Leon" should be merged. The stations themselves are indicated on the map in red. Considering the distance between the two stations, the reason these stations were not merged by the automatic merge is most likely because the distance exceeds the set limit. In order to determine whether these stations should indeed be merged, the official map of the metro network should be consulted. If these stations are indeed one and the same, the analyst confirms this in the prompt and the stations are merged. This process is followed for all stations meeting the conditions of the two merge recommender steps. The output provided is once again a list of the merged stations, along with the updated graph representation.

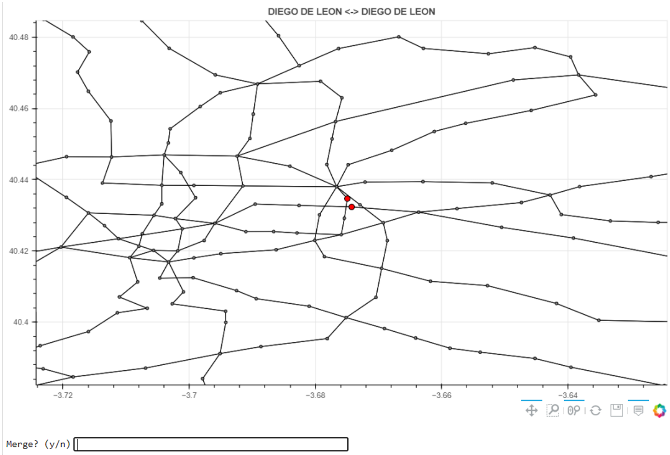


Figure 4.6: An example of a prompt from the merge recommender algorithm

5. **Manual merge** - While the automatic merge and merge recommender capture nearly all stations to be merged, there sometimes are cases which require manual intervention. To facilitate this, the manual merge algorithm is created. This algorithm provides the analyst with a visual interface in which two nodes can be selected for merging.

   The Madrid network is once again used as an example of stations to be merged that are not captured by the earlier two algorithms. An example of stations like this, is provided in Figure 4.7. The two stations of Plaza de Espana and Noviciado appear as two separate stations on two separate metro lines on the map. They are however connected via two black lines, which in the case of Madrid, indicates an out-of-network transfer (with a long walking time specifically). As such, even though these are technically two different stations, the operator treats them as one transfer station and indicates it as such. Considering the importance of transfers in the calculation of shortest paths, it is essential to include transfer possibilities such as these in the implementation as well. As such, the analyst can manually decide to merge these two stations. The basis on which to merge stations such as these differs per network. The general rule applied however is that if the official map, provided by the operator themselves, indicates a transfer between stations is possible, the stations are merged. More detailed information per network on exactly which stations are and are not merged, along with the argumentation, can be found in Appendix F. Similarly to the other merging algorithms, the manual merge algorithm also provides an overview of the merged stations along with the final L-space representation.

Figure 4.7: A separated transfer station in the Madrid network[2]

4

6. **Final checks** - After the three merging steps, the L-space is generally considered to be verified and completed. In order to further confirm its correctness, three final steps are taken. Firstly, it is checked whether the network consists of one connected component. With very few exceptions, metro networks are generally one connected component. As such, if the final L-space graph is not one connected component, a mistake is generally present. It sometimes occurs that some stations are out-of-order or ghost stations are present in the network, thus causing disconnected components. These stations are then deleted. Secondly, the bi-directionality of the network is confirmed. Generally, each edge should exist in both directions. The example presented in Figure 3.3 shows that exceptions do exist, which should be taken into account. This second check confirms whether each edge exists in both directions and outputs those that do not. The analyst can consequently manually confirm (using the official map) that the presented exceptions are in fact correct. Lastly, a manual inspection must be performed to confirm each station is indeed present in the L-space representation. This is done through a side-by-side comparison of the official map and the L-space representation. After these final checks, the curated L-space is stored for further usage.

4.3.2. P-SPACE

Following the creation of the L-space representation, the P-space representation can also be created. Fortunately, the curated L-space representation can be used to facilitate this process. The process of creating this P-space representation is described sequentially below.

1. **Copy nodes from L-space** - In order to create the nodes for the P-space representation, the nodes of L-space are simply copied. Considering that set of nodes is already verified, and that the representations use the same stations as nodes, no further steps are needed.

[2] Adapted from: https://www.metromadrid.es/sites/default/files/documentos/Viaja%20en%20Metro/Planos/Planoesquematico.pdf

2. **Create edges** - The creation of the P-space edges is a much more intricate process and requires further explanation. As explained in Section 3.2.3, nodes in P-space have edges between them when they are connected by a line. In order to create these edges, firstly the list of lines is retrieved from the GTFS data. Considering the fact that P-space also uses a directed implementation, these lines are split into two: one for each direction. The lines are then considered one by one and edges are drawn between all stations on a line. To retrieve the desired weight for an edge (the average waiting time for that line), a few more steps are necessary. When a network does not have parallel lines, this is a fairly simple process. As described in Section 4.3.1, the number of vehicles between each node in L-space is known, which is required to calculate the average waiting time. In order to determine what the approximate number of vehicles is between two edges that are not directly connected in L-space (but are on the same line and hence connected in P-space), two edges from L-space are required. These are the outgoing edge from the first node in the direction of the second node, and the incoming edge from the direction of the first node to the second node. The lowest number of vehicles is then chosen as that is the maximum number of vehicles possible between these two nodes. The average waiting time is then calculated as described in Section 3.2.3 and used as the edge weight.

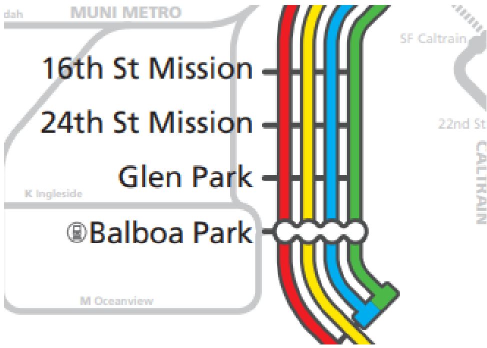


Figure 4.8: Four lines running in parallel in the San Francisco BART system[3]

For nodes that have multiple lines between them, this calculation is slightly more complicated. Consider the stations of Saint-Charles and Castellane in Figure 4.4. To get from one to the other, either the blue or red line can be taken. Another, more common example, is the one in the San Francisco network as can be seen in Figure 4.8. Here four lines run in parallel between four stations. Nevertheless, only one edge (for each direction) will exist between any two stations in P-space. For this study, it was decided that the number of vehicles for parallel lines are added together, and thus a lower average waiting time is achieved. This relies on the assumption that travellers will take the first train that arrives regardless of the line it runs on. For the San Francisco example, this is a sensible assumption considering

[3]Based on: https://www.bart.gov/sites/default/files/images/basic_page/system-map-everyday-until-9pm.png

the vehicles arrive at the same platform and taking the first vehicle is the fastest way to reach the next station. For the Marseille example however, this assumption is less valid. While Saint-Charles can be reached from Castellane using both the blue and red line, these lines are not parallel at Castellane, but actually cross at a right angle. As such, the red and blue line vehicles arrive at different platforms at this station. It is therefore much more difficult to take the first vehicle that arrives, considering the vehicles arrive at different platforms. For this case the expected behavior for a traveller is most likely to wait at the red line considering that it most likely has a shorter in-vehicle travel time (only having to cross two stations as opposed to the blue line's three). Considering the fact that the parallel case of the San Francisco network is more common than the orthogonal case for the Marseille network, the number of vehicles for both lines are added together. In terms of the visualization, edges are created in the official color of their line as much as possible (i.e., this data is frequently present in the GTFS data but not always). When multiple lines run between two stations, these edges are instead made black to indicate this. The created P-space representation for the Marseille network is provided in Figure 4.9.

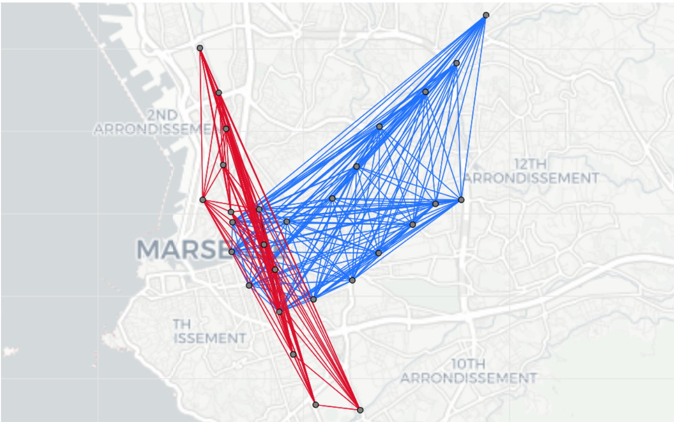


Figure 4.9: The P-space representation for the Marseille network

3. **Final checks** - Finally, the P-space representation should be manually verified in order to ensure no mistakes are present. This can be done by checking the waiting time for a random selection of edges and confirming it is within expected ranges. In addition, the presence of black edges where expected can also be used to confirm the correctness. After these final checks, the P-space can be saved for further calculations.

4.4. SHORTEST PATH CALCULATION

As described in Section 3.3, the primary indicator of this study relies on shortest path calculations. Considering that the exact implementation of these calculations can differ, more details about it are provided in this section. Firstly, the manner in which the attribute coefficient values are determined is briefly explained in Section 4.4.1. Afterwards, the shortest path implementation itself is explained in Section 4.4.2.

4.4.1. ATTRIBUTE COEFFICIENT VALUES

The shortest path travel time in this study consists of a combination of the in-vehicle travel time, waiting time and number of transfers. As explained in Section 3.3.2, both the waiting time and the number of transfers are weighted using a constant, positive integer penalty/coefficient value. This coefficient translates the waiting time and number of transfers to a representative amount of in-vehicle time. Considering that the exact value of these coefficients has been studied extensively in other works, this study does not aim to redefine these values. Instead, based on these other studies, representative values are chosen. Since the topic has been studied throughout the past two decades, various studies from throughout this period are taken (Lee and Vuchic, 2005; Guo and Wilson, 2011; Garcia-Martinez et al., 2018; Jara-Diaz et al., 2022). Lee and Vuchic (2005) researched different penalties per transfer to discover how it affected the transit demand and willingness to transfer. They use penalties varying between 0 and 30 minutes per transfer. The researchers do not draw any conclusions about which of these values is most realistic, but simply evaluate the effects on demand, logically decreasing with increasing transfer penalties. Guo and Wilson (2011) actually research the topic specifically for the London Underground, which logically has a large overlap with this study. Their study concludes that in the London Underground this penalty is around 5min per transfer. They do however mention a lot of factors which cause this value to be much lower than most likely is actually experienced by travellers. Garcia-Martinez et al. (2018) try to form a more generic value for waiting and transfer penalties in multimodal transit networks. Their conclusion is that the average transfer penalty is between 15.2 and 17.7 for the multimodal network of Madrid. Four years later, this work was followed upon by the same authors (Jara-Diaz et al., 2022). This latest study further evaluates the earlier results for Madrid and also compares them to other studies. They conclude 13 to 18 minutes to be the most realistic range for the transfer penalty for planning purposes, while their investigation into related works show 2 to 3 minutes to be most realistic as a penalty for the waiting time. Considering the relevance of their study compared to this one and its recent nature, these values are deemed most realistic. For this study, the lower bounds of both of these values are taken, thus two in-vehicle equivalent minutes per minute waiting time and thirteen in-vehicle equivalent minutes per transfer. This is because other studies (among Guo and Wilson (2011) and Lee and Vuchic (2005)) show significantly lower values and this study does not focus on multimodal networks, but on metro networks only. As Jara-Diaz et al. (2022) also show, transfers are experienced better by travellers within the same mode, considering it is not necessary to leave the system. Since these are the only transfers in this study, it is sensible to take the lower bound values.

4.4.2. IMPLEMENTATION

The two penalty values of two and thirteen minutes for waiting time and number of transfers respectively are consequently used in the implementation of the shortest path calculation. As explained before in Section 3.3.2, the in-vehicle time is calculated in L-space while the waiting time and number of transfers are calculated in P-space, which are then combined into the total travel time. Three sequential steps are followed for this process, which are described below. These steps are also visualized in Figure 4.10 for two example nodes in the network of Marseille.

1. **Calculate k-shortest paths in L-space** - Firstly, the actual travel paths between two nodes in the network must be determined. For this, the L-space is necessary as that contains information on the network topology not present in P-space. Since the information about waiting time and transfers is not present in L-space, these conditions cannot be used to calculate the shortest path in L-space. The shortest paths in L-space are purely based on the in-vehicle time. As such, the shortest path in L-space might not actually be the total shortest path when also accounting for waiting time and transfers. To mitigate this as much as possible, the k-shortest paths from L-space are considered as total shortest path candidates. For this study, a k of 5 was used. This value is a balance of considering as many shortest paths as possible while also keeping the runtime of the algorithm restricted to reasonable values. Using this k, for 90% of the networks, all possible paths are actually considered. Take for instance the example of Marseille in Figure 4.10: between these two nodes only two possible paths exist. For the few networks where more than five paths between some nodes exist (e.g., New York, Madrid, Paris), a fair number of all possible shortest paths is still considered. The runtime of the algorithm for these networks is somewhere between 10 minutes (Paris) and an hour (New York). Unfortunately, this means not all shortest paths are considered and the actual shortest total path might not be chosen. This is deemed an acceptable simplification considering the individual shortest paths are not used heavily in the further analysis, only an average value.

2. **Calculate corresponding P-space information for k paths** - For these k-shortest paths from L-space, the total shortest path travel time must be calculated. This is done by finding the corresponding waiting time and number of transfers in P-space for the different paths between these nodes.

3. **Pick shortest total path** - From the k-shortest paths, the one with the shortest total travel time is consequently chosen as the shortest path between the two nodes.

The example in Figure 4.10 shows how the inclusion of the waiting time and number of transfers from P-space changes which path is the shortest. While the orange path is the shortest based on in-vehicle time in L-space, it requires an extra transfer and a longer waiting time when also considering P-space. This extra transfer and more waiting time, along with the penalties assigned for those, means the light blue path actually has a shorter total travel time than the orange path.

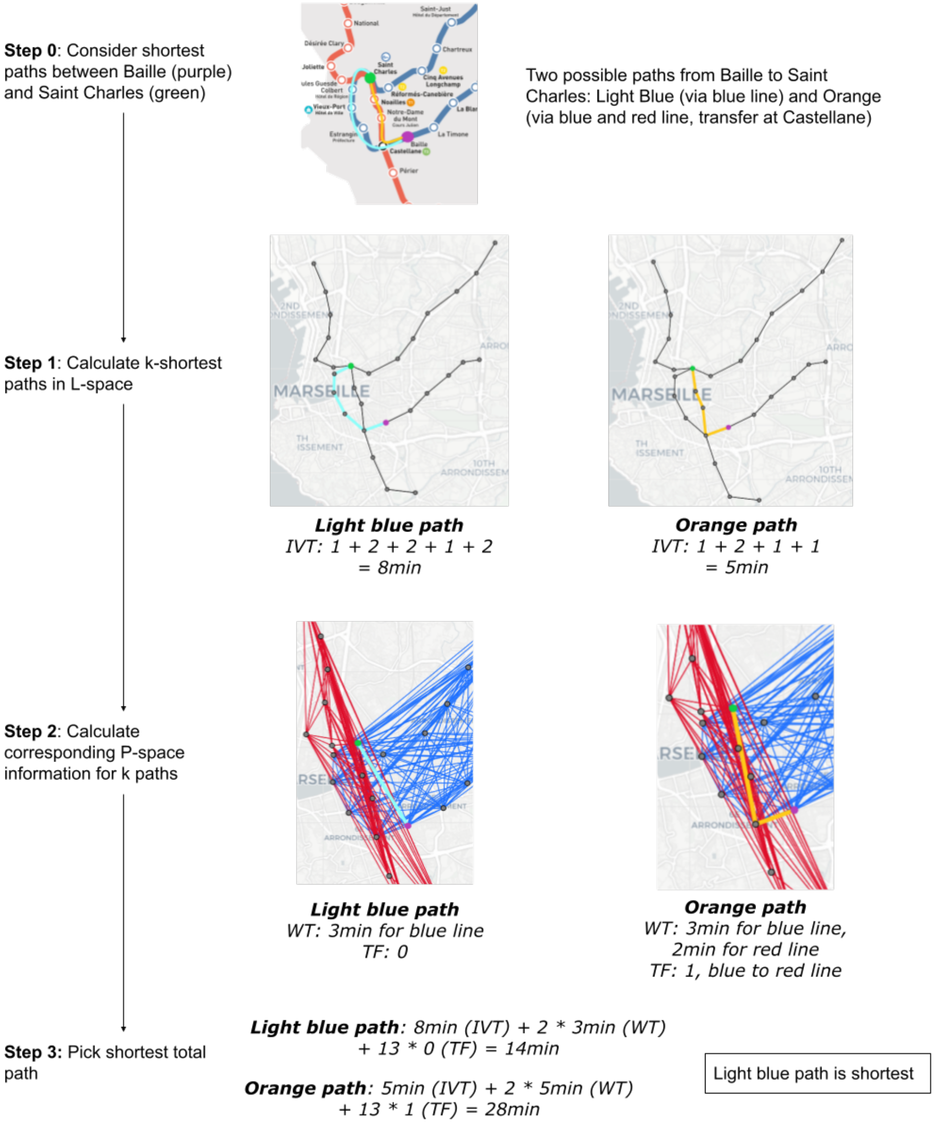


Figure 4.10: A visualization of the shortest path calculation

4.5. ADDITIONAL DATA

In addition to the data calculated based on the GTFS data, some other data is also necessary for this study. The different data sets and how they are retrieved is briefly discussed in this section.

4.5.1. ADDITIONAL NETWORK INFORMATION

While most network information comes directly from the GTFS data, the official network maps are used to confirm this information. These maps are, as much as possible, retrieved from the websites of the operators of the networks to ensure the intention of the operator is captured as much as possible. Examples of relevant information in these maps is the stations that can be used for transfers, the lines of the network or the existence of express lines.

4.5.2. POPULATION

The population information for each city is retrieved from a few different sources. Note should be made that the exact demarcation of a city or urban area is not always easy to define. As such, the intention is to use only a few databases with similar data to ensure the data is properly comparable between networks. In addition, official government resources were used as much as possible. Considering that the vast majority of networks in this data set are in either Europe or North-America, databases for those regions are most important. For Europe, the official EU database Eurostat is used (https://ec.europa.eu/eurostat). The data from 2019 is used as that is most complete for all cities in this study. Consequently, for the other databases 2019 is also taken as the representative year (where possible). The population data for the United States was taken from the official 2020 Census via www.censusreporter.org. For the remaining networks, a variety of different sources were used. The exact sources used can be found in Appendix G.

4.5.3. GDP

The GDP data for each city was more difficult to find. Considering that GDP is primarily calculated for countries, not individual cities, this data is simply less available. The intention for this data is similar: to use a low amount of sources, preferably governmental. Once more, a caveat should be made about the exact demarcation of cities. For especially the North-American cities, the GDP is usually of a much larger metropolitan region than just the individual city. This naturally might inflate the GDP values for those cities compared to cities from other regions. The exact sources used can also be found in Appendix G.

5

Results

In this chapter the results from the study are described in detail. Firstly, the primary indicator based on the total travel time is discussed in Section 5.1. Afterwards, the correlations between the travel time and network- and city-related factors are discussed in Section 5.2. Based on these correlations both a simple as well as multiple regression model is created in Sections 5.3 and 5.4 respectively. Further comparisons between the travel time and other factors are briefly described in Section 5.5. In Section 5.6, this chapter is concluded with a benchmark analysis comparing the total travel time methodology applied in this study to other state-of-the art methods.

5.1. Primary indicator

In this section, the results based on the primary indicator used for comparison (as described in detail in Section 3.3) are provided. Firstly, the representative value taken as an indicator of the total travel time is described in Section 5.1.1. Afterwards, in Section 5.1.2, the exceptions discovered for some networks are shortly described.

5.1.1. Representative value

The primary indicator in this study is based on the total shortest path travel time between each node pair in the network. Whereas most studies take either simple hops or only in-vehicle time, this study uses the total travel time which is a combination of the in-vehicle time, waiting time and transfers, providing a more realistic image of the actual travel time experienced by travellers. In order to be able to compare networks based on this metric to other factors (e.g. total network length), all shortest path travel times in a network must be distilled into one value. In order to gain a representative value for these travel times in networks, the distribution per network must be analyzed first. This distribution of the shortest total travel time between all node pairs is captured in a histogram for each individual network. These histograms for all 51 networks can be found in Appendix E (N.B.: In order to display figures that can be directly compared, the travel time is normalized in terms of the maximum travel time of that network in Appendix E).

Looking at the figures in Appendix E, it becomes immediately obvious that not all networks follow the same distribution of travel time in the figure. The four examples in Figure 5.1 clearly show this variation (N.B.: The distributions in Figure 5.1 are presented unnormalized, with the total travel time on the x-axis). Whereas London could be classified as normally distributed, Cleveland is exponential, while Lisbon is bimodal and Vancouver is quite difficult to define in terms of its distribution.

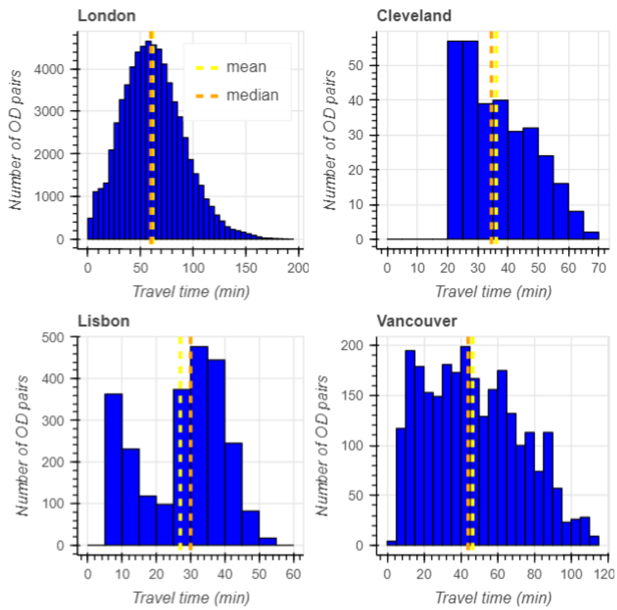


Figure 5.1: Four networks with different travel time distributions

To represent these distributions with one value, there are various options such as the mean, median, mode, standard distribution and variance. For the intended comparisons in this study, the mean or median would make most sense as these indicators explicitly retain the information about the absolute values of travel time the best. These two options are also less sensitive to outliers than the mode, which also retains information about the absolute values. In Figure 5.1, the mean and median are indicated with vertical dashed lines in yellow and orange respectively. As described before, these figures follow significantly different distributions, most being not normally distributed. Therefore, the mean is less suitable as a representative point as it might be skewed in distributions other than normal distributions. This indeed becomes clear from Figure 5.1 where the mean and median overlap for London, but do not for the other three networks. As such, the median is chosen as the single data point to represent these distributions.

The median travel time for each network can be found in Figure 5.2. Each network in the data set is represented using a colored dot to indicate the value of the median travel time. The lighter (i.e. yellow, orange) dots represent shorter travel times whereas the darker (i.e. red, purple) dots represent longer travel times. The detailed legend can be found in Figure 5.3. Considering the high density of networks in Europe and the eastern part of North-America, each region is also displayed separately in Appendix B, which also includes the exact travel time values for each network.

