Quantifying hierarchy in public transport networks: Developing a new metric

Case-study of the Amsterdam and Rotterdam public transport networks

A.E. Buijtenweg

TU Delft

Royal HaskoningDHV
Enhancing Society Together

Quantifying hierarchy in public transport networks: Developing a new metric

Case-study of the Amsterdam and Rotterdam public transport networks

by

A.E. Buijtenweg

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Monday March 20, 2020 at 9:30.

Student number:	4703219	
Project duration:	September 1, 2019 – March 20, 2020	
Thesis committee:	Dr. O. Cats,	Chairman TU Delft
	Dr. ir. T. Verma,	Daily supervisor TU Delft
	Dr. H. Wang,	External supervisor TU Delft
	Ir. B. Donners,	Daily supervisor Royal HaskoningDHV

In collaboration with:

TU Delft Delft University of Technology

Royal HaskoningDHV Enhancing Society Together

Cover picture retrieved from: `http://www.urbantransportgroup.org/about-us/who-does-what-transport` on 15-01-2020

Preface

This thesis is the final product of the graduation product, finalizing the MSc. program Transport, Infrastructure & Logistics (TIL). This report is the final step of my study at the TU Delft and concludes my stage of life as a student. Throughout this research, the possibilities to quantify hierarchy in public transport networks are explored. The research is carried out for Royal HaskoningDHV to whom I am grateful for this opportunity of carrying out this study.

Therefore, I would first of all like to thank Royal HaskoningDHV for making this research possible, providing me data and supervising me where required. In particular I would like to thank Barth Donners for supervising my thesis process, discussing complex issues and providing feedback and input for the thesis in general. Furthermore, I would like to thank Marcel Scholten for his help with the demand-data analysis and his efforts to provide me the input for the OD matrices.

I would also like to thank my TU Delft committee for their guidance throughout the past months and their helpful feedback during the meetings. First, I would like to thank Trivik Verma for his help with getting into programming with Python, discussing topological indicators and basically all of his expertise on the subject. In particular during our meetings Trivik helped me with most of the struggles I was having leading to a smoother thesis process in general. Next, I would like to thank Huijuan Wang for her guidance as external supervisor with a refreshing and different perspectives on the matter. I think the meetings we had were very inspirational and gave me a different idea of how to approach and explain the methodology. Lastly I would like to thank Oded Cats for his enthusiasm and involvement as chairman of my committee. With the help of his accurate feedback and broad knowledge I was able to structure my thesis and include relevant elements for the research.

Last but not least I would like to thank my friends and family who have supported me through the process (and way before). In special I would like to thank my parents, brother and girlfriend who have been supporting me and helping me, despite the complex matters, during the thesis. They have shown interest in my work and tried to provide me with feedback whenever possible. Furthermore, I would like to thank any friends, family and colleagues to whom I could complain to at any given moment I was having struggles, to get it off my chest.

Abel Buijtenweg
Amersfoort, March 2020

Contents

List of Figures

List of Tables

1 Introduction

1.1. Problem definition

In a world where total passenger and freight transport is increasing every year (Eurostat, 2019; OECD, 2019), the pressure on transport networks increases correspondingly. Consequently, networks have to adapt to a higher demand by increasing capacity or adding new connections. In order to comprehend how a network should be adjusted, an analysis of the network is required. As the transport networks consist of many connected links, an analysis of the whole network is considered very complex. It is therefore recommended to subdivide a transport network into distinct levels where each level explains a different function of the network. The analysis of these transport networks can therefore be done using separated levels (or layers) to overcome the complexities of the transport system (Gallotti & Barthelemy, 2015) and acknowledge the multilayer structure of a transportation network (Aleta, Meloni, & Moreno, 2017). By dividing the network into separate layers, dynamics within and between layers can be explained. It therefore appears to be important to distinguish between functions and levels within the network to grasp the complexities of transport networks.

For the future development of the transport system of larger cities, public transport (PT) plays an essential part to accommodate passengers (Richards, 2012). The car use in many cities has been attempted to decrease in favor of walking, cycling and PT. However, the organization of a substantial public transport network (PTN) is considered very complex due to the different modalities, infrastructure and passenger flows. To provide an example of the enormous flows, the Paris PTN serves over eight million passengers each day (EMTA, 2019) while the London network serves over seven million passengers on a daily basis (Transport for London, 2015). Hence, the planning and development of the PTN for a sizable city requires well-advised decisions and good understanding of the network dynamics.

The representation of a PTN of an entire city can be done using a multilevel network (Gomez et al., 2013; Min, Do Yi, Lee, & Goh, 2014; Aleta et al., 2017), in which categorical levels are distinguished to represent different functions in the network. One approach to define levels in a PTN is based on the functions of stops, which are referred to as nodes throughout this study. In order to connect the different levels, transfer nodes are used as locations for level interaction. The functioning of these transfer nodes between the different levels in a transport network is becoming an essential aspect of a transport network to facilitate rapid transport throughout the network. The transfer locations are indicated as (transit) hubs, where multiple levels of transport networks are coming together and interaction between the different network levels takes place. As a PTN only requires a limited number of hubs, just a fraction of the nodes functions as a transit hub while the majority functions as a minor hub (e.g. a node located at a location where two lines intersect) or as an entry and exit point only. It is therefore meaningful to understand the different nodes in a network have a different function in which some are considered more important than others. In other words, it is meaningful to distinguish between nodes based on their relative function in the network. After valuing nodes based on their relative function, the nodes can be allocated to a categorical level based on the importance of their function in the network. If for these categorical levels some ordering is applied, leading to a(n) (exponentially) decreasing number of elements for subsequent levels, it may be considered as a hierarchical structure. Following, a hierarchical order in the network can be identified where nodes are allocated in a hierarchical level based on their importance in the network. It is however, hard to quantify the function of a node in a network as an 'important function' can be interpreted in numerous ways. In other words, there is no generally accepted method of determining a hierarchical order in a PTN. Therefore, the assessment of the hierarchical structure in a PTN remains limited.

The transit hubs as well as the hierarchical structure of a network can be connected to the phenomena of bottlenecks in PTN. Intuitively, bottlenecks occur where large flows come together which relates closely to the function of transit hubs in hierarchical PTN. Therefore, bottlenecks can therefore be seen as an external effect of a certain hierarchical structure[1]. In order to increase the performance a network, it is important to understand where the bottlenecks are located (Schmöcker, Bell, & Lam, 2004). For the identification of bottlenecks in public transport networks some research has been done on an operational level (Van Oort, Sparing, Brands, & Goverde, 2015). However, on a more tactical or strategic level, considering the development of a network, the research on the occurrence of bottlenecks in PTN remains underexposed. Furthermore, studies related to the identification of bottlenecks are thus far predominantly related to a dichotomy in network supply and travel demand at a specific location (e.g., Janson, 1995; L. Zhang, 2007) but it neglects the conjunction of different levels of flows which can be understood by analyzing the multilevel structure. Hence, the connection to hierarchy and more specifically to the location of bottlenecks in the hierarchical structure, could provide useful insights in understanding why bottlenecks occur at a certain location.

1.1.1 Hierarchy as a network indicator

While many network indicators have been identified, a quantitative indicator to specifically determine the level of hierarchy in a PTN is in-existent (Gattuso & Miriello, 2005; J. Zhang, Zhao, Liu, & Xu, 2013; Min et al., 2014). Moreover, when comparing different networks across cities (e.g. D. Levinson (2012); Roth, Kang, Batty, and Barthelemy (2012)), the hierarchical structure of a network is left out of the comparison. While the application of indicators for hierarchy in other fields of science is present, in PTN there is no generally accepted indicator.

Connected to this, the influence of transfers has been incorporated in topological indicators but this predominantly relates to the requirement for a transfer in the shortest path rather than the function of transfer locations (Derrible & Kennedy, 2011). Hence, a measure to evaluate the hierarchical structure in a network incorporating the function of transfer locations is generally lacking. Furthermore, a good measure for hierarchy could also be a starting point for a generally accepted definition of hierarchy (Mones, Vicsek, & Vicsek, 2012).

Connecting the possibilities of a hierarchical structure indicator to phenomena in PTN, the vulnerability of a network and cascading failures (Little, 2002) of hierarchical structures seems to be an interesting link. The analysis of the vulnerability of a public transport network can be done by looking into the impact of disruptions on the network (Rodríguez-Núñez & García-Palomares, 2014). Hence, disruptions on certain places of the network cause more problems for the network

[1] E.g. only having one transit hub in a large city inevitably leads to this hub becoming a bottleneck

performance than disruptions on other places. The vulnerability of a network intuitively relates to hierarchy, as disruptions on more hierarchical locations in the network affects less hierarchical locations in another way then the other way round. The relation between the concepts of hierarchy and vulnerability could lead to interesting new insights.

1.1.2 Societal relevance

Next to the value of defining hierarchy in a quantitative way for science, a societal value can be identified too which mainly relates to the practical application of a hierarchy quantification. First, this study attempts to provide a better understanding of the role of nodes in the network and the functioning of the network as a whole. Consequently, the analysis of PTN can be enhanced in terms of evaluating the functioning of specific nodes. By understanding the function of these nodes, adaptations could be made to develop a network towards a desired structure. Secondly, more knowledge about bottlenecks could lead to a better understanding on how to prevent or solve bottlenecks in the future. As bottlenecks are an important cause for delays, this could be beneficial for both the punctuality of the PTN and the appreciation of PT in general. Thirdly, by analyzing the hierarchical structure, weaknesses of short-comings of the network are identified which can lead to improvements in the network for a robust network in the future. Fourthly, as this study uses real city networks as cases to apply the created models to, an analysis for these cities can provide useful insights for the municipalities and PT operators of these cities. It might be particularly useful to analyze how the removal or addition of existing or new lines, by the means of scenarios, affects the hierarchical structure to provide a substantiation of how a network should be adapted in the future. By providing a quantification of the change in hierarchical structure as a result of the addition of a line, arguments for or against the addition of this line may be enhanced.

As this study is executed for the company Royal HaskoningDHV, their interests should be taken into account. Since the company works with PT operators and municipalities as clients, the scientific insights in a method to quantify the hierarchy in a network could be a valuable tool. Furthermore, bottlenecks are an important topic for the company's clients and additional insights in this could help for bottleneck related projects in the future. Throughout the next section, the research objectives of this study are stated.

1.2. Research objective

Throughout this study, the goal is to explore the possibility to quantify the hierarchy in PTN while taking different characteristics of networks into account. A coupling between a topological and an empirical approach to analyze the network is applied to get a comprehensive understanding of the hierarchical structure within the network. This includes a topological oriented approach for the different functions of nodes in a network while applying an empirical approach, based on existing demand data, to understand travel behavior in the network. Combining these two approaches with a definition of hierarchy in PTN, a metric is developed to determine the hierarchy in a network. Furthermore, by understanding the hierarchical structure of a network, improvements for the network are sought. These possible improvements for the network can be reflected in policy directions to develop or adjust networks in the future.

Following, a multilevel network structure, based on the hierarchical value of the nodes can be derived. In this network structure, the spread of the hierarchically important nodes is visualized in a map. Moreover, a network based hierarchy value can be obtained for the entire PTN and, therefore, be used to compare to other networks. This is the end-goal of this thesis despite the subject being a gateway to more related content. However, due to the limited time-frame for the research, further analysis is beyond the objective of this study. In order to clarify the focus of this research, a research question and sub-questions are elaborated throughout the next section.

1.3. Research question and sub-questions

Following from the research objective, a research question is formulated:

How can hierarchy in a public transport network be measured quantitatively and can the hierarchical structure be used to improve the network structure?

In order to provide an answer to the research question, it is divided into the following sub-questions:

1. **How can the hierarchical structure and hierarchy of nodes in a PTN be defined?**

The goal of this sub-question is to provide a clarification for the definition of hierarchy in PTN within the context of this study. As hierarchy, even in the context of PTN, can be interpreted in numerous ways, it is perceived necessary to determine the definition for this study first. Furthermore, by defining elements for which the definition of hierarchy in PTN should account, the requirements for a measure can be outlined. In order to clearly distinguish between the hierarchy within a network and the hierarchy of a network as a whole, these are discussed separately.

2. **How can topological and empirical network analysis be applied to define a measure for hierarchy in PTN?**

This sub-question is seen as complementary to the previous sub-question as this is related to the development of a measure based on the definition of hierarchy for a PTN. The combination of a topological and empirical approach is chosen to emphasize the strengths of either approach in the metric. The topological approach focuses on the nodes within the network from a graph theory perspective. Hence, the network is analyzed without accounting for flows and based on the network structure and line characteristics only. The empirical approach accounts for the passenger flow and the behavior of passengers throughout the network. To answer this sub-question, a shortest-path algorithm to calculate the flows throughout the network would be required based on the demand.

3. **How can a hierarchy metric be applied to analyze the PTN of Amsterdam and Rotterdam?**

The aim of this sub-question is to apply the metrics for node hierarchy and network hierarchy to real networks. While analyzing these cities, the metrics are validated by testing if the hierarchies match expectations and if this is not the case, identify why the outcomes may differ. Furthermore, by analyzing the hierarchical structure of two Dutch cities, a comparison between the cities is done to understand the implication of the structures.

4. **How can bottlenecks and vulnerability in PTN be explained by looking at the hierarchical structure?**

This sub-question attempts to connect the theoretical hierarchical structure to the more practical issues of bottlenecks and vulnerability. By understanding underlying network design characteristics leading to bottlenecks, a practical application of the hierarchical structure analysis is identified. In order to determine which nodes are bottlenecks from a hierarchical point of view, individual analysis on a node-scale is required. For vulnerability, disruptions are tried to be connected to the hierarchical structure by linking the severity of a disruption to the hierarchy of the node. The linkage to hierarchy for either of these two concepts can be used as a substantiation for shaping policy regarding possible adaptations to the network.

5. **How can the impact of PTN scenarios be assessed using the metrics for hierarchical degree and hierarchical structure?**

This final sub-question can be seen as a way to practically apply the metrics to provide a deeper understanding of the impact of changes to the structure of a PTN. Furthermore, by analyzing the changes in hierarchy caused by the implementation of a scenario, a clearer understanding of the benefits and weaknesses of a scenario for the network can be substantiated. This quantifiable assessment of scenarios might be applied to shape the policy for the development of the network. This policy would predominantly be related to the further development and expansion of existing PTN.

1.4. Scope of the study

In order to clearly define what is in the scope of this study and what is beyond it, an overview is perceived necessary. As there are two different approaches taken to analyze the networks, both have a different scope and are elaborated separately. However, first the scope for the networks to be studied has to be set to define which networks qualify for this study.

1.4.1 PTN scope

The networks which are perceived suitable for this study have to meet a number of qualifications regarding network size and data availability. First, the network should have different types of PT links in which different levels levels can be categorized. This is necessary in order to clearly distinguish a hierarchical structure in the network. Furthermore, the network should have a decent size which implies the city in which the network is located should be considered a metropolitan area. The city networks complying with a decent size and available data are, Amsterdam and Rotterdam. As both of these cities are in the Netherlands, they share some similarities in terms of national railway operators but the intra-city PT operator and network structure are significantly different. When drawing conclusions, the different type of data by different PT operators should always be taken into account. For the Amsterdam network specifically, different states of the network are compared to understand how changes to the network structure affect the hierarchy. With the recent addition of the NZL (North-South Line), the hierarchical structure of the PTN may have changed and analyzing this may lead to additional insights. Furthermore, future plans for further developing the network are analyzed to assess how these changes affect the network.

1.4.2 Topological scope

For the topological network analysis, a scope has to be set in order to clearly define how the topological part of the metric for hierarchy is derived. The definition of hierarchy in the context of PTN inevitably leads to a delimitation of relevant topological information. Furthermore, the context of PTN may lead to constraints using topological network analysis as topology is applied to numerous types of networks in different fields of science with different types of nodes and links. Lastly, availability of data can potentially lead to constraints regarding topological analysis due to some types of analysis requiring data which are unavailable for this study or for PTN in general.

1.4.3 Empirical scope

For the empirical model, a demand based passenger flow model is used for this study. For this passenger flow model, some simplifications have to be made to prevent the model becoming too complex without providing much additional value. As this study is mainly an exploration of the possibilities to quantify hierarchy, these simplifications of the model are considered to be acceptable. First, the data used for the empirical model are daily data in which no time of day or morning/evening peaks are considered. Secondly, related to the former, no vehicle congestion is modeled as it would require time of day data especially for peak hours. Thirdly, only PT modes are taken into account and other modes such as cars but also walking and cycling for access and egress are unaccounted for. Consequently, the empirical flow modeling is based on trips between different stations/stops in the PTN and these are considered the beginning and end of the trip. Fourthly, the route choice in the passenger flow model is based on travel time, waiting time and transfer(s) (time). Fare costs and vehicle crowding are beyond the scope of this study.

1.4.4 Structure of the report

The remainder of this report begins with a literature review in the next chapter in which relevant theories and studies are evaluated. Thereafter, the methods for measuring hierarchy in PTN are explained in the methodology. In this chapter, first a way to define hierarchy is presented. Secondly, the metric to determine the hierarchy in a network is outlined and lastly a conceptual overview of input and output is shown to provide an overview of the process. In the fourth chapter the application of the metric to the networks is described in which the implementation, cases and data are outlined. In the fifth chapter, the results for the cases are presented. In the last chapter, the highlights of the results are evaluated in the discussion followed by the conclusion and recommendations. Additionally, in Appendix D a scientific paper based on this report is presented. However, as mentioned above, first the existing literature is reviewed to clearly evaluate how this research relates to previous works and how these insights are applied for this study.

2

Literature review

In order to briefly summarize how this study relates to previous works regarding hierarchy indicators, measures and metrics, an overview of the scientific context is presented in table 2.1 as starting point. In this table, other approaches to quantify hierarchy are presented and their respective science fields are indicated. Throughout the remainder of this chapter, an elaborate overview of the development of the metric for this study is provided.

2.1. Hierarchy in PTN

In this literature review, the topological as well as the empirical approach to analyze PTN, are presented and related to hierarchical structures. However, first a clearer understanding of different approaches to hierarchy in the context of PTN is required to provide an overview of possible interpretations of hierarchy. Besides the approach to hierarchy, the way a hierarchical structure is defined is briefly clarified. Furthermore, different approaches to distinguish levels in a hierarchical structure are elaborated.

2.1.1 Defining hierarchy

Hierarchy is a concept that can be interpreted in numerous ways. Mones et al. (2012) mention there are various definitions to hierarchy and there is no unique, widely accepted definition. Hence, it is important to clearly define how hierarchy is perceived for this study to prevent any misinterpretations. However, explaining other definitions of hierarchy and stating why the context of that definition does not fit this study is at least as important. Hierarchy in its essence is defined as “a generic structure, in which levels are asymmetrically ranked according to a specific type of relation” (J. Luo, Whitney, Baldwin, & Magee, 2009, p. 8). This broad definition implies there are different levels for which a ranking among the levels is determined.

Table 2.1: Approaches to quantification of hierarchy in scientific context

Paper	Perspective on hierarchy	Indicator/Measure/Metric	Science field
Bassolas et al. (2019)	Classification of cities based on dynamical hierarchy in the interaction beween hotspots	Flow-hierarchy (ϕ)	Urban mobility
Trusina et al. (2004)	Classification of any type of hierarchy based on the concept of hierarchical paths	Hierarchical fraction ($\mathscr{F}$)	Physics
Mones et al. (2012)	The impact of a node on the network proportional to the number of nodes reachable from them	Local reaching centrality ($C_R(i)$)	Biological physics
Ru and Xu (2005)	Measuring the hierarchical structure based on dissartotativity and clustering distributions	Hierarchical disassortativity	Aviation
Ravasz et al. (2002)	Measuring small topological modules that combine in hierarchical larger units	Hierarchical modularity	Physics/ Complex networks
Vázquez et al. (2002)	Conditional probability that a link connects two nodes with a certain connectivity	Conditional probability ($p_c(k'k)$	Physics/ Internet science
Corominas-Murtra et al. (2011)	The presence of a pyramidal structure and topological reversibility	Hierarchical index ($f(\mathscr{G})$)	Non-linear science
Eum et al. (2008)	Locate top layer nodes then locate rest of nodes based on distance from top nodes	Hierarchy measure (Q_H)	Information science

The archetypal structure of a hierarchy is explained as a tree (Bassolas et al., 2019) implying few number elements in high levels increasing in number every level downwards. Moreover, the different levels in a hierarchical structure supposedly follow a power-law distribution, increasing exponentially in number in every level below (Barabási, Ravasz, & Vicsek, 2001; Sienkiewicz & Hołyst, 2005). Van Nes (2002) notes a hierarchical transport network is described as a transport network in which functionally different levels are identified. Moreover, each level is suited for a different type of trip length with different demand and speed while also providing access to higher or lower levels. The hierarchy in this sense is created by the flow and speed of the edges in the network (Yerra & Levinson, 2005). Relating this to PTN, defining the hierarchy is a natural way for acknowledging the multilayer structure of a PTN (Van Nes, 2002). To understand the different approaches to define hierarchy in a PTN, link-, line- and node-based hierarchy are distinguished and elaborated separately[1].

Link-based hierarchy

As the first approach to hierarchy relates to the function of different links in the network, this is referred to as link-based hierarchy. Yerra and Levinson (2005) apply link-based hierarchy by defining hierarchy as classified roads in terms of the flow of traffic they carry. In their study, they find results which show hierarchies and roads show emergent behavior, independent from the initial base map's features. This is, among other things, caused by the travel demand from the edges of a bounded network through the center. Furthermore, they note, based on Barabási and Albert's (1999) concept of preferential attachment, that the 'rich get richer' which leads to the formation of hierarchical ordering in the network over time. Lee, Barbosa, Youn, Holme, and Ghoshal (2017) mention a hierarchical ordering based on the function of a link in the road network where the highest level of hierarchy is usually found towards the center of the network. This perception of hierarchy where the function of the link relates to the hierarchy, could be useful for PTN.

Link-based hierarchy is perceived as a good way to determine hierarchical structures in a road network. However, for a PTN, the nodes have a totally different function compared to road networks. For the former, nodes relate to transfer stations where you may have to get off a vehicle, walk to another platform and wait for a different vehicle for a transfer. For the latter, a node is seen as a junction where changing links can be seen as turning at an intersection or traffic light. It is due to the underemphasized function of nodes that hierarchy based on links only, is insufficient to capture the complexity of hierarchical structures in PTN.

Line-based hierarchy

As opposed to link-based hierarchy, a line-based hierarchy approach is able to capture the line bounded nature of PTN by defining direct routes and routes requiring transfers. Jian, Peng, Chengxiang, and Hui (2012) note that each transit lines' function is determined to establish a clear hierarchical structure to study the travel behavior of passengers. Based on the findings of Bagloee and Ceder (2011), the lines are divided into three different types of functions: local route, feeder route and mass route. They explain this distinction by defining local routes as these that traverse through side areas to connect those to the PTN, feeder routes to connect local routes to mass routes and mass routes being the high volume 'skeleton' of the PTN. The different functioning of lines acknowledge the multilevel structure of PTN in general (Van Nes, 2002). It is rather clear that the different functional levels can be connected to the mode type of a line but this presumption is not directly incorporated for this study to avoid modality based assumptions.

One potential drawback of line-based hierarchy would be that the function of a line may not be

[1]This distinction is chosen based on graph theory which is elaborated further on in 2.2.1

evenly distributed over the line. For example, a line going from one edge of a network to another edge traversing through the city center is likely to have a different function in the city center than near the edges of a network. To overcome this drawback, Jian et al. (2012) suggest that one line can cover multiple hierarchies on different sections of the line. Furthermore, there may be multiple lines between two nodes with different features in terms of intermediate stops, infrastructure, frequencies and speed, which implies there may be different lines between nodes with a different hierarchy. It is important to notice the hierarchy of this link between the nodes might be interpreted differently than the lines traversing this link. While line-based hierarchy does capture the transfer based nature of PTN, it is still unable to capture the function of transfer hubs.

Node-based hierarchy

From a more organizational point of view, a hierarchical network is explained as having a scale-free topology and a high degree of clustering (Ravasz & Barabási, 2003). This relates to the hierarchical organization of the nodes in a network. Ahn, Bagrow, and Lehmann (2010) note many networks are known to possess hierarchical organization leading to the grouping of nodes in hierarchical structures. Consequently, the organizational point of view is interpreted as hierarchy in the perspective of nodes or node-based hierarchy. As the name indicates, node-based hierarchy takes the different function of the nodes in the network into account. The nodes in the network are hierarchical in a sense that they have different histories, roles and significance (Mones et al., 2012). In its essence, the different roles can be compared to the function types mentioned by Bagloee and Ceder (2011): local, feeder and mass, where the function mainly depends on the possibility to transfer from one level to another and the function of nodes it is connected to. Ravasz and Barabási (2003) note the presence of a hierarchical structure in real networks can be identified when there are many links with just a few connections and few nodes with many links, following a power-law distribution. De Montis, Barthélemy, Chessa, and Vespignani (2007) note that hubs function as hierarchical nodes by providing connectivity for lower-degree nodes.

A typical network structure in which a node hierarchy is identified easily is the hub-and-spoke network where there are many nodes only connecting to one node, the hub. In this structure, the less hierarchical nodes are connected to the hub (being the spokes) and the hub connects to other hubs. The hub-and-spoke network structure is predominantly observed in aviation networks (S. Zhang, Derudder, & Witlox, 2013) where local nodes generally connect to a feeder or mass node. However, aviation networks are clearly different from PTN since airlines generally connect two airports while PT lines connect a sequence of nodes. Therefore, the hub-and-spoke structure is not seen frequently in PTN. Hence, the definition of hierarchical nodes in a PTN should be different than for aviation networks.

Mones et al. (2012) embrace the node based hierarchy and mention that quantifying the extend to which a node is hierarchical is a crucial step to understand the functionality and controlling of networks. In their view, hierarchical nodes are defined as the nodes with the lowest average distance to every other node. While this does indicate the hierarchy in a sense of a central location, it does not capture the function of the node in the network nor the function of the other nodes. Hence, this definition may not be directly applicable to PTN as it neglects the function, role and significance of the nodes and perceives every node as equal.

In their study, De Montis et al. (2007) come to the conclusion that a hierarchy of nodes[2] exists in which small nodes are highly interconnected locally while larger nodes provide long-distance connections. Despite the context being different for this study, a similar perception of hierarchy in the context of this study can be applied. This implies that smaller nodes are interconnected locally

[2]Which are municipalities on the island of Sardinia in their case

while larger nodes, or hubs, provide longer distance connections. The interconnected local nodes could be seen as a PT line traversing a sequence of smaller nodes in a PTN.

A drawback of node-based hierarchy is that, while it accounts for the nodes' as well as neighboring nodes' properties, it neglects the characteristics of the links connecting these nodes. Therefore, a high demand, high capacity PT connection to a hub is perceived equivalent to a low demand low capacity PT line to the same hub.

2.1.2 Hierarchical structure

Connecting to the previously introduced hierarchy of segments in the network, the hierarchy of the network itself is given by analyzing the hierarchical structure. A definition for the hierarchical structure is explained by "different network parts having different functional levels suited for facilitating specific trip speeds and trip lengths" (Fiorenzo-Catalano, 2007, p. 142). Furthermore, hierarchy can be described as part of the network structure in which a good hierarchical structure leads to higher efficiency (Scheurer & Porta, 2006). As a hierarchical structure only implies having different functional levels, the structure between two hierarchical networks can vary significantly. However, while different hierarchical structures can be compared, there is a lack of methods to determine which structure is more hierarchical. Mones et al. (2012) do propose a measure, the global reaching centrality (GRC), to determine the hierarchical properties of a network. This measure is, however, based on network with a directed graph, to which PTN generally do not belong. Therefore, the interpretation of the hierarchy of a PTN is limited for now due to the lacking of a measure to determine the hierarchical value of a undirected graph.

2.1.3 Hierarchical levels in PTN

As stated earlier, defining hierarchy in a network leads to different hierarchical levels in a network. In order to get a good understanding of how these levels are defined for PTN, it is important to distinguish two approaches on which these levels are based. On the one hand, the different levels are based on lines and links, where higher levels relate to lines or links higher in the hierarchical order and nodes can be present in multiple layers. On the other hand, levels are defined based on the hierarchical value of nodes, where higher levels contain nodes that are more hierarchical and links can either connect nodes within the same level or between different levels. This dichotomy of approaches can be linked to the similar distinction in defining hierarchy in general. Comparing the different levels of hierarchical networks from a link or line approach, lower-levels have lower network speeds, higher network densities and higher access densities (Van Nes, 2002). From a node based hierarchy approach, Mones et al. (2012) note that a network structured in a flow hierarchy, the nodes are layered on different levels to emphasize which higher level nodes influence lower level nodes. Hereafter, the two different approaches to hierarchical levels are explained separately but first the types multilayer networks are briefly outlined.

Multilayer network types

The representation of complex networks is often done using multilayer networks. Kivelä et al. (2014) have done extensive research in defining any type of multilayer network and in what context with which features it can be applied. Based on their findings, two types of multilayer networks appear to be particularly useful for the context of PTN in this study, multilevel and multiplex networks. Multilevel network are characterized by being layer-disjoint, which implies nodes can only be present in one layer. Links can connect nodes in the same layer as well as in different layers. This multilayer type of network representation appears to be especially useful to distinguish between nodes in which every level presents a hierarchical function or threshold value. Furthermore, multilevel networks are used to shape the interaction across layers (Zappa & Lomi, 2016) which in the case of PTN could be explained as links between different nodes with a different degree of hierarchy. Kivelä et al. (2014) define multiplex networks as a representation in which each link is categorized by its type. In this definition, nodes, and even links between two nodes, can be present in multiple levels but in every layer the link represents a different type of link. In this definition of a multiplex network, nodes are not node-aligned, which implies nodes may not be present in every layer and also not layer-disjoint which means nodes are present in only one layer. Consequently, nodes might be present in multiple layers but do not have to be present in every layer.

Link-and line-based hierarchical levels

For link- and line-based levels, higher-levels are seen as high speed and low density levels. Relating this to modalities in PTN, lower-levels would relate to local services such as buses or trams, while higher-levels could be (inter)national trains or metros in a metro in a metropolitan area. Hence, the service area of the network is an indication of the level the network belongs to. An overview of an example of different modalities in different levels is given in figure 2.2.

Table 2.2: Hierarchical levels in (inter)urban networks (Van Nes, 2002)

Travel distance	Urban			Interurban		
	Network	**Stop spacing** [km]	**Freq.** [veh/h]	**Network**	**Stop spacing**	**Freq.** [veh/h]
0 - 8 km	Bus/tram	0.4	4 - 8	Local bus	1 – 2 (0.4)	1 - 2
3 - 10 km	Metro/ light rail	0.6 - 0.8	6 - 12			
5 - 30 km	Local train/ light rail	3.0- 5.0	4 - 8	Local train/ light rail/ Interliner	3 - 5 (1 – 2)	2 - 4
30 - 100 km				Interregional/ Intercity	20 – 30	2
30 - 300 km				Intercity (Randstad only)	30 – 40	2

In partial contrast to the figure above, levels can also be based on the function of the link in the network. Recalling from defining line-based hierarchy, lines but also links, can have the function of a local, feeder or mass route (Bagloee & Ceder, 2011). Defining levels based on this distinction could lead to a somewhat different outcome where the function of a link is leading instead of the modality or the stop spacing. Implicitly, this would still be included as e.g. intercity trains have in general higher capacity and travel speed and would therefore qualify as mass transport and be in the highest level.

Based on the definition of Kivelä et al. (2014) for multilayer networks, a multiplex network would be most suitable for link- and line-based hierarchical levels. Multiplex networks are particularly useful to define levels based on the function of the links in the network. In the context of hierarchy in PTN, hierarchical nodes would be nodes that occur in different levels of the multiplex network to facilitate transfers between different functional levels. For both multiplex and multilevel networks, inter-layer edges (Gallotti & Barthelemy, 2015) are used to shift from one layer to another.

Node-based hierarchical levels

Opposed to link-and line-based hierarchical levels, for node-based hierarchical levels, the nodes are ordered in levels. For this type of ordering, nodes are allocated to a level based on their hierarchy value where nodes with a comparable hierarchical degree belong to the same level. The links between the nodes indicate that the link connects nodes with a comparable hierarchical degree for links within a level or with different hierarchical degrees across levels.

For node-based hierarchical levels, a multilevel network representation would be the obvious choice to order the nodes based on the hierarchical degree while links across levels are possible to connect nodes across levels. Furthermore, in the context of Bagloee and Ceder's (2011) definition of functions of links, local routes would be links within a lower level, feeder routes between a lower level and a higher level and mass routes within higher levels. In order to facilitate a good understanding of how hierarchy is defined and represented, the topological network analysis is outlined throughout the next section.

2.2. Topological network analysis

Regarding network topologies in PTN, metrics based on graph theory are applied for network analysis (Cats, 2017). Emphasizing this, most efforts by researchers on studying the topological properties of complex networks, such as PTN, are based on graph theory (Reggiani, Nijkamp, & Lanzi, 2015). Therefore, a more elaborated overview of how a network can be analyzed topologically could lead to insights regarding analyzing the hierarchy of nodes and the hierarchical structure. Furthermore, by distinguishing and analyzing network properties, different PTN are compared (Von Ferber, Holovatch, Holovatch, & Palchykov, 2009). The comparison of topological networks could provide a method to understand the differences in terms of hierarchy between different networks. Moreover, changes in topology are widely used to examine the development of a PTN (Cats, 2017). In other words, changes in topology could be used to understand the effects of a change in the network structure such as e.g. the addition of the NZL to the PTN of Amsterdam.

A limitation of topological metrics that has to be taken into account is that these metrics are usually focused on the structure of a transportation network while it neglects the dynamic features (Y. Zhou, Wang, & Yang, 2019). Hence, it provides an overview of the network without flow going through the network.

2.2.1 Graph theory

To provide a little background for graph theory, some generic aspects are briefly elaborated. Graph theory is applied in many fields of science while it was originally confined to the field of mathematics (Derrible & Kennedy, 2011). A graph of a network is defined as $G = (V, E)$, where a finite set of nodes is defined $V = \{i, j, ..., k\}$ while a finite set of links is defined $E = \{(i, j) : i, j \in V, i \neq j\}$ (Xia-Miao, Ming-Hua, Jin, & Ke-Zan, 2010). Furthermore, N generally indicates the number of nodes in the network (Berche, Von Ferber, Holovatch, & Holovatch, 2009).

In order to define whether nodes are connected by a link, a node is called adjacent to another node if a link between the nodes exists (Nystuen & Dacey, 1961). Moreover, an adjacency matrix $\mathbf{A} = (a_{ij}) \in \{0,1\}^{N*N}$ with a_{ij} is one for $(i, j) \in E$ and zero otherwise (Xia-Miao et al., 2010) can be drafted to determine which nodes are linked directly. Within the field of PT, many studies have preceded in the use graph theory (e.g. Gattuso and Miriello (2005); Derrible and Kennedy (2009); Von Ferber, Holovatch, and Holovatch (2009); Mishra, Welch, and Jha (2012) to mention a few early prominent works).

2.2.2 Topological model representation

Following from graph theory, a variety of network representations are distinguished which all have different properties (Von Ferber, Holovatch, Holovatch, & Palchykov, 2007). For the different indicators, which are introduced in the subsequent section, using different network representations leads to diverging interpretations of the values for the indicators. Therefore, it is necessary to provide an overview of the different network representations to clearly indicate what is referred to. Based on which aspect of the network is the focus of study, being nodes, edges or lines, different representations are favorable for a specific study. For hierarchical networks, different approaches regarding the network representation can be taken. For example, using the 𝕃-space, the node degree (see 2.2.3) of a station is used to analyze how many different lines of public transport service a station (Von Ferber, Holovatch, Holovatch, & Palchykov, 2009). An overview of the relevant spaces is presented in figure 2.1.

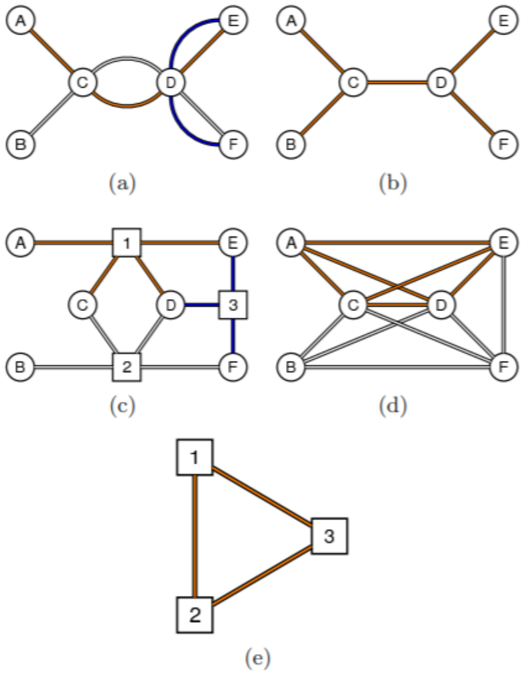


Figure 2.1: Representation of an example of space representations where (a) indicates the existing nodes and lines, (b) the 𝕃-space, (c) the 𝔹-space, (d) the ℙ-space and (e) the ℂ-space (Von Ferber, Holovatch, Holovatch, & Palchykov, 2009)

For this study in particular, a combination of the 𝕃- and ℙ-space can be applied. On the one hand, the 𝕃-space is used to determine the location, neighbors and directions of the links for a node to understand the position in the network. Based on the 𝕃-space, the function of a node in the network is derived. For example, hub nodes are easily spotted in the 𝕃-space by having many links attached. On the other hand, the ℙ-space is used to identify which node pairs are directly connected. This is particularly useful to clarify the number of transfers in a shortest path. Furthermore, by using the ℙ-space, nodes that are passed through along the route but do not require any transfer, are identified. This is useful to determine transit passengers for a node to which passengers passing through on one line do not belong. Hence, the ℙ-space can be used to determine if a node located on the shortest path of a node pair functions as a transfer station or not. Furthermore, as direct connections in PTN are considerably more informative than direct neighbors (which may just be a passing through node), using the ℙ-space generally provides more information in general.

2.2.3 Local network indicators

In order to characterize transport networks topology, network indicators are used extensively (Cats, 2017). For the analysis of PTN, a number of network indicators are reevaluated to make those suitable for the specific context of PTN (Derrible & Kennedy, 2009). Analyzing the topological network indicators can lead to new insights regarding common features between different types of networks (Cats, 2017). Therefore, this study proposes a new network indicator to quantify the hierarchy of a network. In order to do so, existing network indicators are evaluated briefly to identify how a hierarchy indicator is connected to this. On the one hand, indicators to examine a specific node (i) in the network are used to analyze the functioning of nodes within the network. On the other hand, Wang, Mo, Wang, and Jin (2011) mention some network indices to examine the overall network structure such as degree distribution and average path length. Therefore, it is sensible to discuss the implications for indicators both on a global[3] and local[4] level. Consequently, the local indicators are explained first throughout the following paragraphs in which their global counterparts are discussed too. Thereafter, some more general global network indicators are briefly elaborated.

Node degree

In order to provide some background to the indicators in the remainder of this section, the node degree has to be explained first as it partly applied for other indicators. The node degree (k_i) is defined as the number of nearest neighbors (Von Ferber, Holovatch, & Holovatch, 2009)[5]. Consequently, nodes with many links attached have a higher node degree which makes the interpretation of this indicator intuitively easy. By analyzing the degree of nodes in a network, hub nodes can be identified (Gu, Yang, Wang, & Wang, 2008).

On the global level, the distribution of the degrees of the nodes in a network is defined as $P(k)$ (Berche et al., 2009). Lu and Shi (2007) note the degree distribution is seen as an important geometric property of a network. For the degree distribution, a power-law distribution (Newman, 2003; T. Zhou, Yan, & Wang, 2005; Stephen & Toubia, 2009; Sienkiewicz & Hołyst, 2005) is described which indicates an exponential decrease in nodes with a higher degree. This implies there are many nodes with a low degree while there are few nodes with a higher degree. Sienkiewicz and Hołyst (2005) note that whether distribution of the degree follow a power-law depends on the type of space-representation that is applied, which is elaborated in 2.2.2. Berche et al. (2009) emphasize the power-law distribution by mentioning PTN as scale-free networks, which are determined as obeying a power-law following $P(k) \sim k^{-\gamma}$ (Barabási et al., 2001).

In order to determine hierarchy in a PTN, the node degree could be used as an input to determine which nodes are important based on their degree alone. However, that would be jumping to conclusions too quickly as the node degree only determines the number of links of a node rather than the function of these links in the network. Van Mieghem, Wang, Ge, Tang, and Kuipers (2010) note that degree distribution alone provides only limited information as networks with a similar distribution may still differ significantly in terms of topological characteristics. Therefore, studying more elaborate indicators, based on node degree, can lead to additional insights in this regard.

Assortativity

In order to probe the hierarchical structure of a network, Hu and Zhu (2009) propose the concept of assortativity which is defined as the average degree of the nearest neighbors. Based on this indicator,

[3] Referring to the overall network structure level
[4] Referring to indicators for the nodal level
[5] Yerra and Levinson (2005) mention the formation of hierarchies of node-connectivity are referred to as the number of adjacent nodes, which is interchangeable with node degree in this definition

it can be determined whether a network shows assortative of disassortative mixing on the degree (Newman, 2002). The former indicates the assortativity increases for nodes with a higher degree while the latter indicates vice versa. Hu and Zhu (2009) define the assortativity of node i as

$$k_{nn,i} = \frac{1}{k_i} \sum_j a_{ij} k_j, \tag{2.1}$$

where k_{nn} refers to the degree of the nearest neighbors of i. Therefore, a high assortativity value indicates a node is connected to other nodes with a high average degree while a low assortativity value indicates the node is attached to nodes with a low average degree. To provide an example how assortativity should be interpreted intuitively, figure 2.2 shows nodes with high assortativity in box A and nodes with low assortativity in box B.

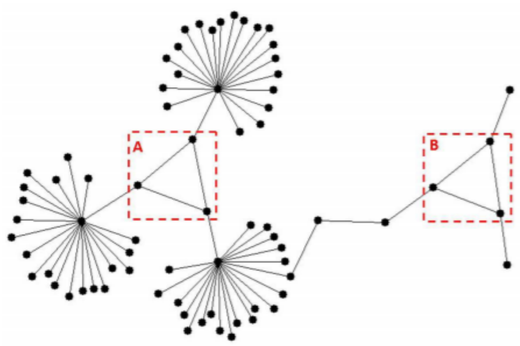


Figure 2.2: Illustrative example of assortativity (G.-Q. Zhang et al., 2012)

From a global point of view, assortative networks in this definition have nodes with a high tendency to attach to other high degree nodes wile in disassortative networks high degree nodes attach to low-degree nodes (Van Mieghem et al., 2010; Qing, Zhenghu, Zhijing, Zhang, & Zheng, 2013). In order to analyze the assortativity, the assortativity coefficient is calculated as a function of degree k where an increasing function implies an assortative network while a decreasing function implies a disassortative network (Xia-Miao et al., 2010; Hu & Zhu, 2009). Sienkiewicz and Hołyst (2005) conclude that the assortativity coefficient for Polish PTN remains positive in the $\mathbb{L}$-space for the whole range of nodes, implying a correlation between nodes of similar degree. On the other hand, Qing et al. (2013) come to the conclusion that assortativity transits at networks with thousand nodes and that networks with $N < 1000$ tend to be disassortative while most networks with $N > 1000$ are assortative. From a hierarchy perspective, assortativity can be interpreted in numerous ways. On the one hand, a high assortativity coefficient could imply nodes with a high degree form an upper layer to travel between different clusters. On the other hand, a high assortativity coefficient value[6] could imply disconnectivity between lower degree and higher degree nodes (Van Mieghem et al., 2010) as nodes mainly attach to nodes with a similar degree. A high disassortative coefficient value[7], could mean that nodes with a low degree attach to at least one node with a high degree (the hub) which connects to other hubs, such as a hub-and-spoke network (Hu & Zhu, 2009; Wang et al., 2011). A high disassortativity coefficient could also mean nodes with a high degree are not connected among themselves. An assortativity value close to zero could imply mixing of nearest neighbor node's degrees in a sense that high degree nodes attach both to low degree nodes and high degree nodes which suits a hierarchical network design. However, a coefficient close to zero could

[6] A high assortativity coefficient is a value between 0 and 1 where 1 implies perfect assortativity
[7] The disassortativity value is similar to the assortativity value but with a negative value between 0 and -1

also mean some high degree nodes attach to other high degree nodes only while other high degree nodes attach to low degree nodes. Therefore, the value of the assortativity coefficient value can be interpreted in different ways for positive and negative values as well as values close to zero. Consequently, while the assortativity coefficient can provide valuable information about the hierarchy in a network, conclusions regarding hierarchy should be drawn carefully due to the ambiguity of the interpretation of the coefficient.

Clustering coefficient

The clustering coefficient is used to measure the degree to which nodes are interconnected (Xia-Miao et al., 2010). In other words, this metric determines the share of neighboring nodes which are mutually connected. Ru and Xu (2005) define the clustering coefficient for each node *i* with k_i edges as

$$c_i(k_i) = \frac{n_i}{k_i(k_i-1)/2}, \tag{2.2}$$

in which n_i indicates the number of existing links between the neighbors of the node *i*. For a hierarchical node, it would make sense the clustering coefficient is low[8] as the importance of the node would decrease with a higher clustering coefficient due to possible alternative routes. A clustering coefficient of zero would indicate any route between neighboring nodes would go through either node *i* or through a node not connected to *i*. To provide a clarification of how the clustering coefficient is determined, Bettinardi (2016) provides an example of the yellow node with a degree of 4 and a c_i of 4/6 as 4 out of 6 links between the neighbors exist. The example is visualized in figure 2.3.

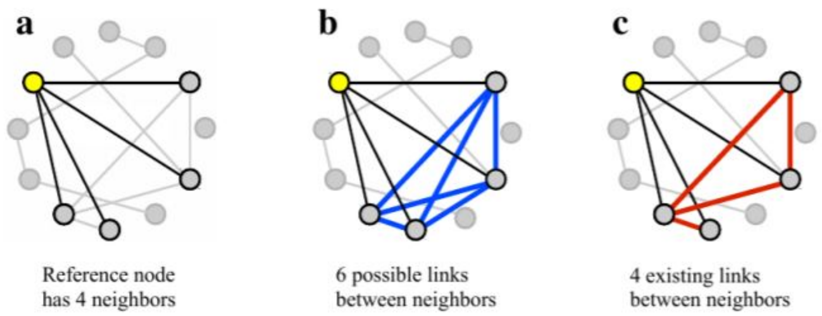


Figure 2.3: Illustrative example of clustering coefficient where (a) indicates the neighbors, (b) indicates possible links among neighbors and (c) indicates existing links among neighbors (Bettinardi, 2016)

On the global scale, a clustering coefficient *C(k)* is defined where the value of c_i for all nodes with degree *k* is averaged. Ru and Xu (2005) define the clustering coefficient of nodes with same degree *k* as

$$c(k) = \frac{\sum_{i;k_i=k} c_i(k_i)}{N(k)}, \tag{2.3}$$

in which *N(k)* is the number of nodes having a degree *k*. Hu and Zhu (2009) notes that *C(k)* exhibits highly nontrivial behavior with a decay curve as a function of *k*. They mention this behavior

[8] Low in both the $\mathbb{L}$-space and $\mathbb{P}$-space as the 'hierarchical node' is then an important hub for transfers

indicates a hierarchical structure in which low-degree nodes are generally well connected mutually while high degree nodes have many neighboring nodes that are not directly connected. For the clustering coefficient in the ℙ-space, described by Sienkiewicz and Hołyst (2005); Hu and Zhu (2009), it makes sense that nodes with a low degree with $k_i \leq 2$ have a high clustering coefficient. This is because the low-degree nodes are often located on one PT line and all nodes on one line are connected mutually in the ℙ-space.

