HCU | HafenCity Universität Hamburg

Classifying Transit-Oriented Development Neighborhoods Based on Network Analysis

by

Hsiao-Hui Chen

Under the supervision of

Prof. Udo Dietrich

A thesis submitted for the degree of Doctor of Philosophy of
Department of Resource Efficiency in Architecture and Planning of
HafenCity Universität Hamburg, Germany

November 2021

Abstract

The aim of this research is to propose a method for developing and exploring the typology of public transportation stations based on network analysis. Transit Oriented Development (TOD), a type of urban development that aims at increasing both the usage of public transportation and the walkability within the neighborhood around the station, considers the influences of urban form, built environment, traffic flow and movement patterns in order to integrate transportation, land use and environmental policies. This notion has later been extended into the "Network TOD", which is the network approach on a broader geographical scale than TOD, as it presents the potential of both creating the livable and attractive neighborhoods around the stations and at the same time shaping the polycentric cities in order to mitigate urban sprawl.

The strength of the network approach consists in offering a more holistic perspective of evaluating the node, i.e. the station, based on the role it plays in the whole network system. In the case of urban and transportation planning, the benefit of employing this approach is that it not only focuses on the quality of urban environment within the boundary of the TOD neighborhood, but also considers the relationship between TOD neighborhoods.

In the last few years, due to the development of information technology and the increased availability of data sources, data-driven urban morphology research has vastly advanced. By employing a machine learning approach, this thesis develops an automatic, parametric, scalable and reproducible method based on network analysis and aims to meet the following objectives:

- Assessing the existing models in network analysis and dealing with limitations of their validity (such as the size effect and the placement effect) on the values of the indicators. The results can be helpful for determining the optimal size and the location of the catchment area in order to mitigate the size and placement effects on network analysis indicators.
- Evaluating the connectivity and resilience of two types of networks, namely public transit network and street network, by exploring their geometric and topological properties.
- Quantitatively classifying and evaluating the importance of stations in the railway public transport system based on 1) their structural importance in the transportation network; 2) topological characteristics of street network in the neighborhoods; and 3) accessibility, density, and diversity of the points of interest in the neighborhoods.
- Determining a suitable location of an intervention that matches the challenges faced by different types of neighborhoods in improving the walkability in the neighborhood.

Finally, being motivated by the outbreak of COVID-19 in 2020, we hope that this research can help urban planners to make informed and data-driven decisions on the location of facilities and services, such as temporary test stations and mobile vaccination centers, during emergencies and crises (for example, a pandemic).

Zusammenfassung

Das Ziel dieser Forschung ist es, eine Methode zur Entwicklung und Erforschung der Typologie von ÖPNV-Stationen basierend auf der Netzanalyse vorzuschlagen. Transitorientierte Entwicklung (TOE), eine Form der Stadtentwicklung, die darauf abzielt, sowohl die Nutzung öffentlicher Verkehrsmittel als auch die Begehbarkeit des Bahnhofsquartiers zu erhöhen, berücksichtigt die Einflüsse von Stadtform, gebauter Umgebung, Verkehrsfluss und Bewegungsmustern, um Verkehrs-, Landnutzungs- und Umweltpolitik zu integrieren. Dieser Begriff wurde später als „Netzwerk TOE“ erweitert, um den Netzwerkansatz auf eine breitere geografische Ebene als TOE anzuwenden, da er das Potenzial bietet, sowohl lebenswerte und attraktive Viertel rund um die Bahnhöfe zu schaffen als auch die polyzentrischen Städte zu gestalten, um die Zersiedelung der Städte einzudämmen.

Die Stärke des Netzwerkansatzes besteht darin, eine ganzheitlichere Perspektive der Bewertung des Knotens, d. h. der Station, basierend auf seiner Rolle im gesamten Netzwerksystem, anzubieten. Bei der Stadt- und Verkehrsplanung liegt der Vorteil dieses Ansatzes darin, dass er sich nicht nur auf die Qualität des städtischen Umfelds innerhalb der Grenzen eines TOE-Viertels konzentriert, sondern auch die Beziehung zwischen den TOE-Vierteln berücksichtigt. In den letzten Jahren hat die datengetriebene Stadtmorphologieforschung aufgrund der Entwicklung der Informationstechnologie und der zunehmenden Verfügbarkeit von Datenquellen enorme Fortschritte gemacht.

Mithilfe des maschinellen Lernens entwickelt diese Arbeit eine auf Netzwerkanalyse basierende parametrische Methode, die darauf abzielt, die folgenden Ziele zu erreichen:

- Bewertung der bestehenden Modelle in der Netzwerkanalyse und Umgang mit den Einschränkungen ihrer Gültigkeit (z.B. Größeneffekt und Platzierungseffekt) hinsichtlich der Werte der Indikatoren. Die Ergebnisse können helfen, die optimale Größe und Lage des Einzugsgebietes zu bestimmen, um die Größen- und Platzierungseffekte auf Indikatoren der Netzanalyse abzumildern;
- Bewertung der Konnektivität und Belastbarkeit von zwei Arten von Netzen, und zwar öffentlichem Nahverkehrsnetz und Straßennetz, durch Untersuchung ihrer geometrischen und topologischen Eigenschaften;
- Quantitative Einordnung und Bewertung der Bedeutung von Bahnhöfen im ÖPNV-System anhand von 1) ihrer strukturellen Bedeutung im Verkehrsnetz; 2) topologischen Eigenschaften des Straßennetzes in den Stadtvierteln und 3) Zugänglichkeit, Dichte und Vielfalt der Anziehungspunkte in den Vierteln;
- Bestimmung eines geeigneten Ortes für eine Intervention, die den Herausforderungen entspricht, denen sich unterschiedliche Stadtvierteltypen bei der Verbesserung der Begehbarkeit im Viertel gegenübersehen.

Schließlich hoffen wir, motiviert durch den Ausbruch von COVID-19 im Jahr 2020, dass diese Forschung Stadtplanern helfen kann, fundiertere und datengestützte Entscheidungen über die Platzierung von Einrichtungen (z.B. temporären Teststationen und mobilen Impfzentren) bei Notfällen und Krisen zu treffen.

Keywords

network analysis, street network, transit network, General Transit Feed Specification, size effect, placement effect, topology, centrality, Transit Oriented Development, machine learning

Acronyms

Application Programming Interface	API
Bike Rental Station	BRS
General Transit Feed Specification	GTFS
Hamburger Verkehrsverbund	HVV
Open Street Map	OSM
Transit-Oriented Development	TOD
Transit Catchment Area	TCA
Pedestrian Catchment Area	PCA
Point of Interest	POI
Public Transport Networks	PTN
Railway Public Transport	RPT

Nomenclature

Symbol	Name	Unit
γ	Degree of Connectivity	link / node
N	Number of nodes	node
M	Number of links	link
C_k	Degree Centrality	
C_c	Closeness Centrality	
C_b	Betweenness Centrality	
C_N	Normalized Closeness Centrality	
V	Average path length (in network distance)	link / path
C	Clustering Coefficient	
C_avg	Average Clustering Coefficient	Clustering Coefficient / node
k	Node Degree	links (of the specific node)
k_avg	Average Node Degree	degree / node
$p(k)$	Node Degree Distribution (i.e. fraction of nodes have k degree)	
F_{vu}	Service frequency	train / hour
d	Length of a link of the idealized network	
D	Side length of the idealized network	link
L	Number of links on each side of the catchment area of the idealized network	link

Acknowledgement

I would like to express my deep appreciation and indebtedness to my supervisor, Professor Udo Dietrich, for always offering valuable advice and timely feedback to guide me from the very beginning until the last moment of this research. He has inspired me with his research and extensive knowledge of this topic and aided me with technical challenges. I was given a lot of flexibility to explore and ask questions. He has always been positive and encouraging, while also carefully guiding me through the process and pointing out the potential risks and problems. Under his guidance and supervision, my ideas were able to be developed and proceeded, while at the same time benefitting from the occasional warnings leading to the readjustment of the direction. I especially appreciate his welcoming attitude, which encouraged me to present and discuss any thoughts even before they were fully developed. The difference in our backgrounds, applied physics and social science, did not hinder our communication but rather became a source of inspiration for exploring interesting perspectives and approaches. Also, I have greatly benefited from his original insights and superb ability to comprehend the essence of different subjects within very short time. I am extremely grateful for having not just a thesis supervisor but also a mentor for my academic career.

I am very much indebted to my soulmate and partner, Andriy Ilyuk. This thesis would not have been possible without his sense of humor, constant support and understanding by my side.

My deepest thanks go to my parents, Wei-Zen Chen and Hui-Chu Yang, for their unconditional love and support. This thesis is dedicated to them.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Complex urban system & spatial network

From the perspective of complex network theory, cities can be viewed as complex, dynamic and self-organizing systems (Batty & Longley, 1994) (Portugali, 2000) (Batty M. , 2005). “Under the seeming disorder of the old city, wherever the old city is working successfully, is a marvelous order for maintaining the safety of the streets and the freedom of the city. It is a complex order” (Jacob, 1961). Cities have originally emerged, developed, and eventually resulted in complex networks that consist of many links. They can be viewed as a network of networks, where different structures are tailored in the web and closely intertwined (Blumenfeld-Lieberthal & Portugali, 2010).

“Urban spatial networks, such as streets, paths, and transit lines, organize the human dynamics of complex urban systems” (Boeing, 2017) and are changing in the process of urban development. Therefore, theory of complex networks has been increasingly adopted to model and analyze the processes in urban systems (Masucci, Smith, Crooks, & Batty, 2009).

Adopting the network analysis to the study of urban spatial networks as complex systems can improve our understanding on the structure and evolution of a city. Results of network analysis can also “open new perspectives in the scientific relation between city planning and complex networks, stimulating the debate on the effectiveness of the set of knowledge that statistical physics can contribute for city planning and urban morphology studies” (Strano, Hao, Everson, & Evans, 2013). These new perspectives are particularly important to the challenges we are facing today.

1.2 The challenges

Urban form and land use are shaped by the pattern of roads and streets that represents an essential backbone of the cities. Therefore, the current level of street connectivity has strong association with the increase in the energy use and CO_2 emissions in the future

(Barrington-Leigh & Millard-Ball, 2019). By studying the urban networks to reveal the evolution of street structures and the patterns of movement flows in the city, previous research has found that currently there are two trends of the development that can be observed.

1) **Increasing connections and the changes in urban structure.** On the one hand, increased connections in the street networks are changing the land use pattern. The investigation of the highly urbanized areas found that cities with a long history, like London, are gradually evolving from a more loop-like structure towards a more tree-like structure (Masucci, Stanilov, & Batty, 2013). The loop-like structure is the typical feature of a planned city, and the tree-like structure is the typical feature of a self-organized city. Insertion of new short-cuts for route optimization, which leads to a more tree-like structure, also contributes to a more pedestrian-oriented system (Hespe & Sanders, 2019).
2) **Decreasing connectivity of street network and increasing energy use and CO_2 emissions.** On the other hand, in the less urbanized area, it is found that the disconnected street-network sprawl is increasingly found in more and more cities in the world (Barrington-Leigh & Millard-Ball, 2019) as the street network patterns reveal worrying trend towards decreased connectivity of street network. A less connected street network is clearly car oriented because it lacks density necessary for having a successful pedestrian system. This trend may continue to exist due to the rapid development of "global urbanization and its immediate consequences, including changes in patterns of food demand, circulation and land use" (Strano, et al., 2017).

Previous research in Europe and North America has shown that the fast development of urbanization, disconnected urban street networks and urban sprawl have led to the imbalances in cities, more frequent vehicle travel and the increase in traffic congestion, pollution, energy use and CO_2 emissions (Barrington-Leigh & Millard-Ball, 2019). This trend also has the effect of discouraging the use of public transportation (Zhang, Song, Nes, He, & Yin, 2019). Therefore, the relationship between urbanization and urban spatial network requires continuous monitoring, surveying, and investigation.

By combining complexity theory and network analysis, this research intends to propose a framework for Transit-Oriented Development (TOD) in order to investigate the challenges brought about by the urbanization and urban sprawl.

1.3 Transit-Orientated Development (TOD) and Node-Place Model as a framework for classifying TOD neighborhoods

The trend of urban sprawl and its effect of discouraging the use of public transportation make the TOD an even more pressing issue. “TOD emphasizes the development and opportunities provided by public transportation. It also underlines the integration and cooperation of transportation and land use” (Zhang, Song, Nes, He, & Yin, 2019).

One of the main targets of TOD is minimizing the walking distance between the public transport stations and the places for residential, business and leisure purposes (Finkenbinder, Britt, & Blair, 2010). This approach facilitates local living because of the proximate access to facilities, amenities, services, functions and works. At the same time, the residents can access other destinations by public transportation (Newman, Baum, Javanparast, O'Rourke, & Carlon, 2015). By shortening the walking distance, TOD facilitates local living and contributes to sustainable urban growth by reducing the use of private single-occupancy vehicles, the traffic congestion and the transportation costs (Renne, 2009) (Stutzer & Frey, 2008) and by increasing the public transport ridership (Jeffrey, Boulangé, Giles-Corti, Washington, & Gunn, 2019).

Typologies of public transportation stations and the surrounding catchment areas, i.e. the TOD neighborhoods, are commonly used to simplify their complex characteristics in order to study the spatial distribution of the urban structure (Zhang, Song, Nes, He, & Yin, 2019) and to assess their TOD potential. Classifying the stations enables policymakers, urban planners and designers to understand the strengths and weaknesses of the transit stops and to identify the train stations that require further development and intervention (Lyu, Bertolini, & Pfeffer, 2016) (Zemp, Stauffacher, Lang, & Scholz, 2011).

To develop typologies, TOD is often assessed using the Node-Place Model proposed by Bertolini (1999). This model provides a framework and an analytical tool for assessing the transit stop (“the node”) in the transportation system, the land use system in its surrounding areas (“the place”) and the relationship between the two (Monajem & Nosratian, 2015). According to Bertolini, “the Node Value indicates intensity and diversity of transportation system, which also indicates the potential for physical human interaction. The Place Value indicates to what extent this potential has been realized by the intensity and diversity of human activities” (Bertolini, 1999). Figure 1.1 illustrates that, under the Node-Place model, the TOD neighborhoods can be broadly categorized into two scenarios: the balanced and unbalanced TOD neighborhoods. And further differentiation can be made among five categories based on the Node and Place Values.

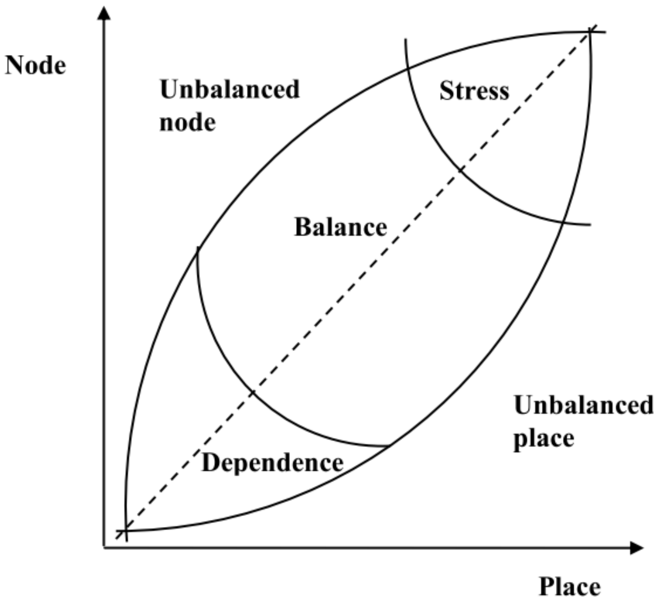


Figure 1.1 Node-Place Model (Bertolini, 1999).

i) Balanced scenarios

The balanced neighborhoods belong to the central part along the middle diagonal line of Figure 1.1, where both Node and Place Values are comparatively equal and are close to the average value of the study area because the interaction between the intensity and diversity of transportation system and human activities is "balanced" (Papa, Moccia, Angiello, & Inglese, 2013). The development potential of these neighborhoods are also higher because they match transit accessibility with demand (Jeffrey, Boulangé, Giles-Corti, Washington, & Gunn, 2019). The balanced neighborhoods can be further articulated into the following two sub-scenarios (Monajem & Nosratian, 2015).

- *Balanced - stressed:*

The "balanced-stressed" neighborhoods can be found at the top right of the line in the model. Both of their Node and Place Values are very high. The combination of a strong node and a strong place indicates very high level of the intensity and diversity of transportation system and human activities. Therefore, these neighborhoods are characterized by transportation congestion and very intensive land use. Furthermore, the great concentration of flows and activities in these neighborhoods also means that there could be conflicts and competition for the limited space. Intervention suggesting further development in these neighborhoods may lead to incompatibility in transportation and land use systems (Monajem & Nosratian, 2015) due to restricted space. In addition, it could also be argued that such neighborhoods have a high potential to be developed into new sub-centers, where all of the residents' everyday needs are satisfied by the local amenities and, therefore, reduce the need for transportation.

- Balanced-dependent:

The "balanced-dependent" neighborhoods are at the bottom left of the line, which indicates that both their Node and Place Values are very low. The combination of a weak node and a weak place indicates that the intensity and variety of both the transit system and the human activities are minimal. The actions suggested for these neighborhoods would be to increase both the node and the place aspects.

ii) Unbalanced scenarios

Besides the balanced scenarios, there are also two unbalanced ones, as one can see at the top left and bottom right of Figure 1.1. These scenarios also require further intervention in order to optimize the TOD outcomes. They can be articulated into the following sub-categories:

- *Unbalanced nodes:*

"Unbalanced nodes" are at the top left of Figure 1.1, where the Node Value is higher than the Place Value, representing the neighborhoods where the intensity and diversity of transportation system is more dominant than that of human activities. They have a great potential to be transformed into TOD neighborhoods if their Place Value can be increased. The current study will also emphasize on increasing their Place Value.

- *Unbalanced places:*

By contrast, "unbalanced places" are at the bottom right of Figure 1.1, where Place Value is higher than the Node Value. In these neighborhoods, the intensity and diversity of human activities is more dominant than that of the transportation system.

1.4 Walkability and TOD neighborhoods

Based on the framework of the Node-Place model, the current research focuses on the measurement of walkability in the model. Successful TOD maximizes accessibility by active transportation modes, such as cycling or walking. The walkability level of the neighborhood plays an important role for achieving this goal and, therefore, representing
the potential of a transit station to become a TOD neighborhood (Jeffrey, Boulangé, Giles-Corti, Washington, & Gunn, 2019).

A walkable neighborhood is characterized by greater street connectivity, high residential densities and mixed land uses (Giles-Corti, Foster, Koohsari, Francis, & Hooper, 2015). Such characteristics facilitate the convenient, proximate and safe access to and from the essential and local destination, such as supermarkets, jobs, retail shops, health and community services in order to meet the local residents' daily living (Jacobson & Forsyth, 2008) (Val, 2015). By encouraging the residents to walk and use the public transportation, the TOD neighborhood can achieve its full potential (Jeffrey, Boulangé, Giles-Corti, Washington, & Gunn, 2019)..

- Walkability and street network analysis

To sustain the walkability of a catchment area around the stations, the streets in the neighborhood and the streets that are connected to a station should be well-integrated (Monajem & Nosratian, 2015). Network analysis helps to explore the level of connectivity and also to identify the important nodes in the neighborhood street networks. Furthermore, for the catchment area around the station to become a TOD neighborhood, the importance of the station itself in the whole transit network should also be considered. If a station plays an important role in the transit network by providing higher accessibility to other areas in the city, the residents would be more motivated to use it and, therefore, increase the frequency of walking to the station. Therefore, this research intends to enhance our understanding of the walkability of TOD neighborhood by carrying out network analysis of two types of networks: the transit network and the street network in the catchment area around the station.

- Walkability and accessibility to point of interest (POI)

In addition, points of interest (POI) in the neighborhood act as the hubs of socio-economic activities and generate movement that facilitate the development commercial land use (Bernick & Freilich, 1998) (Bertolini & Spit, 1998). The combination of higher accessibility, diversity and density of POI makes the catchment area around the station a favorable location for setting up the business and, therefore, bring in more job opportunities (Bertolini & Spit, 1998) (Monajem & Nosratian, 2015). Therefore, accessibility, diversity and density of POI will also be considered when measuring the Place Value of the TOD neighborhoods.

1.5 Objective and research questions: Classification of TOD based on network analysis

Despite the importance of network analysis for TOD, to date the complexity of the TOD typologies has not been fully explored (Jeffrey, Boulangé, Giles-Corti, Washington, & Gunn, 2019). And there is still room to explore how the network analysis of transit network and street network in the TOD neighborhoods could be used for the planning and management of TOD (Val, 2015) (Schlossberg & Brown, 2004).

In the last few years, due to the powerful tool of information technology and the increased availability of data sources, the data-driven and computational urban morphology research has vastly advanced in the following aspects: 1) the ability to examine the patterns and structure of urban form, circulation and spatial order; 2) the comparison among the networks of different location, scales and types through Big Data; 2) the acceleration of the retrieval and analysis of the vast amount of data, especially geodata and data from social network software, in order to visualize network and create a quick and intuitive overview of a complex system (Boeing, 2017).

The current study will take advantage of the advancements in information technology and Big Data to systematically characterize a network in a structured and automated manner. With the computational application of a network analysis, the current research aims at answering the following questions.

- How to decide the size and location of the catchment area in order to mitigate the size and placement effects on the indicators of network analysis?
- What is the level of connectivity and resilience of the public transit network and the street network in Hamburg?

- How can one quantitatively classify and evaluate the importance of transportation stations in the railway public transport system?
- How to determine a suitable location for a new facility or an intervention that matches the characteristics and challenges of a particular type of neighborhood?

In short, this research project intends to apply computational geometry and network analysis to the urban planning with a special emphasis on the TOD. The methods, the typology framework and the results shall help planners in defining the catchment areas around the stations in line with the sustainable development of transport infrastructure and service. They shall also contribute to the development of the transit system in combination with spatial development strategies of the TOD neighborhoods.

1.6 Structure of the thesis

Using Hamburg as the study case, network analysis will be applied in this thesis to analyze two types of networks: transit network and street network. The thesis is organized according to the framework shown in

Figure 1.2 and the workflow shown in Figure 1.3. The analysis will consist of three parts.

Part I investigates the transit network. This part is further divided into two chapters. Chapter 2 employs network analysis to scrutinize the topological properties, including geometric properties, small-world properties and scale-free properties, of the public transit network. Chapter 3 uses clustering analysis to group the railway public transport stations based on their connectivity, their importance in the network and the service frequency. The results of this classification will be used as the Node Value in the Node-Place Model.

Part II examines the street network. Before carrying out the empirical street network analysis in the neighborhood of the individual stations, there are a few challenges that need to be dealt with. Chapter 4 deals with the first challenge, namely deciding on the size of the study area. Assessment of an urban network requires the determination of the center point and the catchment area around it, and the size of the catchment area exerts significant influence on the values of the network analysis indicators. This influence is referred to as the size effect, and chapter 4 aims to provide a guideline for determining the appropriate size for the street network analysis.

The second challenge is the placement effect, which means that the measurements of network analysis, such as Closeness Centrality, depend on the placement of the catchment area. In chapter 5, we will examine the placement effect and propose a method for choosing the location of the catchment area if the Closeness Centrality of different nodes has to be compared.

In chapter 6 we apply the findings of chapters 4 and 5 to the empirical study in order to classify TOD neighborhoods by the Place Value, which is measured by the street network analysis and the accessibility, diversity and density of the POI. Cluster analysis will then be applied to further group the neighborhoods around the train stations in order to assign their Place Value in the Node-Place Model. After the Node and Place Values of each station are determined, we will be able to identify the neighborhoods that are already successful from the TOD point of view and, on the other hand, the neighborhoods with the potential for further improvement. In addition, the exemplar neighborhoods of each group will also be identified.

Part III examines the network-based location allocation of the proposed intervention in the selected exemplar neighborhoods. In chapter 7 we choose the bike rental station (BRS) as an example of a possible intervention and demonstrate how to choose an appropriate location for such a station by utilizing the network analysis.

In the final chapter, the limitations of the current research are identified and the potential directions for further research are proposed accordingly. The limitations explained in this chapter can be the starting point for a significant improvement.

Finally, in the context the outbreak of COVID-19 in 2020, we hope that this research might be useful for the planners by helping them to make better informed and data-driven decisions on the circulation design and the location of various facilities like temporary test stations and mobile clinics for providing immunizations during the pandemics.

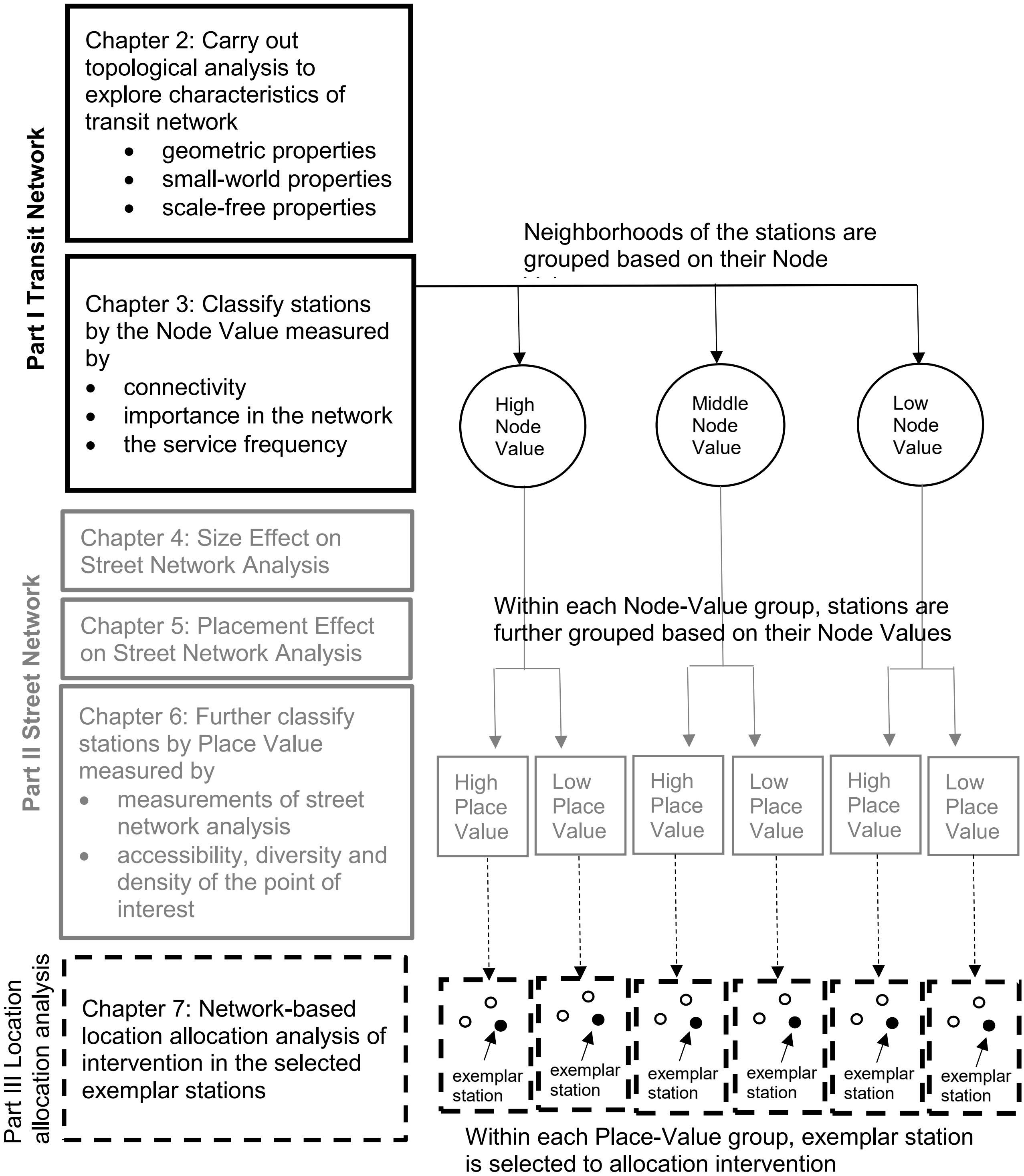


Figure 1.2 Thesis framework.

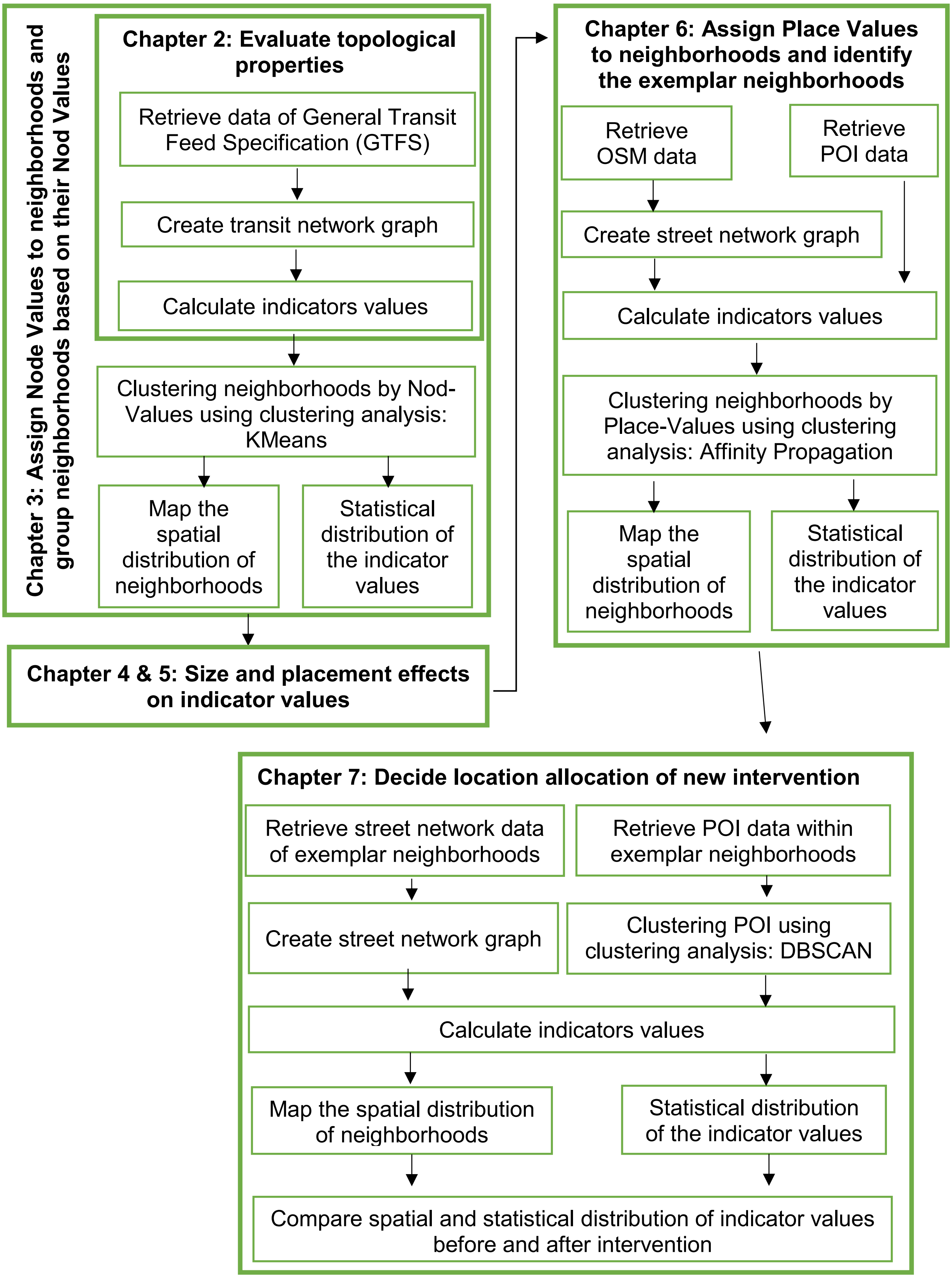


Figure 1.3 Chart of workflow.

2 Topological characteristics of the public transportation network in Hamburg

2.1 Introduction

In many cities, public transportation modes have been integrated in order to increase accessibility and reduce commute times. For the management of public transport networks (PTN), it is important to periodically evaluate the network's connectivity and ability to handle congestion and to adjust its spatial configurations in accordance with such evaluations. As the PTN systems have become vital for increasing the ridership, it would be of great interest to policy makers and planners to examine the topology of PTN in order to have more data-driven and graph-based knowledge of the structure and characteristics of PTN and to assess strategies for improving their performance.

Network analysis is an important aspect of transport geography because it describes the distribution of edges[1] and the arrangement of nodes as well as the relationships between them. Viewing transit systems as complex networks and applying the network-science perspective and the graph theory can have many benefits, including enabling a holistic view of the system, exploring network strengths and weaknesses and carrying out the comparisons between networks in different geographical locations (Derrible, 2012). Recommendations adopted by planners and designers for future plans and the optimization of performance of the transportation system can be developed based on the results of such an analysis (Hong, Tamakloe, Lee, & Park, 2019).

However, the complexities of urban public transportation networks make their examination rather challenging. This problem arises because transportation networks often consist of a large number of stations and links. Due to this complexity, research applying the graph[2]-based measures to systematically analyze and compare the networks remain to be quite limited.

In the last few years, increasing availability of comprehensive data sources and powerful tools of information technology enable to combination of network science and

1. "Edges" is a term in graph theory referring to the connection between a pair of nodes. In the current research of transportation networks, we also use "links" as an identical term to edges.
2. In graph theory, a network is considered as a graph. The basic units for constructing a graph are "nodes", which are represented by a point, and "edges", which are the connection between nodes and are represented by lines.

computational geometry in order to accelerate the automatic retrieval and analysis of extremely large amounts of data (Boeing, 2017). The geodata can then be used to construct network graphs, visualize networks and analyze the attributes of networks by calculating the metric and topological indicators.

The objective of this chapter is to analyze the topological and structural properties of PTN in Hamburg metropolitan area by taking advantage of the recent developments of the network science tools, including the computer programming techniques, implementation of known algorithms and Big Data analysis. Based on the analysis of the topological characteristics, the current study examines whether the PTN in Hamburg possess the required features for advancing the efficient operation and the resilience to failure or disruption.

2.2 Literature review

Public transportation systems are typical examples of complex networks. They are made up of a set of nodes (i.e., stations) and links. The stations are distributed across the network and are connected by the links. The combination of several links between stations forms the routes or paths. Therefore, many ideas in graph theory and network analysis can be employed to explore the implications of the PTN structure, patterns and properties on public transportation (Shi, Wen, Zhao, & Wu, 2019) (Soh, et al., 2010) (Zhang, Li, Deng, & Wang, 2014) (Mishra, Welch, & Jha, 2012).

One of the important elements in understanding transport networks is the topological properties, which explain the physical arrangement, connection and relationship between the elements of a network (Zahedi, Mawengkang, Masri, Ramon, & Putri, 2019). The advantages of the application of topological properties include "simplifying complex scenarios and modeling numerous complex interactions within systems in order to allow the engineer to make modifications to the existing network for enhancement of services" (Hong, Tamakloe, Lee, & Park, 2019). In the current study, three different topological properties, i.e. geometric, small-world and scale-free properties, are explored and the indicators for studying these properties are introduced.

The first type of topological properties is the geometric properties, which can be measured by the attributes like the size and density of the network. Measurements for scrutinizing geometric properties include **number of nodes (*N*), number of links (*M*), link-node ratio, and degree of connectivity(γ)**.

The other types of topological properties, namely small-world and scale-free properties, have been identified in several studies as the two unique features required for efficient operation and network resilience (Zhang L. , Lu, Fu, & Li, 2019). First of all, a network has small-world properties if the following two conditions are satisfied:

1) the **average path length (*V*)** is close to that of a random network with the same number of nodes;

2) the **average Clustering Coefficient** is significantly higher than that of a random network with same number of nodes.

Secondly, if the **degree distribution (*p(k)*)** of a network follows the power law distribution, it is considered to have the scale-free property. Because this means that, compared to the exponential distribution, the degree distribution of a network reduces more gradually. This also indicate the possible presence of cluster that have many links and, therefore, have large degree (Rommel, 2014) (Humphries & Gurney, 2008).

Transport networks with these two properties have been recommended for urban transit network because they possess desirable features, including robustness and resilience to failure or disruption (Chopra, Dillon, Bilec, & Khanna, 2016) (Wu, Gao, Sun, & Huang, 2004) (Huang, Grigolon, Madureira, & Brussel, 2018).

2.3 Data and computational tools

In this chapter we will use the above-mentioned indicators for the systematic analysis of transit networks using data from the General Transit Feed Specification (GTFS). This section introduces the background GTFS and describes the structure of the files in the dataset.