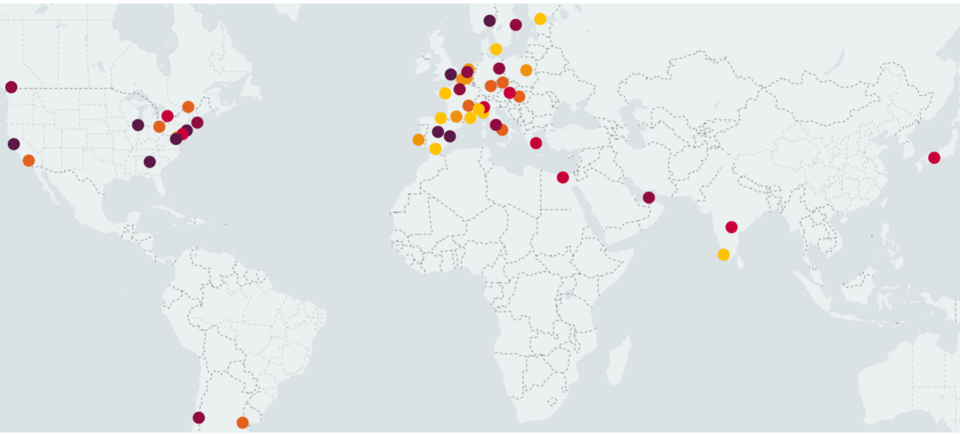


Figure 5.2: The median travel time for all 51 metro networks

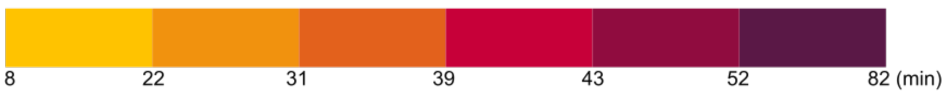


Figure 5.3: The legend for Figure 5.2 detailing the color for each range of values of the median travel time

5.1.2. EXCEPTIONS

Unfortunately, the different ways in which operators implement their GTFS data in combination with the implementation decisions of this study means that the results of the shortest path algorithm sometimes contain mistakes. There is a set of six networks with these mistakes which can be found in Figure 5.4. The difference in coloring in these figures is caused by the different number of bins that have to fit in figures of the same size(i.e., Paris has about 40 bins while New York has 160).

As becomes clear from Figure 5.4, these networks have some very unrealistic travel times for some OD-pairs up to 800 minutes in the case of New York. While New York is a large network, a travel time of over ten hours is clearly unrealistic. The reasons for these exceptionally long travel times are based on exceptionally long waiting times, as becomes clear from Figure 5.5. These anomalies in waiting time stem from the P-space implementation for this study. Specifically, a high waiting time is directly related to a line with a low frequency. The details of exactly which lines this entails for each individual network and why these mistakes appear in the implementation, is described in more detail in Appendix F. Fortunately, the median is not very susceptible to outliers such as these and thus these exceptions provide no hindrance for the further analyses.

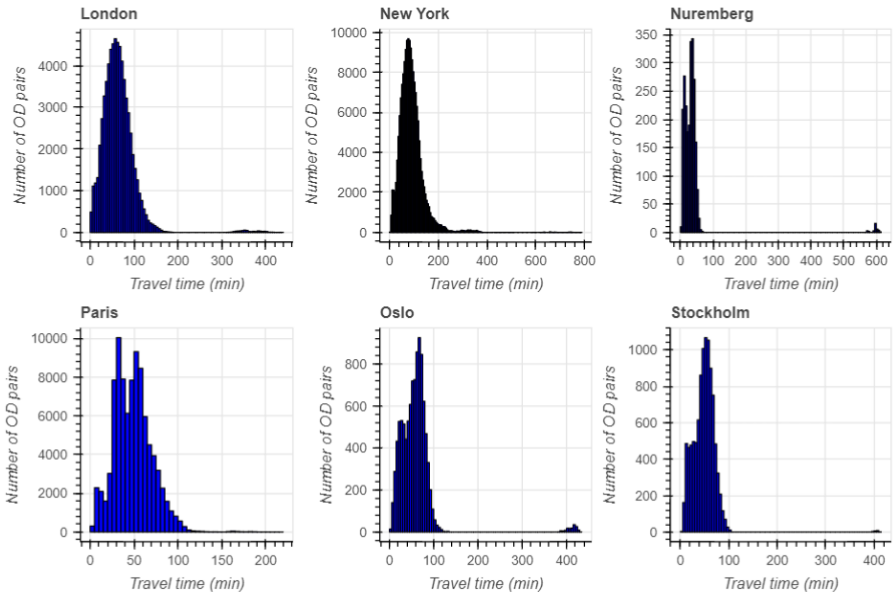


Figure 5.4: Six networks that have anomalies in their total travel time data

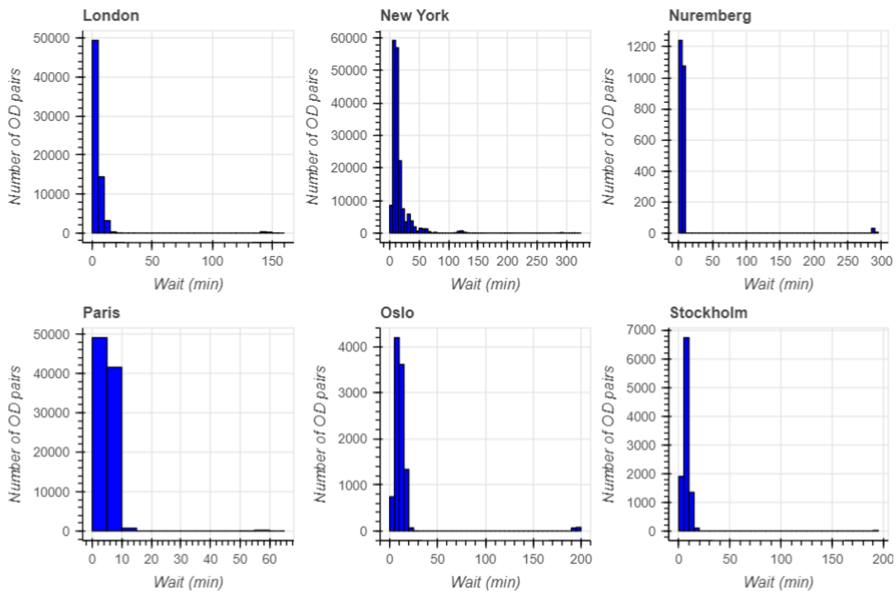


Figure 5.5: Six networks that have anomalies in their waiting time data

5.2. CORRELATIONS BETWEEN THE INDICATORS

In this section, the correlations between the indicators used in this study will be explored. Considering the fact that the relationship between the different indicators is unknown, both the Spearman and Pearson correlation will be explored in Sections 5.2.1 and 5.2.2 respectively. The individual values of the different indicators for each network can be found in Appendices C and D.

5.2.1. SPEARMAN

The first correlation to explore is the Spearman correlation. The reasoning for first exploring Spearman is twofold. Firstly, there is no reason to assume the different indicators used in this study are distributed normally. As the Pearson correlation assumes this to be the case, while Spearman does not, Spearman is preferable. Secondly, whereas Pearson only tests linear correlation, Spearman is a bit more lenient and tests for a monotonic correlation. Considering the fact that the discovered correlations might not necessarily be linear, this too favors the usage of the Spearman correlation. The Spearman correlation between the different indicators can be found in Figure 5.6.

The first thing to note from Figure 5.6 is that all factors are positively correlated. This is in line with expectation, considering all factors relate to the increase of either the network or the city, which has been hypothesized to correlate with an increasing average travel time. Secondly, most correlations are fairly strong with only a few having a Spearman correlation of 0.50 or lower.

The total travel time is especially strongly correlated with network length, having the second-highest correlation in the whole heatmap. The highest correlation is actually between the network length and the number of stations with a Spearman correlation of 0.91. This is fairly sensible considering how both of these indicators are descriptors of the network size and are very directly related. An increase in stations generally leads to an increase in the network length and viceversa. The number of lines actually has a surprisingly high correlation with both the number of stations and network length, being 0.86 for both. This correlation is notably weaker with the travel time, being only 0.75. Interestingly, the average direct station distance has the strongest correlation with travel time, and a weaker one with the other six factors. This is fairly sensible considering how both the travel time and average direct station distance describe the average shortest path, in terms of duration and direct distance respectively. However, a key difference is that the average direct station distance does not account for the infrastructure of the network but instead considers the direct (i.e. geodesic) distance between stations.

Overall, it can be concluded that all four network factors have a strong positive correlation with the travel time, all having a Spearman correlation of 0.75 or higher. The network length clearly has the strongest correlation and can thus best considered as the best single explanatory parameter for the total travel time.

The travel time also has a positive correlation with both the population and GDP. Both of these correlations are however notably weaker than the network correlations, with 0.65 for population and 0.57 for GDP. This is fairly sensible, considering that these two city aspects and the travel time in the metro network are not directly related. In fact, the correlations discovered here are both most likely spurious and instead are correlations between the city factors and network size. Interestingly however, the correlation

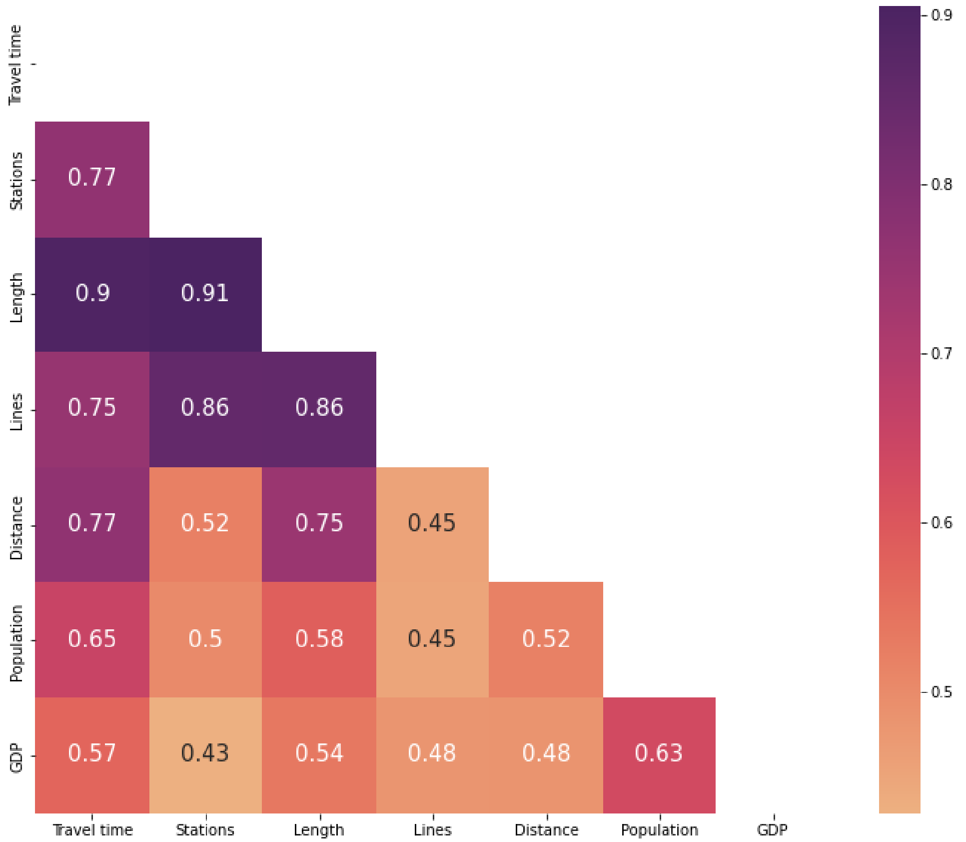


Figure 5.6: A heatmap of the Spearman correlations between the seven indicators

between travel time and population is actually the strongest correlation for population, being even higher than that between population and GDP. Out of all correlations with GDP, GDP and travel time is actually the second highest. The correlation with travel time being comparatively high with both city indicators suggests a connection between the travel time and the city indicators independent of the network size. More in-depth research will need to be performed to draw conclusions about this relationship.

5.2.2. PEARSON

Considering the strong positive Spearman correlations found in Section 5.2.1, the expectation is that the Pearson correlations will show a similar pattern. The Pearson correlations between the different indicators can be found in Figure 5.7. From this figure it becomes clear that the same strong positive correlations indeed exist in terms of Pearson correlation as well. Logically, some differences in the exact height of the correlation do arise, especially for the GDP compared to the other factors. For the most important correlations, those between the travel time and the other factors, there are no notable differences. These high Pearson correlations between the travel time and other indicators, as well as some low correlations among the other indicators themselves, mean there is evidence to use a regression model to estimate the travel time.

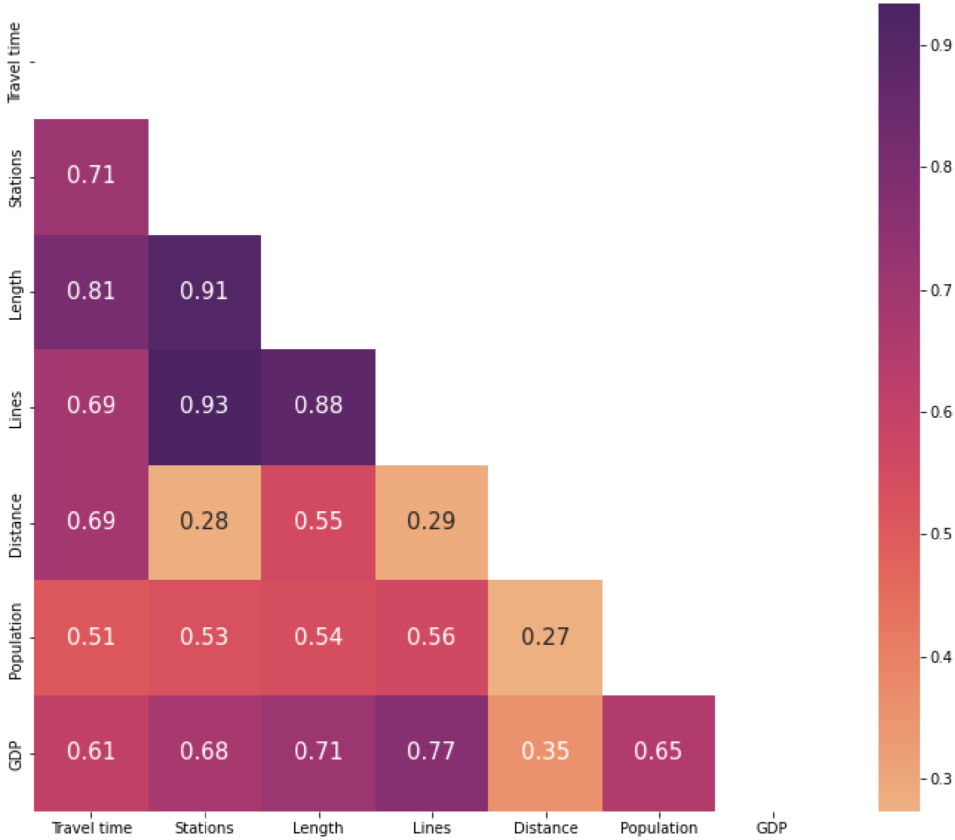


Figure 5.7: A heatmap of the Pearson correlations between the seven indicators

5.3. SIMPLE LINEAR REGRESSION

Considering the high correlation between the travel time and the network/city factors, it is sensible to try and estimate the travel time using these factors. In this section, the possibility of linear regression using a single independent parameter will be explored.

The Pearson correlation heatmap presented in Figure 5.7 naturally provides the best indication of what secondary parameter to use as an indicator for the travel time. The network length has the highest Pearson correlation of 0.81 and can thus best be used as a single estimator of the travel time. An Ordinary Least Squares (OLS) regression model is applied to find the best fit between the dependent and independent variable. To compare this model to others, the R-squared, adjusted R-squared and Bayesian Information Criterion (BIC) are used. Both the adjusted R-squared and BIC account for an increase in the number of parameters and thus help avoid overfitting and ensuring the model is parsimonious. The performance of this simple regression model on three criteria can be found in Table 5.1, while the coefficient values and information about their significance can be found in Table 5.2. The best fit line between the travel time and network length can be found in Figure 5.8.

Criterion	Value
R-squared	0.653
Adj. R-squared	0.646
BIC	375.2

Table 5.1: The performance of the simple regression model

Parameter	Coefficient	P-value
Intercept	26.990	0.000
Length	0.132	0.000

Table 5.2: The parameters of the simple regression model including coefficients and p-values

As expected from the high correlation, this linear regression model fits reasonably well. The R-squared of this regression is 0.653 and both the intercept and the coefficient for network length are significant at a 5% confidence interval. From Figure 5.8 it becomes clear however that this model is far from perfect. While the line fits fairly well, there is still a fair number of outliers. To be precise, there are ten networks that are either over- or underestimated by more than 10 minutes in this model. The five networks that are overestimated by this model are Rennes (-20.1), London (-18.5), Genoa (-14.8), Turin (-13.8) and Helsinki (-10.2). On the other hand, Oslo (+19.5), Atlanta (+19.25), Boston (+15.3), Chicago (+14.0) and San Francisco (+11.2) are the networks that are most strongly underestimated. These networks are indicated in Figure 5.8 using the abbreviation labels from Appendix C. Their outlier nature is discussed briefly below. In addition, the regional differences that can be identified in Figure 5.8 are also shortly described.

5.3.1. OUTLIERS

The first outliers to discuss are the four networks towards the lower bottom of Figure 5.8: Rennes, Genoa, Turin and Helsinki. All of these networks have a fairly short network length but an especially low average travel time. In order to explain these relatively low travel times, the total travel time is split up into the three components, which can be found in Figure 5.9. From Figure 5.9a, it becomes clear that Rennes and Genoa have the lowest in-vehicle travel time out of all networks while Turin and Helsinki also have a relatively low in-vehicle travel time. A similar note can be made about the waiting time from Figure 5.9b. Lastly, the average number of transfers made can be noted in Figure 5.9c. From this figure, it becomes clear that Rennes, Genoa and Turin have no transfers at all (since they have only one line) while Helsinki has a very low average number of transfers. Helsinki has such a low average number of transfers since it has only two lines that run in parallel over the same infrastructure for nearly 75% of their length. Combin-

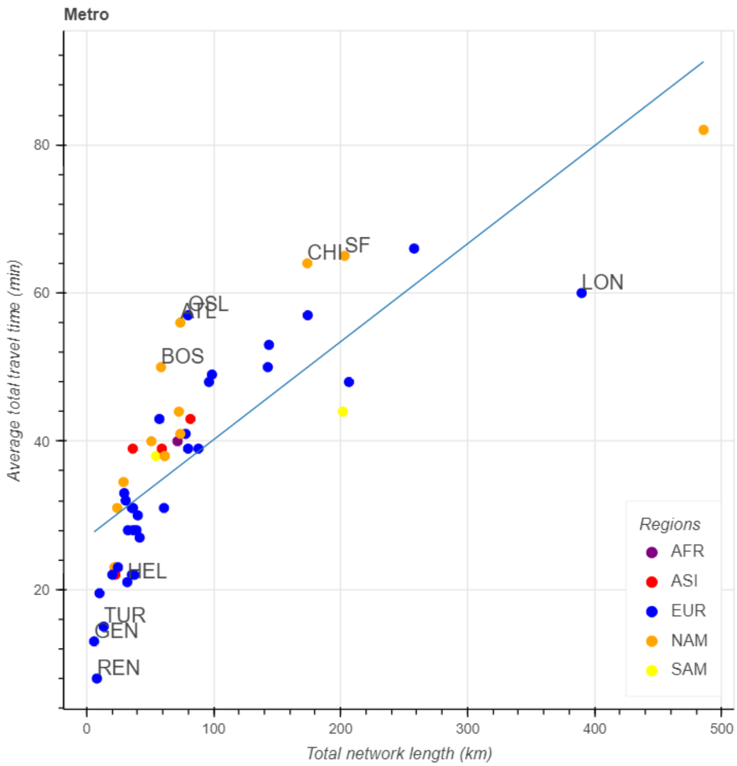
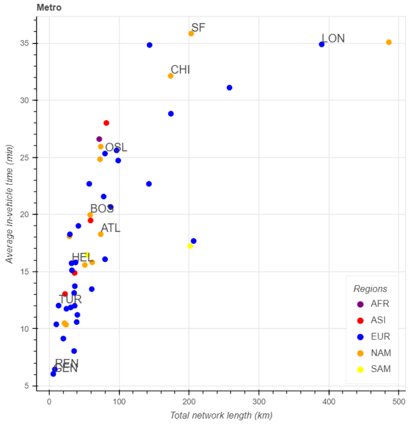


Figure 5.8: The best fit line between the total travel time and the network length

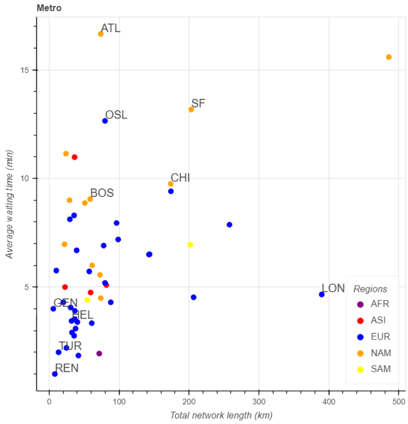
ing the data points from these three figures, it becomes clear that none of these networks have notably lower performance on one of the components than the others. Instead, the networks perform very well on all of these components, which combined leads to them having a much lower travel time than estimated by the linear regression model.

The fifth network that is significantly overestimated by the model is London. This is interesting to note considering London has a much larger network length than the other four overestimated networks. Looking firstly at Figure 5.9a, it becomes clear that London does not have a notably low in-vehicle travel time. On the contrary: it has the second highest in-vehicle travel time out of all networks. The same can be noted for the number of transfers in Figure 5.9c, where London has the fourth highest number of transfers. Figure 5.9b shows a completely different image however, with London having an exceptionally low waiting time, especially considering its size. Since the waiting time is weighted twice as heavily as the in-vehicle time in this model, this relatively low waiting time compensates for the higher in-vehicle time and number of transfers.

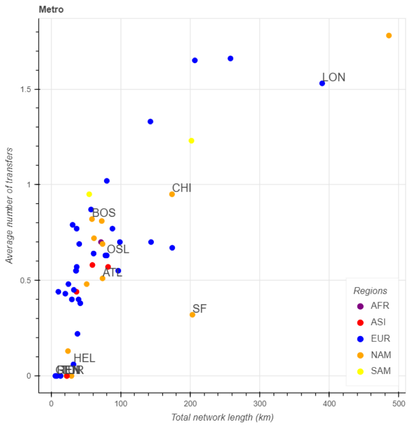
The outlier nature of the five most underestimated networks is best explained by the waiting time in Figure 5.9b, since these networks are among the networks with the highest average waiting time. Considering the relatively higher weight of the waiting time, it is sensible that this significantly influences the total travel time of these networks. In



(a) In-vehicle time



(b) Waiting time



(c) Transfers

Figure 5.9: The travel time versus the network length, split into three components

terms of in-vehicle travel time and number of transfers, Oslo, Boston and Atlanta do not perform notably well or poorly. San Francisco and Chicago on the other hand have a very high in-vehicle travel time, while Chicago also has a relatively higher number of transfers. San Francisco interestingly actually has a comparatively low number of transfers, especially for its network size. However, this low number of transfers is not sufficient to compensate its high in-vehicle and waiting time.