Relating the clustering coefficient to hierarchy, there are some interesting connections. First, it should be noted that that the model representation for the clustering coefficient is very important as a link between neighbors in the 𝕃-space indicates a triangle of links between three nodes while a link between neighbors in the ℙ-space indicates a line passing through the three nodes. Therefore, choosing either representation has implications for the generalizations regarding hierarchy. Secondly, in general, a lower clustering coefficient of a node *i* means the node is more inevitable when traveling on a route through the neighbors of *i* as there is no direct link between them implying a lower clustering. A higher clustering coefficient on the other hand indicates many of the neighbors are mutually connected which makes traveling through [9] or transferring[10] at the node unnecessary. Consequently, a lower clustering coefficient of a node could be related to being more hierarchical. From a network hierarchy point of view, a decreasing function of *c(k)* could imply a more hierarchical structure.

Betweenness centrality

The betweenness centrality is an indicator which is particularly capable of highlighting the importance of a node as a transfer point (or passing through point) for a node pair (Derrible, 2012). Betweenness is defined as the share of shortest paths between every pair of nodes that passes through a given node *i* (Hu & Zhu, 2009). This indicator is especially useful for PT to understand the function of a transfer station in the network (Derrible, 2012). Hence, the betweenness centrality captures a special feature for a node in the network: the pass-through point (Porta et al., 2009). The betweenness centrality is defined as

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{j,k} \frac{n_{jk}(i)}{n_{jk}}, \tag{2.4}$$

in which n_{jk} is the number of shortest paths between *j* and *k* while $n_{jk}(i)$ is the number of shortest paths between *j* and *k* that contains node *i* (Hu & Zhu, 2009). A higher value for the betweenness centrality of a node *i* indicates the node is present in a higher share of node pairs and is therefore referred to as more central (Derrible, 2012). For the 𝕃-space representation, a higher value for the betweenness centrality may be an indication the shortest path and passes through or transfers at node *i* while for the betweenness centrality in the ℙ-space representation (De Bona, Fonseca, Rosa, Lüders, & Delgado, 2016), it only relates to transfers at node *i.* To determine the number of shortest paths going through a specific node from a topological perspective, different algorithms can be used (Brandes, 2008).

On the global scale, the distribution of the betweenness centrality in large complex networks obeys a power law (Barthelemy, 2004). Derrible (2012) notes that betweenness consistently becomes more distributed over nodes with network growth. Intuitively this makes sense as network growth often occurs near the edge of a network or by the means of creating missing links between existing clusters, relieving central nodes.

[9] For the 𝕃-space representation
[10] For the ℙ-space representation

From a hierarchy perspective, betweenness can partially capture the importance of nodes in a hierarchical structure. The betweenness centrality indicator can identify which nodes are important with many shortest paths crossing a certain node. However, betweenness centrality, in particular in the 𝕃-space, is sensitive to unjust conclusions as, e.g. a node very close to a hub, which has a lot of lines 'passing-through' while not many transfers take place at this node (as the transfers occur in the hub), would have a high betweenness centrality while the node may not be that important in the hierarchical structure. Consequently, this node would be overemphasized in terms of its betweenness centrality. On the one hand this can be seen as a conceptual pitfall of topological network analysis of PTN in general, by neglecting demand, frequencies, etc. but for betweenness centrality more specifically the near hub nodes would unjustly 'take away' the betweenness. In order to determine the hierarchy, betweenness centrality could therefore provide some insight but is prone to misconceptions. It should be noted that the misconceptions would relate to the 𝕃-space representation and not necessarily apply to the ℙ-space representation of the betweenness centrality. However, even for the ℙ-space betweenness, every node pair is valued equally while the demand for between each node pair varies significantly.

Eigenvector centrality

Connected to the betweenness centrality, eigenvector centrality is based on the centrality of nodes. However, instead of just the centrality of a node itself, it is based on the centrality of directly connected nodes. In other words, the quality of the links for a node are taken into account by valuing links to a more central node a higher than links to a node with a low centrality (Soh et al., 2010). Soh et al. (2010) have defined the eigenvector centrality x_i for a node i as

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} a_{i,j} x_j, \tag{2.5}$$

in which λ is a constant value which is written in vector-matrix notation,

$$\lambda \mathbf{x} = \mathbf{A} \cdot \mathbf{x}. \tag{2.6}$$

In this vector-matrix notation, $\mathbf{A}$ is an adjacency matrix, $\mathbf{x}$ is the eigenvector with eigenvalue λ. By applying the eigenvector centrality to the adjacency matrix of a network in the ℙ-space, the ranking of nodes in a PTN is determined from the perspective of PT lines (Majima, Katuhara, & Takadama, 2007). In other words, it ranks nodes based on having direct lines to important hubs. For the global scale, the only reference to its application is by taking the average based on the work of Soh et al. (2010). The usefulness of the average of this indicator is, however, rather limited as it does not explain the distribution in the network.

Relating the eigenvector centrality to hierarchy in PTN, this measure is excellent in indicating the function of a node in combination with the function of neighboring nodes. By implicitly incorporating the function of directly connected nodes (for the eigenvector in the ℙ-space), the influence of a node can be determined. In general, nodes that have higher influence in the network are those that would be the more hierarchical nodes.

Overlapping degree

This indicator can be seen as the odd one out as it relates to the different number of lines or modes a node is served by and can, therefore, not be derived from the degree in the 𝕃-space or ℙ-space. It would require a mode aggregated 𝔹-space representation in which the links to a node explain if a node is served by a mode or not. For this indicator, the network has to be split into different layers. The different layers are based on different modalities or lines (Aleta et al., 2017). The overlapping degree is defined as

$$o_i = \sum_{j \in M} k_i^j \tag{2.7}$$

in which k_i^j is the degree in layer j and M is the set of modes. In other words, this would be the degree of node i in the $\mathbb{B}$-space with numbered nodes as aggregated modes[11]. This indicator is particularly useful when modal layers are defined and available and transfer locations between modes are known. If this is not the case, specifying which links and lines belong to which layer requires extensive manual work for sizable networks.

On a global scale, the distribution of the overlapping degree is unknown although a declining distribution function is implicitly suggested by Aleta et al. (2017). However, due to the absence of the research on the curves of these distribution functions, no conclusions can be drawn.

Regarding hierarchy in PTN, the overlapping degree could provide some useful insights, especially because of the different nature of this indicator compared to the other indicators. However, one important remark should be made, as this indicator only indicates whether a node is covered by a certain mode, other conclusions about the importance and functioning of the node within the network cannot be drawn based on this indicator alone.

Overview of local indicators

To summarize the discussed indicators in the context of hierarchy, an overview is presented in table 2.3. In this table, the indicators are briefly summarized and some of the previous studies used for this study are mentioned.

2.2.4 Global network indicators

In table 2.4 an overview of the global counterparts for the local indicators is given. For the global indicators, the relevance for the hierarchy coefficient and previous studies are mentioned once more. Furthermore, the type of function to describe the indicator distribution is given.

In order to determine the hierarchy of the network, or on the global scale, it makes sense to consult similar methods of determining global network indicators as explained in table 2.4. Applying a generally used approach to determine the hierarchical coefficient, a distribution function of the hierarchical degree within the network could be used to analyze the number of nodes for a certain hierarchy value. The curve of the hierarchy value distribution could then be examined provided it obeys a power-law from which the alpha value can be derived. The alpha value is used as a negative coefficient in the formula

$$p(x) = x^{-\alpha}, \tag{2.8}$$

in which x would be the share of nodes with that value for hierarchy or higher and p(x) would be the number or share of nodes in the range. The alpha value is used as a negative power to give an exponentially decreasing function for increasing hierarchy. Based on this alpha value, the distribution of the hierarchy in the network can be analyzed.

[11] This implies every bus, metro, train etc. is seen as one node and every station served by the mode is connected to this node

Table 2.3: Overview of local network indicators

Local network indicator	**Notation**	**Explanation**	**Explaining hierarchy**	**Previous studies**
Node degree	k_i	Number of links that are directly connected to a node	+-	Hu and Zhu (2009) Gu et al. (2008) Van Mieghem et al. (2010)
Assortativity	$k_{nn,i}$	Average degree of nearest neighbors	+-	Hu and Zhu (2009) Xia-Miao et al. (2010) Newman (2002) Qing et al. (2013)
Clustering coefficient	C_i	Share of node's direct neighbors that are mutually connected	+	Xia-Miao et al. (2010) Ru and Xu (2005)
Betweenness centrality	$C_B(i)$	Share of shortest paths go through a certain node	+-	Derrible (2012) Hu and Zhu (2009) Porta et al. (2009) De Bona et al. (2016)
Eigenvector centrality	x_i	Influence of a node and its directly connected neighbors	+	Soh et al. (2010) Majima et al. (2007)
Overlapping degree	o_i	Presence of nodes in different layers	+-	Aleta et al. (2017)

Table 2.4: Overview of global network indicators

Global network indicator	**Function type**	**Relevance for hierarchy coefficient**	**Previous studies**
Degree distribution	Truncated power-law distribution	+-	Hu and Zhu (2009) Sienkiewicz and Hołyst (2005) Berche et al. (2009) Lu and Shi (2007) Newman (2002)
Network assortativity coefficient	Assortativity coefficient as function of degree k	+-	Hu and Zhu (2009) Sienkiewicz and Hołyst (2005) Qing et al. (2013)
Clustering coefficient distribution	Clustering coefficient as function of degree k	+	Hu and Zhu (2009) Sienkiewicz and Hołyst (2005) Berche et al. (2009)
Betweenness distribution	Power-law obeying distribution	+-	Derrible (2012) Hu and Zhu (2009) Barthelemy (2004)
Average weighted centrality	Average	-	Soh et al. (2010)
Overlapping degree distribution	Undetermined	-	Aleta et al. (2017)

Trusina et al. (2004) have come up with a method to determine the hierarchy of a network. They indirectly use the degree to determine hierarchy by defining a link as hierarchical up path if it goes from a link with a lower degree to a link with a higher degree and down path vice versa. Following, a shortest path is considered hierarchical if the path follows a sequence of up paths followed by a sequence of down paths[12]. The fraction of hierarchical shortest paths and total number of shortest paths then determines the hierarchical structure. This idea provides some interesting possibilities within the context of PTN. The shortest path is analyzed in terms of the hierarchy of the nodes on its path. However, the shortest path in the $\mathbb{L}$-space is likely to be a non-hierarchical path in PTN as there are "passing through" stops in between hubs with no hierarchical degree. Therefore, instead the $\mathbb{P}$-space could be used for which the shortest paths only takes the transfer nodes into account. The issue with this is that for most shortest paths in a PTN a maximum of two transfers is required for which a hierarchical path cannot be determined. Hence, this approach does not provide the desired tools to determine the hierarchy of a PTN with the size of a city.

Another approach to analyze the distribution in a network structure is the Gini inequality index (Lämmer, Gehlsen, & Helbing, 2006; Xie & Levinson, 2009; Reynolds-Feighan, 2001; D. M. Levinson, Xie, & Zhu, 2007). The Gini inequality index is applied to test the inequality in a distribution. Based on an approximation, the discrete form of the formula for the Gini index is defined by Xie and Levinson (2009) as

$$G = 1 - \sum_{k=1}^{e} (X_k - X_{k-1})(Y_k + Y_{k+1}) \qquad \{G \in \mathbb{R} : 0 \leq G \leq 1\} \tag{2.9}$$

where X_k represents the cumulative portion of links, or nodes, with $k = 0, 1, ..., e$ while Y_k would be defined as the cumulative portion of hierarchy in a network in this sense. The links or nodes are ranked ascendingly based on their hierarchical value. The lower bound for the Gini-index is 0 which implies the hierarchy is spread out evenly among every node and there is no in-equality which would be the case when every node would have exactly the same hierarchy. The upper bound would be a value of 1 which implies a completely in-equal spread of the hierarchy. In other words, one node would have all of the hierarchy and the other nodes nothing. This measure is especially useful to determine how in-equally the share of an indicator is distributed throughout the network. For a hierarchical network in particular, an in-equal spread of hierarchy among the nodes would be expected. This is due to a uniformly distributed network in terms of hierarchical degree cannot be considered as hierarchical as no different hierarchical levels are distinguished. On the other hand, a completely in-equal distribution would imply that one node has all of hierarchy in the network, such as a hub-and-spoke network with one hub.

2.2.5 Hierarchy related to network characteristics

Besides the obvious function of defining the hierarchy of a network, a hierarchy indicator can also be applied to get a better understanding of other network characteristics. Hence, the hierarchical structure could also be an explaining factor for other measures next to its intended function of determining hierarchy in a network. To give some insight in how a hierarchy indicator can be relevant for other factors, vulnerability and cascading failures are used as an example to indicate the applicability of a hierarchy indicator in a wider perspective.

Vulnerability and cascading failures

The societal importance of a robust and reliable transport system has led to considerable research regarding the vulnerability of a network (Mattsson & Jenelius, 2015). Furthermore, regarding PTN,

[12]Where either sequence can be 0 as long as there is no mixing of up and down (e.g. up path followed by down path followed by up path)

topological indicators have most widely been used for network vulnerability (Cats, 2017; Reggiani et al., 2015). The concept of vulnerability relates to the sensitivity of networks as a result of disruptions to nodes (Angeloudis & Fisk, 2006) and to the potential damage to a network from a negative perspective (Y. Zhou et al., 2019). Cats and Jenelius (2015) note there is a lack of knowledge to overcome risks relating to disruptions. Murray, Matisziw, and Grubesic (2008) relate the concept of vulnerability to policy by indicating the challenge to develop strategic plans for a reduction of vulnerability to disruptions. This is an interesting link how additional insight in vulnerability could be applied for shaping policy regarding a future-proof network.

Latora and Marchiori (2001) developed a measure for vulnerability by defining it as the importance of a node and measuring it as the decrease in efficiency if the node is deactivated. In the context of hierarchy, the decrease in efficiency can be transformed to the change of the hierarchical structure, leading to a measure for the importance of a node regarding the hierarchy of a network. However, Criado, Flores, Hernández-Bermejo, Pello, and Romance (2005) note that the concept of node importance is unable to distinguish networks that should have different vulnerabilities based on their network structure[13]. Relating this to the hierarchical structure, intuitively a hierarchical network is more or less sensitive to disruptions based on the location of a disrupted node in the hierarchical structure. A more hierarchical node is likely to cause a higher decrease in network performance than a node with a lower hierarchy. On the same note, a disruption to a link between two hierarchical nodes or a feeder node to a hierarchical node is likely to cause more problems than a disruption to a link connecting two low hierarchical nodes.

Cascading failures relate to the failure of one node or edge, spilling over to other nodes and edges triggering chain reactions that could lead to disruptions throughout the entire network (Huang, Zhang, Guan, Yang, & Zong, 2015) or result in a large-scale collapse within the network (Dong et al., 2014).

Connected to the concept of cascading failures, load redistribution is described as relocating passenger flows to follow an alternative route after an attack or random failure (Dong et al., 2014). An illustration of load redistribution is visualized in figure 2.4. Load redistribution can be linked to hierarchical structure by the means of defining alternative routing through different hierarchical levels. If, for example, a hierarchical node is unavailable, alternative routings through lower hierarchical levels may be desired while the unavailability of a lower hierarchical node may require routing adaptation on the highest level.

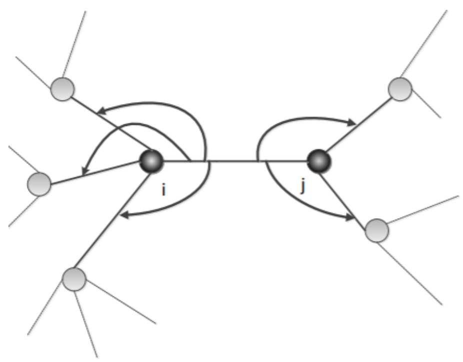


Figure 2.4: Visualization of load redistribution (Dong et al., 2014)

[13]They mention a full graph and a cycle graph both have a vulnerability of zero despite the former being more robust intuitively

2.3. Empirical network analysis

Diverging from the topological approach to analyze PTN, a more empirical method, incorporating passenger data, is applied based on travel demand. Using demand data from the network, an analysis of the network can be done to determine the hierarchy of the network based on passenger flows. First, an overview of travel demand models is presented. Thereafter, a link to hierarchy and how to connect hierarchy with demand is elaborated. Finally, the concept of bottlenecks, which is explained in terms of demand and supply, is clarified and outlined.

2.3.1 Travel demand models

In order to understand how travel demand models are applied for this study, first a clarification of its definition should be specified. Each node in the network is assigned as an *origin* and a *destination* (Gentile & Noekel, 2016). Based on this definition, the concept of an OD matrix (origin-destination matrix) may be used to represent the travel patterns in a network (Tamin & Willumsen, 1989). For every OD pair, a certain demand is known or can be calculated depending on the availability of data. Furthermore, between every OD pair a set of routes connects the pair where every route has different characteristics in terms of travel time, transfers, fare price, etc. (Gentile & Noekel, 2016; Fiorenzo-Catalano, Van Nes, & Bovy, 2004). As there are numerous assignment methods to estimate the flow, an assignment method, which is perceived optimal for determining the hierarchy, has to be chosen.

The most straightforward way of flow assignment is based on static demand implying the notion of time is unaccounted for (Schweizer, Danesi, Rupi, & Traversi, 2012; Flötteröd, 2015). By using static assignment, the advantages of more comprehensive flow assignment methods such as dynamic assignment are not taken into account. Advantages of such an approach include considering the effect of congestion (Sumalee, Tan, & Lam, 2009) and variations for different time intervals (Schmöcker et al., 2004) are neglected, leading to a limitation in terms of understanding the dynamics in travel behavior. Neglecting the congestion of vehicles, and nodes in general, may lead to a bias in terms of travel time as well as a preference to travel around the congested nodes (e.g. to avoid busy stations during peak hour).

The least comprehensive traffic assignment method is based on a all-or-nothing perspective. Opposed to a stochastic perspective, which would take the variety in route preference into account (Celikoglu & Cigizoglu, 2007), all-or-nothing assignment assumes all passengers choose the fastest route and are unaffected by congestion or preference. While stochastic route choice has the advantages that different route choices and preferences are taken into account, to account for variability in preference regarding traveling (Nielsen & Jovicic, 1999), all-or-nothing assignment has significantly lower computation times as only one shortest path between nodes has to be defined instead of a set of possible routes between nodes. Some of the relevant factors for determining the shortest path between a node pair on which the assignment is based, are in-vehicle time, waiting time and transfer time. All of these times are calculated by looking at average transfer time, vehicle speed and line frequency.

Hierarchical demand

For this study, total numbers of demand and flow are not directly relevant for determining the hierarchy. As has been explained in the topological network analysis, it is the functions of transfer hubs that are of particular interest. Therefore, hierarchical demand implicitly relates to transfers and more specifically, transfers between different lines at a node. The concept of transfer demand is relatively unbeknownst especially in this interpretation. Jiancong, Shaokuan, Yuqiang, and Liping (2006) note the importance of an increase of transfer demand in transportation hubs but focus towards a perspective of car to PT transfer. Yu, Yang, Jin, Wu, and Yao (2012) mention transfer demand from a link perspective, where the demand for a specific link consists of demand for the OD pair plus the transfer demand. Chen, Pel, Chen, Sparing, and Hansen (2012), on the other hand, study transfer demand from a node perspective, focusing on the transfer between access & egress modes and PT. Consequently, the concept of transfer demand, from a node perspective regarding different PT levels, remains underexposed as of now.

2.3.2 Bottlenecks

When the demand of a PTN is higher than the supply, bottlenecks tend to arise. Bottlenecks in PTN are in many cases caused by capacity restrictions (Rothengatter, 1996). While bottlenecks may appear as a main policy issue due to the loss of time it causes, the knowledge of its effect is limited (Witte, Wiegmans, van Oort, & Spit, 2012). Witte et al. (2012) also note the lack of capacity of infrastructure is only part of the problem and upgrading the capacity would only solve the problem partially. Furthermore, neither a micro nor a macro approach is able to capture the full complexity of bottlenecks (Lami, 2014). Consequently, a one-size-fits-all approach for bottlenecks appears impossible.

The causes for bottlenecks in a network are described as temporal obstruction, a capacity constraint or a stochastic fluctuation in demand (Wu, Gao, & Sun, 2009). As indicated in the previous paragraph, static models are incapable of modeling these factors while these do affect travel times and also the functioning of the hierarchical structure. If for example, the most hierarchical node in a network is in fact a huge bottleneck, causing severe delays, it would hardly be desirable to have this node this high in the hierarchical structure. To further elaborate the approach of this study to bottlenecks, types of bottlenecks and a link between bottlenecks and hierarchy are elaborated.

Types of bottlenecks

The two types of bottlenecks that are distinguished are structural bottlenecks (Li et al., 2015) and bottlenecks appearing randomly. Rodrigue, Comtois, and Slack (2016) defines these two types of bottlenecks as chronic and temporary. Rodrigue et al. (2016) further elaborates this by emphasizing chronic constraints are caused by physical restrictions while temporary constrains can be caused construction, accidents, events, etc.. Consequently, the two different types of bottlenecks have different relations to a hierarchical structure. A chronic bottleneck can be caused by hierarchy by, for example, having a hierarchical node at a capacity restricted location. A temporary bottleneck is more random of nature and its cause is harder to predict. However, random bottlenecks could be solved by having a hierarchical structure which is resilient in terms of redistributing passengers (Dong et al., 2014). This does, however, depend on the location in the network where the temporary bottleneck occurs where in general, the higher in the hierarchy the bottleneck appears, the more passengers are affected and have to be redistributed.

Bottlenecks and hierarchy

As mentioned in the previous paragraph, the concepts of bottlenecks and hierarchy can be linked as the hierarchy of a network can both be the cause of a bottleneck and provide a solution to overcome (or limit the delay caused by) bottlenecks. While hierarchy and bottlenecks in PTN are not explicitly linked in literature, it provides an interesting societal aspect this study can clarify. Furthermore, by linking these concepts, policy related implications could be identified. Hence, throughout the process of this research, the link to bottlenecks is tried to be maintained and emphasized.

2.4. Conclusion of literature review

Throughout this chapter, the relevant literature and theories for this study have been outlined. Based on this literature review, a definition for hierarchy in the perspective of this study is given. Furthermore, based on the review of different topological indicators and empirical network analysis, a tailor-made metric, incorporating both perspectives into a complementary metric is developed. Throughout the following chapter, the findings of this literature review are applied to define hierarchy and develop the metric in the methodology.

3

Methodology: Measuring hierarchy

Throughout this chapter, a definition for local and global hierarchy is given following from the literature review. Thereafter, a metric to determine hierarchy in PTN is introduced and elaborated in which different elements of the metric are distinguished. Local hierarchy, or the hierarchy of nodes, is referred to as the hierarchical degree and is explained first. Global hierarchy, or network hierarchy, is referred to as the hierarchical coefficient and is explained secondly as it can be derived from the distribution of the local hierarchy in the network. Furthermore, some potential limitations of the metric are elaborated to critically notice and explain its drawbacks. Stating its limitations is perceived necessary as the metric cannot incorporate everything, yet understanding what is excluded from the metric may benefit the evaluation and refining of the metric in a later stage. In the last part of this chapter, an overview of the methodology is shown in which the input, process and output of determining the metric are visualized in a conceptual overview. Based on this conceptual overview, the data requirements are briefly elaborated.

3.1. Definition for hierarchy in this study

It can be accepted that neither of the three approaches introduced in section 2.1 fully captures the hierarchical structure of a PTN. While each of these approaches has its strengths, there are some weaknesses for every approach as well. Therefore, this study suggests a combined approach to define hierarchy. To ensure the definition captures all perceived aspects of hierarchy in a PTN, a number of elements have to be included to determine the local hierarchy, or the hierarchy of a node:

- The node should be measured based on passenger transfers, simply passing through a node does not make it hierarchical.
- The function of the node should be evaluated, hierarchical nodes should facilitate changes between different functional levels and have a high influence in the network.
- Directly connected nodes have to be evaluated, ideally, a hierarchical node is connected to a mixture of hub and feeder nodes comparable to a hub-and-spoke network.
- The node has to be evaluated in terms of connectivity among neighbors, if neighbors share many mutual lines, the node becomes expendable as a transfer point.

In terms of the approaches, the definition in this study can be seen as a node-based hierarchy approach enhanced with link-based travel demand and line-based redundancy. It is a node-based approach as the output of the hierarchical degree only values the hierarchy of the nodes in the network. Link-based travel demand mainly relates to the transfer element of hierarchy, connecting feeder and mass routes. Line-based redundancy is related to the necessity for transferring at a certain node where there are no other direct lines between directly connected nodes. Hence, the perspective of the three approaches are included but the output is purely based on valuing nodes. Based on the list of elements introduced, the definition of a hierarchical node in a PTN would then be: "**a node which has high influence in the network by being directly connected to a wide range of nodes and operating as a transfer hub**". In this definition, connected to a wide range of nodes implies nodes have direct lines to many other nodes. However, as a direct connection to another important node could be seen as more important than a connection to a minor node, not all connections should be valued equally. Functioning as a transfer hub relates to the share of passengers transferring at the specific node and the number of directions to transfer to. Furthermore, by being a transfer hub, the node should have a degree of inevitability for the neighboring nodes. In other words, neighboring nodes should have a low number of mutual connections.

In order for a network to be hierarchical, the structure should show functional levels. Following from the hierarchical degree, the nodes should be ordered in such a way that there is a high share of nodes with a small hierarchical degree and a low share of nodes with a high hierarchical degree. Therefore, for a hierarchical network structure, or global hierarchy the definition has to include the following elements:

- The distribution of nodes in terms of their hierarchical degree should obey a power-law with many low hierarchical nodes and only a few very hierarchical nodes decreasing exponentially with increasing hierarchy.
- In general, the hierarchy of nodes in the network should increase for more connected nodes.
- The hierarchy of nodes should not be spread out equally among nodes in the network. In other words, the variety in hierarchy should be high to clearly distinguish more hierarchical nodes from less hierarchical nodes.

Based on these elements, the definition for a hierarchical network structure in the context of this study is given by: "**a network structure which shows distinguishable functional layers, which come together at transfer hubs, and has a power-law distribution of nodes in terms of hierarchy**". In this definition, the functional layers can be interpreted as levels of nodes in a certain hierarchy range in which every layer underneath contains nodes with a lower hierarchical degree.

Table 3.1 provides an overview of the approaches introduced previously and how the combined approach of this study relates to those approaches. Hereafter, the adaptation of the definitions towards a metric are elaborated.

3.2. Hierarchical degree

Following the approach of Lee et al. (2017) to define a new measure, the metric is created in sequential steps. This is done to evaluate the metric after its initial use and adjust it if desired. Furthermore, this allows for adding features of hierarchy step-by-step, based on the definition stated earlier. In order to assure the metric aligns with the definition for hierarchy, three separate elements are included in the metric which are: topological influence, redundancy and transfer potential. The three elements are chosen as these are directly derived from the definition for hierarchy and have a complementary function. For these three elements, the first two relate to the topological approach of the node while the last relates to the empirical approach. The first element of topological influence can be interpreted as the function of the node with respect to the functions of directly connected nodes based on the eigenvector centrality. The second element of redundancy can be interpreted

Table 3.1: Existing approaches to hierarchy in present literature

Approach to hierarchy	Explanation	Strength(s)	Weakness(es)	Previous studies
Link-based hierarchy	Define the hierarchy of each link in the network	• Evaluates every section of a line separately • Can be used to explain emergent behavior	• Unable to capture the importance and dynamics of transfer(s) (hubs)	Yerra and Levinson (2005), Lee et al. (2017)
Line-based hierarchy	Divide (sections of) lines into hierarchical levels	• Defines the role of transfers in the PTN	• Only captures the requirement for transfers but not the function of transfer locations	Jian et al. (2012), Bagloee and Ceder (2011), Van Nes (2002)
Node-based hierarchy	Determine the hierarchy of nodes based on their function	• Defines the function of hubs in the network	• Does not distinguish characteristics of links connected to the node	Ravasz and Barabási (2003), Mones et al. (2012) De Montis et al. (2007)
Combined approach	Node-based hierarchy enhanced with link/line travel demand	• Complementary view of hierarchy	• Requires both topological and demand data	-

as the superfluity of a node in the network. The third element relates to the demand flows for transfers in a node. Hence, the metric is defined in three separate elements which are elaborated first separately to clarify how the metric is established. It should be noted in advance that the metric is based on a $\mathbb{P}$-space representation for the first and second element while a $\mathbb{L}$-space representation is used as input for the third element. An additional adjusted space is implicitly applied to the third element to determine where transfers take place and to value transfer time realistically. Moreover, the hierarchical degree is only calculated for nodes with more than two directions to travel to, as it would not be able to function as a transfer location otherwise. Hence, the degree of the nodes in the $\mathbb{L}$-space should be higher than two or else the hierarchical degree should be set to zero, where

$$H_i = 0, \quad \text{if} \quad k_i \leq 2, \tag{3.1}$$

which holds for all three elements. In this equation k_i represents the degree of node i in the $\mathbb{L}$-space. Throughout the subsequent paragraphs, the different elements are explained and elaborated.

3.2.1 Element A: Topological influence

The first element relates to the topological influence of a node in the network, incorporating the function of nodes it is directly connected to. This element is derived from the eigenvector centrality with a small adjustment to normalize for the network. The inclusion of the function of directly connected nodes relates to the value of having a hub in close proximity, or direct connection, being beneficial for the node in question too. The eigenvector centrality is exceptionally useful to look at the function of a node while valuing connections to other important nodes higher than connections to less important nodes. If there are for example two relatively comparable nodes but one is connected to a major hub while the other is only connected to less connected nodes, intuitively the node connected to the hub has a more significant influence in the network by being the gateway to the hub. Moreover, by having more influence in the network, a node is considered more hierarchical as the function of the node, as stated in the definition, is more significant. In comparison to other indicators such as assortativity or a type of categorization of connected nodes, this indicator is less sensitive to outliers compared to assortativity while it does not require subjective choices which would be required for categories. Furthermore, this indicator is capable of identifying influence beyond a first degree neighbor to prevent unjustified high values for nodes being connected to a relatively low important node with a high degree. Hence, this indicator provides an objective way, not sensitive to the impact of outliers while taking not only first degree neighbors but also further neighbors into account.

The normalization of the eigenvector centrality is done in order to facilitate cross-network comparison as larger networks tend to have a more spread centrality with generally lower eigenvector centrality values. Furthermore, normalizing the eigenvector centrality enables a clear range for output values of the topological influence in the interval [0,1] where a value of 0 indicates the node is separated from the network while a value of 1 indicates the node is considered the most influential node in the network. As indicated above, the $\mathbb{P}$-space representation is applied to emphasize the importance of direct lines between nodes rather than the node being a direct neighbor. To clarify how this element can be obtained it is described as e_A where

$e_i^A \sim$ *the **topological influence** of node i, based on connections to other **influential nodes***

$$e_i^A = \begin{cases} \dfrac{x_i}{x_{max}}, & \text{if} \quad k_i > 2; \\ \\ 0, & \text{otherwise}, \end{cases} \qquad \forall i, \qquad \{e_i^A \in \mathbb{R} : 0 \le e_i^A \le 1\}, \tag{3.2}$$

and recalling from 2.2.3

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} a_{i,j} x_j, \tag{3.3}$$

in which $a_{i,j}$ is the adjacency matrix and x_j is the eigenvector centrality of neighboring nodes. Figure 3.1 visualizes how the eigenvector centrality can be determined for several nodes. 3.1a illustrates an example of node i with a degree of two, leading automatically to a $H_i = 0$. 3.1b is an example of a small network where node i is connected to all nodes and all opposite nodes are connected in the $\mathbb{P}$-space. The x_i value for node i is approximately 0.62 based on solving the eigenvector for the adjacency matrix[1]. This value is the highest of the network resulting in x_{max} being the value for x_i. Example 3.1c is a comparable network but one of the neighboring nodes is now a hub serving five other lines to five other nodes. Consequently, node i is no longer the most influential node in the network as node i is connected to four out of nine possible nodes directly while the hub is connected to seven out of nine possible nodes. Therefore, the hub now has the x_{max} value of 0.60. Another interesting aspect of this network is that the top node served by the red line has a higher value than the side nodes connected by the blue lines. This is due to this node being connected to more influential nodes (both i and the hub). The e_i^A value for i can still be seen as relatively high as only the hub would score higher. The final example (3.1d) shows a node i which would not be required for any transfers but still gets a relatively high value as it is connected to the two most influential only. Therefore, this example provides an example of how the topological influence is unable to fully capture the hierarchy of nodes on itself as node i gets a high value while it would not be required for any transfers. Consequently, the other elements should be seen as complementary to this element in order to fully capture the hierarchy of a node.

[1] It is solved using a Python code which is further elaborated in chapter 4

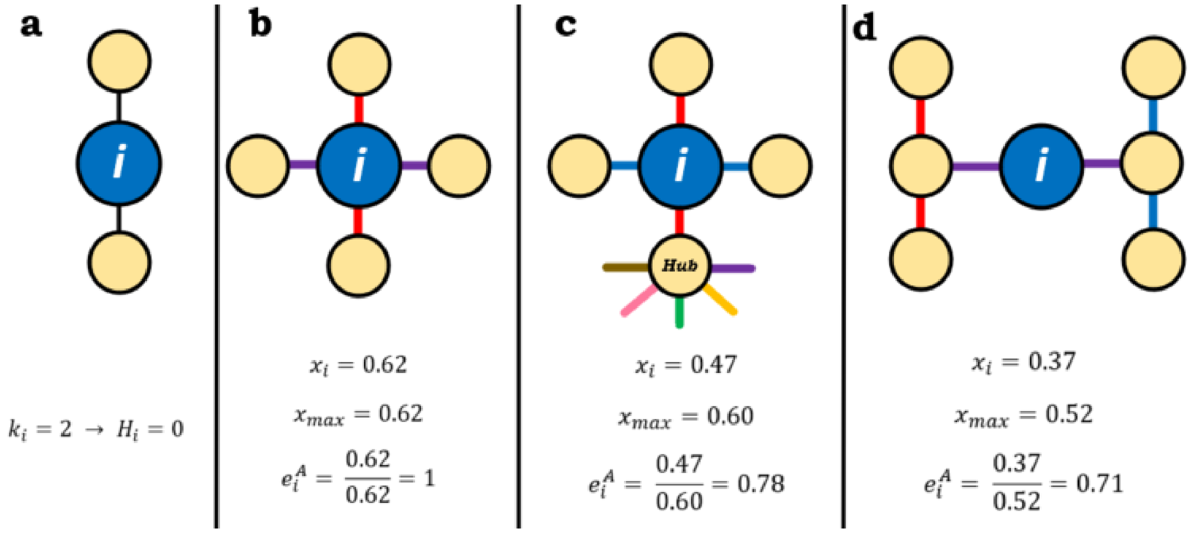


Figure 3.1: Illustrative example of topological influence values in small networks

3.2.2 Element B: Redundancy

The second element of the formula for the metric is directly derived from the clustering coefficient as explained in 2.2.3. This element is added to the metric in order to analyze the necessity of a node in terms of transfers. This element is used to test whether a node is required for the connection between its directly connected nodes. If many of these directly connected nodes were to share another mutual connection, only a limited number of OD pairs would require a transfer at the given node. Therefore, the less direct connections of a node are mutually connected, the more hierarchical it is perceived.

The clustering coefficient value of a node is determined to test the redundancy of a node in the $\mathbb{P}$-space. The $\mathbb{P}$-space is applied as this indicates which of the directly connected nodes share a mutual line between them and require, therefore, no transfer at the node in question. Consequently, a higher value for the clustering coefficient indicates many of the directly connected nodes share a line and only few transfers have to be made at the node in question. The clustering coefficient is subtracted from one, leading to a lower value if the clustering coefficient is higher. As explained in the definition for hierarchical nodes, less mutual lines among neighbors should contribute to a higher hierarchical degree which is why the clustering coefficient is subtracted from one. To explain what this element looks like as a function, it is described as e_i^B, where

e_i^B ~ *the **(non-)redundancy** of node i, based on **mutual connections** among neighboring nodes*

$$e_i^B = \begin{cases} 1 - c_i(k_i), & \text{if} \quad k_i > 2; \\ 0, & \text{otherwise}, \end{cases} \qquad \forall i, \qquad \{e_i^B \in \mathbb{R} : 0 \leq e_i^B \leq 1\}, \tag{3.4}$$

and recalling from 2.2.3

$$c_i(k_i) = \frac{n_i}{k_i(k_i - 1)/2} \qquad \forall i, \tag{3.5}$$

in which n_i indicates the number of links between neighboring nodes in the $\mathbb{P}$-space and k_i in this equation refers to the degree in the $\mathbb{P}$-space while it refers to the $\mathbb{L}$-space in equation 3.4. Figure

3.2 visualizes how clustering affects e_i^B. Once more 3.2a illustrates the example of node i with a ($\mathbb{P}$-space) degree of two, leading to $H_i = 0$ automatically. 3.2b is basically the same example as 3.1b but with four different lines to each of the neighboring nodes. This has significantly different implications for e_i^B as there are no mutual lines between neighboring nodes. Consequently, traveling from any of the neighboring nodes to one of the other nodes is only possible through transferring at node i leading to a high e_i^B value for the node. Therefore, the clustering is zero and e_i^B reaches its maximum value of one. 3.2c and d show how mutual connections affect e_i^B with two lines among neighbors in 3.2c and all connections among neighbors in 3.2d respectively. If all of the neighboring nodes are mutually connected, transferring through node i becomes redundant which leads to e_i^B being zero and eventually the hierarchical degree being zero as well.

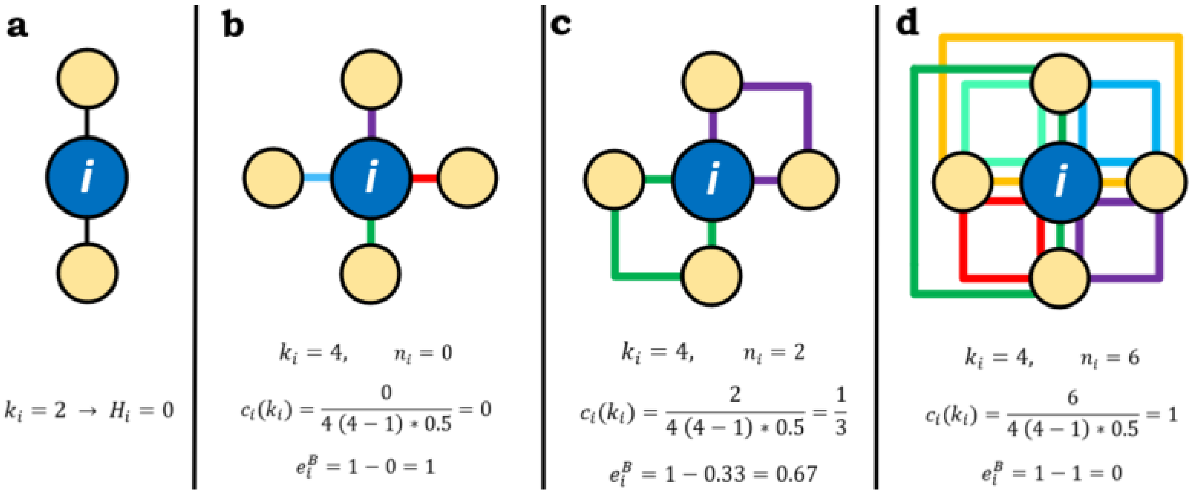


Figure 3.2: Illustrative example of (non-)redundancy values in small networks

3.2.3 Element C: Transfer potential

The third part of the metric is based on the share of transferring passengers at a node. Transferring in this sense means changing lines with possibly a different mode. This element is significantly different from the prior element as this one requires demand data and therefore incorporates the demand for transfers at a node. Consequently, the empirical aspect of hierarchy in PTN is complied with by including a passenger perspective with the share of transfers that take place at a certain node. The share of transferring passengers is multiplied by the share of potential transfer directions which is defined as the degree minus two divided by the degree in the $\mathbb{L}$-space. This definition of transferring directions is based on the fact that transferring here relates to getting off the vehicle and the vehicle generally has an in- and outgoing link. To explain this logic, assume a vehicle traveling from A to C through B where B offers transfers towards D, E, F and G. Then $\frac{4}{6}$ outgoing links require a transfer as traveling to C requires no transfer and A is the origin. e_i^C is defined

$e_i^C \sim$ *the **transfer potential** based on **transfer passenger share** and **transfer directions** for node i*

$$e_i^C = \begin{cases} \dfrac{log(P_i^{transfer})}{log(P_{max}^{transfer})} \dfrac{k_i - 2}{k_i}, & \text{if} \quad k_i > 2; \\ \\ 0, & \text{otherwise}, \end{cases} \qquad \forall i, \qquad \{e_i^C \in \mathbb{R} : 0 \le e_i^C < 1\}, \tag{3.6}$$

in which $P_i^{transfer}$ are passengers transferring at node i, $P_{max}^{transfer}$ is the the maximum number of passengers transferring at a node in the system. The degree k_i is here based on the $\mathbb{L}$-space of the network. $P_{max}^{transfer}$ is used to normalize the values where the node with the highest number of transfers has the highest transfer potential thus a value of one and the other nodes a fraction between zero and one. The logarithm is used as it is expected many nodes probably have just a small share of transferring passengers compared to the highest scoring node. Therefore, using the logarithm prevents getting values for other nodes becoming very close to zero for any node with a just a small fraction of the transfer passenger share. Intuitively it makes sense to use the logarithm in order to prevent the transfer share from having too much impact on the hierarchical degree. If non-logarithm transfer passenger numbers would have been used, any node besides a few crucial transfer hubs would have a value of practically zero.

To provide an example of how this part of the metric should be interpreted, figure 3.3 visualizes an example of e_i^C. Related to the previous figures, 3.3a provides the example of node *i* with a degree of two leading to the node being considered non-hierarchical without further consideration. 3.3b and 3.3c show how an increase in the share of transfer passengers increases e_i^C within the same network structure. Furthermore, 3.3d shows that while it offers way more transfer directions than 3.3c, the value of e_i^C remains the same due to a lower share of transfer passengers.

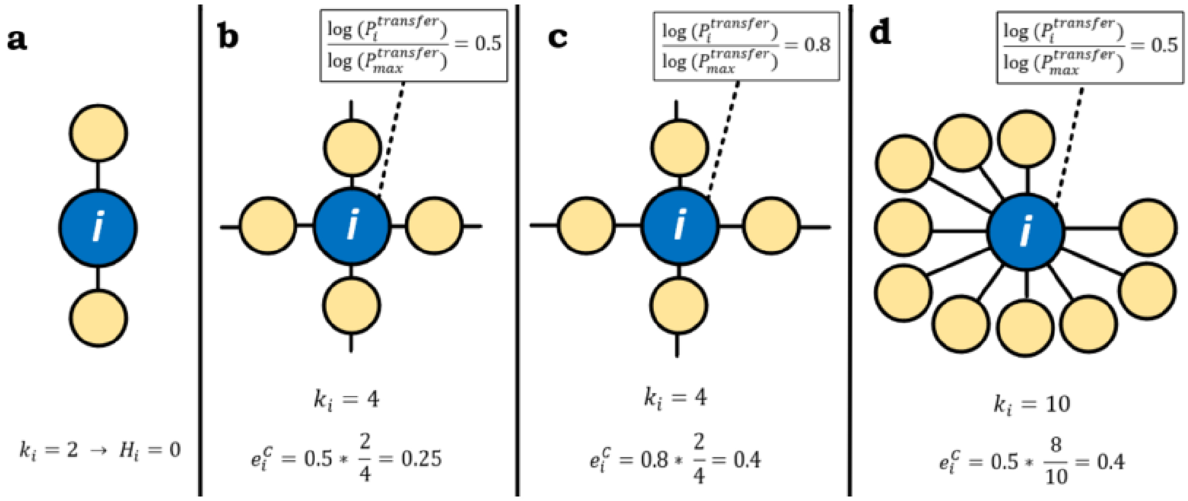


Figure 3.3: Illustrative example of transfer potential values in small networks

3.2.4 Combining the elements

In order to combine the three elements it is crucial to understand how these complement each other. For example, if a node would show high values for topological influence and non-redundancy but barely serves any transfer passengers, should it be considered as hierarchical? Based on the definition for local hierarchy, nodes should comply with different characteristics of hierarchy in PTN and only having a high value for one element does not imply that. To provide an example, a node connected to four hubs would have a high value for the topological influence since it has a high eigenvector centrality because of neighboring nodes. However, it is likely these hubs are directly connected themselves making the node in question score high on redundancy (low value for element B). Based on the definition for element B, a node which is redundant should not be considered as hierarchical. Therefore, this study suggests the elements should be combined by multiplying the three elements. By using multiplication to combine the elements, a high score for one of the elements does only lead to a high hierarchical degree if the other two elements have a decent value as well. Consequently, nodes with a high value for only one of the elements, are flattened in terms of hierarchical degree. Combining all three elements into one, the metric for the hierarchical degree reads

$$H_i = e_i^A * e_i^B * e_i^C, \quad \forall i, \qquad \{H_i \in \mathbb{R} : 0 \le H_i < 1\}. \tag{3.7}$$

with all the elements written out, the hierarchical degree can be defined as

$$H_i = \begin{cases} \frac{x_i}{x_{max}} * (1 - c_i(k_i)) * \frac{log(P_i^{transfer})}{log(P_{max}^{transfer})} \frac{k_i - 2}{k_i} & \text{if} \quad k_i > 2; \\ 0, \qquad \text{otherwise,} \end{cases} \quad \forall i, \qquad \{H_i \in \mathbb{R} : 0 \le H_i < 1\}. \tag{3.8}$$

The range for the hierarchical degree is between zero and one. The maximum value for a node can approach one but could only theoretically be one due to the redundancy and the transfer direction part of the transfer potential. For the redundancy, the value can only be one if no line passes through a node and every line only has one other stop (e.g. at figure 3.1b). This is generally not possible for PTN. The value for the second part of the transfer potential could only be one if the degree reaches infinity, which is impossible. Furthermore, the transfer share and topological influence can be one for the node with the most transfers and influence respectively.

After the relatively straightforward examples for the separate elements, one would wonder how these different elements combine for a real example, based on the comprehensive metric. To include the functioning of all elements in an example which is (partially) manually workable and also part of the subsequent case-study, the station Amsterdam Lelylaan is chosen as an example. It is worth noting this station is served by every PT modality considered for the Amsterdam PTN[2]. Based on timetables and maps of GVB (2019), figure 3.4 provides an overview of the node Lelylaan and its directly neighboring nodes in the Amsterdam PTN where unique lines between neighbors have a different color.

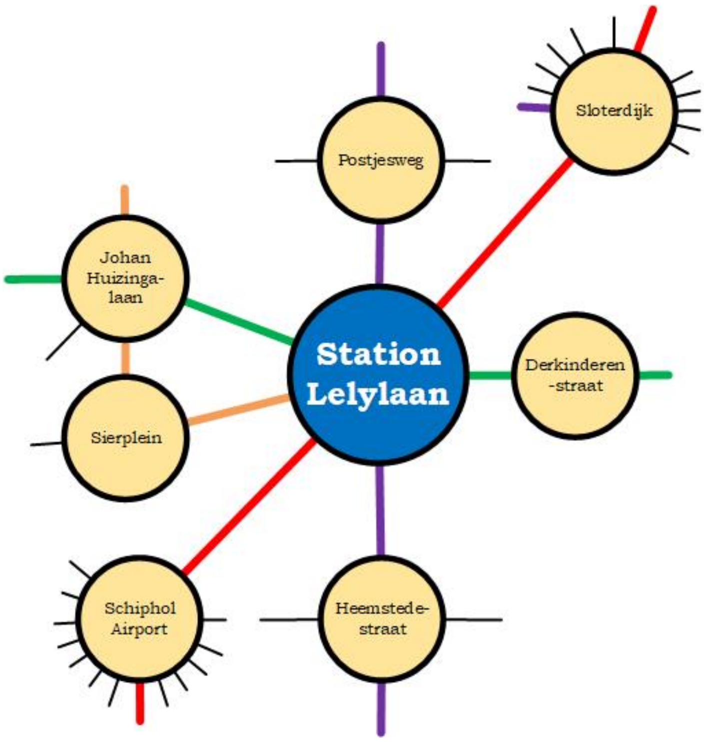
In this figure, the 1* value for transfer passenger share indicates an assumption that Station Lelylaan has the most transfers in the system as the transfer data are not applicable at this point. Furthermore it should be noted that topological influence and redundancy cannot be directly derived from this figure as it is drawn in the $\mathbb{L}$-space and a $\mathbb{P}$-space representation is required to determine the eigenvector centrality and clustering coefficient. Furthermore, a $\mathbb{P}$-space representation for Station Lelylaan would have a degree of 117 with 6786 neighbor pairs out of which 1691 are connected[3]. The calculation for the eigenvector cannot be computed manually either but can be checked afterwards. The hierarchical degree value of 0.324 for the node does not mean much on itself as there are no nodes for comparison but the node is considered high-hierarchical from the definition.