GTFS is an open platform and a data model initiated by Google and the Portland, Oregon, public transit agency (TriMet), in order to allow transit organizations to publish their schedules for routing and visualization purposes at a low cost (McHugh, 2013).

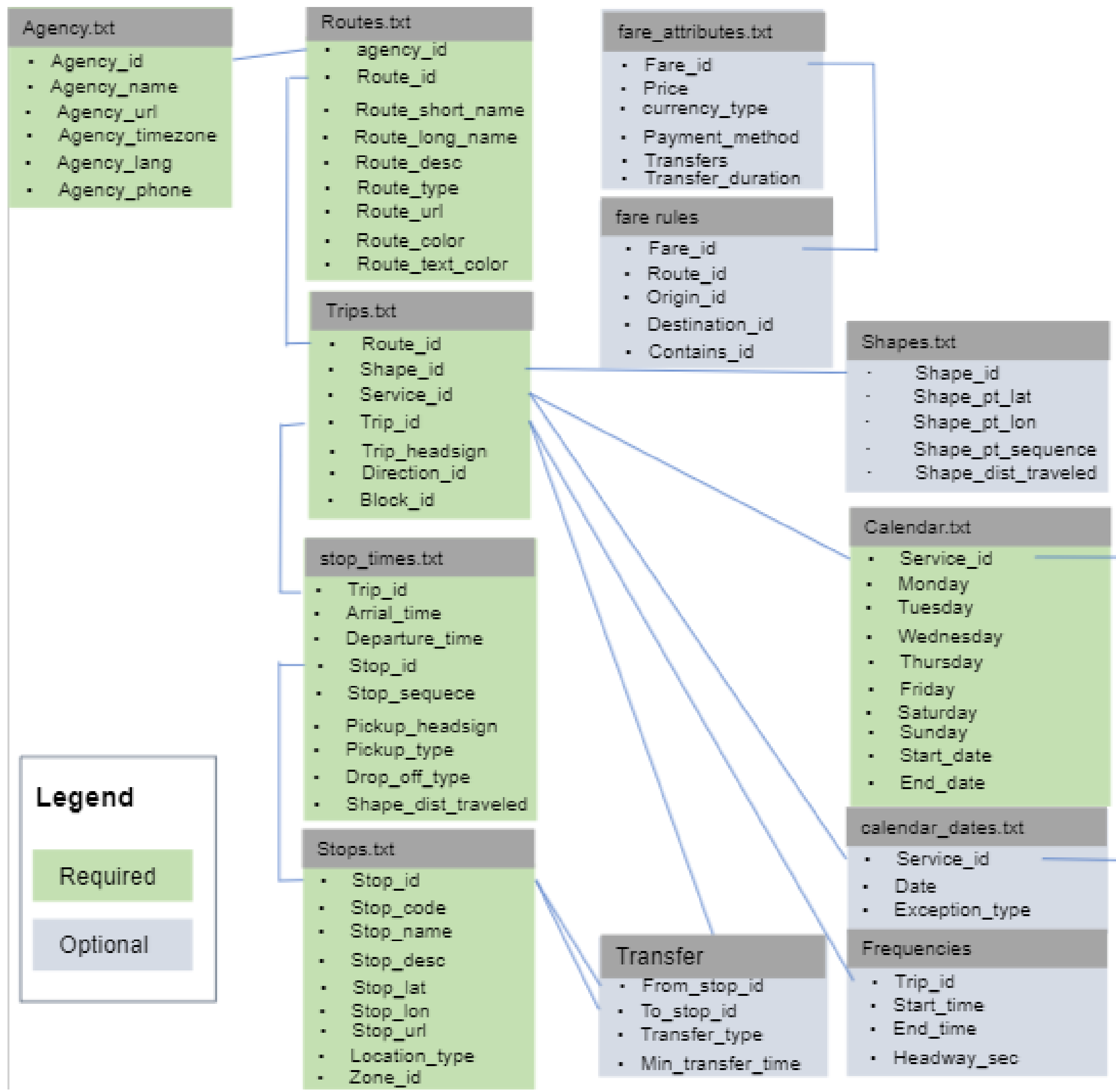


Figure 2.1 GTFS files and relationships.

Google defined a publishing standard and a common format for transit agency operational data, which include public transportation schedules and associated geographical information (e.g., stops, stop times, routes). The GTFS Feed provided by a public transit agency is a zip file containing six mandatory files with comma-separated values (CSV) and seven optional files. Each CSV file contains information about a part of the transport operator's system.

Figure 2.1 shows these files and their relationship. Combining these files together, we can describe the stops, routes, and schedules of an entire transit system and the outcome consists of many tables of a relational database (McHugh, 2013).

In this chapter, the Python library of peartree 0.6.4 and partridge 1.1.1 are employed to 1) check the Transit.Land API and query it for any and all operators that serves Hamburg; 2) retrieve the zip location of the original GTFS; 3) download the zip file to a local temporary directory; 4) convert the GTFS data into a directed multigraph network.

Through the entire thesis, OSMnx version 0.15.1 (Boeing, 2017) (Boeing, 2019), which is built on top of Python3 libraries of NetworkX 2.4, matplotlib 3.3.0, and GeoPandas 0.8.1, serves as an automatic, scalable, reproducible tool that performs the network graph analytics operations on GTFS data and calculates the indicator values. Pandas 1.0.5 and GeoPandas 0.8.1 are used in this research to carry out statistical analysis of the calculated indicator values. The figures in this research are generated using Matplotlib 3.3.0.

2.4 Indicator calculate and results: topological properties of the network

2.4.1 Case study area

The city of Hamburg in Germany is selected as the study area in the current research in order to exemplify the process of the proposed methodology framework. With the population of more than 1.84 million inhabitants and the size of 755.22 km^2, Hamburg is the second largest city in Germany.

The Hamburg metropolitan region, which includes the city of Hamburg and its neighboring district, has a population of more than 3 million people[3]. The Hamburg metropolitan region corresponds roughly to the service area of the Hamburger Verkehrsverbund (HVV), which can be translated as Hamburg Transport Association. The complete and integrated HVV transit network, including both railway public transport (U- and S-Bahn[4]) and bus networks in Hamburg metropolitan region, is shown in Figure 2.2, which is created by OSMnx using the retrieved GTFS data.

[3] http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=urb_lpop1&lang=en, Retrieved 15 June 2021

[4] U-Bahn, which literally translates as “underground railway” is the urban rapid transit or mass rapid transit (MRT) in German cities. S-Bahn is the German urban-suburban rail serving a metropolitan region and connecting the nearby districts and towns to the city.

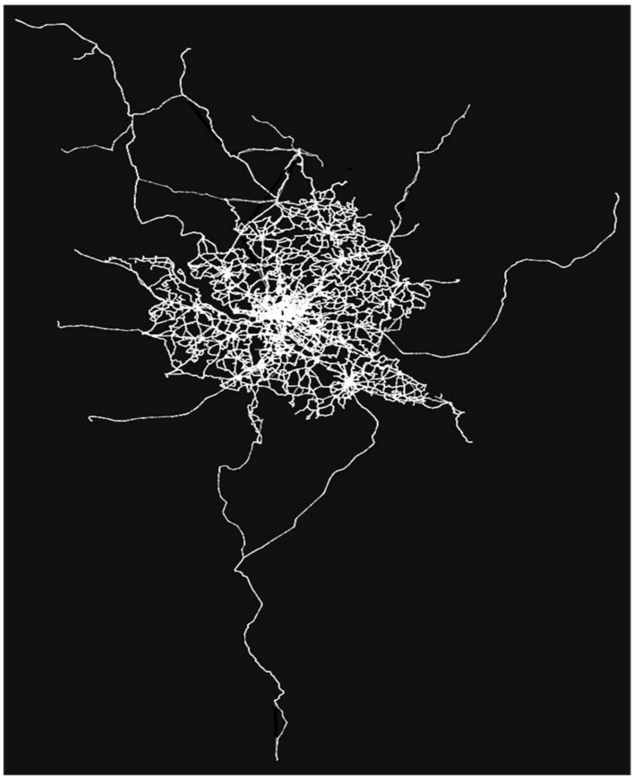


Figure 2.2 HVV network.

Table 2.1 Computed properties of Hamburg integrated transit network.

	Property	Hamburg	Seoul
Topological characteristics	Number of nodes, *N*	14416	12873
	Number of links, *M*	17914	19354
	Link-node ratio	1.24	1.5
	Degree of Connectivity, γ	0.000172	0.000233
Small-world properties	Average path length (in network distance), *V*	11.396	28.169
	Average Clustering Coefficient, *C_avg*	0.038	0.028
Scale-free network properties	Average degree, *k_avg*	2.485	3.007

Table 2.1 provides the results of the computed network statistics. In order to evaluate the HVV network, we also include the previous research of the public transportation network in Seoul (Hong, Tamakloe, Lee, & Park, 2019) as a reference for comparison. In the following sections, we explain the relationship between these indicators and the level of network connectivity.

2.4.2 Geometric properties

1) Size

In the current research, the size of the networks is measured by the **number of nodes (*N*)**, i.e. stations, and the number of **links (*M*)**. Naturally, there exists a positive correlation between the size of the network and the level of accessibility. If the covered area remains constant, a higher number of stations and links (which also means a higher number of connections and routes) provides higher accessibility to destinations.

A higher number of nodes does not necessarily mean that the network also has a higher number of links than another comparable network. For example, as one can see in Table 2.1, the total number of stations in the integrated system (including the U-Bahn, the S-Bahn and the bus networks) is 14,416, and the number of links is 17,914. In comparison with the PTN in Seoul, the HVV network has more stations but less links. As this example shows, it is not possible to evaluate the connectivity by using only the number of nodes or only the number of links. Rather, such evaluation requires the use of a complex indicator, namely the **link-node ratio**.

2) Link-node ratio

The link-node ratio is an index of connectivity which divides the total number of links by the total number of nodes in a catchment area (Cerin, et al., 2013) (Molaei, Tang, & Hardie, 2021). This ratio formally expressed by the following formula:

$$link\text{-}node\ ratio = N\ /\ M \tag{1}$$

where *N* is the total number of nodes and *M* is the total number of links in the network. Naturally, higher values indicate better connectivity. Ewing (1996) suggests setting the threshold for high street connectivity at the link-node ratio of 1.4. In the HVV network case, the link-node ratio is 1.24, which is lower than the threshold score and therefore falls short of reaching the high connectivity level. In comparison with the Seoul, HVV is also lower than the score of the PTN in Seoul, which is 1.5. Therefore, in terms of the link-node ratio, the connectivity of the Seoul PTN is better than that of the HVV network.

3) Degree of Connectivity, γ

The Degree of Connectivity (γ) describes the relationship and interaction between the elements, i.e. nodes and links, in a network system. In this research we use the gamma index to measure the Degree of Connectivity, which is defined by dividing the observed number of links by the possible maximum number of links in the network (Rommel, 2014). It can be calculated as follows:

$$\gamma = \frac{M}{M_{max}} = \frac{M}{N*(N-1)/2} \quad (2)$$

where N is the total number of nodes in the network, M is the total observed number of links in the network and M_{max} is the possible maximum number of links in the network. The value of gamma lies between 0 and 1, where the value of 1 indicates a completely connected network. This means that each node is connected with one link to all other nodes (Czerkauer-Yamu, 2012). In practice, such a perfectly connected network would of course be extremely unlikely.

In the HVV network case, the gamma index (i.e., the Degree of Connectivity) is 0.000172, which is smaller than the Seoul network's value of 0.000234. Based on the results of the underlying basic network properties displayed in Table 2.1, we infer that the HVV network does not provide higher connectivity compared to the theoretical reference and the transportation network in Seoul.

2.4.3 Small-world properties

A network has the small-world properties, proposed by Watts and Strogatz (1998), if it possesses the following features: 1) most nodes are not neighbors of one another; 2) most nodes can be reached by a small number of nodes by small number of steps (Mohmand & Wang, 2014). From the perspective of transportation research, investigating the small-world properties of a transportation network can be very useful for the researchers evaluating a transportation network. First of all, due to its short average path length, a network with the small-world properties means it has high capability and connectivity to link the nodes (stations) more effectively. Therefore, analyzing the small-world properties informs us about **how efficient in communicating the network is**. Secondly, a small-world network is structurally more resilient to failure or disruptions (Chopra, Dillon, Bilec, & Khanna, 2016). Therefore, analyzing the small-world properties can be very useful when evaluating **how robust a network is**, i.e. how good it will perform if one of its nodes (stations) stops functioning for some reason.

Typical features of a spatial structure with the small-world properties are, compared to a random network with the same number of nodes, its small average shortest path length and large Clustering Coefficient. Previous research (Humphries & Gurney, 2008) (Mohmand & Wang, 2014) (Telesford, Joyce, Hayasaka, Burdette, & Laurienti, 2011) (Chopra, Dillon, Bilec, & Khanna, 2016) has proposed to compare the

average path length *(V)* and **average Clustering Coefficient *(C_avg)*** of the real network to an Erdos–Renyi (E–R) random network constructed with the same number of nodes as the real network.

The steps are as the following.

i) Calculate the average path length, V_{real}, and the Clustering Coefficient, C_{real}, of the real network.

ii) Calculate the average path length, V_{random}, and the Clustering Coefficient, C_{random}, of a random network with the same number of nodes.

iii) Calculate the normalized shortest path $\lambda = V_{real}/V_{random}$ and $\gamma = C_{real}/C_{random}$.

iv) If λ and γ fulfil the criteria, $\lambda \approx 1$ and $\gamma > 1$, the network can be identified as a small-world network.

In the following, this method is applied to see if the small world properties are present in the HVV network.

1) Average Path Length (*V*)

In network analysis, the shortest path length is the minimum number of links along the shortest path in order to travel from the origin node to the destination node. The average path length[5], *V*, indicates the average number of links passed through the shortest paths out of all possible pairs of nodes. It is an indicator measuring the progression of a network in time, i.e. how easy it is to move from one node to another node. It can be formally expressed as

$$V = \frac{1}{N(N-1)} \sum_{u \neq v} d_{uv} \qquad (3)$$

where *N* is to the total number of nodes in the network, *V* is the average path length and d_{uv} corresponds to the length of the shortest path between stations *u* and *v*. A small average path length indicates that there is good connectivity and efficient communication among the stations in the network, regardless of geographical distance.

In the case of the HVV network, the average path length is 11.396, which is smaller than the value of 28.169 in the case of Seoul. This means that, on average, a passenger has to transverse 28.169 links in Seoul but only 11.396 links in Hamburg in order to travel to his or her destination.

[5] The length here refers to number of links (i.e., the network distance), rather than the number of meters (i.e., the metric distance). In other words, the measurement unit for the path length is the number of links.

2) Clustering Coefficient (*C*) and average Clustering Coefficient (*C_avg*)

The Clustering Coefficient, *C*, determines the connectivity among neighbors of a node-*u*. Assuming that the node-*u* has k_u neighbors, the Clustering Coefficient of node-*u*, C_u, can be calculated by the following equation

$$C_u = \frac{2E_u}{k_u(k_u - 1)}$$

(4)

where E_u is the number of links between node-*u*'s neighbors and $k_u\ (k_u - 1)/2$ is the normalization factor, which is the maximum number of links that can possibly exist among the neighbors of the node-*u*. Further averaging C_u by dividing it by the total number of nodes, *N*, the aggregated level of clustering within the network can be defined by

$$C\_avg = \frac{\sum_u^N C_u}{N}$$

(5)

As shown in Table 2.1, in the case of the HVV network, the average Clustering Coefficient is 0.038, which is larger than the value of 0.028 in the case of Seoul. In terms of average Clustering Coefficient, there are more "hubs" and "clusters of stations" in HVV network than in the case of Seoul.

As shown in Table 2.1, the average path length, V, of the HVV network constitutes 11.396 and its average Clustering Coefficient, C_avg, is 0.038 for HVV network.

According to *step iv* in the previous section, the values of average path length and average Clustering Coefficient need to be compared to the corresponding values of an E–R random network with the same number of nodes. The results of such a comparison show that:

- $\lambda = V_{real} / V_{random} = 6.3$, which does not meet the criteria of $\lambda \approx 1$
- $\gamma = C_{real} / C_{random} = 5.28$, which does not meet the criteria of $\gamma > 1$

Based on this comparison, we conclude that the small-world properties cannot be identified in the HVV network. This indicates that the connectivity and robustness of the HVV network are not very resilient to failure or disruptions. If one of the stations stops functioning for some reason, it is possible that there will be a significant adverse effect on the average speed and ease of achieving the destination. In other words, the degree of fault-tolerance of the HVV network is rather low.

2.4.4 Scale-free properties

1) Node Degree (*k*) and average Node Degree (*k_avg)*

Node Degree, *k*, is the number of links connected to the node. It shows how accessible a node is and is used to characterize the local features of the network (Wang & Fu, 2017). It is both the **connectivity measure,** which is why it is a basic parameter for examining the scale-free properties of the network, and the most straightforward **centrality measure**, which indicates the importance of a node in the network.

In the transportation research, Node Degree is an indicator that express the accessibility level of the stations in the transit network. A station with higher Node Degree means that its connectivity is higher, and it is, therefore, more accessible and important in the network. Because this indicates that the station has a higher chance to increase its capacity for receiving many passenger flows. However, the stations with higher Node Degree are also the most vulnerable to attacks or other disruptions because the failure of these most connected stations can cause a failure of a large portion of the network.

Another indicator, the **average Node Degree, *k_avg***, of a network is a measure that shows the overall level of connectivity of the network because it indicates how connected each node in the network is. The network-wide average Node Degree, *k_avg*, can be expressed as

$$k\_avg = 2 * M / N \tag{6}$$

where *M* refers the total number of links and *N* refers to the total number of nodes.

As shown in Table 2.1, in the case of Hamburg, the HVV network has N = 14416 nodes and M = 17914 links among stations. The average Node Degree of the HVV network, *k_avg*, is, therefore, 2 *M / N = 2.485, which indicates that, on average, a station is directly connected to 2.485 other stations. This value is smaller than 3.007 in the PTN in Seoul.

2) Node Degree Distribution, *p(k)*, and Cumulative distribution of Node Degree

The **Node Degree Distribution, *p(k),*** of a network helps us to understand the structure and the topology of a network and to identify whether or not the network is scale-free. A network has the scale-free properties when it is insensitive to the change of scale. In other words, irrespective of whether or not the network size increases, its underlying structure remains unchanged (Wang & Chen, 2003).

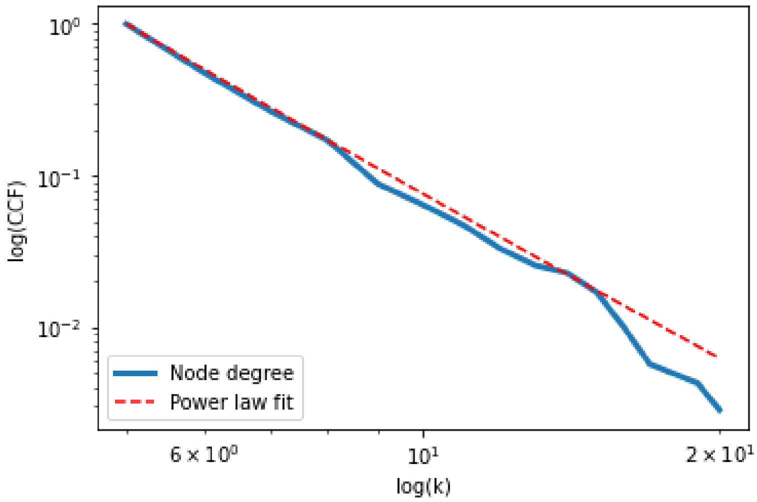


Figure 2.3 Log–log cumulative distribution of station nodes of HVV network.

If a network possesses the scale-free properties, its cumulative Node Degree Distribution should follow the power law distribution. There are two steps in evaluating how well our network data fit the power law distribution. Firstly, by plotting the cumulative Node Degree distribution of a network in a logarithmic scale, like the blue line in Figure 2.3, we can examine how it fits with the line of power law, indicated by the red dash line. If the distribution follows the power law, the blue line would be characterized by a more gradual fall. This means that there is a large number of nodes with only a few links to other nodes, while there are relatively few nodes that are connected to many other nodes. As shown in Figure 2.4 Node Degree Distribution, p(k), of stations in HVV network. Figure 2.4, which presents the Node Degree Distribution of the HVV network, it is shown that the stations in the HVV network have a minimum degree of 1 and a maximum degree of 20. The majority of the stations have 2.5 links connected to them. Again, the nodes with a larger number of links have a stronger influence on the entire network's structure and dynamics. Because the study of scale-free networks helps us to determine whether a network has a large number of important nodes, it is often used for determining how resilient or robust a network is (Wu, Gao, Sun, & Huang, 2004).

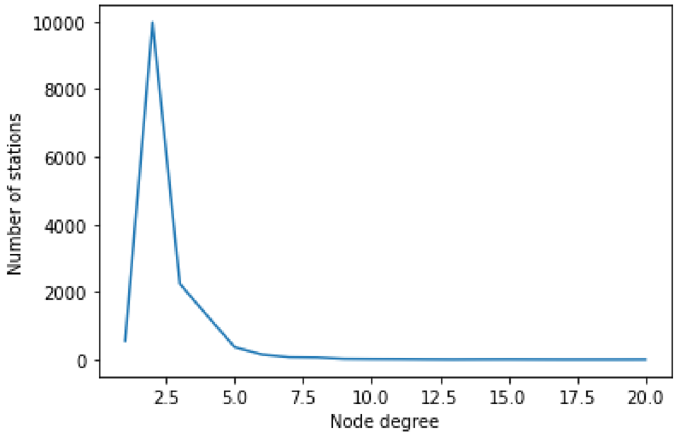


Figure 2.4 Node Degree Distribution, p(k), of stations in HVV network.

Secondly, in order to evaluate whether the power-law distribution itself is a statistically plausible model, we employ the Kolmogorov–Smirnoff tests and compute the *p*-value for the fitted power-law model to test how well the data (i.e. the blue line in Figure 2.3) fit the power law distribution. Since the p-value of our data is 0.365 and the *p-value* > 0.1, we have relatively strong support for a conclusion that the Node Degree Distribution follows the power-law distribution. And the probability of finding a station with *k* connections is proportional to $k^{-1.28}$, which means that there are stations with very high degree in the network. The results indicate that the HVV network is vulnerable to significant failures or disruptions if one of the stations with a large number of links stops operating for some reasons (e.g., due to construction or maintenance works).

2.5 Conclusion and discussion

In this chapter we have examined the topological properties of Hamburg metropolitan area's public transportation network. First of all, the geometric properties have been measured by the number of nodes, number of links, link-node ratio and the Degree of Connectivity. Secondly, the small-world properties have been measured by the average path length and Clustering Coefficient. Thirdly, the scale-free properties have been measured by average and cumulative distribution of the Node Degree. Based on the analysis of the topological properties, the current study examines the level of robustness and resilience to failure or disruption and whether any unique features required for advancing the efficient operation exist in the network.

Wherever possible, results of theoretical models or previous research have been provided, serving as reference for comparison when investigating the characteristics and determining the levels of connectivity, performance and resilience of the network.

Our examination of the spatial configuration of the HVV network has found that the network has low fault tolerance. First of all, in terms of geometric properties, the HVV network does not facilitate easier movement between stations than the theoretical reference network or the Soul transportation network. Furthermore, the small-world properties cannot be identified in the HVV network. This indicates that the connectivity and robustness of the HVV network have low level of resilience in case of failures or disruptions. In other words, if one of the stations stops functioning for some reason, the speed and ease of getting to a destination can be severely affected. The examination of scale-free properties also reveals that the HVV network is vulnerable to significant failures or disruptions if one of the stations with a large number of links stops operating due to, for example, construction or maintenance work.

In this chapter, we have shown that analysis of the structural composition of the PTN can help us to evaluate the level of fault tolerance based on the topological properties and spatial configuration of the network. The existence or absence of the agglomerated community structure can be used to determine the robustness of the network and to identify structural defects that require to be improved. These results are imperative for the future enhancement of mobility, such as strategically constructing new stations, relocating the existing stations, creating more links.

In future research, the analysis in this chapter can be used as a basis for comparing multiple PTNs in different cities. Further investigation can also consider the effects of other factors, including the population density or the size, structural pattern and shape of the catchment area. One can also compare the small-world properties of the entire network with that of the sub-network in the inner city, where the network density is high. By identifying the clusters of sub-networks within the entire network, it might also be useful to explore various characteristics of these sub-networks and examine how these characteristics affect the connectivity or resilience of the entire network.

Another interesting investigation would be testing the scalability of the network shape and structure. This means examining whether the existing network can scale both in size and connectivity. “A network’s shape dictates its scalability and connectivity. A good shape and structure allow a network to scale with growth and still maintain the focus of its vision and purpose through connectivity. If the topology of the network discourages scalability and connectivity or has an unhealthy obsession of one over the other, then it may not be shaped for movement” (Yang, 2018). For example, Hamburg has recently planned the new route, U5 (https://www.hamburg.de/u5/). Further investigation can test to what extend an additional route improves the connectivity of the entire network. If adding one additional route significantly improves the connectivity of the network, the shape and structure of the HVV network can be considered to have good scalability.

3 Classifying stations by Node Value

3.1 Introduction

In chapter 2, we have examined the integrated HVV network, which includes both the railway public transport (RPT) and the bus stations. This chapter focuses on the network consisting only of the RPT stations. In addition, in the previous chapter, we examined the **topological properties of the entire network**. In the current chapter, the focus is on comparing the **importance of individual nodes**, which are stations in the transportation network. The stations will be classified by their **centrality indices** and **service frequency**. The results of this classification will be used as the Node Value of each station in the Node-Place Model, which is a typology framework that we use for the assessment of Transit Oriented Development (TOD).

Stations in the transit system are the access points with which the users interact on a very regular basis. Therefore, examining the stations in order to determine their level of importance in a transportation network can have significant practical benefits. First of all, it helps to identify which stations in the system are the “key stations” that have greater impact on the efficiency and productivity of the network (Fortin, Morency, & Trépanie, 2016). Secondly, for the management of public transport networks (PTN), identifying the key stations in a transit network facilitates the design of a system with more accurate and effective traffic control, which helps to distribute the flow of passengers more evenly. Thirdly, in times of disasters when the traffic is disrupted, it also helps to improve network connectivity and robustness by adjusting the spatial configurations of the network. Finally, a method for determining the importance of stations can be developed as “a tool to forecast ridership increase linked with the opening of new lines” (Derrible, 2012).

However, development of effective strategies for determining the importance of stations in a public transportation network is often constrained by the lack of data and of the cost-effective modeling (Jayasinghe, Kasemsri, Abenayake, & Mahanama, 2019). This study is an attempt to develop a practicable framework and methodology for identifying the key stations. At the core of this method lie the spatial network analysis and the statistical analysis, which are performed on a selected set of components based on two main concepts: connectivity and service frequency.

3.2 Literature review

Previous research on the evaluation of TOD has focused on individual transit stations. For example, Singh et al. (2014) has aggregated multiple spatial indicators of the area around the station of the city region of Arnhem-Nijmegen in the Netherlands in order to measure the TOD levels of the stations.

However, such spatial indicators are not sufficient for understanding the role of each station within a transportation network. A station may not have high performance when evaluating its individual TOD level, but it may serve as a significant transit node within the transportation network (Huang, Grigolon, Madureira, & Brussel, 2018). In other words, a station should be evaluated not only by its spatial indicators but also by its relationship with other stations in the transit network. Therefore, the emphasis of the current research will focus on the network centrality measurement, which is one of the most important features of transportation network systems.

There is already some research attempting to develop applications based on centrality parameters in the spheres of transport planning. For example, centrality measures have been employed in the literature on public transit demand. They are used as a tool for explaining fundamental qualities related to transit ridership, transfer at stops and through-traffic in networks. Janasinghe and Munshi (2014) apply the centrality measure in order to explore whether there is some regular relationship between transit demand and network centrality in the context of Indian cities, and if so, what is the extent of this relationship. Porta et al. (2006) studied the movement flow in the network by using several centrality indices, instead of the single best-fit centrality measure. Kazerani and Winter (2010) studied the dynamics and temporal aspects of people's travel demand through a modified version of Betweenness Centrality. Finally, by introducing a GIS-based tool designed to evaluate the application of centrality and connectivity in urban public transportation networks, Scheurer et al. (2008) used Degree Centrality, Closeness Centrality and Betweenness Centrality to scrutinize and visualize the strengths and weaknesses regarding the geographical coverage, network connectivity, competitive speed and service levels of public transport networks.

In the current study, the centrality of stations is applied to identify the importance and the value of stations in the transportation network. Nodes with higher centrality would be described as more "influential" nodes in the network.

3.3 Indicators for identifying key stations

3.3.1 Centrality

In order to identify the key nodes in the transit network, Porta et al. (2006) (2008) developed the Multiple Centrality Assessment (MCA). The MCA model considers both the topology (how links or edges are connected to each other) and the metric of the system (distance is in fact computed metrically, rather than just topologically). The importance of a node is measured by multiple criteria, including being close to all the other nodes, being the intermediary between the selected pair of nodes, and being critical for effective traffic control of the system as a whole. In this study, we identify major stations in the Hamburger Verkehrsverbund (HVV) railway public transport (RPT, including U- and S-Bahn) networks based on the commonly employed centrality indices: Degree Centrality, Betweenness Centrality and Closeness Centrality.

1) Degree Centrality, C_k

The Node Degree, k, measures the number of links connected to the node. On the one hand, it is a **connectivity measure** that is employed to characterize the connectivity and the local features of the network, as we have seen in the previous chapter. The Node Degree Distribution, $p(k)$, is the parameter for identifying the scale-free properties in a network. On the other hand, Node Degree is also a **centrality measure**, which reflects how accessible and important a node (station) is in the network. A higher value of Degree Centrality, C_k, indicates higher accessibility of the station. It also means that the station is a hub of the transit system. Therefore, in the current chapter, the Degree Centrality, C_k, is employed as a measure for identifying key stations.

For any node (station) in the PTN, the Degree Centrality of node-v indicates the number of nodes that are directly connected to node-v. In a directed network with N nodes, the Degree Centrality of node-v, $C_k(v)$ can be calculated by the following formula:

$$C_k(v) = \frac{k_v}{N-1} \tag{1}$$

where k_v is the number of nodes directly adjacent to the node-*v*, *N* is the number of nodes in the network, and *N-1* is the maximum possible degree value of a node in the network.

2) Closeness centrality, C_c

Another parameter for identifying the key stations is the Closeness Centrality, C_c, which measures the average shortest distance from a given starting node-*v* to all the other nodes. In the public transport network, Closeness Centrality reflects how close it is from one station to all the other stations in the network. The higher the values of C_c, the easier and faster that a station can be reached in terms of speed and frequency (Hong, Tamakloe, Lee, & Park, 2019) (Mohmand, Mehmood, Amjad, & Makarevic, 2015) because it means that it takes less steps to reach other stations. This also means that this is a station with greater influence and has wider range of service.

Closeness Centrality is defined as the inverse of the sum of the shortest distances between the chosen node-*v* and all other nodes. The C_c of a node-v can be formally expressed as

$$C_c(v) = 1 \,/ \sum_{u=1}^{N} S(v,u) \qquad (2)$$

where *N* refers to the total number of nodes in the chosen catchment area and *S(v, u*) refers to the length of the shortest distance between the chosen node-v and another node-*u*.

3) Betweenness Centrality, C_b

Betweenness Centrality, C_b, is “based on the idea that a node is more central when it is traversed by a larger number of the shortest[6] paths connecting all couples of nodes in the network” (Porta, et al., 2012). For the purpose of identifying key stations in a transportation network, Betweenness Centrality is unique among all other indicators in the sense that it highlights the frequency of a station that lies on the shortest paths between any pairs of stations. A station with very high Betweenness Centrality means that, in order to reach other stations, many flows need to pass through this station. Therefore, Betweenness Centrality reflects the importance of a station as a transfer node or a connector among many stations and regions within the network. This

[6] Here the shortest refers to the distance and not the time required for travelling.

transfer characteristic is especially relevant in transit systems. The Betweenness Centrality of a node-*v* can be calculated by

$$C_b(v) = \frac{1}{(N-1)(N-2)} \sum_{\substack{i,j \in N \\ v \neq j, i \neq j}} \frac{m_{ij}(v)}{m_{ij}} \tag{3}$$

where N is the total number of nodes, m_{ij} is the number of shortest paths connecting node-*i* and node-*j*, and $m_{ij}(v)$ is the number of shortest paths connecting node-*i* and node-*j* and going through node-*v*.

4) Comparison between centrality measures

The three centrality measures described above can be used to identify the level of importance of stations in the transit network, but each of these measures has a different focus. Degree Centrality is measured by the number of links connected to the node, which reflects the direct influence of a node on other nodes in the network. Closeness Centrality measures the average shortest distance from one node to another node, which can be used to determine how accessible and how close a node is to all the other nodes. Betweenness Centrality reflects the extent to which a node comes in between other nodes, reflecting the load capacity of this node (Wang & Fu, 2017). Both Closeness and Betweenness Centralities are employed to identify the stations with high traffic and congestion.

3.3.2 Service frequency, F_{vu}

Service frequency reflects the temporal aspect of accessibility. For example, a station with one train every 5 minutes offers a higher level of accessibility than a station with one train per hour and, therefore, is more important in the network. Gomez-Ibanez (1996) find that service frequency is strongly associated with more frequent use of public transportation systems. In the current study, we use the data of the General Transit Feed Specification (GTFS) to calculate service frequency, F_{vu}, defined as service frequency between stations-*v* and -*u* in departures per hour per direction from 6 a.m. to 9 a.m. on weekdays.

3.4 Methods

The Python library *gtfs_function 1.0.1* takes the path of the GTFS zip file as argument and retrieves 5 dataframes / geodataframes of *routes, stops, stop_times, trips and shapes*. It also creates a geodataframe of the stop frequency where each row in the geodataframe includes a Point geometry. OSMnx version 0.15.1 (Boeing, 2017) (Boeing, 2019) is used to create a network graph, perform the network graph analytics operations on GTFS data and calculate and normalize the indicator values.

Descriptive statistics for the measures defined in the previous section are computed and cluster analysis is used to group the railway public transport (RPT) stations in order to determine their Node Values.

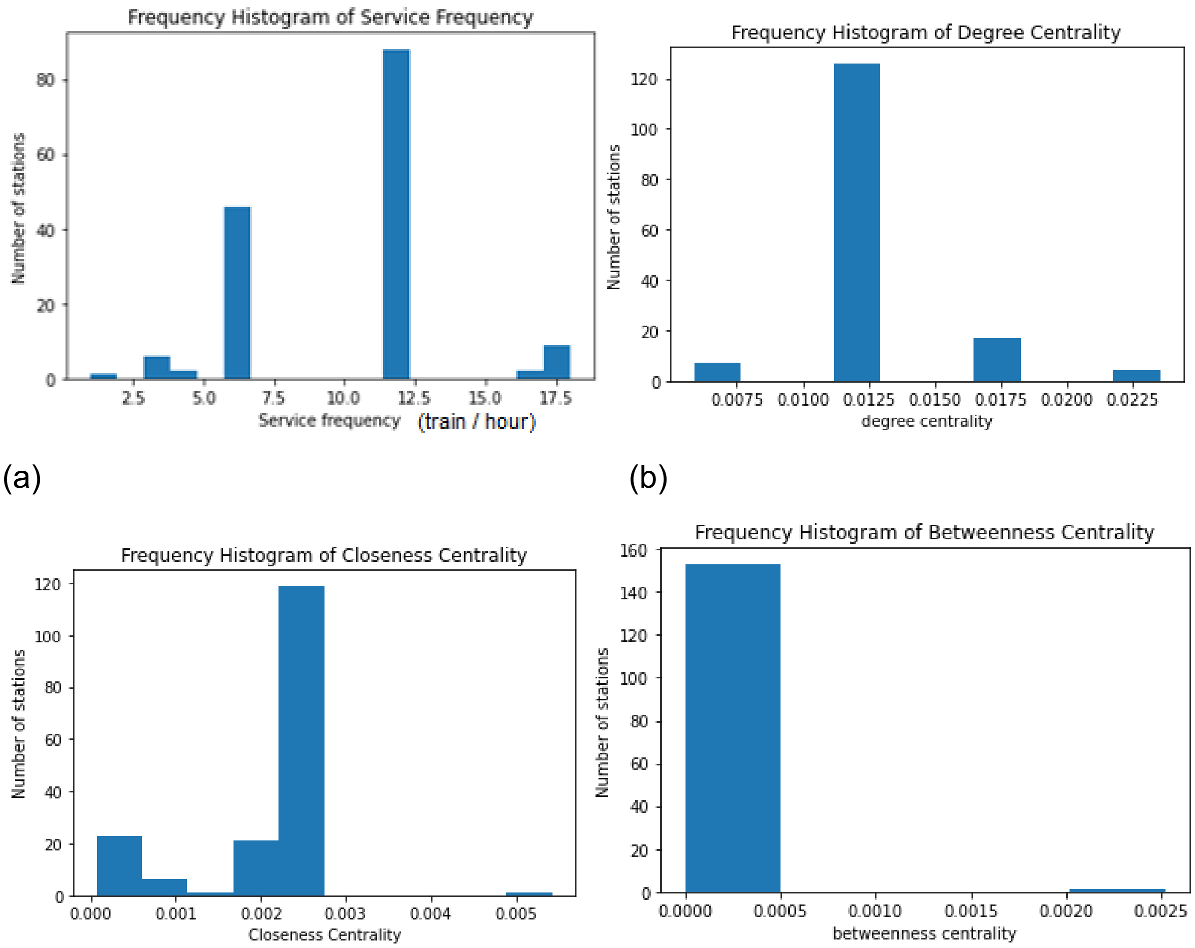
First of all, Principal Component Analysis (PCA), which is an unsupervised, non-linear technique, is employed as a tool for data exploration and for visualizing high-dimensional data in a low-dimensional (2- or 3-dimensional) space. PCA captures the largest variation of the data with a few principal components. It enables the readers to get a clearer idea of how the data is arranged in a high-dimensional space.