5.3.2. REGIONS

Besides the specific outliers mentioned in the previous section, a few notes can also be made about the more structural regional differences that can be identified in Figure 5.8. Firstly, it is interesting to note how ten out of the twelve North-American networks are underestimated by the model. It can thus be concluded that North-American networks have a relatively high total travel time compared to other networks. The opposite can be said for European networks, considering nearly all of the networks overestimated by the model are in fact European. The other three regions have a relatively low number of networks and no notable outliers. The travel time components from Figure 5.9 can help provide more insight into where exactly this difference arises. Once more, the in-vehicle time (Figure 5.9a) and number of transfers (Figure 5.9c) provide no real insight, with no clear distinction between the performance of networks by region. The waiting time in Figure 5.9b however provides more insight. Eight out of the twelve North-American networks are among the networks with the highest waiting time. As such, it becomes clear that the waiting time is the main contributor to North-American networks' relatively long total travel time. Considering this waiting time is directly derived from the networks' frequency, North-American networks have a relatively low frequency compared to other networks. This apparent regional differentiation can also be accounted for in the model using dummies for each region. Experimentation shows that only Europe and North-America provide significant parameters (at a 5% level), when considered separately. The results for different models including North-America and Europe as regions separately as well as combined can be found in Table 5.3. From this table, it becomes clear that as expected, including either region improves the model slightly. The effects of North-America are slightly stronger than those of Europe, which is in line with the intuitive findings from Figure 5.8. The BIC of the model including North-America is 3.3 points higher which, while positive, is not especially strong.

Region	R^2	Adj. R^2	BIC	Coeff.	P-value
None	0.653	0.646	375.2	-	-
North-America	0.699	0.686	371.9	7.70	0.009
Europe	0.695	0.682	372.6	-6.38	0.014
N-Amer. + Eur.	0.704	0.685	374.9	5.04, -3.23	0.223, 0.368

Table 5.3: A comparison of regression models including regional dummies

5

5.4. MULTIPLE LINEAR REGRESSION

In order to improve the predictive power of the model, a model with multiple parameters is considered in this section. While including more parameters generally improves the predictive power of models (since more of the variance can be explained), a caveat should be made about this. Firstly, including more parameters can lead to overfitting and as such should be done with care. As mentioned before, the adjusted R-squared and BIC account for this. Secondly, including more parameters is only sensible if they are not too highly correlated amongst themselves. High multicollinearity means the independent variables are in fact partly dependent on each other and will make the model volatile. As can be seen in Figure 5.7, in addition to high correlations between the travel time and the other factors, the other factors are also highly correlated amongst themselves. Naturally, this is sensible especially for the network factors considering all of these are an indicator of the network size and thus directly linked. This high multicollinearity thus makes it difficult to combine these factors with multiple linear regression.

5.4.1. FULL MODEL

The initial multiple linear regression model is one including all six explanatory factors from Figure 5.7 along with the categorical factor "Region". The results of this regression model can be found in Tables 5.4 and 5.5.

Criterion	**Value**
R-squared	0.801
Adj. R-squared	0.751
BIC	382.3

Table 5.4: The performance of the full multiple regression model

Parameter	**Coefficient**	**P-value**
Intercept	-8.282	0.658
Stations	0.078	0.208
Length	0.0152	0.755
Lines	0.419	0.557
Distance	1.631	0.001
GDP	-0.019	0.141
Population	1.236	0.159
Asia	20.486	0.219
Europe	23.399	0.204
North-America	29.228	0.105
South-America	13.222	0.283

Table 5.5: The parameters of the full multiple regression model including coefficients and p-values

Comparing the results from the full model in Table 5.4 to those of the simple model in Table 5.1, the full model as expected has a higher regular as well as adjusted R-squared.

Comparing the BIC however, shows that even though the full model has higher explanatory power, it uses many more parameters, leading to a higher (i.e., worse) score. Additionally, with the exception of the average direct station distance, none of the parameters in Table 5.5 are significant to the 5% confidence interval. This is sensible considering the large multicollinearity in the data set. From this, it becomes clear that including all parameters does not provide an improved model.

5.4.2. IMPROVED MODEL

Instead, several other models are explored with far fewer parameters. It is most sensible to consider parameters that have a high correlation with travel time but a low one with each other. Figure 5.7 shows that the number of stations and average direct station distance are two good candidates. These factors have a correlation with the travel time of 0.71 and 0.69 respectively while they have a notably low correlation of 0.28 amongst each other. The results of a model including these two parameters can be found in Tables 5.6 and 5.7.

Criterion	Value
R-squared	0.768
Adj. R-squared	0.758
BIC	358.6

Table 5.6: The performance of the two-factor regression model

Parameter	Coefficient	P-value
Intercept	15.527	0.000
Stations	0.109	0.000
Distance	1.819	0.000

Table 5.7: The parameters of the two-factor regression model including coefficients and p-values

From Table 5.6 it becomes clear that while the two-factor model has a lower R-squared than the full model, it has a better adjusted R-squared and much better BIC. In addition, all three parameters are significant at the 5%. As a network with no stations and no length does not exist, the coefficient value of the intercept cannot be properly interpreted. What can be concluded however, is that metro networks have a base total travel time of 15.5 minutes increasing by 0.1min per extra station and 1.8min per extra kilometer of average direct distance between stations. This means that as networks get bigger by adding more stations and getting more spread out, the travel time increases accordingly. This finding is in line with that of other studies such as Luo et al. (2019).

This improved model can also be used to make estimations for the total travel time of the networks in the data set. By comparing these estimated travel times to the actual travel times, an image can be gained about the accuracy of the model as well as which networks perform well or poorly based on this two-factor model. The difference between the model estimation and actual travel time can be found in Figure 5.10.

5

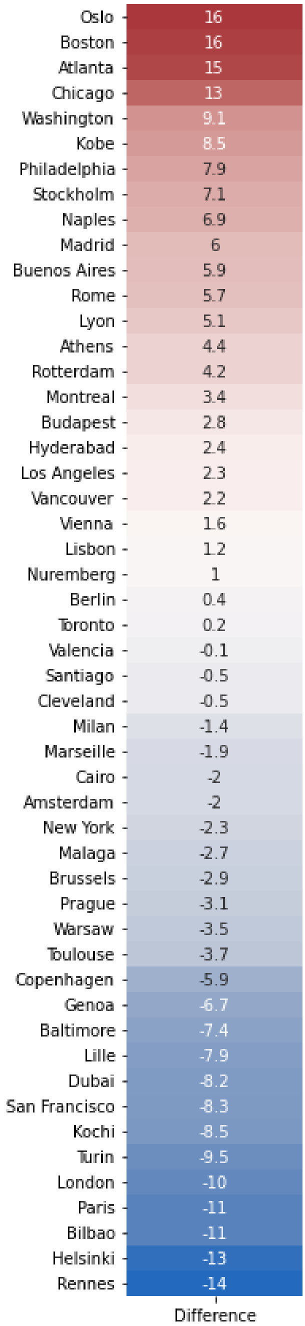
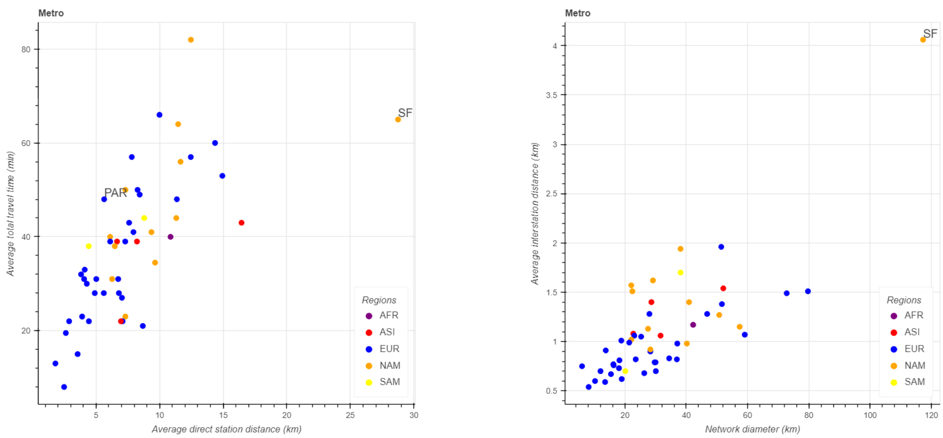


Figure 5.10: The difference between the estimated travel time and actual travel time

From Figure 5.10 it becomes clear that the estimation is fairly accurate, with over 50% of the networks having an estimated travel time that is only five minutes higher or lower than the actual travel time. There are four networks that have a much higher actual travel time than their estimated travel time, being Chicago, Atlanta, Boston and Oslo. These are the same four networks as identified in Section 5.3. As such, in addition to having a long travel time relative to their network length, these networks also have a long travel time relative to their number of stations and average direct station distance.



(a) Travel time versus average direct station distance

(b) Network diameter versus interstation distance

Figure 5.11: Additional visualizations related to the average direct station distance

Interesting to note is the San Francisco network, which was greatly underestimated by the length-based model (+11.2) but is actually strongly overestimated by this model (-8.3). Figure 5.11a makes it clear that San Francisco's exceptionally long average direct station distance is the cause of this overestimation. San Francisco in fact has an average direct station distance that is 13km longer than any other network's. This average direct station distance metric can be seen as a combination of the network diameter and the average interstation distance. These two factors are compared in Figure 5.11b. From this figure it becomes immediately obvious that San Francisco has both a much higher network diameter as well as interstation distance than all other networks. This begs the question whether the San Francisco networks should even be classified as a metro network. While the network diameter and interstation distance were not specifically defined as characteristics for metro networks in Section 3.1.1, they are related to other concepts mentioned there. Considering these factors and San Francisco's relatively low amount of transfers, it can be better be described as a commuter rail network.

There are also five networks that have a much lower actual travel time than estimated, being London, Paris, Bilbao, Helsinki and Rennes. Helsinki, Rennes and London were also overestimated outliers in the length-based model, and as such can simply be considered as networks that have a low travel time for their network size. Bilbao was also an overestimated outlier in the length-based model (-9.98), albeit slightly less so than Helsinki and Rennes. Genoa and Turin, two networks that were greatly overestimated by the length-based model (-14.8 and -13.8 respectively) are still overestimated here but much less so (-6.7 and -9.5). It thus seems that while this model is better at estimating the travel time for these two networks, they are still outliers having a relatively low travel time for their network size. Paris was also overestimated by the length-based model (-6.3) but is overestimated considerably more in this model (-11). Figure 5.11a does not provide a clear explanation for this behavior as Paris does not have an especially low average direct station distance. Figure 5.12 on the other hand shows how Paris has a relatively low travel time for its number of stations. It can thus be concluded that Paris performs especially well for its number of stations. Figure 5.9 provides a reasonable explanation for this. Paris has exceptionally low in-vehicle travel time and waiting time, also in comparison to its network length. In turn, the number of transfers is actually the second-highest out of all networks. It thus appears that this high number of transfers is offset by the low in-vehicle and waiting times.

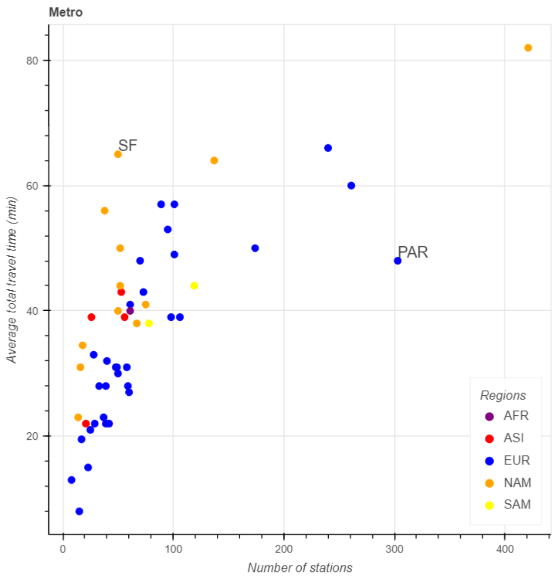


Figure 5.12: The travel time compared to the number of stations

5.4.3. CONCLUSION

In conclusion, it can be noted that the model based on the number of stations and average direct station distance can estimate the total travel time of networks to an acceptable degree. In this model, there are still networks that are significantly over- or underestimated in terms of their travel time. These networks can thus be said to perform well or poorly respectively based on their network size metrics. As becomes clear from the more in-depth look at the travel time components in Figure 5.9, this outlier nature can often be attributed to a significantly higher or lower average waiting time. This confirms the importance of including the total travel time and its components in comparisons between networks, as they capture effects not fully explained by network-related indicators.

5.5. FURTHER COMPARISONS

In this section, the comparisons with the three factors that were not extensively explored in Sections 5.3 and 5.4 — number of lines, population and GDP — will be shortly described.

5.5.1. NUMBER OF LINES

The remaining unexplored network factor of comparison is the total number of lines in the network. As hypothesized in Section 3.4.3, the number of lines is indeed positively correlated with the travel time, with a Pearson correlation of 0.69 as can be seen in Figure 5.7. The direct comparison between the travel time and number of lines can be found in Figure 5.13.

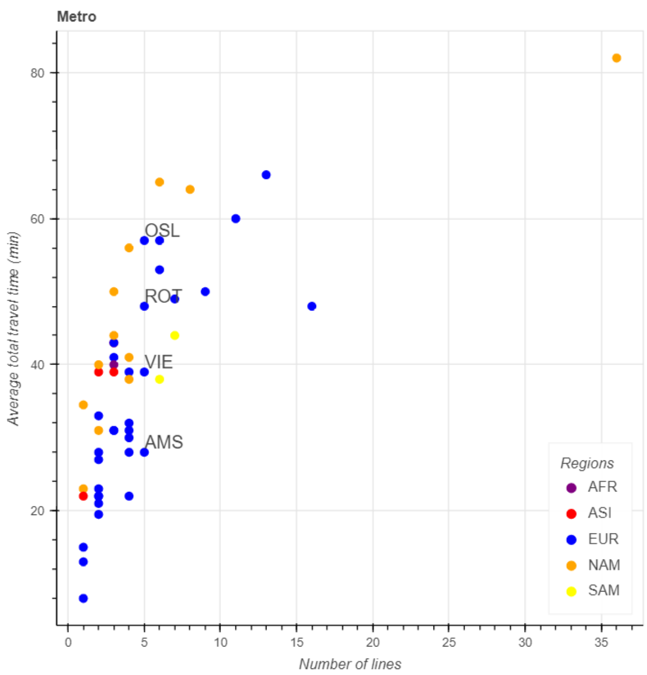


Figure 5.13: A comparison between the travel time and the number of lines

Taking a detailed look at this figure reveals that the intuitive notion about Amsterdam and Vienna, mentioned in Section 3.4.3, indeed holds true. Amsterdam and Vienna, while on the same vertical axis (both having five lines), have a difference in travel time of nearly twelve minutes. This pattern can be seen for other networks too, even within the same group of five line networks. Rotterdam and Oslo for example, have even higher travel times than Vienna and Amsterdam.

Considering the discrete nature of the number of lines, further investigation can be performed into patterns between networks with the same number of lines. As described in Section 5.1.1, networks appear to have very different travel time distributions. Considering the large effect of transfers on the waiting time (i.e., adding a travel time of thirteen minutes extra per transfer), a relationship can be expected between the number of lines and these distributions. As an example, the six networks with one line will be considered. The travel time distributions for this group of networks can be found in Figure 5.14. What can be immediately noted from Figure 5.14 is that even though these networks have different travel times (as already became clear from Figure 5.13), the distribution of their travel time is very similar. All six networks with one line have a travel time that is exponentially distributed. Naturally, this is very sensible considering none of these networks have transfers and consequently never have a 13min penalty for a transfer. As such, all paths through the graph are simply the waiting time for the one line plus the in-vehicle travel time, hence leading to this exponential pattern. Similar patterns can be found for networks with more lines as well, but this will not be discussed in further detail here.

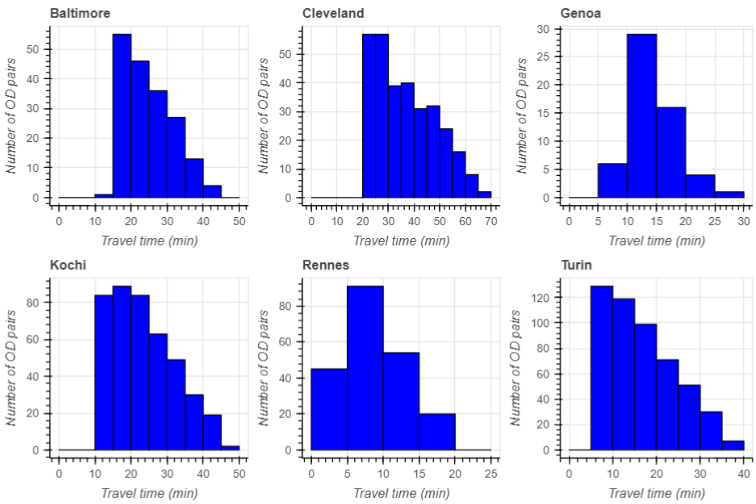


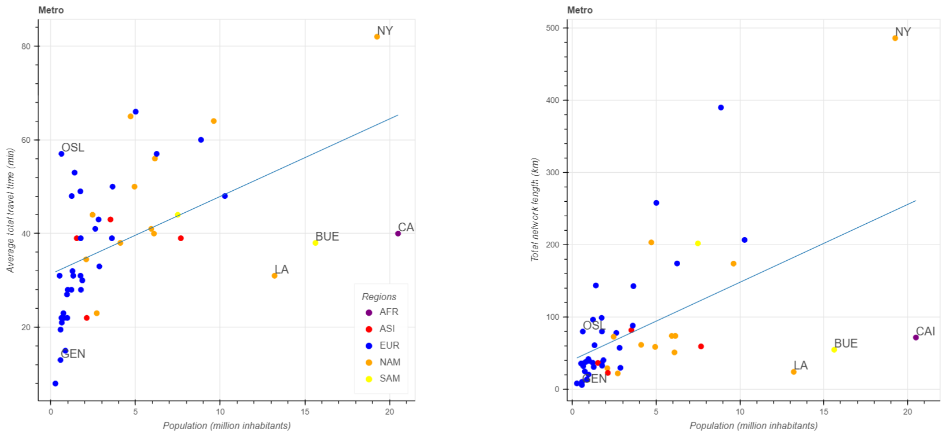
Figure 5.14: Travel time distributions of the six networks with one line

In conclusion, while the number of lines is also fairly strongly correlated with the travel time, its ability to meaningfully explain differences in travel time is much weaker. Significant differences in travel time can be identified between networks with the same number of lines. As hypothesized, a possible reason for these significant differences is the network planner's freedom to define what constitutes a line, as visualized by Figure 3.4. Interestingly however, the distribution of the travel time does show significant similarities for networks with the same number of lines, even if the values are different.

5.5.2. POPULATION

The first of the city-related factors for comparison is the city population. The direct comparison between the travel time and population can be found in Figure 5.15a.

The Pearson correlation between the population and travel time is quite low with a value of 0.51. This lower correlation also becomes clear from the patterns in Figure 5.15a having a lot of vertical disparity, a large fan-out towards the top-right and a lot of



(a) Travel time versus population

(b) Network length versus population

Figure 5.15: Direct comparisons between the city population and travel time/network length

outliers. Even though a correlation of 0.51 can still be considered moderately strong, the figure clearly shows that the correlation is quite unreliable. While there is a positive trend between the population and the travel time as hypothesized, the behavior is far too erratic to assign any true conclusions to this. Considering the strong correlation between travel time and network size as described extensively in Section 5.2, this positive correlation between the travel time and the population is most likely spurious and can be primarily attributed to larger cities having larger networks which in turn have longer travel times. Figure 5.7 confirms this positive correlation between the population and network size. with there being a Pearson correlation of 0.54 between the population and network length. This relationship is also visualized in Figure 5.15b.

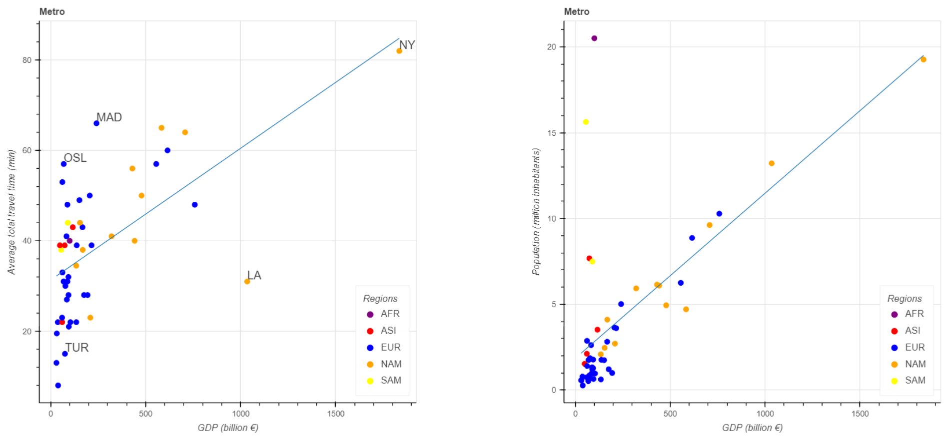
This comparatively weak correlation and the relationship with the network length can be best be explained by looking at a few notable outliers. Firstly, it is interesting to note the vertical disparity for networks up to one million inhabitants. The travel time for networks in this category varies from less than ten minutes up to nearly sixty minutes. As an example, consider Oslo and Genoa, which have a very similar number of inhabitants (~620,000 and ~570,000 respectively) but have a travel time of 57 and 13 minutes respectively. This significant difference in travel time is logically explained by the size of their network, Oslo having a network length of 80 kilometers whereas Genoa has a network length of only 13 kilometers. This example indicates how the population is not a very good determinant of the travel time in a network.

In addition to this vertical disparity, there is also a large number of outliers. These outliers are different from the ones identified in Sections 5.3 and 5.4 and include for example New York, Los Angeles, Buenos Aires and Cairo. Logically, their horizontal outlier nature is caused by their large population compared to most other cities in the data set. Los Angeles, Buenos Aires and Cairo however clearly do not follow the pattern of increasing travel time with an increasing population. Compared to their large size, the travel time for these cities is relatively low. Looking again at Figure 5.15b, this is because these networks are comparatively small for the size of the city, which in turn leads to lower travel times. New York on the other hand has a relatively high travel time for its population, also primarily caused by its high network length.

In conclusion, while the travel time and population are positively correlated, they have a much weaker relation than the previously explored factors. This is in line with expectation considering that the size of the population, while in part a determinant for the size of the network as proven by Figure 5.15b, is not directly related to the travel time of a metro network. The population does not seem to provide any additional meaningful information that network length does not already, and is thus a relatively poor indicator of the travel time.

5.5.3. GDP

The second of the city-related factors for comparison is the GDP of the city in which the metro network operates. The direct comparison between the travel time and the GDP can be found in Figure 5.16a.



(a) Travel time versus GDP

(b) GDP versus population

Figure 5.16: Direct comparisons between the city GDP and travel time/city population

As becomes clear from Figure 5.7, the correlation between the travel time and GDP is one of the weakest out of the six indicators with a Pearson correlation of 0.61. It is however positive like the others, and still reasonably strong. The relatively weaker relationship is mostly caused by the significant vertical disparity of the networks, as was similarly the case for population. An example of this vertical disparity are the networks of Oslo and Turin. While these networks have nearly the same GDP (68 and 75 billion euros respectively), their travel time differs by more than forty minutes. While there is a definite inclination to networks with higher GDP having higher travel times, there are many networks that have very different travel times while they have the same GDP.