3.2.5 Multi-level representation

In order to subdivide the nodes into different functional levels, three sub-levels are distinguished which are the high-hierarchical, low-hierarchical and non-hierarchical for levels 1, 2 and 3 respectively. A node scoring zero is considered as non-hierarchical, while nodes with a hierarchical degree of higher than 0.125 (which is based on each element having a value of 0.5) are considered high-hierarchical and values in the interval (0, 0.125) are considered low-hierarchical. The value of 0.5 as threshold for level 1 is chosen as this is exactly in the middle of each of the range of the elements.

[2] including train, metro, tram and bus but excluding e.g. ferry

[3] This calculation is done using a Python script which is also applied for the case-study



$$x_i = 0.12$$

$$x_{max} = 0.20$$

$$\boldsymbol{e_i^A = \frac{0.12}{0.20} = 0.61}$$

$$c_i(k_i) = \frac{1691}{117\,(117-1)*0.5} = \frac{1691}{6786}$$

$$\boldsymbol{e_i^B = 1 - \frac{1691}{6786} = 0.75}$$

$$\frac{\log\,(P_i^{transfer})}{\log\,(P_{max}^{transfer})} = 1^*$$

$$\boldsymbol{e_i^C = 1 * \frac{5}{7} = 0.71}$$

$$\boldsymbol{H_i =\ e_i^A\, e_i^B\, e_i^C} = 0.61 * 0.75 * 0.71 = \boldsymbol{0.324}$$

Figure 3.4: Elaboration of hierarchical degree for Amsterdam Lelylaan

3.2.6 Limitations of the hierarchical degree metric

A limitation of the metric is the lack of demand perspective incorporated in the metric. While the demand data is used to determine the share of passengers transferring at a node, the dynamics in travel behavior and its influence on the hierarchical structure are only partially captured. Furthermore, as explained earlier, some travel behavior elements such as route preference, vehicle crowding, fare costs and time-of-day are not included in determining the route and therefore not included in determining the transfers either. These elements are excluded as this study focuses on aggregated daily demand to analyze the structure of the network from a static point of view. Dynamics throughout the day such as peak-hour variance in demand patterns and peak-hour lines changing travel routes, are beyond the scope of this study. Nevertheless, this simplification should be taken into account when drawing conclusions and could potentially be overcome in further research.

Another limitation of the hierarchical degree relates to the multiplication of the different elements, which on itself is an arbitrary choice. Hence, there is a boundary to the objectiveness of the metric. However, other methods of combining the elements such as taking the average or summing the values and normalizing based on the maximum, are less desirable. This is due to the definition for a node hierarchical being a combination of characteristics instead of scoring high on one element only. For example, taking the average could still lead to a potential score of 0.67 if one of elements would be zero. This would imply that the node has either no influence, is redundant or has no transfer potential and should not be considered hierarchical. Therefore, multiplication, in which a low value for an element affects the hierarchical degree significantly, is preferred despite its limitations.

3.3. Hierarchical coefficient

To determine the hierarchical structure of a network, a combination of the alpha value for the power-law distribution and the value of the Gini-index is used to incorporate both the decrease of nodes with a higher hierarchy and the in-equal spread of hierarchy throughout the network. These can be seen as complementary as the power-law distribution is used to analyze the decrease in share of nodes for increasing hierarchy while the Gini coefficient is used to analyze the spread of hierarchy in the network. Consequently, the power-law distribution is used to analyze the share of nodes while the Gini-index is used to analyze the share of hierarchy. Therefore, this study perceives these coefficients are complementary rather than overlapping. To define the hierarchical structure of a network the following function is used

$$H_I = G_I^{\frac{1}{|\alpha_I|}} \qquad \{H_I \in \mathbb{R} : 0 \leq H_I < 1\}, \tag{3.9}$$

in which H_I is the hierarchical coefficient of network I and G_I and α_I are the Gini index and curve of the power-law distribution function of network I respectively. The value for H_I indicates the hierarchy of the network where a value close to 0 indicates either a uniform hierarchical degree or a uniform hierarchical degree distribution. A value close to 1 on the other hand indicates either in-equal hierarchical degree or a very in-equal distribution. While a theoretical value of 1, where one node would have all hierarchical degree, would be beyond a hierarchical structure, the closer the value gets to one, the more hierarchical a network becomes. It is chosen to combine the alpha value and Gini-coefficient in this way as this function would provide a value between 0 and 1. For this function a higher Gini-coefficient, indicating a more in-equal distribution, would increase the hierarchical coefficient. Furthermore, a lower (higher negative) alpha value, indicating a steeper distribution function of nodes, would also increase the hierarchical coefficient. Intuitively this makes sense as a steeper function indicates a faster decreasing share of nodes with higher hierarchy where different hierarchical levels can be identified. It should be noted that a more hierarchical structure may not be desired in all cases and some adaptations to a network to increase the spread of hierarchical degree and, therefore, decrease the hierarchical coefficient could be advisable.

3.4. Conceptual overview

In figure 3.5 an overview of the in- and output for the hierarchical degree metric is given. This figure summarizes the required input for each of the elements followed by the way this input is processed to determine the output for each of the elements and the hierarchical degree. Based on the distribution of the hierarchical degree among the nodes, the hierarchical coefficient can be determined. In order to elaborate what the generic requirements for the data of the input are, these are elaborated in the following section.

3.5. Data requirements

In order to clearly distinguish which data are required as input for the metric, the input for the conceptual overview from figure 3.5 is used. First, the network of the PTN in study has to be defined for which every node, link and line within the boundary of the network has to be known. Based on the data for nodes and lines the $\mathbb{P}$-space can be determined while the data for nodes and links is used for the $\mathbb{L}$-space of the PTN. Each node is given a four numbered digit as node ID while each link is a combination of two node ID's which both serve as begin and end as the PTN is considered bidirectional. For each of the links, the link distance is added as link weight based on the greater circle distance between the nodes it connects. This could potentially be done using real link distance provided these data are available. For the lines, a numbered sequence of links is used which indicates the route of the line. Furthermore, each line is given an index and modality which are used

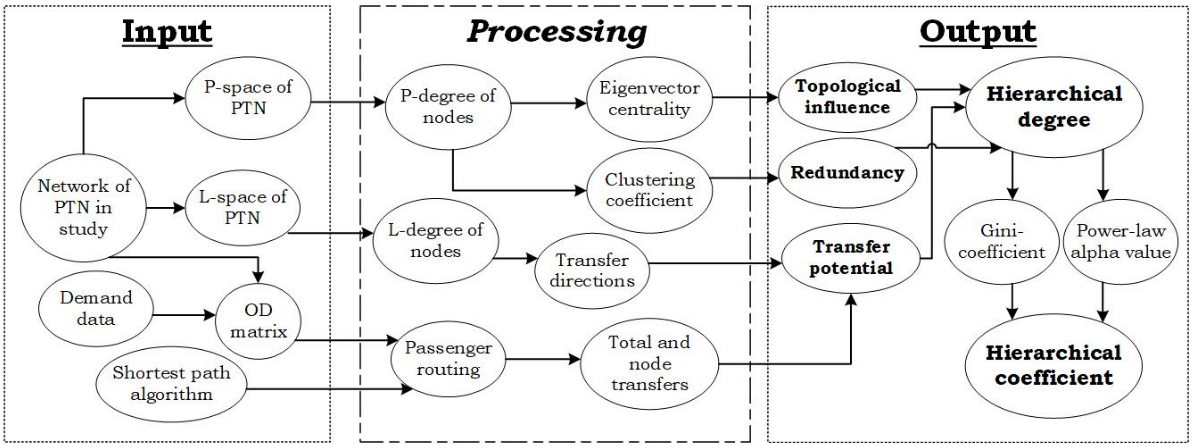


Figure 3.5: Conceptual overview of the hierarchical degree and hierarchical coefficient

to determine the line characteristics such as speed and frequency. While there can only be one link between a set of nodes, there are potentially infinite lines using the same link.

Secondly, the demand data is required for determining the transfer share of nodes. As the transfer share of nodes is often not directly obtainable, it is derived from travel patterns based on an OD matrix. The OD matrix can be acquired from the demand data which can potentially include demand to and from external areas going through the network. The demand data can be coupled to zones of the network or directly be linked to nodes of the PTN. In the former case, the demand of each zone is assigned to a node or a set of nodes in that specific zone while in the latter case this is not required.

The demand data can basically be the demand for any desired time-frame (e.g. peak-demand, hourly demand, daily demand) but for this study it is based on daily demand. It should be noted that a different type of demand data (e.g. smart card data or GPS data) would abolish the requirement for OD matrix estimation as it would be based on true observations rather than estimations. Furthermore, having observed travel data including route choice would nullify the requirement for passenger routing estimation and therefore overcome its limitations. This could be a potential application for the metric in the future provided that these data are available.

Following from the OD matrix, the routing between nodes can be determined using a shortest path algorithm which is based on an AON assignment. As noted above, this method of estimating demand and route choice is considered less realistic than using observed data but is chosen for this study due to the unavailability of smart card and GPS data. Based on the OD matrix and the shortest path algorithm, the route of each considered passenger can be determined from which the number of transfers per node is derived. With all of the data requirements clarified, the next step is the application of the metric to networks, which is explained throughout the following chapter.

4 Application

4.1. Implementation of the metric

In order to calculate the hierarchical degree for a large set of nodes, a programming code has been developed which can be applied to determine the hierarchical degree for a large number of nodes at once. This code is developed in Jupyter Notebook using a Python library called NetworkX which is particularly useful for studying graphs. As this library has built-in degree, eigenvector centrality and clustering coefficient functions, the topological analysis is relatively straightforward in that regard. The one thing that has to be taken into account is that both these elements are calculated in the $\mathbb{P}$-space. Therefore, a separate graph in the $\mathbb{P}$-space is used to calculate these elements. To verify the code, a small test network (of approximately 20 nodes) and the networks used as examples in 3.2 are evaluated to ensure the code works as intended. As the eigenvector centrality is quite hard to calculate manually even for simple networks, the output is multiplied with the adjacency matrix to verify the eigenvalues are equal. Therefore, this verification is the other way round compared to the verification of the clustering coefficient which is manually workable for small networks. Within the test network, some links can be added and removed to test the functioning of the code for the metric. The coding and verification of the code for transfer demand is discussed separately as this requires additional data to be imported. After the code for the metric excluding the demand is verified, the first case-study PTN can be imported to apply the coding to a more extensive network.

4.2. Coding the transfers for the metric

In order to determine the number of transfers for each node and the total network, demand data are added to the network. The demand is defined for every OD pair implying the number of transfers are not directly retrievable. Calculating the transfer potential not as straightforward as the other elements of the metric as the data regarding transfer demand is not directly derivable. First, based on a shortest path algorithm, the fastest route between each node pair is determined in which the stops at which transfers occur can be determined. Secondly, for each node pair a route and a demand can be determined which implies the transfer demand for each shortest-path of node pairs can be determined as well. Lastly, summing the transferring passengers for each shortest path transferring at a specific node (i) leads to the transfer demand for that given node (i). Based on the transfers at a specific station and the total number of transfers in the system, the transfer share can be determined. Note that transfers at external nodes are excluded from the total number of transfers as these do not occur in the system.

4.2.1 Defining the shortest path between nodes

In order to define the shortest path, the NetworkX library offers a built-in tool to determine the shortest path for each node pair based on node distance. However, to provide a more realistic shortest path determination and to analyze where transfers occur, this shortest path tool has to be used in a different way. First, to give a realistic view of the distance, each link is assigned a weight based on the great-circle distance of the two nodes it connects. To differentiate between the modality that operates a link, the in-vehicle travel time is calculated by the length divided by the speed of the mode serving the link. The speed of different modes may vary from network to network. Furthermore, the frequency for each of the modes which is used to determine the waiting time during a transfer[1]. For each stop on a route, a stopping time of 60 seconds is used which is added to the in-vehicle travel time.

4.2.2 Additional graph representation

In order to facilitate transfers, an additional and directed graph is used in which each node is split into a node for every line it serves. If e.g. a node x would be served by line 1, 2 and 3, there would be a node x_1, x_2, and x_3,. Furthermore, each node has an origin and destination node which can only be used as starting and end location respectively. This is implemented by giving the origin node outgoing links only and the destination node incoming links only. To clarify this, x_{origin} would have links to x_1, x_2, and x_3, while x_1, x_2, and x_3, would have links to $x_{destination}$ and to each other. By doing this, the origin and destination can be used as starting and end point for each transfer while the nodes with a line-number subscripted indicate the stop for that specific line. Nodes with a certain subscript can only be connected to other nodes with the same subscript and only if they are directly linked (using the weight of the travel time between the nodes). Nodes with the same base but a different subscript indicate possible transfers at a certain node. To clarify this type of network representation, figure 4.1 provides an example of two nodes and its notation in the $\mathbb{L}$-space and how this is adjusted to facilitate transfers and transfer times as link weights. This approach of defining and coding transfers is inspired from (D. Luo, Cats, van Lint, & Currie, 2019).

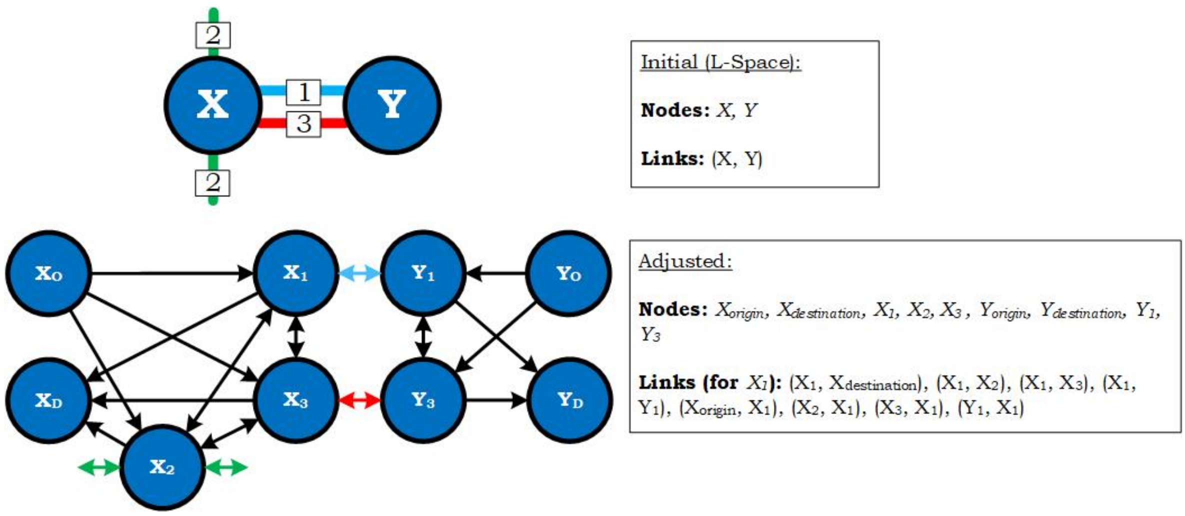


Figure 4.1: Adjusted graph representation of nodes to facilitate transfers where the top half represents the real network and the bottom half the adjusted space to incorporate transfers

The weight of the link between nodes with the same base indicates the transfer time which is set

[1]Both speed and frequency are based on assumptions for the average and can actually vary for every line, link and the time of day

to a default of 180 seconds plus the average waiting time in seconds. To determine the average waiting time, the frequency of the new line is used: Average waiting time = $\frac{3600}{Frequency} * \frac{1}{2}$ in which the frequency is based on the frequency of the incoming end of the link[2]. The average transfer time between two lines and vice versa may vary as the frequency of the incoming line may be different. Hence, the use of directed graphs is crucial to indicate the different transfer times.

A notable limitation of this definition of transfers is transferring to a line which shares the route for subsequent links with (an)other line(s). Based on the current definition, the waiting time is based on the frequency of one line, choosing one of the lines for the continuing of the route. However, as the link is shared by multiple lines, it does not matter which line is chosen for the route. Hence, the combined frequency of every line serving the following link could be used. To illustrate this intuitively confusing example, figure 4.2 shows why this on the one hand would provide a more realistic travel time to destination A while it would lead to issues for destination B. This problem is also known as the 'common corridor' problem. In the figure, the origin node is only served by one line, the red one. To reach either destination A or B, a transfer has to be made at the node transfer hub. For destination A, the chosen line does not matter as each of the three line serves the destination A node leading to a 'common corridor' with a frequency of 14 and an average waiting time of around two minutes. However, for destination B, only the purple line is an option as the other two lines do not serve this node. Therefore, the principle of the common corridor does not apply for this case as only 2 out of the 14 vehicles from this node go to destination B. For the other 12/14 vehicles, another transfer has to be made at destination A. Consequently, the average waiting time based on the shared frequency does not provide a realistic waiting time for the shortest path between origin and destination B.

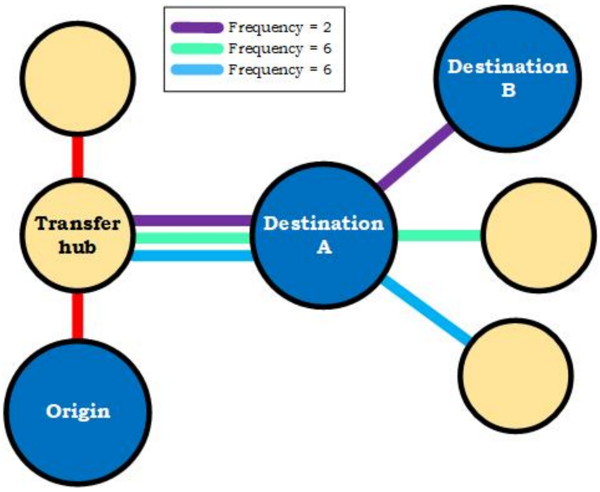


Figure 4.2: Example of common corridor issue for shared links

Considering this study focuses on the functions transfers at a certain node, the common corridor problem is handled by not adding the frequency of multiple lines in a common corridor. If the common corridor would have a shared frequency, a free transfer with no penalty time would have been given at Destination A in the example. To prevent the overestimation of transfers after a common corridor, it is preferred to avoid summing the frequencies of common corridors and accept that the transferring time for some routes may be overestimated. It should be noted that frequencies of PTN lines in city networks are generally high and waiting times relatively low. Therefore, even without

[2]The factor $\frac{1}{2}$ is used as e.g. a frequency of 4 indicates the line operates every 15 minute thus when you arrive you have an equal chance it will leave at that moment or you will have to wait for 15 minutes. Consequently, the average waiting time is 7.5 minutes, half of the maximum waiting time

summing frequencies of common corridors, transfers are generally preferred over detours leading to a very limited loss of transfer data. Related to this, nodes located on common corridors, being neither the first nor the last node of the corridor, cannot facilitate transfers. This is done to prevent randomly assigning transfers to any node on a corridor if a change between two lines sharing a part of the route has to be made. Instead, the transfer has to take place at the either the location where the lines converge or where the lines diverge.

4.2.3 From shortest path to transfer demand

Based on the adjusted graph with nodes for every node-line combination, the shortest path between every origin and destination node can now be computed. In order to reduce calculation time, only for node pairs having a demand of more than zero between them, this computation is executed. To illustrate the difference in computation time, without the removal of non-demand OD pairs there would be 344569 pairs for the Amsterdam PTN while removing these OD pairs leads to a total of 24583. The assignment of the shortest path is done based on an All or Nothing (AON) principle where in-vehicle time, waiting time and transfer time are all valued equally. This could be altered in future research to add weights to transfers or to stochastically/probabilistically assign the flow to routes based on stochastic or probabilistic preference attributes. For each node pair with a demand, the shortest path can be used to determine where the transfers take place by analyzing the nodes with a change in the subscript (e.g. x_1 to x_3). The demand of the node pair can then be assigned as transferring at node x. Summing the transfer demand for each shortest path for every node leads to a total number of transfers in the system and a total number of transfers for each node. By dividing the log of the transfers for each node by the log of the highest number of transfers, the transfer share can now be determined. With all the elements of the formula now determinable, the hierarchical degree for each node can be calculated.

4.2.4 Verification of demand and transfers

In order to verify the coding for the shortest path works as intended, the shortest paths are compared to those advised in a real PTN [3]. Table 4.1 shows the results for the verification of ten randomly generated node pairs in the Amsterdam PTN which is further discussed in the subsequent section. In this table, nine out of ten routes are identical while the last route has the same transfer but to a different line. As the transfer location is the only output from which the hierarchical degree is derived, the results would not change from transferring at the same location but to a different line. Therefore, the output for the hierarchical degree from the coding and the 9292 travel planner are considered exactly the same.

[3] Using https://9292.nl/ to determine the fastest route between an origin and destination for off-peak trips on 11-12-2019.

Table 4.1: Verification of shortest path algorithm for ten randomly generated node pairs

Origin ID	Destination ID	Origin label	Destination label	Line(s) in shortest path	Transfer location(s)	9292 advice
9094	2129	Ceintuurbaan	Hugo de Grootplein	3	-	Identical
7098	432	Olympiaplein	Diemerbrug	15 - 908 - 44	Station Zuid Station Diemen-Zuid	Identical
7324	8281	Rijksmuseum	Vennepluimstraat	2 - 26	Centraal Station	Identical
4406	3026	Inaristraat	Jan van Galenstraat(W)	1 - 50 - 7	Station Lelylaan Burg.de Vlugtlaan	Identical
2259	8281	Contactweg	Vennepluimstraat	22 - 26	Centraal Station	Identical
9026	196	Amstelstation	Station Ganzenhoef	53	-	Identical
9849	8222	Paasheuvelweg	Peter Martensstraat	120 - 66	Station Bijlmer ArenA	Identical
7108	515	Olympiaweg	Station Duivendrecht	15 - 50	Station Zuid	Same transfer, different line after
9810	1058	Sportlaan	Purmerplein	348 - 52 - 35	Station Zuid Noorderpark	Identical
9805	9807	Kalfjeslaan	Kruiskerk	347	-	Identical

4.3. Overview of case-study Amsterdam PTN

The first case-study elaborated in this study is the PTN of the Amsterdam including the municipalities of Amsterdam, Amstelveen, Diemen, Ouder-Amstel, the suburb region Badhoevedorp and the intersection at the north side of Schiphol Airport. Summarizing, the city of Amsterdam and its directly attached suburb areas are included for the PTN. Consequently, some PT lines, in particular regional buses and trains, cross the border and leave the PTN at a certain point. To realistically determine the hierarchical degree of these border nodes, one node beyond the boundary of the system is included. Furthermore, neighboring towns and cities with direct links are implemented as external nodes to facilitate demand from outside the PTN. These external nodes do not have a hierarchical degree but do have demand to and from other nodes. The defined PTN is shown in figure 4.3[4] and contains five different PT operators which are GVB, NS, Connexxion, Syntus/Keolis and EBS[5]. The latter three of these operate regional buses while the NS operates trains. Any PT in the city itself is operated by GVB. For this study, only PT lines with a regular timetable are included which excludes night lines and peak-hour lines. Furthermore, ferries and international trains and buses (such as Eurostar, Thalys and FlixBus) are considered beyond the scope of this study.

Nevertheless, 107 PT lines are considered including 14 tram-lines, 5 metro-lines, 20 local bus-lines, 48 regional bus-lines and 21 train-lines. Regarding the stops, 587 nodes are considered out of which 194 have a $k_i > 2$ and 172 are internal nodes and therefore considered for hierarchical degree. The other nodes are required to determine the hierarchical degree for these nodes but can be given a hierarchical degree of zero in advance. The line and node data are imported from a NDOV dataset which is elaborated in detail in Appendix A. Furthermore, this Appendix contains a list of lines included in the network and which lines are partially included and at which stop these are cut off.

[4] Adapted from: `https://www.gvb.nl/sites/default/files/lijnenkaart2020.pdf`

[5] A more elaborate overview, of which operator operates which lines, can be found on: https://wiki.ovinnederland.nl/wiki/Openbaar_vervoer_in_Amsterdam

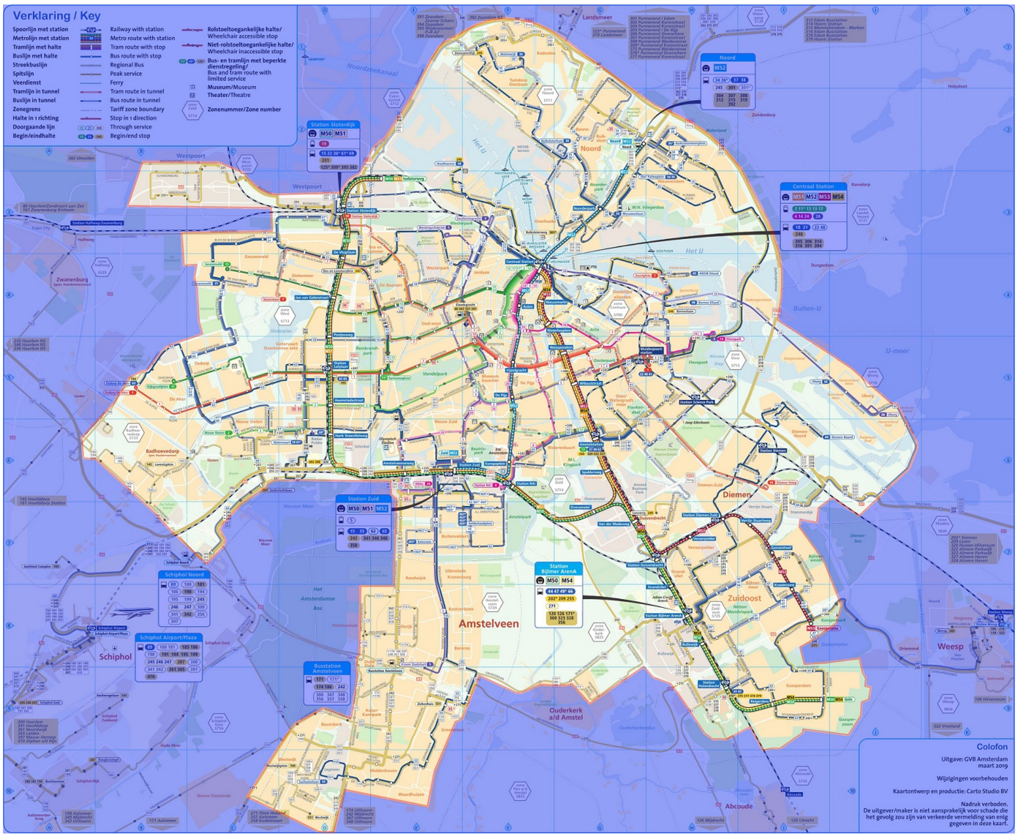


Figure 4.3: Boundary of the selection of the Amsterdam PTN

4.3.1 Scenario analysis

In order to analyze how different states of the network affect the hierarchical degree, scenarios are studied. These scenarios can be either past states of the network based on implemented changes or forthcoming states based on future plans. By analyzing how the hierarchical degree changes for affected nodes and the network as a whole, some conclusions can be drawn about the impact of the changes. For these scenarios, the contemporary state is used as a base year. This implies that for the scenarios changes are made to the network structure and the output is compared to that of the base year. For the past state, the network before the implementation of the North-South line (NZL) is used. This previous state is not merely the same network without the NZL but also includes different routes for trams and buses. Therefore, multiple changes in the network have to be accounted for. The differences in the network are elaborated in Appendix B.

For the future states, there are a couple of interesting options for scenarios which are relatively straightforward to test as these only include the addition or extension of one line. By analyzing the addition of one line at a time, the causal benefits for the network structure could be identified. The future lines considered are[6]:

- The extension of metro line 50/51 from Isolatorweg to Noorderpark
- The extension of metro line 50/51 from Isolatorweg to Centraal Station
- The addition of an east-west line from Schiphol to Lelylaan to Zeeburgereiland

[6]these future scenarios are based on a report of the municipality of Amsterdam retrieved from: `http://noordzuidlijn.wijnemenjemee.nl/images/2017/07/II__Uitbreiding_metronet_Amsterdam_-_aanvullende_verkenningen_2017.pdf`

- The addition of a west line from Schiphol to Lelylaan and Centraal Station

Another future scenario, extending the NZL to Schiphol is left out of this analysis. This scenario is very likely to be realized (e.g. de Volkskrant (2019)) where the NZL would connect to Schiphol and Hoofddorp. However, this additional line would barely affect the network structure as there already is an existing train connection between Amsterdam Zuid and Schiphol and a metro line from Amsterdam Zuid to Amstelveenseweg. Therefore, this scenario would not necessarily be redundant, but its impact on the hierarchical structure would be minimal. Consequently, analyzing this specific scenario is disregarded for this study. A map of the future scenarios and the exact route for the scenarios are elaborated in Appendix B. For all of these scenarios, the same demand model is used as for the contemporary situation.

4.4. Overview of case-study Rotterdam PTN

The second case-study elaborates on another Dutch city as a means of comparison to the network of Amsterdam. For this case-study there are no additional scenarios as this case-study is mainly done to provide possibilities for cross-network comparison of hierarchy. The network of Rotterdam includes the city of Rotterdam as well as directly attached neighboring towns and cities, similarly to the Amsterdam PTN. The network includes the municipalities of Rotterdam, Schiedam, Vlaardingen, Barendrecht, Capelle aan de IJssel & Krimpen aan de IJssel. The neighboring towns of Ridderkerk and Spijkenisse are excluded as these are separated from the city by a natural barrier (similar to Weesp in the Amsterdam PTN). The exclusion of these two sub-urbs makes sense on the one hand as the same methodology is applied to the Amsterdam PTN but, in particular for Spijkenisse, this comes with some limitations as there is a considerable transfer point located for regional buses to the metro. The limitation is an inevitable result of comparing two networks with a different network structure. An overview of the selection for the Rotterdam PTN is shown in figure 4.4[7].

For any line crossing the boundary of the PTN, external nodes are used in the same manner as the Amsterdam PTN. For the regional metro towards Hoek van Holland and the Hague, external nodes are used for each of the urban centers it passes. For the Rotterdam PTN the most important PT operator is RET which operates the metros, trams and most buses. NS operates the trains and the regional buses which are not operated by RET are operated by Connexxion, Arriva, Qbuzz and EBS. In line with the Amsterdam PTN, only PT lines with a regular timetable are included which excludes night, peak-hour and student lines. Furthermore, ferries and international trains and buses (such as Eurostar, Thalys and FlixBus) are considered beyond the scope of this study.

For the Rotterdam PTN, 638 nodes are considered out of which 147 have a $k_i > 2$ and 123 are internal nodes. Comparing this to Amsterdam, there are slightly more nodes in total but less nodes with a $k_i > 2$. The Rotterdam PTN has 79 different lines. These lines consist out of 8 tram-lines, 5 metro-lines, 34 local bus lines, 19 regional buses and 13 train-lines. Comparing this to the Amsterdam PTN, the main difference is between local buses and regional buses where Amsterdam has significantly less local buses but more regional buses. The line and node data are once more imported using a NDOV data-set which is elaborated in detail in Appendix A. Furthermore, this Appendix contains a list of lines included in the network and also which lines are partially included and at which stop these are cut off.

[7]Adapted from: `https://www.ret.nl/fileadmin/user_upload/Documenten/PDF/Kaarten_en_plattegronden/RET_lijnennetkaart_2020.pdf`

Figure 4.4: Boundary of the selection of the Rotterdam PTN

4.5. Data

Throughout this section, the different data (sources) for the case studies are elaborated. Based on the different data sources, some potential issues and limitations can be identified in advance. However, in order to clearly indicate which data are applied, first the data compilation is briefly elaborated after which it is further clarified for topological and demand data as well as the different cases separately.

4.5.1 Data compilation

For the computation of the metric, topological data and demand data are required. While the data source for the topological data is the same for both cases, the demand data source is region specific. For the topological data, NDOV open source data-sets are used in combination with a GTFS Ovapi data-set for coordinate data. These sources provide the required topological data for each operator in the Netherlands. NDOV is basically an institute stimulating the use of open source data for Dutch open PT data with different types of information. This information varies from stop accessibility, fares and on-demand travel information to network structure, line information, direct link information and stop codings. GTFS is an international data specification which allows PT agencies to publish data. As both of these data-sets use the same method for coding nodes, these can easily be connected to link the coordinate data to the nodes retrieved the NDOV data-set. The NDOV data-sets for Amsterdam[8] and Rotterdam[9] and GTFS Ovapi data-set for both networks [10] can be found using the URL links provided.

From the NDOV data-sets, node and line data are retrieved. The line data also include the links

[8]Data-set *KV1_GVB_1629_1.zip* retrieved from: `http://data.ndovloket.nl/gvb/`
[9]Data-set *retKV01_v813_20191230-0434.zip* retrieved from: `http://data.ndovloket.nl/ret/`
[10]Data-set *gtfs-nl.zip* retrieved `from:http://gtfs.ovapi.nl/nl/`

on the route for that particular line. For the node data, every unique stop (with possibly multiple modalities) is considered as one node. By defining nodes in this way, both nodes with only one stop for one line and nodes with stops for many lines and many different modalities are considered as one node. Furthermore, every link regardless of the modality is perceived equal and only one link between every node pair can exist[11]. This definition of the network is in line with the $\mathbb{L}$-space representation and can also be transformed to a $\mathbb{P}$-space as the lines serving each node are known.

For the line data, duplicate lines and lines outside the scope of this study are removed or cut-off. Further details on the cleaning of these data-sets are shown in Appendix A. For the demand data, VENOM and V-MRDH data-sets are used for the Amsterdam and Rotterdam PTN respectively. The data-sets that are used, contain OD matrices for the nodes selected for the network. These nodes are linked to the nodes from the topological data for which the processing is also further elaborated in Appendix A. Throughout the following paragraphs, the data-sets are further elaborated for the specific cases. Every part of coding and each data-set required for the results can be found on the following link: `https://github.com/Abel287/MasterThesis.git`.

4.5.2 Topological data

As stated above, for the topological data, the sources for both the Amsterdam and the Rotterdam PTN are the same from the same institute. However, as these are still different data-sets they are briefly explained separately.

Amsterdam

From the data-sets mentioned earlier, the number of nodes, link and lines are imported. In table 4.2 an overview of the total, internal and external nodes, links and lines is given for the Amsterdam PTN.

Table 4.2: Overview of the nodes, lines and link in the Amsterdam PTN

	Total	**Internal**	**External**
Nodes	587	550	37
Lines (bidirectional)	107	93	14
Links (unique)	793	713	80

For the scenarios, the previously described data are used as a basis where adjustments are made to. For the future scenarios these adjustments are only related to the addition of the line for that specific scenario which makes the changes to the network relatively easily comprehensible. On the other hand, for the past state before the NZL, a different schedule was used in which many trams and buses used different routes. Consequently, besides the obvious adjustment of removing the NZL, the changes to other trams and buses have to be added manually based on the schedule before the NZL opened[12]. Since the changes in the network are not only related to the NZL, the outcomes are not only a result of the opening of the NZL but also to other changes in the network. Furthermore, the stops Station Noord and Noorderpark were opened with the NZL and are not present in the earlier state. On the other hand due to rerouting of trams and buses, some other stops are no longer in use.

[11] This is not the case for the graph determining the transfer demand as explained earlier

[12] This is the schedule until 21-07-2018 and the changes are retrieved from: `https://wiki.ovinnederland.nl/wiki/Dienstregeling_concessie_Stadsvervoer_Amsterdam_2018`

Rotterdam

An overview of the number of internal and external nodes, lines and links for the Rotterdam PTN is shown in table 4.3.

Table 4.3: Overview of the nodes, lines and link in the Rotterdam PTN

	Total	**Internal**	**External**
Nodes	638	591	47
Lines (bidirectional)	79	65	14
Links (unique)	776	696	80

4.5.3 Demand data

VENOM model

The demand data-set for the Amsterdam PTN is retrieved from an institute called VENOM.The VENOM data-set is a regional traffic model for the metropolitan area of Amsterdam and can be applied for calculations for projects and policies regarding infrastructure and mobility. While VENOM is focused on any kind of modality (including pedestrians, bikes and cars) it can also be used for PT only as is required for this study. In order to connect the demand data to the topological data, the data in these different data-sets have to be aligned as these are not coded similarly. The VENOM model used, has the base year 2030 which is not reflecting the current situation. However, the alternative base year would have been before the opening of the NZL (2016) leading to a demand for a different network structure. For the scenarios, the same demand data is used in order to compare between the network states.

First, the data-sets are aligned in a manner that every considered node from the (VENOM) is directly linked to a node in the topological data-set which overcomes the different way the data-sets are built. Then, in the VENOM model trips are generated by the attraction and production for zones in the network which are calibrated after. Coupling these zones to PT nodes, the demand in the network can be estimated based on the calibrated model. Thereafter, for every OD pair, the daily demand (aggregated morning/evening peak and rest of the day) is derived from the VENOM model leading to a value for the average number of passengers traveling between nodes daily. With the demand for every OD pair in the network, the coding for the transfers can be applied to the Amsterdam PTN leading to the output for the network. In table 4.4 the estimates for mode speed and frequencies for the Amsterdam PTN are presented which are required to determine the shortest path. The estimates for train[13], metro[14], tram[15], local bus[16] and regional bus[17] are based on different sources as no sources covered all modes. However, despite possibly creating a bias by using different sources, this is perceived acceptable as the speed of the vehicles has only minor influence.

[13] Based on estimate from: https://www.clo.nl/indicatoren/nl214003-aanbod-van-openbaar-vervoer
[14] Based on estimate from: https://madeby.tfl.gov.uk/2019/07/29/tube-trivia-and-facts/
[15] Based on estimate from: https://api1.ibabs.eu/publicdownload.aspx?site=vervoerregio&id=100029836
[16] Based on estimate from: https://www.researchgate.net/publication/275020231_Measurement_of_the_average_speed_of_city_buses_and_the_possibilities_of_increasing_it
[17] Based on estimate from: http://www.metrolinx.com/thebigmove/Docs/big_move/RTP_Backgrounder_Transit_Technologies.pdf

Table 4.4: Characteristics of the lines for the different modalities it serves

Mode	Speed (km/h)	Frequency (veh/h)
Train	60	4
Metro	35	10
Tram	15	8
Local bus	25	6
Regional bus	30	4

V-MRDH model

For the Rotterdam PTN, a V-MRDH model is used as source for the demand data. Similarly to the VENOM model, V-MRDH is a regional traffic model for the metropolitan areas of Rotterdam and The Hague. Furthermore, this model includes different modalities too which are left out for this analysis. The applied base year for this model is 2030, which is the same as the VENOM model for Amsterdam which increases the comparability of the networks.

The aligning of the data-sets is done in a similar way to the VENOM data-set by linking the nodes from the V-MRDH data-set to nodes in the topological data-set. The generations of trips is done in a similar way too, by computing the attraction and production for zones in the network which are calibrated after and coupled to the nodes in the V-MRDH data-set. The data are aggregated for daily data (by combining morning/evening peak and rest of the day) but also for access and egress modes (walk-bike, walk-walk, bike-walk, bike-bike) as these are determined separately in the V-MRDH model. Consequently, every OD pair is provided an average demand comparable to the method applied to the Amsterdam network. The mode characteristics for the Rotterdam PTN are the same as for the Amsterdam PTN as described in table 4.4. With all the data elaborated, throughout the following chapter the results are presented. The results for the different cases and scenarios are compared, and some societal implications of these results are elaborated.

5 Results

Throughout this chapter, the results of the case-studies are elaborated. First, the Amsterdam case-study is discussed followed by the scenarios for this case-study. For these scenarios, the results are compared to the results for the Amsterdam PTN in the base year. Thereafter, the Rotterdam case-study is elaborated and the results are compared to the results of the Amsterdam PTN to analyze how two different networks compare. Lastly, the practical applications for the results are outlined by analyzing how the output for the hierarchical degree can be used to prevent or solve bottlenecks and overcome issues such as vulnerability and cascading failures. While these are not direct results from the model, it is important to emphasize the opportunities for more general PTN issues. Related to this, the implications of this study with regards to developing policy, are elaborated. Therefore, these results should be interpreted as more general compared to the results for the case-studies.

5.1. Results of the case-study for the Amsterdam PTN

Throughout this section, the results of the Amsterdam case-study are elaborated. First the output for each element is discussed separately. Thereafter, the combined results are discussed more in depth including a correlation graph between the hierarchical degree and each element as well as the correlation between the elements.

5.1.1 Results for the different elements

Output for the topological influence in the Amsterdam PTN

For the first element, the topological influence of the nodes is determined. In table 5.1 the five highest scoring nodes are shown, indicating that Centraal Station, as would be expected, scores highest. Furthermore, two other train stations are among the five highest scoring nodes while also Leidseplein and Elandsgracht are up there. The latter two are considered as important tram and bus stops connected to a lot of influential nodes.

Table 5.1: Overview of the highest scoring nodes for the topological influence in the Amsterdam PTN

Rank	Node label	e^A value
1.	Centraal Station	1
2.	Leidseplein	0.815
3.	Station Sloterdijk	0.727
4.	Station Zuid	0.705
5.	Elandsgracht	0.686

In figure 5.1, the cumulative distribution function of the topological influence is shown. In this figure it can be noted that almost seventy percent of the nodes has no influence at all. This is most likely caused by the nodes having a degree of two or lower. Furthermore, around twenty percent of the nodes has a topological influence value between 0 and 0.35 and only ten percent has a higher value. Therefore, it can be stated that there are only a few influential nodes while the other nodes have only little influence or no influence at all.

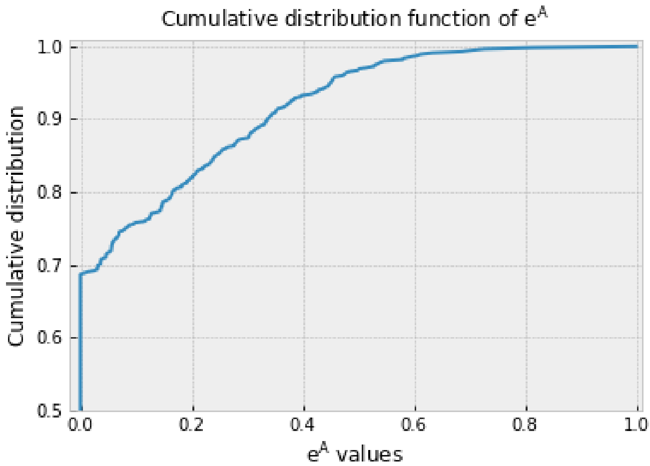


Figure 5.1: Cumulative distribution function of the topological influence for the Amsterdam PTN

Output for the redundancy in the Amsterdam PTN

For the (non-)redundancy, the clustering coefficient of the nodes is calculated. Table 5.2 shows the five highest scoring nodes in which it is striking that each of these are intercity train stations. While only internal nodes are considered for this element, the intercity stations appear to be nodes that have the least neighbors with mutual connections. Intuitively, this results makes sense as these intercity stations serve as hubs where transfers between local and regional buses to trains occur. Therefore, these stations are considered as local hubs to connect to the rest of the network.

Table 5.2: Overview of the highest scoring nodes for the (non-)redundancy in the Amsterdam PTN

Rank	Node label	e^B value
1.	Centraal Station	0.853
2.	Station Bijlmer ArenA	0.805
3.	Station Sloterdijk	0.802
4.	Station Zuid	0.770
5.	Amstelstation	0.756

In figure 5.2 the cumulative distribution function of the (non-)redundancy is shown in which nearly seventy percent appears to be redundant. However, opposed to the topological influence, there are relatively few nodes with a value between 0 and 0.4 as the highest share of non-zero scoring nodes (around twenty percent) are in the range between 0.4 and 0.6. This implies for most non-zero scoring nodes around half of the neighbors have a direct line between them and half require a transfer. However, this does not automatically imply that each of these transfers have to take place at the considered node. In retrospect it makes sense that the values for non-zero scoring nodes are not very close to zero. This can be explained by using an example node on an intersection (with degree 4) on which two lines cross. The $\mathbb{P}$-space for this node includes all nodes on both lines and while it makes sense that each of the nodes on one line are connected mutually, any node with a degree of two cannot be connected to nodes of the intersecting line. Only nodes with a degree of higher than two can be connected to nodes on the intersecting line by the means of a third line. In short, it could have been expected that a lot of the non-zero scoring nodes would have a value close to 0.5 if it is located on an intersection.

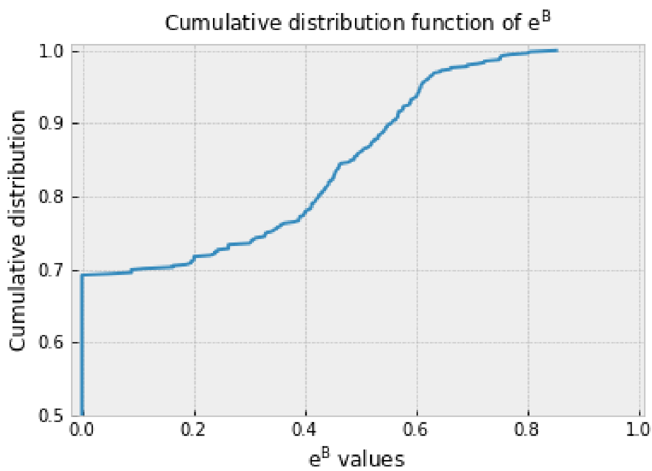


Figure 5.2: Cumulative distribution function of the (non-)redundancy for the Amsterdam PTN

Output for the transfer potential in the Amsterdam PTN

In table 5.3 the highest scoring nodes for the transfer potential are shown on the left side while it shows the highest scoring nodes for transfer share on the right side. It should be noted that four out of five nodes appear in both lists while Station Bijlmer ArenA is in the top five for the transfer potential score and Station Lelylaan is in the top five for the transfer share. This implies the number of transfer directions for Bijlmer ArenA is significantly higher than for Lelylaan. Furthermore, as Lelylaan is ranked second in terms of transfer share, the number of transfer directions has to be significantly lower than any of the other high ranking nodes to be dropped out of the top five for the final score of the transfer potential. Looking at the type of nodes in the list, it can be noted that comparable to the (non-)redundancy, all of these nodes are intercity train stations. As the passengers from external nodes through the system are considered, it makes sense a lot of transfers occur at intercity stations where the external nodes are connected to.

Table 5.3: Overview of the highest scoring nodes of the transfer potential and transfer share in the Amsterdam PTN

Rank	Node label	e^C value	Rank	Node label	log transfer share
1.	Centraal Station	0.867	1.	Centraal Station	1.000
2.	Station Sloterdijk	0.749	2.	Station Lelylaan	0.909
3.	Station Zuid	0.733	3.	Amstelstation	0.887
4.	Station Bijlmer ArenA	0.702	4.	Station Sloterdijk	0.885
5.	Amstelstation	0.690	5.	Station Zuid	0.879

In figure 5.3 the cumulative distribution function of the transfer potential is shown while figure 5.4 shows the cumulative distribution function for the transfer share. First, it should be noted that the distribution function of the transfer potential has a slightly higher share of zero-scoring nodes than the distribution function for transfers. As the value for the transfer potential is automatically assigned a value of zero for nodes with a degree of two, this means that these nodes have a degree of two and do facilitate transfers. The cause of this is that nodes which are a terminal station for two lines have a degree of two but are still able to facilitate transfers for passengers transferring between these lines. The transfer share for any given node should be higher than the value of the transfer potential as the value for transfer directions is in the range between zero and one with one being the theoretical case of infinite transfer directions. The highest share of non-zero scoring nodes for the transfer share is between 0.4 and 0.6 which is around ten percent. A transfer share of 0.5 for a node indicates the number of transfers is equal to the square root of the highest transfer share[1].