Next, the k-means clustering algorithm is used because it is easily scalable based on the number of samples and can efficiently handle the large and high-dimensional datasets. It is a partition-based clustering algorithm and one of the simplest and most popular unsupervised machine learning algorithms. The algorithm partitions all the points into several clusters of equal variances minimizing the within-cluster Euclidean distance. As an input the algorithm requires the samples and the number of clusters. It is sensitive to the number of clusters specified. K-means clustering meets the purpose of this chapter, which aims at classifying the stations into a few groups in order to assign their Node Values. We do not want to have too many clusters at this stage because a further detailed classification based on the Place Value will be carried out in chapter 5.

3.5 Results

3.5.1 Statistical distribution of indicators

Before grouping stations into different clusters, Figure 3.1 presents the statistical distribution of the four indicators of all stations. The results in Figure 3.1 (a) show that, first of all, most stations have the service frequency of 12 trains per hour from 6 to 9 a.m., which means there is one train every 5 minutes. The second largest group have the service frequency of 6 trains per hour, which means there is a train every 10 minutes. There are only few stations with less than 6 trains per hour from 6 to 9 a.m.



(a) (b)

(c) (d)

Figure 3.1 Frequency distribution (a) service frequency; (b) Degree Centrality; (c) Betweenness Centrality; (d) Closeness Centrality of all stations in the HVV railway public transport (RPT) network.

Pearson correlation	Betweenness Centrality	Closeness Centrality	Degree Centrality	Service frequency
Betweenness Centrality	1	0.4928***	-0.0315	0.187982**
Closeness Centrality	0.4928***	1	-0.0136	0.0944
Degree Centrality	-0.0315	-0.0136	1	-0.1997**
Service frequency	0.187982**	0.0944	-0.1997**	1

*** Correlation is significant at 0.01 level
** Correlation is significant at 0.05 level
* Correlation is significant at 0.1 level

Table 3.1 Correlation between the indicators.

Figure 3.1 (b) and 3.1 (c) show that the distribution of centrality measurement values is extremely uneven. Majority of the stations have low Degree, Closeness and

Betweenness Centrality. These means that only a few stations play important roles because they have more connections, are more accessible and more frequently passed by. In fact, this is not unusual in a complex network with a lot of nodes and links.

Table 3.1 presents the correlation between the selected indicators. Service frequency is positively correlated with all three centrality measures but only significantly correlated to both Betweenness Centrality and Degree Centrality. This means that stations with higher service frequency are usually the ones with more connections, are more accessible and more frequently passed by.

Regarding the relationship among the three centrality measures, Betweenness Centrality and Closeness Centrality are significantly correlated with each other, whereas Degree Centrality is negtively but not significantly correlated with the other two centrality measurements. This means that stations that are more accessible or more frequently passed by do not necessarily have more connections. By contrast, Closeness Centrality and Betweenness Centrality are positively correlated with each other, which means that the stations that are more accessible are also the ones that are more frequently passed by.

3.5.2 Principle Component Analysis (PCA)

A loading plot shows how strongly each indicator, which is shown as vector in Figure 3.2, influences the clustering process. The loading plot in Figure 3.2 shows how the vectors are pinned at the origin of the Principal Components (PCs), where PC1 = 0 and PC2 = 0. The direction of the arrows shows that Betweenness Centrality and Closeness Centrality have more say than Degree Centrality and service frequency in PC1, while Degree Centrality and service frequency strongly influence PC2 but in different direction.

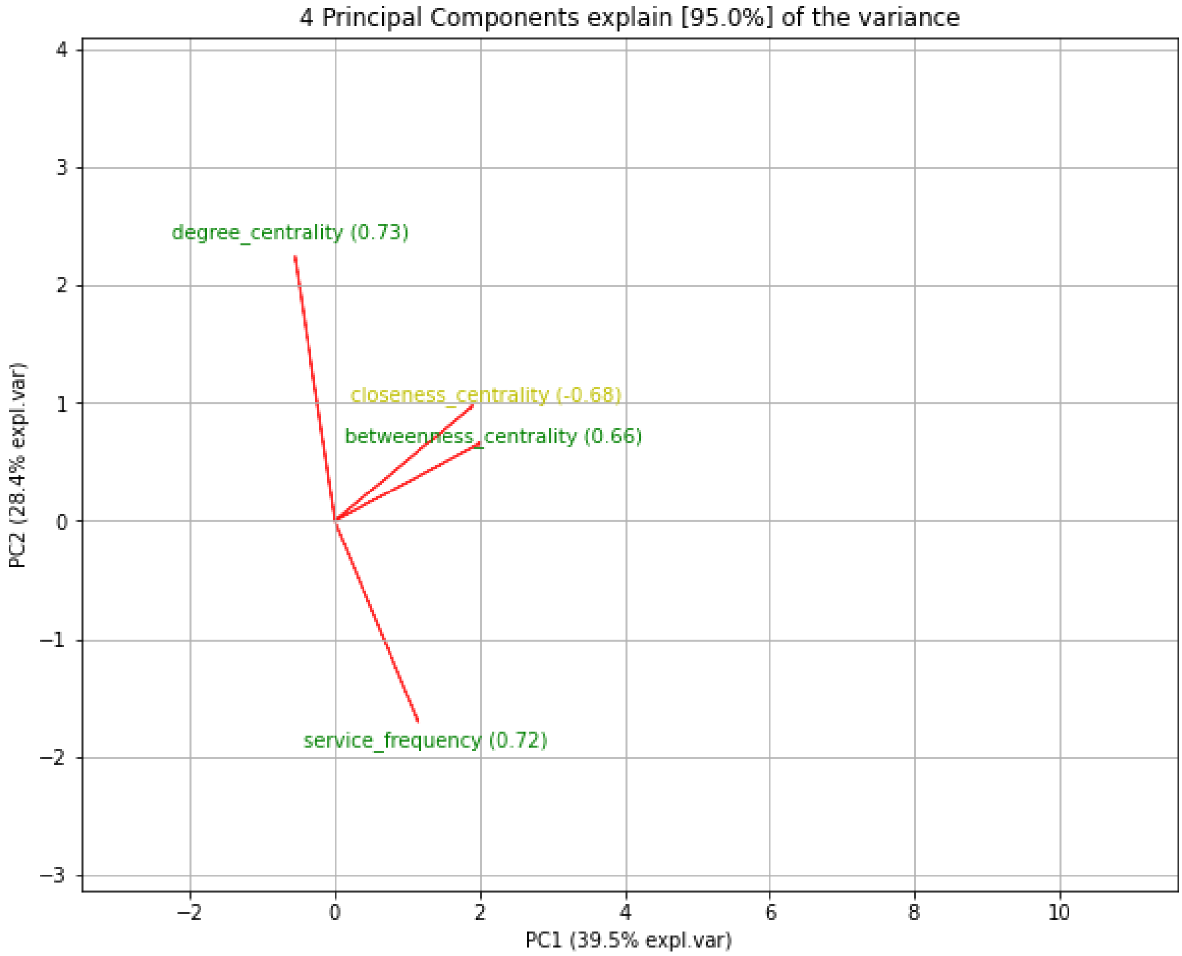


Figure 3.2 Loading plot of the indicators.

The indicators' project values (shown by the lengths of the arrows) on each PC demonstrate how much weight they have on that PC. The longer the red arrow, the more influence the indicator has on the PCs. Figure 3.2 shows that Degree Centrality and service frequency have more influence on the PCs than Closeness Centrality and Betweenness Centrality.

3.5.3 K-means for clustering stations

Next, k-means clustering algorithm is employed to group the data. The ideal number of clusters for the k-means model can be determined by Elbow Method, which measures the sum of the squared distances to the nearest cluster center.

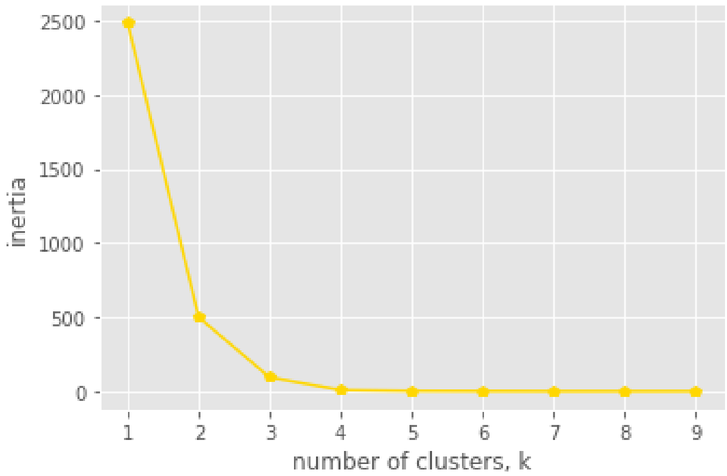


Figure 3.3 Inertia plot for choosing the optimal number of clusters.

The first step of this method is to plot the explained variation as a function of the number of clusters in the inertia plot, as shown in Figure 3.3. The second step is to choose the number of clusters by finding the elbow point in the inertia plot where adding another cluster does not improve very much the modeling of the data (Ketchen & Shook, 1996). It may be intuitive to think that dividing the stations into more clusters will explain more of the variation among the stations. However, a model with too many clusters will also be over-fitting. Therefore, the inertia plot can inform us how much additional information will be added as the number of clusters increases.

Figure 3.3 shows that the change of inertia is becoming less and less as the number of clusters increases. We have found that analyzing three clusters can be a good strategy, since it allows focusing both on the clusters themselves and on the distribution of stations within those clusters. Therefore, we consider 3 as the elbow point and this will be the number of valid clusters for our data set. The results of k-means clustering and the stations in each cluster are presented in Table 3.2.

Table 3.2 List of stations in each cluster.

Cluster 0
Agathenburg, Ahrensburg Ost, Ahrensburg West, Altona, Aumühle, Billstedt, Blankenese, Buchenkamp, Buckhorn, Buxtehude, Dammtor (Messe/CCH), Diebsteich, Dollern, Elbbrücken, Fischbek, Großhansdorf, HafenCity Universität, Halstenbek, Hamburg Airport (Flughafen), Hammerbrook (City Süd), Harburg, Harburg Rathaus, Heimfeld, Hoheneichen, Hoisbüttel, Holstenstraße, Horneburg, Iserbrook, Joachim-Mähl-Straße, Kiekut, Krupunder, Mümmelmannsberg, Neu Wulmstorf, Neugraben, Neukloster, Neuwiedenthal, Niendorf Nord, Norderstedt Mitte, Ohlsdorf, Ohlstedt, Pinneberg, Reinbek, Rissen, Sülldorf, Schippelsweg, Schmalenbeck, Stade Sternschanze (Messe), Thesdorf, Überseequartier, Veddel(BallinStadt), Volksdorf, Wedel, Wilhelmsburg, Wohltorf
Cluster 1
Allermöhe, Alsterdorf, Alte Wöhr (Stadtpark), Alter Teichweg, Bahrenfeld, Barmbek, Baumwall (Elbphilharmonie), Bergedorf, Berliner Tor, Berne, Billwerder-Moorfleet, Borgweg (Stadtpark), Christuskirche, Dehnhaide, Eidelstedt, Elbgaustraße, Emilienstraße, Eppendorfer Baum, Farmsen, Feldstraße (Heiligengeistfeld), Friedrichsberg, Fuhlsbüttel, Fuhlsbüttel Nord, Gänsemarkt, Garstedt, Habichtstraße, Hagenbecks Tierpark, Hagendeel, Hallerstraße, Hauptbahnhof, Hamburger Straße, Hasselbrook, Hauptbahnhof Süd, Hochkamp, Hoheluftbrücke, Hudtwalckerstraße, Kellinghusenstraße, Kiwittsmoor, Klein Borstel, Klein Flottbek (Botanischer Garten), Klosterstern, Kornweg (Klein Borstel), Landwehr, Langenfelde, Langenhorn Markt, Langenhorn Nord, Lattenkamp (Sporthalle), Lohmühlenstraße, Lübecker Straße, Lutterothstraße, Meiendorfer Weg, Merkenstraße, Meßberg, Messehallen, Mittlerer Landweg, Mönckebergstraße, Mundsburg, Nettelnburg, Niendorf Markt, Ochsenzoll, Oldenfelde, Osterstraße, Othmarschen, Poppenbüttel, Rübenkamp (City Nord), Rathaus, Richtweg, Ritterstraße, Rödingsmarkt, Rothenburgsort, Saarlandstraße, Schlump, Sengelmannstraße (City Nord), Sierichstraße, St. Pauli, Steinfurther Allee, Steinstraße, Stellingen (Arenen), Stephansplatz (Oper/CCH), Straßburger Straße, Tiefstack, Trabrennbahn, Uhlandstraße Wandsbek Markt, Wandsbeker Chaussee, Wandsbek-Gartenstadt, Wartenau
Cluster 2
Wellingsbüttel Burgstraße, Hammer Kirche, Hauptbahnhof Nord, Horner Rennbahn, Jungfernstieg, Königstraße, Landungsbrücken, Legienstraße, Rauhes Haus, Reeperbahn, Stadthausbrücke

3.5.4 Spatial distribution of clusters of stations

One advantage of studying transit systems as networks is that, compared to other types of networks, their size is relatively small and it is easier to make an initial visual inspection. Figure 3.4 illustrates the location and the patterns of the spatial distribution of different clusters. The map shows that cluster 2 is predominantly distributed in the inner city. Cluster 1 is mainly located in both inner and middle suburbs. Cluster 0 is mainly distributed in the middle and outer suburbs.

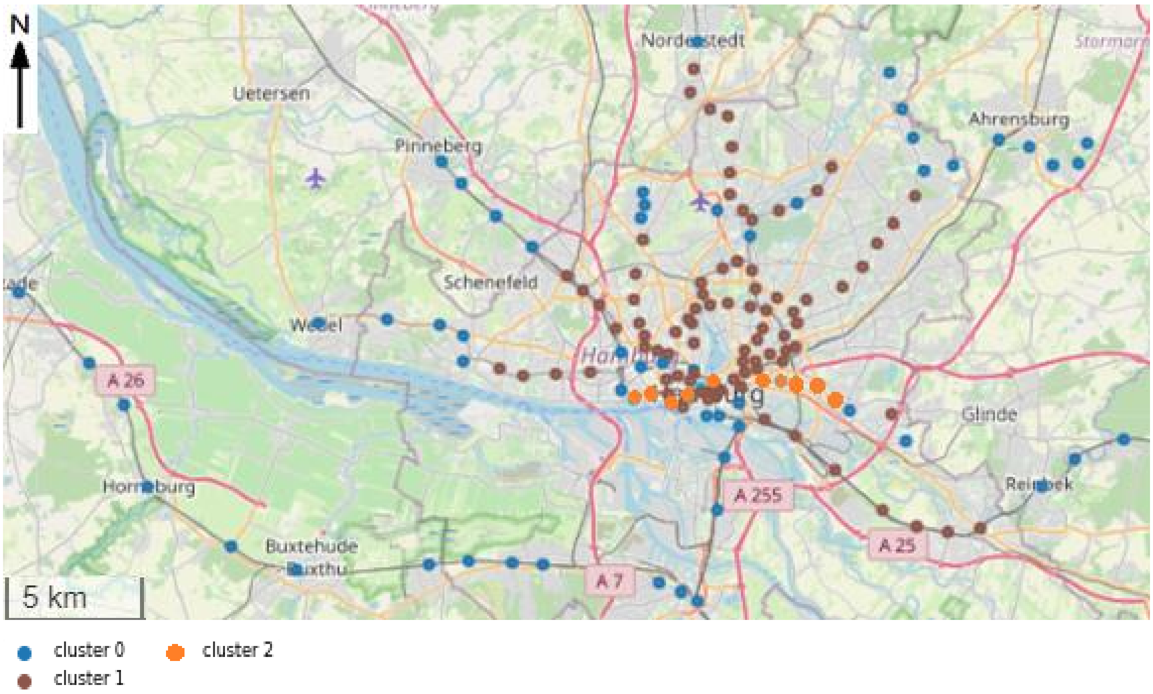


Figure 3.4 Spatial distribution of stations and clusters.

3.5.5 Characterizing clusters with descriptive statistics

The characteristics of each cluster are described by the descriptive statistics, including the mean and the standard deviation of the indicators. Results in Table 3.3 shows that cluster 2 has the highest mean value of all 4 indicators except for Degree Centrality. Both cluster 0 and cluster 1 have very similar average Betweenness Centrality. Cluster 0 has the second highest average value of Closeness Centrality than cluster 1 but it has lower average service frequency than cluster 1.

Figure 3.5 plots the distribution of indicators of the stations in each cluster. Stations in cluster 2 show distinctive characteristics in all indicators, which makes them stand out from the other two clusters. Stations of cluster 2 have the highest service frequency as well as the highest Closeness and Betweenness Centrality. Also, the stations of this cluster are more homogeneous in terms of Closeness and

Betweenness Centrality because they all concentrate on the right side of the figure, which indicates higher values.

Table 3.3 Descriptive statistics of indicators by individual cluster

	All stations		Cluster 0		Cluster 1		Cluster 2	
	mean	std[7]	**mean**	std	**mean**	std	**mean**	std
Service frequency	**10.0974**	3.79847	**5.50909**	1.16861	**12.0000**	0.00000	**17.8181**	0.40452
Average Degree centrality	**0.01245**	0.00293	**0.01326**	0.00325	**0.01197**	0.00259	**0.01230**	0.00317
Average Closeness centrality	**0.00217**	0.00093	**0.00217**	0.00083	**0.00211**	0.00095	**0.00271**	0.00107
Average Betweenness centrality	**0.00011**	0.00020	**0.00009**	0.00005	**0.00009**	0.00005	**0.00034**	0.00073

The biggest difference between cluster 0 and 1 is the service frequency. Cluster 0 has the lowest service frequency among the three clusters. Although the average Closeness and Betweenness Centralities are similar, as shown in Table 3.3, the median values, indicated by the orange lines in Figure 3.5, are different among the two clusters. The median values of Closeness and Betweenness Centrality are both higher in cluster 1 than in cluster 0.

Based on these results, one can determine that stations of cluster 2 have the highest Node value. This is followed by cluster 1 and cluster 0.

[7] Std = standard deviation

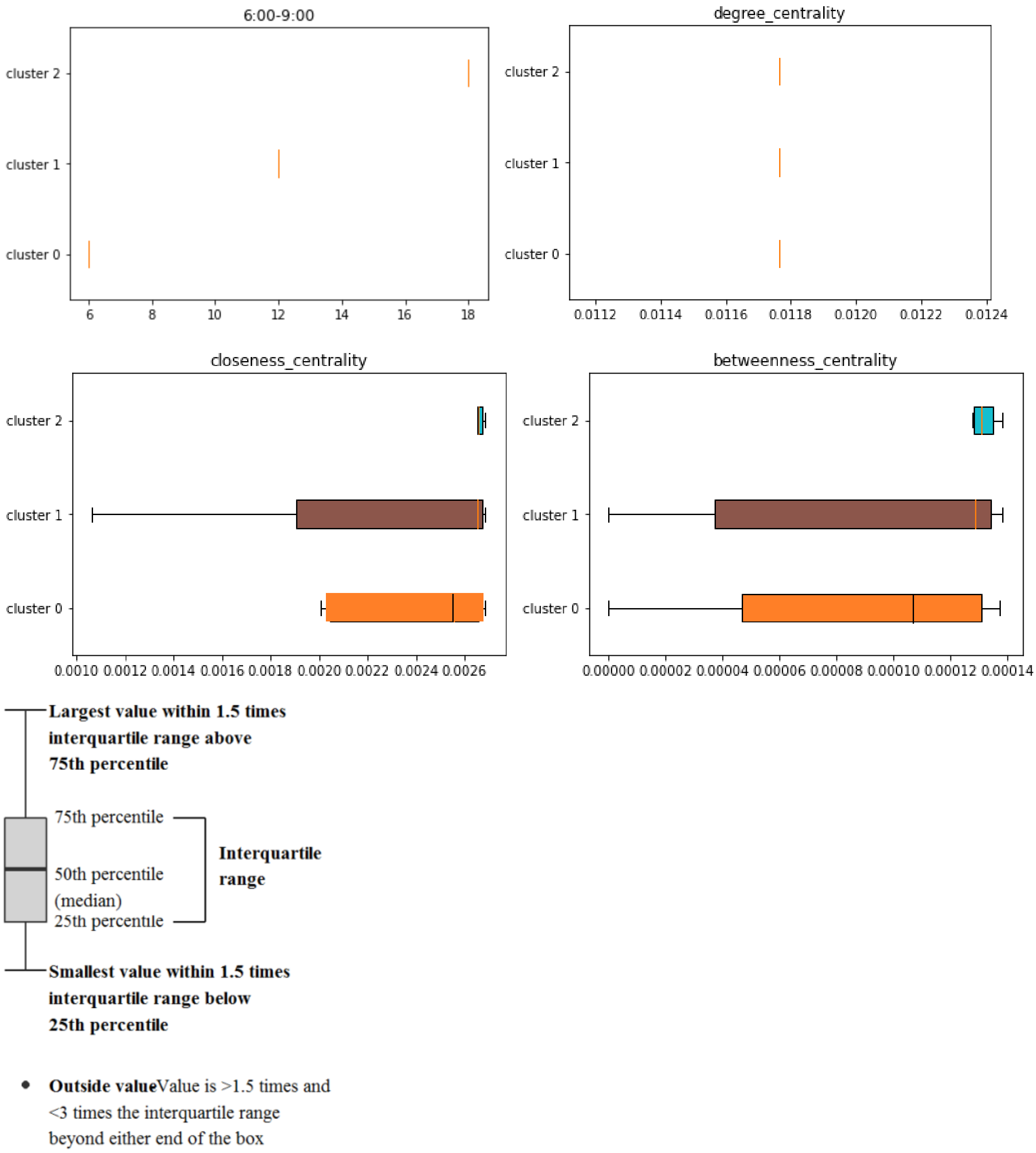


Figure 3.5 Characteristics of each cluster.

3.6 Conclusion and discussion

Both chapter 2 and chapter 3 investigate the public transit network in Hamburg. In chapter 2, the topological properties of the integrated transit network, which include both railway public transport and bus stations, are investigated. In chapter 3, the target is to assign each station with its Node Value in the Node-Place Model.

Using the four measurements of Degree Centrality, Closeness Centrality, Betweenness Centrality and service frequency, three clusters of stations can be identified and characterized. The results show that cluster 2 is predominantly located in the inner city and have the highest Node Value. Cluster 1 is mainly distributed in both inner and middle suburbs. Cluster 1 has the second highest Node Value. Cluster 0 is mainly distributed in the middle and outer suburbs. Cluster 0 has the lowest Node Value.

In the current study, we calculate the departures per hour per direction from 6 a.m. to 9 a.m. on the weekdays. This is certainly not enough to fully explore spatial-temporal interactions in transportation networks. Further studies can carry out a variety of comparisons by differentiating various time slots of service frequency, such as “busy and quiet periods”, “busy periods before and after working hours”, “weekday and weekend”, “daytime and nighttime”. The results may reveal some interesting patterns and requirements, e.g. minimum required frequency for different times in a well-used transportation system. Also, this deeper investigation might be necessary if the internal organization of the network (e.g., most common routes, transferring stations, etc.) differs between the weekday and the weekend. Finally, special events with massive crowds, such as football matches or concerts in a big stadium, can also be investigated additionally and be compared with the regular routine pattern.

The Node Value defined in this chapter is the first step in constructing the Node-Place Model and it also provides a foundation for the later chapter, where each cluster will be further classified by their Place Value. After each station is assigned with both the Node and the Place Value, it will be possible to construct the Node-Place Model and identify the stations that need some improvement. The measurements are used for classifying the stations also indicate the direction for improvement in each cluster.

4 Comparability of urban street networks: Consideration of the size effect in the evaluation of network characteristics and a proposal for determining an appropriate size of the catchment area for pedestrian networks[8]

4.1 Introduction

4.1.1 The challenge

When evaluating the real-world spatial networks, the central point and the catchment area around it need to be defined before carrying out network analysis, so that the indicators of different networks can be compared with each other. However, the real-world spatial networks are not strictly discreet systems, and the boundaries of their catchment areas are often arbitrarily defined. This arbitrarily defined size of the catchment area often exerts significant distortion on the values of network indicators (Greenberg, 2020) (Boeing, 2020) (Marshall, Gil, Kropf, Tomko, & Figueiredo, 2018) (Gil, 2017) (Van Meter, et al., 2010) (Pipley, 1981) for the following reasons.

If the size of the catchment area is too large, it is more likely that it contains multiple sub-networks or sub-structures with different characteristics. In other words, the mixed characteristics of the entire network are often a combination of sub-structures, such as grids, trees, hubs and spokes, or lines structures. And each type of network structure may differ in complexity (Żochowska & Soczówka, 2018) (Soczó wka, Żochowska, & Karoń, 2020). Since the value of the indicator is the average value for the entire network, the complexities of the sub-structures within the network may affect the calculation of the indicator for the evaluation of the characteristics and attributes.

On the other hand, if the catchment area is too small, the network model excludes the characteristics beyond the arbitrarily defined boundary of the model. Since the analytic algorithms are relational, the world outside the catchment area of

[8] This chapter has been submitted to the International Journal of Sustainable Development and Planning in May 2021, accepted in October 2021 and is expected to be published in Issue 7.

analysis still affects the world inside it and the movement patterns within (Gil, 2017) (Greenberg, 2020).

Concerns over the size effect on the reliability or significance of the network analysis results have been expressed and shared by many researchers (Gil, 2017) (Greenberg, 2020) (Ratti, 2004) (Park, 2009) (Krafta, 1994) (Sadler, Gilliland, & Arku, 2011). They test the size effect on the performance of spatial network models by varying the threshold radius of the catchment area. For example, Yoshimura et al. (2020) have investigated Betweenness Centrality with radius of the catchment area varying from 300 to 5000 meters with 100 meters step, and the results show that the indicator value is sensitive to the size of the catchment area. To be more specific, the indicator value of the same node or link in the network may change with the size of the catchment area. And this distortion is particularly pronounced for the nodes and links at the border of the catchment area (Gil, 2017) (Okabe & Sugihara, 2012). For example, Gil (2017) has found that nodes or links near the center of the catchment area tend to have higher degree of Betweenness Centrality compared with those close to the border. Accordingly, this influence has been called the “edge effect” or “boundary effect” (Gil, 2017) (Greenberg, 2020) (Park, 2009) (Ratti, 2004) (Okabe & Sugihara, 2012).

In this research, we focus on the effect of size of the catchment area on the indicator value. This effect is hereafter referred to as the “size effect”.

4.1.2 Types of research where size effect should be considered

Consideration of the size effect may be particularly crucial in research projects associated with the following purposes.

Research aiming at comparing and classifying multiple networks distributed in different locations. For any comparative study, where more than one network is included, the size of the catchment area will either have to be identical or it must be examined as one of the factors that may explain the variation of the measurement values among the multiple networks.

Research requiring the distinction among different types of mobility networks based on the mode of transportation, such as the networks for pedestrian, cyclists or motorized vehicles. For example, “a small road segment in a residential area might be important for pedestrians, but it is almost negligible for motorized transportation. In this sense, the road segment should have at least two kinds of importance: one for pedestrians and the other for motorized transportation" (Yamaoka, Kumakoshi, & Yoshimura, 2021). In other words, the value of the indicator varies depending on whether the catchment area is at the human or vehicle scale. Without considering the size of the study area, the degree of Betweenness Centrality cannot represent multiple aspects of the city that is perceived by people in the real world (Porta, Crucitti, & Latora, 2006). To avoid these problems, Yoshimura *et al.* (2020) propose that, in addition to the global Betweenness Centrality for the entire city, a set of local Betweenness Centrality values at a smaller neighborhood scale shall also be calculated.

Research where an area needs to be divided into sections of smaller size with regular shape. The city area is often divided into smaller units due to the structural complexity and size associated with the scale problem (Ahuja, 1983) (Bell, Diaz, Holroyd, & Jackson, 1983) (Boots, 1980). For example, the analysis of connectivity of the road and street network structure is often conducted with the application of topological measures and typically requires that a large geographical space be divided into smaller parts with regular shape (Soczówka, Żochowska, & Karoń, 2020). In a recent study, in order to determine the size of the catchment area on the value of the selected measures, Soczówka *et al.* (2020) have carried out a comparative analysis by dividing the analyzed area into different sizes with regular shape and comparing the values of connectivity measurements of the road and street network structure. In such a case, the size of a single basic catchment area is very important because it can significantly affect the computational results of these measures.

Research measuring the spatial accessibility. Arbitrary administrative boundaries (such as census tracts or block groups) are often used in the studies of spatial accessibility. The arbitrarily defined border may lead to a methodological limitation (Wan, Zhan, Zou, & Chow, 2012) (Ngui & Apparicio, 2011) because, first of all, accessibility involves movement and a given boundary does not actually prevent people or vehicles traveling across the border from reaching the facilities, services, amenities or any points of interest (Fortney, Rost, & Warren, 2000). Secondly, a given boundary may exclude behavior outside the catchment area (Sadler, Gilliland, & Arku, 2011) (Wan, Zhan, Zou, & Chow, 2012) (Ngui & Apparicio, 2011) (Fortney, Rost, &

Warren, 2000) (Salze, et al., 2011) (Luo, Tian, Luo, Yi, & Wang, 2017) (Donohoe, et al., 2016) (Bissonnette, Wilson, Bell, & Shah, 2012) (Van Meter, et al., 2010) (Sharkey & Horel, 2008) (Wang & Luo, 2005) (Vidal Rodeiro & Lawson, 2005) (Jordan, 2004). Thirdly, the resource or services beyond a defined arbitrary boundary may influence the behavior within the catchment area (Van Meter, et al., 2010). Many research projects studying the spatial accessibility have pointed out the risk that the accessibility of facilities may be biased (Fortney, Rost, & Warren, 2000) (Donohoe, et al., 2016) (Van Meter, et al., 2010) (Iredale, Jones, Gray, & Deaville, 2005) or under-reported (Sadler, Gilliland, & Arku, 2011) (Salze, et al., 2011) (Sharkey & Horel, 2008) (Wang & Luo, 2005) (Zhang, Lu, & Holt, 2011) because "areas close to the boundary may be classified as having poor geographic access even though they may in fact be proximate to resources across the boundary" (Gao, et al., 2017).

4.1.3 Proposed solutions for mitigating the distortion

In an attempt to mitigate, contrast or diminish the distortions stemming from the size effect, the proposed features and principles underlying further classification can be broadly categorized into the following two approaches.

Adding a buffer zone with the assigned threshold radius. The first approach is to create buffer zones around the catchment area in order to compute the indicator value (Greenberg, 2020) (Gil, 2017) (Pezzica, Cutini, & De Souza, 2019) (Penn, Hillier, Banister, & Xu, 1998) (Hillier, Penn, Hanson, Grajewski, & Xu., 1993). For example, the Python library OSMnx (Boeing, 2017) (Boeing, 2019) automatically creates a buffer of half a kilometer around the requested area so that each node has a correct street count. The buffer zones are then trimmed from the constructed network model. The indicator values of nodes and links within the buffer zone are excluded from the analysis because they are not reliable.

Using a homogeneous feature as the criterion for the division. Soczówka et al. (2020) propose to divide the city area into smaller catchment areas with homogeneous features before conducting the analysis. They suggest several possible criteria of the homogeneous features including, firstly, administrative criteria, such as administrative district boundaries; secondly, structural criteria, such as the spatial distribution of density of inhabitants in households; and, thirdly, technical and functional criteria, such as the public transport management system.

In short, any divisions and classifications of geographical space should define the boundary of the catchment area. In response to these considerations, the current research intends to contribute to this decision process and to quantify the effect on the value of the indicators by 1) investigating whether and to what extend indicators, such as the average Node Degree, change with the size of the catchment area; 2) proposing a methodology to decide the size of the catchment area based on the average street length; and 3) offering the recommendation for the appropriate size of the catchment area for the investigation of pedestrian networks.

A theoretical model representing an idealized regular network (see Figure 4.1) has been created as a reference model for the clear mathematical deduction of the relationship between the size of the catchment area and the changes in the indicator values. This allows us to answer the following research questions in a more controlled environment, where the differences in the values of the indicator will only be caused by the size effect[9].

- How do we decide the appropriate size for the investigation of pedestrian networks?
- Are there upper and lower limits of the size effect on the value of an indicator?
- How big should the network catchment area be in order to be able to compare it with the reference model?

4.2 Method

4.2.1 The investigated indicator

The indicator we have chosen to examine the size effect is the average Node Degree, *k_avg*, which is expressed by

$$k\_avg = 2 \times \frac{M}{N} \quad (1)$$

where *M* refers to the total number of links and *N* refers to the total number of nodes. In addition, the average Node Degree provides the information of the network pattern and can also be used to evaluate the level of "gridness." A network with a lot of nodes connecting to 4 links, i.e. $k \cong 4$, means that the network is more likely to be grid-

[9] This is not ignoring the fact that, in the case of real network, the size effect may still exist regardless of the size of the catchment area. Because a real network is not regular to eternity and, therefore, its characteristics change.

pattern. And "more-gridded cities have higher connectivity (i.e., higher average Node Degrees, more four-way intersections, fewer dead-ends etc.) and less-winding street patterns" (Boeing, 2019). This means that, holding the number of nodes constant, if the average Node Degree of a real network is smaller than that of a grid-pattern network, the connectivity of the real network is worse than the connectivity of a grid-pattern network.

Before examining the size effect on the average Node Degree, we need to provide some definitions concerning our theoretical model.

4.2.2 The theoretical mathematical model

For the investigation of the size effect, a real network might not be a good basis to be the reference for comparison. The main problem is that the size effect cannot be clearly separated from other effects. Therefore, a theoretical model for the prognosis of the size effect was created in order to provide the method of mathematical analysis of an idealized regular network. In this way, the differences in the values of the indicator will only be caused by the size effect. Our definition of an idealized network consists of the following components.

Definition of a single quadrat as the basic element. To illustrate the rules of how the network is created, we start with a network that is just a quadrat (square). As shown in (a) in Figure 4.2 and Table 4.1, a quadrat has four links as its boundary, and it serves as the basic unit of the quadratic network with four nodes ($N = 4$) and four links ($M = 4$). We use d to refer to the length of a link in the idealized network, which is also the side length of a quadrat.

Definition of a network. A quadrat is expanded and extended in the same quadratic pattern, as shown in (b), (c) and (d) in Figure 4.2 and Table 4.1. The size of the network is measured by the total number of nodes, N, and the total number of links, M. Element (b) in Figure 4.2 and Table 4.1 shows that a network consisting of four quadrats has nine nodes ($N = 9$) and 12 links ($M = 12$).

Definition of a catchment area. In the case of real networks, a network is different from a catchment area and the selected catchment area will be smaller than the entire network. As illustrated in Figure 4.1(a) and (b), the black solid line indicates the links and also the boundary of a quadratic network and the red dashed line indicates the boundary of a catchment area. A catchment area is like the section that cuts off links connecting nodes from inside of the catchment area to nodes outside of the catchment area[10]. The link cut off by the catchment area could be counted as a complete or a half link. And the value of the indicator will be affected by how these links are counted. Alternatively, they might be excluded from the calculation altogether. For the purpose of this paper, we assume that these cut-off links can be left out, and they are excluded from the calculation.