In terms of outliers, New York and Los Angeles appear here too as was similarly the case for population. Considering the relatively high correlation between the GDP and population (Pearson correlation of 0.65) it is sensible these networks show a similar outlier nature here. Madrid has a fairly high travel time for its GDP compared to networks with similar GDP. This is caused by Madrid's large network size, which was already shown to have a high correlation with the population in Section 5.5.3. Figure 5.16b illustrates this correlation between the GDP and population.

In terms of regional patterns, the most interesting regions in the comparison between travel time and GDP are actually Africa, Asia and South-America. Interestingly, networks from these regions are all grouped together, with the exception of Kochi which has a lower travel time. The networks from these regions have a fairly high travel time for their GDP. Figure 5.16b actually provides some more insight into possible reasons behind this pattern. As can be noted from that figure, the single African network (Cairo), the two South-American networks (Buenos Aires and Santiago) and one of the Asian networks (Hyderabad) all have a very high population compared to their GDP. This means these cities have a relatively low GDP per capita, which is sensible considering these cities are part of developing regions.

In conclusion, out of the six relationships investigated, the GDP is, as expected, one of the weakest. Even though there is a positive correlation with travel time, networks with similar GDP can have very different travel times. Considering the correlation between the GDP and population, it is sensible that the patterns here are similar to those found for the population. However, there is one major difference between the population and GDP in the grouping of Asian, African and South-American networks, which is caused by the comparatively low GDP per capita of these networks.

5.6. TRAVEL TIME CALCULATION BENCHMARK ANALYSIS

To complete the results section, a benchmark analysis is performed to evaluate whether the total travel time method of this study provides significantly different results compared to other commonly used methods. In order to evaluate this, the average shortest path total travel time as calculated in this study is compared to the shortest path based purely on in-vehicle travel time as well as the shortest path based on hops. For brevity, these three methods will be referred to as the "total", "in-vehicle" and "hops" method respectively. The methods are compared both directly as well as indirectly. The direct comparison consists of plotting the average values for total compared to in-vehicle and hops separately for all networks in the data set. The indirect comparison consists of plotting the three methods to another factor (network length in this case) in order to evaluate how different the patterns and results are. Considering the network length's high correlation with the total travel time, this was considered as the most sensible factor for comparison, as also indicated in Section 5.2. Firstly, the correlations between the three methods are shortly discussed in Section 5.6.1. Afterwards in Sections 5.6.2 and 5.6.3, the aforementioned comparisons are plotted and explored using different figures. Lastly, the conclusions from this benchmark analysis are discussed in Section 5.6.4.

5.6.1. CORRELATION BETWEEN THE METHODS

Before going into the comparisons between the three methods, it is interesting to first note how correlated they are amongst themselves. The Pearson correlations between the three methods can be found in Table 5.8. From this table it becomes clear that the

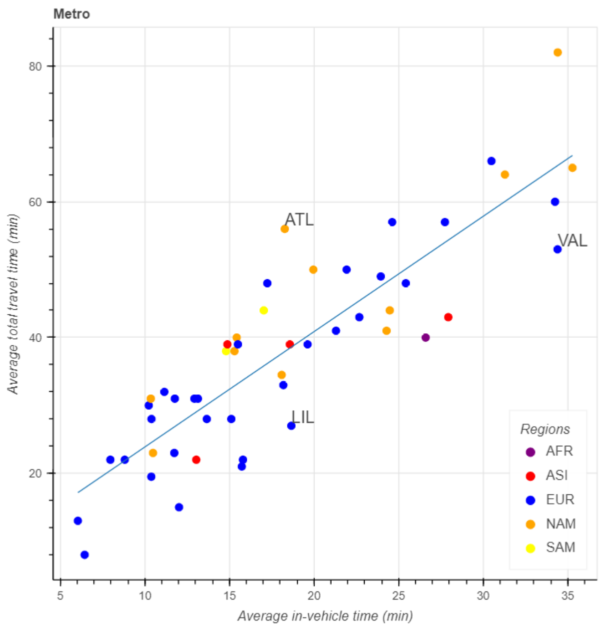
Method 1	Method 2	Pearson correlation
Total	In-vehicle	0.87
Total	Hops	0.68
In-vehicle	Hops	0.80

Table 5.8: The Pearson correlation between the three methods

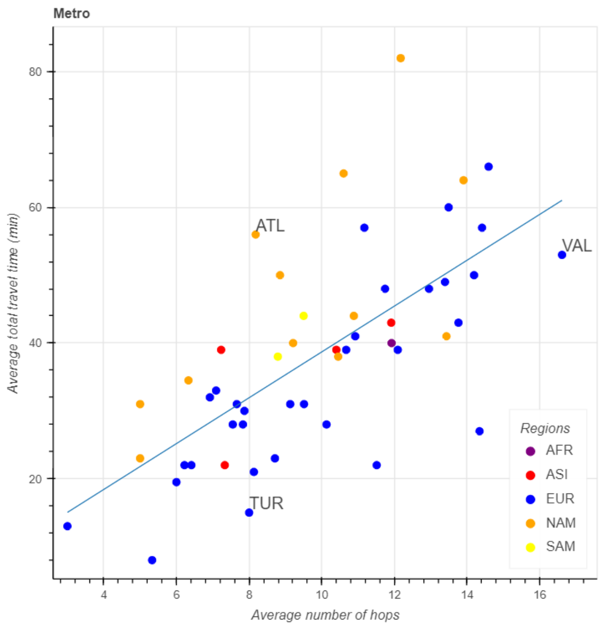
total method from this study has a very strong correlation with the in-vehicle method of 0.87. Naturally, this is a very logical result considering the fact that the total method is in part based on the in-vehicle time. The next strongest relationship is between the in-vehicle and hops method of 0.80. This is quite interesting to note, as this relationship is still quite strong. The fact that this relationship is so strong is fairly sensible however, considering that both methods do not regard transfers and only account for the number of links crossed, either including actual travel time or not. Lastly, the relationship between the total and the hops method must be noted, which is the lowest of the the three with a value of 0.68. It is sensible that this is the lowest, considering these two methods have the least in common. This comparatively weaker correlation is a first indication that the total and hops method are not completely interchangeable and describe networks differently. The exact differences between the two methods are further explored in the next sections.

5.6.2. DIRECT COMPARISON BETWEEN THE METHODS

In this section, the total travel time method is compared directly to the in-vehicle and hops methods. The comparison between the total and in-vehicle and hops methods can be found in Figures 5.17a and 5.17b respectively. Both of these figures support the findings from Section 5.6.1 with both looking clearly positively correlated. The difference in strength of correlation also becomes clear immediately with the hops method comparison having a considerably higher spread than the in-vehicle method comparison. Nevertheless, the spread is quite significant for both of these methods in both the vertical and horizontal direction. Some examples are shortly highlighted to indicate this difference. Atlanta is an excellent example in both comparisons of how the two alternative methods can misrepresent networks. In Figure 5.17a it can be seen how Atlanta has nearly the same in-vehicle travel time as Lille but a completely different total travel time (56 and 27 minutes respectively). Using the in-vehicle method would suggest Atlanta and Lille are similar, while the total travel time method shows they are most certainly not. On the other hand, Atlanta and Valencia have nearly the same total travel time, but a very different in-vehicle travel time (18 and 34 minutes respectively). Using the in-vehicle method would suggest that Valencia has much longer travel times while this is in fact not the case when incorporating waiting time and transfers. This pattern can also be seen for Atlanta in the comparison with the hops method, where the variance is even higher. Valencia also has a significantly higher number of hops than Atlanta, while having the same total travel time. In terms of other networks with the same number of hops as Atlanta, Turin is a notable example with a much lower total travel time (57 vs 15 minutes). These examples show how differently these three methods describe networks.



(a) Total vs in-vehicle



(b) Total vs hops

Figure 5.17: The total travel time method compared to the in-vehicle time and hops method

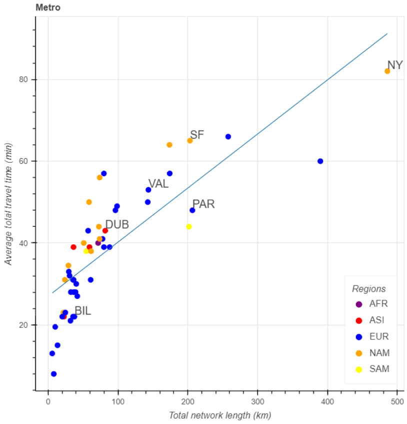
5.6.3. INDIRECT COMPARISON BETWEEN THE METHODS

To further illustrate the differences between them, the three methods are also all compared to the network length in Figure 5.18.

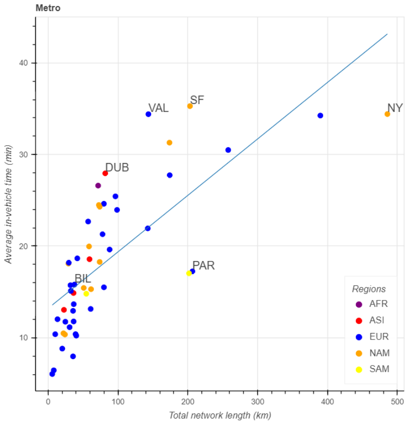
It can be immediately observed that the three figures show fairly different patterns. While all three figures show a positive correlation between the two variables, the correlation is obviously strongest between the total travel time and the length. This is confirmed by the Pearson correlations which are 0.81, 0.74 and 0.54 for the total, in-vehicle and hops method respectively. The correlation is especially unreliable between hops and network length. As can be seen in Figure 5.18c, between a length of 0 and 100km the networks still follow a fairly linear positive pattern. Afterwards this pattern completely disappears and even turns into a negative correlation for Madrid, London and New York. None of the other comparisons in this study have shown a negative correlation as appears here for these three networks.

In regards to earlier identified patterns, differences between the three methods can also be discovered. The poor performance of North-American networks compared to European ones as proven in Section 5.3, and visualized in Figure 5.18a is not visually present for the other two methods. This intuitive notion is confirmed by regression analysis, where including a North-American/European parameter is not significant for both the in-vehicle and hops methods.

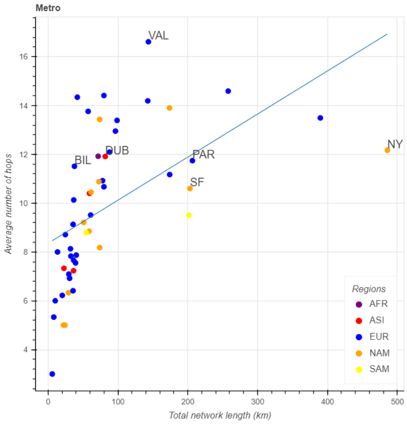
Another manner to indicate how the in-vehicle and hops method misrepresent some networks is through examples of individual networks. The most striking example is New York, as it performs very differently for the hops method than for the total method. According to the hops method, New York has the same number of hops as networks like Bilbao, Dubai and Paris. Considering the earlier analyses performed and the significant differences identified between these networks, it is clear the hops method misrepresents these networks. Especially considering how New York has a much higher total travel time than those networks, it is clear that the number of hops represents shortest path travel times with a much lower accuracy. Something similar can be noted for San Francisco which, as opposed to the total and in-vehicle method, actually outperforms Paris in the hops method. This is sensible considering its high interstation distance and consequently lower number of hops necessary to cover similar distances. A network that also clearly illustrates the differences between the total and in-vehicle method is Valencia. Its position in Figure 5.18a is not especially notable, being situated towards the top-left of the group near Berlin and Washington. In both Figure 5.18b and 5.18c it is however situated far towards the top of the figure, fairly isolated. This is naturally caused by its relatively high in-vehicle travel time and number of hops respectively. From Figures 5.9b and 5.9c it becomes clear that Valencia compensates this high in-vehicle travel time with low waiting times and a lower number of transfers required.



(a) Total travel time vs network length



(b) In-vehicle travel time vs network length



(c) Number of hops vs network length

Figure 5.18: The three different shortest path methods compared to the network length

5.6.4. CONCLUSION

In conclusion, it becomes clear that the three methods compared in this section represent networks very differently. As expected, the hops method has the weakest correlation with the network size (represented as network length). From the comparisons it also becomes clear that this method severely misrepresents some networks due to its lack of incorporating service information such as in-vehicle time, waiting time and the number of transfers. The in-vehicle method is a little more accurate, properly representing the actual travel time spent in vehicles. It does however still significantly over- or underrepresent some networks because of its exclusion of other service aspects such as waiting time and transfers. The total travel time method incorporates all three of these aspects and also has the strongest correlation with the network size. As discussed in Chapter 2, the total travel time method applied in this paper is not frequently used in literature. Considering the more accurate nature of the total travel time method, the importance of using this method in further studies becomes obvious.

6

Conclusion

In this final chapter, the conclusions and recommendations based on the study are provided. Firstly, the key findings answering the research questions are discussed in Section 6.1. The planning and policy implications based on these findings are detailed thereafter in Section 6.2. The limitations of this study are addressed in Section 6.3. Finally, this study concludes in Section 6.4 with recommendations for further research.

6.1. Key findings

The first findings to discuss are related to the main research question: "How can service information be included into a comprehensive comparison of metro networks worldwide?". This study shows how service information can be efficiently included as concepts of in-vehicle time, waiting time and number of transfers in the calculation of shortest path travel times through the network. Through this method, a comparison between metro networks was made that includes this service information.

Naturally, in order to perform these comparisons, data is necessary for each network. The data format used for this study is the so-called GTFS static format, providing information about the physical lay-out of the network as well as the schedule of operation. Relevant data in this format was found for 51 different metro networks. In order to process this data, a pipeline was created. L- and P-space representations were used to represent the topological infrastructure and service of the network respectively. Creating these representations based on GTFS data turned out to be fairly time-consuming, especially for larger networks. The created pipeline aims to facilitate this process by providing methods requiring less manual intervention to arrive at the proper representations. A set of curated L- and P-space representations was created for 51 metro networks and is publicly available for further research. This set of representations can be used in future studies with a variety of different applications such as network theory studies or further transportation research into metro networks. By combining the data from both the L- and P-space representations, composite shortest paths were calculated based on a combination of in-vehicle travel time, waiting time and the number of transfers.

Calculating these shortest path travel times for all networks provided insight into the travel time patterns of each network. This revealed that these patterns were not distributed similarly for each network with some being normally distributed and others following an exponential or bimodal distribution. Considering this variation in distribution, it became clear that the median of these travel times would be the best indicator to characterize networks. As such, in order to incorporate service information, the median total travel time was used as the primary indicator for this study.

A variety of factors related to the network size and city were used to help explain differences in the total travel time between networks. Firstly, the correlations between the travel time and these other factors were investigated. As hypothesized, all factors have a positive correlation with the travel time. From these factors, the network length has the strongest correlation with the travel time. This shows the pattern of increasing travel time with network size found in other studies (Luo et al., 2019) applies to this set of networks as well. Considering its strong correlation, the network length was used in a simple regression model to estimate the travel time. Through this, a variety of outlier networks could be identified that perform either much better or worse than predicted according to the network length. By analyzing the three different travel time components (in-vehicle time, waiting time and number of transfers), the causes for this outlier nature could be discovered. It turned out that the waiting time was the main cause for discrepancies between networks. This thus highlights the importance of including the waiting time in comparisons between metro networks. In addition, this regression model showed that North-American networks perform structurally worse compared to other networks, especially worse than European ones. Including this regional variation in the regression model turned out to be significant and improved the model slightly.

Multiple regression was also applied in order to further improve the model. The improved model has two explanatory variables with a high correlation with travel time, but low amongst each other: the number of stations and the average direct station distance. This model provided significant coefficients and performed much better than the simple regression model. Nevertheless, this model showed some of the same outliers as the simple model, albeit reduced. Paris, Bilbao, Helsinki, Rennes and London seem to perform much better than predicted by the model while Chicago, Atlanta, Boston and Oslo perform much worse. The most notable outlier is San Francisco, which is such an anomaly in terms of both its average direct station distance and average interstation distance, that it should not be considered a metro network, but a commuter rail instead. The regional difference discovered for the simple model was not found to be significant in this model. The other three explored factors — number of lines, city population and city GDP — while also positively correlated with the travel time, provided no further insights.

Lastly, it was evaluated how the total travel time method applied in this study compares to the more common methods in literature which consider only in-vehicle time or hops in shortest path calculations. While all three methods show a positive correlation with network size, the method proposed in this study has a stronger correlation than the other methods. In addition, it becomes clear that networks are represented very differently by these three methods. Some networks have a similar amount of average hops while their average total travel time differs by more than thirty minutes. These examples show how different these methods are in representing networks. Considering the more realistic nature of the total travel time method, it becomes clear that applying this method is desirable for better comparisons of networks.

6.2. Planning and policy implications

The key findings as described above lead to certain implications for planning and policy which are shortly discussed in this section.

In a general sense, the total travel time indicator turned out to be a suitable factor to compare the accessibility of networks. Clear differences between networks appeared from this indicator and network planners can get an idea of how well their network compares to others. In this way, the results of this study provide a benchmark for network planners. Through these results, planners can understand which networks share the same characteristics in terms of different indicators, especially average travel time. This provides them with a clear idea of which networks to consider for how to improve their own network. The method applied by this study also provides a more accurate image of the actual shortest path travel times through the network and as such is a better basis for comparison. In terms of network planning, planners can gain better understanding of which travel time factors are the weakest for their network. A high in-vehicle travel time could be improved by purchasing faster vehicles while long waiting times could be mitigated by increasing the frequency.

This study also provides insights into which metrics accurately represent travel time and which do not. Whereas the network length and the combination of number of stations and average direct station distance were good metrics for comparison, the other explored factors were less so. This means network planners have a better idea of which factors to focus on for comparison. In addition, the city factors, while positively correlated with travel time, do not provide much additional information of interest. The variation in travel time for both the city and population is quite large and does not provide planners with a clear direction of improvement.

The total travel time method applied in this study also emphasizes the importance of considering all travel time aspects of the travellers' journey. Including waiting time and number of transfers in network representations provides better insights into which aspects of the travel time have most room for improvement.

6.3. Limitations

While this study aims to be as comprehensive as possible, decisions had to be made that limited the scope.

The first obvious limitation is the consideration of metro networks as a distinct, separated mode. Naturally, planners frequently attempt to integrate PT modes such that they connect well and that certain gaps of one mode are captured by the others. Considering the complicated nature of multi-layer networks, the decision was made to restrict this study exclusively to metro networks.

In terms of the actual networks included in this analysis, further restrictions were made. Ideally, all 190 metro networks should have been included in this analysis to get a full comparison of all networks in existence. Considering the heavy reliance of this study on scheduling data, this turned out to be not feasible. Instead, a set of 51 networks was considered based on the availability and correctness of the discovered data. Unfortunately, this data set is slightly skewed towards small- and mid-sized networks as well as including primarily networks from Europe and North-America. These limitations show the importance of having correct, publicly-available and standardized data.

Considering the reliance of this study on provided data and the specific implementation, there are quite an amount of assumptions and decisions that affect the results. In processing the data, decisions had to be made about the merging of stations in L-space and combining edges in P-space. Both of these decisions are choices made for this study but might be made differently for other studies, providing different results. In addition, a lot of manual correction was required in order to arrive at the final curated network representations. While this process was performed diligently and each step was documented, the process is still prone to errors from either the data itself or human error in the manual edition. Any errors in the data such as missing stations and incorrect schedules logically can lead to mistakes in the final results. Further, while a set data format was used (GTFS), the exact implementation of this format differs per network in various aspects such as the rounding of the time schedule to minutes or seconds. For each network the schedule is considered only between 05:00 and 23:59 which means potential night networks are not included. The implementation of the shortest path is also a simplification, considering only the five shortest paths are considered in L-space, which are then combined with the shortest path in P-space. Naturally, a consideration of all shortest paths would be more accurate, yet also much more time-consuming.

While the total travel time method applied in this study is more representative than the in-vehicle or hops method, it still relies on some assumptions and simplifications. The waiting time might in fact be much lower for some networks as transfer possibilities are created by planners that are well-connected in the schedule. Walking times are also not specifically considered in this study which can also affect the actual transfer times experienced by travellers. In addition, all transfer stations are considered equal in size for this study, while in actuality some stations are much bigger than others, leading to longer transfer times.

Finally, the median value taken as the average of the travel time distributions might not be representative of the average traveler's experience. As this value is the median of all possible OD-pairs, this might overrepresent some hardly used OD-pairs while underrepresenting frequently used OD-pairs.

6.4. RECOMMENDATIONS FOR FURTHER RESEARCH

Based on the results and limitations of this study, some recommendations can be made for further research.

Firstly, the results of this study can be explored into more detail. An example of this is applying principal component analysis which can improve the approximation of the travel time. Cluster analysis would also provide more detailed insights into the groups of networks that exist, giving more information to planners about their peer networks. In addition, other indicators can be explored in combination with the total travel time. Network science indicators such as directness and degree can be investigated both locally as well as globally to discover further relations with the travel times. Questions such as: "Which stations are most often used for transferring?" can also be answered using the provided data set.

Secondly, the methodology of this study can be expanded upon. Currently, metro is the only mode considered which could be expanded upon with other PT modes such as tram and bus. The pipeline could be adjusted to contain multi-layer networks in order to incorporate these modes. This would allow for the comparison of multimodal PT systems, providing an even more comprehensive image of the PT systems of these cities. In addition, the applied total travel time metric relies on a generic image of the waiting time. Including transfer and walking time separately, using for example split transfer stations, could provide much more detail on specific sections of the network as well as the network as a whole. Other factors such as operation strategies and network structure can also be considered in order to provide more meaningful insights into differences between networks.

Thirdly, the data provided by this study could be combined with actual traveler data. The data from this study could help explain certain travel behavior or identify bottlenecks in the network. An example of traveler data like this is demand data between OD-pairs, which would make the average total travel time more representative of actual travel behavior. In addition, including demand data can lead to further research in terms of accessibility. While both the total travel time and network length are already related to accessibility, further inclusion of other accessibility factors can provide more in-depth insights.

Lastly, the results of this study could be related to existing benchmark studies between networks to get a more comprehensive image of the relationships between networks. An example of this is the McKinsey report (McKinsey, 2021) which provides a complementary qualitative assessment that could be combined with this study's quantitative assessment.