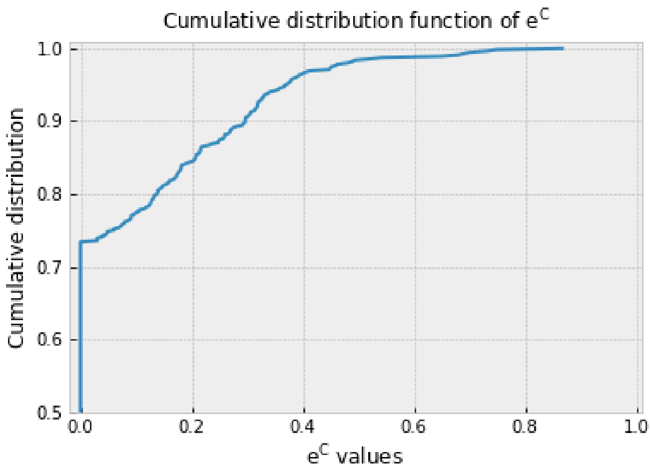


Figure 5.3: Cumulative distribution function of the transfer potential for the Amsterdam PTN

[1] Using basic algebra for solving the formula of the transfer potential

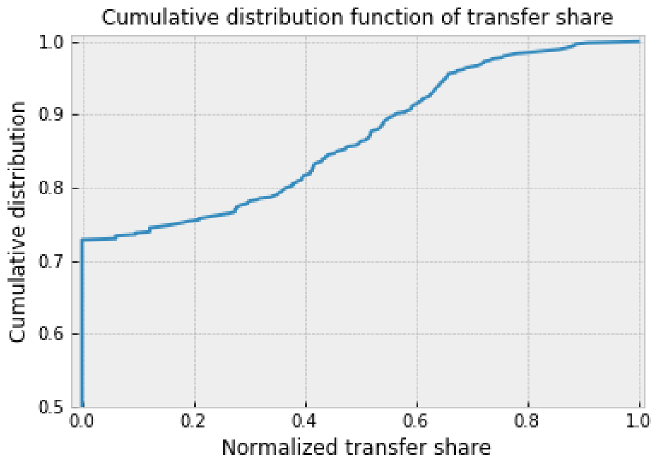


Figure 5.4: Cumulative distribution function of the transfer share for the Amsterdam PTN

5.1.2 Results for hierarchical degree

After looking at the elements separately, the values and ranking for the hierarchical degree can be determined. To provide some further insight than the previous tables, the ten highest ranking nodes are shown in table 5.4. In this table it is first of all clear that Centraal Station, considered the busiest station is right on top for the hierarchy. Furthermore, most of the other train stations with intercity train stops are in the top ten while only three non-train stations are in there (Amstelveenseweg, Leidseplein & Vijzelgracht). However, these other three nodes are considered as important transit point between regional buses and the metro and tram network which makes their presence in the list more or less expected. There are no real notable absentees besides some of the central nodes in the suburbs (e.g. Station Noord, Station Diemen & Amstelveen Busstation). One thing to note from the values for the different elements is that especially the value for the topological influence shows a high decline even for the highest ranked nodes. Furthermore, the non-train serving nodes appear to score relatively high on the topological influence while the train stations score high on the non-redundancy. This can be explained as the train stations are connecting the edges of the network to the center while the non-train nodes are connected in the center to many influential nodes but are also more redundant.

ll in all, this list of nodes shows some expected nodes but also nodes scoring high on different elements emphasizing the complementary characteristics of the different elements. In figure 5.5 the geographical location of the nodes with their hierarchical degree is visualized. In this figure, it is clearly visible the hierarchical nodes are spread out through the network with hierarchical nodes to every side of the network with the exclusion of the northern part of the city. In figure 5.6, an overview of the nodes in the network, divided into different sub-levels, is shown. In this overview, 13 nodes are allocated to level 1, 132 to level 2 and 405 to level 3. Furthermore, a more detailed overview of the highest scoring nodes and the key elements for these nodes is shown in Appendix C.

Table 5.4: Results for 10 nodes with the highest hierarchical degree (H_i) in the Amsterdam PTN

Rank	Node	H_i	e^A	e^B	e^C
1.	Centraal Station	0.739	1.000	0.853	0.867
2.	Station Sloterdijk	0.437	0.727	0.802	0.749
3.	Station Zuid	0.398	0.705	0.770	0.733
4.	Amstelstation	0.309	0.593	0.756	0.690
5.	Station Lelylaan	0.299	0.612	0.751	0.649
6.	Leidseplein	0.283	0.815	0.752	0.462
7.	Muiderpoortstation	0.227	0.580	0.752	0.521
8.	Vijzelgracht	0.183	0.606	0.628	0.481
9.	Station Bijlmer ArenA	0.176	0.312	0.805	0.702
10.	Amstelveenseweg	0.157	0.449	0.646	0.542

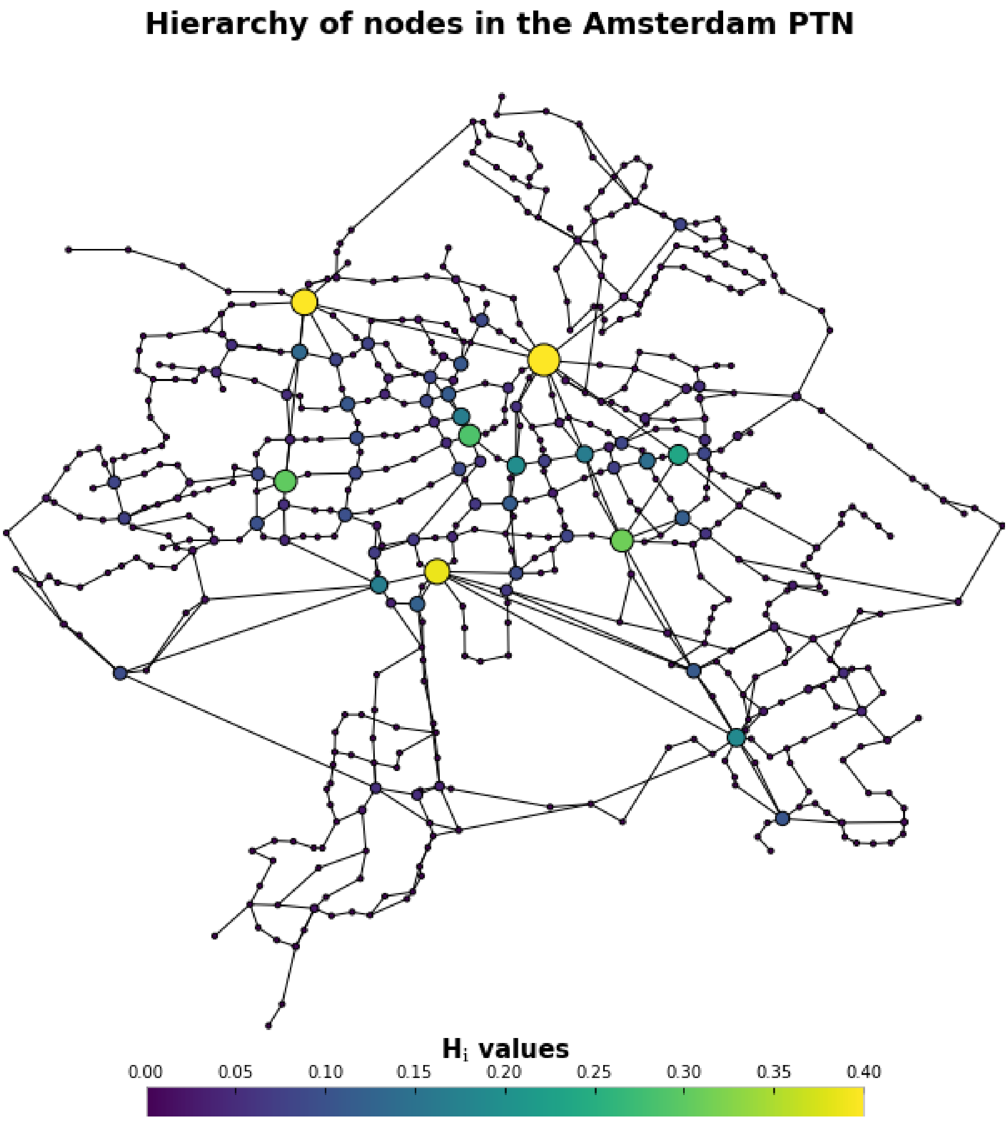


Figure 5.5: Map of the spatial distribution of the hierarchy of nodes in the Amsterdam PTN

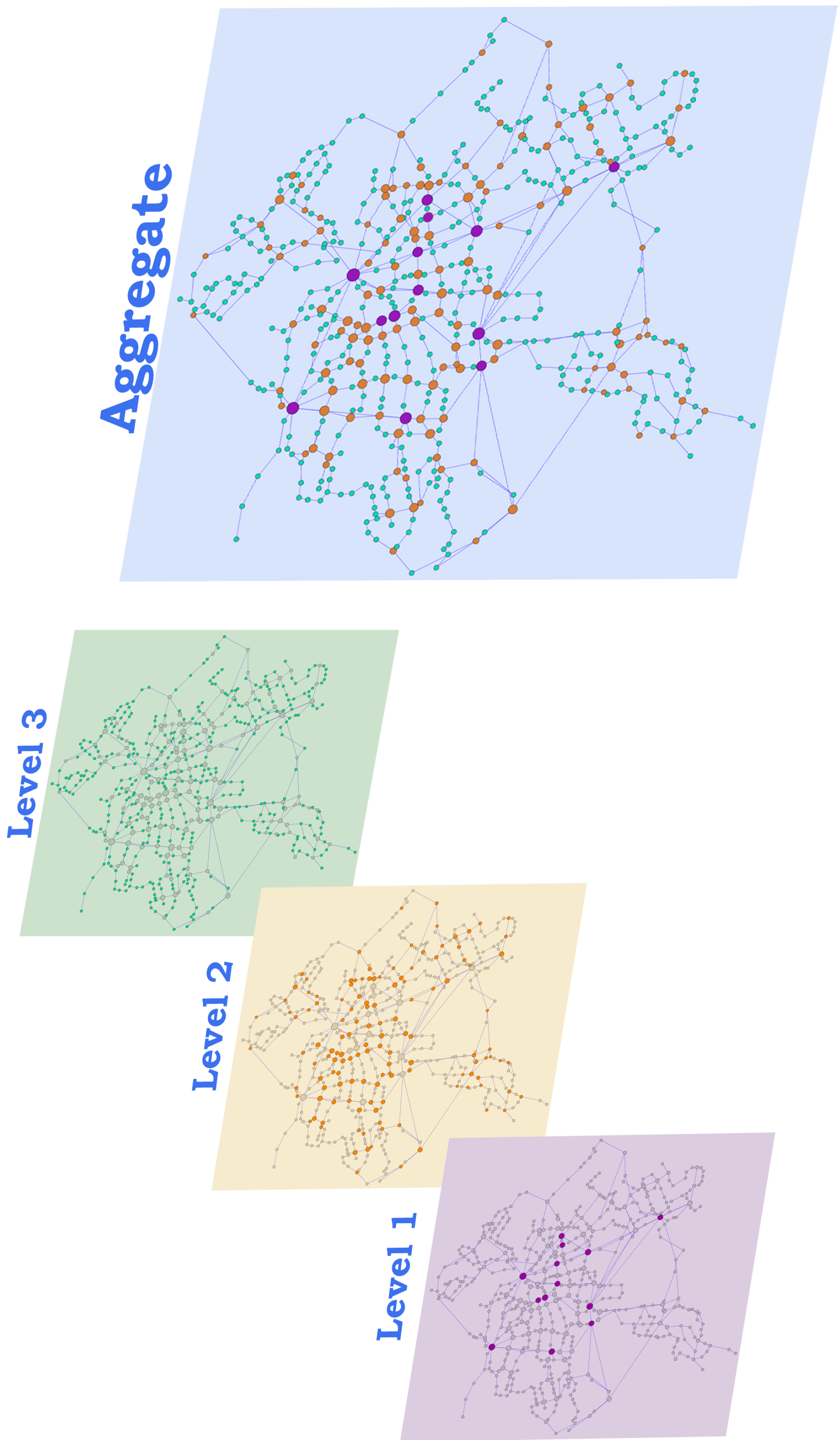


Figure 5.6: Overview of nodes in the Amsterdam PTN divided into three hierarchical levels

5.1.3 Correlation analysis of hierarchical degree and elements in the Amsterdam PTN

For the different elements in the Amsterdam PTN, a correlation analysis is done to clarify how the different elements and the hierarchical degree are correlated. Figure 5.7 visualizes the correlation between each element in graphs including a regression line. It is notable that the transfer potential (e^C) correlates most with the hierarchical degree closely followed by the topological influence (e^A) while the (non-)redundancy (e^B) correlates least with the hierarchical degree. Among the elements, the topological influence and (non-)redundancy have the highest correlation coefficient while the topological coefficient and transfer potential are correlated the least. Therefore, it is remarkable that the non-redundancy correlates least with the hierarchical degree but correlates most with the other elements. There is no clear explanation that can clarify this.

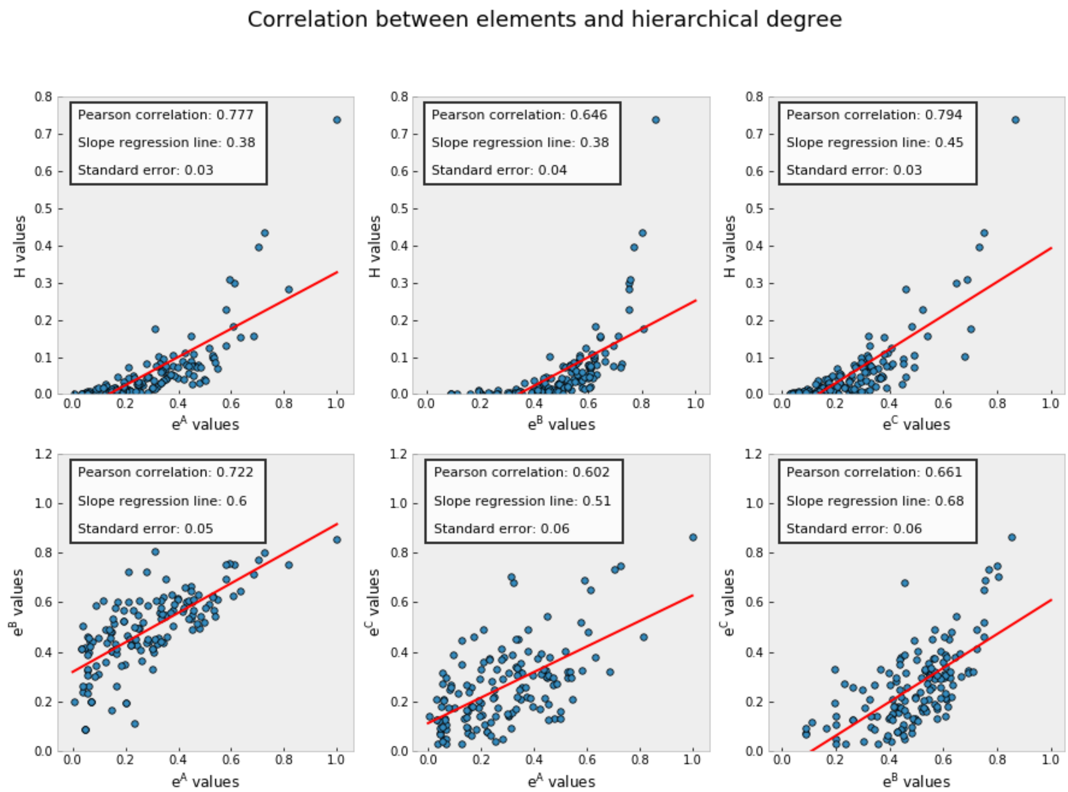


Figure 5.7: Correlation between hierarchical degree and elements for the Amsterdam PTN

5.1.4 Results for network hierarchy

Based on the program code for estimating the power-law parameters, developed by Alstott, Bullmore, and Plenz (2014), the hierarchical coefficient of the Amsterdam PTN can be determined. In figure 5.8 and figure 5.9 the power-law coefficient and Gini-coefficient are shown respectively. With the alpha coefficient of -1.643 and a Gini-coefficient of 0.902, the hierarchical degree of the network can be determined with

$$H_I = 0.902^{\frac{1}{1.643}} = 0.939, \tag{5.1}$$

indicating the hierarchy in the Amsterdam PTN is quite close to one. Hence, the structure of the network can be considered as hierarchical with a high spreading of hierarchical degree. In order to compare this result, different states of the network and different networks have to be evaluated.

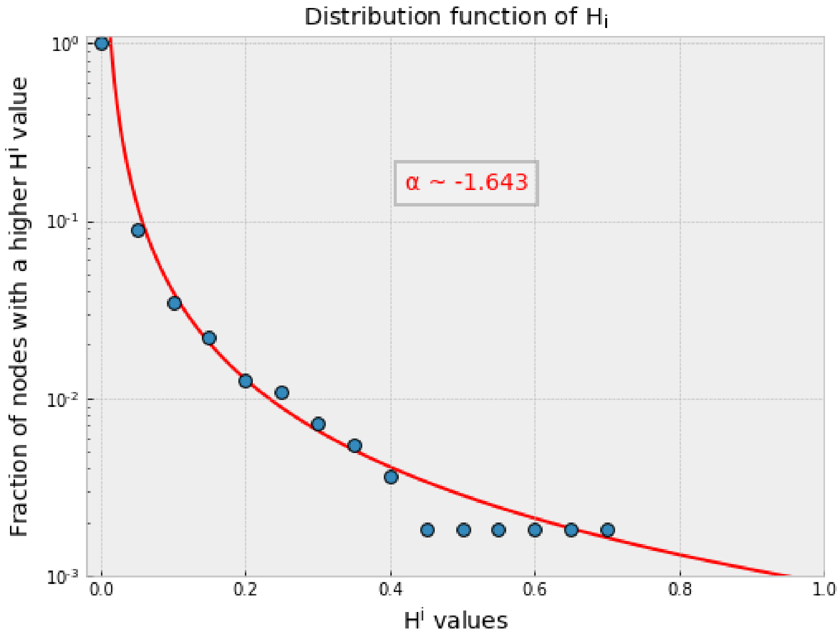


Figure 5.8: Distribution function of hierarchical degree in the Amsterdam PTN

5.1.5 Comparing to other indicators

In table 5.5 the correlation between the hierarchical degree and previously introduced indicators is shown for nodes with a hierarchical degree of higher than zero. From this table it can be interpreted that the betweenness centrality in the ℙ-space correlates most with the hierarchical degree. Thereafter, the overlapping degree correlates most with the hierarchical degree closely followed by the betweenness in the 𝕃-space and the degree. The assortativity correlates the least. The clustering coefficient has a negative correlation which is the negative value of the correlation with the non-redundancy. This makes sense as the (non-)redundancy is basically one minus the clustering coefficient. Following from the significance values, based on the hypotheses that the hierarchical degree is not correlated to any of the other indicators, it can be assumed all of the indicators are correlated to hierarchical degree. It is noteworthy the betweenness centrality and overlapping degree correlate most while these are not included in the metric. These indicators even have a higher correlation coefficient than the indicators that are included. Therefore, these two indicators explain hierarchy, as perceived by this study, best. Following, the results for the scenario study are elaborated.

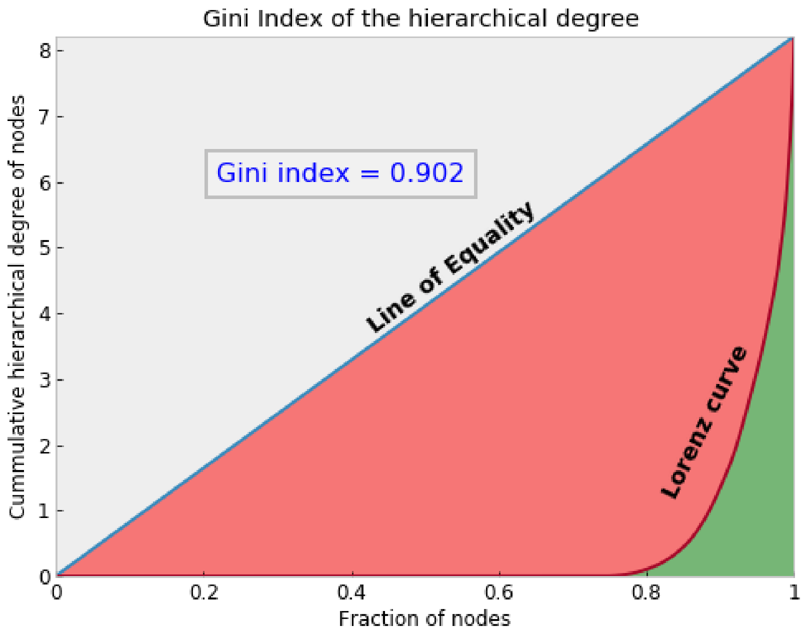


Figure 5.9: Gini-index for hierarchical degree in the Amsterdam PTN

Table 5.5: Correlation of other indicators with the hierarchical degree for the Amsterdam PTN

Indicator	Pearson correlation coefficient	Significance (based on N=141)
Degree (𝕃-space)	0.787	<0.001
Assortativity (𝕃-space)	0.285	<0.001
Clustering coefficient (ℙ-space)	-0.646	<0.001
Betweenness centrality (𝕃-space)	0.791	<0.001
Betweenness centrality (ℙ-space)	0.840	<0.001
Eigenvector centrality (ℙ-space)	0.777	<0.001
Overlapping degree	0.810	<0.001

5.2. Scenario study for the Amsterdam PTN

For the different scenarios, the output is evaluated and compared to the base year. For each scenario, the highest scoring nodes are evaluated for each element and the hierarchical degree. Furthermore, some notable increases or decreases in the score for any node are highlighted. Thereafter, the results for the network hierarchy coefficients (power-law coefficient and Gini-index) are compared to determine which state should be considered more hierarchical. To provide some more practical output, the average travel times of both network states is compared. It should be noted, however, that the removal or addition of new nodes could also potentially lead to changes in average travel time and conclusions should be drawn carefully.

5.2.1 Scenario before NZL

Comparison of elements results

For the past scenario, before the opening of the NZL, the results for each of the elements are compared to the base year. For the topological influence, table 5.6 provides the results for the scenario and the base year. It is striking that except for Centraal Station, none of the stops in the five highest scoring nodes are the same. Beneath the five highest scoring nodes, the rank of the five highest scoring nodes for the other situation are shown. It is notable that the topological influence values for the nodes besides Centraal Station are much higher for the base year, indicating the influence of nodes is more spread out in the contemporary situation. As the eigenvector centrality value is normalized, the value for each node indicates its share of the highest scoring node. This leads to an indication that the values for nodes in the base year are closer to the value of Centraal Station than before the opening of the NZL.

Furthermore, for the highest scoring nodes only the nodes Muiderpoortstation and Molukkenstraat score much lower while in the base year while the other nodes show either a small decrease of less than two hundredths of an increase. Looking at the individual nodes, most of the nodes decreasing in rank and value (Linnaeusstraat, Muiderpoortstation & Molukkenstraat) are located on the previous route of tram-line 7[2]. It is hard to explain why these nodes would score much higher with the original path of line 7 but a plausible explanation would be that the combination of tram-lines 7 and 14 together with bus-line 22[3] would connect these nodes to most influential nodes while the replacement line, line 3 does not. Another plausible and maybe complementary explanation is that the rerouting of other lines and addition of the NZL, increased the influence of other nodes to which these three are not connected.

For the nodes showing a significant increase in rank after the opening of NZL, it is notable that three of these now are directly connected by tram-line 19 (Station Sloterdijk, Leidseplein & Elandsgracht) while Station Zuid got additional connections as the terminal station of the NZL. The value for Leidseplein in particular increased significantly which could be explained by the node being served by two additional tram-lines.

Table 5.6: Highest topological influence values for the scenario before the NZL and the base year in the Amsterdam PTN

Rank	**Node**	**e^A value before NZL**	**Rank**	**Node**	**e^A value base year**
1.	Centraal Station	1.000	1.	Centraal Station	1.000
2.	Muiderpoortstation	0.636	2.	Leidseplein	0.815
3.	Amstelstation	0.605	3.	Station Sloterdijk	0.727
4.	Molukkenstraat	0.603	4.	Station Zuid	0.705
5.	Linnaeusstraat	0.599	5.	Elandsgracht	0.686
6.	Station Sloterdijk	0.566	9.	Amstelstation	0.593
8.	Leidseplein	0.517	10.	Linnaeusstraat	0.582
13.	Station Zuid	0.493	11.	Muiderpoortstation	0.580
23.	Elandsgracht	0.430	49.	Molukkenstraat	0.353

The results for the (non-)redundancy are shown in table 5.2 indicating the rank of the five highest

[2] This is shown in the map in Appendix B
[3] Or tram-line 9 in the case of the node Linnaeusstraat

scoring nodes remains the same. The only changes that are really worth explaining are the increasing value for Station Zuid as a result of new connections from the NZL and the decreasing value of Centraal Station which is very likely caused by some of the regional buses and buses to northern Amsterdam going to Station Noord in the base year instead.

Table 5.7: Highest (non-)redundancy values for the scenario before the NZL and the base year in the Amsterdam PTN

Rank	**Node**	**e^B value before NZL**	**Rank**	**Node**	**e^B value base year**
1.	Centraal Station	0.882	1.	Centraal Station	0.853
2.	Station Bijlmer ArenA	0.804	2.	Station Bijlmer ArenA	0.805
3.	Station Sloterdijk	0.798	3.	Station Sloterdijk	0.802
4.	Station Zuid	0.760	4.	Station Zuid	0.770
5.	Amstelstation	0.758	5.	Amstelstation	0.756

The results for the transfer potential are shown in table 5.3 where the only change in rank between the past scenario and the base year is the increase of Station Zuid. This increase makes sense as the opening of the NZL offers new transfer opportunities at Station Zuid. Looking at the values, Station Zuid and Centraal Station are the only increasing nodes while the other nodes show some decrease. The increasing value of Centraal Station is likely caused by more transfer directions and for Station Zuid a higher share of transfer passengers[4]. The slightly decreasing value for the other nodes may be an indication the number of transfers at Centraal Station increased and the relative share of the other nodes is therefore lower. Table 5.9 shows the normalized transfer share for the highest scoring nodes. This table aligns well with table 5.3 as both indicate an increase for Station Zuid and a relative decrease for other nodes suggesting an increase of Centraal Station as well. As both Centraal Station and Station Zuid are served by the NZL, this is an indication of the transfer potential of the new line.

Table 5.8: Highest transfer potential values for the scenario before the NZL and the base year in the Amsterdam PTN

Rank	**Node**	**e^C value before NZL**	**Rank**	**Node**	**e^C value base year**
1.	Centraal Station	0.846	1.	Centraal Station	0.867
2.	Station Sloterdijk	0.762	2.	Station Sloterdijk	0.749
3.	Station Bijlmer ArenA	0.720	3.	Station Zuid	0.733
4.	Amstelstation	0.713	4.	Station Bijlmer ArenA	0.702
5.	Station Zuid	0.704	5.	Amstelstation	0.69

[4]The transfer share for Centraal Station is highest in both situations and as the value is normalized this would not change the value for e^C

Table 5.9: Highest transfer share values for the scenario before the NZL and the base year in the Amsterdam PTN

Rank	Node	Transfer share before NZL	Rank	Node	Transfer share base year
1.	Centraal Station	1.000	1.	Centraal Station	1.000
2.	Station Lelylaan	0.927	2.	Station Lelylaan	0.909
3.	Amstelstation	0.916	3.	Amstelstation	0.887
4.	Station Sloterdijk	0.901	4.	Station Sloterdijk	0.885
5.	Muiderpoortstation	0.895	5.	Station Zuid	0.879
7.	Station Zuid	0.860	6.	Muiderpoortstation	0.868

Comparison of hierarchical degree results

Combining all of the elements into the hierarchical degree, the difference in output between the pre-NZL scenario and the base year can be done. Table 5.10 shows the ten highest ranking nodes in both cases. The first thing to note is that the H_i scores are generally higher in the base year and that the five highest scoring nodes remain almost the same with the exception Muiderpoortstation dropping a few ranks. As the values for the hierarchical degree are more equally distributed in the base year, this indicates the hierarchy is more spread out implying more nodes become important. Looking back at the results for the separate elements, mainly the increase in topological influence has led to a more evenly spread of hierarchy among the high scoring nodes. Taking a look at individual nodes, the increase of Vijzelgracht (from rank 38 with a H_i value of 0.064 to 8^{th} with a H_i value of 0.183) is notable. This change is very likely a result of the opening of the NZL which serves the node Vijzelgracht. Another notable increase in value is for the node De Pijp which increased from rank 59 to rank 17. Station Noord and Noorderpark, which are newly added nodes after the opening of the NZL do not appear to be scoring high in terms of their hierarchical degree.

Table 5.10: Highest hierarchical degree values for the scenario before the NZL and the base year

Rank	Node	H_i value before NZL	Rank	Node	H_i value base year
1.	Centraal Station	0.747	1.	Centraal Station	0.739
2.	Station Sloterdijk	0.344	2.	Station Sloterdijk	0.437
3.	Amstelstation	0.327	3.	Station Zuid	0.398
4.	Station Zuid	0.264	4.	Amstelstation	0.309
5.	Muiderpoortstation	0.256	5.	Station Lelylaan	0.299
6.	Station Lelylaan	0.209	6.	Leidseplein	0.283
7.	Leidseplein	0.176	7.	Muiderpoortstation	0.227
8.	Linnaeusstraat	0.165	8.	Vijzelgracht	0.183
9.	Burg.de Vlugtlaan	0.137	9.	Station Bijlmer ArenA	0.176
10.	Station Bijlmer ArenA	0.130	10.	Amstelveenseweg	0.157

The power-law coefficient and Gini-coefficient for the network pre-NZL are -1.588 and 0.891 respectively which are a little lower than in the base year. The hierarchical coefficient can now be calculated and is 0.930 for the pre-NZL network, about a hundredth lower than for the base year network. Relating this to the previous findings, this is likely caused by the nodes around rank two to ten becoming more hierarchical rather than Centraal Station. The nodes which are called 'second order nodes', which score highest right after the main hub (Centraal Station), could therefore be seen as more important in the contemporary network compared to the situation before the NZL. The slight increase in hierarchy of the network is therefore interpreted as a positive improvement towards the development of a network with a clear multi-level structure. This change in network structure could be seen as positive output for the policy regarding the development of the NZL[5]. Hence, in aftermath the changes to the network structure can be evaluated as feedback on the decision making for the policy. The policy regarding the network structure for the NZL has been focused on adjusting the network structure from a spiderweb structure towards a herringbone structure to relieve Centraal Station[6]. Despite the node Centraal Station remaining the most hierarchical node in the contemporary situation, other nodes have become more hierarchical, underlining the effectiveness of the policy based on the hierarchical changes.

The average travel time in the system for the pre-NZL scenario is 25 minutes while this is 24 minutes in the base year. Consequently, leaving out any other reasons why the travel time could have decreased, the opening of the NZL has led to a slight decrease in average travel time. It should be noted that additional demand, in particular for Station Noord could be a partial explanation for this decrease.

5.2.2 Future scenarios

For the future scenarios, there are two different alternatives for two metro-lines evaluated. A more detailed description of each of these scenarios is shown in Appendix B. Each of these scenarios is evaluated briefly first, after which the differences in results for the scenarios are compared.

Scenario G1: Extension metro 50 to Centraal Station

For the first scenario, the metro to Isolatorweg is extended towards Centraal Station which finishes the ring line for metros. The results for this scenario are shown in table 5.11 from which can be derived that most nodes in the top 10 decrease in hierarchical degree value besides the nodes located on the extended line (Centraal Station, Station Bijlmer ArenA, Amstelveenseweg, Station Zuid & Station Lelylaan). It makes sense that these nodes would increase in hierarchical degree as these are all served by the extended line. However, another node on the line, Station Sloterdijk decreased in rank which is caused by a slight decrease for the non-redundancy and transfer potential. The decrease for the non-redundancy can be explained as the additional part of the line overlaps with the train-line between Amsterdam Centraal and Schiphol Airport and creates a new line among directly connected nodes. The decrease for the transfer potential can be explained by passengers on and towards the metro line no longer having to transfer at Sloterdijk but being able to transfer at Centraal Station instead.

[5]Purely looking at the network structure rather than the construction process which faced many issues

[6]See: `https://www.parool.nl/nieuws/de-noord-zuidlijn-wordt-de-ruggengraat-van-de-stad~bbaa2d6e/`

Table 5.11: Highest hierarchical degree values for the scenario G1 and the base year

Rank	Node	H_i value with G1	Rank	Node	H_i value base year
1.	Centraal Station	0.745	1.	Centraal Station	0.739
2.	Station Sloterdijk	0.432	2.	Station Sloterdijk	0.437
3.	Station Zuid	0.404	3.	Station Zuid	0.398
4.	Amstelstation	0.309	4.	Amstelstation	0.309
5.	Station Lelylaan	0.305	5.	Station Lelylaan	0.299
6.	Leidseplein	0.278	6.	Leidseplein	0.283
7.	Muiderpoortstation	0.227	7.	Muiderpoortstation	0.227
8.	Station Bijlmer ArenA	0.187	8.	Vijzelgracht	0.183
9.	Vijzelgracht	0.180	9.	Station Bijlmer ArenA	0.176
10.	Amstelveenseweg	0.164	10.	Amstelveenseweg	0.157

The power-law coefficient for this scenario is -1.648 while the Gini coefficient is 0.902. This is an indication the hierarchy of the network has increased by a minimal margin. The average travel time is with 24 minutes equal to the base year indicating no real effect of the line can be seen based on the way this scenario is evaluated.

Scenario G2: Extension metro 50 to Noorderpark

For the scenario G2, the metro-line is extended from Isolatorweg as it is in scenario G1. However, for this scenario it is connected to Noorderpark instead of to Centraal Station. Consequently the line has to cross the river 'Het IJ' by the means of a new tunnel. In table 5.12 the results for the 10 highest scoring nodes are shown indicating only a few minor changes to hierarchy occurred for this scenario. The node Noorderpark to which the new line connects, has increased in hierarchical degree but still only ranks 44th, opposed to 88th in the base year. While this node has increased on any of the elements, the increase for the transfer potential is less than a tenth which is caused by only a minor increase in transfer passengers. Therefore, it should be noted that the desired increase for northern Amsterdam by routing the line towards Noorderpark is disappointing.

Table 5.12: Highest hierarchical degree values for the scenario G2 and the base year

Rank	Node	H_i value with G2	Rank	Node	H_i value base year
1.	Centraal Station	0.738	1.	Centraal Station	0.739
2.	Station Sloterdijk	0.440	2.	Station Sloterdijk	0.437
3.	Station Zuid	0.401	3.	Station Zuid	0.398
4.	Amstelstation	0.310	4.	Amstelstation	0.309
5.	Station Lelylaan	0.304	5.	Station Lelylaan	0.299
6.	Leidseplein	0.281	6.	Leidseplein	0.283
7.	Muiderpoortstation	0.227	7.	Muiderpoortstation	0.227
8.	Station Bijlmer ArenA	0.183	8.	Vijzelgracht	0.183
9.	Vijzelgracht	0.182	9.	Station Bijlmer ArenA	0.176
10.	Amstelveenseweg	0.162	10.	Amstelveenseweg	0.157

The power-law coefficient for this scenario is -1.638 while the Gini coefficient is 0.900. This is an indication the hierarchy of the network has decreased a little despite no major changes in hierarchical degree for the 10 highest scoring nodes. The average travel time is with 24 minutes equal to the base year indicating no real effect of the line is seen based on the way this scenario is evaluated.

Scenario R1: New metro from Schiphol to Centraal Station

For the scenario R1, a new line is added to serve between Schiphol, the western suburbs of Amsterdam, to Lelylaan, Leidseplein and then towards Amsterdam Centraal. In table 5.13 an overview of the hierarchical degree is given for the 10 highest scoring nodes in scenario R1 and the base year. The value for the number one node, Centraal Station has increased by a minimum which is caused by an increased value for the non-redundancy. Further down, in particular the node Leidseplein has increased significantly which makes sense as this node is an important transfer station for the new line. It comes to no surprise that this increase is completely caused by an increase in value for the transfer potential which is both the result of additional transfer directions and an increased share in transfer passengers. In general, however, most nodes besides Leidseplein and Station Lelylaan appear to be dropping in hierarchical degree which could be an indication that Centraal Station has become even more influential and other nodes score lower compared to Centraal Station. This could then be reflected in the hierarchical degree for the other nodes as for the topological influence and transfer share, normalized values are used.

Table 5.13: Highest hierarchical degree values for the scenario R1 and the base year

Rank	Node	H_i value with R1	Rank	Node	H_i value base year
1.	Centraal Station	0.740	1.	Centraal Station	0.739
2.	Station Sloterdijk	0.429	2.	Station Sloterdijk	0.437
3.	Station Zuid	0.384	3.	Station Zuid	0.398
4.	**Leidseplein**	**0.352**	4.	Amstelstation	0.309
5.	Station Lelylaan	0.321	5.	Station Lelylaan	0.299
6.	Amstelstation	0.305	6.	**Leidseplein**	**0.283**
7.	Muiderpoortstation	0.224	7.	Muiderpoortstation	0.227
8.	Vijzelgracht	0.174	8.	Vijzelgracht	0.183
9.	Station Bijlmer ArenA	0.173	9.	Station Bijlmer ArenA	0.176
10.	Elandsgracht	0.153	10.	Amstelveenseweg	0.157

The values for the hierarchical coefficient are -1.647 and 0.905 for the power-law coefficient and the Gini coefficient respectively. This is an indication that the network has become a very little more hierarchical and the hierarchical structure is barely affected. The average daily travel time for this scenario is 24 minutes.

Scenario R2: New metro from Schiphol to Muiderpoortstation

For this scenario, the metro line follows the same path as for R1 until Leidseplein. Thereafter, it follows a different route towards Muiderpoortstation including intersections with the other metrolines in between. It is to be expected this metro would relieve some of the pressure of Centraal Station as it connects many of its surrounding nodes. Table 5.14 shows the results for the hierarchical degree of the scenario with the line added compared to the base year. First of all it should be noted that the hierarchical degree of the nodes after the first three nodes has increased significantly which may be an indication the nodes on the ’second order’ become more important. The value for Centraal Station remains unchanged indicating it is still most influential, has the most transfers and has not become more redundant. This makes sense as the new line does not serve Centraal Station and overlaps for a large part with line 1, thus not creating new connections among mutual neighbors.

The highest increase in rank and hierarchical degree is the node Leidseplein, which is served by the new line. This change in hierarchical degree is caused the most by a significant increase in the value for the transfer potential, indicating an increase in transfer potential. This makes sense intuitively as Leidseplein was already an important tram and bus hub and with the addition of a metro-line the node offers transfers towards further located nodes as well. Besides Leidseplein, every node among the 10 highest ranking nodes showing significant increase are the nodes served by the new line. Furthermore, it should be noted, that the nodes Surinameplein and Alexanderplein, both covered by the new line, greatly increase in rank to just outside the ten highest scoring nodes.

Table 5.14: Highest hierarchical degree values for the scenario R2 and the base year

Rank	Node	H_i value with R2	Rank	Node	H_i value base year
1.	Centraal Station	0.739	1.	Centraal Station	0.739
2.	Station Sloterdijk	0.441	2.	Station Sloterdijk	0.437
3.	Station Zuid	0.396	3.	Station Zuid	0.398
4.	**Leidseplein**	**0.347**	4.	Amstelstation	0.309
5.	Station Lelylaan	0.333	5.	Station Lelylaan	0.299
6.	Amstelstation	0.312	6.	**Leidseplein**	**0.283**
7.	Muiderpoortstation	0.265	7.	Muiderpoortstation	0.227
8.	Vijzelgracht	0.238	8.	Vijzelgracht	0.183
9.	Weesperplein	0.225	9.	Station Bijlmer ArenA	0.176
10.	Station Bijlmer ArenA	0.179	10.	Amstelveenseweg	0.157

The values for the hierarchical coefficient are -1.684 and 0.908 for the power-law coefficient and the Gini coefficient respectively. This is an indication that the network has become more hierarchical and there are more nodes with a relatively high hierarchical degree. The average daily travel time for this scenario is 23 minutes.

Comparison of the future scenarios

First of all it should be noted that while the analysis of the future scenarios does provide interesting results, only limited conclusions can be drawn as the additional demand by realizing new lines cannot be accounted for in the scenarios. For these four scenarios, the R2 scenario appears to have the most impact on the PTN, where in particular nodes in the rank between four and ten seem to increase the most in terms of hierarchy. This could be seen as a desirable change as it both decreases the pressure on the most hierarchical node(s) while it increases the hierarchy of the network as a whole.

In order to evaluate and compare the effectiveness of the scenarios, it is important to define the desired changes in the network structure. Based on the future vision for the metro network in the Amsterdam PTN, there are a few relevant policy points that should be considered for evaluating changes to the network structure:[7]

[7]Adapted from https://www.noordzuidlijnkennis.net/wp-content/uploads/2013/05/metronetstudie.pdf

- Overcome capacity limitations, offering alternative routes
- Focus on a robust and reliable network, able to deal with disruptions
- Close the ring line of the metro
- Improve the connectivity to Schiphol

While the former two policy points can (in)directly be evaluated based on the changes in hierarchical structure, the latter two depend on the spatial location of the line rather than the hierarchical structure. Furthermore, the scenarios G1 and G2 comply with the third policy point while the scenarios R1 and R2 comply with the fourth policy point. Therefore, neither of the scenarios clearly distinguishes itself for these two policy points.

For the capacity limitations, the G1 and R1 scenario offer alternative lines for Centraal Station, which is considered the busiest node. In this sense, these scenarios do offer additional capacity to overcome capacity limitations. However, for both of these scenarios, the hierarchical degree of this node increased even further, emphasizing the node becomes even more important. Considering the second policy point, of being robust and reliable, having one crucial node in the network is undesirable. If, for example, an incident happens and this crucial node becomes unavailable, the whole network would potentially collapse. It is therefore, taking the robustness and reliability of the network into account, not desirable to adapt the network by mainly increasing the importance of Centraal Station. Nevertheless, the additional capacity these scenarios offer could potentially overcome limitations in the future. Comparing these two scenarios mutually, the R1 scenario performs better in terms of desirable hierarchical change by increasing the number of alternative routes to multiple nodes while the G1 scenario only offers an additional connection to Sloterdijk. It is therefore the R1 scenario is perceived more desirable.

The G2 and R2 scenario are different than the former two as these do not add an additional line to Centraal Station. The hierarchical degree for this node does therefore not change or decreases slightly. One thing to consider is that, mostly for the R2 scenario, the hierarchical degree of other nodes among the ten highest scoring nodes increases significantly. It can therefore be interpreted that the 'second order of hierarchical nodes' become more important which is considered desirable in terms of robustness. Furthermore, the capacity offered by the additional lines for both the G2 and the R2 provide an alternative route to avoid and relieve Centraal Station. For the G2 scenario this would be a direct connection between Sloterdijk and Noorderpark. For the R2 scenario this would offer a faster connection between Schiphol, Amsterdam West and Amsterdam East while some of the current shortest routes between these regions would be through Centraal Station. While both of these scenarios do offer alternative routes to relieve Centraal Station, the R2 scenario induces significant and desirable changes to the hierarchical structure while the G2 scenario is unable to do so.

Comparing the G2 scenario to the G1 and R1 scenario, it does offer an additional alternative route to avoid Centraal Station but does not relieve Centraal Station significantly. While it does increase the hierarchy of 'second order nodes' it does less so than scenario R1. Furthermore, due to only a slight decrease in hierarchy for Centraal Station, the G2 is perceived as slightly more desirable than scenario G1 but less desirable than scenario R1 and R2

An overview of the comparison of the results for each scenario is shown in table 5.15.

In the last column, an estimation of the implementation costs are shown based on the routing, existing infrastructure, length of the route and necessity of additional civil engineering works. These values are merely rough estimations and are not based on any kind of source. It should be noted that any of these additional lines would be a significant investment for the region.

Table 5.15: Comparison of the future scenarios for the Amsterdam PTN

Scenario	New route	Desirable hierarchical change	Implementation costs
G1	Isolatorweg - Centraal Station	+-	+
G2	Isolatorweg - Noorderpark	+-	-
R1	Schiphol Airport - Centraal Station	+	-
R2	Schiphol Airport - Muiderpoort	++	- -

The R2 scenario is perceived to be the most desirable in terms of the hierarchical structure but is the most expensive in terms of implementation. The benefits of this line would be decreasing average travel time, relieving the pressure on Centraal Station, offering a faster alternative for tram 1 and a new connection from the airport to the city. Following from the results for the hierarchical changes, a substantiation for policies towards an expansion of the metro-network can be augmented.

The R1 scenario would be the second most desirable but does not relieve the pressure on Centraal Station. The implementation costs are however, much lower than for the R2 scenarios as it can use existing infrastructure. The G2 scenario offers some desirable hierarchical change to relieve Centraal Station but this is only limitedly reflected in hierarchical change. While the 'second order of nodes' increases slightly, the reduction in hierarchy for Centraal Station is very small. Furthermore, this scenario is rather costly in terms of implementation costs as a new tunnel has to be constructed. The G1 scenario offers the least desirable hierarchical changes as the network structure does not change much (since the connection between Sloterdijk and Centraal Station exists already) but is also the cheapest option. An overview of the scenarios in terms of the implementation costs and hierarchical changes is shown in figure 5.10.

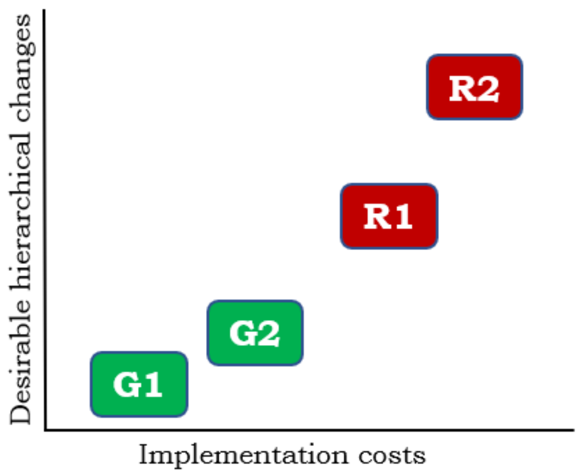


Figure 5.10: Comparison of the hierarchy changes and implementation costs of the future scenarios

These scenarios have shown how the metric and its output could be applied to create a foundation for the development (or rejection) of new lines based on the desirable changes from policy. Therefore, it could be incorporated into the decision making regarding PT lines to test whether a new line would go hand in hand with the desired hierarchical changes.

5.3. Results of the case-study for the Rotterdam PTN

Throughout this section, the results for the Rotterdam case-study are outlined. In order to compare the results to the output of the Amsterdam network, the same structure, going through the output for each of the elements first, is applied. Furthermore, the analysis is supplemented by a comparison to the Amsterdam network in each step as well as a more elaborate comparison for the network structure.

5.3.1 Results for the different elements

Output for the topological influence in the Rotterdam PTN

In table 5.18, the results for the five highest scoring nodes for the topological influence in the Rotterdam network are shown. First of all, unsurprisingly the central station comes out on top as the most influential node, which is similar to the Amsterdam network. Furthermore, the other highest scoring node rank much closer in value to the most influential node than for the Amsterdam network (for Rotterdam the nodes in the rank 2 to 5 are all rounded up to 0.9 where these are rounded up to 0.7 and 0.8 for the Amsterdam network). Furthermore, it should be noted that four out of five most influential nodes (Marconiplein as the exception) in the Rotterdam network are located in close proximity of each other, separated by only a few hundred meters. Comparing this to the Amsterdam PTN, the most influential nodes are much closer together while only one train station is included in the top 5 for Rotterdam opposed to three train stations for Amsterdam.

Table 5.16: Overview of the highest scoring nodes for the topological influence in the Rotterdam PTN

Rank	Node label	e^A value
1.	Rotterdam Centraal	1.000
2.	Stadhuis/Weena	0.937
3.	Beurs	0.934
4.	Marconiplein	0.874
5.	Kruisplein	0.870

In figure 5.11 an overview of the cumulative distribution function of the topological influence for the Rotterdam network is shown. The figure shows that almost 80% of the nodes in the network do not have any influence at all, which is around 10% more than for the Amsterdam network. This is caused by these nodes having a degree of one or two for which the share is therefore higher than for the Amsterdam PTN. This is an indication that there is a smaller share of hierarchical nodes in the Rotterdam PTN. It should also be noted that only approximately 10% of the nodes has a topological influence value higher than 0.2 indicating the higher influence is shared by only a small share of nodes.