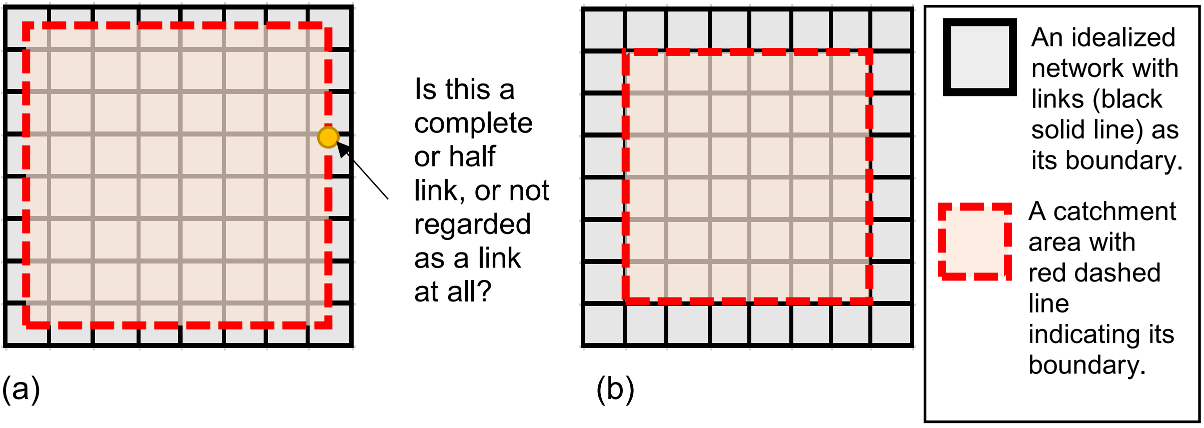


Figure 4.1 Abstracted expression of the boundaries of the idealized network and the catchment area.

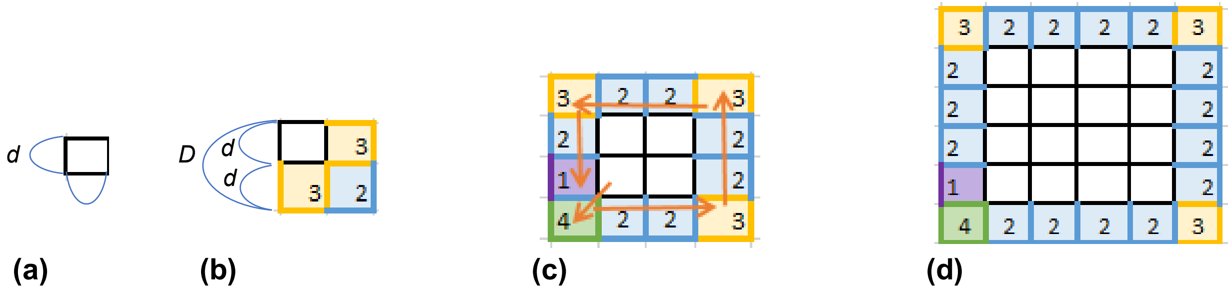


Figure 4.2 Abstracted expression of the process of increasing the size of the catchment area. (a) One quadrat (square). The network with one quadrat is the smallest network. (b) A network consists of 4 quadrats. Firstly, 3 links are added to create yellow quadrats.

This means that the boundary of the catchment area consists of the links of the network as shown in Figure 4.1(b). Therefore, the side length of a catchment area, *D*, consists of one or multiple links of the network. And the size of the catchment area is *D x D*.

There can be two kinds of relationships between the network and the catchment area.

[10] In a real network, it is not necessarily the case that a catchment area cuts off links connecting nodes from inside of the catchment area to nodes outside of the catchment area. Sometimes, the boundary of the catchment area may coincidently lie exactly on a link as shown in Figure 4.1(b).

When the catchment area consists of one quadrat and D consists of single d (i.e. $D = d$). In the case of the smallest catchment area (as shown in element (a) in Figure 4.2 and Table 4.1), there is only one quadrat and, therefore, the side length, D, of this catchment area is equal to d. The size of the catchment area is $D \times D = d^2$. The network in this catchment area has 4 links ($M = 4$) and 4 nodes ($N = 4$). The average Node Degree, *k_avg*, is accordingly 2.

When the catchment area consists of multiple quadrats and D consists of multiple d. In element (b) in Figure 4.2 and Table 4.1, there are 4 quadrats and, therefore, the side length, D, of this catchment area equals to $2d$. The size of the catchment area is $D \times D = 4d^2$. The network in this catchment area has 12 links. The total number of links, M, in this network is 12 and the total number of nodes, N, is 9. The average Node Degree, *k_avg*, is accordingly 2.67.

The relationship between the quadrat, the network and the catchment area of various sizes are summarized in Table 4.1 and it shows that, with the increasing side length, D, of the catchment area, the average Node Degree, *k_avg*, also increases. This means that there is a size effect on the indicator of average Node Degree. This trend is noteworthy, so we break down the process into steps in order to investigate the details. The following discussion is divided into two parts: the additional links and the additional nodes.

Table 4.1 Relationship between the quadrat, network and catchment area of various sizes.

Elements in Figure 4.2		(a)	(b)	(c)	(d)
Number of quadrats		1	4	16	36
Network	Total number of links (M)	4	12	40	84
	Total number of nodes (N)	4	9	25	49
	Average Node Degree (*k_avg*)	2	2.67	3.2	3.43
Catchment area	Side length of catchment area (D)	d	$2d$	$4d$	$6d$
	Size of catchment area ($D \times D$)	d^2	$4d^2$	$16d^2$	$36d^2$

4.2.3 Process of increasing the size of the quadratic network

Figure 4.2 presents the abstracted expression of the process of increasing the size of the catchment area. Figure 4.2 (a) shows one quadrat (square). The network with one quadrat is the smallest network.

i) Increasing the number of links

Figure 4.2 (b) shows a network consists of 4 quadrats. To extend the network from (a) to (b) in Figure 4.2 and Table 4.1, three links are added to create the yellow quadrats. And then two more links are added to create the blue quadrat.

For the network in (c) in Figure 4.2 and Table 4.1, which is a network consisting of 16 quadrats, we begin with a corner and, firstly, create the green quadrate with four links. Next, two more links are added to create another two blue quadrats. Thirdly, three links are added to create the yellow quadrat in the corner. Finally, these three steps are repeated until the final purple quadrate, which needs only one additional link to be created.

As the network becomes bigger and bigger, there are more and more blue quadrats, which are formed by two additional links, among the newly created quadrats. In other word, the number of blue quadrats increases much faster than other quadrats. Eventually this type of quadrat becomes more and more important and thus dominant the pattern of the increasing side length, *D*. Meanwhile, the yellow corner quadrat, which consists of 3 links, becomes less dominant.

ii) Increasing the number of nodes

Since the blue quadrat is more dominant than the other types of quadrats on the increasing side length, *D*, the focus of the investigation is on this type of quadrats. For every blue quadrat, it takes two additional links to crease one additional node. Eventually the total number of nodes, *N*, will be half of the total number of links, *M*. Therefore,

$$N = \frac{1}{2}M \tag{2}$$

or, in other words,

$$M = 2\,N \tag{3}$$

This means that average Node Degree, *k_avg*, will eventually approach the final limit.

$$k\_avg = 2 \times \frac{M}{N} = 2 \times \frac{2N}{N} = 4 \tag{4}$$

The results are shown in Figure 4.3 and Table 4.2.

.

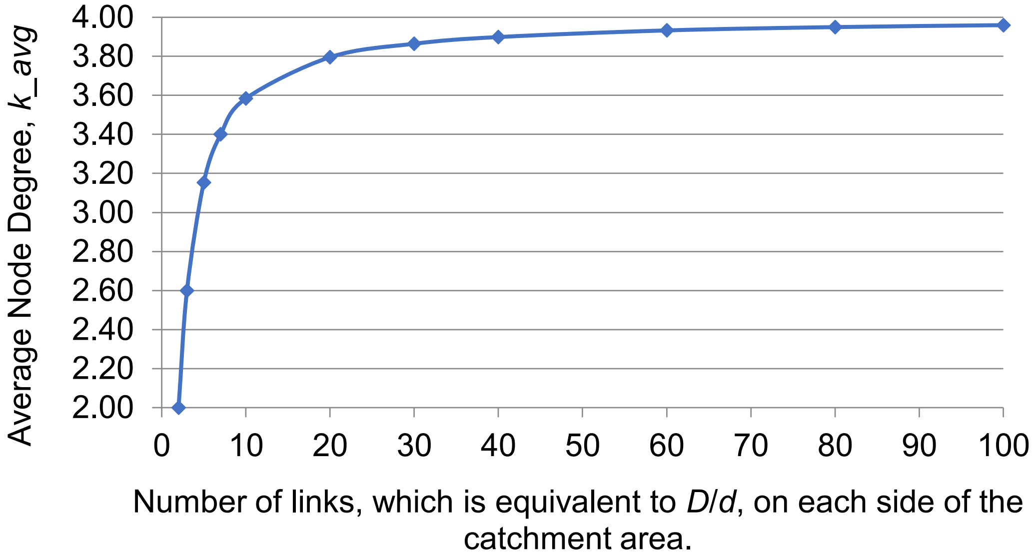


Figure 4.3 Relationship between the value of the average Node Degree, *k_avg*, and the size of the catchment area, which is expressed by the number of links on each side of the catchment area.

4.2.4 Appropriate size of the catchment area for pedestrians

For the investigations of pedestrian networks, it makes no sense to explore a large area. If we consider that the pedestrians' maximum acceptable walking time is about 15 to 20 minutes, the size of the catchment area would be between 1500x1500m² to 2000x2000m². Also, following the findings regarding the size effect, the comparison between different "pedestrian" networks is only correct and thus possible if all catchment areas have the same size.

Table 4.2 Relationship between size of the catchment area and the values of indicators. The size of the catchment area is indicated by its side length, D, and D is indicated by the number of links, which is equivalent to D/d.

Side length, *D*, of the catchment area	Links in the network			New quadrat				Indicators			
Number of links on each side of the catchment area	Total number of links, *M*	Total additional links of base quadrat, indicated by the black links	Total additional links required to create the new quadrat, indicated by colorful links	Number of new green quadrats formed of 4 additional links	Number of new yellow quadratic formed of 3 additional links	Number of new blue quadratic formed of 2 additional links	Number of purple quadrats formed of 1 additional links	Total number of nodes, *N*	Average Node Degree, k_avg	nodes / area	Area
1	4	4	4	1	0	0	0	4	2.00	4.00	1
2	12	4	8	0	2	1	0	9	2.67	2.25	4
4	40	12	28	1	3	7	1	25	3.20	1.56	16
6	84	40	44	1	3	15	1	49	3.43	1.36	36
8	144	84	60	1	3	23	1	81	3.56	1.27	64
10	220	144	76	1	3	31	1	121	3.64	1.21	100
12	312	220	92	1	3	39	1	169	3.69	1.17	144
14	420	312	108	1	3	47	1	225	3.73	1.15	196
16	544	420	124	1	3	55	1	289	3.76	1.13	256
18	684	544	140	1	3	63	1	361	3.79	1.11	324
20	840	684	156	1	3	71	1	441	3.81	1.10	400
40	3280	2964	316	1	3	151	1	1681	3.90	1.05	1600
50	5100	4704	396	1	3	191	1	2601	3.92	1.04	2500
100	20200	19404	796	1	3	391	1	10201	3.96	1.02	10000

4.3 Results and implications

The benefit of investigating the size effect with a theoretical model is clear because, with the controlled condition, we can find out the scenario when the size effect is (nearly) saturated and deliver a correct prognosis about the total size of the network. Figure 4.3 shows the relationship between the value of the average Node Degree, *k_avg*, and the size of the catchment areas, which is expressed by the number of links on each side of the catchment area.

It can be concluded that the size effect is very notable until the side length of the catchment area, *D*, equals 10 times of the length of the link, i.e. *D* = 10*d*. And it is still notable with a derivation of 3.8/4 = 5% when *D* = 20*d*. The derivation is reduced to 2.5% when *D*=40*d*, which can be regarded as a threshold and recommended as the size that is big enough to allow comparisons between different catchment areas and/or different networks. Eventually, it comes close to saturation where D = 50d. For values D >= 50d the average Node Degree, *k_avg*, reaches the ideal value of the extended network and the size does not play a role after that.

The principal implication of this theoretical model is twofold. First of all, the above investigation provides evidence to show that the size effect is remarkable and cannot be neglected until D = 40d.

Secondly, Table 4.3 shows the relationship between the average street length and the catchment area size range. The results offer a principal guideline for determining the size of the catchment area where the real network can be compared with the theoretical model. In the real street network, the average length of a street, which is the length of the link, d, in the theoretical model, is mostly between 50m and 100m. Assuming that d = 100m, the size of the selected catchment area has to be at least 20d x 20d = 2000x2000m^2 in order to be able to compare it with the theoretical model. If the side length of the catchment area is 40d, i.e. 4000m, and the size of the catchment area is at least 4000x4000m^2, the size effect will be even less significant in the theoretical model. Therefore, we provide the evidence to support that the lower and upper limits of the size of the catchment area should be 2000x2000m^2 and 4000x4000m^2.

Table 4.3 Recommendations for the lower and upper limits of the side length of the catchment area

Average street length, *d*	**50m**	**100m**
Lower limit of the side length of the catchment area	20d = 1000m	20d = 2000m
Upper limit of the side length of the catchment area	40d = 2000m	40d = 4000m
Range of the sizes of catchment areas	1000x1000m^2 ~ 2000x2000m^2	2000 x 2000m^2 ~ 4000 x 4000m^2

To sum up, the results in this research contribute to the argument that a catchment area with an area size that is too large would not be practical due to the following reasons.

1) The characteristics and patterns of the street networks in a real city may vary from quarter to quarter. For example, the characteristics of the network in the historical center would be different from its surrounding areas. Therefore, if the size of the catchment area is too big, the values of the indicators would reflect not the information about one type of network but rather about a sum of multiple types of networks in several neighboring and connected quarters. Such mixed information

would be less valuable for investigating the relationship between the street network structure and the indicators or for classifying the street networks.

2) The average street length can be one of the indicators for determining the size of the catchment area. According to our analysis, in the theoretical network, the size effect on the indicator is not very significant when the size of the catchment area is larger than 4000 x 4000m^2. Therefore, any size larger than 4000 x 4000m^2 would not be necessary.
3) The average Node Degree of an idealized regular network changes with the size of the catchment area. But the variation becomes nearly neglectable when the size of the catchment area is larger than D = 40d and vanishes with D >= 50d. This means that, if the size of the catchment area of the real network is larger than D = 40d to 50d and the average Node Degree is 4, the network pattern has a grid-like character. However, if the average Node Degree of a real network is smaller than 4, its connectivity is worse than that of the theoretical grid-pattern network and vice versa.
4) We suggest that in all future investigations the size of the catchment area should be defined before carrying out further analysis.
5) When comparing the indicators of multiple catchment areas, all of the catchment areas should have the same size as long as D is less than 40d to 50d.
6) Results from the previous literature are only comparable if they refer to the same size of chosen catchment area.
7) Average Node Degree has been chosen because its behavior can be calculated for the theoretical and idealized networks proposed in the current research in order to quantify and to demonstrate the size effect on the indicator values step by step. To what extent other indicators reach the saturation until the size effect vanishes should be part of further investigations. It would be interesting to test whether their threshold is also around D = 40 to 50 d. In future research, the analysis of the variability and the sensitivity of indicators shall facilitate decisions on which indicators should be used for a given size of the catchment area.

5. Placement effect and Closeness Centrality of urban street network: evaluating the network characteristics by different normalization methods

5.1 The challenge

One of the first decisions in the process of street network analysis is that the boundary and the location of the center point of the catchment area needs to be defined at the early stage of the investigation. The challenge of this decision lies in the fact that the structure and attribute of the network change with the location of the catchment area. This change directly influences the number of nodes and the characteristics and topologies of network included in the selected catchment area. Subsequently, this change also influences the measurement of the indicators. In this research, the term *placement effect* refers to the phenomenon that the attributes and characteristics of the network depends on where the catchment area is retrieved.

One of the indicators that could be most influenced by the placement effect is Closeness Centrality, $C_c(v)$, which is the reciprocal of the sum of the shortest distance between the chosen node-*v*, and all other nodes in the catchment area.

$$C_c(\mathrm{v}) = 1 \,/ \sum_{u=1}^{N} S(v,u) \quad (1)$$

where $S(v, u)$ refers to the length of the shortest distance between the chosen node-*v* and other nodes, *u*, and *N* refers to the total number of nodes in the chosen catchment area.

Closeness Centrality measures how fast a node exerts influence on all other nodes. For example, if the target is to spread the information in the network, a node with strong Closeness Centrality means that it is in a position to spread information quickly. Nodes with higher value of Closeness Centrality can be important influencers in the network.

Depending on the location of the selected catchment area, the value of Closeness Centrality may change. A node in the center of the network has the advantage to have more influence on other nodes and has higher value of Closeness

Centrality than when it is on the borders of the network (Gil, 2017). The Closeness Centrality of a chosen node may not be necessarily small in the entire city street network, but it may be small in the selected catchment area only because it is not close to the center of the catchment area. In other words, the Closeness Centrality is not an absolute value and depends on where we place the center of the catchment area in the entire network.

The aim of this research is, firstly, to investigate how the attributes of the network and the indicator values of nodes change with the placement of the catchment area and, secondly, propose a method to choose the location of the catchment area if the Closeness Centrality of different nodes shall be compared.

5.2 Presenting the placement effect on Closeness Centrality

In order to demonstrate the placement effect, we use Plaza Luceros in Alicante as the example. In this research the size of the catchment area is 3000x3000m^2 and the Closeness Centrality has the unit of *1/km*. We choose the area size that is larger than the acceptable walking distance for the pedestrian because there will be more space to move the chosen node further away from the center in order to investigate the placement effect.

Another important consideration is the selection of network types. In this research the targeted network type is pedestrian network. Therefore, the criteria of the selection should reflect the perception of the pedestrian. For example, from the perception of the pedestrians, a street with two sidewalks is perceived as one link and will only be perceived as two links if the street is very wide. In order to produce a "pedestrian-like" network, the investigations of the authors lead to the following proposal of the highway tags of the Open Street Network (OSM) are selected: primary, secondary, tertiary, residential, pedestrian, steps, path and unclassified.

Eight catchment areas with different centers were selected and presented in Table 5.1. The center of the catchment area is indicated by a blue center point. Since we are interested in the placement effect on the Closeness Centrality of the nodes, we only focus on one chosen node, which is indicated by the red node in each catchment area[11].

[11] This red node is also the node that is closest to Plaza Luceros.

These catchment areas differ in the distance between the blue center point and the red chosen node. In catchment area 1, the blue center point and the red chosen node are identical with each other. Starting from catchment area 2 to catchment area 8, the blue center point gradually moves 50m, 100m, 200m, 300m, 500m, 1000m, 1500m towards the north of the red chosen node. Taking catchment area 2 as an example, the distance between the blue center point and the red node chosen node is 50m.

The results in Table 5.1 show that the Closeness Centrality of the red chosen node changes from 0.000678 *1/km* in catchment area 1 to 0.000438 *1/km* in catchment area 8. This change shows that the Closeness Centrality of the red chosen node is affected by its distance to the blue center point in all eight catchment areas. Hence, there is a placement effect on the value of the Closeness Centrality of the red chosen node.

5.3 Normalization of the Closeness Centrality

In order to evaluate and compare the characteristics of the networks, an indicator, ***Normalized Closeness Centrality***, $\boldsymbol{C_N}$, is created so that different number of nodes in different catchment areas will be balanced with the normalization. A common standard of Normalized Closeness Centrality is the multiplication of Closeness Centrality of the chosen node with number of nodes. By normalizing Closeness Centrality, number of nodes is now connected with the Closeness Centrality and, therefore, the fact that number of nodes changes with the location of the catchment area is now taken into the consideration. In the following, we use the results in Table 5.1 to explain the process of normalization of Closeness Centrality.

To normalize the Closeness Centrality, the multiplication with the *total number of nodes* minus 1, i.e. *N-1*, is often recommended in literature[12]. When the network is big enough it is not influenced by the “-1” and it can be neglected. Therefore,

$$C_N(v) = N \times C_c(v) \tag{2}$$

[12] It means that all the nodes except the chosen node are counted, which is precisely the definition of centrality.

where $C_N(v)$ refers to Normalized Closeness Centrality of node-v, N refers to total number of nodes in the catchment area, $C_c(v)$ refers to Closeness Centrality of the chosen node-v.

Take catchment area 1 in Table 5.1 as an example, the total number of nodes, N, is 1606 nodes and the Closeness Centrality of the red chosen node, $C(v)$, is 0.0006784 *1/km*. The Normalized Closeness Centrality of the red center node, $C_N(v)$, would be 1606×0.006784 = 1.0895 *1/km*.

The procedure seems to be a bit arbitrarily but has a deeper sense in the background. Firstly, we know that:

Step 1 $\quad N = 1606$ nodes

Step 2 $\quad C_c(v) = 0.0006784\ 1/km$

Since Closeness Centrality, $C_c(v)$, is the reciprocal of the sum of the shortest distance between the chosen node and all other nodes in the catchment area, we know that

$$\text{Step 3} \qquad \sum_{u=1}^{N} S(v,\ u) \ = \ \frac{1}{C_c(v)} \ = \ \frac{1}{0.0006784} \ = \ 1474.06\ \text{km} \qquad (3)$$

where $S(v, u)$ refers the sum of the length of the shortest paths between the chosen node-v and all other nodes, N refers to the number of nodes in the chosen catchment area.

If we divide $\sum_{u=1}^{N} S(v,u)$ by number of nodes, N, the result is equivalent to the average of the shortest distances from each node to the chosen node-v,

$$\text{Step 4} \qquad \sum_{u=1}^{N} S(v,u)\ \ /\ N \ = \ 1474.06\ \text{km} \ /\ 1606 \ = \ 0.917\ \text{km} \qquad (4)$$

In order to normalize different number of nodes in different catchment areas, the number of nodes, N, is divided by the sum of the shortest distance between the chosen node-v, and all other nodes.

$$\text{Step 5} \qquad N/\sum_{u=1}^{N} S(v,u) \ = 1606/1474.06km = 1.0895\ 1/km \qquad (5)$$

We found that the value in step 5 is equivalent to multiplying Closeness Centrality, $C_c(v)$, and number of nodes, N, in step 6. Therefore, this process is referred to as normalization of Closeness Centrality. By normalizing Closeness Centrality, number of nodes is now connected with the Closeness Centrality and, therefore, the fact that number of nodes changes with the placement of the catchment area is now taken into the consideration.

Step 6 $$C_N(v) = N \times C_c(v) = 0.006784\ 1/km \times 1606 = 1.0895\ 1/km \quad (6)$$

Furthermore, the reciprocal of Normalized Closeness Centrality is equivalent to the reciprocal of average of the shortest distances between the chosen node-*v*, and all other nodes, therefore

Step 7 $$\frac{1}{C_N(v)} = \frac{1}{1.0895}\ 1/km = \sum_{u=1}^{N} S(v,u)/\,N = 0.917km \quad (7)$$

Although the value of 0.917km, i.e. the reciprocal of the Normalized Closeness Centrality in step 7, is the same as the average of the shortest distances between the chosen node and all other nodes in step 4, it explains the meaning of the normalization.

Table 5.1 Visualization the network catchment areas with different center nodes (blue node) and the values of the indicators of the chosen node under investigation (red node).

Indicators		Unit	Catchment area 1	Catchment area 2
				
Distance between red chosen node and blue center node		m	0	50
Closeness Centrality of the red chosen node	$C_c(v)$	*1/km*	0.000678	0.000674
Average distance between all nodes to chosen node	$\sum_{u=1}^{N} S(v,u)/N$	m	917	906
Total number of nodes	N		1606	1637
Normalized Closeness Centrality	$C_N(v)$	*1/km*	1.0895	1.103

Table 5.1 (continued) Visualization the network catchment areas with different center nodes (blue node) and the values of the indicators of the chosen node under investigation (red node).

Indicators		Unit	Catchment area 3	Catchment area 4
				
Distance between red chosen node and blue center node		m	100	200
Closeness Centrality of the red chosen node	$C_c(v)$	*1/km*	0.000668	0.000649
Average distance between all nodes to chosen node	$\sum_{u=1}^{N} S(v,u)/N$	m	899	891
Total number of nodes	N		1665	1728
Normalized Closeness Centrality	$C_N(v)$	*1/km*	1.112	1.12

Table 5.1 (continued) Visualization the network catchment areas with different center nodes (blue node) and the values of the indicators of the chosen node under investigation (red node).

Indicators		Unit	Catchment area 5	Catchment area 6
				
Distance between red chosen node and blue center node		m	300	500
Closeness Centrality of the red chosen node	$C_c(v)$	*1/km*	0.000635	0.000611
Average distance between all nodes to chosen node	$\sum_{u=1}^{N} S(v,u)/N$	m	874	854
Total number of nodes	N		1802	1916
Normalized Closeness Centrality	$C_N(v)$	*1/km*	1.145	1.170

Table 5.1 (continued) Visualization the network catchment areas with different center nodes (blue node) and the values of the indicators of the chosen node under investigation (red node).

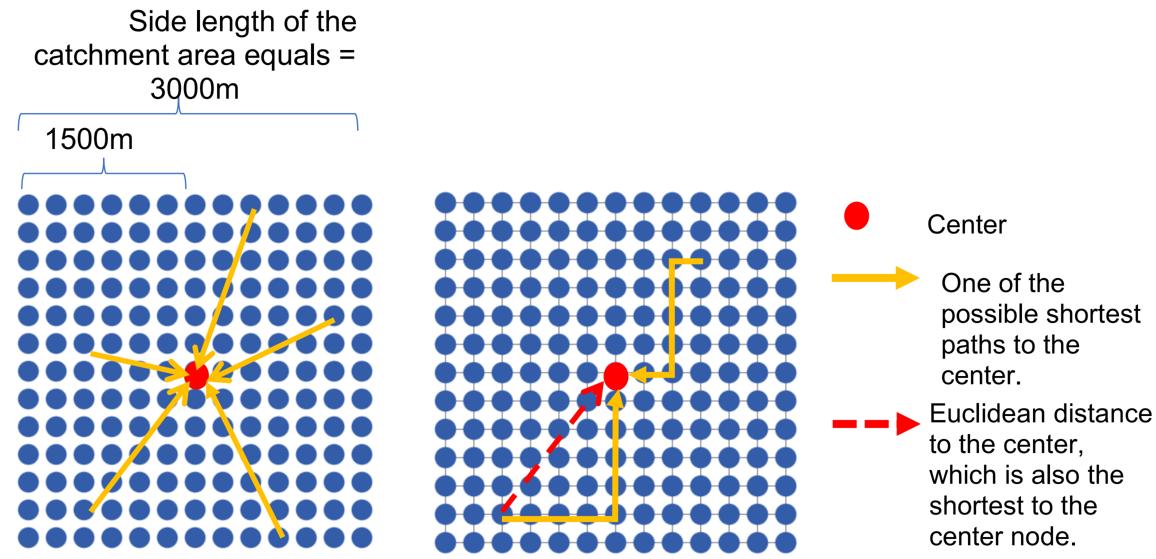
Indicators		Unit	Catchment area 7	Catchment area 8
				
Distance between red chosen node and blue center node		m	1000	1500
Closeness Centrality of the red chosen node	$C_c(v)$	*1/km*	0.000528	0.000438
Average distance between all nodes to chosen node	$\sum_{u=1}^{N} S(v,u)/N$	m	918	1190
Total number of nodes	N		2062	1919
Normalized Closeness Centrality	$C_N(v)$	*1/km*	1.088	0.840

5.4 Using two types of idealized network as the reference for comparison

Knowing the value of the Normalized Closeness Centrality of the chosen node-*v*, $C_N(v)$, in the real network, we need to have a reference for different networks to be compared with in order to evaluate this value. In this research we propose to use an “idealized” network to be the reference.

For the definition of an idealized network, we propose to use the network that is mathematically ideal and where the nodes are evenly distributed in a quadratic grid. Because the idealized network is ideal in the sense of mathematics, it can be calculated more easily than the real networks. In fact, this is the type of network that can also be frequently observed in reality and is in that sense more than a theoretical ideal case.

The idealized network can be further divided into two types, the star-like pattern (Figure 5.1 (a)) and the quadratic pattern (Figure 5.1 (b)), depending on how the chosen node-*v* is connected with all other nodes. In other words, the idealized network is differentiated depending on the type of distance. The first type of the idealized network uses Euclidean Distance, and the second type can be referred to as Network Distance [13] (Levinson & El-Geneidy, 2007).



(a) Star-like pattern **(b) Quadratic pattern**

Figure 5.1 Two types of idealized network.

In this research the center node of network is the chosen node-*v* (which we explore), and we are checking how its Closeness Centrality changes with 1) different number of nodes in the catchment area, i.e. different density of nodes; and 2) the corresponding sum of the shortest distance between the center node-*v* and all other nodes.

[13]Comparing with the Euclidean Distance, which is indicated by the red dash arrow in Figure 5.1(b), the Network Distance is by far not the shortest distance between any node to the center node. But for the sum of all shortest paths between all nodes it might present an optimal version because, due to the multiple alternative shortest connection between two nodes shown in Figure 5.7, it is very robust in case of a failing link.

TYPE 1 – the star-like pattern

Figure 5.1(a) presents the first type of idealized network. In this type of idealized network, there is a direct connection between the center node and all other nodes. The nodes and links in such a network form a star-like pattern. The shortest distance between any node to the center node is indicated by the yellow arrow.

Such a network is only favorable to one node, which is the center node in this case. For all other nodes, such a network is not ideal because the sum of the shortest distance between the center node and all other nodes is smaller than the sum of the shortest distance between all nodes and any of the non-center node.

The sum of the shortest distance between the center node and all other nodes can serve as a reference for comparison. We can compare this value in the idealized network with that in the real network to assess to what extend the nodes in the real network are evenly distributed or grouped in several clusters.

Table 5.2 Relationship between indicator values and network pattern

		Value of the indicators	
		Sum of the length of the shortest paths between the center node and all other nodes	**Closeness Centrality**
Topology of real network	**Close to evenly distributed nodes**	Real network >= idealized network type 1	Real network <= idealized network type 1
	Many nodes concentrated around center nodes	Real network < idealized network type 1	Real network > idealized network type 1

If the nodes in the real network are not evenly distributed, it is expected that the sum of the shortest distance between the center node and all other nodes in the real network would be larger than that in the idealized network. This also means that the Closeness Centrality of center node in a real network would not be larger than the Closeness Centrality of the center node in the idealized network.

On the contrary, if the sum of the shortest distances between the center node and all other nodes in the real network is smaller than that in an idealized network and, therefore, the Closeness Centrality of center node is bigger than the Close Centrality of the center node in the idealized network, we might expect to observe that that the nodes concentrate and form clusters around the central node in the real network. Table 5.2 summarizes the relationship between indicator values and network pattern.

TYPE 2 – the quadratic pattern

Figure 5.1 (b) presents the second type of idealized network. Like type 1 network, the nodes in type 2 network are also evenly distributed but they are connected in a different way. Each node has 4 links to its direct neighbors, and all links and nodes form a quadratic pattern. The shortest distance between any node to the center node is indicated by yellow arrows in Figure 5.1(b). There are multiple alternative and equivalent shortest connections between two nodes, which makes this type of network very robust in case of a failing link.

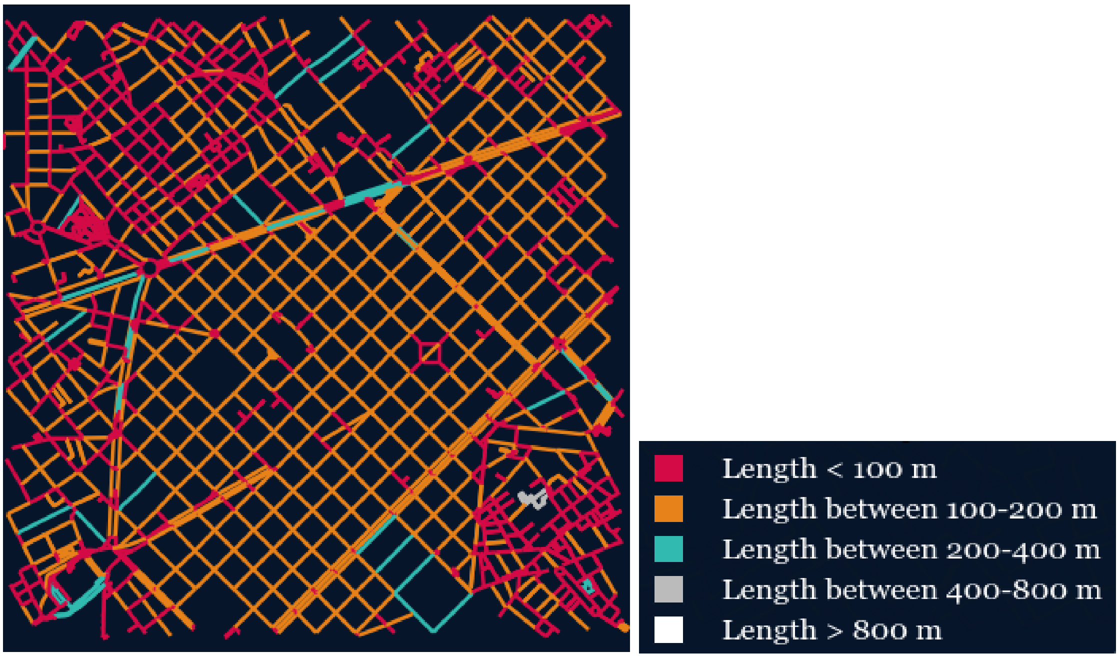


Figure 5.2 Quadratic pattern of street network in Barcelona

Since all nodes are integrated in the network in the same way, any hierarchy or difference between the nodes can be avoided. This type of network is often found in real urban networks and it corresponds to the structure of classical street block, such as the quadratic pattern of street network in the city Barcelona shown in Figure 5.2.

5.5 Closeness Centrality in type 1 of idealized network

5.5.1 Normalization procedures using type 1 of idealized network as the reference

To proceed with the normalization of Closeness Centrality, we first need to calculate the **average distance of all nodes to the center node**. Although the side length of the catchment area, which is indicated by the black rectangular in Figure 5.3, is 3000m, for the reason of symmetry it is sufficient to focus on just one quadrant of the catchment area, which is indicated by the red rectangular in Figure 5.3, with the edge length of 1500m.

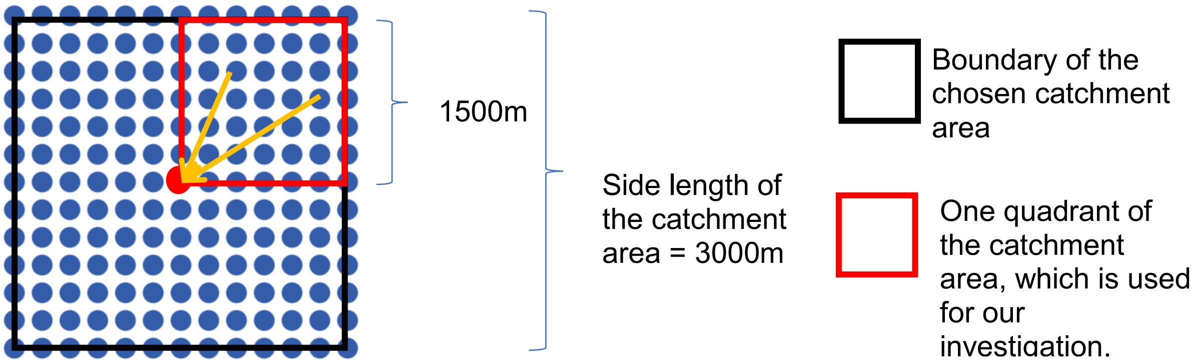


Figure 5.3 Relationship between the catchment area and one quadrat of the catchment area.

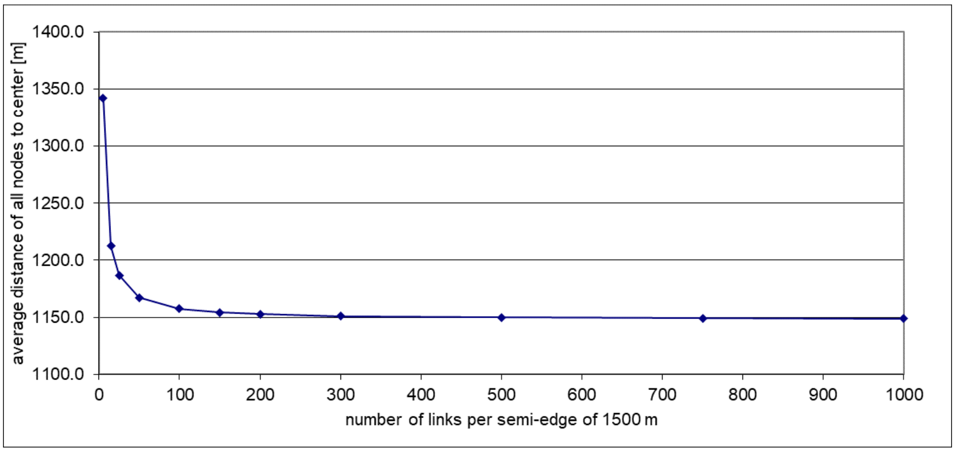


Figure 5.4 The relationship between the number of links on each side of the 1500m×1500m quadrant of the catchment area and the average distance between the red center node and all other blue nodes.