BIBLIOGRAPHY

Acampa, G., Contino, F., Grasso, M., & Ticali, D. (2019). Evaluation of infrastructure: Application of TOD to Catania underground metro station. *AIP Conference Proceedings*, *2186*. https://doi.org/10.1063/1.5138078

Amini, B., Peiravian, F., Mojarradi, M., & Derrible, S. (2016). Comparative analysis of traffic performance of urban transportation systems. *Transportation Research Record*, *2594*, 159–168. https://doi.org/10.3141/2594-19

Bastidas-Zelaya, E. (2021). Metro systems in Latin America, comparison of planning and development models versus other regions in the world. *Transportation Research Procedia*, *58*, 614–621. https://doi.org/10.1016/j.trpro.2021.11.081

Bokeh Development Team. (2018). *Bokeh: Python library for interactive visualization*. https://bokeh.pydata.org/en/latest/

Derrible, S. (2012). Network centrality of metro systems. *PLoS ONE*, *7*. https://doi.org/10.1371/journal.pone.0040575

Derrible, S., & Kennedy, C. (2009). Network analysis of world subway systems using updated graph theory. *Transportation Research Record*, 17–25. https://doi.org/10.3141/2112-03

Derrible, S., & Kennedy, C. (2010a). Characterizing metro networks: State, form, and structure. *Transportation*, *37*, 275–297. https://doi.org/10.1007/s11116-009-9227-7

Derrible, S., & Kennedy, C. (2010b). The complexity and robustness of metro networks. *Physica A: Statistical Mechanics and its Applications*, *389*, 3678–3691. https://doi.org/10.1016/j.physa.2010.04.008

Derrible, S., & Kennedy, C. (2011). Applications of graph theory and network science to transit network design. *Transport Reviews*, *31*, 495–519. https://doi.org/10.1080/01441647.2010.543709

Dimitrov, S. D., & Ceder, A. (2016). A method of examining the structure and topological properties of public-transport networks. *Physica A: Statistical Mechanics and its Applications*, *451*, 373–387. https://doi.org/10.1016/j.physa.2016.01.060

Garcia-Martinez, A., Cascajo, R., Jara-Diaz, S. R., Chowdhury, S., & Monzon, A. (2018). Transfer penalties in multimodal public transport networks. *Transportation Research Part A: Policy and Practice*, *114*, 52–66. https://doi.org/10.1016/j.tra.2018.01.016

Google. (2022). GTFS Static Overview. https://developers.google.com/transit/gtfs/

Guo, Z., & Wilson, N. H. (2011). Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground. *Transportation Research Part A: Policy and Practice*, *45*, 91–104. https://doi.org/10.1016/j.tra.2010.11.002

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th python in science conference* (pp. 11–15).

Haznagy, A., Fi, I., London, A., & Nemeth, T. (2015). Complex network analysis of public transportation networks: A comprehensive study. *Models and Technologies for Intelligent Transportation Systems (MT-ITS)3-5. June 2015. Budapest, Hungary*, 536.

Hong, J., Tamakloe, R., Lee, S., & Park, D. (2019). Exploring the topological characteristics of complex public transportation networks: Focus on variations in both single and integrated systems in the Seoul Metropolitan Area. *Sustainability (Switzerland)*, *11*. https://doi.org/10.3390/su11195404

Igualada, J. P. The metro and the city: Interaction with the urban landscape in the new underground stations in Valencia 1995–2010. In: *146*. 2015, 364–374. https://doi.org/10.2495/UT150291.

Ingvardson, J. B., & Nielsen, O. A. (2018). How urban density, network topology and socio-economy influence public transport ridership: Empirical evidence from 48 European metropolitan areas. *Journal of Transport Geography*, *72*, 50–63. https://doi.org/10.1016/J.JTRANGEO.2018.07.002

Jara-Diaz, S., Monzon, A., Cascajo, R., & Garcia-Martinez, A. (2022). An international time equivalency of the pure transfer penalty in urban transit trips: Closing the gap. *Transport Policy*, *125*, 48–55. https://doi.org/10.1016/j.tranpol.2022.04.019

Jin, F., Jiao, J., Qi, Y., & Yang, Y. (2017). Evolution and geographic effects of high-speed rail in East Asia: An accessibility approach. *Journal of Geographical Sciences*, *27*, 515–532. https://doi.org/10.1007/s11442-017-1390-8

Jupyter. (2022). Jupyter Project Documentation. https://docs.jupyter.org/en/latest/

Kujala, R., Weckström, C., Mladenović, M. N., & Saramäki, J. (2018). Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys. *Computers, Environment and Urban Systems*, *67*, 41–54. https://doi.org/10.1016/j.compenvurbsys.2017.08.012

Latora, V., & Marchiori, M. (2002). Is the Boston subway a small-world network? *Physica A: Statistical Mechanics and its Applications*, *314*(1), 109–113. https://doi.org/https://doi.org/10.1016/S0378-4371(02)01089-0

Lee, K., Jung, W. S., Park, J. S., & Choi, M. Y. (2008). Statistical analysis of the Metropolitan Seoul Subway System: Network structure and passenger flows. *Physica A: Statistical Mechanics and its Applications*, *387*, 6231–6234. https://doi.org/10.1016/j.physa.2008.06.035

Lee, Y. J., & Vuchic, V. R. (2005). Transit network design with variable demand. *Journal of Transportation Engineering*, *131*, 1–10. https://doi.org/10.1061/(ASCE)0733-947X(2005)131:1(1)

Li, W., & Cai, X. (2007). Empirical analysis of a scale-free railway network in China. *Physica A: Statistical Mechanics and its Applications*, *382*, 693–703. https://doi.org/10.1016/j.physa.2007.04.031

Lin, J., & Ban, Y. (2013). Complex network topology of transportation systems. *Transport Reviews*, *33*, 658–685. https://doi.org/10.1080/01441647.2013.848955

Luo, D., Cats, O., & van Lint, H. (2020). Can passenger flow distribution be estimated solely based on network properties in public transport systems? *Transportation*, *47*, 2757–2776. https://doi.org/10.1007/s11116-019-09990-w

Luo, D., Cats, O., van Lint, H., & Currie, G. (2019). Integrating network science and public transport accessibility analysis for comparative assessment. *Journal of Transport Geography*, *80*, 102505. https://doi.org/10.1016/J.JTRANGEO.2019.102505

Majima, T., Katuhara, M., & Takadama, K. (2007). *Analysis on transport networks of railway, subway and waterbus in Japan.*

McKinsey. (2021). *Urban transportation systems of 25 global cities - Elements of success.*

MobilityData. (2021). Mobility Database. https://database.mobilitydata.org/

Musso, A., & Vuchic, V. R. (1988). Characteristics of metro networks and methodology for their evaluation. *Transportation Research Record*, *1162*, 22–33.

Python. (2022). Python 3.10.7 documentation. https://docs.python.org/3/

Saidi, S., Wirasinghe, S. C., Kattan, L., & Esmaeilnejad, S. (2017). A generalized framework for complex urban rail transit network analysis. *Transportmetrica A: Transport Science*, *13*, 874–892. https://doi.org/10.1080/23249935.2017.1348401

Shanmukhappa, T., Ho, I. W. H., & Tse, C. K. (2018). Spatial analysis of bus transport networks using network theory. *Physica A: Statistical Mechanics and its Applications*, *502*, 295–314. https://doi.org/10.1016/j.physa.2018.02.111

Shanmukhappa, T., Ho, I. W.-H., Tse, C. K., & Leung, K. K. (2019). Recent development in public transport network analysis from the complex network perspective. *IEEE Circuits and Systems Magazine*, *19*, 39–65. https://doi.org/10.1109/MCAS.2019.2945211

Sienkiewicz, J., & Hołyst, J. A. (2005). Statistical analysis of 22 public transport networks in Poland. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *72*. https://doi.org/10.1103/PhysRevE.72.046127

SQLite. (2022). SQLite Documentation. https://www.sqlite.org/docs.html

Taecharungroj, V. (2022). An analysis of Tripadvisor reviews of 127 urban rail transit networks worldwide. *Travel Behaviour and Society*, *26*, 193–205. https://doi.org/10.1016/j.tbs.2021.10.007

UITP. (2022). *World Metro Figures 2021.*

Von Ferber, C., Holovatch, T., Holovatch, Y., & Palchykov, V. (2009). Public transport networks: Empirical analysis and modeling. *European Physical Journal B*, *68*, 261–275. https://doi.org/10.1140/epjb/e2009-00090-x

Weckström, C., & Mladenović, M. N. (2020). Evaluation of public transport policy formulation and implementation: Case study of 24 mid-sized Nordic cities. *Transportation Research Procedia*, *45*, 979–986. https://doi.org/10.1016/j.trpro.2020.02.068

Wei, S., Teng, S. N., Li, H. J., Xu, J., Ma, H., Luan, X. L., Yang, X., Shen, D., Liu, M., Huang, Z. Y., & Xu, C. (2019). Hierarchical structure in the world's largest high-speed rail network. *PLoS ONE*, *14*. https://doi.org/10.1371/journal.pone.0211052

Wu, X., Dong, H., Tse, C. K., Ho, I. W., & Lau, F. C. (2017). Analysis of metro network performance from a complex network perspective. *Physica A: Statistical Mechanics and its Applications*, *492*, 553–563. https://doi.org/10.1016/j.physa.2017.08.074

Xu, Q., Mao, B. H., & Bai, Y. (2016). Network structure of subway passenger flows. *Journal of Statistical Mechanics: Theory and Experiment*, *2016*. https://doi.org/10.1088/1742-5468/2016/03/033404

Xu, Q., Zu, Z. H., Xu, Z. J., Zhang, W. D., & Zheng, T. (2013). Space P-based empirical research on public transport complex networks in 330 cities of China. *Journal of Transportation Systems Engineering and Information Technology*, *13*, 193–198. https://doi.org/10.1016/s1570-6672(13)60100-2

Zhang, J., Zhao, M., Liu, H., & Xu, X. (2013). Networked characteristics of the urban rail transit networks. *Physica A: Statistical Mechanics and its Applications*, *392*, 1538–1546. https://doi.org/10.1016/J.PHYSA.2012.11.036

A

Scientific paper

Including service information in a topological comparison of metro networks worldwide

Sam Vijlbrief[1]

[1] *Delft University of Technology, Delft, The Netherlands*

October 20, 2022

Metro networks are a vital public transport (PT) mode in urban areas around the world because of their high-capacity and high-speed operation. Comparing different metro networks to one another is a suitable manner for transport planners to gain insights into the characteristics of their networks and which areas of improvement exist. While metro networks have been studied extensively in the field of network science, the practical relevance for the field of transport science often remained limited, caused by the lack of realism of the network representations used. This study proposes a topological comparison of metro networks worldwide including service information. Service information is included in the form of the total travel time indicator for shortest path calculations, which combines the in-vehicle time, waiting and number of transfers. The median of this total travel time is taken for each network and compared to other networks. This metric in turn is contrasted with network size characteristics in order to explore relations between these factors and to explain the patterns discovered. From this analysis, it is revealed that the travel time increases with network size. The total travel time methodology applied in this study shows significantly different results to other commonly used methods that rely only on in-vehicle time or hops to calculate shortest path travel times. Future studies can expand on this by considering other network science indicators and looking further into local indicators. In addition, the methodology could be expanded with more detailed transfer information and other PT modes.

1 Introduction

Public transport (PT) plays a vital role in commuting billions of travellers in cities all over the world. Considering the rapid threat of climate change, improving sustainability and equity are crucial factors to keep cities habitable. PT plays a crucial role in providing transport that is both sustainable as well as accessible. A variety of different modes to transport volumes of people throughout cities exist, such as bus, tram, metro and train. In an urban environment, metro networks have a variety of factors that make them especially attractive. In terms of operation, a metro is considerably high-capacity and high-speed compared to most other PT modes. In terms of infrastructure, their (primarily) underground nature makes their location highly flexible while also having a minimal impact on the above-ground urban infrastructure. Considering these advantages and the rapid advance of technology to construct metro networks, their popularity has increased greatly over the past decade. In the last decade, almost sixty new systems have opened, nearly a third of the total number of metro networks worldwide. On the other hand, existing networks also frequently expand, seeing as the total amount of metro network infrastructure has increased by 25%, or 3,300km in total, in the past three years alone (UITP, 2022).

Metro networks have also received much attention in literature in the recent years. The field of complex network science has been using public transport systems as a field of application since the early 2000s (Latora and Marchiori, 2002; Sienkiewicz and Hołyst, 2005). PT systems as a complex network have been extensively explored around the year 2010 with the works of Derrible and Kennedy (Derrible and Kennedy, 2009; Derrible and Kennedy, 2010a; Derrible and Kennedy, 2010b; Derrible and Kennedy, 2011; Derrible, 2012). These researchers noted how suitable PT networks were to investigate various concepts from complex network theory. While this provided many new insights in the field of network science, the practical relevance for the field of transport science often remained limited. This limited relevance is primarily caused by the lack of realism of the network representations used. Frequently, these studies used only simple hops or in-vehicle travel time to calculate travellers' paths through the network. In doing so, service concepts such as transfer possibilities and waiting time are completely ignored. Evidently, this can lead to a misrepresentation of networks.

PT scientists recognized this gap and tried to integrate concepts from transport science into network analyses. This led to a large variety of examples such as: creating a weighted graph with passenger flows (Xu, Mao, and Bai, 2016); investigating the relationship between network topology and ridership (Ingvardson and Nielsen, 2018); or integrating network science and accessibility analysis (Luo et al., 2019). These studies, albeit limited in number and scope, provided more insights into the relationship between the structure of PT networks and their actual service and usage patterns, providing information for policy makers on the performance of their networks.

With this field of combining network infrastructure and service information upcoming, various aspects are still left unexplored. For example, while studies as the aforementioned do contribute new methodologies in the field, these have not been extensively used to compare different networks yet. It is vital for PT planners to understand the performance of their network on these aspects compared to other networks, in order to learn where room for improvement exists within their networks. In addition, even when comparison studies are performed, the set of networks used is usually similar: a set of a few dozen, large-sized metro networks. Logically, using only the same few dozen networks means a significant portion of the 190 metro networks worldwide (UITP, 2022) is currently not studied. Additionally, as described before, the number of networks has greatly increased in recent years, meaning many recent networks and expansions are also still unexplored. Using primarily large-sized networks also skews the results towards networks of that size, thus producing results that might be less applicable for small- and mid-sized networks. It can thus be concluded that a comprehensive comparison of a large set of varied networks is currently missing.

To fill these gaps, this study proposes a comprehensive, topological comparison study of metro networks worldwide of varying size. This comparison study includes service characteristics mathematically in the topological representations of the networks in order to provide a more realistic representation of these networks. This analysis uses an indicator that incorporates service information to compare networks to each other. This metric in turn is contrasted with network- and city-related characteristics in order to explore relations between these factors and to explain the patterns discovered. Specific attention is also paid towards regional differences in order to discover what potential patterns there might be.

Based on the research gaps as described above, the question this study aims to answer can best be formulated as follows:

How can service information be included into a topological comparison of metro networks worldwide?

This research question is supported by the following sub-questions that provide more detail on the possible methodology and results:

- What indicator can best be used to account for service information?
- Which regional differences and outliers are identified when examining networks' total travel time?
- Can (a combination) of network factors be used to meaningfully explain the travel time of a network?
- Does including service information provide significantly different results to commonly used methods?

The rest of this paper is structured as follows: Firstly, the current state of the literature is reviewed in Chapter 2. In Chapter 3, the methodology for answering the research questions is further detailed. The results are reported in Chapter 4. Finally, the conclusions and recommendations of this study are detailed in Chapter 5.

2 Related works

In this section, the related works are discussed. The related works are split up into two distinct categories: PT network topology and service information.

2.1 PT network topology

The first of topic to discuss, PT network topology, is arguably the most significant and comprehensive one. This significance becomes clear from the fact that it has been an interest of study for several decades and is continually expanding with new findings and applications. The network topology of PT networks has been of interest for transport scientists and network scientists alike, giving the field a broad variety of studies from different directions.

One of the earliest examples of literature on network infrastructure is by Musso and Vuchic (1988). In this paper, the authors aim to identify the most important geometric characteristics of metro networks.

This study can be regarded as the foundation on which many future studies expanded in the field of network structure analysis. Studies continuing and expanding on this work arose especially in the 2000s, when different studies were performed on various locations and PT modes around the world. One of the earliest examples is the work by Latora and Marchiori (2002). In their work, the authors aim to bridge the gap between the theoretical paradigms of complex network science and real complex networks by applying network science methodologies on the Boston Subway network. A continuation of this work came three years later in the highly cited study by Sienkiewicz and Hołyst (2005) analyzing bus systems in Poland. This study focuses heavily on the application of complex network science concepts on a real-world network. Unlike Musso and Vuchic (1988) in their work, the authors do not focus on the transport characteristics of these bus networks at all, but simply treat them as complex networks. A very similar analysis was performed on the Chinese rail network (Li and Cai, 2007) which confirmed that railway network to be scale-free. Majima, Katuhara, and Takadama (2007) instead took a more transport-related approach and performed a complex network analysis to evaluate the potential of a waterbus in the Tokyo area when combined with other modes. Researchers from Korea performed a similar study applied to the subway system of Seoul in 2008 (Lee et al., 2008). Their study focused on the analysis of statistical and topological properties and also included passenger flow.

Von Ferber and Derrible took the field to a new level with extensive studies on various network aspects of public transport networks worldwide. One of the earlier works is by Ferber et al. (2009). The goal of their work was twofold: define public transport networks based on statistical properties; and build a model that can create a network that would reproduce these properties based on a few simple rules. These statistical properties were identified for fourteen public transport networks. The study identified some unexpected similarities and differences between networks and created a model aimed to capture those rules. Derrible and Kennedy (2009) analyze nineteen metro networks worldwide, showing a relationship between network design and ridership. Network design was modeled using three indicators: coverage, directness and connectivity. The study shows a strong correlation between these aspects and the ridership of the metro networks, suggesting the effect of network design on ridership is significant. Derrible and Kennedy (2010a) similarly to Derrible and Kennedy (2009) try to characterize metro networks using indicators. This work specifically focuses on adapting graph theory concepts into well defined public transport-specific applications and use this method to characterize networks. The three dimensions used in this study are: state, form and structure. In similar work from the same year (Derrible and Kennedy, 2010b), the authors aim to address topological networks in complex network science using the same data set of 33 networks. The interest here lies in the complexity of the networks and the effects thereof on the robustness. The networks in this study are not actually compared to each other nor does the study consider service characteristics. The third of the works by Derrible on network science in metro networks specifically focuses on network centrality (Derrible, 2012). An alternative graph representation using only terminal and transfer nodes was used to compare 28 metro networks on their betweenness centrality. While thematically similar, the work by Derrible and Kennedy (2011) takes a different approach and instead reviews the existing literature on applying graph theory and network science to the field of transit network design. At the time, this application was still fairly new and unexplored but has since seen a huge increase in popularity. The study provides a useful overview of the indicators found in other papers that could be applied to public transport networks. A similar review was created by Lin and Ban (2013).

In later years, the field was also further explored by transport scientists appying network science concepts in order to compare PT networks to each other or to better understand specific networks. A good example of this approach is the paper by Haznagy et al. (2015) where the urban PT systems of five Hungarian cities are analyzed. In their work, they find interesting similarities even in cities with varying morphologies. Their discoveries of this independence within Hungary are supported by different researchers in other regions such as China (Xu et al., 2013) and Poland (Sienkiewicz and Hołyst, 2005) suggesting that these findings apply globally. Wu et al. (2017) compare six of the world's biggest metro networks using a new centrality measure called "node occupying probability". The study finds that these networks perform better under random attack than targeted attack and that some variation exists between these networks on the performance of these measures. Shanmukhappa et al. (2019) similarly to Derrible and Kennedy (2011) and Lin and Ban (2013) review the current developments in PT network analysis in 2019. They recommended performing more in-depth analysis of PT networks to get proper insights, as simple topological analyses of the graph simplification of a network does not provide many practical insights.

2.2 Service information

The second topic of interest in literature, is that of service information. This field has been upcoming for the past few years where the network theory as described in Section 2.1 is combined with the actual service or operation of a PT network. The studies in this field are quite varied in their nature and approach as the topic can be addressed from many different perspectives.

One approach of how to combine network topology and service information is by considering the operation or ridership of a metro system. Xu, Mao, and Bai (2016) incorporate trip data in order to create a weighted passenger flow network of the Beijing subway system. Through this new methodology, the researchers were able to identify the spatial mobility patterns in the urban area, providing useful insights for policy makers. Saidi et al. (2017) similarly use generalized passenger travel costs combined with network theory to compare the urban rail transit systems of six large cities. Amini et al. (2016) take a slightly different approach. Instead, they investigate which type of urban network structure is most suitable for a city, depending on its traffic conditions. Their study reveals that the most suitable network structure depends on the amount of traffic in the city, indicating a relationship between network structure and traffic volumes. Ingvardson and Nielsen (2018) look how network topology (among other factors) influences the ridership of public transport systems. Their study finds that extensive rail network coverage (among which metro was counted as well) is positively correlated with ridership. This correlation seemed largest for metro, because of its high passenger-carrying capacity. Their study does not find significant influence of network topology indicators on ridership. Having significant transfer possibilities does however increase the mobility of the system, leading to a higher attractiveness of the entire system.

Another approach is to consider how the structural topology and service of a metro system relate to its accessibility. A first example of this is by Luo et al. (2019) who integrate accessibility and network science, using a network science-based approach. They formulate an accessibility indicator of travel impedance based on the generalized travel cost (GTC), a combination of in-vehicle travel time, waiting time and the number of transfers. Their analysis reveals that, when including these service properties, there is a higher spatial disparity in PT accessibility. In addition, their work shows that larger networks exhibit a larger average travel impedance. A work in the following year by the same authors takes this integration a step further by evaluating whether passenger flow distribution can be estimated by looking solely at the network properties of a PT system (Luo, Cats, and Lint, 2020). Their work reveals that some network indicators can indeed be used to estimate passenger flow with reasonable accuracy. This is especially the case for indicators using weighted graphs and those related to the space-of-service, indicating the higher level of realism of those representations. Jin et al. (2017) take a more comprehensive approach and consider the accessibility of the whole high-speed rail (HSR) network of East Asia, comparing multiple countries. They suggest that further development of an integrated East Asia HSR system could reduce the travel time between the major cities but admits the impacts are lower than national developments and that there are many hurdles to realize this.

3 Methodology

In this chapter, the methodology and implementation applied for this study will be shortly discussed. Firstly, the network representations for this study will be discussed in Section 3.1. Afterwards, the total travel time method applied in this study is described in Section 3.2. Lastly, this chapter is concluded in Section 3.3 with an explanation of the comparison to the network and city aspects.

3.1 Network representations

The focus of this study is to create a comparison of metro networks that also includes the service information quantitatively. Similar to other studies, this is done by representing the metro networks topologically using concepts from graph theory. From these graph representations, metrics can be computed that can be used to compare the networks. In this section, the chosen representations and their implementation are extensively described.

This study will make use of two network representations: L-space and P-space. L-space, or space-of-infrastructure, is used to define the infrastructural lay-out of the network. It is explained in more detail in Section 3.1.1. The P-space, or space-of-service, is used to define the service layer of the network. This representation is especially important for this study as it can incorporate additional service information not captured in L-space. The P-space is discussed in more detail in Section 3.1.2. The data used for this study is briefly described in Section 3.1.3.

3.1.1 L-space

As mentioned in the introduction of this section, the L-space or alternatively space-of-infrastructure is a topological representation of the infrastructure of the network. In this representation, stations are represented as nodes while the tracks/tunnels between them are represented as edges. This is the common way to represent L-space where the information about the different metro lines/routes is lost and only the infrastructure remains. Within this basic concept of L-space, there are two factors that can differ per implementation: edge weight and edge direction.

Firstly, for this study it was decided to use average station-to-station in-vehicle travel time as the L-space edge weight. In this way, the L-space represents not just the physical lay-out of the network but also part of its operation, namely the travel time between stations. This weight is a combination of the distance, the type of vehicles used (i.e., as they generally determine the maximum speed of operation) as well as the infrastructural intricacies of the network (e.g., sharp bends, speed limits because of external factors).