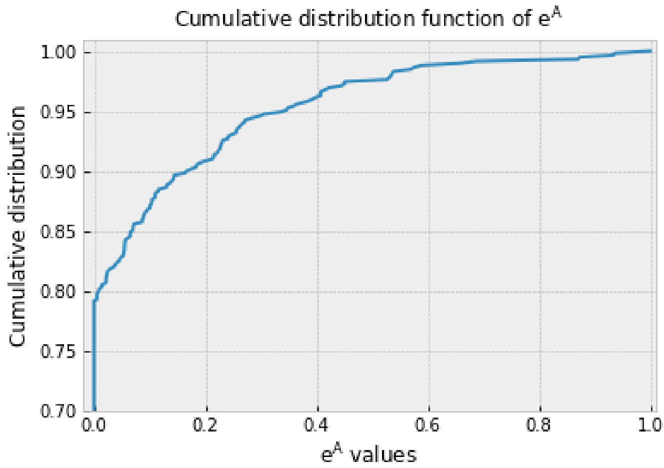


Figure 5.11: Cumulative distribution function of the topological influence for the Rotterdam PTN

Output for the (non-)redundancy in the Rotterdam PTN

In table 5.17 the five highest values for the non-redundancy in the Rotterdam PTN are shown. First of all, not the central station but Zuidplein, the hub in the southern part of Rotterdam is the least redundant. Three of the four other nodes in the top 5 are important train stations while Stadhuis/Weena is an important intersection of trams and metros near Rotterdam Centraal. Comparing this to the Amsterdam PTN, the values for the non-redundancy for the five highest ranking nodes seems to be more or less in the same range. One notable difference is that for the Amsterdam PTN, every node in the top 5 is a train station while only three out of five are so for the Rotterdam PTN. This could be an indication that the train stations on the edge of the network for the Amsterdam PTN are better connected to the network than for the Rotterdam PTN.

Table 5.17: Overview of the highest scoring nodes for the (non-)redundancy in the Rotterdam PTN

Rank	Node label	e^B value
1.	Zuidplein	0.881
2.	Rotterdam Centraal	0.833
3.	Stadhuis/Weena	0.774
4.	Schiedam Centrum	0.771
5.	Station Alexander	0.754

In figure 5.12, the cumulative distribution function of the non-redundancy for the Rotterdam network is shown. It shows a similar share of 0 scoring nodes as for the topological influence and also shows similar patterns as the Amsterdam PTN with a high share of non-zero scoring nodes in the range between 0.2 and 0.6. This implies that a relative high share of non-zero scoring nodes has low redundancy to a certain extent.

Output for the transfer potential in the Rotterdam PTN

For the transfer potential, the results for the Rotterdam PTN are shown in table 5.18 where the left side shows the results for the transfer potential in total and the right side shows the results for the transfer share excluding transfer directions. Comparing the right side to the left side of the table, it can be noted that the node Beurs has a very high transfer share but still does only score 5th for the

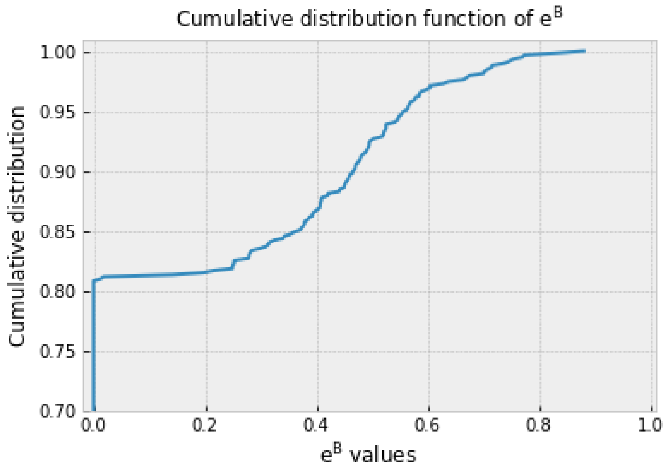


Figure 5.12: Cumulative distribution function of the non-redundancy for the Rotterdam PTN

transfer potential. This can only be caused by a lack of transfer directions for Beurs. In comparison to the Amsterdam PTN, there are less train stations for Rotterdam among the highest scoring nodes. This can be explained as Zuidplein is an important transfer hub for regional buses, trams and metros while Beurs is the only node where all metros intersect. The value for the central station in Amsterdam is higher than for Rotterdam which is an indication it offers more transfer directions. For the transfer share, in Rotterdam there are two nodes really close in terms of highest transfer share while for Amsterdam there is one node standing out. This could be an indication that the network of Rotterdam has more spread out transfers than the Amsterdam PTN.

Table 5.18: Overview of the highest scoring nodes for the transfer potential and transfer share in the Rotterdam PTN

Rank	Node label	e^C value	Rank	Node label	log transfer share
1.	Rotterdam Centraal	0.833	1.	Rotterdam Centraal	1.000
2.	Schiedam Centrum	0.782	2.	Beurs	0.990
3.	Zuidplein	0.739	3.	Schiedam Centrum	0.939
4.	Station Blaak	0.674	4.	Station Blaak	0.899
5.	Beurs	0.660	5.	Zuidplein	0.874

In figures 5.13 and 5.14 the cumulative distribution functions of the transfer potential and transfer share for the Rotterdam PTN are shown respectively. First of all, it should be noted that the share of zero scoring nodes is higher than for the topological influence and (non-)redundancy. This is caused by nodes having a degree of three or higher but do not facilitating any transfers. For the non-zero scoring nodes the highest share for the transfer potential values is in the range between 0.05 and 0.3 while it is more spread out for the normalized transfer share. This is an indication that the transfer directions, based on the degree, affects the value of the transfer potential significantly. Compared to the Amsterdam PTN, there is a smaller share of nodes with a normalized transfer share higher than 0.6 which indicates a majority of the transfers occur at only a few nodes.

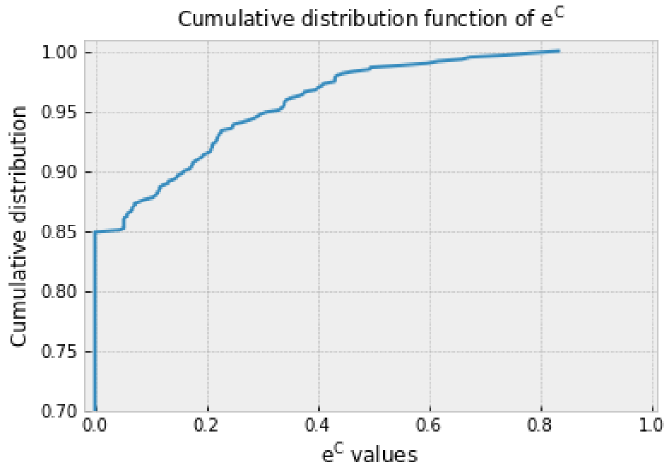


Figure 5.13: Cumulative distribution function of the transfer potential for the Amsterdam PTN

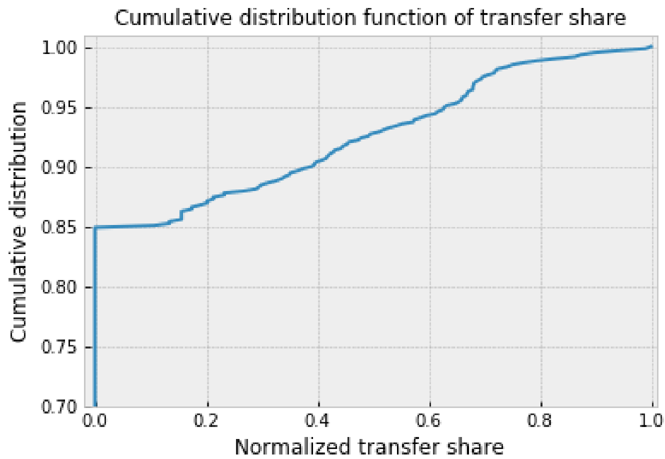


Figure 5.14: Cumulative distribution function of the transfer share for the Amsterdam PTN

5.3.2 Results for the hierarchical degree

By combining the elements, the hierarchical degree of the nodes for the Rotterdam PTN can be determined. For the ten highest scoring nodes in the network, table 5.19 provides the results for the hierarchical degree and the elements. Unsurprisingly, Rotterdam Centraal is the highest scoring node for the Rotterdam PTN, as was the case for the central station in the Amsterdam PTN. However, the value for the central station in Rotterdam is slightly lower than in Amsterdam (0.694 opposed to 0.739) which is mainly caused by a higher redundancy and a lower number of transfer directions. The second node for the Rotterdam PTN, Beurs, scores higher than the second node for Amsterdam which is mainly the result of a high topological influence of the node. Further down, nodes show comparable values to the Amsterdam network in terms of hierarchical degree. One notable difference between the Amsterdam and Rotterdam PTN, which could already be identified at the results for the elements, is the location within the network of the high scoring nodes. Out of the ten highest scoring nodes, five are located very close to each other in the city center while the high scoring nodes for the Amsterdam network are more spread out over the network. Furthermore, what could be related to the former, for the Rotterdam network, among the ten highest scoring nodes only four are train stations while for the Amsterdam network there are seven.

In figure 5.15 a map showing the nodes and their hierarchical degree is shown in which it is clearly visible most hierarchical nodes are located in the center. Furthermore, the northern, eastern and southern areas of the city appear to have barely any hierarchical nodes. In figure 5.16, an overview of the nodes in the network, divided into different sub-levels, is shown. In this overview, 11 nodes are allocated to level 1, 78 to level 2 and 502 to level 3. Furthermore, a more detailed overview of the highest scoring nodes and the key elements for these nodes is shown in Appendix C.

Table 5.19: Results for 10 nodes with the highest hierarchical degree (H_i) in the Rotterdam PTN

Rank	Node	$\mathbf{H_i}$	$\mathbf{e^A}$	$\mathbf{e^B}$	$\mathbf{e^C}$
1.	Rotterdam Centraal	0.694	1.000	0.833	0.833
2.	Beurs	0.461	0.934	0.747	0.660
3.	Schiedam Centrum	0.397	0.658	0.771	0.782
4.	Stadhuis/Weena	0.313	0.937	0.774	0.432
5.	Marconiplein	0.300	0.874	0.738	0.465
6.	Station Blaak	0.268	0.589	0.676	0.674
7.	Eendrachtsplein	0.190	0.536	0.715	0.496
8.	Zuidplein	0.177	0.272	0.881	0.739
9.	Schiedam Nieuwland	0.150	0.534	0.569	0.494
10.	Station Alexander	0.142	0.306	0.754	0.616

5.3.3 Correlation analysis of hierarchical degree and elements

For the different elements in the Rotterdam PTN, a correlation analysis is executed to provide an overview of how the different elements and the hierarchical degree are correlated. Furthermore, by comparing the correlation analysis of the Rotterdam network to the Amsterdam network, there might be some patterns to be identified. Figure 5.17 shows the results for the correlation analysis with the top row indicating the correlation between the elements and the hierarchical degree. For easier comparison, the correlation graph for the Amsterdam PTN is shown once more in figure 5.18 where it is greyed out for distinctness. The correlation between both the topological influence and the non-redundancy with the hierarchical degree for the Rotterdam PTN is quite similar to the Amsterdam network but the correlation between the transfer potential and the hierarchical degree is considerably lower. This could be explained by some nodes in the Rotterdam PTN being located further from the center and having a relatively low hierarchical degree, but being an important transfer hub for regional passengers. These nodes do generally have a lower influence as many of the influential nodes are all located in the city center and the peripheral transfer hubs are not connected to all of these.

Emphasizing this assumption is the correlation between the topological influence and the transfer potential, which is much lower for the Rotterdam PTN than for the Amsterdam PTN. This is likely caused by influential nodes being located in the city center while transfer hubs are also located towards the edges of the network. The correlation between the topological influence and redundancy is also much lower for the Rotterdam PTN which could be explained as many of the influential nodes in the city center are also redundant due to many mutual connections. Lastly, the correlation between the redundancy and transfer potential appears to be higher for the Rotterdam PTN which could mean the important transfer hubs in the Rotterdam PTN are less redundant than in Amsterdam, offering fewer locations for transfers.

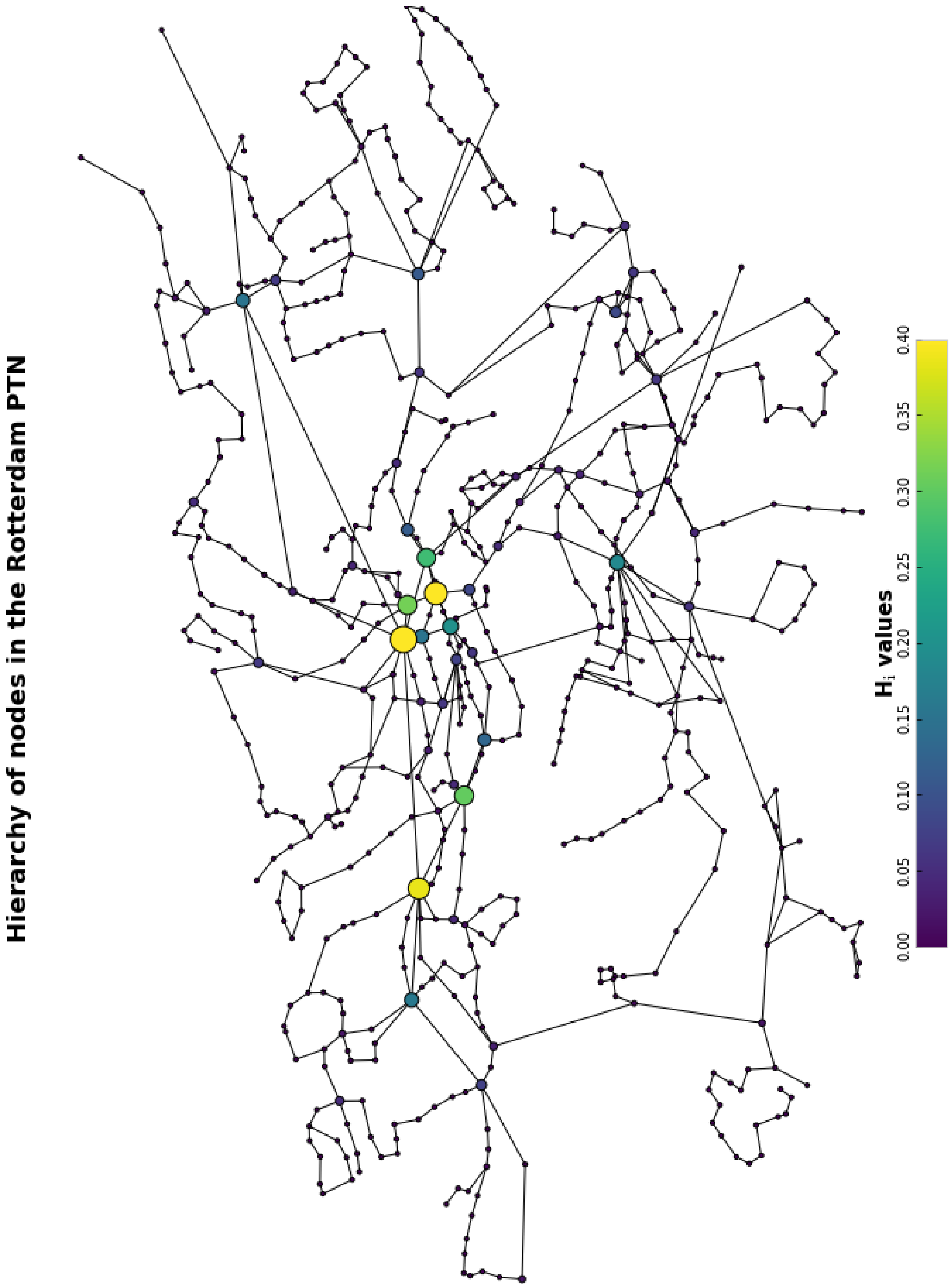


Figure 5.15: Map of the spatial distribution of the hierarchy of nodes in the Amsterdam PTN

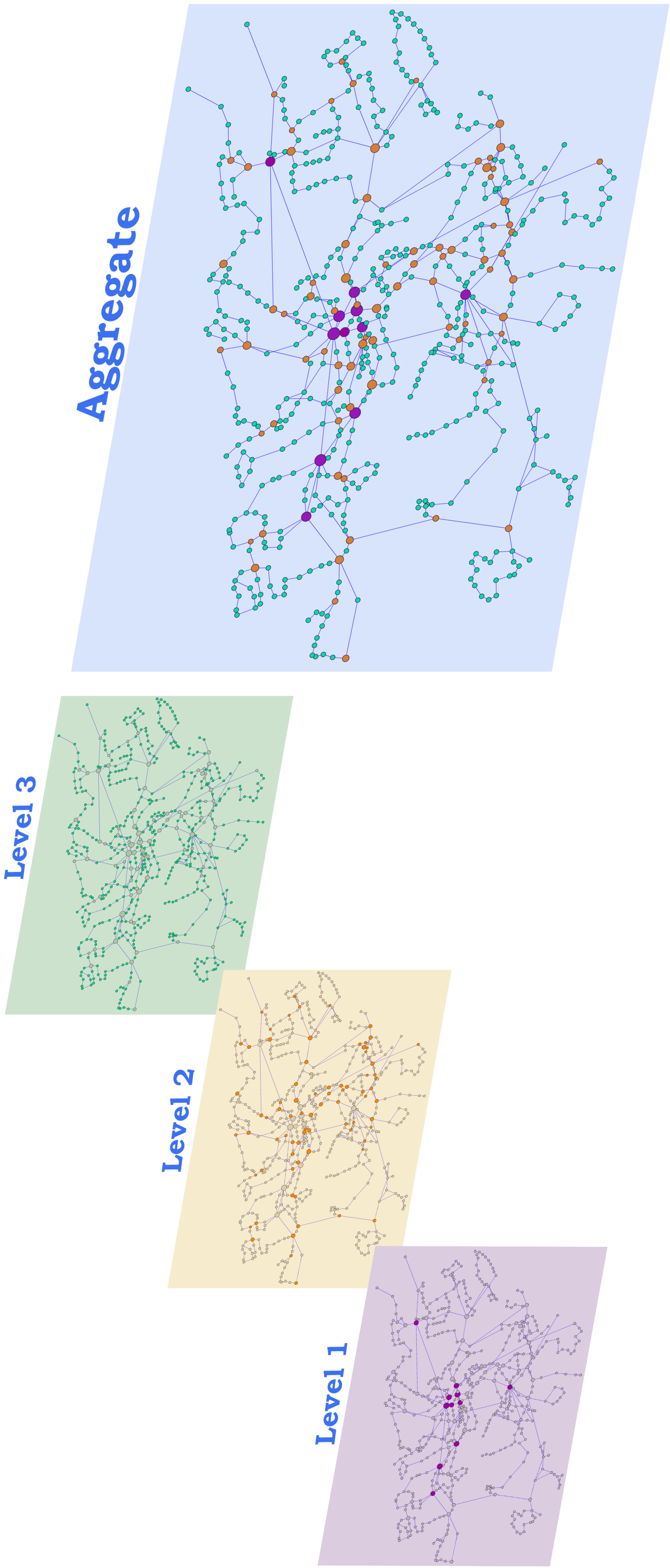


Figure 5.16: Overview of nodes in the Rotterdam PTN divided into three hierarchical levels

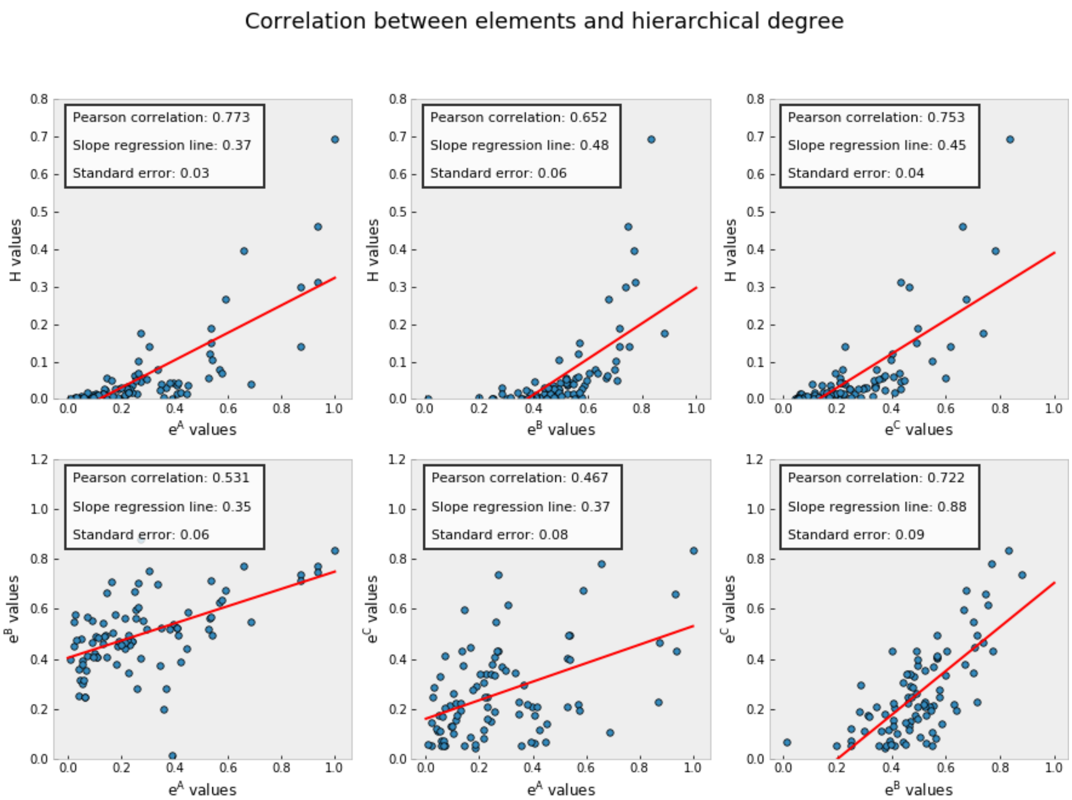


Figure 5.17: Correlation between hierarchical degree and elements for the Rotterdam PTN

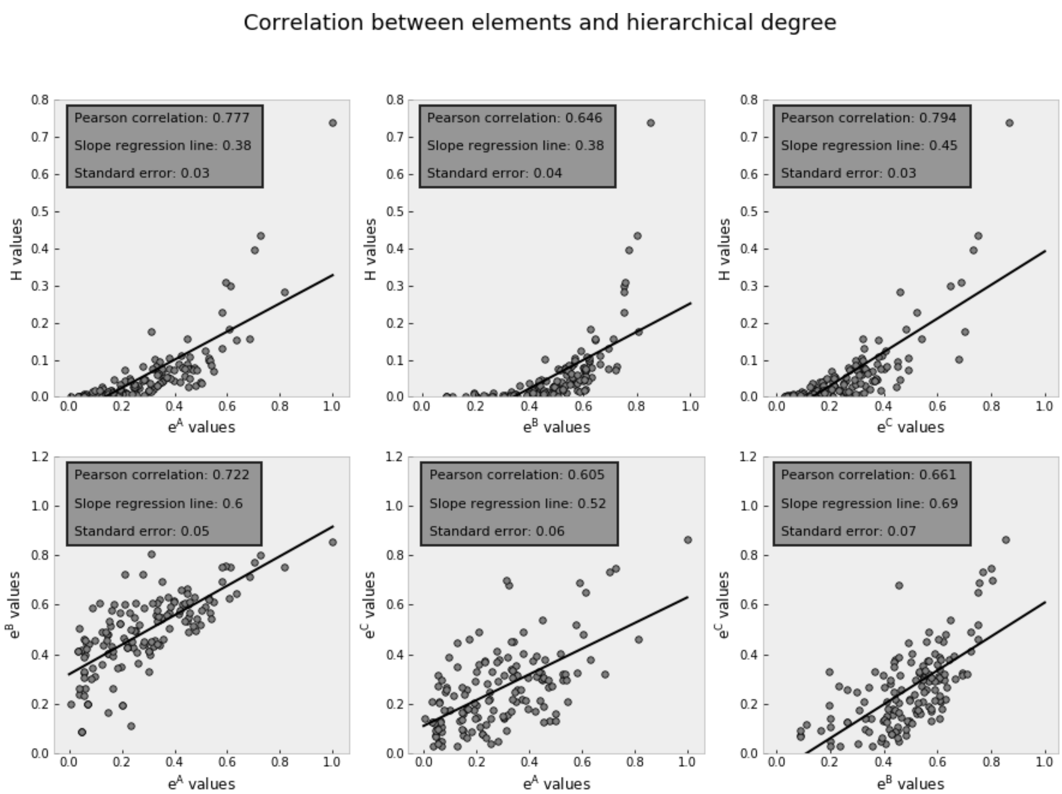


Figure 5.18: Correlation between hierarchical degree and elements for the Amsterdam PTN (greyed)

5.3.4 Results for network hierarchy

Based on the program code for estimating the power-law parameters, developed by Alstott et al. (2014), the hierarchy of the Rotterdam PTN can be determined. Figure 5.19 and figure 5.20 show the results for the power-law coefficient and the Gini-coefficient respectively. With the alpha coefficient of -1.391 and a Gini-coefficient of 0.954, the hierarchical can be determined with

$$H_I = 0.954^{\frac{1}{1.391}} = 0.967, \tag{5.2}$$

indicating the Rotterdam PTN can be considered more hierarchical than the Amsterdam PTN. Looking closer at the values for both coefficients, it is notable that the power-law coefficient is closer to zero (lower negative value) than for the Amsterdam PTN while the Gini-coefficient is higher. One of the reasons why the Gini-coefficient for the Rotterdam PTN is likely to be higher is the higher fraction of zero scoring nodes for the Rotterdam PTN compared to the Amsterdam PTN which are 85% and 74% respectively. The higher value for the power-law coefficient for the Amsterdam PTN is mainly a results of the higher share of nodes with a relatively low hierarchical degree. Linking this to the former, the Amsterdam PTN appears to have a higher share of nodes with a low hierarchical degree while the Rotterdam PTN has a higher share of zero scoring nodes. In the end, the higher share of zero scoring nodes in the Rotterdam PTN outweighs the higher power-law coefficient for the Amsterdam PTN in terms of the hierarchical coefficient. It is, however, debatable which network structure should be valued as preferable based on these outcomes. This is further elaborated in the upcoming section 5.3.6.

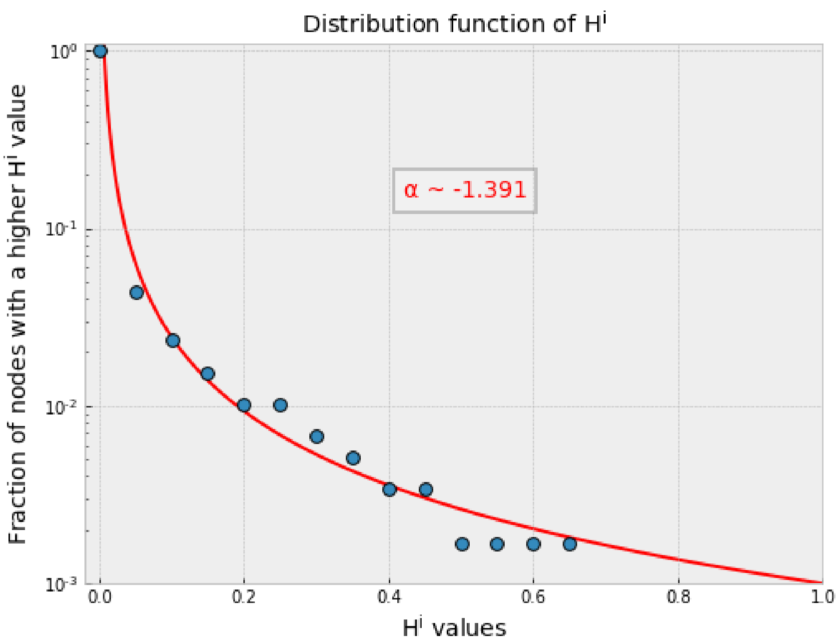


Figure 5.19: Distribution function of hierarchical degree in the Rotterdam PTN

5.3.5 Comparing to other indicators

In table 5.20 the correlation between the hierarchical degree and other indicators for the Rotterdam network is shown for nodes with a hierarchical degree of higher than zero. From this table it can be

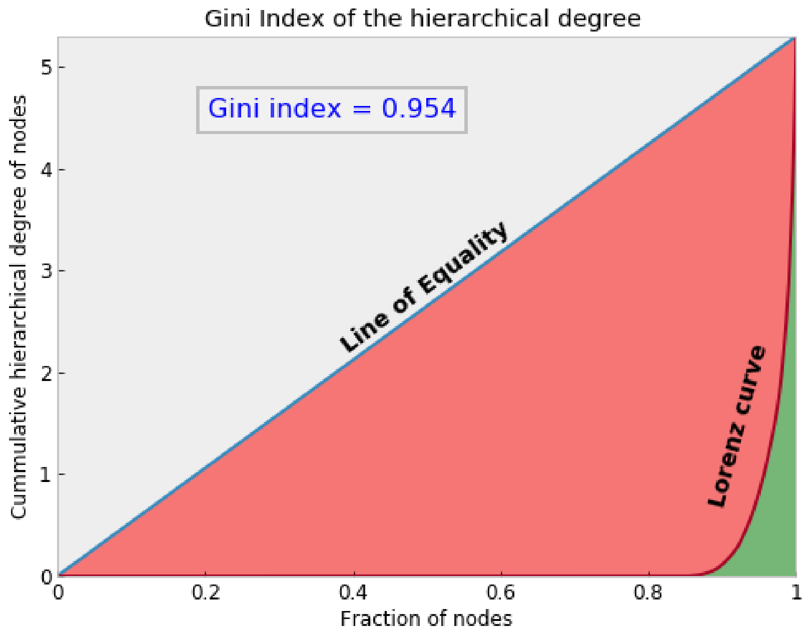


Figure 5.20: Gini-index for hierarchical degree in the Rotterdam PTN

interpreted that the eigenvector centrality correlates most with the hierarchical degree, which is different than for the Amsterdam PTN. This could imply the influence of the nodes, which is measured by the eigenvector centrality, is more prominently represented in the hierarchy for the Rotterdam PTN than for the Amsterdam PTN. Furthermore, the correlation between these indicators and those for the Amsterdam PTN are generally lower for which no clear explanation is found.

Table 5.20: Correlation of other indicators with the hierarchical degree for the Rotterdam PTN

Indicator	**Pearson correlation coefficient**	**Significance (based on N=89)**
Degree (𝕃-space)	0.730	<0.001
Assortativity (𝕃-space)	0.343	<0.001
Clustering coefficient (ℙ-space)	-0.652	<0.001
Betweenness centrality (𝕃-space)	0.722	<0.001
Betweenness centrality (ℙ-space)	0.735	<0.001
Eigenvector centrality (ℙ-space)	0.773	<0.001
Overlapping degree	0.761	<0.001

5.3.6 Differences between the Amsterdam and Rotterdam network

Throughout this section, the case-study of Rotterdam has been elaborated. To clearly emphasize why the results of the Rotterdam PTN are different from the Amsterdam PTN, what the causes are for these differences and which network structure would be preferable, these are elaborated here. Furthermore, based on the outcomes of the comparison, a possible direction of network structure development for the less desirable network structure can be identified. First, it should be noted that while for both networks the same method is applied to determine the size of the network, notable differences between the layout of the networks are present. This is mainly caused by Rotterdam having significantly more suburbs directly attached to the city compared to Amsterdam. Consequently, when drawing conclusions, the implications of the network layout should be taken into account

For the distinct elements, the differences in values between the networks are only minor while the real difference is found in the location of the high scoring nodes. For the Amsterdam PTN, in general, the high scoring nodes after the central station are located at train stations around four to five kilometers from the central station itself. On the other hand, the high scoring nodes for Rotterdam after the central station are located in close proximity to the central station while train stations further away from the central station score lower in general. Furthermore, the Rotterdam PTN has a much higher share of zero scoring nodes in terms of hierarchy indicating that a smaller share of nodes functions as a transfer location.

The consequence of a smaller share of transfer locations would be having more transfers at the hubs, enhancing the pressure and essential function of the hubs. On the one hand, this could be a strength of the network on the condition that, if all hubs are well connected, no superfluous lines are realized. On the other hand, the vulnerability of the network increases as it becomes reliant on these transfer hubs which can be disrupted. The Amsterdam PTN appears to have better connections between the decentralized nodes by the means of a (nearly complete) ring structure, which relieves the pressure on the center of the network. For the Rotterdam network, no such ring structure can be identified which could be a reason a lot of the hierarchical nodes are located in the center. Therefore, even though the hierarchical coefficient of the Rotterdam PTN is higher than the coefficient of Amsterdam, the network structure of Amsterdam appears to be more balanced and robust in this case, by having a more scattered layout of its hierarchical nodes. Hence, for this comparison, the hierarchical coefficient does not explain the full picture while analyzing the location and scattering of the hierarchical structure does provide additional value.

Looking at the potential application of these findings, there are some possible policy directions for the Rotterdam PTN to develop a more robust network which relieves some of the pressure on the nodes in the center such as Rotterdam Centraal and Beurs. First, a ring shaped metro connecting suburban train stations or hubs such as Schiedam Centrum, Rotterdam Alexander and Zuidplein could be a good addition to the network, comparable to the ring line in Amsterdam. Furthermore, connecting the suburbs locally could lead to more of a hierarchical sub-structure or a 'second order of hierarchical nodes'. Coincidentally, an independent[8] source has drawn a 2050 vision for the Rotterdam network which incorporates these elements. This vision for 2050 is visualized in figure 5.21[9]. This is another example of how the metric can be applied as substantiation for policy regarding the development of a network structure. This might be aligned with the policy for social security in PT, developed by MRDH, which is the overarching institute for V-MRDH[10]. In this policy document, the desire for a robust and complete PTN is emphasized which requires a network structure with alternative routes and low vulnerability. To elaborate the link between the metric and the potential application for policies regarding the network structure, this is elaborated separately throughout the subsequent section.

5.4. Potential applications

While these results provide interesting outcomes, the linkage to its general potential applications for policy is rather unclear. For the case-studies and scenarios, conclusions can be drawn, but for

[8]Independent as in not connected to any involved company, municipality or province

[9]Retrieved from: `http://www.3develop.nl/blog/wp-content/uploads/2020/01/metrokaart-rotterdam-2050-max.jpg`

[10]Adapted from: `https://mrdh.nl/sites/mrdh.nl/files/files/Beleidslijn%20SVOV%20MRDH.pdf`

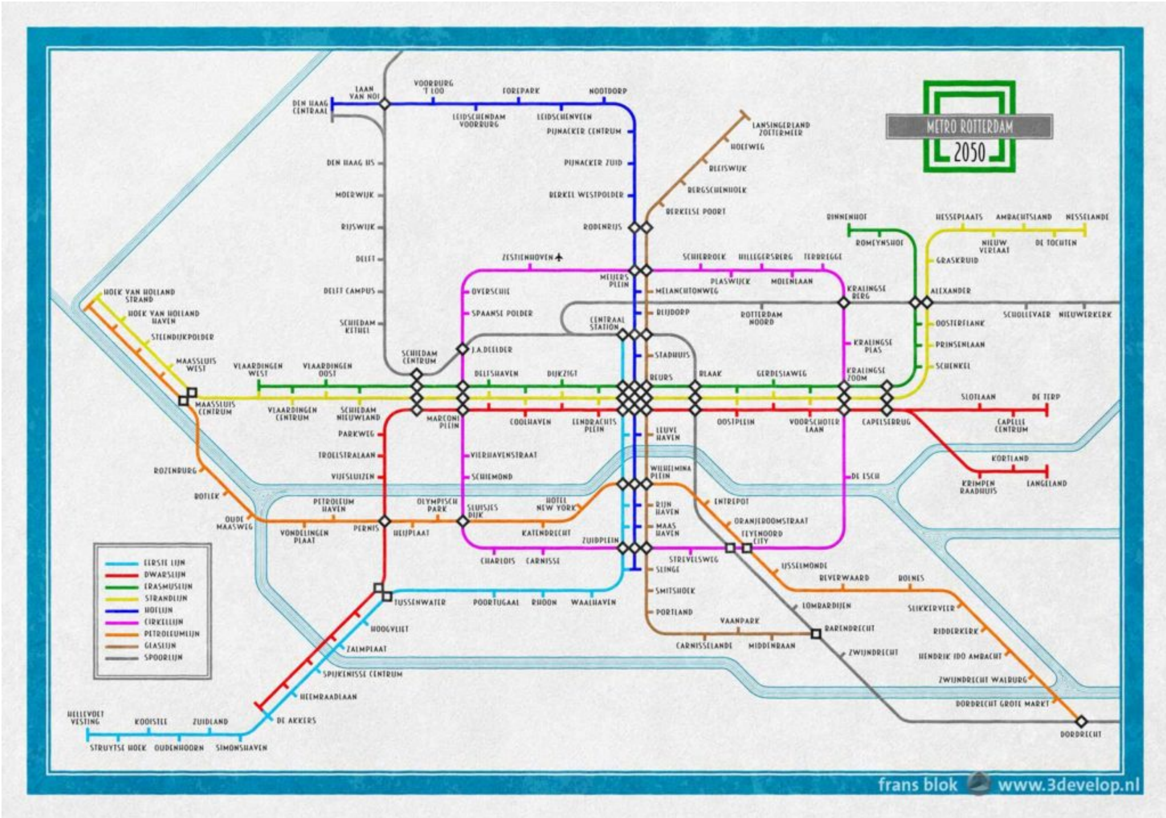


Figure 5.21: Future vision Metromap Rotterdam 2050 © Frans Blok

network structures and hierarchical structures in general, the results remain limited. Furthermore, the linkage to how policies and strategical decisions can be determined based on the results for the hierarchical degree, could be clarified more. Therefore, this section provides some potential applications of the results in terms of bottlenecks and vulnerability/cascading failures to generalize the findings from the case-studies. While the former is mainly an interpretation of how hierarchy could be spread out more among nodes in case one node is overcrowded, the latter relates to how a network can overcome disruptions by looking at the hierarchical structure. To clarify and elaborate how these potential applications should be interpreted, these two concepts are discussed separately and with the addition of examples from the case-studies.

5.4.1 Bottlenecks

For bottlenecks, each of the three elements for the hierarchical degree offers a different perspective on the identification of bottlenecks. The topological influence provides a method to identify which nodes are most influential in which a high difference between influence in nodes could be an indication one node has too much influence and is prone to becoming a bottleneck. Connected to this, the (non-)redundancy determines the redundancy of a node where nodes with low redundancy are crucial as transfer location and prone to being a bottleneck too. Lastly, the transfer potential is used to determine where the highest share of transfers take place in a network. If one node offers the most transfers in the network and other nodes only have a fraction of its transfer share, also taking into account logarithms are used, this node is susceptible to being a bottleneck.

While the elements can partially be applied for indicating which nodes are (prone to being) bottlenecks, there is one crucial aspect the hierarchy metric does not account for, the capacity of a node. As described in chapter 2.3.2, bottlenecks tend to occur where the demand is higher than the supply. The perspective on the supply side is largely neglected in this study as e.g. vehicle crowding and time of day are not taken into account for the metric. Consequently, the capacity of nodes remains unknown which implies only limited conclusions for nodes being bottlenecks can be drawn. There are, however, possible ways to deal with these limitations.

First, if a bottleneck is an issue for a PTN, its location is likely to be known, making the identification of it superfluous. With the location known, the hierarchical degree can be used to determine which of the elements might be the cause for this node being a bottleneck. If the value for the non-redundancy would be high for a bottleneck node, implying low redundancy, it might be a good idea to create mutual connections among directly connected nodes from the bottleneck. If the influence or transfer share of a node is high but the capacity on the supply side is insufficient, it may be a consideration to either increase the capacity of the node or to improve the influence or transfer share of nodes in its close proximity in order to relieve the pressure on the bottleneck. On the other hand, if a node has high capacity but has a limited hierarchical degree, the same logic can be applied vice versa. If, for example, a high capacity node has only a limited transfer potential, it might be a good consideration to increase the transfer directions of the node.

To provide a practical example, Station Noord in the Amsterdam PTN has been developed for mass transport towards the city center (by the means of the NZL) from both the northern part of the city and from the region north of the city. The policy for the network structure has been focused on reducing the number of buses from the region going to Centraal Station and therefore relieving the pressure on this node [11]. Following from the relatively low hierarchical degree value for Station Noord, there is at least one element withholding this node from reaching its potentially crucial function of gateway from the region to the city. Taking a closer Look at the values for the different elements, the topological influence seems to be the cause for the low hierarchical degree. Evaluating its direct connections, only the nodes on the NZL are influential connections, leading to the relatively low value for the topological influence. A potential increase for the topological influence would be a direct connection to other influential nodes such as Sloterdijk or Leidseplein. In potential this could be realized by developing scenarios G2 and G1/R1 with an extension from Noorderpark or Centraal Station respectively. Whether this may or may not be a realistic option, the metric proves to be useful in explaining why a node could be a bottleneck, or in this case, not reaching its potential. This could be incorporated in policies for enhancing Station Noord or the connectivity of northern Amsterdam in general.

Secondly, while the capacity of a node is not included in the metric for this study, it could potentially be applied for the metric. This could be linked to a more time focused application of the metric where elements such as time of day and vehicle crowding are incorporated to identify bottlenecks during peak hours. The peak demand could be used to determine the gross number of transfers in the network including in-flows and out-flows of nodes for specific lines. Combining this with vehicle capacity and frequency (or shared capacity for common corridors), the share of capacity used for a specific link can be determined. Based on the previous, capacity issues leading to bottlenecks are identified and narrowed down to lines and nodes. Thereafter, by computing the hierarchy in the network for the peak hour demand, the hierarchical structure can be analyzed. Furthermore, causes for bottlenecks can be analyzed similarly to the previous example and to the examples in the scenario study to derive solutions. This would basically be a combination of traditional transportation models to identify the bottlenecks and the hierarchy metric to identify policy directions to solve the bottlenecks. By providing a quantifiable foundation for the solution to bottlenecks, it is possible to develop strategic policy to overcome the impact and delays caused by bottlenecks in future adaptations of the network. This type of policy aligns with the policy developed by the European

[11] This policy is further elaborated on `https://vervoerregio.nl/nieuwlijnennet`

Parliament promoting sustainable transport and removing bottlenecks.[12]

5.4.2 Vulnerability and cascading failures

The concepts of vulnerability and cascading failures can be linked to every element of the hierarchical degree as each of these elements implicitly values vulnerability. A higher value for the influence indicates the network would be more sensitive to disruptions at this specific node. Furthermore, the less redundant a node is, the higher the vulnerability and the higher the transfer potential the more prone it is to disruptions. As the different elements can clearly be linked to the concept of vulnerability, it is important to understand how vulnerability can be assessed based on values for the hierarchical degree. It should be noted that vulnerability of a network is considered a normative characteristic while the hierarchy of a network is descriptive. This implies that a higher vulnerability would be considered a worse condition while a higher hierarchy value cannot be valued as better or worse in advance.

Building on the assumption that the severity of a disruption is correlated with the hierarchy of a node at which the disruption takes place, it is assumed that nodes with a higher hierarchical degree are more vulnerable in general. This can be explained as nodes with a high hierarchical degree are likely to be more critical in the network due to its influence, non-redundancy and transfer potential. Therefore, the hierarchical nodes are considered more vulnerable in general for this study.

Consequently, the vulnerability of a node can be derived from its hierarchical degree and the vulnerability of a network could be understood better by evaluating the hierarchical structure. To provide an example for this, the Rotterdam PTN is evaluated with its most hierarchical node, Rotterdam Centraal, disrupted and skipped for every line. By analyzing how the hierarchical structure changes with the disruption of one node, the vulnerability of the network to the disruption of a certain node can be evaluated. While some other nodes decrease in hierarchy, nodes close to Rotterdam Centraal, such as Stadhuis/Weena and Station Blaak, show a significant increase in hierarchical degree. This is an indication these nodes serve as an alternative for the disrupted node. Based on the capacity of these alternative nodes, the ability to adapt to the disruption of the specific node can be assessed. In table 5.21 the five highest scoring nodes in the disrupted network are shown where the values in brackets indicate the values for the non-disrupted situation. In this table it is clear that in terms of transfers, mainly Stadhuis/Weena increases significantly, indicating this is the most crucial alternative transfer location. Furthermore, the topological influence of each of these five nodes increases as the most influential node is removed and the second most influential node becomes the new most influential node.

Table 5.21: Results for 5 nodes with the highest H_i in the Rotterdam PTN with a disruption at Rotterdam Centraal

Rank	Node	$\mathbf{H_i}$	$\mathbf{e^A}$	$\mathbf{e^B}$	$\mathbf{e^C}$
1.(2.)	Beurs	0.503 (0.461)	1.000 (0.934)	0.755 (0.747)	0.667 (0.660)
2.(4.)	Stadhuis/Weena	0.446 (0.313)	0.974 (0.937)	0.783 (0.774)	0.585 (0.432)
3.(3.)	Schiedam Centrum	0.412 (0.397)	0.720 (0.658)	0.775 (0.771)	0.739 (0.782)
4.(5.)	Marconiplein	0.317 (0.300)	0.934 (0.874)	0.745 (0.738)	0.451 (0.465)
5.(6.)	Station Blaak	0.295 (0.268)	0.643 (0.589)	0.680 (0.676)	0.676 (0.674)

The former example also illustrates how the evaluation of the hierarchical degree of a disrupted

[12] This policy is further elaborated on http://www.europarl.europa.eu/RegData/etudes/ATAG/2018/620234/EPRS_ATA(2018)620234_EN.pdf

node is used to determine the effect of cascading failures. If a node is removed from the network and its demand is redistributed over neighboring nodes, the nodes increasing in influence and in particular increase in transfer share could become overcrowded. To prevent these cascading failures from affecting the network performance, the use of additional capacity could be estimated using the hierarchical degree. The robustness of the network could be increased by anticipating the effects of a certain random node failure and taking precautions in case of the occurrence of this disruption. In this way, the redistribution of the load could be devised in advance and additional capacity could be reserved.

Relating this to policy, in order to develop a robust network on a strategic level, it is important to take the possibilities of disruptions into account. For policies regarding the robustness of a network for the future[13], where the ability of a network to deal with the impact of incidents is crucial, the metric could be applied in a similar manner. By taking the effects of disruptions into account during the development and evolution of a network, a more robust network structure can be created.

[13] e.g. https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2019/06/07/schets-mobiliteit-naar-2040/schets-mobiliteit-naar-2040.pdf

6
Discussion & Conclusion

Throughout this study, the goal has been to quantify the hierarchy in PTN by developing a new metric. The results have indicated the developed metric is capable of determining the hierarchy in PTN where nodes that would be expected to be hierarchical were valued as so. Throughout chapter, the results of this study are evaluated and the questions composed in the introduction are answered. The metric developed in this study has been applied to two case-studies and several scenarios, but the question of how effectively it is to assess the hierarchy in PTN compared to previous studies is not yet determined. Hence, in this chapter the interpretation and implications of the results are elaborated first, which is done following the structure of the report. Following, the conclusions for this study are elaborated separately in which the research question and sub-questions are answered. Thereafter the limitations of the study are explained. Lastly, the recommendations based on the findings of this study are provided.

6.1. Discussion of the key findings

The results of the different case-studies have shown that the metric is capable of doing what it is developed to do, evaluating the hierarchical structure of a PTN. In order to define hierarchy for this study the approach of Mones et al. (2012), where the hierarchy in PTN is defined from the perspective of nodes, is followed for this study. Furthermore, related to the findings of Corominas-Murtra et al. (2011), the hierarchy in a network showed the presence of a clear pyramidal structure. The output for cities can be connected to the work of Bassolas et al. (2019) where hierarchical patterns within a city are analyzed. It is interesting to see that different states of networks and different cities show significantly different outcomes and patterns. Connected to the work of Ru and Xu (2005), the distribution of hierarchy in the network is used to understand the hierarchical structure. Deviating from the previous, this study finds only limited understanding of hierarchy in networks is derived from a coefficient due to the absence of a spatial element.