Figure 5.4 shows the relationship between the number of links on each side of the 1500m×1500m quadrant and the average distance between the red center node and all other blue nodes in type 1 of idealized network.

Table 5.3 The average distance between the red center node and all the blue nodes, with an increasing number of links on each edge.

Number of nodes on each side of the quadrant	Average distance between the red center node and all the blue nodes (m), $\sum_{u=1}^{N} S(v,u)/N$	Sum of shortest possible distance between each node and the central point (m), $\sum_{u=1}^{N} S(v,u)$	Total number of nodes, N	**Normalized Closeness Centrality, $C_N(v)$**
5	1342.176	33554	25	**0.000745**
15	1212.577	272830	225	**0.000825**
25	1186.662	741664	625	**0.000843**
50	1167.228	2918069	2500	**0.000857**
100	1157.511	11575105	10000	**0.000864**
150	1154.272	25971110	22500	**0.000866**
200	1152.652	46106082	40000	**0.000868**
300	1151.033	103592930	90000	**0.000869**
500	1149.737	287434240	250000	**0.000870**
750	1149.089	646362654	562500	**0.000870**
1000	1148.765	1148765266	1000000	**0.000871**

Figure 5.4 shows the relationship between the number of links on each side of the 1500m×1500m quadrant and the average distance between the red center node and all other blue nodes in type 1 of idealized network.

Table 5.3 presents the average distance between the red center node and all the blue nodes, with an increasing number of links on each side of the quadrant. As the number of links on each edge increases, the node density of the network also increases. The results show that, for an increasing number of links on the side of the quadrant, the average distance between the red center node and all the blue nodes, i.e.

$$\sum_{u=1}^{N} S(v,u)\,/N$$

, decreases and reaches a saturation at 1148m[14].

[14] This value holds only for a section 3000x3000m^2. For all other sizes of a catchment area, the distribution pattern, i.e. the behavior, of the average distance between the red center node and all the blue nodes will be equivalent to the blue line in Figure 5.4 but the final value will be different from 1148 m.

This means that, if this idealized network had 1606 nodes, which is also the number of nodes of the real network in catchment area 1 in Table 5.1, in theory we should receive the following results

$$\sum_{u=1}^{N} S(v,u)/N = 1148m \times N = 1148 \times \frac{1606}{1000} = 1844\ km$$

$$C_c(v) = 1/1844 = 0.000542\ 1/km$$

$$C_N(v) = C_c\,(v) \times N = 0.000542\ 1/km \times 1606 = 0.871\ 1/km$$

$$1/\sum_{u=1}^{N} S(v,u) = \frac{1}{1.148km} = 0.871 \quad 1/km$$

This Normalized Closeness Centrality, $C_N(v)$, of an idealized network type 1 with 1606 nodes, which is 0.871 *1/km*, can now be compared directly with that of the real network with 1606 nodes, which is 1.0895 *1/km*.

5.5.2 Comparing Normalized Closeness Centrality of the real network with that of type 1 of idealized network

As Figure 5.4 shows the relationship between the number of links on each side of the 1500m×1500m quadrant and the average distance between the red center node and all other blue nodes in type 1 of idealized network.

Table 5.3 presents, under the precondition that the nodes are distributed evenly, the Normalized Closeness Centrality of the center node in the real network cannot be greater than that in the idealized network. In other words, if the 1606 nodes in the real network were distributed evenly, the Normalized Closeness Centrality of the center node is expected not be greater than 0.871 *1/km*.

However, the value of the Normalized Closeness Centrality of the real network in catchment area 1 of Table 5.1 is 1.0895 *1/km*, which is greater than 0.871 *1/km*. This means the 1606 nodes are not evenly distributed in the real network and the majority of the nodes are closer to the central node, leading to a smaller sum of shortest distance and thus to a higher Closeness Centrality.

The situation that nodes are clustering around the center nodes of the catchment area could be commonly observed in the scenarios when areas like parks, water bodies, or the newly developed area in the outskirt of urban area are included in the boundary of the catchment area. This means that, in some places of the chosen catchment area, there are no or only very few nodes. If these places occupy large part of the catchment area, the value of the Closeness Centrality can be greater than the reference value of the idealized network.

For example, as shown in the selected catchment area presented in Figure 5.5, there is a large proportion of water body (port) in the bottom-right corner, which is far from the center node at Plaza Luceros. The values of Closeness Centrality of the red center node in all eight catchment areas in Table 5.1 are influenced by the proportion of this water body as the blue center point of the catchment area moves more and more to the north.

Figure 5.5 A catchment area of 3000x3000m2 with Plaza Luceros as the center node. There is a big water body (port) at the bottom-right corner and it is far from the center node at Plaza Luceros.

In a hypothetically idealized network, the nodes should have been evenly distributed and cover the area where the water body is. However, those nodes that should have covered the sea are now missing and there are many nodes that are near to the red center node.

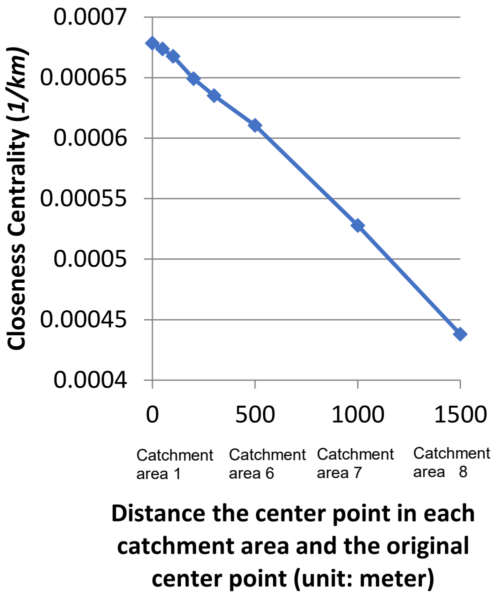
It is in such case that the Normalized Closeness Centrality is expressing something about the characteristics of the network. Because Normalized Closeness Centrality indicates whether the nodes are evenly distributed, more concentrated, or there are some “white” spots, such as water body, in the catchment area. Further explanation is provided in the next catchment area.

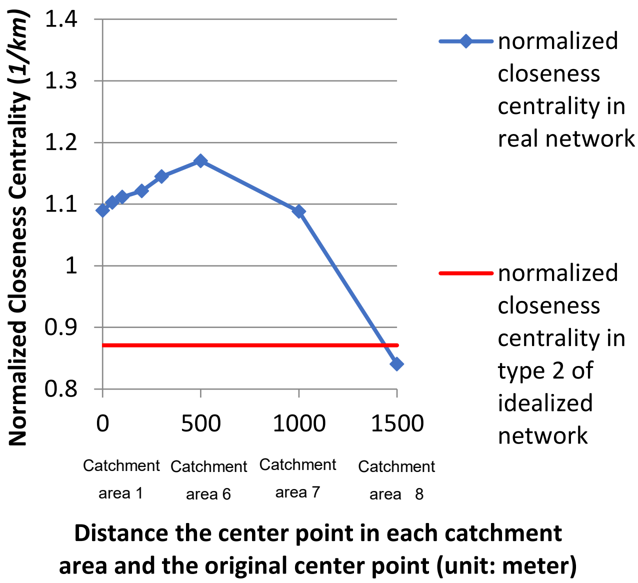
5.5.3 Comparing Closeness Centrality and Normalized Closeness Centrality of the real network

How can the Normalized Closeness Centrality help to minimize the placement effect? To answer this question Figure 5.6(a) presents the Closeness Centrality of the red center node in the eight catchment areas of the real network in Table 5.1. In Figure 5.6(b), the blue line represents the Normalized Closeness Centrality of the red center node in the eight catchment areas of the real network and the red line represents the Normalized Closeness Centrality of the center node in type 1 of idealized network. To be more specific, both blue lines in Figure 5.6(a) and Figure 5.6(b) belong to the real network. The red line in Figure 5.6(b) is the Normalized Closeness Centrality of the idealize network type 1 and serves as a reference for comparison.

As shown in Figure 5.6(a), the Closeness Centrality of the red chosen node is not normalized by the number of nodes and the value reduces as the catchment area moves from catchment area 1 to catchment area 8 in Table 5.1. There can be different explanations of this observation.

- Explanation 1**:** As the catchment area moves, the Closeness Centrality of the red center node really becomes less. For example, the catchment area might include the less connected network due to the defects, such as bad spots or missing nodes in the network, which cause detours and increase the length of the shortest paths.
- Explanation 2: As the catchment area moves, the nodes become less evenly distributed and the majority of the nodes are further and further away from the red center node. That increases the sum of the shortest paths and reduces the Closeness Centrality, even if the network might be very well-connected.
- Explanation 3**:** As the catchment area moves, the number of nodes increase, which leads to a higher node density, and thus a larger length of the shortest paths. This leads to a smaller value of Closeness Centrality because Closeness Centrality is the reciprocal of the sum of the shortest distance. As mentioned before, one of the contributions of normalizing Closeness Centrality is that the number of nodes can be included into the consideration and explanation 3 is more relevant to the Normalized Closeness Centrality.





(a) Closeness Centrality of the red center node in the eight catchment areas of real network in Table 5.1.

(b) Normalized Closeness Centrality of the red center node in the eight catchment areas of real network in table 5.1 (blue line) and in idealized quadrate network type 1(red line).

Figure 5.6 Comparing Closeness Centrality and Normalized Closeness Centrality of the real network.

Take catchment areas 6, 7, and 8 in the Table 5.1 as the examples. There is a higher number of nodes (1916, 2062 and 1919 nodes) in catchment areas 6, 7, and 8 in comparison with 1606 nodes in catchment area 1. And comparing with catchment area 1, the Closeness Centrality of catchment areas 6, 7, and 8 are lower. In other words, the increasing number of nodes in the catchment areas 6, 7, and 8 leads to a decreasing Closeness Centrality.

Why do we observe a decreased Closeness Centrality as the number of nodes increases? In order to examine whether explanation 3 can be applied to explain the observed results, it is better to use Normalized Closeness Centrality as the indicator to investigate the relationship. By multiplying the Closeness Centrality with number of nodes, explanation 3 is now widely balanced and connected with number of nodes after the normalization. And the results are presented by the blue line in Figure 5.6(b). The curvy blue line shows the effect of normalization. The Normalized Closeness Centrality of catchment area 1 is now smaller than that of the catchment area 6.

Therefore, comparing with Closeness Centrality, the normalized values provide additional information about the characteristics of the network, such as how well it is connected and how even the nodes are distributed.

In addition, as shown in Figure 5.6(b), the Normalized Closeness Centrality of catchment area 8, whose center point is 1500m away from the chose node, is below the red line, which represents the Normalized Closeness Centrality of the idealized network. This is where explanation 1 can help. Because now the area at the outer city is included in catchment area 8 and there is a smaller number of nodes in this area.

5.6 Closeness centrality in type 2 of the idealized network

5.6.1 Normalization procedures using type 2 of the idealized network as the reference

Further normalization procedures could be created for the type 2 of the idealized network. Similar to type 1, type 2 of the idealized network is also under the precondition that the nodes are evenly distributed. But they differ in how the nodes are connected. In other words, the types of distance are different in these two types of the idealized network. Type 1 is an idealized network where there is a direct and straight connection from all nodes to the central one without any detour. Type 2 is shown in Figure 5.7, where the blue arrows show the path from a random blue node to the red center node. As the blue arrows show, the movement corresponds to how a rook(castle) moves on a chessboard.

For a rook on a chessboard, there can be many shortest paths for moving from a random blue node to the center node (as indicated by the blue arrows in Figure 5.7) and the length of all paths is identical. Thus, the network is quite robust against failure and disturbances. For example, if there is a missing link, there are alternative shortest paths with the same length. In that sense, such a regular network has high robustness which had to be reached by real ones. Indeed, this type of rectangular, quadratic network could be a better, more realistic assumption than type 1 of the idealized network.

In order explain the normalization procedure, we again focus on the red center node in type 2 of the idealized network in Figure 5.7 and use it as the chosen node under investigation. If we multiply the Closeness Centrality of the red center node in the real network by the sum of the shortest distances between the chosen node and all the other nodes in the idealized network type 2, this procedure is equivalent to a normalization and the value should be around one.

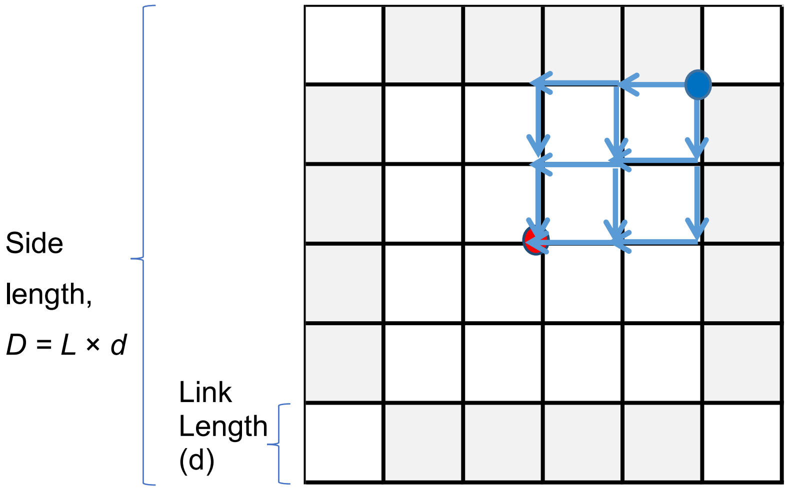


Figure 5.7 Type 2 of idealized network. The blue arrows show the movement of a castle on the chess board and also present the shortest path from a random blue node to red center node.

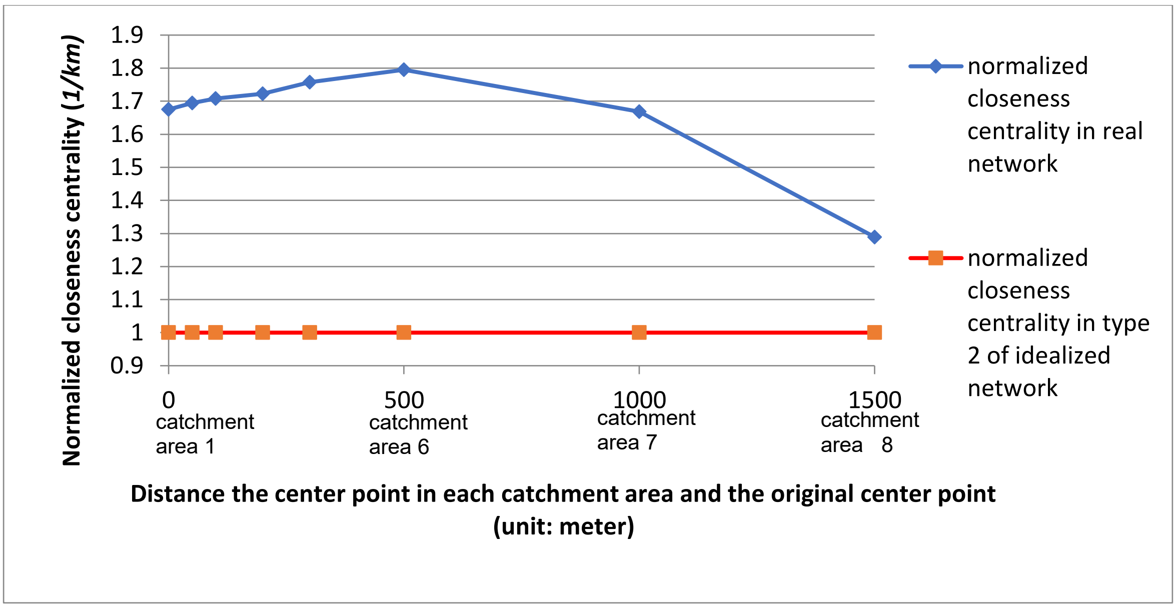


Figure 5.8 Normalized Closeness Centrality in the real network in the 8 catchment areas (blue line) and the idealized quadrate network type 2(red line).

The results of the idealized network type 2 are shown in Figure 5.8. Both blue lines in Figure 5.6(b) and Figure 5.8 show the same pattern and behavior but the value of the red line in Figure 5.6(b) is 0.817 *1/km*, whereas the value of the red line in Figure 5.8 is 1, which serves as a reference for the following comparison.

- Values above 1 indicate a well-connected network with shortest paths. This better-connected network could be realized if the diagonal links exist in addition to the horizontal and vertical links. This would (partly) allow the movement like the queen on a chessboard, leading to shortening of the shortest connections.
- Values below 1 indicate a weaker Closeness Centrality of the red center node in real network than that in the idealized network type 2. A possible explanation can be that there are missing links or there are very few or even no diagonal links.

In order to calculate the sum of the shortest distances between the chosen node and all the other nodes, it is helpful to have some further definitions about sides length and links. As shown in Figure 5.7,

- A catchment area of the network is surrounded by four sides.
- Each side consists of a certain number of links. Therefore,

$$D = L \times d \tag{8}$$

where *D* refers to the side length, *L* refers to the number of links on each side of the catchment area, *and d* refers to length of a link.

For example, catchment area 1 in Table 5.4 shows a catchment area with each side consisting of two links, i.e. *L* = 2 and *D* = 2*d*, and catchment area 2 shows a catchment area with each side consisting of 4 links, i.e. *L* = 4, and, therefore, *D* = 4*d*. The general formulas for the castle-pattern network are:

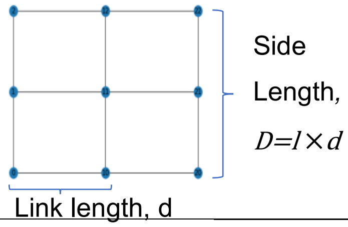
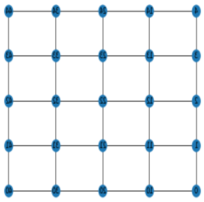
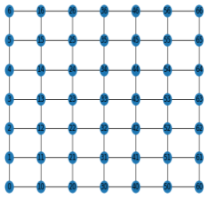
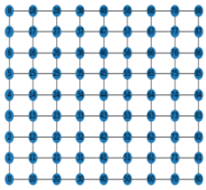
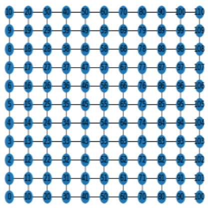
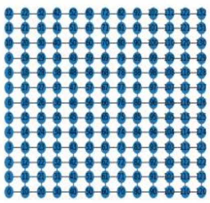
$$\sum_{u=1}^{N} S(v,u) = [(4 \times (L+1)) \times (\frac{L}{2}+1) \times (\frac{L}{2})/2] \times (d) \tag{9}$$

$$\text{Total number of links} = (L) \times (L+1) \times 2 \tag{10}$$

$$N = (L+1)\ ^\wedge\ 2 \tag{11}$$

where $\sum_{u=1}^{N} S(v,u)$ refers to the sum of shortest distance between all nodes and chosen node-*v*, *L* refers to the number of links in on the edge or on each side of the catchment area, *d* refers to the length of one link and *N* refers to the total number of nodes.

Table 5.4 The edge length, D = L ×d, the shortest distance between all nodes and chosen node, and the total number of links in catchment areas with different number of nodes, N.

		Edge length, $D = L \times d$	Sum of shortest distance between all nodes and chosen node $\sum_{u=1}^{N} S(v,u)$	Total number of links	Total number of nodes, N
Catchment area 1		2d	4×d +4×2d	12	9
Catchment area 2		4d	4×d + 8×2d + 8×3d + 4×4d	40	25
Catchment area 3		6d	4×d + 8×2d + 12×3d + 12×4d + 8×5d + 4×6d	84	49
Catchment area 4		8d	4×d + 8×2d + 12×3d + 16×4d +16×5d + 12×6d + 8×7d + 4×8d	144	81
Catchment area 5		10d	4×d + 8×2d + 12×3d+16×4d + 20×5d+20×6d + 16×7d+12×8d + 8×9d + 4×10d	220	121
Catchment area 6		12d	4×d+8×2d+ 12×3d+16×4d + 20×5d+ 24×6d + 24×7d+20×8d +16×9d+12×10d +8×11d + 4×12d	312	144

5.6.2 Comparing Normalized Closeness Centrality of the real network with that of idealized network type 2

Take catchment area 1 of Table 5.1 as an example again. We know that the number of nodes is 1606 and Closeness Centrality is 0.000678 *1/km*. Hypothetically, if these

1606 nodes are evenly distributed like the type 2 of idealized network, using the formula (1) to (3) we can calculate that:

$Total\ number\ of\ nodes,\ N = (L + 1)\ ^\wedge\ 2 = 1606\ nodes$

$Side\ length,\ D = 39.0749d = 3000m\ (i.e.\ 39.1\ links\ on\ each\ side\ of\ the\ catchment\ area)$

$L = 39.0749$

$d = 3000m\ /\ 39.0749 = 76.7756m$

$Number\ of\ nodes\ in\ each\ edge = L + 1 = 40.0749\ nodes\ per\ edge$

$Total\ number\ of\ links = (L) \times (L + 1) \times 2 = 3131.8501$

$$\sum_{u=1}^{N} S(v,u) = [(4 \times (L+1)) \times (\frac{L}{2} + 1) \times (\frac{L}{2})/2] \times (d)$$

$$= [(4 \times (39.0749 + 1)) \times ((39.0749\ /\ 2) + 1) \times (39.0749\ /\ 2)\ /\ 2] \times (76.7756)$$

$$= 2469.053\ km$$

$$Closeness\ Centrality = C_c(v) = 1/\sum_{u=1}^{N} S(v,u) = 0.000405\ 1/km$$

Obviously, this Closeness Centrality of type 2 of idealized network, which is 0.000405 *1/km*, is different from the Closeness Centrality of the real network, which is 0.000678 *1/km*. But we can use the former as a reference. If we divide the later with the former, the normalization of the Closeness Centrality of the real network results in:

$$C_c(v)\ of\ real\ network\ /\ C_c(v)\ of\ type\ 2\ of\ idealized\ network = 0.000678\ /\ 0.000405 = 1.67$$

This result is the same as multiplying the sum of shortest distance between all nodes and the chosen node in type 2 of ideal network with Closeness Centrality of the real network and this procedure is equivalent to a normalization procedure.

Applying the same procedure to all eight catchment areas in Table 5.1, the results in Figure 5.8 shows the same pattern like the Normalized Closeness Centrality of type 1 of idealized network in Figure 5.6(b). However, Figure 5.6(b) and Figure 5.8 provide different information. Under the assumption that in the real network the nodes would be distributed also evenly, the result in Figure 5.8 can be interpreted that the Closeness Centrality of the real network is 1.675 times higher than that of type 2 of idealized network with the same number of nodes (1606 nodes). This might indicate that the real network contains more diagonal links than the idealized type 2 and, therefore, is a more connected network than the idealized network type 2.

Finally, it should be noted that the obvious disadvantage of the normalization methods is that the calculation of the sum of shortest distance is only correct for the center point and not for all other nodes. However, the calculation allows the comparison of the Closeness Centrality between the idealized network and the real network.

5.7 Conclusions and discussion

For the assessment of an urban network the location of the center point and a catchment area around it need to be defined. The location of the catchment area, which is a sample of the entire network, exerts significant influence on the value of the indicator and in this research these influences are referred to as the placement effect in the current research. This effect becomes even more significant when multiple catchment areas are sampled to be compared and classified.

This chapter examines one of the most affected indicators, Closeness Centrality. The Closeness Centrality of a chosen node changes remarkably depending on its position in the catchment area. If the value of Closeness Centrality of a chosen node has the highest value among all nodes, we cannot be sure whether it is because this chosen node happens to be placed in the center of the catchment area or it is because it really is the most important node in the entire street network.

In this research we propose that this problem can be balanced to a certain degree by a normalization process because the normalization takes the number of nodes into consideration. In addition, by using normalizations, additional information about the characteristics of the network, such as network pattern, can be acquired. We propose to use two kinds of Normalized Closeness Centrality in different types idealized networks as the references for the evaluation of the characteristics of the real network. In short, by comparing Normalized Closeness Centrality in the real network with that in the idealized network, we can 1) evaluate the placement effect on Closeness Centrality and decide whether the real network is better or worse connected than the idealized ones and 2) explore to what extend the nodes in the real network are evenly distributed.

The results show that the Closeness Centrality of the same node varies remarkably depending on its position and how central it is in the chosen catchment area. In other words, the Closeness Centrality of a chosen node may be not necessarily small in the entire city street network, but it may be small in the selected catchment area just because it is not close to the center of the catchment area.

Furthermore, we can use the Normalized Closeness Centrality of the center node in catchment area 1 as a reference and plot the deviation of the Normalized Closeness Centrality of the other 7 catchment areas. The results in Figure 5.9 show that, if the red node is placed within 100m of the original place, the deviation is less than 2%[15], which is highlighted with a gray belt in Figure 5.9. If the center point of a catchment area is moved by more than 100m away from the original center point, the Closeness Centrality of the same node starts to be significantly influenced by the placement effect. As shown in Figure 5.9, the percentage of deviation starts to rise after 100m. This means the values of the indicators of different chosen node can only be compared directly, if they are not more than 100m away from each other. However, this hypothesis needs to be further investigated with other networks and catchment areas.

The implications are, firstly, as a consequence of this placement effect, a direct comparison of Closeness Centrality between different nodes in the same catchment area is only possible if these nodes are less than 100m away from each other. Secondly, if we want to compare the value of Closeness Centrality of two nodes that are more than 100m away from each other, we need to create two catchment areas with these two nodes being the center in each of these catchment areas.

As a consequence, a map like Figure 5.10 that is an output of software cannot be used to compare the absolute values of Closeness Centrality of different nodes in the entire street network at the global level. But such a map shows well the situation in a relative context at local level. It shows whether certain nodes in the catchment area are the highlight or not in the direct neighborhood.

[15] The researcher may decide what percentage is the acceptable range and, in this research, we decide that 2% is the acceptable deviation.

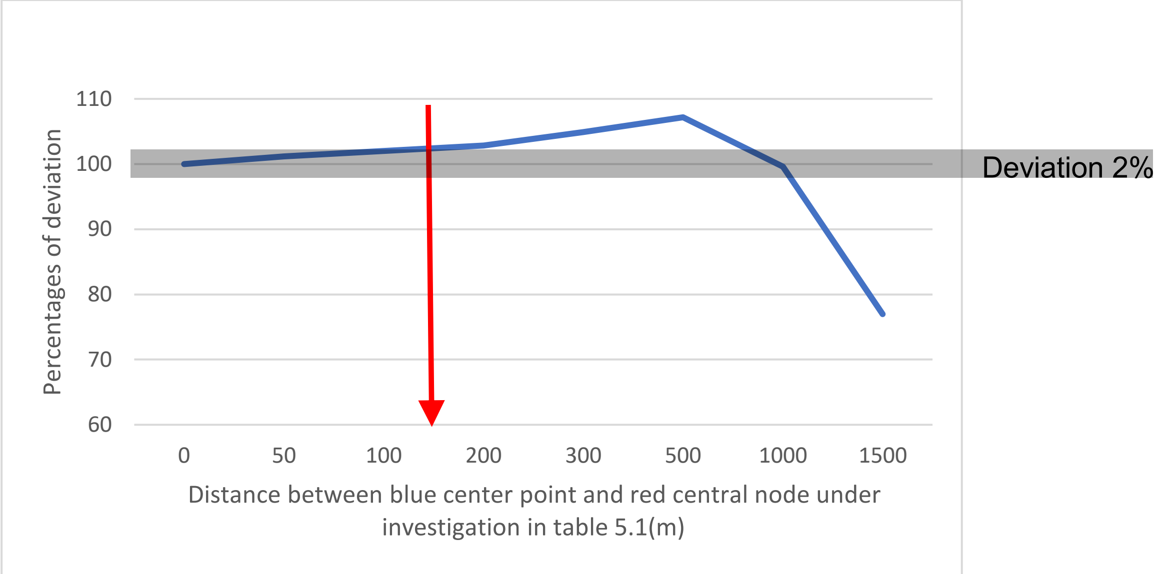


Figure 5.9 Relationship between the percentage of deviation and the distance between blue center point and red central node.

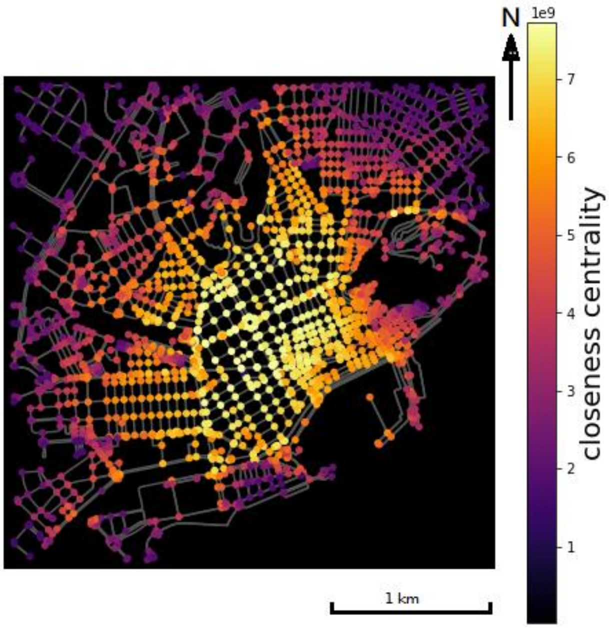


Figure 5.10 Nodes colored by Closeness Centrality.

6. Classifying stations by Place Value

In order to identify the extent to which a transit station can function as a transit-oriented development (TOD) neighborhoods, the stations have been clustered by their Node Values in chapter 3. The main aim of the current chapter is to further classify the stations in each cluster by their Place Values. The other target of this chapter is to select stations that can serve as examples of each sub-cluster so that in the next chapter we can zoom in to the neighborhoods around the stations and scrutinize the detailed spatial distribution of the street network and urban elements.

6.1 Introduction

Urban spatial structure is very important in urban planning because it reflects both physical and dynamic contexts. It is defined as the relation and interaction of different urban elements (Dadashpoor & Yousefi, 2018) (Bourne, 1982), such as road networks (Spadon, Gimenes, & Rodrigues-Jr., 2017), transportation (Sasaki, 1989) (Sadayuki, 2018), distribution of buildings (Cao, Shi, & Liu, 2016), land-use patterns and economic performance (Zhu & Sun, 2017) (Zhang, Song, Nes, He, & Yin, 2019).

One of the benefits of examining the urban structure is that it allows gaining deeper insights into the evolution of the city. The fast and increasing global urbanization has resulted in the hitherto unprecedented urban growth and aggregation of people and buildings. The economic development and the increase in urban scale often leads to urban structure evolving from the monocentric to the polycentric model. Following the urban growth and the aggregation of people and buildings, balancing the development among different areas becomes a vital task for the policy makers and city planners. Accordingly, one of the purposes of introducing the concept of TOD is to mitigate and remedy the often-unbalanced development of transit and land use among the neighborhoods.

"TOD emphasizes the development and opportunities provided by public transportation" (Zhang, Song, Nes, He, & Yin, 2019). In addition, a walkable neighborhood can help to enhance the success rate of TOD because it facilitate direct and proximate access to and from the transit node (Jacobson & Forsyth, 2008) (Val, 2015) and, therefore, encourages greater use of the transit stop. It also provides the convenient and safe access to more essential destinations, amenities and services for meeting local resident's living, working and playing (Jeffrey, Boulangé, Giles-Corti, Washington, & Gunn, 2019). Therefore, a walkable neighborhood is often characterized by well-connected streets and high density of residential housing (Giles-Corti, Foster, Koohsari, Francis, & Hooper, 2015). By locating high residential density and a variety of local services, utilities, and employment around public transit stations, TOD intends to attract activities to neighborhoods around the stations in order to encourage walking and to contribute to a habitable community (Higgins & Kanaroglou, 2016) (Val, 2015) (Kamruzzaman, Baker, Washington, & Turrell, 2014).

The research on the evaluation of TOD has focused on individual transit nodes (Huang, Grigolon, Madureira, & Brussel, 2018) because the TOD plans and their implementation may vary depending on the conditions in the catchment areas[16] around different stations. Developing an approach for classifying the TOD neighborhoods is an important step for exploring the urban spatial structure and for cultivating a more profound understanding of the integrated relationships and connection among different urban elements. Classifying stations and the surrounding areas into categories is very useful for policymakers, planners and urban designers, as it allows them to simplify the complex characteristics of the areas in order to understand which neighborhoods have similar relevant properties and so to better assess their TOD potential (Reusser, Loukopoulos, Stauffacher, & Scholz, 2008) (Austin, et al., 2010) (Zemp, Stauffacher, Lang, & Scholz, 2011) (Lyu, Bertolini, & Pfeffer, 2016). The scientific assessment of the existing TOD conditions and the measurement of their heterogeneity help the policy maker and planner to uncover the potential of an area is to become transit-oriented and to explore the underlying conditions and reasons (Singh, Fard, Zuidgeest, Brussel, & van Maarseveen, 2014). Results of these analyses are crucial for facilitating the future TOD projects and enhancing the success rate of their implementation (Kamruzzaman, Baker, Washington, & Turrell, 2014).

This chapter aims at developing and exploring the station typology based on the objective features of street networks measured within close proximity to railway public transport (RPT) stations in metropolitan Hamburg, Germany. It is an attempt to refine and improve the measurement for investigating the complexity of street networks and

[16] In this chapter, the 'catchment area' is interchangeable with 'neighborhood' or 'surrounding area' around the station.

Points of Interest (POI) in the TOD neighborhoods and to identify successful TOD developments.

6.2 Measurements and methodologies

6.2.1 Measurements and indicators

The characteristics of TOD can be measured by the density of points of interest (POI), the diversity of POI, the destination accessibility, and the design of street network (Cervero & Kockelman, 1997).

Density is considered to be the one of the core concepts for describing the urban spatial structure (Krehl, Siedentop, Taubenböck, & Wurm, 2016) and it is also an important factor in forming hubs or subcenters in the city (Chen, Hui, Wu, Lang, & Li, 2019). Since the size of the catchment area is fixed for all the stations, density of POI is measured directly by the number of POI.

Diversity is measured by the types of POI in the current study. The types of POI are the POI tags of Open Street Map (OSM) (OpenStreetMap contributors, 2017). By analyzing different types of POI, one can distinguish the sub-centers of neighborhoods, the residential areas, the suburban areas, etc. (Kamruzzaman, Baker, Washington, & Turrell, 2014).

Destination accessibility is often associated with walkability (Gunn, et al., 2017) and with the concept of "local living" (Badlanda, et al., 2017). Facilitating the access to essential destinations allows the residents to spend more time within their immediate neighborhood. Living close to amenities that meet their daily needs is crucial in shifting travel behavior towards more active travel modes, maximizing the use of the TOD neighborhoods and alleviating the road congestion (Jeffrey, Boulang é, Giles-Corti, Washington, & Gunn, 2019). In the current research, accessibility is defined by the average number of POI within 750 m of each node.

Detailed categorization of the indicators is presented in Table 6.1. Definition of POI defined by the tags of Open Street Map are provided in Table 6.2.

Table 6.1 Indicators for measuring the degree of walkability focusing on the street network in the TOD neighborhood.

Category	Measure		Implications to walkability
Diversity	1) Number of POI categories		Bigger is better
Density	2) Number of POI		Bigger is better
Destination accessibility	3) Average number of POI within 750m of each node		Bigger is better
Design of street network	TOPOLOGICAL MEASURE		
	Connectivity	4) Number of nodes 5) Number of links 6) Average segment length 7) Average Node Degree	Bigger is better Bigger is better Smaller is better Bigger is better
	Clustering	8)Average Clustering Coefficient	Higher value indicates more clusters exist in the network.
	Urban spatial order	9) Entropy of street bearing	Street network that is more grid-like exhibits less entropy.
	CENTRALITY MEASURE		
	Centrality	10) Average Closeness Centrality	Higher value indicates close proximity from a node to all other reachable nodes along the shortest path in the network.
		11) Average Betweenness Centrality	High values indicate being more frequently traversed by a larger number of the shortest paths connecting all couples of nodes in the network.

Table 6.2 Tags of point of interest from Open Street Map.