Secondly, it is important to establish whether to use directed or undirected edges in L-space. Whereas most studies detailed in Chapter 2 use an undirected implementation for L-space, a directed implementation is chosen for this study. The reasons for this are twofold. Firstly, while metros nearly always travel in both directions between any given stations, within certain metro networks (e.g., Paris), one-directional segments do exist. Secondly, the travel time might differ per travel direction, because of for example elevation differences. As such, this study will use an L-space with directed edges. An example L-space representation for the network of Marseille is provided in Figure 1.

Figure 1: *The L-space representation for the metro network of Marseille*

In conclusion, the L-space consists of a set of nodes representing stations and a set of directed edges representing the rail infrastructure between the stations. The edges are weighted with the in-vehicle travel time between the two stations for that direction. L-space can thus be defined as a graph $\mathrm{L} = (V, E, w)$ where V = a set of nodes, $E \subseteq \{(i,j) | (i,j) \in V^2 \wedge i \neq j\}$ with main edge weight $w : (i,j) \rightarrow \mathrm{ivt}(i,j)$. An additional edge label $\mathrm{l}(i,j)$ is defined to represent the direct (i.e. geodesic) distance between stations to compute the total network length.

3.1.2 P-space

Whereas L-space describes the physical, infrastructural state of the network, the information about the service (e.g., where the lines are, which stations can be transferred at) is lost. The P-space, or space-of-service, is the chosen space to represent the actual service that is run on the infrastructure described by the L-space. In this representation, similarly to L-space, the stations are represented as nodes. An edge between two stations represents the fact that those two stations share a line and thus have a direct means of travel between them (i.e., without a transfer). Through this representation it is for example possible to represent the waiting time and number of transfers for a travel path. For P-space, the same types of decisions have to be made as for L-space in terms of the edge weight and direction. These choices are briefly described below.

Considering the aim of this study to include service information in a topological comparison, it is sensible to include as much information in the space of service as possible. As such, a weighted P-space implementation will be used. Specifically, the average waiting time will be used as the edge weight. This weight turns the frequency of a line into an average waiting time instead. This is done through the following formula:

$$w : (i,j) \rightarrow \mathrm{wt}(i,j) = \frac{60}{(veh_{tot} \div p)} \div 2$$

In this formula veh_{tot} represents the total amount of vehicles travelling on the line in a set time period p. This could for example be the total amount of vehicles in a full day of 24 hours. This amount of vehicles is then divided by the period to get the vehicles per hour. Dividing 60 by the vehicles per hour, gives the maximum waiting time in minutes. This maximum waiting time is in turn divided by 2 in order to get the average waiting time.

Using this average waiting time as an edge weight means the P-space provides information on both the number of transfers needed as well as the average time travellers have to wait for each transfer they make. As such, the P-space now provides a lot of information on the service and scheduling of the network. Combining this waiting time and number of transfers with the in-vehicle time from L-space, provides a fairly comprehensive image of the travel time that travellers actually experience.

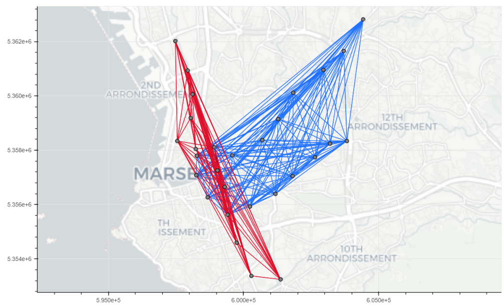


Figure 2: *The P-space representation for the metro network of Marseille*

In addition to the edge weight, the decision must also be made whether to use directed edges or not. Considering that P-space, similar to L-space, uses edge weights that might differ per travel direction, it is sensible to also use directed edges in P-space. Examples of how the average waiting time might be different per travel direction could be extra express trains that only travel in one direction thus lowering the average waiting time for that direction. The P-space representation for the metro network of Marseille can be found in Figure 2.

In conclusion, the P-space contains the same nodes as L-space, representing each station. Edges exist between each station that is connected by the same line. The weight of these edges represents the average waiting time on that line between the two stations. P-space can thus be defined as a graph $P = (V, E, w)$ where V = a set of nodes, $E \subseteq \{(i,j)|(i,j) \in V^2 \wedge i \neq j\}$ and $w : (i,j) \rightarrow \text{wt}(i,j)$. An additional label $\text{r}(i,j)$ exists to represent the route of each edge.

3.1.3 Data

In order to create the network representations as described in Sections 3.1.1 and 3.1.2, a significant amount of scheduling data is required. In addition, as described in Chapter 1 the goal is to include as many different networks as possible. For these reasons, the General Transit Feed Specification (GTFS) data format was chosen for this study (Google, 2022), specifically the GTFS static format. This format contains all the data necessary for this study and is publicly available for many metro networks. In addition, many libraries exist to process this data, easing the data processing stage of this study. 51 networks metro networks were found to have GTFS static data that is both correct and available. All of these networks were included in the final data set and are visualized in Figure 3.

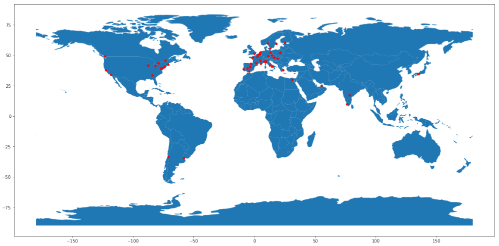


Figure 3: *The 51 metro networks included in the analysis indicated in red*

3.2 Total travel time method

As becomes clear from Section 3.1, information about the operation and service is included in both network representations for this study. In order to compare networks based on these representations, the data from these representations must be transformed into comparable metrics. The primary metric for this study is an accessibility metric based on the average shortest path through the network. For this study, the total travel time method is used which combines several aspects of the travel time through the network. To be more precise, the total travel time is a combination of the in-vehicle travel time, waiting time and number of transfers. This can be mathematically defined as described in Equation 1 below.

$$tt(i,j) = ivt^{\text{L}}(i,j) + \alpha * wt^{\text{P}}(i,j) + \beta * tf^{\text{P}}(i,j) \quad (1)$$

$tt(i,j)$ - The total (shortest path) travel time between node i and j (min)
$ivt^{\text{L}}(i,j)$ - The total in-vehicle travel time between node i and j from L-space (min)
$wt^{\text{P}}(i,j)$ - The total waiting time between node i and j from P-space (min)
$tf^{\text{P}}(i,j)$ - The number of transfers needed in the shortest path between node i and j from P-space (-)
α - A positive integer, constant penalty per minute of waiting time (min/min)
β - A positive integer, constant penalty per transfer (min/transfer)

The shortest path between two arbitrary nodes i and j is thus the path with the lowest total travel time which is a combination of the factors $ivt^L(i,j)$, $wt^P(i,j)$ and $tf^P(i,j)$. Considering these travel time components come from different network representations, they cannot be computed simultaneously. As such, the k-shortest paths based on in-vehicle time in L-space are taken as the base. The corresponding waiting time and transfer information are then retrieved from P-space for all of these paths. The addition of these three components (multiplied by the respective penalty factors) then determines which path has the shortest total travel time and is thus considered the shortest path. For this study, a k of 5 is chosen as that covers all shortest paths for nearly all networks while also keeping the runtime sufficiently low. The penalty values are taken as 2 and 13 for $alpha$ and $beta$ respectively, based on the lower bounds found in recent studies (Jara-Diaz et al., 2022).

Given a directed graph G, a real-valued weight function $w : E \rightarrow \mathbb{R}$, the shortest path between two arbitrary nodes i and j is the path $SP = (v_1, v_2, ..., v_n)$ (where $v_1 = i$ and $v_n = j$) that over all possible n minimizes the sum $\sum_{x=1}^{n-1} w(v_x, v_{x+1})$. This general notation can be adopted for the calculation of the shortest path calculation in L-space as follows:

$$ivt^{\text{L}}(i,j) = \min \sum_{x=1}^{n-1} \text{ivt}(v_x, v_{x+1}) \quad (2)$$

This calculation provides both a shortest path SP in L-space between node i and j as well as the corresponding in-vehicle time $ivt^L(i,j)$. The waiting time and number of transfers corresponding to this path SP can then be gained from P-space to determine the value of the total travel time. This is done through Equations 3 and 4.

$$wt^{\text{P}}(i,j) = \text{wt}(v_1, v_2) + \sum_{x=2}^{n-2} \begin{cases} \text{wt}(v_{x+1}, v_{x+2}), & \text{if } \text{r}(v_x, v_{x+1}) \neq \text{r}(v_{x+1}, v_{x+2}) \\ 0, & \text{if } \text{r}(v_x, v_{x+1}) = \text{r}(v_{x+1}, v_{x+2}) \end{cases} \quad (3)$$

$$tf^{\mathrm{P}}(i,j) = \sum_{x=1}^{n-2} \begin{cases} 1, & \text{if } \mathrm{r}(v_x, v_{x+1}) \neq \mathrm{r}(v_{x+1}, v_{x+2}) \\ 0, & \text{if } \mathrm{r}(v_x, v_{x+1}) = \mathrm{r}(v_{x+1}, v_{x+2}) \end{cases} \tag{4}$$

Using this shortest path calculation, the total travel time between any two stations in the network can be calculated. The average total travel time in a network can in turn be used to define networks by. For this study, the average total travel time is taken as the median of all station-to-station total travel times. The median is taken as it is both less sensitive to outliers (compared to for example the mean) and retains the absolute values for a network (as opposed to the standard deviation or variance). Henceforth this median total travel time is simply referred to as "travel time" for brevity.

3.3 Comparison to network factors

In addition to the comparison using the average total travel time, the metro networks will also be compared using secondary metrics based on the network size. These indicators can help provide context and further insights into the differences found between the networks. In addition, the region of each network will also be included in the analyses to identify potential regional differences.

For the network size, three different indicators are considered: the number of stations (ns^{L}), the network length (nl^{L}) and the average direct station distance (d^{L}). All three of these indicators describe the size of the network in a different manner. The reason to include these indicators is to see whether a relationship exists between the travel time (as described by the primary indicator) and the network size. It is hypothesized that a positive correlation between these factors exists considering the intuitive notion of increasing travel times in larger networks and related studies (Luo et al., 2019) confirming this pattern. The mathematical description of these three indicators is provided below:

$$ns^{\mathrm{L}} = |V(\mathrm{L})| \tag{5}$$

$$l^{\mathrm{L}} = \frac{\sum_{(i,j)\in E(\mathrm{L})} \mathrm{l}(i,j)}{2} \tag{6}$$

$$d^{\mathrm{L}} = \frac{\sum_{i\in V(\mathrm{L})} \sum_{j\in V(\mathrm{L})} d(i,j)}{|V(\mathrm{L})|^2 - |V(\mathrm{L})|} \tag{7}$$

4 Results

In this chapter, the results of this study will be discussed. Firstly, in Section 4.1 the correlation between the travel time and other factors is briefly discussed. Secondly, a simple regression model is applied in Section 4.2 to estimate the travel time. This model is improved on with a multiple regression model in Section 4.3. The results chapter is concluded in Section 4.4 with a benchmark analysis of the total travel time method applied in this study.

4.1 Correlations

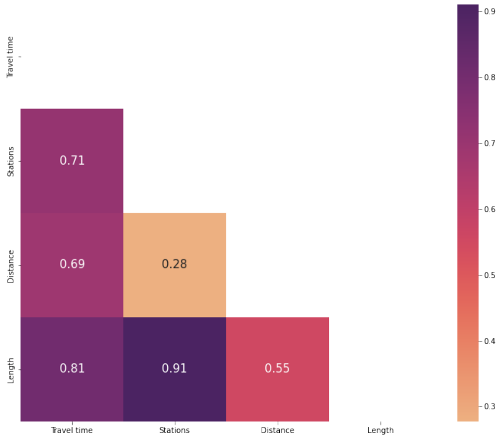


Figure 4: *A heatmap of the Pearson correlations between the four indicators*

In this section, the Pearson correlations between the four indicators are compared. From Figure 4, the correlations between the different indicators can be directly compared. The first thing to note is that all factors are positively correlated. This is in line with expectation, considering all factors relate to the increase of either the network or the city, which has been hypothesized to correlate with an increasing average travel time. Secondly, most correlations are moderate to strong with only the correlation between the number of stations and direct station distance being below 0.50.

The total travel time is especially strongly correlated with network length, having the second-highest correlation in the whole heatmap. The highest correlation is actually that between the network length and number of stations with a correlation of 0.91. This is fairly sensible considering how both of these indicators are descriptors of the network size and are very directly related. An increase in stations generally leads to an increase in the network length and viceversa. Interestingly, the direct station distance has the strongest correlation with travel time, and a weaker one with the other six factors. This is fairly sensible considering how both the travel time and direct station distance describe the average shortest path, in terms of duration and direct distance respectively.

Overall, it can be concluded that all three network factors have a strong positive correlation with the travel time, all having a correlation of higher than 0.65. The network length clearly has the strongest correlation and can thus best considered as the best single explanatory parameter for the total travel time.

4.2 Simple regression model

Considering the high correlation between the network length and the travel time, it is sensible to try and estimate the travel time using this factor. In this section, the possibility of linear regression using this single independent parameter will be explored. The performance of this simple regression model on three criteria can be found in Table 1, while the coefficient values and information about their significance can be found in Table 2. The best fit line between the travel time and network length can be found in Figure 5.

Criterion	Value
R-squared	0.653
Adj. R-squared	0.646
BIC	375.2

Table 1: *The performance of the simple regression model*

Parameter	Coefficient	P-value
Intercept	26.990	0.000
Length	0.132	0.000

Table 2: *The parameters of the singular regression model including coefficients and p-values*

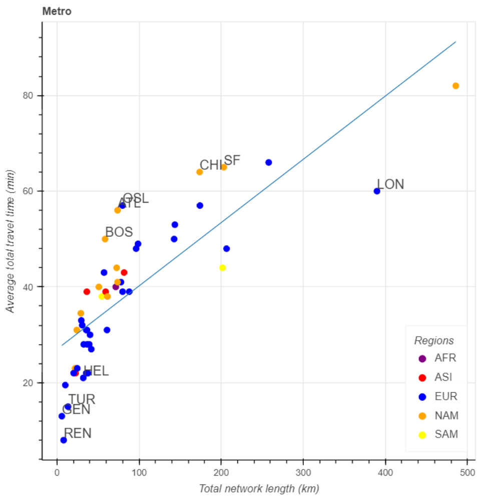


Figure 5: *The best fit line between the total travel time and the network length*

As expected from the high correlation, this line fits reasonably well. The R-squared of this regression is 0.653 and both the intercept and the coefficient for network length are significant at a 5% confidence interval. From Figure 5 it becomes clear however that this model is far from perfect. While the line fits fairly well, there is still a fair number of large outliers. To be precise, there are ten networks that are either over- or underestimated by more than 10 minutes in this model. The five networks that are overestimated by this model are Rennes (-20.1), London (-18.5), Genoa (-14.8), Turin (-13.8) and Helsinki (-10.2). On the other hand, Oslo (+19.5), Atlanta (+19.25), Boston (+15.3), Chicago (+14.0) and San Francisco (+11.2) are the networks that are most strongly underestimated. Looking at these major outliers as well as Figure 5, some regional differences appear to be present. These differences are shortly discussed below.

4.2.1 Regions

Some notes can also be made about the structural regional differences that can be identified in Figure 5. Firstly, it is interesting to note how ten out of the twelve North-American networks are underestimated by the model, which suggests that these networks have a relatively high total travel time compared to other networks. The opposite can be said for European networks, considering nearly all of the networks overestimated by the model are European. The other three regions have a relatively low number of networks and no notable outliers. This apparent regional differentiation can also be accounted for in the model using dummies for each region. Experimentation shows that only Europe and North-America provide significant parameters (at a 5% level), when considered separately. The results for different models including North-America and Europe as regions separately as well as combined can be found in Table 3. From this table, it becomes clear that as expected, including either region improves the model slightly. The effects of North-America are slightly stronger than those of Europe, which is in line with the intuitive findings from Figure 5. The BIC of the model including North-America is 3 points higher which, while positive, is not especially strong.

Region	R^2	Adj. R^2	BIC	Coeff.	P-value
None	0.65	0.65	375	-	-
NAM	0.70	0.69	372	7.7	0.009
EUR	0.70	0.68	373	-6.4	0.014
C. (NAM)	0.70	0.69	375	5.0	0.223
C. (EUR)	0.70	0.69	375	-3.2	0.368

Table 3: *A comparison of regression models including regional dummies*

4.3 Multiple regression model

An improvement over the simple regression model from Section 4.2 is using a multiple regression model with more factors. It is most sensible to consider parameters that have a high correlation with travel time but a low one among other. Figure 4 shows that the number of stations and direct distance are the best candidates. Both of these factors have a correlation of 0.77 with travel time while they have a relatively low correlation of 0.52 amongst each other. The results of a model including these two parameters can be found in Tables 4 and 5.

Criterion	Value
R-squared	0.768
Adj. R-squared	0.758
BIC	358.6

Table 4: *The general information of the two-factor regression model*

Parameter	Coefficient	P-value
Intercept	15.527	0.000
Stations	0.109	0.000
Distance	1.819	0.000

Table 5: *The parameters of the two-factor regression model including coefficients and p-values*

From Table 4 it becomes clear that the two-factor model has a much better R-squared, adjusted R-squared and BIC. In addition, all three parameters are significant at the 5% level. As a network with no stations and no length does not exist, the coefficient value of the intercept cannot be properly interpreted. What can be concluded however, is that metro networks have a base total travel time of 15.5 minutes increasing by 0.1min per extra station and 1.8min per extra kilometer of average distance between stations. This means that as networks get bigger by adding more stations and getting more spread out, the travel time increases accordingly. This finding is in line with that of other studies such as Luo et al. (2019). While this model has a better fit than the single parameter model, there still are some outliers which will be shortly highlighted below.

There are four networks that have a much higher actual travel time than their predicted travel time, being Chicago (+13.0), Atlanta (+15.0), Boston (+16.0) and Oslo (+16.0). These are the same four networks as identified in Section 4.2. As such, in addition to having a long travel time relative to their network length, these networks also have a long travel time relative to their number of stations and direct station distance.

Interesting to note is the San Francisco network, which was greatly underestimated by the length-based model (+11.2) but is actually strongly overestimated by this model (-8.3). Looking into the data, it becomes clear that San Francisco has an exceptionally long average direct station distance which is the cause of this overestimation. San Francisco in fact has a direct station distance that is 13km longer than any other. This begs the question whether the San Francisco network should even be classified as a metro network, considering it is so different from all other networks in terms of its direct station distance. Instead, it can better be defined as a commuter rail.

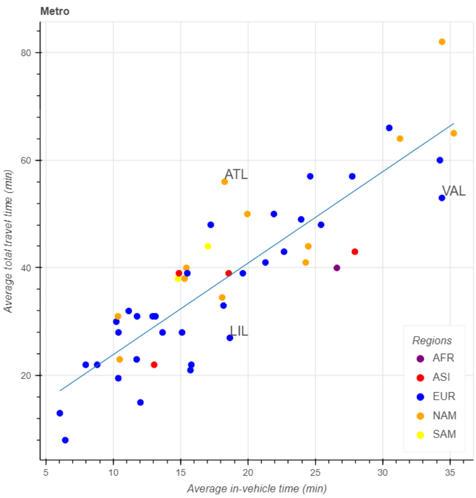
There are also five networks that have a much lower actual travel time than predicted, being London (-10), Paris (-11), Bilbao (-11), Helsinki (-13) and Rennes (-14). Helsinki, Rennes and London also were overestimated outliers in the length-based model, and as such can simply be considered as networks that have a low travel time for their network size. Bilbao was also an overestimated outlier in the length-based model (-9.98), albeit slightly less so than Helsinki and Rennes. Genoa and Turin, two networks that were greatly overestimated by the length-based model (-14.8 and -13.8 respectively) are still overestimated here but much less so (-6.7 and -9.5). It thus seems that while this model is better at predicting the travel time for these two networks, they are still outliers having a relatively low travel time for their network size. Paris was also overestimated by the length-based model (-6.3) but is overestimated considerably more in this model (-11). The data reveals that Paris has a relatively low travel time for its number of stations. It can thus be concluded that Paris performs especially well for its number of stations.

In conclusion, it can be noted that the model based on the number of stations and average direct station distance can predict the total travel time of networks to an acceptable degree. In this model, there are still networks that are significantly over- or underestimated in terms of their travel time. These networks can thus be said to perform well or poor respectively based on their network size metrics.

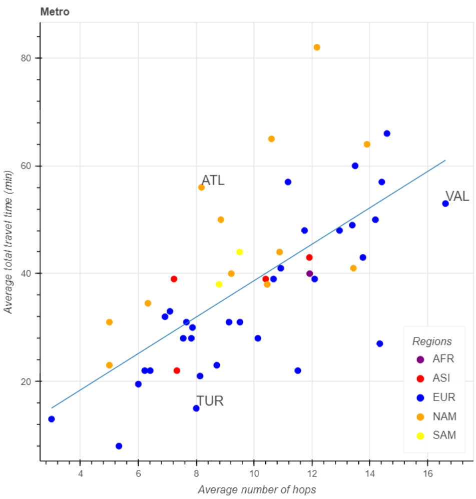
4.4 Travel time calculation benchmark analysis

In addition to comparing the average travel time to other factors, the methodology applied in this study is also compared to other state-of-the-art methods. To be more precise, the total travel time method (henceforth simply referred to as "total" method) is compared to the "in-vehicle" and "hops" methods. The in-vehicle method takes the shortest path travel time based purely on the path with the least in-vehicle travel time. The hops method instead considers only the amount of hops between stations as the shortest path. Both of these methods are widely applied in literature as they do not require as much data as the total method. In this section, these two alternative methods are directly compared to the total method to identify the differences in their representation.

The comparison between the total and in-vehicle and hops methods can be found in Figures 6a and 6b respectively. Both of these figures show a clear positive correlation, which is confirmed by the numerical analysis, with a Pearson correlation of 0.87 for in-vehicle and 0.68 for hops respectively. While these correlations are both strong and positive, the hops method is clearly weaker correlated. The difference in strength of correlation also becomes clear from Figure 6 im-



(a) *Total vs in-vehicle*



(b) *Total vs hops*

Figure 6: *The total travel time method compared to the in-vehicle time and hops method*

mediately with the hops method comparison having a considerably higher spread than the in-vehicle method comparison. Nevertheless, the spread is quite significant for both of these methods in both the vertical and horizontal direction. Some examples are shortly highlighted to indicate this difference. Atlanta (ATL) is an excellent example in both comparisons of how the two alternative methods can misrepresent networks. In Figure 6a it can be seen how Atlanta has nearly the same in-vehicle travel time as Lille (LIL) but a completely different total travel time (56 and 27 minutes respectively). Using the in-vehicle method would suggest Atlanta and Lille are similar, while the total travel time method shows they are most certainly not. On the other hand, Atlanta and Valencia (VAL) have nearly the same total travel time, but a very different in-vehicle travel time (18 and 34 minutes respectively). Using the in-vehicle method would suggest that Valencia has much longer travel times while this is in fact not the case when incorporating waiting time and transfers. This pattern can also be seen for Atlanta in the comparison with the hops method, where the variance is even higher. Valencia also has a significantly higher number of hops than Atlanta, while having the same total travel time. In terms of other networks with the same amount of hops as Atlanta, Turin is a notable example with a much lower total travel time (57 vs 15 minutes). These examples show how differently these three methods describe networks.

5 Conclusion & Discussion

In this final chapter, the conclusions and recommendations based on the study are provided. The first findings to discuss are related to the main research question: "How can service information be included into a comprehensive comparison of metro networks worldwide?". This study shows how service information can be efficiently included as concepts of in-vehicle time, waiting time and number of transfers in the calculation of shortest travel path travel times through the network. Through this method, a comparison between metro networks was made that includes this service information. The median total travel time was used as the indicator to incorporate service information.