For the approach to hierarchy, previous works have mainly chosen either a link-based approach (Yerra & Levinson, 2005; Lee et al., 2017), line-based approach (Van Nes, 2002; Bagloee & Ceder, 2011; Jiancong et al., 2006) or node-based approach (Ravasz & Barabási, 2003; Mones et al., 2012; De Montis et al., 2007). This study has attempted to combine them into a complementary approach based on a enhanced node-based approach. By complementing the node-based approach, line and link specific characteristics as described by Bagloee and Ceder (2011) are incorporated. The combined approach is done using topological network analysis to examine the function of a node in the network and empirical network analysis to evaluate the flows between the nodes and within the nodes. This study can be seen as innovative in terms of combining the different approaches. Furthermore, by combining topological and empirical elements, the idea of using multiple approaches for this study is underlined.

For the topological network analysis, this study finds two indicators in particular, the clustering coefficient (T. Zhou et al., 2005; Ru & Xu, 2005) and eigenvector centrality (Soh et al., 2010; Majima et al., 2007) most suitable for explaining hierarchy. Other indicators such as betweenness centrality and assortativity are useful in explaining elements of hierarchy but can be problematic in some cases due to valuing nodes equally and ambiguity in interpretation respectively. For both the eigenvector centrality and the clustering coefficient, a topological model representation based on the $\mathbb{P}$-space should be applied (Von Ferber, Holovatch, Holovatch, & Palchykov, 2009) to account for the line based nature of PTN.

For the empirical network analysis, the work of Gentile and Noekel (2016) has been consulted to identify how public transport flows can be modeled. For the relation with hierarchy from a network perspective, especially the transfers of passengers in the PTN provide valuable information. Based on OD matrices for the network, the demand between every pair of nodes is determined which can be seen as a commonly applied approach (e.g. Tamin and Willumsen (1989)). The next issue relates to the assignment of flows for which this study follows a static assignment model (see e.g. (Schweizer et al., 2012)). Static assignment is on the one hand relatively straightforward as it can be determined beforehand but it cannot account for phenomena such as congestion and vehicle crowding. For the assignment of traffic, an 'All or Nothing' (AON) approach is applied. AON is perceived as less realistic than stochastic or probabilistic assignment which are applied for numerous studies (e.g. (Celikoglu & Cigizoglu, 2007)). In this sense, this study is regarded as restricted in the use of traffic assignment models compared to other studies.

Nevertheless, AON is deemed sufficient for this study as it requires less computation time and the routing of passengers is only used for determining transfers. Furthermore, the demand data are aggregated for average daily number of passengers which cannot effectively model the variability for stochastic elements such as vehicle crowding and congestion. Consequently, AON assignment may not be the most realistic method for traffic assignment, but within the context of this rather exploratory study, the advantages relating to its ease of use are regarded as more important than determining the most realistic traffic flow. Throughout the application of the model, this judgment is confirmed by the corresponding verification of the shortest path algorithm of the metric.

The concept of bottlenecks has been explained as a dichotomy between supply and demand. Bottlenecks are an extensively discussed topic (e.g. (Rothengatter, 1996; Witte et al., 2012; Wu et al., 2009)) within the context of PTN. However, this study has attempted to create a new link between bottlenecks and the hierarchical structure of a network. Bottlenecks are connected to hierarchy to some extent as bottlenecks are assumed likelier to be located at hierarchical nodes. This can however, not be confirmed with absolute certainty in this study due to the lack of data on the location of bottlenecks in PTN. The element clearly missing to provide a good understanding of bottlenecks is the supply-side of nodes. Without understanding the capacity of a node, no clear conclusions can be drawn on a node being a bottleneck. Therefore, the identification of bottlenecks is beyond the capacity of the hierarchy metric but understanding why a node is a bottleneck and how it could be solved, are seen as possibilities for its use. This use, to solve bottlenecks, could be a potential application of the metric for policy. In particular the application for policies aiming at relieving centrally located bottlenecks caused by a node being 'too hierarchical' would be an interesting application.

For the definition of hierarchy within a network, a node is appointed as hierarchical if it functions as a transfer location, connects to different type of nodes, is influential in the network and is necessary for connections among directly connected nodes. This approach contains some elements which are comparable to existing studies. First, interactions between hotspots (Bassolas et al., 2019) can partially be compared to determining transfer flows as both incorporate aggregated flows throughout the network to determine hierarchy. However, the approach in the study of Bassolas et al. (2019) incorporates urban indicators and characteristics while this study largely neglects those. Secondly, the approach from Mones et al. (2012) can be compared to the topological influence in a network as both evaluate the impact of a node on the network. One notable difference between these two approaches is that this study evaluates the function of nodes directly connected to a node rather than the gross number of nodes reachable. Lastly, the approach of Ru and Xu (2005) shows similarities in using a clustering coefficient to define the hierarchical structure which has been applied for this metric as well. However, as Ru and Xu (2005) apply the indicator to flight networks, the use of a $\mathbb{P}$-space representation, which is used for this study, makes no sense as a line in flight networks is exceptional.

The development of the metric is done in a unique way combining the topological and empirical approach and thereby combining the three different elements. This approach of combining both methods for network analysis is not found in any of the literature consulted for this study. In comparison to other network indicators the dual approach is seen as both a strength and a weakness of the indicator. On the one hand the use of multiple elements is comprehensive in capturing the hierarchy of a network but on the other hand it is less intuitive. Furthermore, using three separate elements provides a consideration of multiple aspects while their multiplication adds an arbitrary element to the metric. Consequently, the way the metric is developed has its benefits but also its constraints. Despite its limitations, this study has thrived to develop a metric which is capable of evaluating the hierarchy in PTN as accurately as possible. The hierarchical coefficient is also defined in a unique way by combining the inequality of hierarchy in the network and the distribution of nodes based on their hierarchy. As both of these are able to determine the spreading of hierarchy in the network, it is chosen to combine them for the hierarchical coefficient. The way the hierarchical coefficient is calculated, by using the Gini-coefficient and power-law coefficient as exponential power can be seen as another arbitrary choice which has shown clear limitations throughout the results.

For the application of the metric, most of the coding is based on existing coding adjusted for this study. For example, the eigenvector centrality, clustering coefficient and shortest path algorithm have existing codes in Python libraries which just have to be applied to the right network representation to provide the desired output. Moreover, for determining the power-law (Alstott et al., 2014) and Gini-coefficient, existing coding could be applied too. The only newly added feature of coding this study has developed is the adjusted representation of the network, where transfers within nodes can be coded as links with a weight based on the estimated duration of the transfer. This code is inspired from the work of D. Luo et al. (2019).

Throughout the results, the usability of the metric has been evaluated based on the case-studies. By evaluating the outcomes for the case-studies, some interpretations of the results are done. However, as this study is considered an exploratory research and no previous studies have been done about the hierarchy in the PTN of Amsterdam and Rotterdam, the credibility of the metric is hardly assessed in terms of validation. Nevertheless, the results for both of the case-studies and the scenarios showed some plausible yet non-trivial results. This is an indication that the metric provides some reliable but no obvious results and therefore provides additional value. Furthermore, the different outcomes for each element for the case-studies indicate there is clearly a distinction between the aspect of hierarchy captured by each element.

The results for the base year in the Amsterdam PTN showed some interesting results on the structure of the network where the central station scored highest on every element and mostly train stations in suburban regions scored high values, forming a ring like structure around the city center. For the scenarios, the network changed drastically with the opening of the NZL, not only affecting nodes served by this line but nodes all over the network. For the future scenarios some interesting new lines have been evaluated where in general it can be concluded that the more impact a new line has on the hierarchical structure, the higher the implementation costs would be.

The differences between the base-year and the scenarios for the Amsterdam network provide a good indication of how the network structure would change if new lines are added or removed. This could be one of the most important practical applications this study has delivered, providing a new method to support policy for adapting the network structure and quantifying the extent to which the hierarchical structure would change. This method could further be applied to analyze radical changes to the network structure by adding an expensive line. Examples in policy where such radical changes are considered are a ring-line, relieving the central station in Utrecht[1] and an additional rapid PT connection between Eindhoven and Eindhoven Airport[2]. These two scenarios, and numerous others, could be potential applications for the metric to analyze scenarios for the future.

The Rotterdam PTN has been evaluated too, indicating a different hierarchical structure than the Amsterdam PTN despite both networks having the same modalities and being located in the same country. The Rotterdam PTN showed a much more centralized focus with most hierarchical nodes being located in the city center while train stations and transfer hubs in the suburbs generally have a lower hierarchical degree in this network. Despite the hierarchical coefficient of the Rotterdam PTN being higher than for the Amsterdam PTN, the hierarchical structure for the latter appears to be more desirable due to the spatial spreading of the hierarchical nodes, located further from the central station. Consequently, the spatial location of hierarchical nodes has shown to be of significant importance too but it cannot be directly captured by the hierarchical coefficient.

In order to compare the elements of the hierarchical degree to each other as well as comparing the hierarchical degree to other indicators, a correlation analysis is done. This correlation analysis is done for every non-zero scoring node. For the elements and the hierarchical degree, the topological influence and transfer potential appear to be more correlated to the hierarchical degree than the redundancy for both Amsterdam and Rotterdam. This may be an indication that the redundancy of the nodes is less reflected in the hierarchy than the other two indicators. Among the elements, the Rotterdam and Amsterdam network showed no consistent correlation as the topological influence and redundancy are most correlated for the Amsterdam PTN while the redundancy and transfer potential are most correlated for the Rotterdam PTN. However, for both networks, the topological influence and transfer potential are least correlated despite these being most correlated to the hierarchical degree. This finding tells on the one hand that being influential does not necessarily mean a node has transfer potential which can be explained as being connected to influential nodes does not imply transfers are facilitated. On the other hand, it is hard to comprehend why the topological influence and transfer potential would correlate most with the hierarchical degree but least with each other. This could potentially be interpreted as both elements explaining a different aspect of hierarchy, what these are eventually designed to do.

The correlation analysis for the values of the hierarchical degree in the network and other network indicators has shown that there appears to be significant correlation between any of these. For the Amsterdam PTN the betweenness centrality in the $\mathbb{P}$-space (De Bona et al., 2016) and overlapping degree (Aleta et al., 2017) appear to correlate most with the hierarchical degree despite these indicator not directly being used for the calculation of the hierarchical degree. Therefore, it should be

[1]See: `https://www.treinreiziger.nl/utrecht-wil-tweede-intercitystation-en-ov-ringlijn/`

[2]See `https://innovationorigins.com/nl/station-eindhoven-airport-ns-directeur-ziet-er-weinig-heil-in/`

considered why these two indicators are best at explaining hierarchy as perceived by this study. The betweenness centrality in the $\mathbb{P}$-space explains the share of shortest paths between OD pair transferring at a node but does not incorporate the demand of the OD pairs. The overlapping degree differs from the other indicators as it distinguishes between modes for every node. The higher the degree for each mode, the higher the overlapping degree. As this indicator correlates most with the hierarchical degree, there appears to be a relation between the number of modes for a node and its hierarchy. This is, however, not explicitly included in the hierarchical degree and could be used to further refine the metric.

For the Rotterdam PTN, the eigenvector centrality correlates most with the hierarchical degree closely followed by the overlapping degree. Furthermore, for both networks the assortativity, despite being significant, correlates least with with the hierarchical degree, emphasizing the ambiguity of the assortativity in terms of hierarchy. The high correlation between the overlapping degree and hierarchical degree for both PTN is an indication that hierarchical nodes tend to be served by multiple modalities which can be seen as an expected output but so far unconfirmed claim.

For the potential application, the additional value of this study should be seen as an additional tool to assess scenarios but also to provide more insight into bottlenecks and vulnerability. Consequently, the use of the metric could be applied to substantiate policy, in particular on a strategic level regarding long-term adjustments to the network structure. Applications for policy that come to mind are related to network robustness assessments, line development evaluation and bottleneck analysis.

As stated earlier, the metric is not directly capable of identifying bottlenecks without the supply-side perspective, but it can be used to indicate why bottlenecks occur at a specific location in the hierarchical structure or how the situation might be improved to relieve the pressure on the bottleneck. For vulnerability and cascading failures, the hierarchical degree can be used to determine the impact of the disruption of a node on the network. By analyzing the differences in hierarchy between the regular situation and the disrupted situation, an indication of the impact of the disruption can be explored. By knowing this in advance, measures to recover and redistribute the passengers throughout the network could be incorporated more effectively in policy to reduce the impact of a disruption and to increase the robustness of the network.

6.2. Conclusion

Throughout this study, the goal has been to find an answer to the question; 'How can the hierarchy in a public transport network be measured quantitatively and can the hierarchical structure be used to improve the network structure?'. To provide a short answer before analyzing the research question and sub-questions in depth, the developed metric has provided a method to quantitatively measure hierarchy in PTN in terms of the definition for hierarchy for this study. Furthermore, the hierarchical structure is used to evaluate the network by looking at how scenarios would affect the network, comparing structures among networks and by looking at bottlenecks.

For the first sub-question, the definition of a hierarchical structure and the hierarchy of nodes in a PTN has been analyzed. While there have been many preceding studies defining hierarchy, there is not one generally accepted definition for hierarchy in PTN. There are three notable approaches to hierarchy in PTN which are link-, line- and node-based hierarchy. While each of the approaches offers its strengths and weaknesses, this study suggests a combined approach where node-based hierarchy is enhanced with link/line travel demand. Furthermore, this study has defined hierarchy in a PTN based on the different functions of nodes where transfer hubs and more influential nodes should be valued higher in terms of hierarchy. For the hierarchical structure, the hierarchy should be spread in-equally throughout the network and follow a power-law distribution. These definitions follow some of the definitions of previous works but do add additional elements such as the combined approach. Therefore, this study has provided some additional insights in the definition of hierarchy in PTN.

For the second sub-question, the way topological and empirical network analysis can be applied to define a measure for hierarchy in PTN has been analyzed. Following from the combined approach for defining hierarchy, a combination of topological and empirical elements appears to be a sensible choice for defining the metric. Furthermore, by combining different aspects of the definition for hierarchy, the metric has been developed from the elements topological influence, redundancy and transfer potential which cover the aspects of the definition. These three elements have common grounds with previous works, as explained in the previous section, but its combination is unobserved in existing literature. The three elements are combined into one metric by multiplying the values for each element. While the multiplication does decrease the objectiveness of the metric, each of the separate elements can explain only a part of the hierarchy. Hence, multiplication is used to value the combination of each element. This metric has provided a unique method to determine the hierarchy in PTN but requires significant effort and data to compute. Therefore, it may not be the most straightforward metric intuitively but it is able to capture a wide variety of elements of hierarchy.

The third sub-question relates to the application of the metric to the Amsterdam and Rotterdam PTN. Throughout the application, these two cases have been studied and compared extensively. The results have shown some expected results but also notable differences between the network structures. It must be taken into account that the processing of the data, in particular the demand data, offers only limitedly realistic outcomes due to simplifications in the data. Nevertheless, the case-studies have provided some unique and non-trivial insights in the hierarchical structures of the Amsterdam and Rotterdam PTN and based on these findings, the networks are compared and evaluated. While the hierarchical coefficient has been developed to compare the networks as a whole, the spatial location of the hierarchical nodes in the network turned out to be more meaningful for the comparison.

The fourth sub-question relates to the connection between hierarchy and the more practical network issues of bottlenecks and vulnerability. The coupling to bottlenecks and vulnerability with examples from the case-studies has shown some insight in the possible applications of the metric. One thing that should be taken into account, mostly for bottlenecks, is the lack of a supply-side perspective the metric offers. For the assessment of bottlenecks, additional information and analysis is required to which hierarchy can add value. Consequently, on the one hand, hierarchy does provide additional insights for these concepts. However, on the other hand, it requires additional data on the supply-side to properly evaluate bottlenecks. For vulnerability, the metric can provide additional insights into how the evaluate the risks of disruptions and how to substantiate the severity of disruptions to be included in policies for a robust network. In general, the metric has shown significant potential for its application, in particular in assessing and substantiating strategic policy regarding future development for the network structure.

The last of the sub-questions relates to the assessment of scenarios based on the hierarchical degree. Throughout the scenario study for the Amsterdam PTN, both a past network state and future options have been evaluated which clearly showed differences in relation to the base-year. In that sense, the impact of scenarios on the structure can be evaluated. The changes in hierarchical degree for (in)directly affected nodes provides some insight in the effects of the removal or addition of a line to the network. It is, however, debatable if an analysis in the changes in hierarchical degree can determine if a scenario would be desirable or not. The increase in a hierarchy for a network could mean only a few nodes increase in hierarchy (the rich getting richer) while it could also mean the second order nodes could increase in function, relieving bottlenecks. Therefore, only evaluating the changes in hierarchical degree does only provide a limited understanding of the effects of a scenario on the network. A more in depth approach to a scenario in order to evaluate its strategic prospects should be applied for which the hierarchy metric can be used as a part. Nevertheless, the metric can be applied for policy regarding large interventions in the network structure where effects of adjusting, adding or removing lines on the hierarchical structure, should be quantified and evaluated.

In general, this study has provided satisfactory answers to the research question and sub-questions considering the exploratory nature of the subject. The metric developed in this study provides an interesting new perspective on hierarchy and a tool to evaluate it further. Compared to previous research, this study is unique in defining a multi-approach for hierarchy in PTN. Throughout this study, it has been attempted to apply the metric to a small variety of networks and network states but it could potentially be used to numerous networks. Additionally, the metric could be adjusted and altered to include elements such as vehicle crowding and time-of-day. The way the metric is developed offers a lot of additional options depending on the input network. If, for example, a peak hour network has to be evaluated, peak lines and other lines only operating during peak hours should be included and frequencies and vehicles can be adjusted too for this specific time frame. Therefore, the metric does not necessarily evaluate PTN based on the daily demand but it can in potential be used for any temporal context.

6.3. Limitations

There are also some notable limitations to this study which are on the one hand inevitable limitations in modeling PTN but on the other hand also possible subjects for further research. The latter are further and more detailed elaborated in the subsequent section 6.4. First, the definition of hierarchy for this study is attempted to be as inclusive as possible but cannot capture every element of hierarchy. In particular the hierarchy of links and lines is only partially captured as the hierarchical degree is an output for nodes. While the characteristics of links and lines are included in the shortest path algorithm, no output for links is provided. It is presumable that links between hierarchical nodes have a different function than links between one hierarchical node and a non-hierarchical node or between two non-hierarchical nodes but this is not reflected by the output metric. Consequently, a distinction between links and types of links is hardly possible based on this study.

Secondly, the way the metric is defined, and in particular the shortest-path algorithm for the transfer potential, comes with limitations in terms of reliability. Despite the verified coding, the AON assignment method is relatively restricted and less reliable compared to e.g. observed smart-card data. Other simplifications for determining the routes, such as only including scheduled lines, aggregating daily data, averaging speed and frequency per mode and averaging transfer time, are decreasing the truthfulness of the outcome and are therefore limitations of this study. These simplifications are, however, considered as beyond the scope of this study.

Thirdly, the common corridor problem has been acknowledged but not solved for this study. Consequently, the metric could provide incorrect outcomes in a few cases where alternative routing is chosen for the shortest path algorithm due to the absence of shared frequency for links. Considering this study is an exploratory research, these type of limitations are to be expected and could be overcome by improving and refining the way the metric is applied.

Fourthly, the data for each of the cases and scenarios does not comply with the year of the case. For each of the cases, a 2030 demand data-set is used which does not correctly reflect on the contemporary situation. Furthermore, these data are based on calibrated model estimates which limits the generalizations of the output for these specific cases. Therefore, the results for each of the case-studies can only be limitedly applied for analyzing the networks. Nevertheless, the goal of this study has been to apply the metric to networks and has not been to provide the most realistic output for these specific cases. Therefore, determining the most realistic output for a case should be considered as beyond the scope of this study.

Fifthly, the spatial element for networks has been neglected in the methodology for evaluating the networks while this does provide interesting additional insights. Throughout the comparison between the PTN of Amsterdam and Rotterdam, the spread of hierarchical nodes throughout the network has been a striking difference. While the metric does not provide any quantification to assess the spatial distribution of hierarchy, a glimpse on the map for both network would give away notable differences. Consequently, the limited possibilities to assess and quantify the spatial spread of the hierarchical nodes in the network is seen as a limitation of this study

Lastly, the limited number of cases and scenarios combined with the limited period of time for each case has led to meager application of the metric. While it has shown interesting differences in results, more cases and scenarios would significantly improve the interpretation of the results. Connected to this, the practical applications for bottlenecks and vulnerability have only be assessed limitedly where, in particular, cases regarding a known bottleneck would have been an interesting addition. However, due to time constraints, this is perceived as beyond the scope of this study.

6.4. Recommendations

For the recommendations, three categories are identified: recommendations for additional scientific research into the nature of hierarchy in PTN, recommendations regarding the methodology of the metric and recommendations for the input and application of the metric. To clearly separate these different types of recommendations, these are discussed separately throughout this section.

6.4.1 Scientific recommendations

The first scientific recommendation relates to the approach of defining hierarchy in a network. Throughout this study, a predominantly node-based approach is applied where additional insights could be achieved using a link- or line-based approach for hierarchy. Therefore, adjusting or developing an additional metric to measure the hierarchy for these perspectives is seen as an interesting idea for further research. Furthermore, connecting the ideas of node-based hierarchy more to these other approaches could reveal patterns in the way the hierarchy of links, lines and nodes are related. To exemplify this, if a hierarchical and a non-hierarchical node are directly connected, what would that say about the hierarchy of that specific link.

Another scientific recommendation would be to further analyze a multi-layer approach for hierarchical networks which is referred to in this study but the definition for layers are missing and could be an interesting subject for further research. As the values for hierarchical degree are on the continuous spectrum between zero and one, threshold values have to be determined to discretize between layers. For the maps in figures 5.6 and 5.16 this study has identified three levels which are high-hierarchical, low-hierarchical and non-hierarchical. The thresholds values between the different level for this study are arbitrary choices where a theoretical value of 0.5 for each element would be the threshold for a node to be considered as high-hierarchical. A more objective discretization between levels could be done by analyzing the cumulative distribution functions to clearly distinguish functional levels for different layers. This definition for functional hierarchical layers could be an interesting idea for future research.

Throughout this study, the concept of vulnerability has been linked to hierarchy and according to this study, vulnerability could be more intelligible by looking at the hierarchical structure. However, this connection between vulnerability and hierarchy is just a small part of this study and could be the subject of further research. Furthermore, the connection between hierarchy and different PTN related phenomena such as e.g. transit-oriented development could be interesting topics for the implications of the metric for developing strategic network related policy in the future.

The empirical perspective for this study has been somewhat in the back seat as the metric consists of two topological elements and the empirical aspect is only reflected in the transfer potential. Furthermore, for the travel demand, the relatively straightforward AON assignment is applied while more sophisticated traffic assignment methods such as the probabilistic or stochastic assignment may lead to additional insights. Connected to this, the use of observed data, such as smart-card or GPS data could overcome the need for estimated demand models in general. Therefore, further research to how the empirical element of hierarchy could be better reflected is suggested.

6.4.2 Methodological recommendations

For this study, the combination of the different element for hierarchy into the hierarchical degree is done by multiplication. This is considered as an arbitrary choice as there are other possibilities to combine the elements. These include normalizing the element values for all nodes and taking the average, adding up the element values or looking at variations between nodes. Therefore, this study has just explored one of the opportunities to combine the elements and numerous other opportunities are possible. This could be an interesting topic for future research.

The way hierarchical coefficient is defined has proved to be limited in terms of comparing networks. Especially the spatial distribution of hierarchy in the network appeared to be a missing aspect of the hierarchical coefficient to effectively compare networks. Furthermore, the coefficient is able to determine the inequality and distribution in the network but unable to further specify how the differences should be interpreted. If a change to a network structure would affect the hierarchical coefficient, it would be hard to assess if this would be desirable based on the hierarchical coefficient alone. Therefore, it is suggested further research is done how the hierarchical coefficient could be more comprehensible in terms of explaining the hierarchical structure of a PTN and also incorporate a spatial element.

The common corridor problem has been identified in the application but no solution for this problem has been found. Therefore, the limitations of this feature have been accepted within the scope of this study. It it, however, recommended to further study how this problem can be overcome in such a way that a shared frequency is used in a common corridor but cannot be used if the destination is located after the common corridor ends[3].

Another element that is missing for the metric and has been mentioned briefly above is a spatial element to indicate how the nodes are spread throughout the network. While the results clearly indicate this element is important to assess networks, this study has not defined any spatial indicator for networks. A spatial element for network hierarchy could be developed in multiple ways for future research. First, an obvious but rather limited method to incorporate spatial spreading would be to include the average distance between the five or ten most hierarchical nodes. This would be a very intuitive method to determine the spread of nodes in the network. To compensate for networks with different sizes, the distance could be divided by the maximum distance between any node pair in the network. A more potentially worthy approach but also arbitrary approach would be to determine the average distance between level 1 node pairs as a fraction of the average distance between

[3] In its essence this problem sounds relatively straightforward but for more complex structures with e.g. partially combined lines sharing a common corridor with other partially combined lines this would become difficult quite fast

each node pair. The arbitrary element in this method for incorporating a spatial element is based on the threshold value for level 1 nodes, which is, at least for this study, an arbitrarily chosen value. This method would, however, provide an intuitive coefficient for the spreading of hierarchy and is independent of network size as distance shares are used.

6.4.3 Input and application recommendations

One very time consuming problem this study has consistently encountered is the lack of generic coding in data-sets. Throughout the different data-sets (NDOV, GTFS, VENOM and V-MRDH) there is generally no consistency in the coding of nodes. Therefore, linking the nodes from different data-sets involves a lot of manual work by coupling the coding from one database to another. By applying an additional and generic way to code the nodes, preferably based on the latitude and longitude of a node, the databases could maintain their own coding but add a property containing the unique generic code. Consequently, this multi-data-set analysis would significantly be less time consuming for comparable studies in the future.

Another recommendation regarding the application is the implementation of different network states based on the time of day. Besides different demand for each period of the day, the network structure could differ by appending peak lines during peak hours and night lines for the night. Consequently, the network would not consist of only the lines operating consistently throughout the day but also include lines dedicated to a specific period. The network layout and hierarchy of the PTN could vary significantly throughout the day. This feature of varying hierarchy is not captured by aggregating over a day and would capture more of the time related dynamics in travel behavior. Therefore, this would be a very interesting subject for further research.

Connected to the previous recommendation, additional types of lines including night-lines and peak lines but also international lines and ferries could potentially add interesting new insights. For this study, these types of lines were perceived beyond the scope but for different networks such as e.g. national train-networks, international trains could be seen as crucial in understanding the hierarchy in the network. One issue regarding these type of lines is the irregularity these often operate with. As the frequency of a line is used to determine the shortest path, lines with a very low frequency would not be chosen as fastest transfer. This is one issue that should be taken into account when adding irregular or low frequent lines. Furthermore, for each of the modes, average frequencies and speeds are used, by defining these for each line specific, more realistic results could be obtained.

Throughout this study, the metric has been applied to one type of PTN, the metropolitan area of a Dutch city. It would be an interesting option for future research to apply the metric to different sized networks and to networks in different countries. The hierarchical structure in smaller networks could vary significantly from the metropolitan networks while foreign networks could show different patterns too. Two specific relevant Dutch cases where the network structure is a current issue are the cities of Utrecht and Eindhoven. Both of these cities, as mentioned before, are considering an expansion of the current network structure towards a more 'wheel with spokes' structure with a ring line around the central node. The metric could be applied to determine if such a structure would change the structure for these networks in a desired way. Therefore, it could be a substantiation for the policy regarding the development and adjustment of the PTN structure. Furthermore, the application of the metric to different types of network, such as a national PTN or aviation network rather than a city PTN could be another interesting application of the metric.

For this study, only a few scenarios have been studied and only for the Amsterdam PTN. It would be recommendable to further apply the metric to different scenarios to get a better understanding how changes in the network affect the hierarchical structure. Aligning this with the cities mentioned in the previous paragraphs, doing a scenario study for the Utrecht and Eindhoven network would be an excellent way to further apply the metric and to test the scenarios for the future structure of the network. Furthermore, for scenarios in particular, it would be advisable to evaluate plausible scenarios for the future in order to understand how findings for scenarios could be used as substantiation for policy.

In this study the concept of bottlenecks has been linked to hierarchy. However, a case where one node would clearly be a bottleneck in the network, has not been studied. This could be an interesting topic for further research to exemplify how bottlenecks can be evaluated with the metric for hierarchy. The same applies to the vulnerability where it would be recommended to study a case in which the a vulnerable node is comprehensively studied. Within such a study, the possibilities to apply the hierarchical metric for different uses are better evaluated.

All in all, there are plenty of recommendations for future research as this has just been an exploration of the possibilities of quantifying hierarchy. It makes therefore sense to further apply and evaluate the metric to accurately study the implications of hierarchical structures in PTN.

References

Ahn, Y.-Y., Bagrow, J. P., & Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *nature*, *466*(7307), 761.

Aleta, A., Meloni, S., & Moreno, Y. (2017). A multilayer perspective for the analysis of urban transportation systems. *Scientific reports*, *7*, 44359.

Alstott, J., Bullmore, E., & Plenz, D. (2014). powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one*, *9*(1), e85777.

Angeloudis, P., & Fisk, D. (2006). Large subway systems as complex networks. *Physica A: Statistical Mechanics and its Applications*, *367*, 553–558.

Bagloee, S. A., & Ceder, A. A. (2011). Transit-network design methodology for actual-size road networks. *Transportation Research Part B: Methodological*, *45*(10), 1787–1804.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, *286*(5439), 509–512.

Barabási, A.-L., Ravasz, E., & Vicsek, T. (2001). Deterministic scale-free networks. *Physica A: Statistical Mechanics and its Applications*, *299*(3-4), 559–564.

Barthelemy, M. (2004). Betweenness centrality in large complex networks. *The European physical journal B*, *38*(2), 163–168.

Bassolas, A., Barbosa-Filho, H., Dickinson, B., Dotiwalla, X., Eastham, P., Gallotti, R., ... others (2019). Hierarchical organization of urban mobility and its connection with city livability. *Nature Communications*, *10*(1), 1–10.

Berche, B., Von Ferber, C., Holovatch, T., & Holovatch, Y. (2009). Resilience of public transport networks against attacks. *The European Physical Journal B*, *71*(1), 125–137.

Bettinardi, R. G. (2016). *Spontaneous brain activity: how dynamics and topology shape the emergent correlation structure.*

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, *30*(2), 136–145.

Cats, O. (2017). Topological evolution of a metropolitan rail transport network: The case of stockholm. *Journal of Transport Geography*, *62*, 172–183.

Cats, O., & Jenelius, E. (2015). Planning for the unexpected: The value of reserve capacity for public transport network robustness. *Transportation Research Part A: Policy and Practice*, *81*, 47–61.

Celikoglu, H. B., & Cigizoglu, H. K. (2007). Modelling public transport trips by radial basis function neural networks. *Mathematical and computer modelling*, *45*(3-4), 480–489.

Chen, L., Pel, A. J., Chen, X., Sparing, D., & Hansen, I. A. (2012). Determinants of bicycle transfer demand at metro stations: Analysis of stations in nanjing, china. *Transportation Research Record*, *2276*(1), 131–137.

Corominas-Murtra, B., Rodríguez-Caso, C., Goni, J., & Solé, R. (2011). Measuring the hierarchy of feedforward networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *21*(1), 016108.

Criado, R., Flores, J., Hernández-Bermejo, B., Pello, J., & Romance, M. (2005). Effective measurement of network vulnerability under random and intentional attacks. *Journal of Mathematical Modelling and Algorithms*, *4*(3), 307–316.

de Volkskrant. (2019). *Noord-zuidlijn rijdt na investering van 3 miljard door tot schiphol, aldus sector en regio.* Retrieved from `https://www.volkskrant.nl/nieuws-achtergrond/noord-zuidlijn-rijdt-na-investering-van-3-miljard-door-tot-schiphol-aldus`

-sector-en-regio~b37a54a1/?referer=https%3A%2F%2Fwww.google.com%2F (accessed: 03-02-2020)

De Bona, A., Fonseca, K., Rosa, M., Lüders, R., & Delgado, M. (2016). Analysis of public bus transportation of a brazilian city based on the theory of complex networks using the p-space. *Mathematical Problems in Engineering, 2016*.

De Montis, A., Barthélemy, M., Chessa, A., & Vespignani, A. (2007). The structure of interurban traffic: a weighted network analysis. *Environment and Planning B: Planning and Design, 34*(5), 905–924.

Derrible, S. (2012). Network centrality of metro systems. *PloS one, 7*(7), e40575.

Derrible, S., & Kennedy, C. (2009). Network analysis of world subway systems using updated graph theory. *Transportation Research Record, 2112*(1), 17–25.

Derrible, S., & Kennedy, C. (2011). Applications of graph theory and network science to transit network design. *Transport reviews, 31*(4), 495–519.

Dong, Y., Yang, X., & Chen, G. (2014). Robustness analysis of layered public transport networks due to edge overload breakdown. *International Journal of Information Technology & Computer Science, 6*(3), 30–37.

EMTA. (2019). *Paris ile-de-france : Public transport networks.* Retrieved from https://www.emta.com/spip.php?article81&lang=en (accessed: 09-12-2019)

Eum, S., Arakawa, S., & Murata, M. (2008). A new approach for discovering and quantifying hierarchical structure of complex networks. In *Fourth international conference on autonomic and autonomous systems (icas'08)* (pp. 182–187).

Eurostat. (2019). *Passenger transport statistics.* Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Passenger_transport_statistics&oldid=271714 (accessed: 02-09-2019)

Fiorenzo-Catalano, M. S. (2007). Choice set generation in multi-modal transportation networks. *Trail*.

Fiorenzo-Catalano, M. S., Van Nes, R., & Bovy, P. H. (2004). Choice set generation for multi-modal travel analysis. *European journal of transport and infrastructure research EJTIR, 4 (2)*.

Flötteröd, G. (2015). *Traffic assignment for strategic urban transport model systems.*

Gallotti, R., & Barthelemy, M. (2015). The multilayer temporal network of public transport in great britain. *Scientific data, 2*, 140056.

Gattuso, D., & Miriello, E. (2005). Compared analysis of metro networks supported by graph theory. *Networks and Spatial Economics, 5*(4), 395–414.

Gentile, G., & Noekel, K. (2016). Modelling public transport passenger flows in the era of intelligent transport systems. *Gewerbestrasse: Springer International Publishing*.

Gomez, S., Diaz-Guilera, A., Gomez-Gardenes, J., Perez-Vicente, C. J., Moreno, Y., & Arenas, A. (2013). Diffusion dynamics on multiplex networks. *Physical review letters, 110*(2), 028701.

Gu, Q., Yang, X.-H., Wang, W.-L., & Wang, B. (2008). Research on urban public transport networks based on complex networks [j]. *Computer Engineering, 20*.

GVB. (2019). *Lijnenkaart 2019.* Retrieved from https://www.gvb.nl/sites/default/files/lijnenkaart2019-2.pdf (accessed: 01-11-2019)

Hu, Y., & Zhu, D. (2009). Empirical analysis of the worldwide maritime transportation network. *Physica A: Statistical Mechanics and its Applications, 388*(10), 2061–2071.

Huang, A., Zhang, H. M., Guan, W., Yang, Y., & Zong, G. (2015). Cascading failures in weighted complex networks of transit systems based on coupled map lattices. *Mathematical Problems in Engineering, 2015*.

Janson, B. N. (1995). Network design effects of dynamic traffic assignment. *Journal of Transportation Engineering, 121*(1), 1–13.

Jian, G., Peng, Z., Chengxiang, Z., & Hui, Z. (2012). Research on public transit network hierarchy based on residential transit trip distance. *Discrete Dynamics in Nature and Society*, *2012*.

Jiancong, W., Shaokuan, C., Yuqiang, H., & Liping, G. (2006). Simulation of transfer organization of urban public transportation hubs. *Journal of Transportation Systems Engineering and Information Technology*, *6*(6), 96–102.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of complex networks*, *2*(3), 203–271.

Lami, I. M. (2014). *Analytical decision-making methods for evaluating sustainable transport in european corridors* (Vol. 11). Springer.

Lämmer, S., Gehlsen, B., & Helbing, D. (2006). Scaling laws in the spatial structure of urban road networks. *Physica A: Statistical Mechanics and its Applications*, *363*(1), 89–95.

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical review letters*, *87*(19), 198701.

Lee, M., Barbosa, H., Youn, H., Holme, P., & Ghoshal, G. (2017). Morphology of travel routes and the organization of cities. *Nature communications*, *8*(1), 2229.

Levinson, D. (2012). Network structure and city size. *PloS one*, *7*(1), e29721.

Levinson, D. M., Xie, F., & Zhu, S. (2007). The co-evolution of land use and road networks. *Transportation and traffic theory*, 839–859.

Li, D., Fu, B., Wang, Y., Lu, G., Berezin, Y., Stanley, H. E., & Havlin, S. (2015). Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences*, *112*(3), 669–672.

Little, R. G. (2002). Controlling cascading failure: Understanding the vulnerabilities of interconnected infrastructures. *Journal of Urban Technology*, *9*(1), 109–123.

Lu, H., & Shi, Y. (2007). Complexity of public transport networks. *Tsinghua Science and Technology*, *12*(2), 204–213.

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.

Luo, J., Whitney, D. E., Baldwin, C. Y., & Magee, C. L. (2009). *Measuring and understanding hierarchy as an architectural element in industry sectors.* Harvard Business School.

Majima, T., Katuhara, M., & Takadama, K. (2007). Analysis on transport networks of railway, subway and waterbus in japan. In *Emergent intelligence of networked agents* (pp. 99–113). Springer.

Mattsson, L.-G., & Jenelius, E. (2015). Vulnerability and resilience of transport systems–a discussion of recent research. *Transportation Research Part A: Policy and Practice*, *81*, 16–34.

Min, B., Do Yi, S., Lee, K.-M., & Goh, K.-I. (2014). Network robustness of multiplex networks with interlayer degree correlations. *Physical Review E*, *89*(4), 042811.

Mishra, S., Welch, T. F., & Jha, M. K. (2012). Performance indicators for public transit connectivity in multi-modal transportation networks. *Transportation Research Part A: Policy and Practice*, *46*(7), 1066–1085.

Mones, E., Vicsek, L., & Vicsek, T. (2012). Hierarchy measure for complex networks. *PloS one*, *7*(3), e33799.

Murray, A. T., Matisziw, T. C., & Grubesic, T. H. (2008). A methodological overview of network vulnerability analysis. *Growth and Change*, *39*(4), 573–592.

Newman, M. E. (2002). Assortative mixing in networks. *Physical review letters*, *89*(20), 208701.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167–256.

Nielsen, O. A., & Jovicic, G. (1999). A large scale stochastic timetable-based transit assignment model for route and sub-mode choices. In *Transportation planning methods. proceedings of seminar f, european transport conference, 27-29 september 1999, cambridge, uk.* (Vol. 434).

Nystuen, J. D., & Dacey, M. F. (1961). A graph theory interpretation of nodal regions. In *Papers of the regional science association* (Vol. 7, pp. 29–42).

OECD. (2019). *Freight transport.* Retrieved from `https://data.oecd.org/transport/freight-transport.htm` (accessed: 02-09-2019)

Porta, S., Strano, E., Iacoviello, V., Messora, R., Latora, V., Cardillo, A., ... Scellato, S. (2009). Street centrality and densities of retail and services in bologna, italy. *Environment and Planning B: Planning and design*, *36*(3), 450–465.

Qing, X., Zhenghu, Z., Zhijing, X., Zhang, W., & 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*(1), 193–198.

Ravasz, E., & Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Physical review E*, *67*(2), 026112.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551–1555.

Reggiani, A., Nijkamp, P., & Lanzi, D. (2015). Transport resilience and vulnerability: The role of connectivity. *Transportation research part A: policy and practice*, *81*, 4–15.

Reynolds-Feighan, A. (2001). Traffic distribution in low-cost and full-service carrier networks in the us air transportation market. *Journal of Air Transport Management*, *7*(5), 265–275.

Richards, B. (2012). *Future transport in cities.* Taylor & Francis.

Rodrigue, J.-P., Comtois, C., & Slack, B. (2016). *The geography of transport systems.* Routledge.

Rodríguez-Núñez, E., & García-Palomares, J. C. (2014). Measuring the vulnerability of public transport networks. *Journal of transport geography*, *35*, 50–63.

Roth, C., Kang, S. M., Batty, M., & Barthelemy, M. (2012). A long-time limit for world subway networks. *Journal of The Royal Society Interface*, *9*(75), 2540–2550.

Rothengatter, W. (1996). Bottlenecks in european transport infrastructure. In *Pan-european transport issues. proceedings of seminar a held at the 24th european transport forum, brunel university, england, 2-6 september 1996. volume p401.*

Ru, W., & Xu, C. (2005). Hierarchical structure, disassortativity and information measures of the us flight network. *Chinese Physics Letters*, *22*(10), 2715.

Scheurer, J., & Porta, S. (2006). Centrality and connectivity in public transport networks and their significance for transport sustainability in cities. In *World planning schools congress, global planning association education network,*.

Schmöcker, J.-D., Bell, M. G., & Lam, W. H. (2004). Importance of public transport. *Journal of Advanced Transportation*, *38*(1), 1–4.

Schweizer, J., Danesi, A., Rupi, F., & Traversi, E. (2012). Comparison of static vehicle flow assignment methods and microsimulations for a personal rapid transit network. *Journal of advanced transportation*, *46*(4), 340–350.

Sienkiewicz, J., & Hołyst, J. A. (2005). Statistical analysis of 22 public transport networks in poland. *Physical Review E*, *72*(4), 046127.

Soh, H., Lim, S., Zhang, T., Fu, X., Lee, G. K. K., Hung, T. G. G., ... Wong, L. (2010). Weighted complex network analysis of travel routes on the singapore public transportation system. *Physica A: Statistical Mechanics and its Applications*, *389*(24), 5852–5863.

Stephen, A. T., & Toubia, O. (2009). Explaining the power-law degree distribution in a social commerce network. *Social Networks*, *31*(4), 262–270.

Sumalee, A., Tan, Z., & Lam, W. H. (2009). Dynamic stochastic transit assignment with explicit seat allocation model. *Transportation Research Part B: Methodological*, *43*(8-9), 895–912.

Tamin, O., & Willumsen, L. (1989). Transport demand model estimation from traffic counts. *Transportation*, *16*(1), 3–26.

Transport for London. (2015). *Annual report.* Retrieved from `https://tfl.gov.uk/`

info-for/media/press-releases/2015/june/record-passenger-numbers-on-london-s-transport-network (accessed: 09-12-2019)

Trusina, A., Maslov, S., Minnhagen, P., & Sneppen, K. (2004). Hierarchy measures in complex networks. *Physical review letters*, *92*(17), 178702.

Van Mieghem, P., Wang, H., Ge, X., Tang, S., & Kuipers, F. A. (2010). Influence of assortativity and degree-preserving rewiring on the spectra of networks. *The European Physical Journal B*, *76*(4), 643–652.

Van Nes, R. (2002). Design of multimodal transport networks: A hierarchical approach. *Citeseer*.

Van Oort, N., Sparing, D., Brands, T., & Goverde, R. M. (2015). Data driven improvements in public transport: the dutch example. *Public transport*, *7*(3), 369–389.

Vázquez, A., Pastor-Satorras, R., & Vespignani, A. (2002). Large-scale topological and dynamical properties of the internet. *Physical Review E*, *65*(6), 066130.

Von Ferber, C., Holovatch, T., & Holovatch, Y. (2009). Attack vulnerability of public transport networks. In *Traffic and granular flow'07* (pp. 721–731). Springer.

Von Ferber, C., Holovatch, T., Holovatch, Y., & Palchykov, V. (2007). Network harness: Metropolis public transport. *Physica A: Statistical Mechanics and its Applications*, *380*, 585–591.

Von Ferber, C., Holovatch, T., Holovatch, Y., & Palchykov, V. (2009). Public transport networks: empirical analysis and modeling. *The European Physical Journal B*, *68*(2), 261–275.

Wang, J., Mo, H., Wang, F., & Jin, F. (2011). Exploring the network structure and nodal centrality of china's air transport network: A complex network approach. *Journal of Transport Geography*, *19*(4), 712–721.

Witte, P. A., Wiegmans, B. W., van Oort, F. G., & Spit, T. J. (2012). Chokepoints in corridors: Perspectives on bottlenecks in the european transport network. *Research in Transportation Business & Management*, *5*, 57–66.

Wu, J., Gao, Z., & Sun, H. (2009). Topological-based bottleneck analysis and improvement strategies for traffic networks. *Science in China Series E: Technological Sciences*, *52*(10), 2814–2822.

Xia-Miao, L., Ming-Hua, Z., Jin, Z., & Ke-Zan, L. (2010). Hierarchy property of traffic networks. *Chinese Physics B*, *19*(9), 090510.

Xie, F., & Levinson, D. (2009). Topological evolution of surface transportation networks. *Computers, Environment and Urban Systems*, *33*(3), 211–223.

Yerra, B. M., & Levinson, D. M. (2005). The emergence of hierarchy in transportation networks. *The Annals of Regional Science*, *39*(3), 541–553.

Yu, B., Yang, Z.-Z., Jin, P.-H., Wu, S.-H., & Yao, B.-Z. (2012). Transit route network design-maximizing direct and transfer demand density. *Transportation Research Part C: Emerging Technologies*, *22*, 58–75.

Zappa, P., & Lomi, A. (2016). Knowledge sharing in organizations: A multilevel network analysis. In *Multilevel network analysis for the social sciences* (pp. 333–353). Springer.

Zhang, G.-Q., Cheng, S.-Q., & Zhang, G.-Q. (2012). A universal assortativity measure for network analysis. *arXiv preprint arXiv:1212.6456*.

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*(6), 1538–1546.

Zhang, L. (2007). Traffic diversion effect of ramp metering at individual and system levels. *Transportation Research Record*, *2012*(1), 20–29.

Zhang, S., Derudder, B., & Witlox, F. (2013). The impact of hub hierarchy and market competition on airfare pricing in us hub-to-hub markets. *Journal of Air Transport Management*, *32*, 65–70.

Zhou, T., Yan, G., & Wang, B.-H. (2005). Maximal planar networks with large clustering coefficient and power-law degree distribution. *Physical Review E*, *71*(4), 046141.

Zhou, Y., Wang, J., & Yang, H. (2019). Resilience of transportation systems: concepts and comprehensive review. *IEEE Transactions on Intelligent Transportation Systems*.

A

Data cleaning

The cleaning of the data is done to ensure it provides the desired input and outputs for the metric. As the data are obtained from different databases, it is important to ensure the data align well. As various data-sets require different methods for cleaning these are elaborated separately. Based on the documentation in this appendix, the case-study outputs should be reproducible. As mentioned before, all of the programming codes and cleaned data-sets are available on `https://github.com/Abel287/MasterThesis.git`.

A.1. Amsterdam network

Throughout this section, the data cleaning for the Amsterdam PT network is elaborated. For the topological structure of the network, different sources are used for different modalities. For the rail network, a 2019 rail map is used[1] for which the lines are manually added. For the PT in the city operated by the GVB, a NDOV data set from August 2019 is used[2] as this was the most recent complete data set. For the latitude and longtitude coordinates of the stops, a GTFS Ovapi data-set from 2018 is used[3] as this is the most recent one available containing coordinate data. For the regional buses, the coding of the nodes is done differently in the GTFS Ovapi data-set. Therefore, the IDs are manually changed and the coordinates are linked to the new IDs.

A.1.1 Cleaning of the GVB data-set

The cleaning of the data-set has been done in an order which is followed throughout this section. This sequence of steps is required to obtain the same output as this study. Furthermore, the current schedule of the GVB (2019) is used to validate the data. Based on temporal obstructions, some minor adjustments are made which would not significantly affect the outputs.

Same node occurs twice on a line

Some nodes are not connected to any other node as these are occurring twice on the same line and the code does not allow for nodes with the same label on one line. In general, this is going to be filtered out when nodes with the same label are merged later on but there could be cases in which this would not be desired. The duplicates of this data-set which are considered are:

[1] Retrieved from: https://nieuws.ns.nl/download/611967/spoorkaart2019-231351.pdf

[2] Data-set *KV1_GVB_1629_1.zip* retrieved from: http://data.ndovloket.nl/gvb/

[3] Data-set *gtfs-kv1gvb-20180621.zip* retrieved from: http://gtfs.ovapi.nl/gvb/

- Baden Powellweg (Line 63): two different nodes on the same line with other nodes in between, having the same name. Added characteristic north (N) to the second node (4024) which does lead no additional issues as this node is only served by line 63 (GVB, 2019).