Amenities (https://wiki.openstreetmap.org/wiki/Key:amenity)	**Sustenance**	bar, BBQ, biergarten, cafe, drinking_water, fast_food, food_court, ice_cream, pub, restaurant
	Education	college, driving_school, kindergarten, language_school, library, toy_library, music_school, school, university
	Transportation	bicycle_parking, bicycle_repair_station, bicycle_rental, boat_rental, boat_sharing, bus_station, car_rental, car_sharing, car_wash, vehicle_inspection, charging_station, ferry_terminal, fuel, grit_bin, motorcycle_parking, parking, parking_entrance, parking_space, taxi
	Financial:	atm, bank, bureau_de_change
	Healthcare:	baby_hatch, clinic, dentist, doctors, hospital, nursing_home, pharmacy, social_facility, veterinary
	Entertainment, Arts & Culture:	arts_centre, cinema, community_center, fountain, planetarium, public_bookcase, social_center, studio, theatre
	Other:	bench, childcare, conference_center, marketplace, place_of_worship, police, post_box, post_depot, post_office, recycling, townhall, vending_machine, waste_basket, waste_disposal, waste_transfer_station, watering_place, water_point
Leisure (https://wiki.openstreetmap.org/wiki/Key:leisure)		amusement_arcade, bandstand, beach_resort, bird_hide, bowling_alley, dance, disc_golf_course, dog_park, escape_game, firepit, fishing, fitness_centre, fitness_station, garden, golf_course, hackerspace, horse_riding, ice_rink, marina, miniature_golf, nature_reserve, outdoor_seating, park, picnic_table, pitch, playground, resort, sauna, slipway, sports_centre, stadium, summer_camp, swimming_area, swimming_pool, tanning_salon, track, trampoline_park, water_park, wildlife_hide

Table 6.2(continued) Tags of point of interest from Open Street Map

Shops (https://wiki.openstreetmap.org/wiki/Key:shop)	**Food, beverages:**	alcohol, bakery, beverages, brewing_supplies, butcher, cheese, chocolate, coffee, confectionery, convenience, deli, dairy, farm, frozen_food, greengrocer, health_food, ice_cream, organic, pasta, pastry, seafood, spices, tea, wine, water
	Clothing, shoes, accessories:	department_store, general, kiosk, mall, supermarket, wholesale, baby_goods, bag, boutique, clothes, fabric, fashion, fashion_accessories, jewelry, leather, sewing, shoes, tailor, watches, wool
	Discount store, charity:	charity, second_hand, variety_store
	Health and beauty:	beauty, chemist, cosmetics, drugstore, erotic, hairdresser, hairdresser_supply, hearing_aids, herbalist, massage, medical_supply, nutrition_supplements, optician, perfumery, tattoo
	Do-it-yourself, household, building materials, gardening	agrarian, appliance, bathroom_furnishing, doityourself, electrical, energy, fireplace, florist, garden_centre, garden_furniture, gas, glaziery, hardware, houseware, locksmith, paint, security, trade
	Furniture and interior:	antiques, bed, candles, carpet, curtain, doors, flooring, furniture, household_linen, interior_decoration, kitchen, lamps, lighting, tiles, window_blind
	Electronics:	computer, electronics, hifi, mobile_phone, radiotechnics, vacuum_cleaner
	Outdoors and sport, vehicles:	atv, bicycle, boat, car, car_repair, car_parts, caravan, fuel, fishing, golf, hunting, jetski, military_surplus, motorcycle, outdoor, scuba_diving, ski, snowmobile, sports, swimming_pool, trailer, tyres
	Art, music, hobbies:	art, collector, craft, frame, games, model, music, musical_instrument, photo, camera, trophy, video, video_games
	Stationery, gifts, books, newspapers:	anime, books, gift, lottery, newsagent, stationery, ticket
	Others:	bookmaker, cannabis, copyshop, dry_cleaning, e-cigarette, funeral_directors, laundry, money_lender, party, pawnbroker, pet, pet_grooming, pest_control, pyrotechnics, religion, storage_rental, tobacco, toys, travel_agency, vacant, weapons, outpost

Design aspect is evaluated by the measures that are frequently used in the study of street network studies. It is included in the measurement of the TOD characteristics because the network structural characteristics have been found to have significant relationship with the functional or social urban aspects, such as services localization (Peponis, Allen, French, Scoppa, & Brown, 2007), population density (Tang, 2003), as well as walking flow and urbanity (Omer & Jiang, 2008) (Van Nes & ZhaoHui, 2009) (Hamaina, Leduc, & Moreau, 2011). Computational network analysis will be carried out to calculate the topological and centrality measures of the street network, which is retrieved from the OSM (OpenStreetMap contributors, 2017). We have used some of the indicators for analyzing transportation network in chapter 2 and 3. In the current chapter, we include more measurement and focus on their meanings to street network and pedestrian movement.

1) Topological measurements

Topological measurements evaluate the configuration, connectivity and robustness of the network structure and show how these characteristics and properties are distributed. They measure connectivity and complexity of the network by assessing how thoroughly the nodes are linked together. Topological measures in the current research are divided into the following three subcategories of attributes: connectivity, Clustering Coefficient and urban spatial order. Each category of the attributes is the combined behavior of a group of indicators.

i) Connectivity measure

Connectivity measure is "the minimum number of nodes or edges that must be removed from a connected graph to disconnect it" (Boeing, 2017) and how easy or difficult it is for any two nodes to form a connection. Street connectivity can also be used as a measure of the following attributes.

Resilience. Higher connectivity is more robust against failures, disruptions or attacks because if offers more alternative routing choices.

Pedestrian accessibility to transit stops, services and utilities. Less-connected streets increase the route distance for the pedestrian. "Appropriate connection between minor and pedestrian routes to major transit streets is critical for facilitating effective and efficient access to public transit" (Sharifi, 2019).

Choice of traffic mode. The decisions to walk or bike are often affected by people's "perception of certain trip length thresholds" (Larco, 2016). Combining higher connectivity with other features, including high levels of mixed-use development and high density, can influence on people's choice of traffic mode, shift their travel behavior toward more active modes and subsequently reduce transport-related Greenhouse Gas (GHG) emissions (Ewing & Cervero, 2010).

Social and health-related benefits. High connectivity also facilitates the access to local amenities and employment by walking. Higher concentration of walking population attracts more job opportunities in the neighborhood. Therefore, higher connectivity is associated with social and health-related benefits (Sharifi, 2019).

The number of nodes, number of links, average segment length and average Node Degree are commonly used measures of connectivity.

- **Number of nodes and links**

Higher number of nodes and links are associated with higher connectivity.

- **Average segment length**

Street segment length is a measure of distance between nodes. Average segment length is defined as the average length of the shortest segments between all possible pairs of nodes. Lower values indicate better connectivity and walkability.

- **Average Node Degree**

Node degree is defined as the total number of links branching off or concurring in the node. The average Node Degree of a network is a measure that gave an indication of the organization, configuration, and overall **level of connectivity** of the network. Higher value of the average node degree means better communication between the nodes. In terms of walkability, average Node Degree provides the information on route option and is found to be positively correlated to walking frequency (Pooley, et al., 2013)

ii) Clustering Coefficient (C)

Within the network, there are usually some subsets of the system, which are referred to as clusters, that share some common set of properties. These clusters often have important influence on the network's makeup. Clustering Coefficient is a measurement that reveals the topological structure and distribution of these clusters. Higher value indicates more clusters exist in the network. Networks can be categorized according to whether these clusters are centralized, decentralized, or randomly distributed.

iii) Urban spatial order & entropy of street bearing

In this research, urban spatial order, which quantifies the extent to which a network is ordered according to a single grid (Boeing, 2019), is measured by the entropy of the street bearing. Street network that is more grid-like exhibits less entropy.

2) Centrality measures

The degree of connectivity to a given node in a network is closely intertwined with broader concept of centrality, which is a measure that captures how influential or significant a node is within the overall network. Previously in chapter 3, we have employed the same centrality measures to analyze the transportation network. In this chapter, the focus is on the meaning and implication of these measures to the relationship between the street network and the following functions.

Resilience and robustness. Nodes and links with high centrality values play very significant roles in the network system for the service and facilities to be reachable. However, if these important nodes and links are obstructed, the reachability and continuity in the system would be undermined and the traffic volume may not be appropriately distributed by alternative streets (Mattsson & Jenelius, 2015). Therefore, from the perspective of resilience and robustness, the design target of the network is to avoid polarizing the accessibility in the system and the dependence on the highly central nodes and links (Aydin, Duzgun, Wenzel, & Heinimann, 2018). "To avoid such a polarization, it is suggested that the extent of centrality is determined using a hierarchic approach that follows *power law distribution*. This implies having small, medium, and large numbers of high-, moderate-, and low-centrality nodes/links in the system, respectively" (Sharifi, 2019).

Economic activity. Street centrality has been found to have strong and positive correlations with economic activity (Remali, Porta, Romice, & Abudib, 2015) (Porta, et al., 2012) (Liu, Wei, Jiao, & Wang, 2016). "Areas that have higher Betweenness Centrality values are unique locations in the built environment that have a higher potential of being traversed by people and freight trips to other locations in the city. This high potential to attract through traffic increases the possibility of generating business opportunities in areas with high Betweenness Centrality" (Sharifi, 2019).

Each centrality measurement captures different types of importance of a node or link in the street network. The calculation of these indicators has been presented in chapter 3, and here we highlight their implications for the street network.

- **Closeness Centrality (C_c)**

Closeness Centrality is an indicator of **accessibility** because it indicates the ability to reach a node from any location in the network (Porta, Crucitti, & Latora, 2006). Since higher Closeness Centrality indicates higher accessibility, placing services and utilities in the location with high Closeness Centrality reduces spatial disconnection between places (Porta, et al., 2009). This benefit is of significant importance when choosing the location for a service, function, facility or amenity in the street network, which is the topic we will deal with in the next chapter.

- **Betweenness centrality (C_b)**

Betweenness Centrality measures whether a location is the **intermediary** between others. A location with higher Betweenness Centrality means it lies among many other locations and is critical for maintaining the functionality of the street network. However, this also means that if such a node or link is disrupted, this will cause serious failure across the system (Akbarzadeh, Memarmontazerin, Derrible, & Salehi Reihani, 2019). Therefore, from the perspective of resilience and robustness, street network structure with extremely high Betweenness Centrality values, which exist in star- or wheel-shaped street networks with a single dominant node, should be avoided (Kermanshah & Derrible, 2017).

In total, 11 measures listed in Table 6.1 are derived and categorized based on the proposed framework for each catchment area around the station. The determination of the size of the catchment area has been discussed in chapter 4. As explained there, the size of the catchment areas around stations is defined as a square of 1500*1500m^2 because previous researches have shown that this distance is related to measures like typology analyses, physical activity and walking trips (Val, 2015) (Gunn, et al., 2017).

3) Correlation of indicators

It should be noted that the correlations between indicators is only based on the case of Hamburg. In the future, it would be interesting to compare these correlations among cities with different spatial configuration.

It should be noted that the correlations between indicators is only based on the case of Hamburg. In the future, it would be interesting to compare these correlations among cities with different spatial configuration.

Table 6.3 presents the statistical correlation between indicators in order to find whether they are correlated and influential to each other. The patterns of correlation can be broadly divided into the following groups.

First of all, the correlation patterns of the three indicators related to POI are quite similar to each other. They are **significantly** related to all indicators except for **average Node Degree** and **average Clustering Coefficient**. They are positively correlated with the **number of nodes, number of links** and **entropy of street bearing**. They are negatively correlated with **average Betweenness Centrality**. Only **POI diversity** is significantly and negatively correlated with **average Closeness Centrality.** This means that in the RPT neighborhoods in Hamburg where there are larger numbers of nodes and links, higher level of street bearing entropy and lower level of average Betweenness Centrality, there are also higher levels of accessibility, density and diversity of POI.

Secondly, the following three indicators are significantly correlated to only a few of other indicators.

- **Average Node Degree** is only significantly and positively related to one indicator, the **number of links.** This means that neighborhoods with more links, i.e. street segments, also have higher average node degree.
- **Average Clustering Coefficient** is only significantly and positively related to **POI accessibility, POI density, and POI diversity.** This is an interesting observation and deserves further investigation to examine whether larger number of clusters in the street network actually benefits the local business. **Average Clustering Coefficient** is also significantly and negatively correlated to **average segment length**. This may indicate that shorter segments in the neighborhoods generates more clusters in the network.
- **Average Closeness Centrality** is only significantly and positively related to **average Betweenness Centrality.** It is negatively related to the **number of nodes**, **entropy of street bearing** and **POI diversity**.

Thirdly, the following indicators are significantly correlated to most other indicators except for **average Node Degree** and **average Clustering Coefficient.**

- **Number of nodes** is positively correlated with **number of links**, **entropy of street bearing**, **POI accessibility**, **POI density**, and **POI diversity**. It is negatively correlated with **average segment length**, **average Closeness Centrality** and average **Betweenness Centrality**.

- **Entropy of street bearing** is positively correlated with **number of nodes, number of links, POI accessibility, POI density, and POI diversity.** It is negatively correlated with **average segment length, average Closeness Centrality, average Betweenness Centrality.**
- **Average Betweenness Centrality** is positively correlated with **average segment length.** It is negatively correlated with **number of nodes, number of links, entropy of street bearing, average Closeness Centrality, average Betweenness Centrality, POI accessibility, POI density, and POI diversity.**

Finally, the following indicators are significantly correlated to most of the indicators except for **average Closeness Centrality**.

- **Number of links** is positively correlated with **number of nodes, average Node Degree, entropy of street bearing, POI accessibility, POI density, and POI diversity**. It is negatively correlated with **average segment length and average Betweenness Centrality.**
- **Average segment length** is positively correlated with **average Betweenness Centrality.** It is negatively correlated with **number of nodes, number of links, entropy of street bearing, average Clustering Coefficient, POI accessibility, POI density, and POI diversity.**

It should be noted that the correlations between indicators is only based on the case of Hamburg. In the future, it would be interesting to compare these correlations among cities with different spatial configuration.

Table 6.3 Correlation pairs between indicators.

	Number of nodes	Number of links	Average Node degree	Average segment length	Entropy of street bearing	Average Clustering Coefficient	Average Closeness Centrality	Average Betweenness Centrality	POI accessibility	POI density	POI diversity
Number of nodes	1	0.98***	0.18	-0.81***	0.94***	0.20	-0.26*	-0.87***	0.79***	0.77***	0.82***
Number of links	0.98***	1	0.38***	-0.81***	0.92***	0.14	-0.18	-0.83***	0.78***	0.73***	0.76***
Average Node Degree	0.18	0.38***	1	-0.16	0.19	-0.22	0.19	-0.13	0.17	0.05	-0.04
Average segment length	-0.81***	-0.81***	-0.16	1	-0.81***	-0.40***	-0.02	0.76***	-0.70***	-0.67***	-0.68***
Entropy of street bearing	0.94***	0.92***	0.19	-0.81***	1	0.18	-0.35***	-0.92***	0.67***	0.67***	0.81***
Average Clustering Coefficient	0.20	0.14	-0.22	-0.40***	0.18	1	0.05	-0.12	0.30**	0.33**	0.23*
Average Closeness Centrality	-0.26*	-0.18	0.19	-0.02	-0.35***	0.05	1	0.30**	-0.06	-0.16	-0.28**
Average Betweenness Centrality	-0.87***	-0.83***	-0.13	0.76***	-0.92***	-0.12	0.30**	1	-0.63***	-0.66***	-0.78***
POI accessibility	0.79***	0.78***	0.17	-0.70***	0.67***	0.30**	-0.06	-0.63***	1	0.96***	0.80***
POI density	0.77***	0.73***	0.05	-0.67***	0.67***	0.33**	-0.16	-0.66***	0.96***	1	0.82***
POI diversity	0.82***	0.76***	-0.04	-0.68***	0.81***	0.23*	-0.28**	-0.78***	0.80***	0.82***	1

*** Correlation is significant at 0.01 level; **Correlation is significant at 0.05 level; *Correlation is significant at 0.1 level

6.2.2 Methodologies

Descriptive statistics for the measures are computed and cluster analysis is used to group the stations in order to determine their levels of walkability. In this chapter, Affinity Propagation (AP) (Frey & Dueck, 2007) is employed to group the TOD neighborhoods of similar attributes. AP chooses the number of clusters based on the data and it doesn't require specifying the number of clusters in advanced. The AP algorithm forms clusters by sending messages between pairs of instances until convergence. "Each data point sends messages to all other points informing its targets of each target's relative attractiveness to the sender. Each target then responds to all senders with a reply informing each sender of its availability to associate with the sender, given the attractiveness of the messages that it has received from all other senders. Senders reply to the targets with messages informing each target of the target's revised relative attractiveness to the sender, given the availability messages it has received from all targets. The message-passing procedure proceeds until a consensus is reached. Once the sender is associated with one of its targets, that target becomes the point's exemplar. All points with the same exemplar are placed in the same cluster. The exemplars are identified as those most representative of other samples" (Thavikulwat, 2008).

The 11 indicators identified in Table 6.1 are applied for the AP clustering. The "exemplars" defined by AP represents the average characteristics of the neighborhoods in the same cluster. This is an important step because these exemplars will be used as the target neighborhoods and be further investigated in detail in the next chapter.

The process of clustering TOD neighborhoods can be divided into the following steps.

- Step1: Download street network data of the TOD neighborhoods from OSM using the function of *osmnx.graph_from_point* in Python library of OSMnx
- Step2: Calculate the street orientation using the function of *osmnx.add_edge_bearings* from the Python library of OSMnx. Calculate the entropy of street orientation using the function of *entropy* from the python library of scipy.stats.

- Step3: Calculate the number of nodes, number of links, average segments length, average Node Degree using the function of *osmnx.basic_stats* in Python library of OSMnx
- Step4: Calculate the average Clustering Coefficient, average Closeness Centrality, average Betweenness centrality using the function of *osmnx.extended_stats* in Python library of OSMnx
- Step5: Retrieve the coordinate point of interest (POI) using the function of *osmnx.pois_from_point* in Python library of OSMnx
- Step6: Calculate the density of POI by calculating the number of POI within the neighborhood.
- Step7: Calculate the diversity of POI by calculating the types of amenities within the neighborhood
- Step8: Calculate accessibility of POI by searching for the nearest amenities to each node using the function of *pandana.network.nearest_pois* in the Python library of Pandana. Each node then has the information about how many POI are within 750m from it. The sum of this information is then divided by the total number of nodes to form the accessibility of POI of each neighborhood.
- Step9: Cluster TOD neighborhoods using the function of *sklearn.cluster.AffinityPropagation* in the Python library of scikit-learn
- Step10: Plot the POI and the cluster of POI using the Python library of *matplotlib*.

6.3 Result

6.3.1 Numbers of sub-cluster and list of stations in each sub-cluster

By applying the Affinity Propagation, we aim to achieve two targets. Firstly, within each Node-Value cluster, we search for the sub-clusters that can be further grouped together based on their Place Value. Secondly, we identify the “exemplars” that are representative for each sub-cluster and scrutinize the detailed distribution of street network and POI in these exemplar neighborhoods in the next chapter.

Results of Affinity Propagation indicates that stations in Node-Value-cluster-0 can be further divided into 3 sub-clusters, which are named Cluster 0-0, Cluster 0-1 and Cluster 0-2 in

Table 6.4; stations in Node-Value-cluster-1 can be further divided into 4 sub-clusters, which are Cluster 1-0, Cluster 1-1, Cluster 1-2 and Cluster 1-3 in Table 6.5; stations in Node-Value-cluster-2 can be further divided into 2 sub-clusters, which are Cluster 2-0 and Cluster 2-1 in Table 6.6. The exemplar of each sub-cluster is highlighted with gray background and their distribution of street network and POI will be examined further in the following chapter.

Table 6.4 Stations in each sub-cluster of Node-Value cluster 0.
(Exemplar of each group is lighted with gray background.)

Cluster 0-0	**Cluster 0-1**	**Cluster 0-2**
Altona	Agathenburg	Billstedt
Buxtehude	Ahrensburg Ost	Blankenese
Dammtor (Messe/CCH)	Ahrensburg West	Diebsteich
HafenCity Universität	Aumühle	Elbbrücken
Harburg	Buchenkamp	Hamburg Airport (Flughafen)
Harburg Rathaus	Buckhorn	Hammerbrook (City Süd)
Holstenstraße	Dollern	Heimfeld
Neugraben	Fischbek	Horneburg
Norderstedt Mitte	Großhansdorf	Joachim-Mähl-Straße
Sternschanze (Messe)	Halstenbek	Neu Wulmstorf
Überseequartier	Hoheneichen	Neuwiedenthal
	Hoisbüttel	Ohlsdorf
	Iserbrook	Pinneberg
	Kiekut	Rissen
	Krupunder	Stade
	Mümmelmannsberg	Veddel (BallinStadt)
	Neukloster	Wedel
	Niendorf Nord	Wilhelmsburg
	Ohlstedt	
	Reinbek	
	Sülldorf	
	Schippelsweg	
	Schmalenbeck	
	Thesdorf	
	Volksdorf	
	Wohltorf	

Table 6.5 Stations in each sub-cluster of Node-Value cluster 1.
(Exemplar of each group is lighted with gray background.)

Cluster 1-0	Cluster 1-1	Cluster 1-2	Cluster 1-3
Alte Wöhr (Stadtpark)	Allermöhe	Alter Teichweg	Gänsemarkt
Barmbek	Alsterdorf	Bahrenfeld	Hauptbahnhof
Baumwall (Elbphilharmonie)	Billwerder-Moorfleet	Berne	Hauptbahnhof Süd
Bergedorf	Eidelstedt	Borgweg (Stadtpark)	Meßberg
Berliner Tor	Elbgaustraße	Fuhlsbüttel	Mönckebergstraße
Christuskirche	Farmsen	Garstedt	Rathaus
Dehnhaide	Fuhlsbüttel Nord	Hallerstraße	Rödingsmarkt
Emilienstraße	Hagenbecks Tierpark	Hasselbrook	Steinstraße
Eppendorfer Baum	Hagendeel	Hoheluftbrüke	
Feldstraße (Heiligengeistfeld)	Hochkamp	Hudtwalckerstraße	
Friedrichsberg	Kiwittsmoor	Klein Borstel	
Habichtstraße	Klein Flottbek (Botanischer Garten)	Klosterstern	
Hamburger Straße	Kornweg (Klein Borstel)	Landwehr	
Kellinghusenstraße	Langenfelde	Langenhorn Markt	
Lattenkamp (Sporthalle)	Langenhorn Nord	Niendorf Markt	
Lohmühlenstraße	Meiendorfer Weg	Ochsenzoll	
Lübecker Straße	Merkenstraße	Othmarschen	
Lutterothstraße	Mittlerer Landweg	Poppenbüttel	
Messehallen	Oldenfelde	Richtweg	
Mundsburg	Rothenburgsort	Ritterstraße	
Nettelnburg	Steinfurther Allee	Saarlandstraße	
Osterstraße	Stellingen (Arenen)	Sengelmannstraße (City Nord)	
Rubenkamp (City Nord)	Tiefstack	Sierichstraße	
Schlump	Trabrennbahn	Straßburger Straße	
St. Pauli	Wandsbek-Gartenstadt	Wartenau	
Stephansplatz (Oper/CCH)	Wellingsbüttel		
Uhlandstraße			
Wandsbek Markt			
Wandsbeker Chaussee			

Table 6.6 Stations in each sub-cluster of Node-Value cluster 2.
(Exemplar of each group is lighted with gray background.)

Cluster 2-0	Cluster 2-1
Burgstraße	Hauptbahnhof Nord
Hammer Kirche	Jungfernstieg
Horner Rennbahn	Stadthausbrücke
Königstraße	
Landungsbrücken	
Legienstraße	
Rauhes Haus	
Reeperbahn	

6.3.2 Characterizing each sub-cluster based on Place Value

The main purpose of this section is to assign Place Value to each sub-cluster based on their characteristics. The characteristics are identified by 1) spatial distribution and 2) statistical distribution of the measurement.

i) Node-Value-cluster 0

The spatial distribution of the three sub-clusters in Node-Value-cluster-0 is mapped in Figure 6.1. In chapter 3 we have learned that, compared to other clusters, stations in Node-Value-cluster-0 are mostly distributed in the middle to outer suburbs and have the lowest Node Value. After further dividing the Node-Value-cluster-0 into three sub-clusters based on their Place Values, the results show that stations of

- Cluster 0-0 are predominantly located in the inner city, with some exceptions in the middle suburbs.
- Cluster 0-1 are mainly distributed in the middle and outer suburbs.
- Cluster 0-2 are distributed all over, including inner city, middle and outer suburbs.

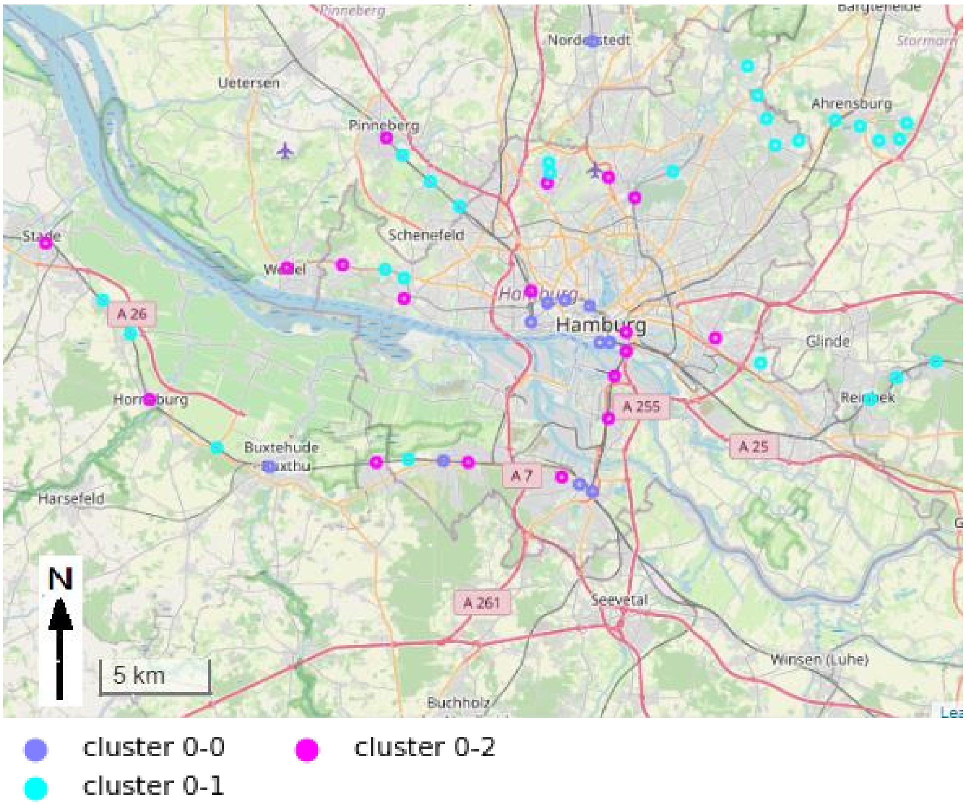


Figure 6.1 Spatial distribution of each sub-cluster in Node-Value-cluster 0.

Table 6.7 Descriptive statistics of each sub-cluster in Node-Value-cluster-0.

	All station in cluster 0		Cluster 0-0		Cluster 0-1		Cluster 0-2	
	mean	std	**mean**	std	**mean**	**std**	**mean**	std
Number of nodes	**165.24**	87.90	**308.09**	50.65	**93.62**	26.28	**181.39**	26.27
Number of links	**406.91**	230.57	**786.18**	152.97	**227.12**	66.27	**434.83**	72.81
Average Node Degree	**4.88**	0.44	**5.12**	0.75	**4.84**	0.30	**4.79**	0.34
Average segment length(m)	**99.22**	27.13	**65.62**	13.21	**120.71**	18.94	**88.73**	12.43
Entropy of street bearing	**5.67**	0.50	**6.33**	0.20	**5.24**	0.30	**5.88**	0.17
Average Clustering Coefficient	**0.07**	0.03	**0.08**	0.02	**0.06**	0.03	**0.07**	0.03
Average Closeness Centrality	**0.0010**	0.0002	**0.0010**	0.0002	**0.0011**	0.0002	**0.0010**	0.0001
Average Betweenness Centrality	**0.07**	0.02	**0.05**	0.01	**0.09**	0.02	**0.06**	0.01
POI_Accessibility	**63.03**	69.03	**171.20**	80.83	**22.63**	14.50	**55.29**	29.23
POI_Density	**146.75**	163.88	**399.82**	193.20	**46.42**	25.48	**137.00**	73.89
POI_Diversity	**23.31**	10.06	**36.18**	6.60	**15.96**	6.73	**26.06**	5.61

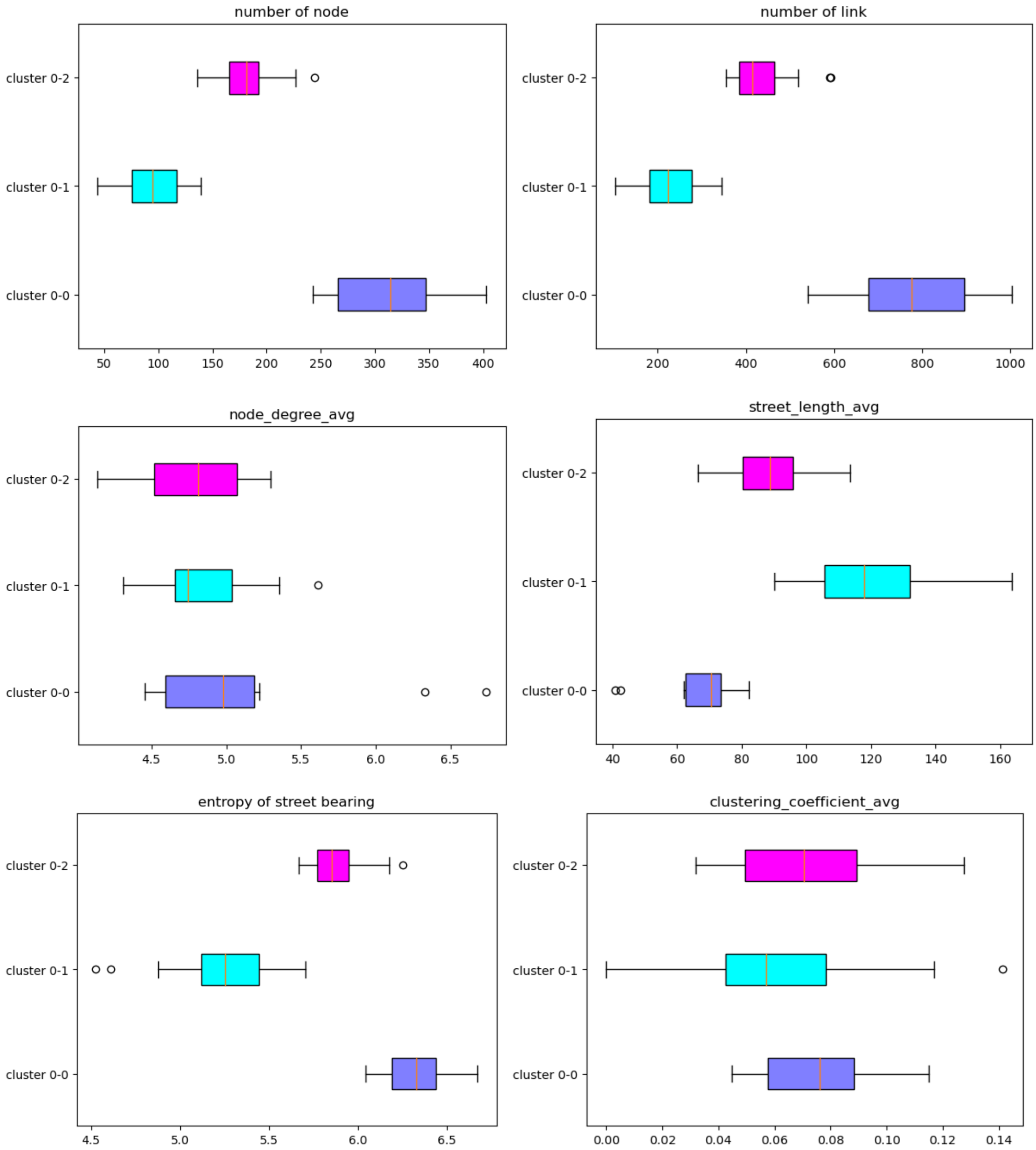


Figure 6.2 Statistical distribution of the characteristics of each group in Node-Value-cluster-0.

Next, the sub-clusters can be characterized by the statistical distribution of the indicators. Based on the criteria we have set up in Table 6.1, the results in Table 6.7 and Figure 6.2 show that cluster 0-0 has the highest Place Value because it has the highest value in terms of number of nodes, number of links, average Node Degree and entropy of street bearing. Also, it has the lowest value of average segment length and Betweenness Centrality. This means that the street networks in the neighborhoods of these stations are well connected and more walkable than the stations in the other two groups. This group of stations also has highest value in terms of accessibility, density and diversity of POI. Cluster 0-1 has the lowest Place Value, while cluster 0-2 has the middle Place Value in cluster 0.

In sum, with regard to Place Value, cluster 0-0 > cluster 0-2 > cluster 0-1.