Various network factors were used to help explain differences in total travel times between networks. Firstly, the correlations between the travel time and these factors were investigated. As hypothesized, all factors have a positive correlation with the travel time. From these factors, the network length has the strongest correlation with the travel time. As such, the network length was used in a simple regression model to estimate the travel time. Through this, a variety of outlier networks could be identified that perform either much better or worse than predicted according to the network length. In addition, this regression model suggested that North-American networks perform structurally worse compared to other networks, especially worse than European ones. Including this regional variation in the regression model turned out to be significant and improved the model slightly.

Multiple regression was also applied in order to further improve the model. A model was created with two factors with a high correlation with travel time, but low amongst each other: the number of stations and the average direct station distance. This model provided significant coefficients and performed much better than the simple regression model. Nevertheless, this model showed some of the same outliers as the simple model, albeit reduced. Paris, Bilbao, Helsinki, Rennes and London seem to perform much better than predicted by the model while Chicago, Atlanta, Boston and Oslo perform much worse. The most notable outlier is San Francisco, which is such an anomaly in terms of both its direct station and interstation distance, that it should not be considered a metro network, but a commuter rail instead. The discovered regional difference was not significant at the 5% level in this model.

Lastly, it was evaluated how the uncommon total travel time method applied in this study compares to the more common methods of using only in-vehicle time or hops in shortest path calculations. While all three methods show a positive correlation with network size, the method proposed in this study clearly has a much stronger correlation than the other methods. In addition, it becomes clear that networks are represented very differently by these three methods. Some networks have a similar amount of average hops while their average total travel time differs by more than thirty minutes. These examples clearly show how different these methods are in representing networks. Considering the more realistic nature of the total travel time method, it becomes clear that applying that method is desirable for better comparisons of networks.

In a general sense, the results of this study provide a benchmark for network planners. Through the results of this study, planners can understand which networks share the same characteristics in terms of different indicators, especially average travel time. This provides them with a clear idea of which networks to consider on how to improve their own network. This study also provides insights into which metrics accurately represent travel time and which do not. Whereas the network length and the combination of number of stations and direct station distance turn out to be good metrics for comparison, the other explored factors are not as much. This means network planners have a better idea of which factors to focus on for comparison. The total travel time method applied in this study also emphasizes the importance of considering all travel time aspects of the travellers' journey. Including waiting time and number of transfers in network representations provides better insights into which aspects of the travel time have most room for improvement.

In terms of limitations, the first obvious limitation is the consideration of metro networks as a distinct, separated mode. Considering the complicated nature of multi-layer networks, the decision was made to restrict this study exclusively to metro networks. In terms of the actual networks included in this analysis, the set of 190 possible metro networks was limited to 51 because of data restrictions. While the total travel time method applied in this study is more representative than the in-vehicle or hops method, it still relies on some assumptions and simplifications. The waiting time might in fact be much lower for some networks as transfer possibilities are created by planners that are well-connected in the schedule. Walking times are also not specifically considered in this study which can also affect the actual transfer times experienced by travellers. In addition, all transfer stations are considered equal in this study while some are much bigger than others, leading to different transfer times. Finally, the median value taken as the average of the travel time distributions might not be representative of the average traveler's experience. As this value is the median of all possible OD-pairs, this might overestimate some hardly used OD-pairs while underestimating frequently used OD-pairs.

For future work, the results of this study could be explored into more detail by applying for example clustering or principal component analysis. Other network science indicators such as directness and degree can be investigated both locally as well as globally to discover further relations with the travel times. The methodology could also be expanded upon by creating multilayer networks that include multiple PT modes. In addition, the data provided by this study could be combined with actual traveler data such as OD-pair demand data. The data from this study could help explain certain travel behavior or identify bottlenecks in the network. Lastly, the results of this study could be related to existing benchmark studies between networks to get a more comprehensive image of the relationships between networks. An example of this is a report by McKinsey (McKinsey, 2021) which provides a complementary qualitative assessment that could be combined with this study's quantitative assessment.

Bibliography

Amini, Behnam et al. (2016). "Comparative analysis of traffic performance of urban transportation systems". In: *Transportation Research Record* 2594, pp. 159–168. ISSN: 03611981. DOI: 10.3141/2594-19.

Derrible, Sybil (July 2012). "Network centrality of metro systems". In: *PLoS ONE* 7 (7). ISSN: 19326203. DOI: 10.1371/journal.pone.0040575.

Derrible, Sybil and Christopher Kennedy (2009). "Network analysis of world subway systems using updated graph theory". In: *Transportation Research Record* (2112), pp. 17–25. ISSN: 03611981. DOI: 10.3141/2112-03.

Derrible, Sybil and Christopher Kennedy (Mar. 2010a). "Characterizing metro networks: State, form, and structure". In: *Transportation* 37 (2), pp. 275–297. ISSN: 00494488. DOI: 10.1007/s11116-009-9227-7.

– (Sept. 2010b). "The complexity and robustness of metro networks". In: *Physica A: Statistical Mechanics and its Applications* 389 (17), pp. 3678–3691. ISSN: 03784371. DOI: 10.1016/j.physa.2010.04.008.

– (July 2011). "Applications of graph theory and network science to transit network design". In: *Transport Reviews* 31 (4), pp. 495–519. ISSN: 01441647. DOI: 10.1080/01441647.2010.543709.

Ferber, C. Von et al. (Mar. 2009). "Public transport networks: Empirical analysis and modeling". In: *European Physical Journal B* 68 (2), pp. 261–275. ISSN: 14346028. DOI: 10.1140/epjb/e2009-00090-x.

Google (2022). *GTFS Static Overview*. URL: https://developers.google.com/transit/gtfs/.

Haznagy, Andor et al. (2015). "Complex network analysis of public transportation networks: a comprehensive study". In: p. 536. ISBN: 9789633131428.

Ingvardson, Jesper Bláfoss and Otto Anker Nielsen (Oct. 2018). "How urban density, network topology and socio-economy influence public transport ridership: Empirical evidence from 48 European metropolitan areas". In: *Journal of Transport Geography* 72, pp. 50–63. ISSN: 09666923. DOI: 10.1016/J.JTRANGEO.2018.07.002.

Jara-Diaz, Sergio et al. (Sept. 2022). "An international time equivalency of the pure transfer penalty in urban transit trips: Closing the gap". In: *Transport Policy* 125, pp. 48–55. ISSN: 1879310X. DOI: 10.1016/j.tranpol.2022.04.019.

Jin, Fengjun et al. (May 2017). "Evolution and geographic effects of high-speed rail in East Asia: An accessibility approach". In: *Journal of Geographical Sciences* 27 (5), pp. 515–532. ISSN: 1009637X. DOI: 10.1007/s11442-017-1390-8.

Latora, Vito and Massimo Marchiori (2002). *Is the Boston subway a small-world network?*, pp. 109–113. URL: www.elsevier.com/locate/physa.

Lee, Keumsook et al. (Oct. 2008). "Statistical analysis of the Metropolitan Seoul Subway System: Network structure and passenger flows". In: *Physica A: Statistical Mechanics and its Applications* 387 (24), pp. 6231–6234. ISSN: 03784371. DOI: 10.1016/j.physa.2008.06.035.

Li, W. and X. Cai (Aug. 2007). "Empirical analysis of a scale-free railway network in China". In: *Physica A: Statistical Mechanics and its Applications* 382 (2), pp. 693–703. ISSN: 03784371. DOI: 10.1016/j.physa.2007.04.031.

Lin, Jingyi and Yifang Ban (Nov. 2013). "Complex Network Topology of Transportation Systems". In: *Transport Reviews* 33 (6), pp. 658–685. ISSN: 01441647. DOI: 10.1080/01441647.2013.848955.

Luo, Ding, Oded Cats, and Hans van Lint (Dec. 2020). "Can passenger flow distribution be estimated solely based on network properties in public transport systems?" In: *Transportation* 47 (6), pp. 2757–2776. ISSN: 15729435. DOI: 10.1007/s11116-019-09990-w.

Luo, Ding et al. (Oct. 2019). "Integrating network science and public transport accessibility analysis for comparative assessment". In: *Journal of Transport Geography* 80, p. 102505. ISSN: 0966-6923. DOI: 10.1016/J.JTRANGEO.2019.102505.

Majima, Takahiro, Mitujiro Katuhara, and Keiki Takadama (2007). *Analysis on Transport Networks of Railway, Subway and Waterbus in Japan*. URL: www.springerlink.com.

McKinsey (2021). *Urban transportation systems of 25 global cities Elements of success 2*.

Musso, Antonio and Vukan R Vuchic (1988). *Characteristics of Metro Networks and Methodology for Their Evaluation*.

Saidi, Saeid et al. (Nov. 2017). "A generalized framework for complex urban rail transit network analysis". In: *Transportmetrica A: Transport Science* 13 (10), pp. 874–892. ISSN: 23249943. DOI: 10.1080/23249935.2017.1348401.

Shanmukhappa, Tanuja et al. (Oct. 2019). "Recent Development in Public Transport Network Analysis from the Complex Network Perspective". In: *IEEE Circuits and Systems Magazine* 19 (4), pp. 39–65. ISSN: 15580830. DOI: 10.1109/MCAS.2019.2945211.

Sienkiewicz, Julian and Janusz A. Hołyst (Oct. 2005). "Statistical analysis of 22 public transport networks in Poland". In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 72 (4). ISSN: 15393755. DOI: 10.1103/PhysRevE.72.046127.

UITP (2022). *World Metro Figures 2021*.

Wu, Xingtang et al. (Feb. 2017). "Analysis of metro network performance from a complex network perspective". In: *Physica A: Statistical Mechanics and its Applications* 492, pp. 553–563. ISSN: 03784371. DOI: 10.1016/j.physa.2017.08.074.

Xu, Q., B. H. Mao, and Y. Bai (Mar. 2016). "Network structure of subway passenger flows". In: *Journal of Statistical Mechanics: Theory and Experiment* 2016 (3). ISSN: 17425468. DOI: 10.1088/1742-5468/2016/03/033404.

Xu, Qing et al. (2013). "Space P-based empirical research on public transport complex networks in 330 cities of China". In: *Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal of Transportation Systems Engineering and Information Technology* 13 (1), pp. 193–198. ISSN: 10096744. DOI: 10.1016/s1570-6672(13)60100-2.

B

Median travel time maps

In this appendix the median travel time for each network can be found split by region. The colors of each network correspond to the value of the travel time as described by the legend in Figure B.1. The respective regions are North-America (Figure B.2), Europe (Figure B.3) and the rest of the world (Figure B.4) which thus includes Africa, Asia and South-America.

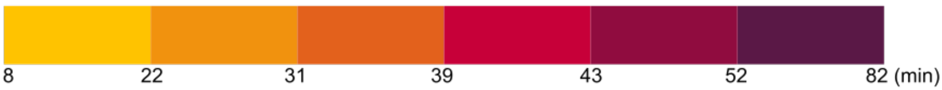


Figure B.1: The legend for Figures B.2, B.3, B.4 with the color for each range of values of the median travel time

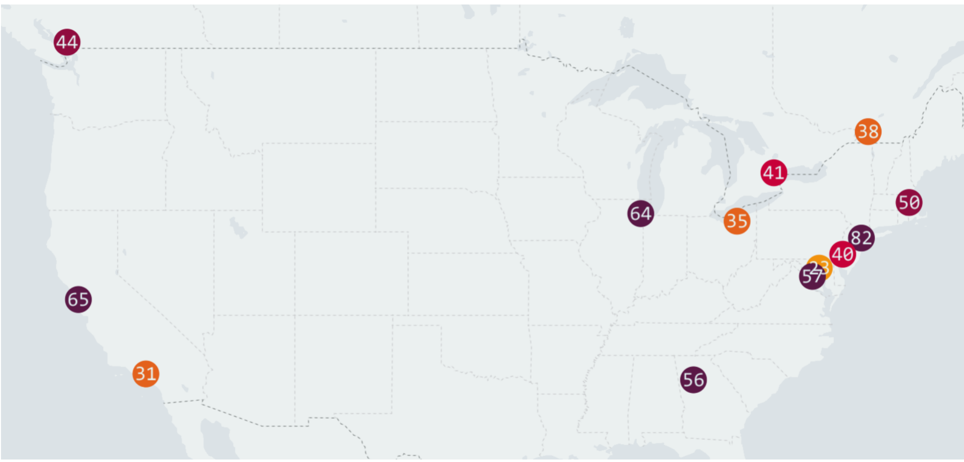


Figure B.2: The median travel time for the thirteen metro networks in North-America

B

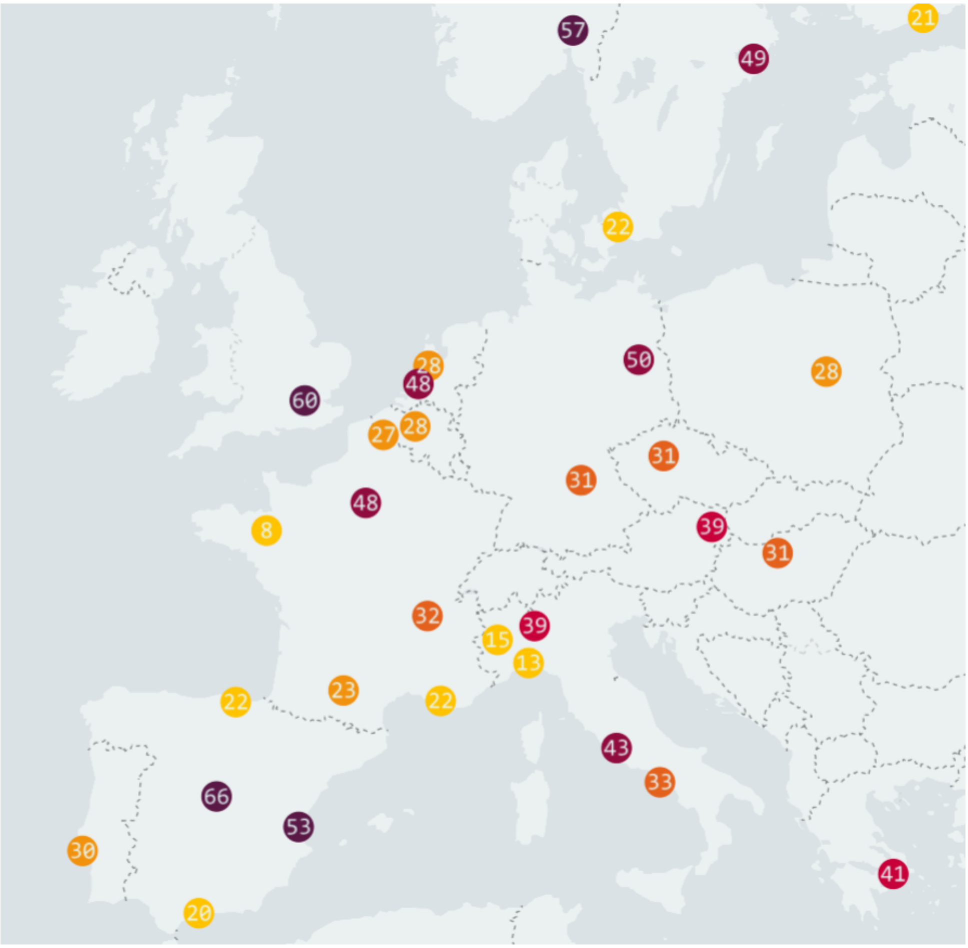


Figure B.3: The median travel time for the 31 metro networks in Europe

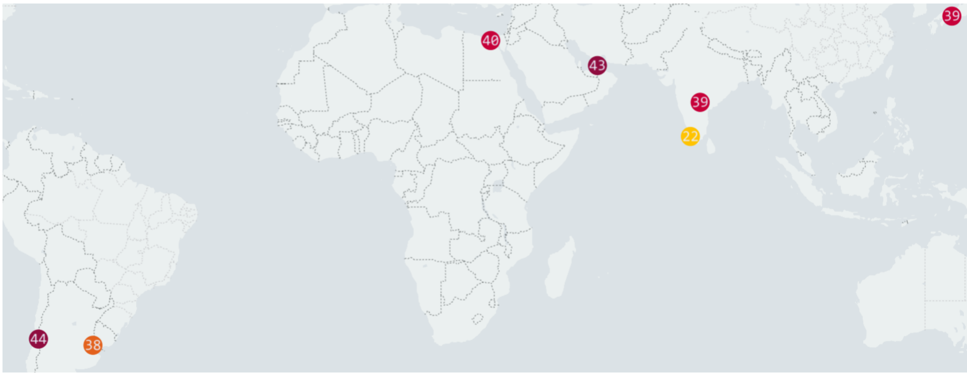


Figure B.4: The median travel time for the seven metro networks in the rest of the world

C

GENERAL NETWORK INFORMATION

Table C.1 contains the 51 networks that were included in the final data set. Along with their names is information about their region, country, abbreviation, number of lines, GDP and population. As described in Section 4.5, this data was all not retrieved from the GTFS data but externally instead.

Table C.1: General information for the 51 networks selected in the final data set

Name	**Region**	**Country**	**Short**	**Lines**	**GDP (Million €)**	**Population**
Amsterdam	EUR	NL	AMS	5	194067	996,915
Athens	EUR	GR	ATH	3	83199	2,622,404
Atlanta	NAM	US	ATL	4	430158.0777	6,144,050
Baltimore	NAM	US	BAL	1	208629.7637	2,710,489
Berlin	EUR	DE	BER	9	205625	3,644,826
Bilbao	EUR	ES	BIL	2	37164	792,617
Boston	NAM	US	BOS	3	478410.8022	4,941,632
Brussels	EUR	BE	BRU	4	175751.23	1,215,289
Budapest	EUR	HU	BUD	4	69658.93	1,752,286
Buenos Aires	SAM	AR	BUE	6	55288.7619	15,625,084
Cairo	AFR	EG	CAI	3	100156	20,500,000
Chicago	NAM	US	CHI	8	707468.811	9,618,502
Cleveland	NAM	US	CLE	1	134416.0513	2,088,251
Copenhagen	EUR	DK	COP	4	134633.4	623,404
Dubai	ASI	AE	DUB	3	116308.61	3,521,816
Genoa	EUR	IT	GEN	1	30436.24	569,184
Helsinki	EUR	FI	HEL	2	95445.5	648,042
Hyderabad	ASI	IN	HYD	3	73696	7,677,018
Kobe	ASI	JP	KOB	2	48589.53	1530000
Kochi	ASI	IN	KOC	1	60564	2,119,724

C

Lille	EUR	FR	LIL	2	85056	955,906
Lisbon	EUR	PT	LIS	4	77439.68	1,859,838
London	EUR	UK	LON	11	615043.14	8,866,541
Los Angeles	NAM	US	LA	2	1034458.982	13,211,027
Lyon	EUR	FR	LYO	4	93402	1,277,584
Madrid	EUR	ES	MAD	13	241040	5,012,504
Malaga	EUR	ES	MAL	2	32207	574,654
Marseille	EUR	FR	MAR	2	103783	969,002
Milan	EUR	IT	MIL	4	215177.43	3,606,653
Montreal	NAM	CA	MON	4	168638.68	4,104,074
Naples	EUR	IT	NAP	2	61721.64	2,867,858
New York	NAM	US	NY	36	1834722.195	19,261,570
Nuremberg	EUR	DE	NUR	3	67888	518,365
Oslo	EUR	NO	OSL	5	68653.21	623,966
Paris	EUR	FR	PAR	16	758527	10,277,625
Philadelphia	NAM	US	PHI	2	441727.3482	6,092,403
Prague	EUR	CZ	PRA	3	88338.82	1,324,277
Rennes	EUR	FR	REN	1	39020	271,686
Rome	EUR	IT	ROM	3	167450.47	2,820,219
Rotterdam	EUR	NL	ROT	5	88121	1,232,747
San Francisco	NAM	US	SF	6	583388.7276	4,709,220
Santiago	SAM	CL	SAN	7	89941.5	7,493,497
Stockholm	EUR	SE	STO	7	151284.4	1,745,766
Toronto	NAM	CA	TOR	4	320473.76	5,928,040
Toulouse	EUR	FR	TOU	2	59930	744,104
Turin	EUR	IT	TUR	1	75076.14	860,793
Valencia	EUR	ES	VAL	6	61499	1,403,247
Vancouver	NAM	CA	VAN	3	154280	2,463,431
Vienna	EUR	AT	VIE	5	137002.48	1,766,746
Warsaw	EUR	PL	WAR	2	93797.85	1,777,972
Washington	EUR	US	WAS	6	555554.1972	6,250,309

D

Network data

Table D.1 contains the data points computed for each of the 51 included networks. Note that the total travel time (TT) is the median value as explained in Section 5.1 while the in-vehicle time (IVT), waiting time (WT) and number of transfers (TF) are listed with their mean. The other data points are the total number of stations (Stations), the total network length (Length) and the average direct station distance (Distance).