Same label, should not be merged

Additionally, some nodes do have the same label but are not the same node. To provide a clarifying example, the node "Prinsengracht" is used for multiple nodes on lines crossing the canal. As these different lines cross the canal at a different location, the nodes should not be considered as one despite the similar name. If this phenomena occurs, to which the Amsterdam network has a habit of, their labels are altered based on their relative location. The alterations are named after compass points being north (N), east (E), south (S) and west (W) respectively. For nodes which should have three different labels a middle (M) category is added. This is done to provide every node which should not be merged with a unique node-label. To emphasize which node gets which label, the node ID's are provided. Some nodes are merged with other nodes which is further elaborated subsequently. The modified nodes are:

- Baden Powellweg (4011 merged with "Tussen Meer" node, 4409 & 4410 become S, 4024 becomes N, as explained earlier
- Hoekenes (4015 N, 4412 S)
- Johan Huizingalaan (4063 N, other node ID's S)
- Derkinderenstraat (4130 N, 4061 S)
- J.P. Heijestraat (6049 S, 6016 N)
- Amstelveenseweg (7034 changed to Zeilstraat)
- Prinsengracht (8063 E, 6071 W)
- Keizersgracht (8061 E, 6073 W)
- Dam (5032 E, other node ID's W)
- Wibautstraat(9020 N, 9512 S)
- Maasstraat(9161 S, 9068 & 9069 M, 9079 N)
- Waalstraat(9083 N other node ID's)
- Prinses Irenestraat (7130 W, 7187 & 7186 E)
- Jan van Galenstraat (3026 W, 2184 & 2222 E, 3169 renamed to match with metro node written as "Jan v.Galenstraat")
- Bilderdijkstraat(6037 merged with "De Clercqstraat", 6021 merged with "Kinkerstraat")
- Dr. H. Colijnstraat(3123 N, 3100 M, 3104 S)
- Jan Tooropstraat (4131 merged with "Postjesweg")
- Postjesweg (6015 & 6115 merged with "Hoofdweg")
- Kattenburgerstraat (8236 merged with "Jan Schaeferbrug"
- Winkelcentrum (985 added Diemen, 1328 added Noord to indicate where the shopping malls are located)

Different label, should be merged

In order to figure which nodes are located next to each other but with different names, a distance algorithm based on the latitude and longitude of nodes is used[4]. For the nodes which should be merged, the label is adjusted to either of the two existing labels. For nodes that would apply to being merged but have a reason to not being merged the reasoning is explained. Based on the algorithm, the following nodes would apply to being merged as these stops are less than 150 meters apart:

- Johan Huizingalaan(N) & Cornelis Lelylaan
- Surinameplein & Curaçaostraat - removed the latter node as this only the terminal station for one line located next to the former station
- 1e Con. Huygensstraat & Overtoom
- Beukenweg & Beukenplein
- Linnaeusstraat & Wijttenbachstraat
- Laan v.Vlaanderen & Antwerpenbaan
- Louwesweg & Louweshoek
- Van Woustraat & Ceintuurbaan
- Hugo de Grootplein & G. v.Ledenberchstraat
- Nw. Willemsstraat & Nassaukade - are not merged as these are separated by a canal
- Drentepark & Station RAI
- Drentepark & Europaboulevard
- De Boelelaan/VU & De Boelelaan
- Stadionweg & Beethovenstraat
- Van Hallstraat & V.d. Hoopstraat
- Adm. de Ruijterweg & Jan Evertsenstraat
- Hugo de Vrieslaan & Middenweg
- De Sav. Lohmanstraat & Troelstralaan
- Floraweg & Slijperweg
- Floraweg & Klaprozenweg
- Slijperweg & Klaprozenweg
- Meidoornplein & Hagedoornplein - are not merged as these are nodes on the same line
- Science Park & Aer Science Park
- Annie Romeinplein & Gooioord - are not merged as the lines serving these stops were running together before this stop and split here
- Station Diemen Zuid & Station Diemen-Zuid
- Station Reigersbos & Reigersbos

[4]Retrieved from: https://stackoverflow.com/questions/19412462/getting-distance-between-two-points-based-on-latitude-longitude

- Station Gein & Gein
- Gaasperplas (uitstap) & Gaasperplas
- Henk Sneevlietweg & Ottho Heldringstraat
- Europaplein & Dintelstraat
- Oostenburgergracht & 1e Coehoornstraat
- Hoogte Kadijk & 1e Coehoornstraat - are not merged as these are nodes on the same line
- Burg. Röellstraat & Slotermeerlaan
- Pieter Calandlaan & Baden Powellweg(S)
- Rozengracht & Marnixstraat
- Raadhuis & Raadhuis/Dorpsstraat
- Handweg & Maalderij - are not merged as these are nodes on the same line
- Binnenhof & Stadstuinen - are not merged as these are nodes on the same line
- Diemerbrug & Beukenhorst - are not merged as these are nodes on the same line
- Keizer Karelplein & Hueseplein

Merging of nodes with the same label

Following from the previous modifications, any nodes with the same label should now be considered as one node. Therefore, it is checked which node ID's have the same label and these are merged subsequently. As any nodes with the same label which should not be merged and any nodes which should be merged but have different labels have been identified, these are left out or included respectively. Following, a graph in which only unique node-labels are present in which nodes with different labels being at (nearly) the same location are considered as one.

Miscellaneous

- Line 18 was unjustly send to the stop "Fred. Hendrikplants." as this stop is temporarily unavailable[5]. Adjusted to the stop "Bloemgracht".
- The route of line 15 was outdated and has been adjusted to match the new route.

A.1.2 Cleaning of the NS data

As the train stations outside the area of study are considered external nodes, only direct links from the stops in the network are considered. Therefore, only limited additional train stations are added, which is done manually. To ensure these nodes are not considered for the hierarchical degree, as these are beyond the scope of this study, these nodes are given an identifiable node ID in the range 9900 - 9999. The train lines are added using these external nodes and the node ID's of train stations in the network on the route. When the train line leaves the PTN area considered, the next node on the line is considered as the external node.

[5]Based on the November schedule (GVB, 2019) to validate the model

A.1.3 Cleaning of the regional bus data

As explained in 4.3, there are four different operators for the regional buses. Unfortunately, the coding of links and lines does not match the GVB data in the NDOV data-sets which implies the additional lines, links and nodes have to be added manually. The manually added nodes have a node ID between 9800 and 9899 and are predominantly located in the suburbs such as Amstelveen and Badhoevedorp. This makes sense as the regional buses are used to travel from the city through suburbs to neighboring cities and villages. As the regional buses leave the PTN at one point in every case, border nodes are added after which the line is further linked to the destination of the regional bus. Consequently, any nodes between the border node and the external destination node are excluded from the data. Based on figure 4.3, the cutting of the regional lines is executed in which the bordering nodes are the nodes just outside the PTN, given a node ID between 9900 and 9999 similar to the external train stations. Terminal bus-stations outside the PTN are given a node ID in this range as well. Therefore, any node with a node ID between 9900 and 9999 can be seen as an external node for this network. Regional buses with a nearly similar route but ending at a different terminal station at an external node[6] are combined as these would be considered similar for this study. If a partially external line passes stops excluded from the data outside the boundary of the network, a time penalty is included to compensate for the stopping time of these nodes. Furthermore, existing lines between external nodes are added to prevent unjustly sending the demand between those nodes through the network.

A.1.4 Cleaning of demand data

For the demand data, a VENOM data-set is used to provide OD matrices based on OmniTRANS. The data-set can either be based on 2012 or 2030 data out of which the latter is preferred due to the inclusion of the NZL. However, due to changes to the network structure throughout the period between 2020 and 2030, the set of nodes shows notable differences compared to the NDOV and GTFS data. Furthermore, as the node IDs in this data-set are different from the previous data-sets, the IDs have to be aligned while taking the adjusting to the NDOV data-set into account. Since, the label and location of some stops has been altered in the 2030 network these are either coupled to existing nodes or left out of the demand. As there is no good method to automatically couple the nodes from both data-sets[7], this is done manually. Since there is no real structure in doing this, nodes with the same label in both data-sets, where changes to the location of the node are unlikely beyond reasonable doubt, are coupled. If there is any doubt for a node regarding new lines, adjusted node labels or ambiguity in terms of the label (e.g hospital, mall or bus station) the location of the node is validated in QGIS and coupled to the appropriate node if it exists. This method is not considered totally foolproof but as there are 1300 nodes in the VENOM data-set it is regarded as sufficient within the time constraints of this research.

Once the nodes from the different data-sets are coupled, the demand for each node pair in the initial set can be determined. To prevent any odd values for the demand, it is rounded to integers. For any node pairs with a demand of more than zero, the shortest paths are determined to assign the demand to a specific route.

[6] e.g. bus 301 and 307 both travel from Amsterdam-Noord to Purmerend where each goes to a different part of Purmerend which is considered as one external node

[7] Mainly due to different ID coding, different spelling in node labels, changes to the previous data-set and the addition, removal or replacement of nodes

A.1.5 Selection of internal and external nodes

Based on the considered lines, the selected internal nodes can be clarified. In the following overview, the lines for each modality are elaborated in which the boundary nodes are outlined if the line leaves the PTN at some point.

Internal nodes

Any nodes on the following train lines (NS):

- Zaandam – Schiphol Airport
- Zaandam – Utrecht Centraal
- Halfweg Zwanenburg – Amsterdam Centraal
- Haarlem – Amsterdam Centraal
- Schiphol Airport – Weesp (Via Central Station)
- Schiphol Airport – Weesp (Via Zuid)
- Schiphol Airport – Hilversum
- Schiphol Airport– Utrecht Centraal
- Schiphol Airport – Amsterdam Centraal
- Hilversum – Amsterdam Centraal
- Almere Centrum – Amsterdam Centraal
- Abcoude – Zaandam

and four additional train lines between external nodes.

Any nodes on the following tram lines (GVB):

- 1, 2, 3, 4, 5, 7, 11, 12, 13, 14, 17, 19, 24, 26

Any nodes on the following metro lines (GVB):

- 50, 51, 52, 53, 54

Any nodes on the following bus lines (GVB):

- 15, 18, 21, 22, 34, 35, 36, 37, 38, 40, 41, 44, 47, 48, 55, 61, 62, 63, 65, 66
- 49 until Reigersbroeck
- 69 until Schiphol Airport

Any nodes on the following regional buses (Conexxion, EBS, Syntus/Keolis):

- 80 until Oranje Nassaustraat
- 171 until Aalsmeerderweg 481

- 174, 347, 348 until Arthur van Schendellaan
- 186 until Nieuwe Meerlaan
- 194, 195, 300, 341, 397 until Schiphol Airport
- 199 until Camping Amsterdamse Bos
- 301, 305, 312, 314, 315, 316 until Schouw)
- 304, 306 until Ilpendam
- 319 until Zuideinde 90
- 346, 356 until Burgemeester Reinaldapark
- 357, 358 until Nieuw Oosteinde
- 382 until Oceaangroep
- 391, 394, 395 until Barndegat
- 392 until Zuideinde 239
- 120 until Weltevreden
- 126 until Viaduct A2)
- 320, 322, 323, 327, 328 until Maxisweg

and ten additional regional bus lines between external nodes.

External nodes

The following cities and villages are considered as external nodes as these are the nodes connected to the city network directly: Zaandam, Schiphol, Utrecht, Halfweg Zwanenburg, Haarlem, Weesp, Hilversum, Utrecht, Almere Centrum, Abcoude, Aalsmeer, Uithoorn, Purmerend, Volendam, Edam, Hoorn, Monnickendam, Landsmeer, Ijmuiden, Mijdrecht, Almere Poort, Almere Haven

A.2. Rotterdam

A.2.1 Cleaning of the RET data-set

The cleaning of the data-set has been done in an order which is followed throughout this section. This sequence of steps is required to obtain the same output as this study. Furthermore, the current schedule of the RET is used to validate the data. The same structure as to the cleaning of the GVB data-set in the Amsterdam network is used.

Same label, should not be merged

- Kreekhuizenlaan (555 & 1452 S, 768 N, 898 M)
- Heemraadsplein (1128 S, 2822 N, 2151 M)
- ’s-Gravendijkwal (2366 S, 1130 N)
- Bergse Dorpsstraat (1147 S, 2058 N)
- Parkweg (8157 S, 1153 & 1289 N

- Persoonsdam (618 W, 789 E)
- Meeuwensingel (2217 Capelle, 2406 Schiedam)
- Beltmolen (2389 & 2827 Capelle, 6227 Barendrecht)
- Sleephellingstraat (2572 N, 2139 S)
- Olivier van Noortstraat (2143 Rotterdam, 2014 Schiedam)
- Sportlaan (373 Rotterdam, 2515 Vlaardingen)

Different label, should be merged

In order to figure which nodes are located next to each other but with different names, a distance algorithm based on the latitude and longitude of nodes is used[8]. For the nodes which should be merged, the label is adjusted to either of the two existing labels. For nodes that would apply to being merged but have a reason to not being merged the reasoning is explained Based on the algorithm, the following nodes would apply to being merged as these stops are less than 150 meters apart:

- Rotterdam Centraal perron D, Rotterdam Centraal perron C, Rotterdam Centraal perron A, Rotterdam Centraal perron BB, Rotterdam Centraal perron FF, Rotterdam Centraal Uitstaphalte & Rotterdam Centraal
- Zuidplein perron J, Zuidplein perron M, Zuidplein Hoog, Zuidplein Perron H, Zuidplein perron K, Zuidplein perron N, Zuidplein perron L, Zuidplein Perron F, Zuidplein Perron G & Zuidplein
- Capelsebrug Hoog, Capelsebrug Metro, Capelsebrug Metro Uitstaphalte & Capelsebrug
- Rhoon Metro Beneden, Rhoon Metro & Rhoon
- Kreekhuizenlaan(N), Kreekhuizenlaan(M) & Kreekhuizenlaan(S) - are not merged as this is the first node after paralel lines separate
- Molenvliet & Pythagorasweg
- Dorpsweg & Katendrechtse Lagedijk
- Van Blommesteynweg & Arendsweg
- Claes de Vrieselaan / Nieuwe Binnenweg & Claes de Vrieselaan / Mathenesserlaan - are not merged as this is the first node afte an intersection of the lines
- Mathenesserlaan & Nieuwe Binnenweg
- Weena & Stadhuis - are two nodes on the same line but are merged nevertheless to facilitate transfers between important lines
- Burg. Van Kempensingel & Burg. van Kempensingel
- Willemsplein & Westerstraat - are not merged as these are two nodes on the same line
- Pompenburg & Stadhuis - are not merged as these are two nodes on the same line
- Erasmus Universiteit & Burgemeester Oudlaan - are not merged as these are two nodes on the same line

[8] Retrieved from: https://stackoverflow.com/questions/19412462/getting-distance-between-two-points-based-on-latitude-longitude

- Harreweg & De Akkers - are not merged as these are two nodes on the same line
- Station Schiedam Centrum Buffer, Schiedam Centrum & Station Schiedam Centrum
- Station Blaak & Blaak
- Station Alexander & Alexander
- Beijerlandselaan & Polderlaan
- Oosterflank Metro/Grote Beer, Oosterflank Metro/Hoeksteen & Oosterflank
- De Linie & De Terp - are not merged as these are two nodes on the same line
- De Terp & Westerlengte - are not merged as these are two nodes on the same line
- Station Zuid/Steenplaat, Station Zuid/Rosestraat & Station Zuid
- Roentgenstraat & Vrij Entrepot - are not merged as these are two nodes on the same line
- Breitnerstraat & Wytemaweg - are not merged as these are two nodes on the same line
- Allard Piersonstraat & Beukelsweg
- Rotterdamse Rijweg & Van Noortwijckstraat - are not merged as these are two nodes on the same line
- Meijersplein Metro & Meijersplein / Airport
- Blijdorp Metro & Blijdorp
- Schenkel Metro & Schenkel
- Dijkzigt Metro & Dijkzigt
- Buys Ballotlaan & Dirk de Derdelaan - are not merged as these are two nodes on the same line
- Ikazia Ziekenhuis & Motorstraat - are not merged as this is the first node after paralel lines separate
- Burgdorfferstraat & Damstraat - are not merged as these are two nodes on the same line
- Fuutstraat & Nachtegaalplein - are not merged as this is the first node after paralel lines separate
- Slinge Metro & Slinge
- G.A. Soetemanweg & Van Byemontsingel - are not merged as these are two nodes on the same line
- G.A. Soetemanweg & Cornelis van Dijckstraat - are not merged as these are two nodes on the same line
- Pernis Metro & Pernis
- Garage Sluisjesdijk & Dokhaven
- Rietdijk, Waalhaven Noordzijde, Garage Sluisjesdijk & Dokhaven - are not merged as these are nodes on the same line (taking the merging above into account)
- Hoogvliet Metro Busperron A, Hoogvliet & Hoogvliet Metro
- Poortugaal Metro & Poortugaal

- Pascalweg & Dumasstraat - are not merged as these are two nodes on the same line
- Van de Woestijnestraat & Spinozaweg - are not merged as these are two nodes on the same line
- Lansingh-Zuid & Lansingh Zuid
- Vijfsluizen Metro & Vijfsluizen

Merging of nodes with the same label

Following from the previous modifications, any nodes with the same label should now be considered as one node. Therefore, it is checked which node ID's have the same label and these are merged subsequently. As any nodes with the same label which should not be merged and any nodes which should be merged but have different labels have been identified, these are left out or included respectively. Following, a graph in which only unique node-labels are present in which nodes with different labels being at (nearly) the same location are considered as one.

A.2.2 Cleaning of the NS data

As the train stations outside the area of study are considered external nodes, only direct links from the stops in the network are considered. Therefore, only a few additional train stations are added, which is done manually. To ensure these nodes are not considered for the hierarchical degree, as these are beyond the scope of this study, these nodes are given an identifiable node ID in the range 9900 - 9999. The train lines are added using these external nodes and the node ID's of train stations in the network on the route. When the train line leaves the PTN area considered, the next node on the line is considered as the external node.

A.2.3 Cleaning of the regional bus data

Besides the main operator RET, there are four additional operators for the regional buses. Unfortunately, the coding of links and lines does not match the RET data in the NDOV data-sets which implies the additional lines, links and nodes have to be added manually. Furthermore, for the external nodes, the coding has to be adjusted in order to be perceived as external. This includes stops on the RET lines, metro-lines B and E in particular. As the regional buses leave the PTN at one point in every case, border nodes are added, after which the line is further linked to the destination of the regional bus or to any urban center if it passes multiple. Consequently, any nodes between the border node and the external destination node are excluded from the data. Based on figure 4.4, the cutting of the regional lines is executed in which the bordering nodes are the nodes just outside the PTN are given a node ID between 9900 and 9999 similar to the external train stations. Terminal bus-stations outside the PTN are given a node ID in this range as well. Therefore, any node with a node ID between 9900 and 9999 can be seen as an external node for this network. If a partially external line passes stops excluded from the data outside the boundary of the network, a time penalty is included to compensate for the stopping time of these nodes. Furthermore, existing lines between external nodes are added to prevent unjustly sending the demand between those nodes through the network.

A.2.4 Cleaning of demand data

For the demand data, a V-MRDH data-set is used to provide OD matrices based on OmniTRANS. In order to match the demand data for the Amsterdam network as much as possible, 2030 is used as the base-year for the demand. Similarly to the VENOM model, the coding of nodes is done in a different way than the NDOV data-set leading to manual linking the nodes. Furthermore, the V-MRDH data-set contains some errors in terms of city labeling for nodes with the same street label in different cities[9]. The same method of linking nodes from one model to the other is applied for the Rotterdam network. If there is no reason to doubt the nodes can be linked they are linked, if there is some possible mix up, such as nodes with similar labels, the V-MRDH nodes are checked in QGis and linked accordingly. One additional step for the V-MRDH model is the merging of access and egress modes which are provided in separate data-files per time of day and access of egress mode (walk-walk, walk-bike, bike-bike, bike-walk). As these data files would be too large to handle at once, the selection of nodes, based on the linking between data-sets, is done before the files are merged. However, this should not lead to any differences in output.

Once the data-sets are linked and merged, the demand for every OD pair can be determined. It is, similarly to the Amsterdam PTN, rounded up to integers and the shortest paths for any non-zero OD pair is determined in order to calculate the transfer demand.

A.2.5 Selection of internal and external nodes

Internal nodes

Any nodes on the following train lines (NS):

- Delft - Breda
- Delft Zuid - Zwijndrecht
- Delft - Dordrecht
- Schiphol Airport - Breda
- Gouda - Rotterdam

and six additional train lines between external nodes.

Any nodes on the following tram lines (GVB):

- 2, 4, 7, 8, 21, 23, 24, 25

and two additional train lines between external nodes.

Any nodes on the following metro lines (GVB):

- A
- B until Maassluis Centrum
- C, D until Spijkenisse Centrum
- E until Rodenrijs

Any nodes on the following bus lines (RET):

- 30, 31, 32, 33, 35, 36, 37, 38, 42, 44, 47, 51, 53, 54, 56, 67, 68, 69, 70, 72, 74, 76, 78, 79, 82, 83, 84, 96, 97, 98, 127, 156

[9] e.g. De Akkers is a stop in both Schiedam and Spijkenisse but the V-MRDH codes all of the nodes in Spijkenisse despite some being located in Schiedam on the QGis map of the model

- 40 until De Tempel
- 126, 143 until Maasland Viaduct
- 144 until Huys ten Donck
- 146 until Goudenregenplantsoen

Any nodes on the following regional buses (Conexxion, EBS, Arriva, Qbuzz):

- 160, 166 until Heinenoord, Busstation
- 310, 436 until Numansdorp, Rijksweg A29
- 392, 491 until Oostendam
- 456 until Maasland Viaduct
- 489 until Sportlaan (Ridderkerk)

and six additional regional bus lines between external nodes.

External nodes

The following cities and villages are considered as external nodes as these are the nodes connected to the city network directly: Den Haag Centraal, Den Haag HS, Den Haag Leyenburg, Den Haag Laan van NOI, Delft, Delft Zuid, TU Delft Campus, Zwijndrecht, Dordrecht, Breda, Gouda, Rodenrijs, Spijkenisse, Hoek van Holland, Maassluis, Ridderkerk, Pijnacker, Nootdorp, Voorburg/Leidschenveen, Schiphol, Oud-Beijerland, 's Gravendeel, Bergen op Zoom, Hendrik-Ido-Ambacht, Naaldwijk, Middelharnis, Alblasserdam, Papendrecht, Sliedrecht, Zoetermeer, Hellevoetsluis. The latter two do not have a direct connection to any internal node but are important locations directly connected to the regional metro in Pijnacker/Rodenrijs and Spijkenisse for Zoetermeer and Hellevoetsluis respectively.

B

Scenario adjustments

B.1. Scenario before the opening of NZL

Figure B.1[1] shows the former map of tram-lines in Amsterdam to which many adjustments were made combined with the opening of the NZL. The network of the scenario before the NZL includes some additional nodes which became unused after the NZL opened. Furthermore, some nodes had a different label in the network before the NZL but these are not changed as the label of the node is of no importance for the hierarchical degree. The nodes which were present in the pre-NZL scenario but are no longer in use have to be manually added to the network and demand matrix. These nodes are:

- Ruysdaelstraat (line 16)
- Valeriusplein (S) (line 16)
- Emmastraat (line 16)
- Jac. Obrechtstraat (line 16)
- Zuideinde (bus 36)
- Lange Vonder (bus 36)
- Eeuwige Jeugdlaan (bus 38)

[1]Retrieved from: https://railsinamsterdam.wordpress.com/mijn-plattegronden/tramplattegronden/

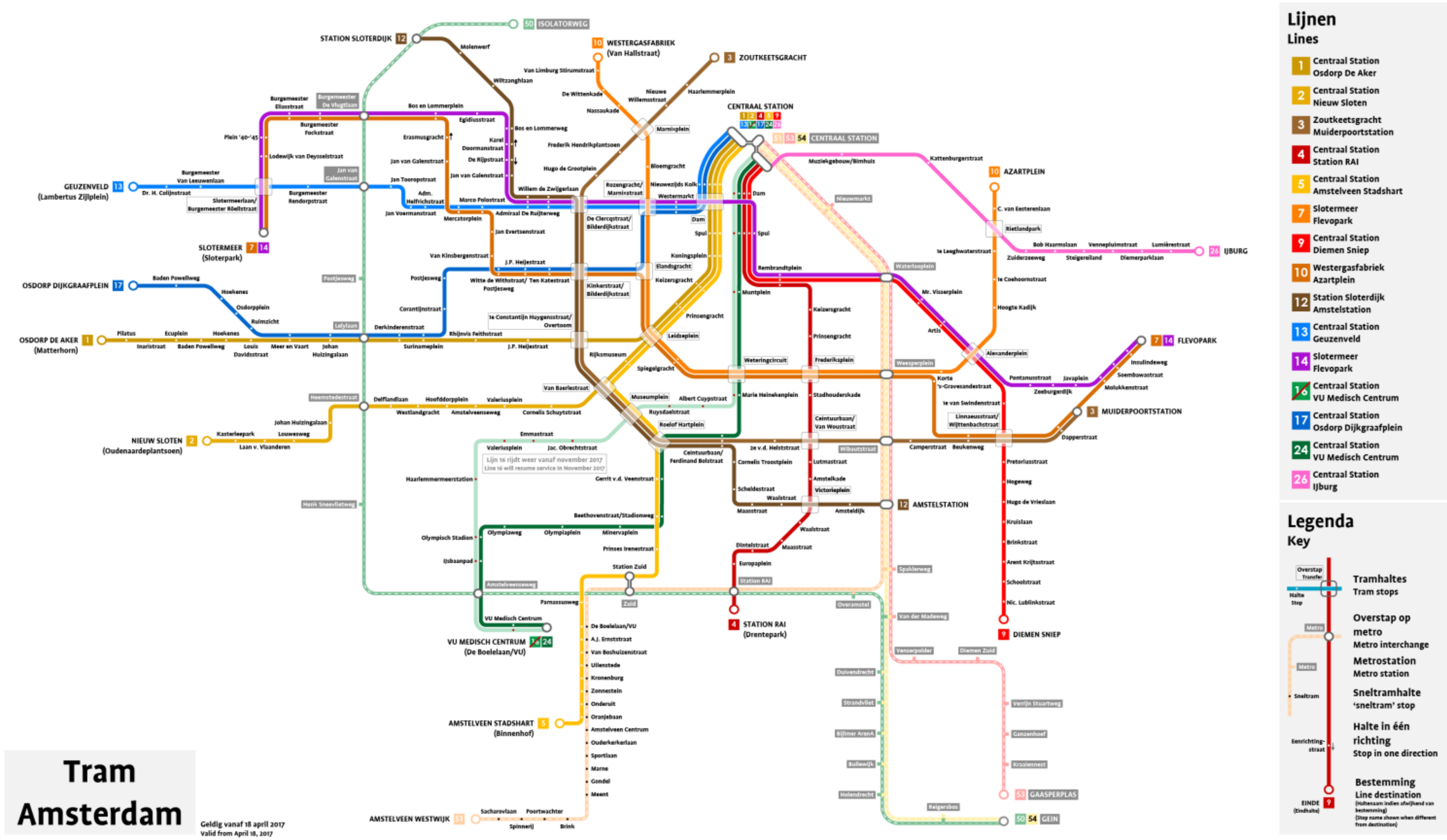


Figure B.1: Former map of tram-lines of Amsterdam

B.2. Future scenarios

For the future scenarios, the options that are assessed are shown in figure B.2[2] which is based on the metro plan study from 2007. In this figure, the red and green dashed lines are the options that are evaluated in the scenarios. The blue dashed lines are excluded as the darker blue line is no longer served by a metro while the lighter blue line does not add any additional lines. As the new lines are connected to existing nodes, the lines are fully linked to existing nodes.

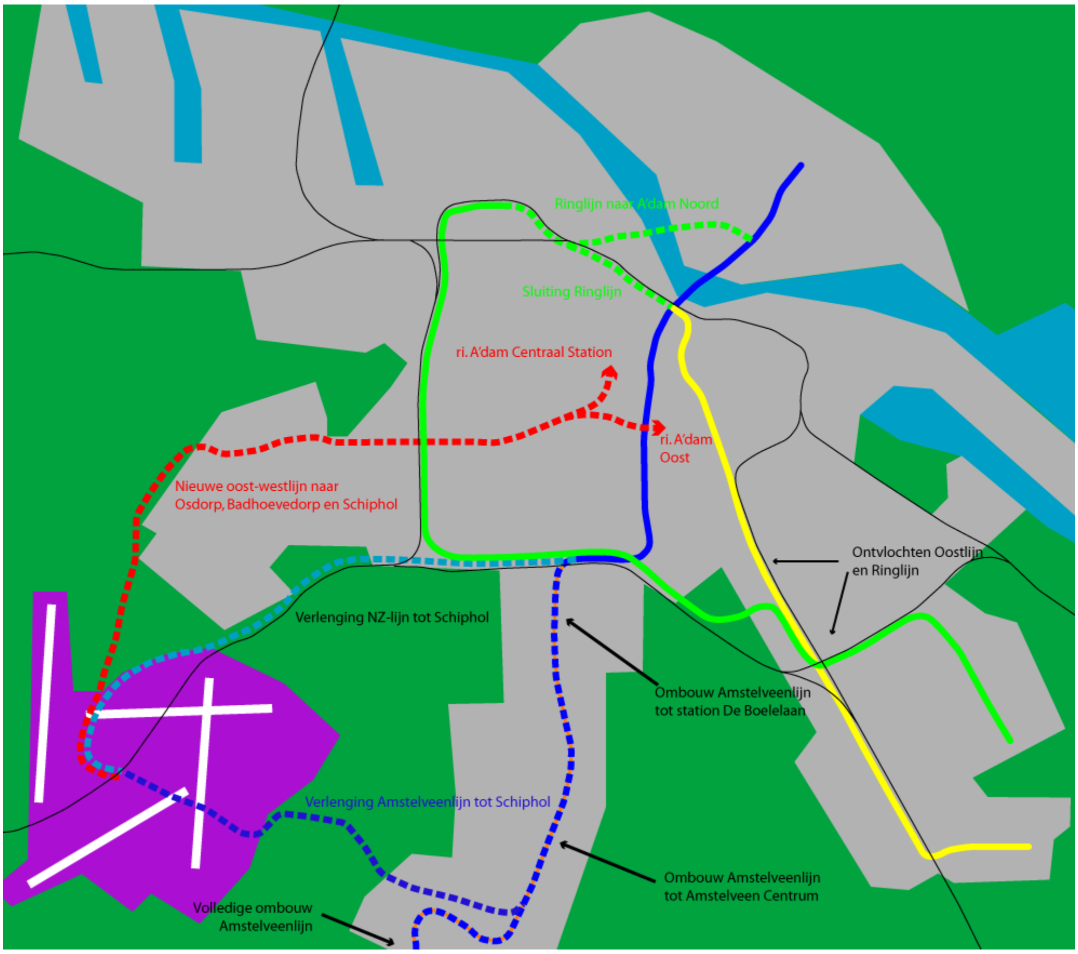


Figure B.2: Future scenarios for the metro-lines in the Amsterdam PTN

The routing for the scenarios is partially based on the Metronetstudie 2007[3] complemented with own interpretation and other sources[4]. The two scenarios for the green line are called G1 and G2 where G1 connects to Centraal Station and G2 connects to Noorderpark while the scenarios for the red line are called R1 and R2 where R1 connects to Centraal Station from Leidseplein and G2 connects to Muiderpoort. The lines would have the following route:

- G1: Isolatorweg - Spaarndammerstraat - Haarlemmerplein - Centraal Station
- G2: Isolatorweg - Spaarndammerstraat - Noorderpark

[2]© Alargule Productions. Retrieved from: `https://upload.wikimedia.org/wikipedia/commons/8/85/Amsterdam_Metronetstudie_2007.png`

[3]Found at: `https://www.noordzuidlijnkennis.net/wp-content/uploads/2013/05/metronetstudie.pdf`

[4]e.g. `https://wijnemenjemee.nl/toekomst/nieuws/komt-er-ooit-een-oost-westlijn`

- R1: Schiphol Airport - Badhoevedorp Oost - Baden Powellweg - Meer en Vaart - Amsterdam Lelylaan - Surinameplein - Overtoom - Leidseplein - Rokin - Centraal Station
- R2: Schiphol Airport - Badhoevedorp Oost - Baden Powellweg - Meer en Vaart - Amsterdam Lelylaan - Surinameplein - Overtoom - Leidseplein - Vijzelgracht - Weesperplein - Alexanderplein - Station Muiderpoort

C

Hierarchical nodes maps

In order to provide a more detailed map for the highest scoring nodes, figure C.1 and figure C.2 visualize the networks for Amsterdam and Rotterdam respectively

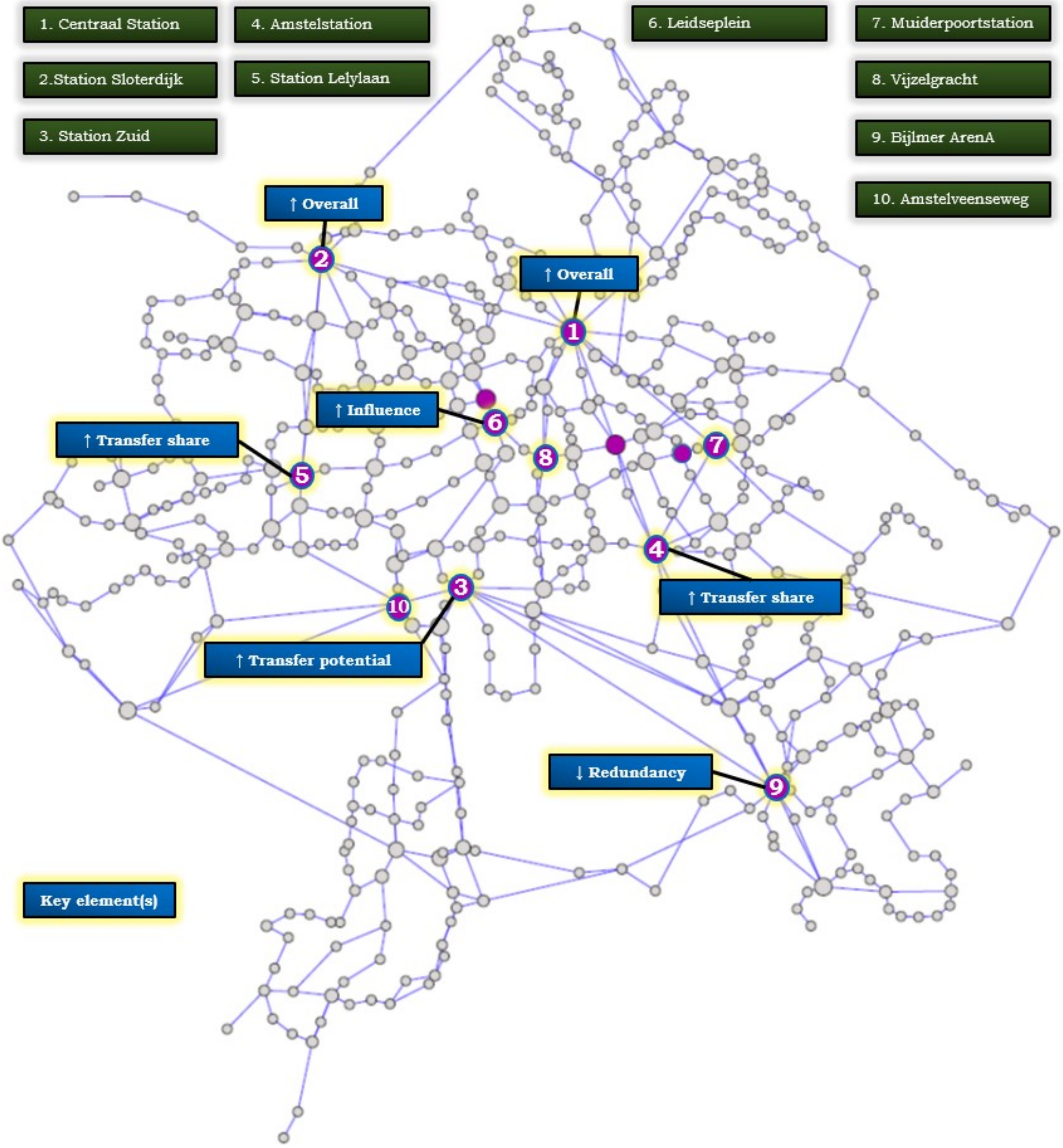


Figure C.1: Overview of the most hierarchical nodes and their key characteristics in the Amsterdam PTN

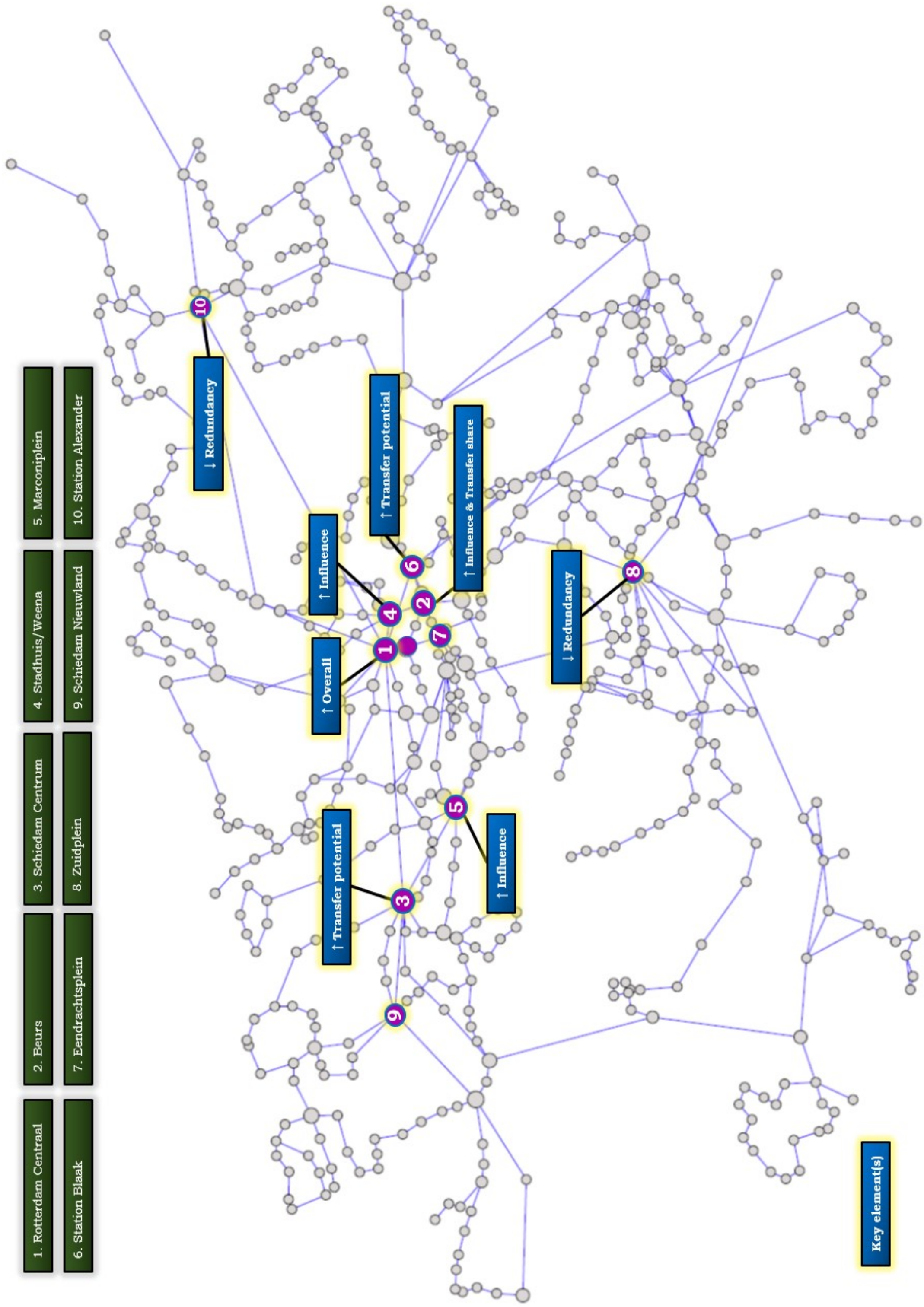


Figure C.2: Overview of the most hierarchical nodes and their key characteristics in the Rotterdam PTN

D

Research paper

Throughout this appendix, a scientific research paper based on this thesis is presented. While it follows the same structure as the report, only highlights and urgent matters are discussed due to the limited size of a scientific article. Hence, in particular the literature review, application and results are only limitedly explained in this article.

Quantifying hierarchy in public transport networks: Developing a new metric

Abel Buijtenweg[a,*], Oded Cats[a], Trivik Verma[b], Huijuan Wang[c], Barth Donners[d]

[a]*Delft University of Technology, Faculty of Civil Engineering and Geosciences. Mekelweg 5, 2628 CD Delft, the Netherlands*
[b]*Delft University of Technology, Faculty of Technology, Policy and Management. Jaffalaan 5, 2628 BX Delft, the Netherlands*
[c]*Delft University of Technology, Faculty of Electrical Engineering, Mathematics & Computer Science. Mekelweg 4, 2628 CD Delft, the Netherlands*
[d]*Royal HaskoningDHV, Laan 1914 no 35, 3818 EX Amersfoort, the Netherlands*

Abstract

Due to increasing pressure on the system, the functioning of public transport networks (PTN) in metropolitan areas is crucial for future mobility. Within these networks, hierarchical levels can be distinguished where different levels have different functions. These hierarchical levels can be analyzed but there is no way to quantitatively determine the hierarchy in a PTN. Therefore, this paper presents a metric to quantify the hierarchy in PTN to increase the understanding of these complex networks. In order to determine the hierarchy, a metric is developed based on a combined topological and empirical approach. The metric is a multiplication of three different elements which are the topological influence, (non-)redundancy and transfer potential. Together these elements are applied to define the hierarchical degree of nodes in a network. Furthermore, to determine the hierarchy of a network as a whole, a hierarchical coefficient, based on the distribution and inequality of the hierarchical degree in the network is developed. The metric has been applied to case-studies for the Dutch cities of Amsterdam and Rotterdam which allows for different state and cross-network comparison. The results show some expected yet non-trivial results identifying different patterns in network structures for network states and different spatial distribution of hierarchy between networks. Furthermore, by dividing the network into functional levels, a hierarchical structure can be identified. Throughout this study, a new method to quantify hierarchy in PTN, based on different approaches, is developed which can be seen as the most important contribution of this research. While this study explores the implications of this metric, it can be applied in numerous different contexts. Furthermore, in potential the metric has numerous network related applications such decreasing vulnerability and solving bottlenecks.

Keywords: Public transport networks (PTN), network structure, hierarchical degree, topological metrics

1. Introduction

For the future development of the transport system in larger cities, public transport (PT) plays an essential part to accommodate passengers [1]. In particular the organization of a public transport network (PTN) is perceived as very complex due to different modalities, infrastructure and passenger flows. A method to represent the PTN of an entire city is by using a multilevel network [2, 3, 4], in which categorical levels are distinguished to tell the different characteristics and functions within the system apart.

One approach to define levels in a PTN is based on the functions of stops, which are referred to as nodes throughout this study. In order to connect the different levels, transfer nodes are used as locations for level interaction. The functioning of these transfer nodes between the different levels in a transport network is becoming an essential aspect of a transport network to facilitate rapid transport throughout the network. The transfer locations are indicated as (transit) hubs, where multiple levels of transport networks are coming together and interaction between the different network levels takes place. As a PTN only requires a limited number of hubs, just a fraction of the nodes functions as a transit hub while the majority functions as a minor hub (e.g. a node located at a location where two lines intersect) or as an entry and exit point only. It is therefore meaningful to understand the different nodes in a network have a different function in which some are considered more important than others. In other words, it is meaningful to distinguish between nodes based on their relative function in the network.

After valuing nodes based on their relative function, the nodes can be allocated to a categorical level based on the importance of their function in the network. If for these categorical levels some ordering is applied, leading to a(n) (exponentially)

*Corresponding author
Email address: abelbuijtenweg@gmail.com (Abel Buijtenweg)

decreasing number of elements for subsequent levels, it may be considered as a hierarchical structure. Following, a hierarchical order in the network can be identified where nodes are allocated in a hierarchical level based on their importance in the network. It is however, hard to quantify the function of a node in a network as an 'important function' can be interpreted in numerous ways. In other words, there is no generally accepted method of determining a hierarchical order in a PTN. Therefore, the assessment of the hierarchical structure in a PTN remains limited.

While many network indicators have been identified, a quantitative indicator to specifically determine the level of hierarchy in a PTN is in-existent [3, 5, 6]. Moreover, when comparing different networks across cities (e.g. [7, 8]), the hierarchical structure of a network is left out of the comparison. Even though the influence of transfers has been incorporated in topological indicators, this predominantly relates to the requirement for a transfer in the shortest path rather than the function of transfer locations [9]. Hence, a measure to evaluate the hierarchical structure in a network incorporating the function of transfer locations is generally lacking. Furthermore, a good measure for hierarchy could also be a starting point for a generally accepted definition for hierarchy [10].

Throughout the remainder of this article, the existing literature is first reviewed in the subsequent section. Thereafter, the methodology to develop a hierarchical metric is elaborated. Next, the application of the metric is briefly clarified after which the results are revealed. Lastly, some conclusions and recommendations are presented.

2. Literature review

In contemporary literature on hierarchy in transport networks, there are three perspectives to be identified which are applicable to determine hierarchy in PTN. *Link-based hierarchy* [11, 12] defines hierarchy for each link of the network and is therefore able to distinguish different functions between each section of a line. However, it is unable to capture the importance and dynamics of transfer(s) (hubs). *Line-based hierarchy* [13, 14, 15] is used to divide (sections of) lines into hierarchical levels. While it emphasizes the role of transfers in PTN, it is unable to determine the role of the locations where the transfers take place. Finally, there is *Node-based hierarchy* [16, 10, 17] which is applied to determine the hierarchy of nodes based on their function in the network. This approach is used to define the function of hubs in the network but does not distinguish characteristics of links connected to the node.

In order to further elaborate how hierarchy may be analyzed in PTN, two approaches are applied, *topological network analysis* and *empirical network analysis*. These approaches are seen as complementary as each of these has a different yet non-contradictory perspective on hierarchy.

2.1. Topological network analysis

Regarding network topologies in PTN, metrics based on graph theory are widely applied for network analysis [18]. Emphasizing this, most efforts by researchers on studying the topological properties of complex networks, such as PTN, are based on graph theory [19]. Therefore, a more elaborate overview of how a network is analyzed topologically, could lead to insights regarding analyzing the hierarchy of nodes and the hierarchical structure. Furthermore, by distinguishing and analyzing network properties, different PTN are compared [20]. The comparison of topological networks could potentially provide a method to understand the differences in terms of hierarchy between different networks. Moreover, changes in topology are widely used to provide insight the development of a PTN [18].

A limitation of topological metrics that has to be taken into account is that these metrics are usually focused on the structure of a transportation network while it neglects the dynamic features [21]. Hence, it provides an overview of the network without demand nor capacity related characteristics.

2.1.1. Topological model representation

Following from graph theory, a variety of network representations are distinguished which all have different properties [22]. An overview of the relevant spaces is presented in figure 1.