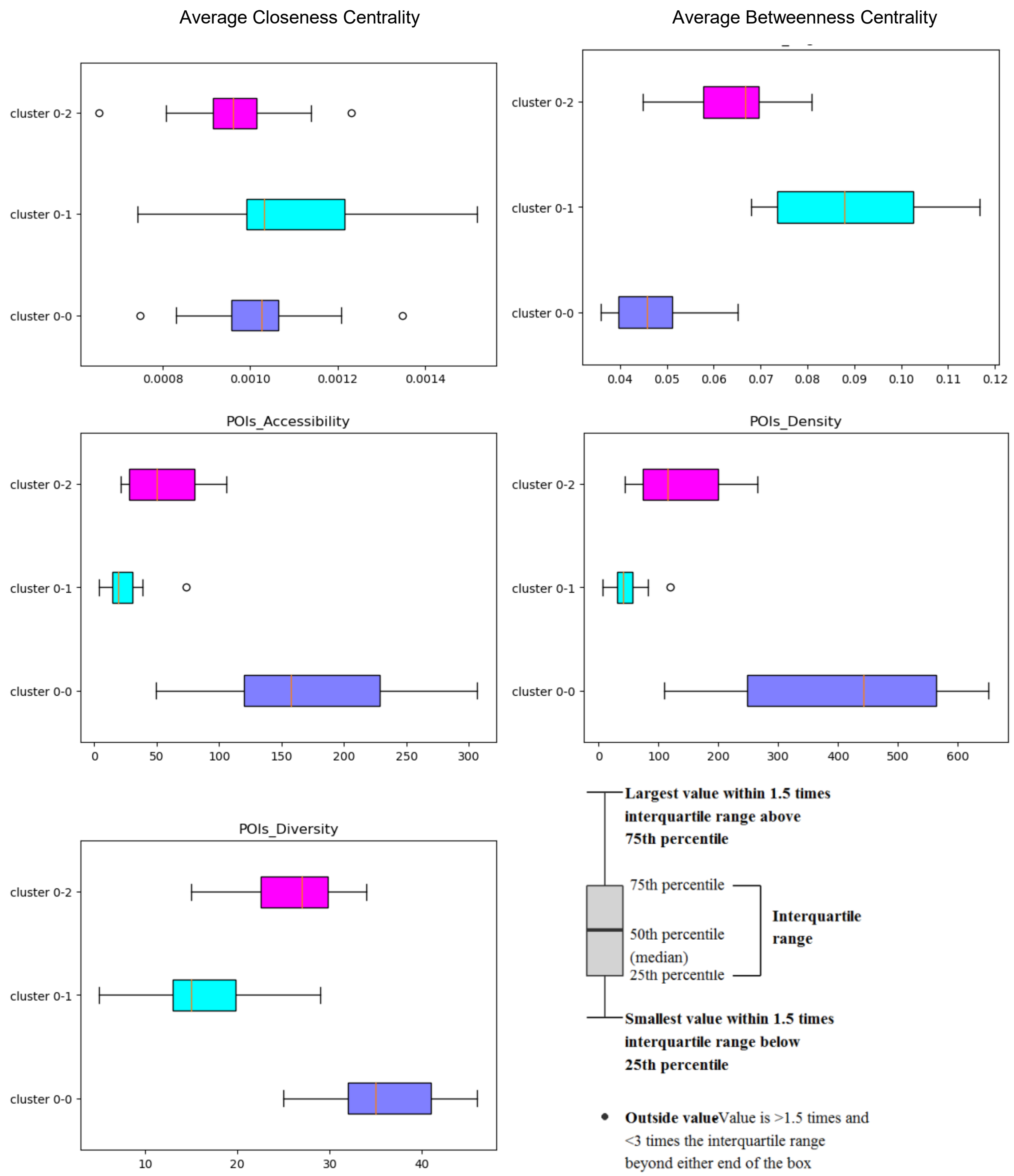


Figure 6.2 (continued) Statistical distribution of the characteristics of each group in Node-Value-cluster-0.

ii) Node-Value-cluster 1

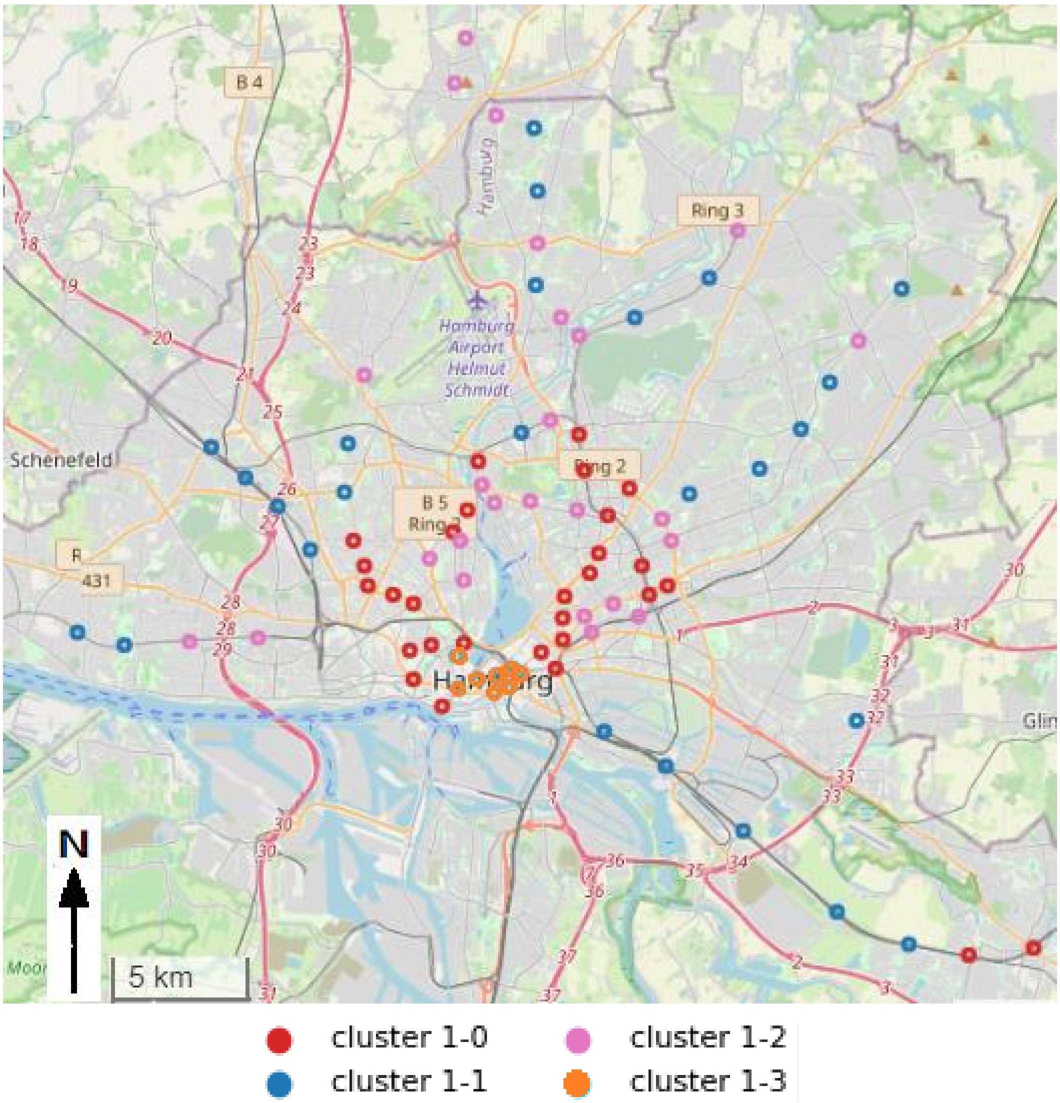


Figure 6.3 Spatial distribution of each sub-cluster in Node-Value-cluster-1.

The spatial distribution of the three sub-clusters in Node-Value-cluster-1 is mapped in Figure 6.3. In chapter 3 we have learned that, compared to other clusters, stations in Node-Value-cluster-1 are mostly distributed in the inner city and middle suburb. After further dividing the cluster into 4 sub-clusters based on their Place Values, the results show that stations of

- Cluster 1-0 are predominantly located in the middle suburb, with two exceptions in the outer suburbs.
- Cluster 1-1 are mainly distributed in the outer suburb.
- Cluster 1-2 are mainly distributed in the middle and outer suburbs.
- Cluster 1-3 are distributed in the inner city.

Next, the sub-clusters can be characterized by the statistical distribution of the indicators. Based on the criteria we have set up in Table 6.1, the results in Table 6.8 and Figure 6.4 show that cluster 1-3 has the highest Place Value because it has the highest value in terms of number of nodes, number of links, average Node Degree and entropy of street bearing. Also, it has the lowest value of average segment length and Betweenness Centrality. This group of stations also has highest value in terms of accessibility, density and diversity of POI. Cluster 1-1 has the lowest Place Value because it has the lowest values in terms of number of nodes, number of links and entropy of street bearing. Also, it has the highest value of average segment length, Closeness Centrality and Betweenness Centrality.

In sum, with regard to Place Value, cluster 1-3 > cluster 1-0 > cluster 1-2 > cluster 1-1.

Table 6.8 Descriptive statistics of each sub-cluster in Node-Value-cluster-1.

	All station in cluster 1		Cluster 1-0		Cluster 1-1		Cluster 1-2		Cluster 1-3	
	mean	std	**mean**	std	**mean**	std	**mean**	std	**mean**	std
Number of nodes	**224.84**	128.56	**264.83**	48.78	**104.19**	33.57	**200.88**	29.67	**546.88**	87.08
Number of links	**523.98**	337.12	**595.86**	123.67	**239.27**	82.16	**455.40**	62.69	**1403.00**	302.31
Average Node Degree	**4.58**	0.37	**4.49**	0.27	**4.54**	0.47	**4.55**	0.21	**5.09**	0.33
Average segment length(m)	**86.53**	21.71	**75.17**	10.75	**108.85**	18.70	**86.32**	13.86	**55.81**	4.99
Entropy of street bearing	**5.88**	0.58	**6.17**	0.16	**5.21**	0.53	**5.94**	0.15	**6.80**	0.11
Average Clustering Coefficient	**0.08**	0.03	**0.09**	0.02	**0.08**	0.04	**0.09**	0.02	**0.08**	0.01
Average Closeness Centrality	**0.00098**	0.00017	**0.00098**	0.00010	**0.00101**	0.00026	**0.00096**	0.00016	**0.00098**	0.00003
Average Betweenness Centrality	**0.06188**	0.02498	**0.04959**	0.00599	**0.08828**	0.02951	**0.05859**	0.00653	**0.03091**	0.00277
POI_Accessibility	**122.82**	105.78	**156.80**	49.32	**35.90**	33.93	**87.67**	28.84	**391.96**	48.21
POI_Density	**332.86**	272.05	**444.48**	139.06	**94.42**	69.79	**240.44**	89.73	**992.00**	87.63
POI_Diversity	**31.73**	10.50	**38.93**	4.55	**19.42**	6.87	**31.60**	5.35	**46.00**	3.12

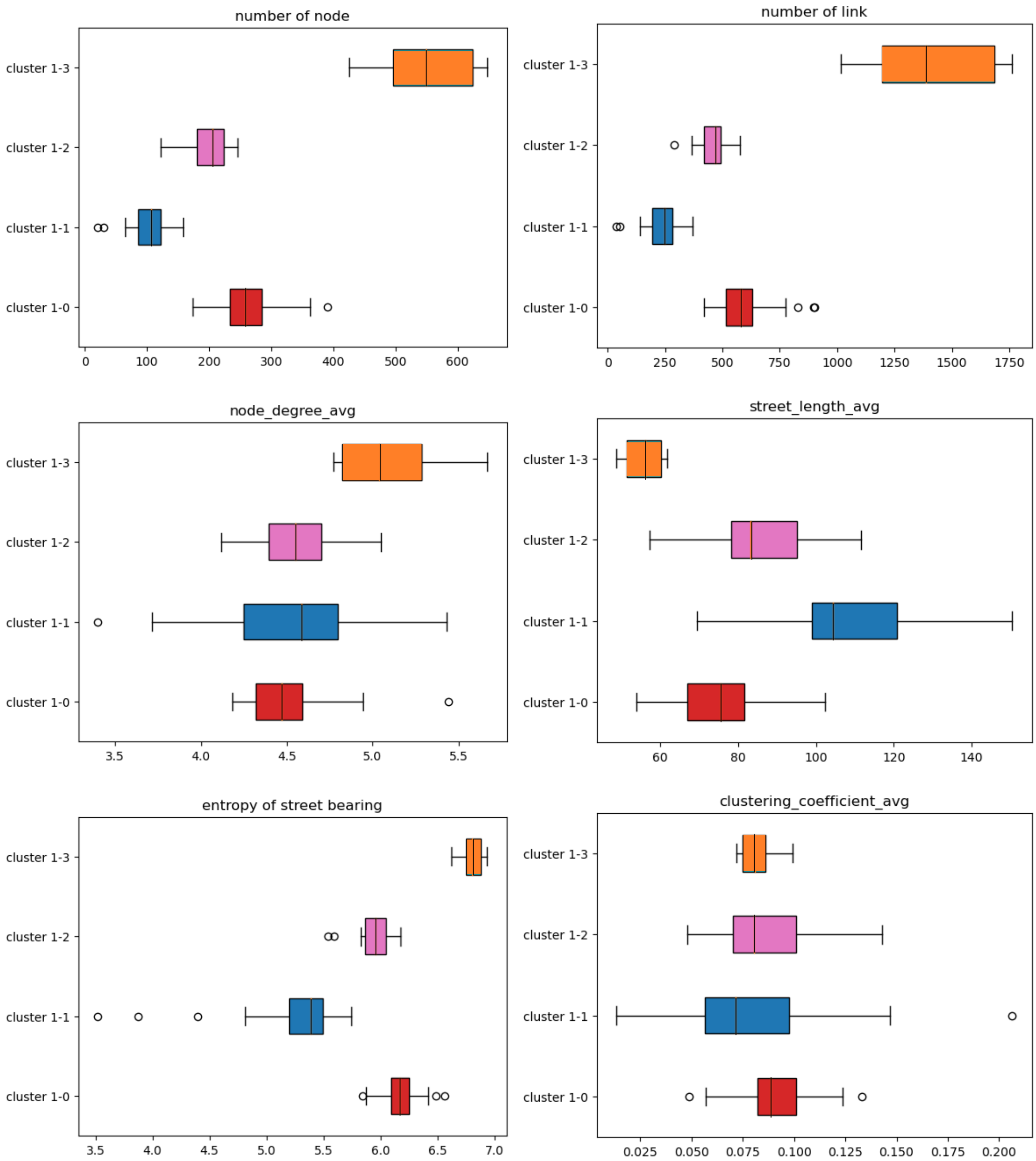


Figure 6.4 Statistical distribution of the characteristics of each sub-cluster in Node-Value-cluster-1.

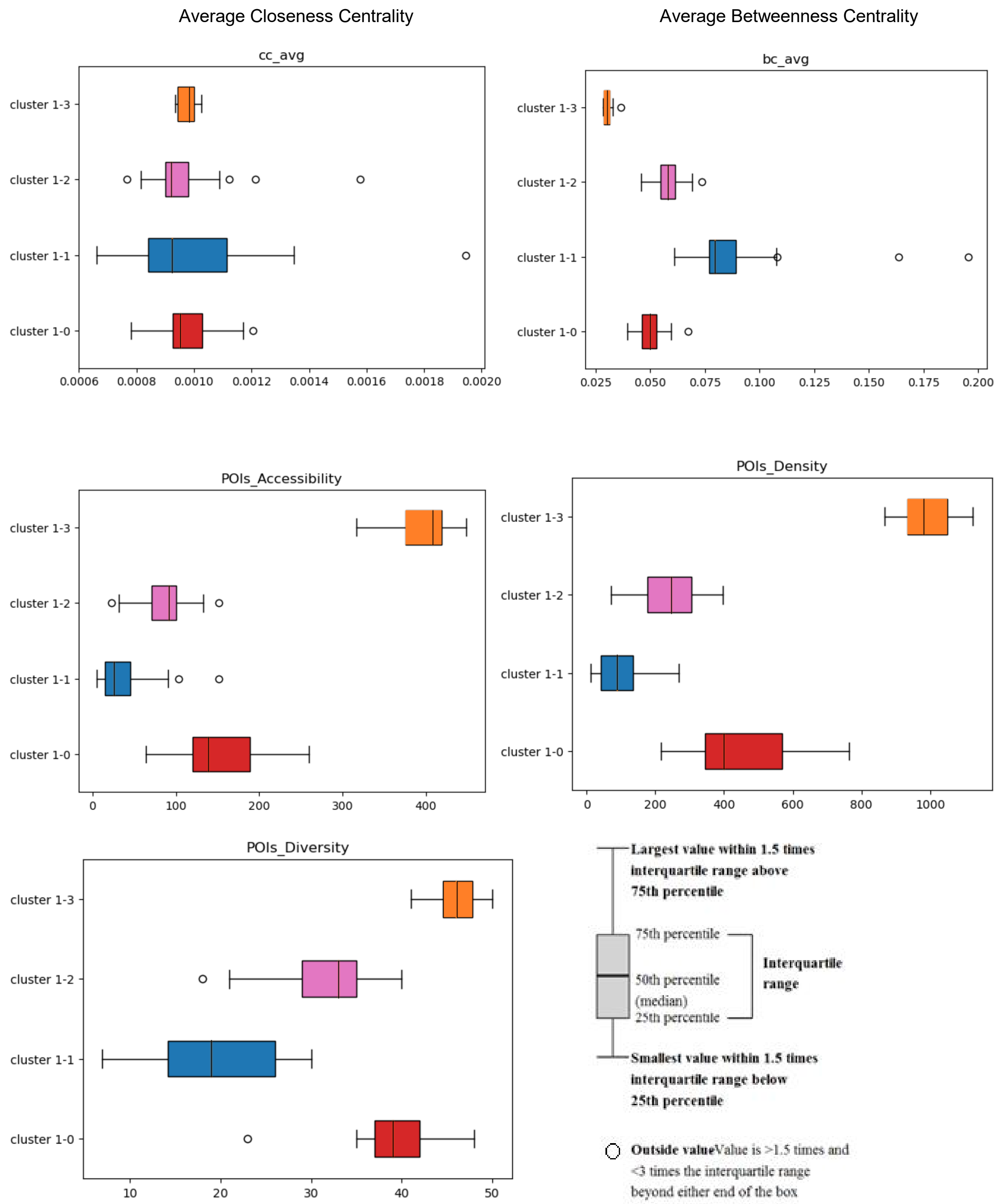


Figure 6.4 (continued) Statistical distribution of the characteristics of each sub-cluster in Node-Value-cluster-1.

iii) Node-Value-cluster 2

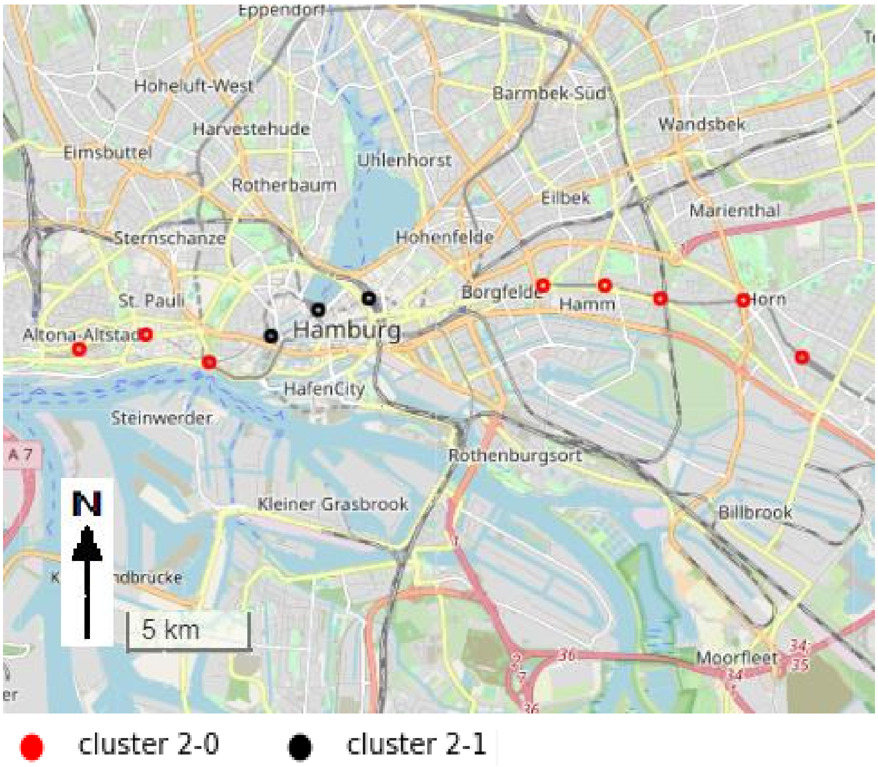


Figure 6.5 Spatial distribution of each sub-cluster in Node-Value-cluster-2.

The spatial distribution of the three sub-clusters in Node-Value-cluster-2 is mapped in Figure 6.5. In chapter 3 we have learned that, compared to other clusters, stations in Node-Value-cluster-2 are mostly distributed in the inner city and has the highest Node Value. After further dividing the cluster into 2 sub-clusters based on their Place Values, the results show that stations of

- Cluster 2-0 are mainly distributed in the inner and middle suburbs.
- Cluster 2-1 are predominantly located in the inner city.

Next, the sub-clusters can be characterized by the statistical distribution of the indicators. Based on the criteria we set up in Table 6.1, the results in Table 6.9 and Figure 6.6 show that cluster 2-1 has the highest Place Value because it has the highest value in terms of the number of nodes, number of links, average Node Degree and entropy of street bearing. Also, it has the lowest value of average segment length and Betweenness Centrality. This group of station also has highest value in terms of accessibility, density and diversity of POI. By contrary, cluster 2-0 has the lowest Place Value among the Node-Value-cluster-2.

In sum, with regard to Place Value, cluster 2-1 > cluster 2-0.

Table 6.9 Descriptive statistics of each sub-cluster in Node-Value cluster 2.

	All station in cluster 2		Cluster 2-0		Cluster 2-1	
	mean	std	**mean**	std	**mean**	std
Number of nodes	**307.45**	154.02	**223.38**	54.78	**531.67**	66.56
Number of links	**707.18**	400.52	**488.13**	128.51	**1291.33**	201.22
Average Node Degree	**4.49**	0.29	**4.36**	0.19	**4.85**	0.15
Average segment length(m)	**79.12**	19.44	**87.21**	16.16	**57.55**	3.90
Entropy of street bearing	**6.17**	0.45	**5.93**	0.19	**6.81**	0.10
Average Clustering Coefficient	**0.08**	0.02	**0.08**	0.02	**0.08**	0.01
Average Closeness Centrality	**0.00092**	0.00012	**0.00090**	0.00013	**0.00099**	0.00004
Average Betweenness Centrality	**0.048**	0.014	**0.055**	0.010	**0.031**	0.003
POI_Accessibility	**196.10**	133.95	**127.62**	71.24	**378.71**	56.31
POI_Density	**492.09**	325.07	**322.75**	174.19	**943.67**	39.51
POI_Diversity	**38.45**	7.65	**35.25**	6.20	**47.00**	2.65

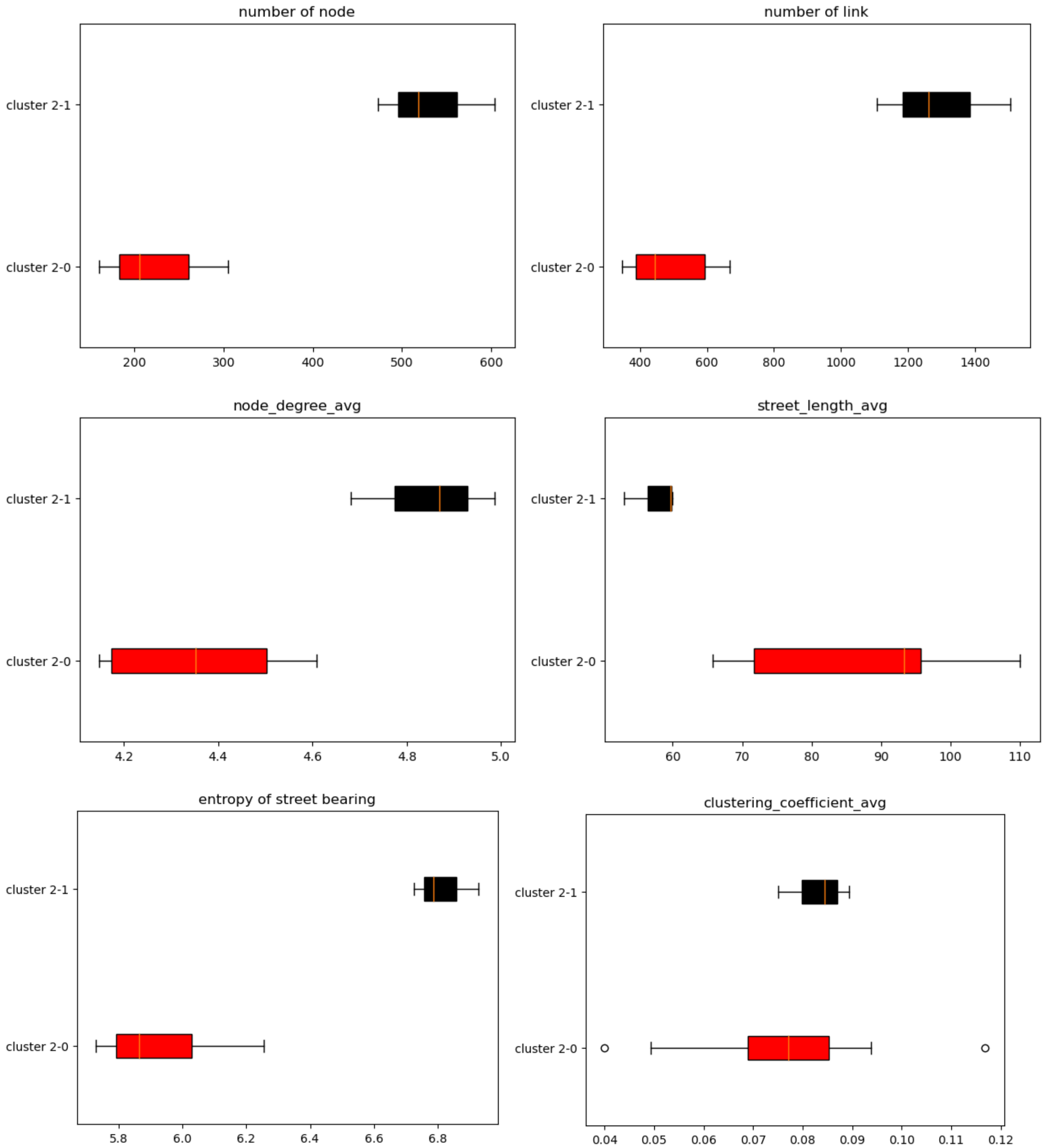


Figure 6.6 Statistical distribution of the characteristics of each sub-cluster in Node-Value-cluster-2.

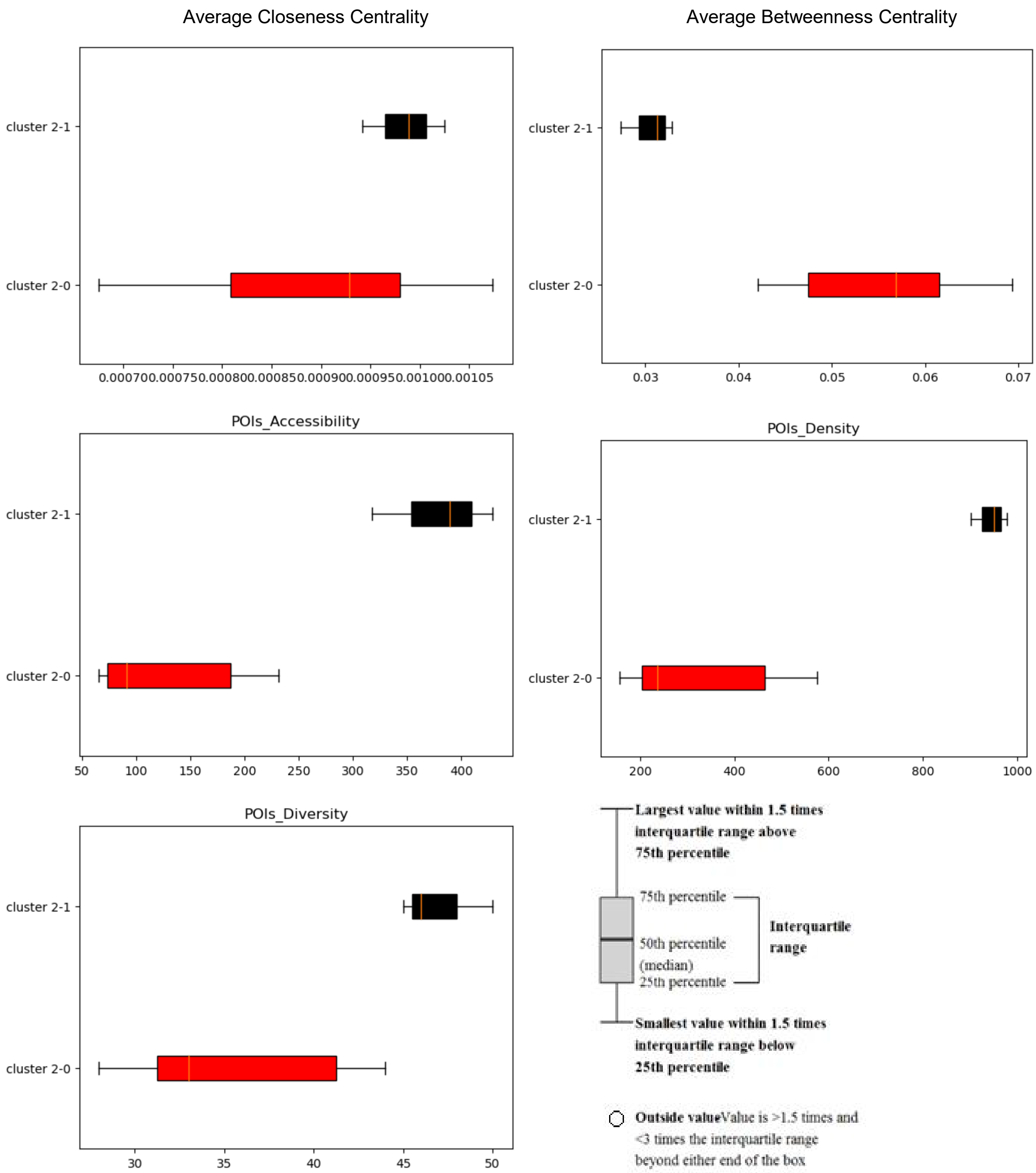


Figure 6.6(continue) Statistical distribution of the characteristics of each group in Node-Value-cluster-2.

6.4 Conclusion and discussion

At the beginning of this chapter, we have set up two tasks as the goals. The first goal is to further divide the stations into sub-clusters based on their Place Values, which are measured by two categories of indicators: network analysis and the accessibility, diversity and density of the POI. Affinity Propagation is used to group the neighborhoods around the train stations in order to determine their Place Value in the Node-Place Model.

After each station is assigned with both the Node and the Place Values, it will be possible to construct the Node-Place Model, as shown in Figure 6.7. The stations can then be compared and ranked by both the Node- and Place-Values. Note that, Figure 6.7 is merely an attempt to relate our results to the Node-Place Model proposed by Bertolini (1999). The values and distance between the clusters are depicted not in absolute but only in relative terms. For instance, the relative position means that cluster 1-2 has lower Node Value than cluster 1-0 and higher Place Value than cluster 0-2.

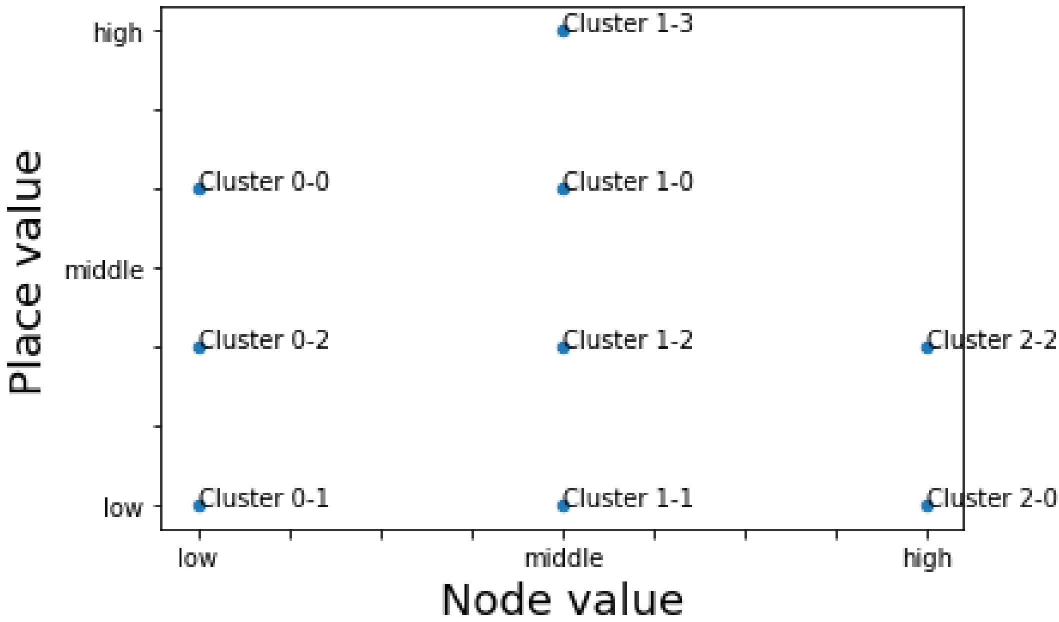


Figure 6.7 Node-Place Model of the HVV railway public transport (RPT) stations.

The results help to identify the neighborhoods that are already successful from the TOD point of view and, on the other hand, the neighborhoods with the potential for further improvement. The values of indicators can be used to indicate the direction for improvement in each sub-cluster.

The other goal of this chapter is to select stations that can be the exemplar of each sub-cluster so that, in the following chapter, we can zoom into the catchment area around the station and scrutinize the detailed spatial distribution of the street network and POI.

Based on the results of network analysis and POI-related indicators, the selected exemplars and the order of their Place Values are:

- Überseequartier (cluster 0-0) > Großhansdorf (cluster 0-2) > Pinneberg (cluster 0-1)
- Steinstraße (cluster 1-3) >Berliner Tor (cluster 1-0)>Klosterstern (cluster 1-2) >Klein Flottbek (Botanischer Garten) (cluster 1-1)
- Hauptbahnhof Nord (cluster 2-1) >Hammer Kirche (cluster 2-0).

In addition, by comparing and Table 6.7, Table 6.8, Table 6.9, one interesting pattern emerges that the high-Node-Value cluster, i.e. cluster 2, also performs better in the measurements of Place Value. For example, comparing to cluster 0 and 1, cluster 2 has the highest average number of nodes, average number of links and average value of the POI-related indicators, i.e. density, diversity and accessibility of POI.

Finally, as mentioned in section 6.2, it should be noted that the correlations between indicators in this chapter is only based on the case of Hamburg. In the future, it would be interesting to compare these correlations among cities with different spatial configuration. In fact, this comparison can also be carried out among TOD neighborhoods in the future studies.

7. Strategic Allocation of Resources in the Transit Oriented Development neighborhoods

In the previous chapter, we have identified the exemplar neighborhood of each sub-cluster. In this chapter, we aim to investigate the current situation of these neighborhoods and propose appropriate interventions for improvement by adding new facilities or services, such as bike rental station (BRS). In order to locate the new BRS at a site that is both accessible and frequently passed by the local residents, the current location and distribution of BRS and the street nodes with the highest Closeness Centrality and Betweenness Centrality in the neighborhood need to be investigated.

At the core of sustainable development is the efficient usage and distribution of limited resources. One of the approaches for achieving sustainable development at the Transit Oriented Development (TOD) neighborhood is the efficient allocation of resources, which is represented by the efficient allocation of Point of Interest (POI) in the current study. Using the bike rental station as an example of POI, this chapter seeks to develop a method for determining the location to set up new BRS in order to increase the accessibility and, therefore, improve the level of walkability in the neighborhood.

We will firstly examine the current allocation of BRS by calculating their coverage rate and by clustering the BRS that are less 300m from each other. The choice of 300m as the radius is based on the study of Kabra et al. (2019). This diagnosis will help us identify areas that are still not being served and are, therefore, suitable for locating new BRS there.

Furthermore, the current distribution of the BRS and the street network characteristics will be investigated as this chapter is an attempt to explore the potential of using network analysis as a method for estimating efficient allocation of POI in TOD neighborhoods. Employing network analysis facilitates the data-driven decision-making process on choosing the location for the services, facilities or amenities that will meet the local demand in the selected exemplar neighborhood and, therefore, helps the planners to effectively handle the urban complexity within the budget constraints. This study employs 1) Closeness Centrality (C_c) to measure **the level of accessibility** (Derrible, 2012) to a POI from surrounding areas via a set of street segments; and 2) Betweenness Centrality (C_b) to measure **the level of movement opportunities** (Derrible, 2012), which measures the frequencies of being passed by, within the TOD neighborhood.

The method proposed in this paper can be applied to any investigation requiring network analysis for determining the suitable location for setting up certain services, such as a mobile clinic, mobile library, test station or a new shop, especially when data that requires large amount of time and cost is not available.

7.1 Location-allocation analysis for determining the optimal location of the bike rental stations (BRS)

The first step of the location-allocation analysis is to decide the **target facility**. In our case the target facility is BRS, and the relationship between the BRSs is not competition but rather complementarity.

Next, we need to identify the **demand point**, which is the location representing people or things that require the functions, services or goods provided by the target facility. In other words, if there is a location where the demand for the BRS is high, this location would be an ideal place to set up the station. In the current study, we do not have the social-demographic data of the residents in the neighborhood. However, a node can be easily reached from different directions or streets, and this makes a node in general a favorite point for a bike station. Since our purpose is to demonstrate the procedure of the proposed methodology, we choose the street nodes to be the demand points.

Finally, we need to think about which problem should be solved by this analysis. Depending on the current situation in the selected neighborhoods, possible problems that need to be solved and the corresponding strategies are listed below.

- **Increase the coverage level →** The corresponding strategy is to increase the number of BRS.
- **Avoid the overlap of the service areas →** The corresponding strategy is to decrease the clusters of BRS.
- **Minimize the distance from the facilities, i.e. BRS, to the demand points, i.e. street nodes with high demand→** The corresponding strategy is to allocate BRS near the nodes with higher Closeness Centrality (because higher Closeness Centrality means higher accessibility).

- **Increase the level of movement opportunities**→ The corresponding strategy is to allocate BRS near the nodes with higher Betweenness Centrality. Higher Betweenness Centrality indicates higher frequencies to be passed by.

7.2 Measurements and methodologies

The identified problems and proposed solutions in the previous section indicate that we need the measurements of coverage level, number of clusters and the Closeness Centrality and Betweenness Centrality of the street nodes. The calculation of centrality measurements has been introduced in the previous chapters. In this section the measurements of coverage level and number of clusters will be explained. After the new BRSs are placed in the proposed locations, the coverage level is recalculated again and compared with the one before the intervention.

7.2.1 Coverage level

The BRS coverage level of each neighborhood is measured by **the percentage of the nodes that are with at least one BRS within 300m from it**. The higher the percentage, the greater the coverage level.

i) Firstly, the computer program calculates how many BRSs are within 300m radius from a selected node. For the selected node, the more BRSs are located within 300m from the node, the higher the coverage level is. The process can be divided into the following steps.
   - **Step1:** Load street network data using the function of *osm.pdna_network_from_bbox* in the Python library of Pandana;
   - **Step2:** Download points of interest (POIs) and network data from OSM using the function of *osm.node_query from* in the Python library of Pandana;
   - **Step3**: Calculate accessibility by searching for the nearest POI to each node using the function of *pandana.network.nearest_pois* in the Python library of Pandana.

ii) After repeating the same calculation for all the nodes, we can count the number of nodes in each coverage level.

iii) This number is then divided by the total number of nodes and forms the coverage level.