Table D.1: Data points computed for all 51 networks

Name	**TT**	**IVT**	**WT**	**TF**	**Stations**	**Length**	**Distance**
Amsterdam	28 min	11 min	7 min	0.4	39	39 km	6 km
Athens	41 min	22 min	7 min	0.6	61	78 km	8 km
Atlanta	56 min	18 min	17 min	0.5	38	74 km	12 km
Baltimore	23 min	10 min	7 min	0	14	22 km	7 km
Berlin	50 min	23 min	7 min	1.3	174	143 km	8 km
Bilbao	22 min	16 min	3 min	0.2	42	38 km	7 km
Boston	50 min	20 min	9 min	0.8	52	59 km	7 km
Brussels	28 min	14 min	4 min	0.6	59	37 km	5 km
Budapest	31 min	12 min	4 min	0.8	48	37 km	4 km
Buenos Aires	38 min	16 min	4 min	1	78	55 km	4 km
Cairo	40 min	27 min	2 min	0.7	61	72 km	11 km
Chicago	64 min	32 min	10 min	1	137	174 km	11 km
Cleveland	35 min	18 min	9 min	0	18	29 km	10 km
Copenhagen	22 min	8 min	3 min	0.6	39	36 km	4 km
Dubai	43 min	28 min	5 min	0.6	53	82 km	16 km
Genoa	13 min	6 min	4 min	0	8	6 km	2 km
Helsinki	21 min	16 min	3 min	0.1	25	32 km	9 km
Hyderabad	39 min	19 min	5 min	0.6	56	59 km	8 km
Kobe	39 min	15 min	11 min	0.4	26	36 km	7 km

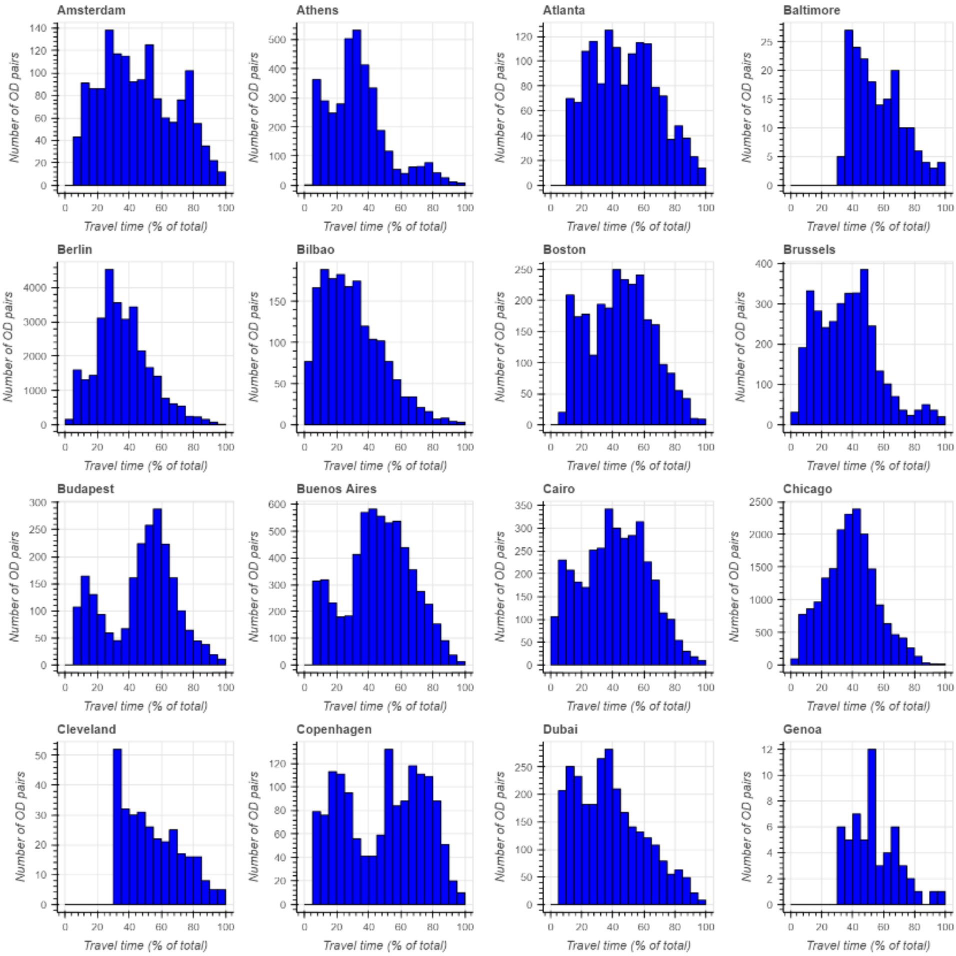
D

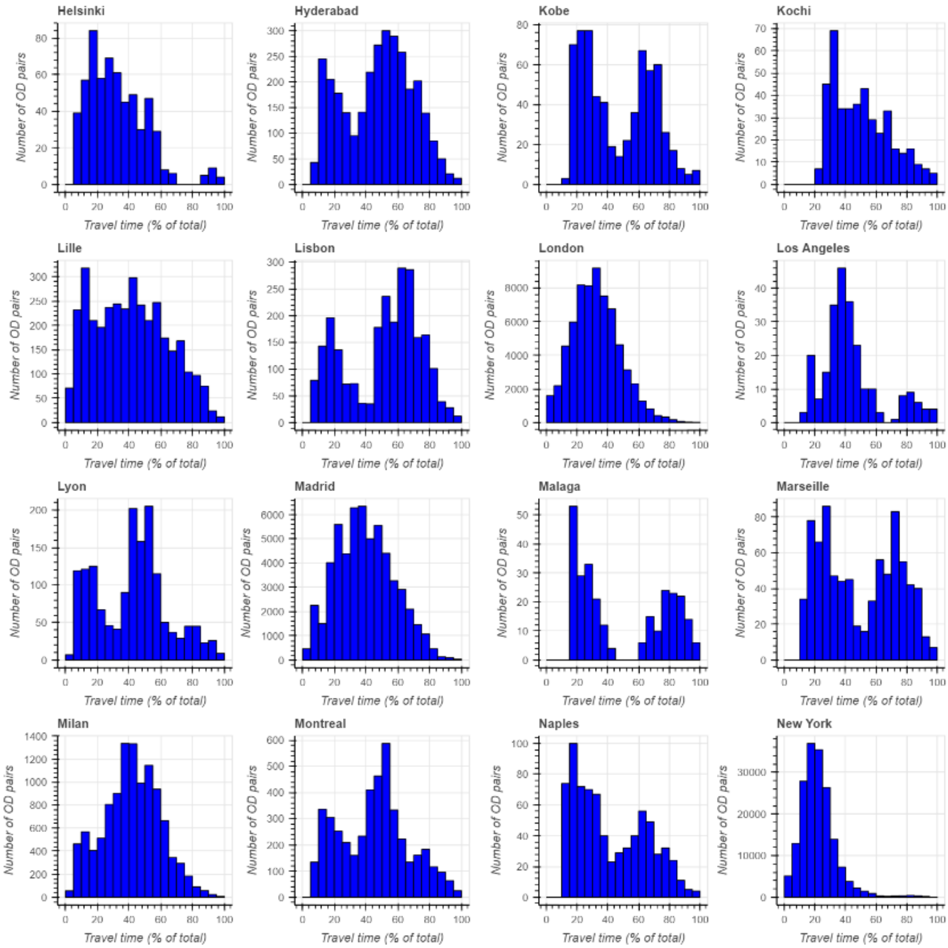
Cochin	22 min	13 min	5 min	0	21	23 km	7 km
Lille	27 min	19 min	2 min	0.4	60	42 km	7 km
Lisbon	30 min	11 min	3 min	0.7	50	40 km	4 km
London	60 min	35 min	5 min	1.5	261	390 km	14 km
Los Angeles	31 min	10 min	11 min	0.1	16	24 km	6 km
Lyon	32 min	12 min	4 min	0.8	40	31 km	4 km
Madrid	66 min	31 min	8 min	1.7	240	258 km	10 km
Malaga	20 min	10 min	6 min	0.4	17	10 km	3 km
Marseille	22 min	9 min	4 min	0.4	29	20 km	3 km
Milan	39 min	21 min	4 min	0.8	106	88 km	7 km
Montreal	38 min	16 min	6 min	0.7	67	61 km	6 km
Naples	33 min	18 min	8 min	0.4	28	30 km	4 km
New York City	82 min	35 min	16 min	1.8	421	486 km	12 km
Nuernberg	31 min	13 min	8 min	0.6	49	36 km	5 km
Oslo	57 min	25 min	13 min	0.6	101	80 km	8 km
Paris	48 min	18 min	5 min	1.7	303	207 km	6 km
Philadelphia	40 min	16 min	9 min	0.5	50	51 km	6 km
Prague	31 min	13 min	3 min	0.6	58	61 km	7 km
Rennes	8 min	6 min	1 min	0	15	8 km	2 km
Rome	43 min	23 min	6 min	0.9	73	57 km	8 km
Rotterdam	48 min	26 min	8 min	0.6	70	96 km	11 km
San Francisco	65 min	36 min	13 min	0.3	50	203 km	29 km
Santiago	44 min	17 min	7 min	1.2	119	202 km	9 km
Stockholm	49 min	25 min	7 min	0.7	101	99 km	8 km
Toronto	41 min	26 min	4 min	0.7	75	74 km	9 km
Toulouse	23 min	12 min	2 min	0.5	37	25 km	4 km
Turin	15 min	12 min	2 min	0	23	13 km	4 km
Valencia	53 min	35 min	7 min	0.7	95	144 km	15 km
Vancouver	44 min	25 min	6 min	0.8	52	73 km	11 km
Vienna	39 min	16 min	5 min	1	98	80 km	6 km
Warsaw	28 min	15 min	3 min	0.5	33	33 km	7 km
Washington	57 min	29 min	9 min	0.7	89	174 km	12 km

E

HISTOGRAMS

Figure E.1 below contains the (normalized) total travel time distributions for all 51 networks.


Amsterdam
Athens
Atlanta
Baltimore
Berlin
Bilbao
Boston
Brussels
Budapest
Buenos Aires
Cairo
Chicago
Cleveland
Copenhagen
Dubai
Genoa
Number of OD pairs
Travel time (% of total)


Helsinki
Hyderabad
Kobe
Kochi
Lille
Lisbon
London
Los Angeles
Lyon
Madrid
Malaga
Marseille
Milan
Montreal
Naples
New York
Number of OD pairs
Travel time (% of total)

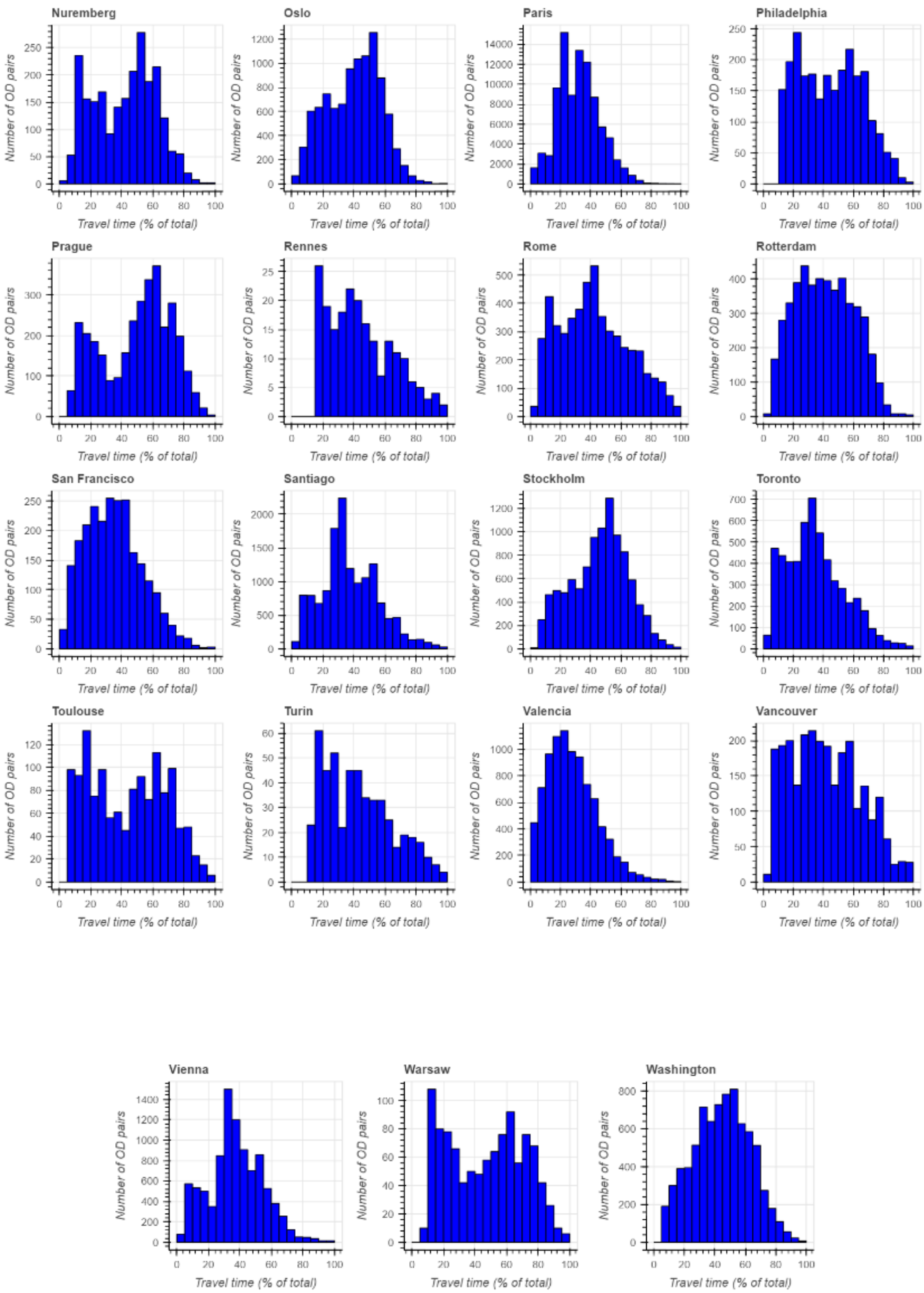


Figure E.1: The total travel time distribution for all 51 networks

F

PROCESSING

Table F.1 provides an overview of all 51 networks included (and nine that were excluded). The "time spent" column refers to the processing time from the gtfspy graph output to the curated L- and P-space graphs. The notes contain information about decisions, assumptions and mistakes. For the nine excluded networks, the reasoning for excluding them from the final data set is shortly explained. These networks are indicated in red.

Table F.1: Processing notes for all included networks

Network	**Date**	**Time spent**	**Notes**
Amsterdam	19-04-2022	5min	GTFS data manually split to remove all routes other than Amsterdam metro routes.
Athens	30-06-2017	20min	Map is mixed with tram and commuter rail, both of which are not included in analysis. Line 3 extension towards Piraeus is missing, first part was completed in 2020, full connection not finished as of 2022.
Atlanta	22-04-2022	20min	Notably high waiting times but seems to match up with schedule (about every 20min)
Baltimore	11-07-2020	15min	There seems to be one train in both directions going from John Hopkins to Reistertown and then to Owings Mills directly, without stopping in-between. As there is only vehicle per day, these edges are deleted.

Barcelona	None	Not included	Split into two files because of two different providers, need to be merged. FGC version works, TMB does not. TMB breaks because of stop_times entries without departure arrival time. After deleting all of these entries (as they're irrelevant for metro), it breaks on frequencies as there are apparently missing stops.
Berlin	06-05-2022	30min	Node 21101 (Rathaus Schoneberg) is closed for reconstruction and is thus not part of the scheduling. There are three triangles where the metro skips a station in one direction but not the other. These are: Seestrasse (U6), Yorckstrasse (U7) and Augsburgerstrasse (U3).
Bilbao	20-05-2022	10min	Even though the system officially has two lines, the GTFS data treats them as a single line. The way the two lines are separated is using headsign (meaning there are eight headsigns for one line). Exception is created in the code such that headsign is taken as differentiator. Data does not entirely correspond with actual case/Gmaps, as from Ariz and Basauri at the end of L2, transfers to L1 are not taken correctly. As this hardly affects the total travel times, this is left unfixed.
Boston	15-04-2022	15min	The green line and mattapan line are included in the system as a subway line but are technically a light rail line, as such are not included in the official analysis. Park St. and Downtown Crossing are left unmerged as the green line is not present.

Brussels	19-04-2022	40min	Elisabeth and Simonis are combined into one station, even though the map does not explicitly mention this. Line 5 strangely extends into a portion of line 3 at the end of the day.
Bucharest	None	Not included	For some reason for all lines, the last edge between the second-to-last and last station is missing in only one direction.
Budapest	04-11-2016	30min	Older GTFS data was taken as reconstruction has taken out line 3, thus making the network incomplete. There are three disconnected stations but they're not actually on the map so unsure exactly what those are, they were deleted.
Buenos Aires	31-12-2021	1hr	On line A, Alberti is only passed on westbound trains and Pasco on eastbound trains.
Cairo	01-01-2016	10min	Is from 2014 so does not contain the line 3 extensions since then.
Chicago	26-03-2022	1hr	Jackson Red, Jackson Blue and Library are all merged into one. State/Lake, Lake and Washington are merged. Directed edges seem to match up with directed edges on the official map.
Cleveland	21-03-2022	5min	Only red line is included as green and blue line are light-rail.
Copenhagen	19-04-2022	5min	-
Dubai	31-12-2021	15min	All lines have the same color in the data. The southern branch of the red line is programmed as an extra line and actually operates like that in reality as well. Colors have been manually adjusted to fit the real life colors.
Genoa	12-04-2022	1min	-

Hamburg	None	Not included	Alsterdorf is missing because of construction. Hauptbahnhof Nord and Sud are merged. There are some extra numbered stations of which the function is unclear. The merge algorithm cannot properly accomodate for merging these into existing stations.
Helsinki	14-04-2022	2min	-
Hyderabad	02-01-2022	5min	-
Istanbul	None	Not included	Many implementation differences in GTFS data make it impossible for gtfspy to load this data.
Kobe	01-12-2006	5min	Data is over 16 years old. While the stations and lines are completely correct, the frequencies are most likely lower than currently.
Kochi	29-10-2019	5min	Petta, Alliance JN and SN JN are missing from the data (were not constructed yet).
Lille	04-02-2022	10min	-
Lisbon	01-03-2022	20min	Roma and Areiro are left unmerged as their merge on the map is related to commuter rail. Line directions are programmed as separate lines, red line has an extra line too. Seems to not cause any problems in P-space.
London	02-01-2017	2.5hr	Added shape_id as a possible direction indicator, correctness not 100% guaranteed. There seem to be some express trains on the Metropolitan line, these were deleted.The 10min walk transfers are merged. Battersea branch of Northern line is missing. The right part of the northern circle of the central (red) line seems to be one-directional, this might be a temporary thing. Kensington station is inaccessible and Gmaps also does not record it, so it's deleted.
Los Angeles	19-04-2022	10min	-

F

Lyon	13-04-2022	10min	-
Madrid	20-07-2021	40min	Stations with "long walking transfer distance" according to the official map, have been merged into one station.
Malaga	11-01-2021	5min	-
Marseille	19-04-2022	5min	-
Mexico city	None	Not included	Pipeline could not process the data for unknown reasons.
Milan	07-09-2021	20min	-
Montreal	21-03-2022	15min	Station Outremont is missing as that is under construction at the current moment.
Munich	None	Not included	Data is from March 22, where part of U6 and U3 are out of service. As this makes the network incomplete, network is not included.
Naples	01-04-2022	30min	Rainbow line not included in the data. As that falls outside of the municipal zone of Naples and is a very short line, this is deemed acceptable. Piazza Cavour was merged into Museo. San Giovanni-Barra station is missing from the data.
New York	20-12-2021	3hr	Used stop distance 0 for automatic merge, turned merge recommender phase 1 into automatic (with 75% overlap and 0m), skipped phase 2. All subway and out-of-system transfers are merged. E 180 st and Gun hill road, Jamaica and Parsons (F), 7 Av and Jay (FG), 7 Av and Church (FG), Burnside and 167 (4) were deleted. Some weird one-directional stuff with the brown line, northeast. Has been kept in, seems fine.
Nuremberg	22-03-2022	20min	Colors overwritten manually as they were all programmed with same color.

Oslo	16-06-2017	30min	Gulerasen is skipped in south-bound direction (as indicated by official map). There are some P-space edges with extremely low frequency that do not appear on the map. These seem to be from end-of-day trains that take alternative routes. This messes up P-space calculation slightly, meaning the shortest paths mean is about 3min too long.
Paris	15-04-2022	3hr	Has some lines that run in only one direction. Urban transfers (according to the map) are considered as the same station. Chatelet and Les Halles are left as separate stations as combining them does not add a transfer option. Saint Augustin is merged into Saint Lazare as that adds line 9 as a transfer option to Saint Lazare. Havre Caumartin and Opera are left split as combining those does not lead to extra, logical travel behavior. La Chapelle is merged into Gare du Nord to connect line 2 to Nord. St. Michel and Cluny are left unmerged as the only reason for their liason is the RER line which is not part of the metro network. Similar reasoning for not merge Gare de L'est and Chateau Landon. Austerlitz is merged into Lyon as that is a sensible urban transfer. Unclear why and edge between Porte d'Auteuil and Michel Ange Molitor exists (probably a late night only connection, according to Wikipedia). Anvers and Abesses are left unmerged as their connection includes a funiculaire which is not included.

Philadelphia	07-06-2021	50min	8th street stations are merged. City hall and 15th are merged. System has express lines which are thus automatically added as extra edges in L-space. The broad-spur section on the broad street line is a bit confusing considering Fairmount and Girard. Fairmount stations are already merged, even though the map does not indicate them as such. This seems to be correct according to the official station information. There is an extra L-Space edge between the terminals of the Market-Frankford line. As this edge has a frequency of 1, it is deemed a last night train and deleted.
Prague	19-04-2022	10min	-
Rennes	25-04-2022	5min	-
Rome	19-04-2022	10min	GTFS did not contain colors so backup colors were used.
Rotterdam	19-04-2022	5min	GTFS data manually split to remove all routes other than Rotterdam metro routes.
San Francisco	14-02-2022	30min	Has notably high waiting times, when compared to actual schedule, seems a bit on the high side. While Millbrae station is part of the red line during regular service, it becomes the terminus of the yellow line after 9PM. This messes slightly with the shortest path calculations, meaning some shortest paths are much longer than necessary.
Santiago	19-09-2020	1h30min	Has lines that have different routes where trains stop at alternating stations. The line 5 express routes seem to be missing in some directions, not exactly clear. The other two express lines (2 and 4) work fine, 5 seems to miss an express route in one direction.
São Paulo	None	Not included	The data could not be processed by the pipeline for unknown reasons.

Singapore	None	Not included	The frequencies from the GTFS data do not load properly, hence important data is missing
Stockholm	21-02-2022	30min	For some unclear reason, the ghost station Kymlinge is included in the data, including the trips/routes. The way this data is programmed causes similar problems as for Oslo.
Tehran	None	Not included	The data seems to not be programmed correctly and cannot be properly processed by the pipeline for unknown reasons.
Toronto	28-03-2022	40min	-
Toulouse	15-04-2022	5min	-
Turin	06-07-2021	1min	-
Valencia	19-04-2022	30min	On line 3, between Rafelbunyol and Albalat there are multiple edges between nodes with an unclear pattern, for unknown reason. Balien and Xativa are left unmerged as their merged status on the map is because of the tram. The data is programmed with a huge amount of lines even though there are 5 in actuality. This is likely related to trains stopping at earlier stations and express trains etc. Shortest path calculations seem to be able to accomodate for this.
Vancouver	03-01-2022	15min	Vancouver City Centre is merged into Granville.
Vienna	02-06-2020	30min	Data contains a double schedule, one until 17/03/2020 and one from 16/05/2020. The P-space algorithm is manually adjusted to ensure the former schedule is taken only (as that one has the correct route colors).
Warsaw	19-04-2022	5min	Because of data age, still missing Bernowo and Ulrychow on western part of red line.

Washington	03-01-2022	1hr	Shady Grove and Rockville on the red line are present as stations but unconnected to the rest of the graph, this is because of construction to those stations in the chosen period. There's an extra edge on blue line that skips Arlington cemetery, this ruins P-space calculation so edge is deleted, meaning the frequency is lowered with about a vehicle an hour. This is because this station closes at 7PM and all trains afterwards skip it.

G

ADDITIONAL DATA SOURCES

Table G.1 below lists the sources for the population and GDP data for all included countries/regions.

Region	Population source	GDP source
Argentina	2010 Census (https://web.archive.org/web/20120901061446/http://200.51.91.231/censo2010/)	Government statistics(www.estadistica.ec.gba.gov.ar)
Canada	StatCan (statcan.gc.ca)	StatCan (statcan.gc.ca)
Chile	Central Bank (https://si3.bcentral.cl)	CEIC (https://www.ceicdata.com/en)
Egypt	CitiesABC (https://www.citiesabc.com)	CitiesABC (https://www.citiesabc.com)
EU	EuroStat (https://ec.europa.eu/eurostat)	EuroStat (https://ec.europa.eu/eurostat)
India	2011 Census (https://censusindia.gov.in/	Government data (https://community.data.gov.in/)
Japan	Japan External Trade Organization (JETRO) (https://www.jetro.go.jp)	JETRO (https://www.jetro.go.jp)
UAE	Government statistics (https://www.dsc.gov.ae/)	Government statistics (https://www.dsc.gov.ae/)
UK	EuroStat (https://ec.europa.eu/eurostat)	Statista (www.statista.com)
USA	2020 Census (censusreporter.org)	US Bureau of Economic Analysis (https://www.bea.gov/)

Table G.1: The data sources for the population and GDP of the included networks