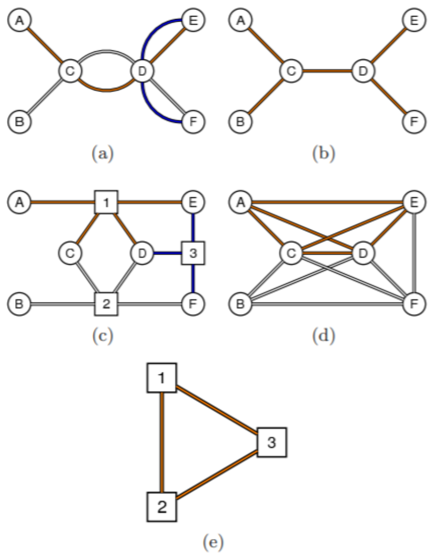


Figure 1: Representation of an example of space representations where (a) indicates the existing nodes and lines, (b) the 𝕃-space, (c) the 𝔹-space, (d) the ℙ-space and (e) the ℂ-space [from 20]

For this study in particular, a combination of the 𝕃- and ℙ-space are applied. On the one hand, the $\mathbb{L}$-space is used to determine the location, neighbors and directions of the links for a node to understand the position in the network. Based on the $\mathbb{L}$-space, the function of a node in the network is derived. For example, hub nodes are easily spotted in the $\mathbb{L}$-space by having many links attached. On the other hand, the $\mathbb{P}$-space can be used to identify which node pairs are directly connected. This is particularly useful to clarify the number of transfers in a shortest path. Furthermore, by using the $\mathbb{P}$-space, nodes that are passed through along the route but do not require any transfer, are identified. This can be useful to determine transit passengers for a node to which passengers passing through on one line do not belong. Hence, the $\mathbb{P}$-space can be used to determine if a node located on the shortest path of a node pair functions as a transfer station or not. Furthermore, as direct connections in PTN are considerably more informative than direct neighbors (which may just be a passing through node), using the $\mathbb{P}$-space generally provides more information in general.

2.1.2. Local network indicators

In order to characterize transport networks topology, network indicators are used extensively [18]. Analyzing the topological network indicators can lead to new insights regarding common features between different types of networks [18]. In order to define a new metric so, existing network indicators are evaluated to identify how a hierarchy indicator can be connected to this. The node degree for any given node is defined as the number of nearest neighbors [23]. The node degree (in the $\mathbb{L}$-space) is in particular useful to identify hub nodes [24]. For determining hierarchy, the node degree could be applied as it indicates the number of links a node has. However, this provides only limited information as the number of links does barely explain the function of a node.

The assortativity which is defined as the average degree of the nearest neighbors in the $\mathbb{L}$-space [25]. This indicator can be applied to determine if a network shows assortative or disassortative mixing on the degree [26]. The former indicates the assortativity increases for nodes with a higher degree while the latter indicates vice versa. The clustering coefficient is used to determine the degree to which nodes are interconnected [27]. In other words, this metric determines the share of neighboring nodes which are mutually connected. The betweenness centrality is particularly capable of highlighting the importance of a node as a transfer point (or passing through point) for a node pair [28]. Betweenness is defined as the share of shortest paths between every pair of nodes that passes through a given node i [25]. Connected to the former, the eigenvector centrality is based on the centrality of nodes. However, instead of just the centrality of a node itself, it is based on the centrality of directly connected nodes. In other words, the quality of the links for a node are taken into account by valuing links to a more central node a higher than links to a node with a low centrality [29]. The last indicator, the overlapping degree, is based on the degree of nodes for different layers. The different layers can be based on different modalities or lines [4]. An overview of these indicators and their applicability for a hierarchy indicator is shown in table 1.

2.1.3. Global network indicator

In order to determine the hierarchy of the network, or on the global scale it makes sense to consult a generally applied approach related to the curve of the distribution function. The curve of the hierarchy distribution could then be examined provided it obeys a power-law from which the alpha value can be derived (see e.g. [36, 25, 26, 37, 38, 39]). Another approach to analyze the distribution in a network structure is the Gini inequality index [40, 41, 42, 43]. The Gini inequality index can be applied to test the inequality in a distribution by examining the Lorenz-curve based on the ascendingly ranked values for an indicator.

2.2. Empirical network analysis

Diverging from the topological approach to analyze PTN, an empirical approach, incorporating passenger data, can be applied based on travel demand. Using demand data from the network, an analysis of the network can be done to determine the hierarchy of the network based on passenger flows.

Each node in the network is assigned as an *origin* and a *destination* [44]. Based on this definition the concept of an OD matrix (origin-destination matrix) can be used to represent the travel patterns in a network [45]. For every OD pair, a certain demand is known or can be calculated depending on the availability of data. Furthermore, between every OD pair a set of routes connects the pair where every route has different characteristics in terms of travel time, transfers, fare price, etc. [44, 46]. As there are numerous assignment methods to estimate the flow, an assignment method, which is perceived optimal for determining the hierarchy, has to be chosen.

The most straightforward way of flow assignment is based on static demand implying the notion of time is unaccounted for [47, 48]. Using static assignment, the advantages of more comprehensive flow assignment methods such as dynamic assignment are unaccounted for. Advantages of such an approach include considering the effect of congestion [49] and variations for different time intervals [50] are neglected, leading to a limitation in terms of understanding the dynamics in travel behavior. Neglecting the congestion of vehicles, and nodes in general, leads to a bias in terms of travel time as well as a preference to travel around the congested

Table 1: Overview of local network indicators

Local network indicator	Notation	Explanation	Explaining hierarchy	Previous studies
Node degree	k_i	Number of links that are directly connected to a node	+ -	[24, 25, 30]
Assortativity	$k_{nn,i}$	Average degree of nearest neighbors	+ -	[27, 25, 26, 31]
Clustering coefficient	C_i	Share of node's direct neighbors that are mutually connected	+	[27, 32]
Betweenness centrality	$C_B(i)$	Share of shortest paths go through a certain node	+ -	[25, 28, 33, 34]
Eigenvector centrality	x_i	Influence of a node and its directly connected neighbors	+	[29, 35]
Overlapping degree	o_i	Presence of nodes in different layers	+ -	[4]

nodes (e.g. to avoid busy stations during peak hour). Nevertheless, as travel behavior is perceived of limited importance for hierarchy, a straightforward assignment method is deemed sufficient.

For the traffic assignment method, the least comprehensive method is based on a all-or-nothing perspective. Opposed to a stochastic perspective, which would take the variety in route preference into account [51], all-or-nothing assignment assumes all passengers choose the fastest route and are unaffected by congestion or preference. While stochastic route choice has the advantages that different route choices and preferences are taken into account, to account for variability in preference regarding traveling [52], all-or-nothing assignment has significantly lower computation times as only one shortest path between nodes has to be defined instead of a set of possible routes between nodes. Hence, choosing either approach is practically a trade-off between accuracy and computation time.

3. Methodology

In order to determine the metric for hierarchy in PTN, first a definition for hierarchy within the context of PTN is required. Based on a combination of node-based hierarchy enhanced with line/link-based hierarchy for demand, the definition for a hierarchical node (or local hierarchy) reads: *a node which has high influence in the network by being directly connected to a wide range of nodes and operating as a transfer hub.* In this definition, connected to a wide range of nodes implies nodes have direct lines to many other nodes. However, as a direct connection to another important node could be seen as more important than a connection to a minor node, not all connections should be valued equally. Functioning as a transfer hub relates to the share of passengers transferring at the specific node and the number of directions to transfer to. Furthermore, by being a transfer hub, the node should have a degree of inevitability for the neighboring nodes. In other words, neighboring nodes should have a low number of mutual connections.

For the hierarchy of a network itself (or global hierarchy) the definition reads: *a network structure which shows distinguishable functional layers, which come together at transfer hubs, and has a power-law distribution of nodes in terms of hierarchy.* In this definition, the functional layers can be interpreted as levels of nodes in a certain hierarchy range in which every layer underneath contains nodes with a lower hierarchical degree.

3.1. Hierarchical degree

In order to assure the metric aligns with the definition for hierarchy, three separate elements are included in the metric which are: *topological influence*, *redundancy* and *transfer potential.* The three elements are chosen as these are directly derived from the definition for hierarchy and have a complementary function. For these three elements, the first two relate to the topological approach of the node while the last relates to the empirical approach. It should be noted in advance that the metric is based on a $\mathbb{P}$-space representation for the first and second element while a $\mathbb{L}$-space representation is used as input for the third element. An additional adjusted space is implicitly applied to the third element to determine where transfers take place and to value transfer time realistically. Moreover, the hierarchical degree is only calculated for nodes with more than two directions to travel to as it would not be able to function as a transfer location otherwise. Hence, the degree of the nodes in the $\mathbb{L}$-space should be higher than two or else the hierarchical degree should be set to zero, where

$$H_i = 0, \quad \text{if} \quad k_i \leq 2, \tag{1}$$

which holds for all three elements. In this equation k_i represents the degree of node i in the $\mathbb{L}$-space. Throughout the subsequent paragraphs, the different elements are explained and elaborated.

3.1.1. Element A: Topological influence

The first element relates to the topological influence of a node in the network, incorporating the function of nodes it is directly connected to. This element is derived from the eigenvector centrality with a small adjustment to normalize for the network. The inclusion of the function of directly connected nodes relates to the value of having a hub in close proximity, or direct connection, being beneficial for the node in question too. The eigenvector centrality is exceptionally useful to look at the function of a node while valuing connections to other important nodes higher than connections to less important nodes.

The normalization of the eigenvector centrality is done in order to facilitate cross-network comparison as larger networks tend to have a more spread centrality with generally lower eigenvector centrality values. Furthermore, normalizing the eigenvector centrality enables a clear range for output values of the topological influence in the interval [0,1] where a value of 0 indicates the node is separated from the network while a value of 1 indicates the node is considered the most influential node in the network. As indicated above, the $\mathbb{P}$-space representation is applied to emphasize the importance of direct lines between nodes rather than the node being a direct neighbor. To clarify how this element can be obtained it is described as e_A where

$e_i^A \sim$ *the* ***topological influence*** *of node i, based on connections to other* ***influential nodes***

$$e_i^A = \begin{cases} \dfrac{x_i}{x_{max}}, & \text{if} \quad k_i > 2; \\ 0, & \text{otherwise,} \end{cases} \qquad \forall i, \qquad (2)$$

and based on the definition for the eigenvector centrality (x_i) [29]

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} a_{i,j} x_j, \qquad (3)$$

in which $a_{i,j}$ is the adjacency matrix and x_j is the eigenvector centrality of neighboring nodes.

3.1.2. Element B: Redundancy

The second element of the formula for the metric is directly derived from the clustering coefficient (see e.g. [27, 53]). This element is added to the metric in order to analyze the necessity of a node in terms of transfers. This element is used to test whether a node is required for the connection between its directly connected nodes. If many of these directly connected nodes were to share another mutual connection, only a limited OD pairs would require a transfer at the given node. Therefore, the less direct connections of a node are mutually connected, the more hierarchical it is perceived.

The clustering coefficient value of a node is determined to test the redundancy of a node in the $\mathbb{P}$-space. The $\mathbb{P}$-space is applied as this indicates which of the directly connected nodes share a mutual line between them and require, therefore, no transfer at the node in question. Consequently, a higher value for the clustering coefficient indicates many of the directly connected nodes share a line and only few transfers have to be made at the node in question. The clustering coefficient is subtracted from one leading to a lower value if the clustering coefficient is higher. As explained in the definition for hierarchical nodes, less mutual lines among neighbors should contribute to a higher hierarchical degree which is why the clustering coefficient is subtracted from one. To explain what this element looks like as a function, it is described as e_i^B, where

$e_i^B \sim$ *the* ***redundancy*** *of node i, based on* ***mutual connections*** *among neighboring nodes*

$$e_i^B = \begin{cases} 1 - c_i(k_i), & \text{if} \quad k_i > 2; \\ 0, & \text{otherwise,} \end{cases} \qquad \forall i, \qquad (4)$$

and based on the definition for the clustering coefficient [32]

$$c_i(k_i) = \frac{n_i}{k_i(k_i - 1)/2} \qquad \forall i, \qquad (5)$$

in which n_i indicates the number of links between neighboring nodes in the $\mathbb{P}$-space and k_i in this equation refers to the degree in the $\mathbb{P}$-space while it refers to the $\mathbb{L}$-space in the former equation

3.1.3. Element C: Transfer potential

The third part of the metric is based on the share of transferring passengers at a node. Transferring in this sense means changing lines with possibly a different mode. This element is significantly different from the prior element as this one requires demand data and therefore incorporates the demand for transfers at a node. Consequently, the empirical aspect of hierarchy in PTN is complied with by including a passenger perspective with the share of transfers that take place at a certain node. The share of transferring passengers is multiplied by the share of potential transfer directions which is defined as the degree minus two divided by the degree in the $\mathbb{L}$-space. This definition of transferring directions is based on the fact that transferring here relates to getting off the vehicle and the vehicle generally has an in- and outgoing link. To explain this logic, assume a vehicle traveling from A to C through B where B offers transfers towards D, E, F and G. Then $\frac{4}{6}$ outgoing links require a transfer as traveling to C requires no transfer and A is the origin. e_i^C is defined

$e_i^C \sim$ *the* ***transfer potential*** *based on* ***transfer passenger share*** *and* ***transfer directions*** *for node i*

$$e_i^C = \begin{cases} \frac{log(P_i^{transfer})}{log(P_{max}^{transfer})}\frac{k_i - 2}{k_i}, \text{ if } & k_i > 2; \\ 0, & \text{otherwise}, \end{cases} \quad \forall i, \tag{6}$$

in which $P_i^{transfer}$ are passengers transferring at node i, $P_{max}^{transfer}$ is the the maximum number of passengers transferring at a node in the system. The degree k_i is here based on the $\mathbb{L}$-space of the network. $P_{max}^{transfer}$ is used to normalize the values where the node with the highest number of transfers has the highest transfer potential thus a value of one and the other nodes a fraction between zero and one. The logarithm is used as it is expected many nodes probably have just a small share of transferring passengers compared to the highest scoring node. Therefore, using the logarithm prevents getting values for other nodes becoming very close to zero for any node with a just a small fraction of the transfer passenger share. Intuitively it makes sense to use the logarithm in order to prevent the transfer share from having too much impact on the hierarchical degree. If non-logarithm transfer passenger numbers would have been used, any node besides a few crucial transfer hubs would have a value of practically zero.

3.1.4. Combining the elements

In order to combine the three elements it is crucial to understand how these complement each other. This research suggests the elements should be combined by multiplying the three elements in order to emphasize the complementary nature of the elements. By using multiplication to combine the elements, a high score for one of the elements does only lead to a high hierarchical degree if the other two elements have a decent value as well. Consequently, nodes with a high value for only one of the elements, are flattened in terms of hierarchical degree. Combining all three elements into one, the metric for the hierarchical degree reads

$$H_i = e_i^A * e_i^B * e_i^C, \quad \forall i, \ \{H_i \in \mathbb{R} : 0 \leq H_i < 1\}. \tag{7}$$

with all the elements written out, the hierarchical degree can be defined as

$$H_i = \begin{cases} \frac{x_i}{x_{max}}(1 - c_i(k_i))\frac{log(P_i^{transfer})}{log(P_{max}^{transfer})}\frac{k_i - 2}{k_i} & \text{if } k_i > 2; \\ 0, \quad \text{otherwise}, \end{cases} \quad \forall i, \tag{8}$$

The range for the hierarchical degree is between zero and one. A node scoring zero is considered as non-hierarchical, while nodes with a hierarchical degree of higher than 0.125 (which is based on each element having a value of 0.5) are considered high-hierarchical and in the interval (0, 0.125) are considered low-hierarchical. Consequently, three different hierarchical levels are identified.

3.2. Hierarchical coefficient

To determine the hierarchical structure of a network, a combination of the alpha value for the power-law distribution and the value of the Gini-index is used to incorporate both the decrease of nodes with a higher hierarchy and the in-equal spread of hierarchy throughout the network. These can be seen as complementary as the power-law distribution is used to analyze the decrease in share of nodes for increasing hierarchy while the Gini coefficient is used to analyze the spread of hierarchy in the network. Consequently, the power-law distribution is used to analyze the share of nodes while the Gini-index is used to analyze the share of hierarchy. Therefore, this paper perceives these coefficients are complementary rather than overlapping. To define the hierarchical structure of a network the following function is used

$$H_I = G_I^{\frac{1}{|\alpha_I|}} \qquad \{H_I \in \mathbb{R} : 0 \leq H_I < 1\}, \tag{9}$$

in which H_I is the hierarchical coefficient of network I and G_I and α_I are the Gini index and curve of the power-law distribution function of network I respectively. The value for H_I indicates the hierarchy of the network where a value close to 0 indicates either a uniform hierarchical degree or a uniform hierarchical degree distribution. A value close to 1 on the other hand indicates either in-equal hierarchical degree or a very in-equal distribution. It should be noted that a more hierarchical structure may not be desired in all cases and some adaptations to a network to increase the spread of hierarchical degree and, therefore, decrease the hierarchical coefficient could be advisable.

3.3. Multi-level representation

In order to subdivide the nodes into different functional levels, three sub-levels are distinguished which are the high-hierarchical, low-hierarchical and non-hierarchical for levels 1, 2 and 3 respectively. A node scoring zero is considered as non-hierarchical, while nodes with a hierarchical degree of higher than 0.125 (which is based on each element having a value of 0.5) are considered high-hierarchical and values in the interval (0, 0.125) are considered low-hierarchical. The value of 0.5 as threshold for level 1 is chosen as this is exactly in the middle of the range of each of the elements.

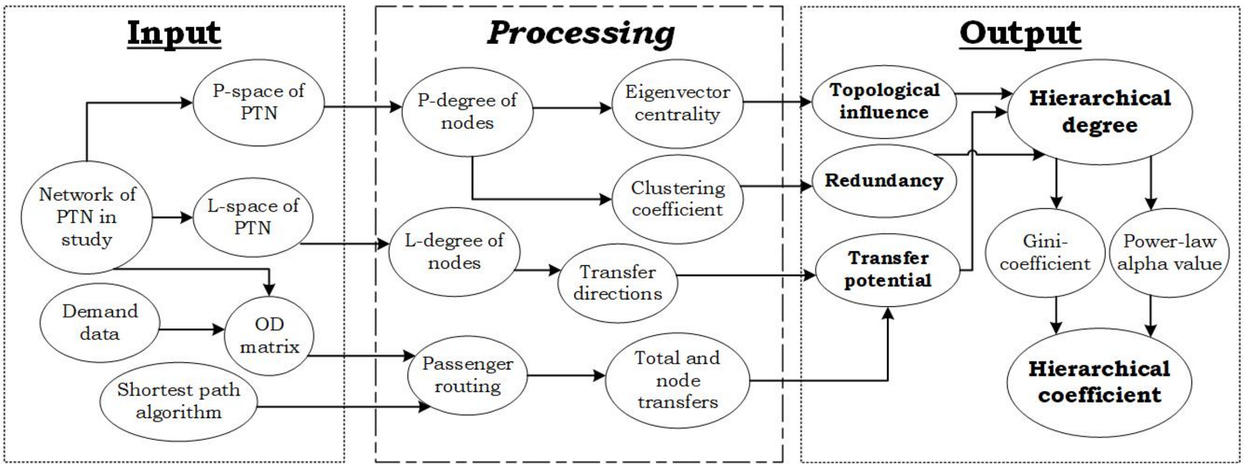


Figure 2: Conceptual overview of the hierarchical degree and hierarchical coefficient

3.4. Conceptual overview

In figure 2 an overview of the in- and output for the hierarchical degree metric is given. This figure summarizes the required input for each of the elements followed by the way this input is processed to determine the output for each of the elements and the hierarchical degree. Based on the distribution of the hierarchical degree among the nodes, the hierarchical coefficient can be determined. In order to elaborate what the generic requirements for the data of the input are, these are elaborated in the following section.

4. Application of hierarchy metric

In order to determine the values for the metric, a program code is developed[1] to apply the metric to case-studies. The case-studies include the two Dutch cities Amsterdam and Rotterdam where the former is subjected to a scenario study as well. For the Amsterdam PTN, one scenario includes the network state before the North-South line (NZL) while for the other scenarios possible future lines are evaluated. The Rotterdam PTN is evaluated to compare both of the city networks.

As explained throughout the methodology, an adjusted space representation is required which is visualized in figure 3 (inspired from [54]). In this figure, the upper network shows the real network with two nodes and three lines while the lower network shows how it is modeled to facilitate transfers and account for origins and destinations. It should be noted that the lower network is considered as a directed graph for which (directional) link weights are used. The link weights for intra-node changes (transfers) are based on the frequency of the incoming node while the links weights between nodes is based on the distance and the speed of the modality serving the link. Based on this network representation, the shortest path, including the location of transfers, can be determined for a demand comprising OD matrix.

[1] Retrievable from: `https://github.com/Abel287/MasterThesis.git`

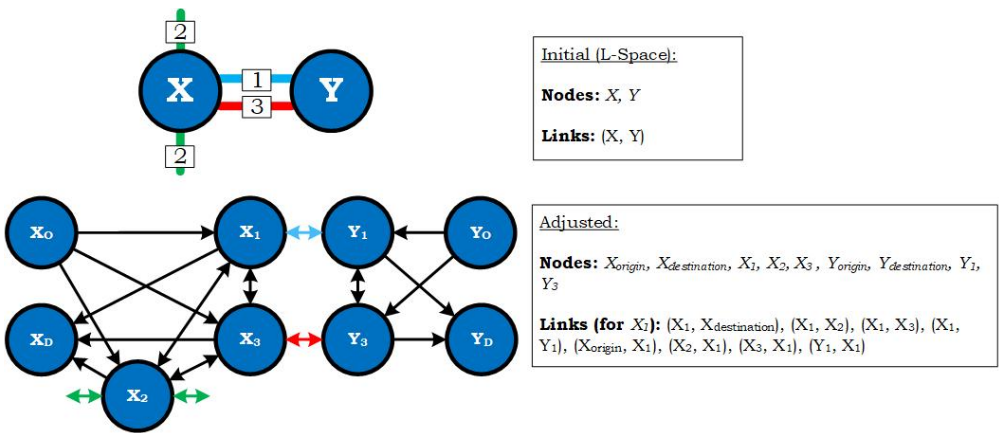


Figure 3: Adjusted graph representation of nodes to facilitate transfers where the top half represents the real network and the bottom half the adjusted space to incorporate transfers

4.1. Amsterdam PTN results

In figure 4 the geographical location of the nodes with their hierarchical degree is visualized.

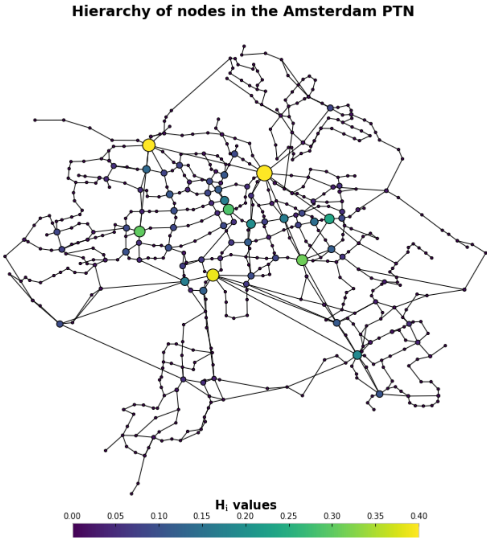


Figure 4: Map of the spatial distribution of the hierarchy of nodes in the Amsterdam PTN

In this figure, it is clearly visible the hierarchical nodes are spread out through the network with hierarchical nodes to every side of the network with the exclusion of the northern part of the city. It is notable there are only a few nodes which score a high value for the hierarchical degree while there are many non-hierarchical nodes. The sub-divided network into levels is shown in figure 5.

By evaluating the alpha coefficient for the Amsterdam PTN, which is -1.643, and the Gini-coefficient of 0.902, the hierarchical coefficient of the network is determine to be 0.939, indicating the hierarchy in the Amsterdam PTN is quite close to one. Hence, the structure of the network can be considered as hierarchical with a high spread of hierarchical degree. In order to compare this result, different states of the network and different networks have to be evaluated.

For the different elements in the Amsterdam PTN, a correlation analysis is done to clarify how the different elements and the hierarchical degree are correlated. An overview of the correlation is shown in table 2. From the correlation analysis it is notable that the transfer potential (e^C) correlates most with the hierarchical degree closely followed by the topological influence (e^A) while the (non-)redundancy (e^B) correlates least with the hierarchical degree. Among the elements, the topological influence and (non-)redundancy have the highest correlation coefficient while the topological coefficient and transfer potential have the lowest correlation.

Table 2: Correlation analysis for the Amsterdam PTN

Correlation	$\mathbf{H_i}$	$\mathbf{e^A}$	$\mathbf{e^B}$
$\mathbf{e^A}$	0.777		
$\mathbf{e^B}$	0.646	0.722	
$\mathbf{e^C}$	0.794	0.605	0.661

Therefore, it is remarkable that the non-redundancy correlates least with the hierarchical degree but correlates most with the other elements.

4.1.1. Amsterdam scenarios

For the scenario regarding the previous network state (before the opening of the NZL) it can be noted is that the H_i scores are generally lower than in the base year. The values for the hierarchical degree are more equally distributed in the base year, indicating the hierarchy is more spread out throughout the network implying relatively more nodes became important in the contemporary situation. Based on the results of the elements, mainly an increase in topological influence has led to a more evenly spread of hierarchy among the high scoring nodes. The power-law coefficient and Gini-coefficient for the network pre-NZL are -1.588 and 0.891 respectively which are a little lower than in the base year. The hierarchical coefficient can now be calculated and is 0.930 for the pre-NZL network, about a hundredth lower than for the base year network.

For the future scenarios, an overview of the comparison of the results for each scenario is shown in table 3. In the second column, the effect of the scenario in relation to existing policy is evaluated. The more positive a score is, the more it aligns with policy which is mainly focused towards becoming robust and relieving the central station. In the last column, an estimation of the implementation costs are shown based on the routing, existing infrastructure, length of the route and necessity of additional civil engineering works. These values are just rough estimations and are not based on any kind of source. It should be noted that any of these additional lines would be a significant investment for the region. The R2 scenario is perceived to be the most desirable in terms of the hierarchical structure but is the most expensive in terms of implementation. Following from the results for the hierarchical changes, a substantiation for policies towards an expansion of the metro-network can be augmented.

4.2. Rotterdam PTN results

In figure 6 a map showing the nodes and their hierarchical degree is shown in which it is clearly visible most hierarchical nodes are located in the center of the city. Furthermore, the northern, eastern and southern areas of the city appear to have barely any hierarchical nodes. In figure 7 the nodes of the network are divided in the same different sub-levels as the Amsterdam PTN.

Table 3: Comparison of the future scenarios for the Amsterdam PTN

Scenario	New route	Desirable hierarchical change	Implementation costs
G1	Isolatorweg - Centraal Station	+ -	+
G2	Isolatorweg - Noorderpark	+ -	-
R1	Schiphol Airport - Centraal Station	+	-
R2	Schiphol Airport - Muiderpoort	++	- -

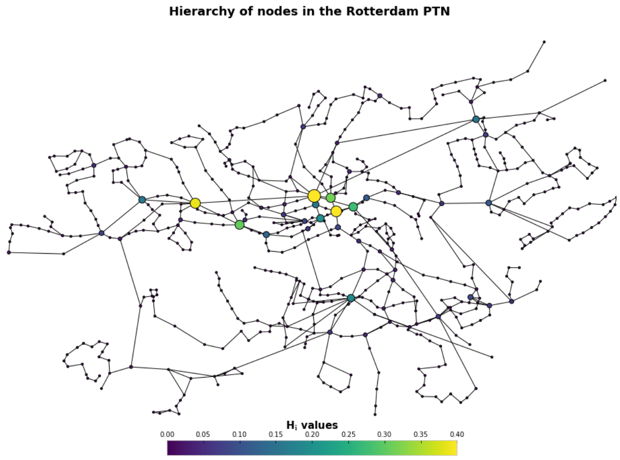


Figure 6: Map of the spatial distribution of the hierarchy of nodes in the Rotterdam PTN

With the alpha coefficient of -1.391 and a Gini-coefficient of 0.954 for Rotterdam PTN, the hierarchical is determined to be 0.967, indicating the Rotterdam PTN can be considered significantly more hierarchical than the Amsterdam PTN. For the different elements in the Rotterdam PTN, a correlation analysis is executed to provide an overview of how the different elements and the hierarchical degree are correlated, which is shown in table 4. Furthermore, by comparing the correlation analysis of the Rotterdam network to the Amsterdam network, there might be some patterns to be identified. The correlation for both the topological influence and the non-redundancy with the hierarchical degree for the Rotterdam PTN is quite similar to the Amsterdam network but the correlation between the transfer potential and the hierarchical degree is considerably lower. This could be explained by some nodes in the Rotterdam PTN being located further from the center and having a relatively low hierarchical degree, but being an important transfer hub for regional passengers. These nodes do generally have a lower influence as many of the influential nodes are all located in the city center and the peripheral transfer hubs are not connected to all of these.

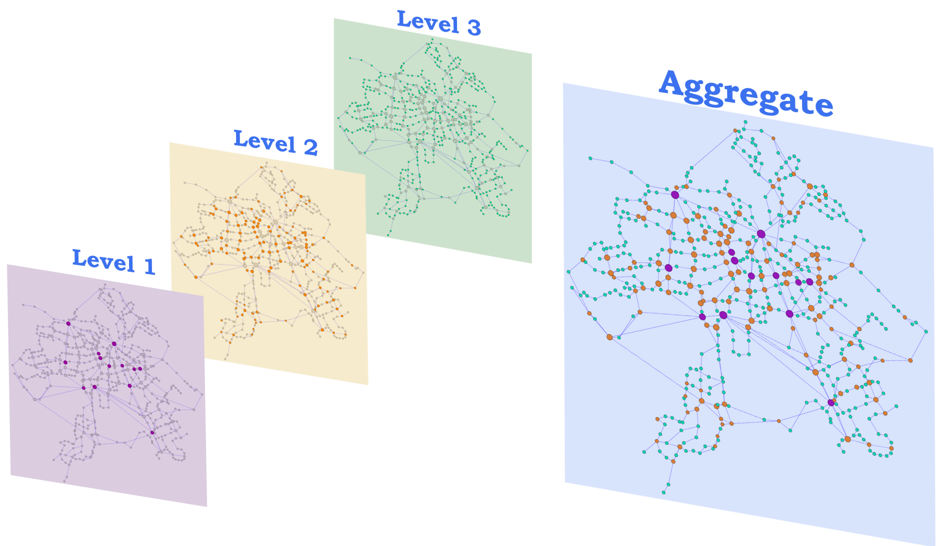


Figure 5: Overview of nodes in the Amsterdam PTN divided into three hierarchical levels

Table 4: Correlation analysis for the Rotterdam PTN with Amsterdam values in parentheses

Correlation	H_i	e^A	e^B
e^A	0.773 (0.777)		
e^B	0.652 (0.646)	0.531 (0.722)	
e^C	0.753 (0.794)	0.467 (0.605)	0.722 (0.661)

Emphasizing this assumption is the correlation between the topological influence and the transfer potential, which is much lower for the Rotterdam PTN than for the Amsterdam PTN. This is likely caused by influential nodes being located in the city center while transfer hubs are also located towards the edges of the network. The correlation between the topological influence and redundancy is also much lower for the Rotterdam PTN which could be explained as many of the influential nodes in the city center are also redundant due to many mutual connections. Lastly, the correlation between the redundancy and transfer potential appears to be higher for the Rotterdam PTN which could mean the important transfer hubs in the Rotterdam PTN are less redundant than in Amsterdam offering fewer locations for transfers.

4.2.1. Differences between the Amsterdam and Rotterdam network

For the distinct elements, the differences in values between the Amsterdam and Rotterdam networks are only minor while the real difference is found in the location of the high scoring nodes. For the Amsterdam PTN, in general, the high scoring nodes after the central station are located at train stations around four to five kilometers from the central station itself. On the other hand, the high scoring nodes for Rotterdam after the central station are located in close proximity to the central station while train stations further away from the central station score lower in general. Furthermore, the Rotterdam PTN has a much higher share of zero scoring nodes in terms of hierarchy indicating that a smaller share of nodes functions as a transfer location.

The Amsterdam PTN appears to have better connections between the decentralized nodes by the means of a (nearly complete) ring structure, which relieves the pressure on the center of the network. For the Rotterdam network, no such ring structure can be identified which could be a reason a lot of the hierarchical nodes are located in the center. Therefore, even though the hierarchical coefficient of the Rotterdam PTN is higher than the coefficient of Amsterdam, the network structure of Amsterdam appears to be more balanced and robust in this case, by having a more scattered layout of its hierarchical nodes. Hence, for this comparison, the hierarchical coefficient does not explain the full picture while analyzing the location and scattering of the hierarchical structure does provide additional value.

4.3. Potential applications

While these results provide interesting outcomes, the linkage to its general potential applications for policy is rather unclear. For the case-studies and scenarios conclusions can be drawn, but for network structures and hierarchical structures in general the results remain limited. Furthermore, the linkage to how policies and strategical decisions can be determined based on the results for the hierarchical degree could be clarified more. Consequently, these practical applications are further elaborated for bottlenecks, vulnerability and cascading failures.

For bottlenecks, each of the three elements for the hierarchical degree offers a different perspective on the identification of bottlenecks. The topological influence provides a method to identify which nodes are most influential in which a high difference between influence in nodes could be an indication one node has too much influence and is prone to becoming a bottleneck. Connected to this, the (non-)redundancy determines the redundancy of a node where nodes with low redundancy are crucial as transfer location and prone to being a bottleneck too. Lastly, the transfer potential is used to determine where the highest share of transfers take place in a network. If one node offers the most transfers in the network and other nodes only have a fraction of its transfer share, also taking into account logarithms are used, this node is susceptible to being a bottleneck.

The concepts of vulnerability and cascading failures can be linked to every element of the hierarchical degree as each of these elements implicitly values vulnerability. A higher value for the influence indicates the network would be more sensitive to disruptions at this specific node. Furthermore, the less redundant a node is, the higher the vulnerability and the higher the transfer potential the more prone it is to disruptions. As the different elements can clearly be linked to the concept of vulnerability, it is important to understand how vulnerability can be assessed based on values for the hierarchical degree. It should be noted that vulnerability of a network is considered a normative characteristic while the hierarchy of a network is descriptive. This implies that a higher vulnerability would be considered worse while a higher hierarchy value cannot be valued as better or worse in advance. Building on the assumption that the severity of a disruption is correlated with the hierarchy of a node at which the disruption takes place, it can be assumed that nodes with a higher hierarchical degree are more vulnerable in general. This can be explained as nodes with a high hierarchical degree are likely to be more critical in the network due to its influence, non-redundancy and transfer potential. Therefore,

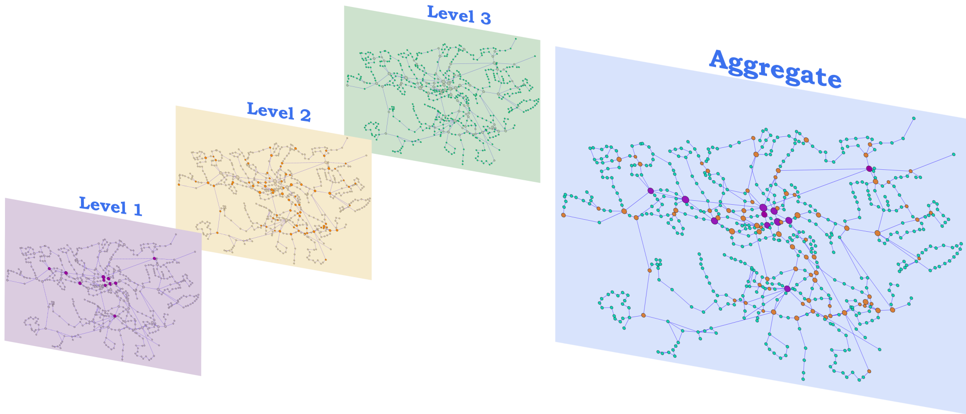


Figure 7: Overview of nodes in the Rotterdam PTN divided into three hierarchical levels

the hierarchical nodes are considered more vulnerable in general.

5. Conclusion

Throughout this article, a newly developed metric to quantify hierarchy in PTN has been elaborated and applied to different case-studies. There are three notable approaches to hierarchy in PTN which are link-, line- and node-based hierarchy. While each of the approaches offers its strengths and weaknesses, this study suggests a combined approach where node-based hierarchy is enhanced with link/line travel demand. Despite this approach being a combination of the three approaches, the node-based approach is reflected predominantly which could lead to underemphasizing the importance of links and lines.

by combining different aspects of the definition for hierarchy, the metric is developed from the elements topological influence, redundancy and transfer potential which cover the aspects of the definition. These three elements have common grounds with previous works but its combination is unobserved in existing literature. While these three elements are seen as complementary, other indicators (such as e.g. overlapping degree or betweenness centrality) could, despite their drawbacks, also be used to define hierarchy in PTN.

The results of the case-studies have shown some expected results but also notable differences between the network structures. The case-studies have provided some unique insights in the hierarchical structures of the Amsterdam and Rotterdam PTN and based on these findings, the networks can be compared and evaluated. While the hierarchical coefficient has been developed based on hierarchy distribution throughout the network, the spatial location of the hierarchical nodes in the network, which is lacking in the coefficient, turned out to be more meaningful for the comparison.

For the scenarios, the changes in hierarchical degree for (in)directly affected nodes provides some insight in the effects of the removal or addition of a line to the network. It is, however, debatable if an analysis in the changes in hierarchical degree can determine if a scenario would be desirable or not. The increase in a hierarchy for a network could mean only a few nodes increase in hierarchy (the rich getting richer) while it could also mean the second order nodes could increase in function, relieving bottlenecks. Therefore, only evaluating the changes in hierarchical degree does only provide a limited understanding of the effects of a scenario on the network. A more in depth approach to a scenario in order to evaluate its strategic prospects should be applied for which the hierarchy metric can be used as a part.

Following from the research, a first recommendation relates to the approach of defining hierarchy in a network. Throughout this study, a predominantly node-based approach is applied where additional insights could be achieved using a link- or line-based approach for hierarchy. Furthermore, a multi-layer approach for hierarchical networks is referred to but the definition for layers are missing for and could be an interesting subject for further research. Hierarchical levels have been discussed multiple times but how these levels can be distinguished has not fully been grasped. Three different levels have been identified which are high-hierarchical, low-hierarchical and non-hierarchical but these are merely based on the average of the interval rather than an objective and scientific substantiated method. Therefore, it would be recommended as a topic for further research how hierarchical levels in PTN can clearly be presented.

Another recommendation relates to the concept of vulnerability which has been linked to hierarchy. According to this study, vulnerability could be more intelligible by looking at the hierarchical structure. However, for the connection between vulnerability and hierarchy only the surface has been scratched and this could be subjected to further research. Furthermore, the connection between hierarchy and different PTN related phenomena such as e.g. transit-oriented development could be interesting topics for the implications of the metric for developing strategic network related policy in the future. Lastly, the empirical perspective for this study has been somewhat in the back seat as the metric consists of two topological elements and the empirical aspect is only reflected in the transfer potential. Furthermore, for the travel demand, the relatively straightforward AON assignment is applied while more sophisticated traffic assignment methods such as the probabilistic or stochastic assignment may lead to additional insights. Connected to this, the use of observed data, such as smart-card or GPS data could overcome the need for estimated demand models in general. Therefore, further research to how the empirical element of hierarchy could be better reflected is suggested.

References

[1] B. Richards, Future transport in cities, Taylor & Francis, 2012.

[2] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C. J. Perez-Vicente, Y. Moreno, A. Arenas, Diffusion dynamics on multiplex networks, Physical review letters 110 (2) (2013) 028701.

[3] B. Min, S. Do Yi, K.-M. Lee, K.-I. Goh, Network robustness of multiplex networks with interlayer degree correlations, Physical Review E 89 (4) (2014) 042811.

[4] A. Aleta, S. Meloni, Y. Moreno, A multilayer perspective for the analysis of urban transportation systems, Scientific reports 7 (2017) 44359.

[5] D. Gattuso, E. Miriello, Compared analysis of metro networks supported by graph theory, Networks and Spatial Economics 5 (4) (2005) 395–414.

[6] J. Zhang, M. Zhao, H. Liu, X. Xu, Networked characteristics of the urban rail transit networks, Physica A: Statistical Mechanics and its Applications 392 (6) (2013) 1538–1546.

[7] D. Levinson, Network structure and city size, PloS one 7 (1) (2012) e29721.

[8] C. Roth, S. M. Kang, M. Batty, M. Barthelemy, A long-time limit for world subway networks, Journal of The Royal Society Interface 9 (75) (2012) 2540–2550.

[9] S. Derrible, C. Kennedy, Applications of graph theory and network science to transit network design, Transport reviews 31 (4) (2011) 495–519.

[10] E. Mones, L. Vicsek, T. Vicsek, Hierarchy measure for complex networks, PloS one 7 (3) (2012) e33799.

[11] B. M. Yerra, D. M. Levinson, The emergence of hierarchy in transportation networks, The Annals of Regional Science 39 (3) (2005) 541–553.

[12] M. Lee, H. Barbosa, H. Youn, P. Holme, G. Ghoshal, Morphology of travel routes and the organization of cities, Nature communications 8 (1) (2017) 2229.

[13] G. Jian, Z. Peng, Z. Chengxiang, Z. Hui, Research on public transit network hierarchy based on residential transit trip distance, Discrete Dynamics in Nature and Society 2012.

[14] S. A. Bagloee, A. A. Ceder, Transit-network design methodology for actual-size road networks, Transportation Research Part B: Methodological 45 (10) (2011) 1787–1804.

[15] R. Van Nes, Design of multimodal transport networks: A hierarchical approach, Citeseer.

[16] E. Ravasz, A.-L. Barabási, Hierarchical organization in complex networks, Physical review E 67 (2) (2003) 026112.

[17] A. De Montis, M. Barthélemy, A. Chessa, A. Vespignani, The structure of interurban traffic: a weighted network analysis, Environment and Planning B: Planning and Design 34 (5) (2007) 905–924.

[18] O. Cats, Topological evolution of a metropolitan rail transport network: The case of stockholm, Journal of Transport Geography 62 (2017) 172–183.

[19] A. Reggiani, P. Nijkamp, D. Lanzi, Transport resilience and vulnerability: The role of connectivity, Transportation research part A: policy and practice 81 (2015) 4–15.

[20] C. Von Ferber, T. Holovatch, Y. Holovatch, V. Palchykov, Public transport networks: empirical analysis and modeling, The European Physical Journal B 68 (2) (2009) 261–275.

[21] Y. Zhou, J. Wang, H. Yang, Resilience of transportation systems: concepts and comprehensive review, IEEE Transactions on Intelligent Transportation Systems.

[22] C. Von Ferber, T. Holovatch, Y. Holovatch, V. Palchykov, Network harness: Metropolis public transport, Physica A: Statistical Mechanics and its Applications 380 (2007) 585–591.

[23] C. Von Ferber, T. Holovatch, Y. Holovatch, Attack vulnerability of public transport networks, in: Traffic and Granular Flow'07, Springer, 2009, pp. 721–731.

[24] Q. Gu, X.-H. Yang, W.-L. Wang, B. Wang, Research on urban public transport networks based on complex networks [j], Computer Engineering 20.

[25] Y. Hu, D. Zhu, Empirical analysis of the worldwide maritime transportation network, Physica A: Statistical Mechanics and its Applications 388 (10) (2009) 2061–2071.

[26] M. E. Newman, Assortative mixing in networks, Physical review letters 89 (20) (2002) 208701.

[27] L. Xia-Miao, Z. Ming-Hua, Z. Jin, L. Ke-Zan, Hierarchy property of traffic networks, Chinese Physics B 19 (9) (2010) 090510.

[28] S. Derrible, Network centrality of metro systems, PloS one 7 (7) (2012) e40575.

[29] H. Soh, S. Lim, T. Zhang, X. Fu, G. K. K. Lee, T. G. G. Hung, P. Di, S. Prakasam, L. Wong, Weighted complex network analysis of travel routes on the singapore public transportation system, Physica A: Statistical Mechanics and its Applications 389 (24) (2010) 5852–5863.

[30] P. Van Mieghem, H. Wang, X. Ge, S. Tang, F. A. Kuipers, Influence of assortativity and degree-preserving rewiring on the spectra of networks, The European Physical Journal B 76 (4) (2010) 643–652.

[31] X. Qing, Z. Zhenghu, X. Zhijing, W. Zhang, T. Zheng, Space p-based empirical research on public transport complex networks in 330 cities of china, Journal of Transportation Systems Engineering and Information Technology 13 (1) (2013) 193–198.

[32] W. Ru, C. Xu, Hierarchical structure, disassortativity and information measures of the us flight network, Chinese Physics Letters 22 (10) (2005) 2715.

[33] S. Porta, E. Strano, V. Iacoviello, R. Messora, V. Latora, A. Cardillo, F. Wang, S. Scellato, Street centrality and densities of retail and services in bologna, italy, Environment and Planning B: Planning and design 36 (3) (2009) 450–465.

[34] A. De Bona, K. Fonseca, M. Rosa, R. Lüders, M. Delgado, Analysis of public bus transportation of a brazilian city based on the theory of complex networks using the p-space, Mathematical Problems in Engineering 2016.

[35] T. Majima, M. Katuhara, K. Takadama, Analysis on transport networks of railway, subway and waterbus in japan, in: Emergent Intelligence of Networked Agents, Springer, 2007, pp. 99–113.

[36] B. Berche, C. Von Ferber, T. Holovatch, Y. Holovatch, Resilience of public transport networks against attacks, The European Physical Journal B 71 (1) (2009) 125–137.

[37] J. Sienkiewicz, J. A. Hołyst, Statistical analysis of 22 public transport networks in poland, Physical Review E 72 (4) (2005) 046127.

[38] H. Lu, Y. Shi, Complexity of public transport networks,

Tsinghua Science and Technology 12 (2) (2007) 204–213.
[39] M. Barthelemy, Betweenness centrality in large complex networks, The European physical journal B 38 (2) (2004) 163–168.
[40] S. Lämmer, B. Gehlsen, D. Helbing, Scaling laws in the spatial structure of urban road networks, Physica A: Statistical Mechanics and its Applications 363 (1) (2006) 89–95.
[41] F. Xie, D. Levinson, Topological evolution of surface transportation networks, Computers, Environment and Urban Systems 33 (3) (2009) 211–223.
[42] A. Reynolds-Feighan, Traffic distribution in low-cost and full-service carrier networks in the us air transportation market, Journal of Air Transport Management 7 (5) (2001) 265–275.
[43] D. M. Levinson, F. Xie, S. Zhu, The co-evolution of land use and road networks, Transportation and traffic theory (2007) 839–859.
[44] G. Gentile, K. Noekel, Modelling public transport passenger flows in the era of intelligent transport systems, Gewerbestrasse: Springer International Publishing.
[45] O. Tamin, L. Willumsen, Transport demand model estimation from traffic counts, Transportation 16 (1) (1989) 3–26.
[46] M. S. Fiorenzo-Catalano, R. Van Nes, P. H. Bovy, Choice set generation for multi-modal travel analysis, European journal of transport and infrastructure research EJTIR, 4 (2).
[47] J. Schweizer, A. Danesi, F. Rupi, E. Traversi, Comparison of static vehicle flow assignment methods and microsimulations for a personal rapid transit network, Journal of advanced transportation 46 (4) (2012) 340–350.
[48] G. Flötteröd, Traffic assignment for strategic urban transport model systems (2015).
[49] A. Sumalee, Z. Tan, W. H. Lam, Dynamic stochastic transit assignment with explicit seat allocation model, Transportation Research Part B: Methodological 43 (8-9) (2009) 895–912.
[50] J.-D. Schmöcker, M. G. Bell, W. H. Lam, Importance of public transport, Journal of Advanced Transportation 38 (1) (2004) 1–4.
[51] H. B. Celikoglu, H. K. Cigizoglu, Modelling public transport trips by radial basis function neural networks, Mathematical and computer modelling 45 (3-4) (2007) 480–489.
[52] O. A. Nielsen, G. Jovicic, A large scale stochastic timetable-based transit assignment model for route and sub-mode choices, in: Transportation Planning Methods. Proceedings of Seminar F, European Transport Conference, 27-29 September 1999, Cambridge, UK., Vol. 434, 1999.
[53] R. G. Bettinardi, Spontaneous brain activity: how dynamics and topology shape the emergent correlation structure (2016).
[54] J. Luo, D. E. Whitney, C. Y. Baldwin, C. L. Magee, Measuring and understanding hierarchy as an architectural element in industry sectors, Harvard Business School, 2009.