For example, the results in Table 7.1 show that Überseequartier has the highest BRS coverage level among the selected exemplar neighborhoods because it has the largest percentage (69%) of the nodes with at least one BRS within 300m from them. In addition, it is also important to know the percentage of nodes with coverage level of 0 because it means there is no BRS within 300m and, therefore, this node does not have easy access to the BRS.

7.2.2 Number of clusters and clustering analysis

Next, we evaluate the allocation and distribution of BRS by the spatial clustering of BRS. The number of BRS clusters indicates whether the service areas of the BRS overlap with each other. As explained in the previous section, if the target is to avoid the overlap of the service areas, the corresponding strategy is to decrease the clusters of BRS.

This measurement is carried out by using Density-Based Spatial Clustering of Applications (DBSCAN), which is a non-parametric algorithm for density-based clustering. Given a set of BRS stations in a certain space, points with many nearby located stations are grouped together. If the stations are too far away from the nearest neighbors and lie in the low-density area, they are marked as outlier stations.

The following three parameters are required in the function of *sklearn.cluster.DBSCAN* in the Python library of scikit-learn to determine the cluster.

- **eps** defines “the maximum distance between two samples for one to be considered as in the neighborhood of the other” (Pedregosa, Varoquaux, Gramfort, & Michel, 2011). In the current study, we choose 300m as the radius - based on the study of Kabra et al. (2019).
- **min_samples** defines “the number of samples in a neighborhood for a point to be considered as a core point. This includes the point itself” (Pedregosa, Varoquaux, Gramfort, & Michel, 2011). In this study we choose 1 to be the minimum number of samples, which means if there are any two BRS that are within 300m will form a cluster.
- **metric** defines ”the metric to use when calculating distance between instances in a feature array” (Pedregosa, Varoquaux, Gramfort, & Michel, 2011) In this study we use “precomputed” as the metric.

The process of clustering BRS can be divided into the following steps.

- Step1: Download network data from OSM using the function of *osmnx.graph_from_point* in Python library of OSMnx
- Step 2: Retrieve the coordinate point of interest (POI) using the function of *osmnx.pois_from_point* in Python library of OSMnx
- Step 3: Attach nearest network node to each POI using the function of *osmnx.get_nearest_nodes* in Python library of OSMnx
- Step 4: Calculate the distances for each pair of nodes that have POI attached to them using function of *network_distance_matrix created* by Boeing (2018)
- Step 5: Cluster nodes using the function of *sklearn.cluster.DBSCAN* in the Python library of scikit-learn
- Step 6: Plot the POI and the cluster of POI using the Python library of *matplotlib*.

7.3 Results

7.3.1 Current situation and proposed location

In the previous chapter, we have identified the exemplar neighborhood of each cluster and ranked them based on their Place Values as indicated below.

- cluster 0-0 > cluster 0-2 > cluster 0-1
- cluster 1-3 > cluster 1-0 > cluster 1-2 > cluster 1-1
- cluster 2-1 > cluster 2-0

In this chapter, we propose to set up new BRS as an intervention to improve the Place Value. In order to understand the current BRS situation in each neighborhood, we examine the coverage level and the number of BRS clusters, as presented in table 7.1 and in (a) from Figure 7.1 to Figure 7.9.

Table 7.1 Strategy targets and the comparison between current and improved situations in the exemplar neighborhoods.

Cluster of stations	Exemplar neighborhood	Number of BRS	Number of BRS clusters	Coverage level: % of nodes with at least one BRS within 300m		Types of current situation and the proposed intervention strategy
				Before	After	
0-0	Überseequartier	10	2	69	69	TYPE I • Increase coverage level • Avoid increasing BRS clustering
0-1	Großhansdorf	0	0	0	40.5	TYPE III • Increase coverage level by allocating BRS in high C_c or C_b
0-2	Pinneberg	0	0	0	6.7	TYPE III • Increase coverage level by allocating BRS in high C_c or C_b
1-0	Berliner Tor	9	1	56.8	60.6	TYPE I • Increase coverage level • Avoid increasing BRS clustering
1-1	Klein Flottbek (Botanischer Garten)	1	0	9.4	10.9	TYPE III • Increase coverage level by allocating BRS in high C_c or C_b
1-2	Klosterstern	6	1	30	38.7	TYPE II • Increase coverage level by allocating BRS in high C_c or C_b • Avoid increasing BRS cluster
1-3	Steinstraße	17	3	67.5	72.9	TYPE I • Increase coverage level • Avoid increasing BRS clustering
2-0	Hammer Kirche	3	0	24.7	33.2	TYPE III • Increate coverage level by allocating BRS in high C_c or C_b
2-1	Hauptbahnhof Nord	15	3	65.1	74.9	TYPE I • Increase coverage level • Avoid increasing BRS clustering

The nine exemplar neighborhoods are categorized into three types based on their current situation and the proposed interventions, as presented in table7.1.

- TYPE I: Neighborhoods of this type are characterized by **high coverage level (higher than 50 % of nodes with at least one BRS within 300m) and at least one BRS cluster.** Furthermore, the existing BRSs are usually already allocated near the nodes with high Closeness Centrality and Betweenness Centrality. Therefore, the goals of the proposed intervention are increasing the coverage level, while at the same time avoiding increased BRS clustering (i.e., increasing the density of the existing BRS clusters). The other noticeable pattern is that this type of neighborhoods is also usually the ones with the highest Place Value in each Node-Value-cluster. Neighborhoods belonging to this type are Überseequartier (cluster 0-0), Berliner Tor (cluster 1-0), Steinstraße (cluster 1-3), and Hauptbahnhof Nord (cluster 2-1).
- TYPE II: Neighborhoods of this type are characterized by **low coverage level (lower than 50 % of nodes with at least one BRS within 300m) and at least one cluster of BRS**. Therefore, the target of the proposed intervention is increasing the coverage level by allocating BRSs near the nodes with high Closeness Centrality or Betweenness Centrality, while avoiding increased BRS clustering. The only neighborhood belonging to TYPE II is Klosterstern (cluster 1-2), which has the lower level of Place Value in Node-Value-cluster-1.
- TYPE III: Neighborhoods of this type are characterized by **low coverage level (lower than 50 % of nodes with at least one BRS within 300m) and no BRS cluster**. In fact, some of these neighborhoods do not have any BRS. Therefore, the target of the proposed intervention is increasing the coverage level by locating BRSs near the nodes with high Closeness Centrality or Betweenness Centrality. Neighborhoods belonging to TYPE III are also usually the ones with the lowest Place Values in each Node-Value-cluster. These neighborhoods are Großhansdorf (cluster 0-1) and Pinneberg (cluster 0-2), Klein Flottbek (Botanischer Garten) (cluster 1-1) and Hammer Kirch (cluster 2-0).

With regard to the location of the demand points, we map the spatial distribution of ten nodes with highest values of Closeness and Betweenness Centrality in order to determine the optimal locations for new BRSs. The other purpose of this step is to allow us to evaluate whether the existing BRSs are located near the nodes with high Closeness or Betweenness Centrality. In addition, the proposed locations for setting up new BRSs should meet at least one of the following criteria:

- most accessible from all street nodes, i.e. node with high Closeness Centrality
- most intermediate or highest movement opportunities between any random pair of street nodes, i.e. node with high Betweenness Centrality
- avoid creating a new cluster

The results are presented in (b) and (c) from Figure 7.1 to Figure 7.9.

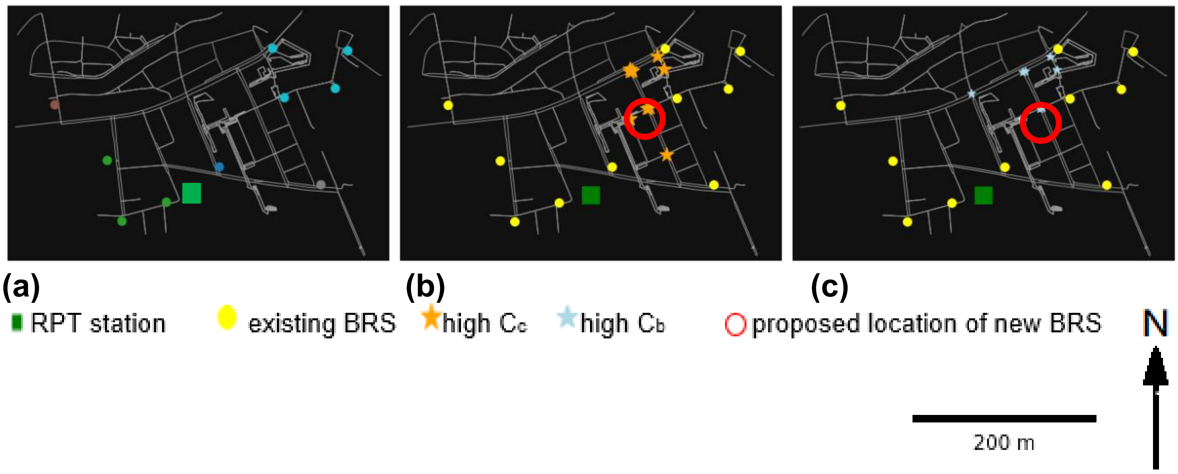


Figure 7.1 Spatial distribution of RPT station, existing BRS and clusters, and ten nodes with highest C_c and C_b in Überseequartier (cluster 0-0).
(proposed new BRS: 53°54'31.8"N 10°00'08.54"E)
(a) spatial distribution of BRS and number of clusters[17];
(b) spatial distribution of RPT station, existing BRS, and ten nodes with highest C_c
(c) spatial distribution of RPT station, existing BRS, and ten nodes with highest C_b

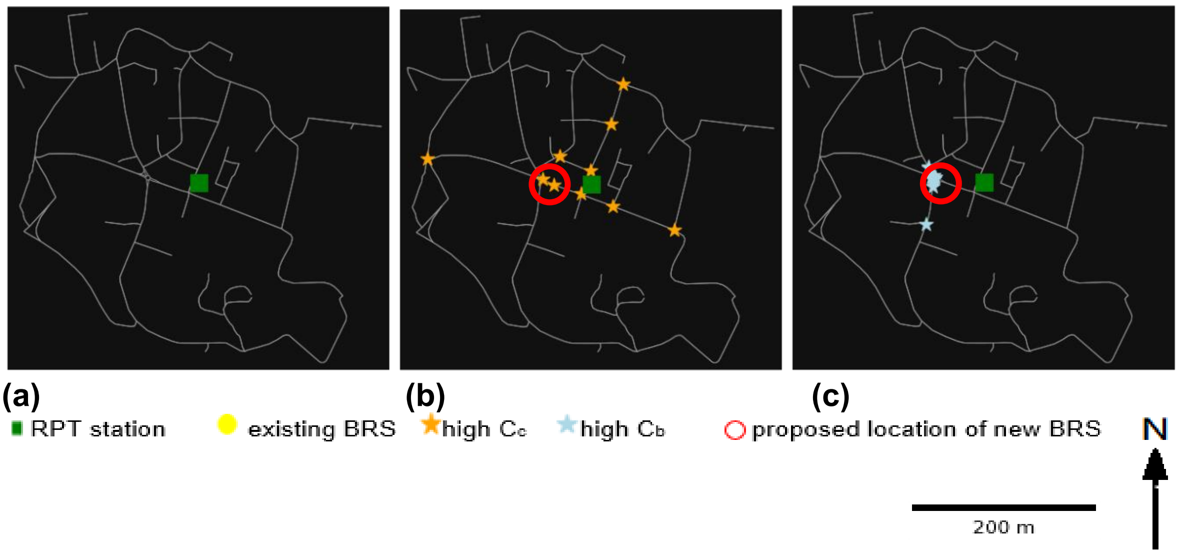


Figure 7.2 Spatial distribution of RPT station, existing BRS and clusters, and ten nodes with highest Cc and Cb in Großhansdorf (cluster 0-1).
(proposed new BRS: 53°66'26.45"N 10°28'36.13"E)
(a) spatial distribution of BRS and number of clusters
(b) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cc
(c) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cb

[17] BRS of the same cluster group are in the same color.

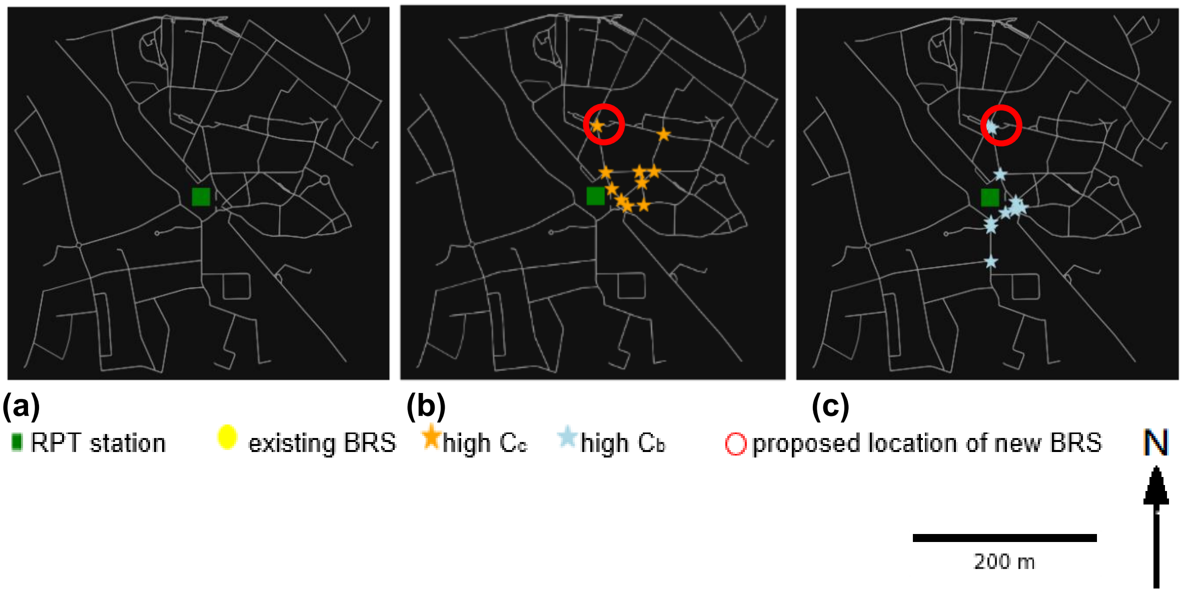


Figure 7.3 Spatial distribution of RPT station, existing BRS and clusters, and ten nodes with highest Cc and Cb in Pinnerberg (cluster 0-2).
(proposed new BRS: 53°65'71.46"N 9°79'88.68"E)
(a) spatial distribution of BRS and number of clusters
(b) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cc
(c) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cb

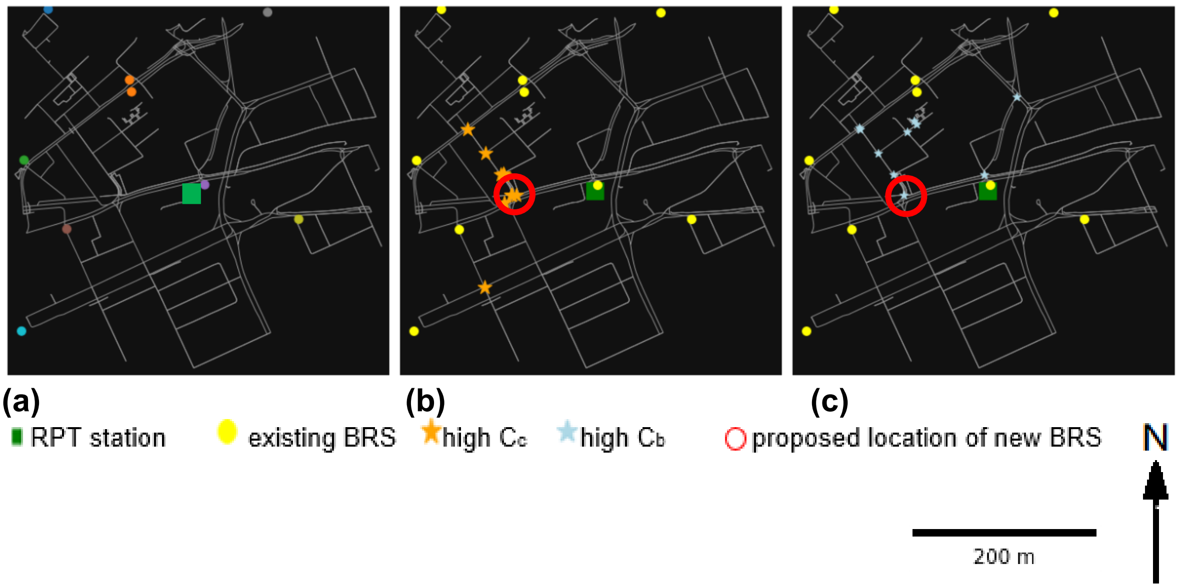


Figure 7.4 Spatial distribution of RPT station, existing BRS and clusters, and ten nodes with highest Cc and Cb in Berliner Tor (cluster 1-0).
(proposed new BRS: 53°55'29.63"N 10°01'92.12"E)
(a) spatial distribution of BRS and number of clusters
(b) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cc
(c) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cb

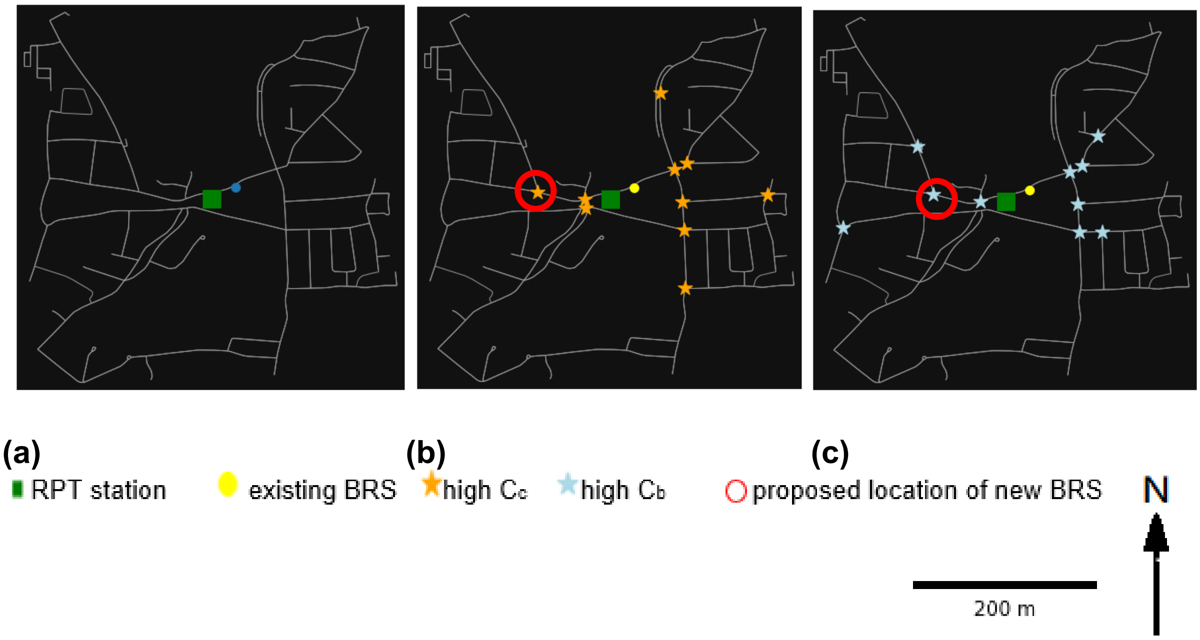


Figure 7.5 Spatial distribution of RPT station, existing BRS and clusters, and ten nodes with highest Cc and Cb in Klein Flottbeck (cluster 1-1).
(proposed new BRS: 53°55'83.2"N 9°85'60.49"E)
(a) spatial distribution of BRS and number of clusters
(b) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cc
(c) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cb

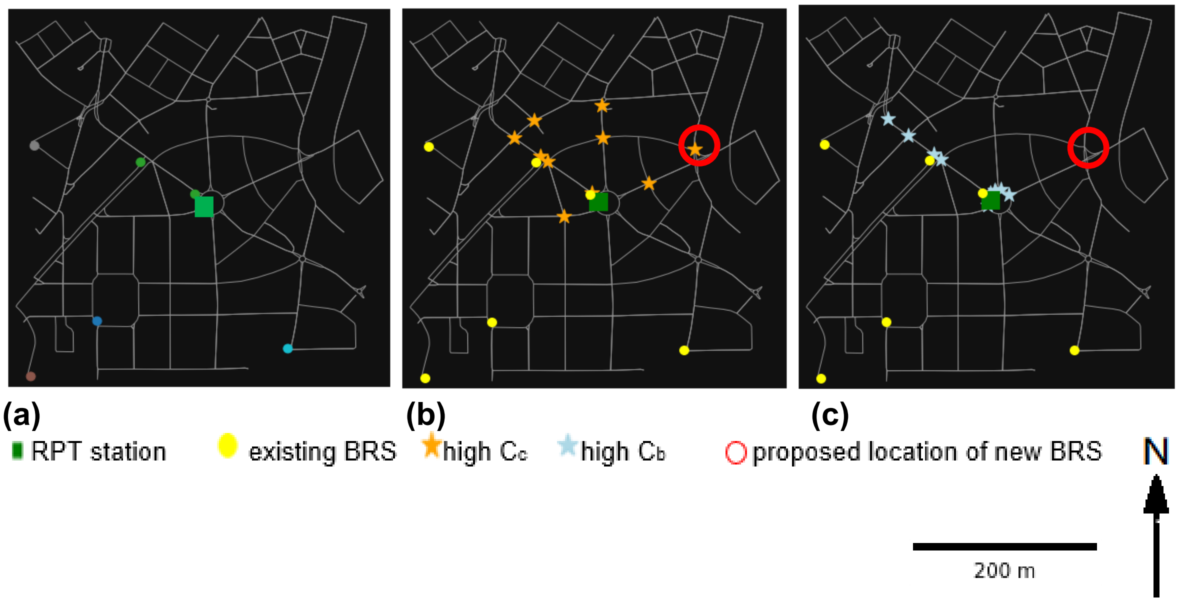


Figure 7.6 Spatial distribution of RPT station, existing BRS and clusters, and ten nodes with highest Cc and Cb in Klosterstern (cluster 1-2).
(proposed new BRS: 53°58'32.16"N 9°99'37.34"E)
(a) spatial distribution of BRS and number of clusters
(b) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cc
(c) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cb

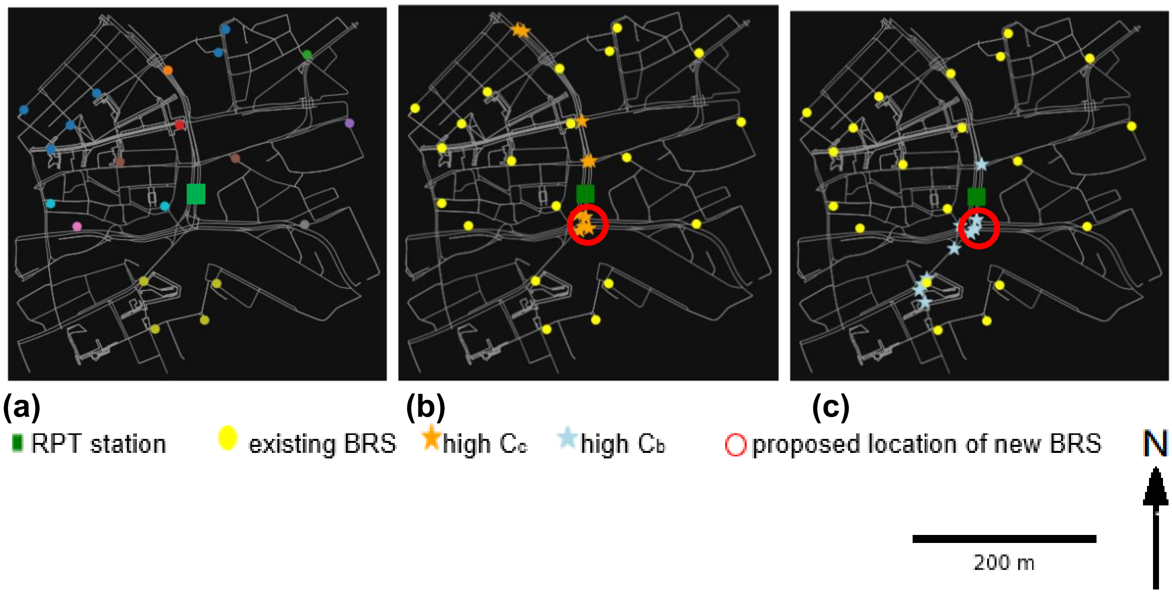


Figure 7.7 Spatial distribution of RPT station, existing BRS and clusters, and ten nodes with highest Cc and Cb in Steinstraße (cluster 1-3).
(proposed new BRS: 53°54'78.22"N 10°00'60.73"E)
(a) spatial distribution of BRS and number of clusters
(b) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cc
(c) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cb

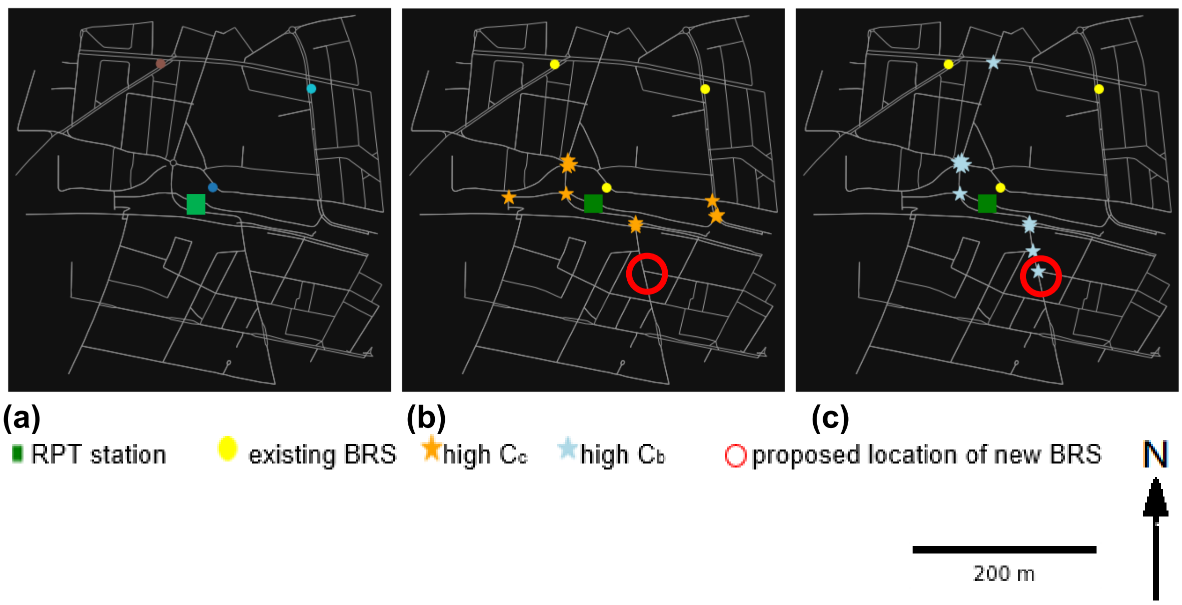


Figure 7.8 Spatial distribution of RPT station, existing BRS and clusters, and ten nodes with highest Cc and Cb in Hammer Kirche (cluster 2-0).
(proposed new BRS: 53°55'40.57"N 10°05'76.51"E)
(a) spatial distribution of BRS and number of clusters
(b) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cc
(c) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cb

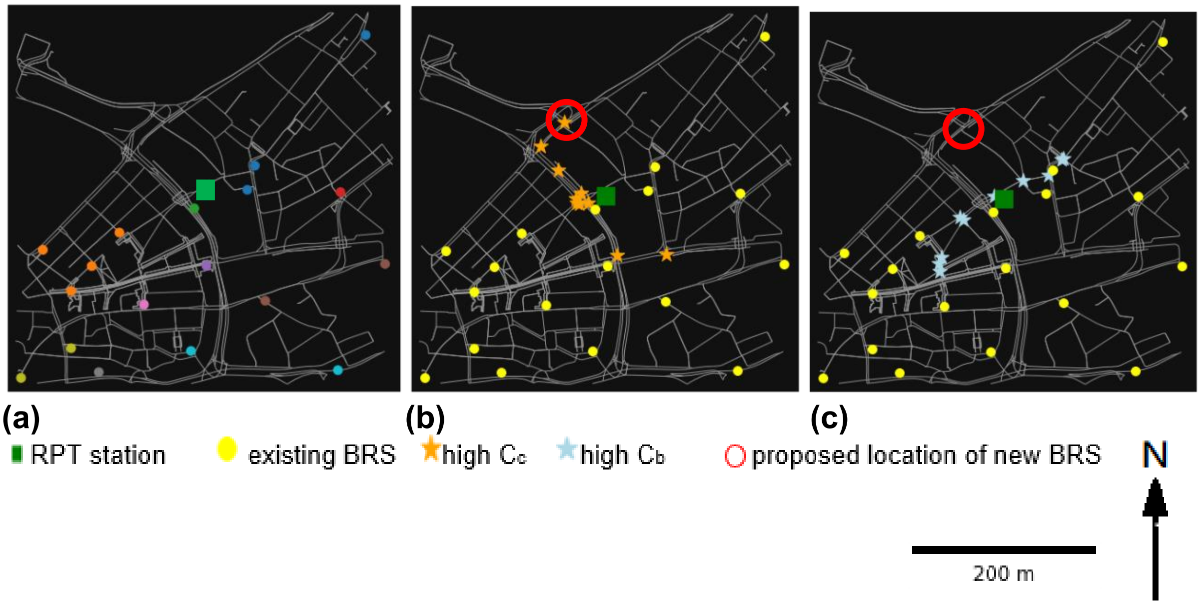


Figure 7.9 Spatial distribution of RPT station, existing BRS and clusters, and ten nodes with highest Cc and Cb in Hauptbahnhof Nord (cluster 2-1).
(proposed new BRS: 53°55'67.76"N 10°00'26.64"E)
(a) spatial distribution of BRS and number of clusters
(b) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cc
(c) spatial distribution of RPT station, existing BRS, and ten nodes with highest Cb

7.3.2 Evaluating the improvement of the intervention

After the new BRSs are placed in the proposed locations, the coverage level is recalculated again and compared with the one before the intervention. The results before and after the recalculation are presented in Table 7.1, Table 7.2, Figure 7.10 and Figure 7.13 to Figure 7.21.

1) Comparing the coverage level before and after the intervention.

The coverage ratio is compared by two measurements. Firstly, Table 7.1shows **the percentage of nodes with at least one BRS within 300m** before and after the intervention. The biggest improvement of the coverage level after the intervention appears in TYPE III neighborhoods. By adding just one additional BRS, the percentage of the nodes that have access to at least one BRS increases from 0% to 6.7% in Pinneberg, 35.3% in Klein Flottbek, 33.2% in Hammer Kirche and 40.5% in Großhansdorf. The smallest improvement of the coverage level after the intervention appears in TYPE I neighborhoods, where the coverage level is already high before the introduction of the intervention.

Secondly, Table 7.2 and Figure 7.10 to Figure 7.12 show **the percentage of nodes with one, two, three or four BRS within 300m**. The most obvious change can be observed in TYPE I neighborhoods. In Überseequartier, the percentage of nodes that have three BRS within 300m increases from 8% to 13.1%. In Berliner Tor, the percentage of nodes that have two BRS within 300m increases from 9.8% to 20.1%. The biggest improvement is in Steinstraße, where the percentages of nodes that have one, two, three and four BRS within 300m all increase after the intervention. In Hauptbahnhof Nord, the percentages of nodes that have one and two BRS within 300m increase after the intervention.

Table 7.2 Comparison of the coverage level of BRS before and after the intervention in each neighborhood.

		TYPE I				TYPE II				TYPE III			
		Before		After		Before		After		Before		After	
		Number of nodes	%	Number of nodes	%	Number of nodes	%	Number of nodes	%	Number of nodes	%	Number of nodes	%
Coverage level		Überseequartier				Klosterstern				Großhansdorf			
	0	78	**31.1**	78	**31.1**	179	**69.9**	157	**61.3**	0	**100**	47	**59.5**
	1	105	**41.8**	96	**38.2**	66	**25.8**	86	**33.6**	0	**0**	32	**40.5**
	2	48	**19.1**	44	**17.5**	11	**4.3**	13	**5.1**				
	3	20	**8.0**	33	**13.1**								
Total		251	**100**	251	**100**	256	**100**	256	**100**	0	**100**	79	**100**
Coverage level		Berliner Tor								Pinneberg			
	0	146	**43.2**	133	**39.3**					0	**100**	112	**93.3**
	1	158	**46.7**	136	**40.2**					0	**0**	8	**6.7**
	2	33	**9.8**	68	**20.1**								
	3	1	**0.3**	1	**0.3**								
Total		338	**100**	338	**100**					0	**100**	120	**100**
Coverage level		Steinstraße								Klein Flottbek (Botanischer Garten)			
	0	194	**32.4**	162	**27.1**					96	90.6	95	86.4
	1	227	**38.0**	233	**39.0**					10	9.4	12	10.9
	2	140	**23.4**	154	**25.8**					0	0	3	2.7
	3	37	**6.2**	47	**7.9**								
	4	0	**0**	2	**0.3**								
Total		598	**100**	598	**100**					0	**100**	190	**100**
Coverage level		Hauptbahnhof Nord								Hammer Kirche			
	0	199	**34.9**	143	**25.1**					203	**73.3**	185	**66.8**
	1	208	**36.5**	256	**44.9**					74	**26.7**	92	**33.2**
	2	105	**18.4**	113	**19.8**								
	3	58	**10.2**	58	**10.2**								
Total		570	**100**	570	**100**					277	**100**	277	**100**

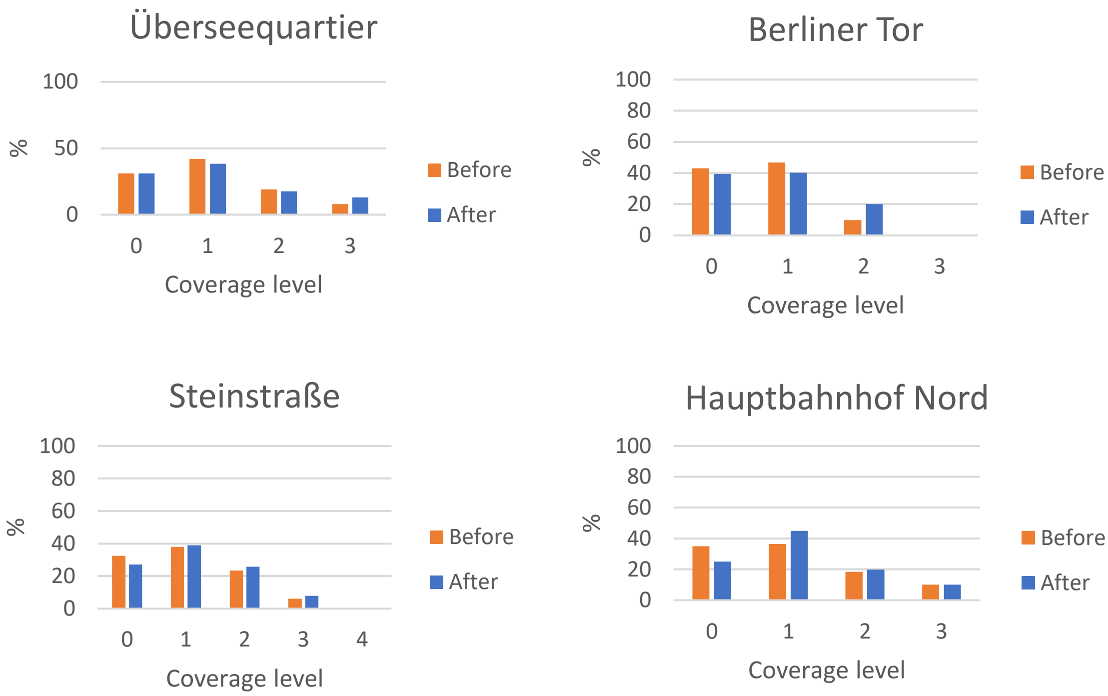


Figure 7.10 Comparing the coverage level before and after the intervention in TYPE I neighborhoods.

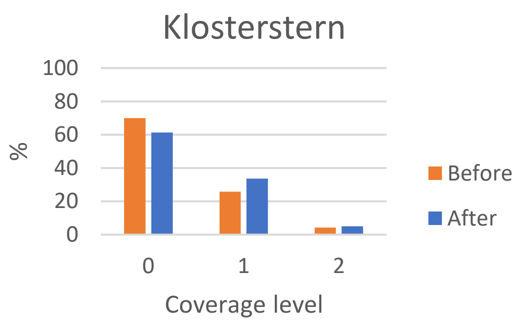


Figure 7.11 Comparing the coverage level before and after the intervention in TYPE II neighborhoods.

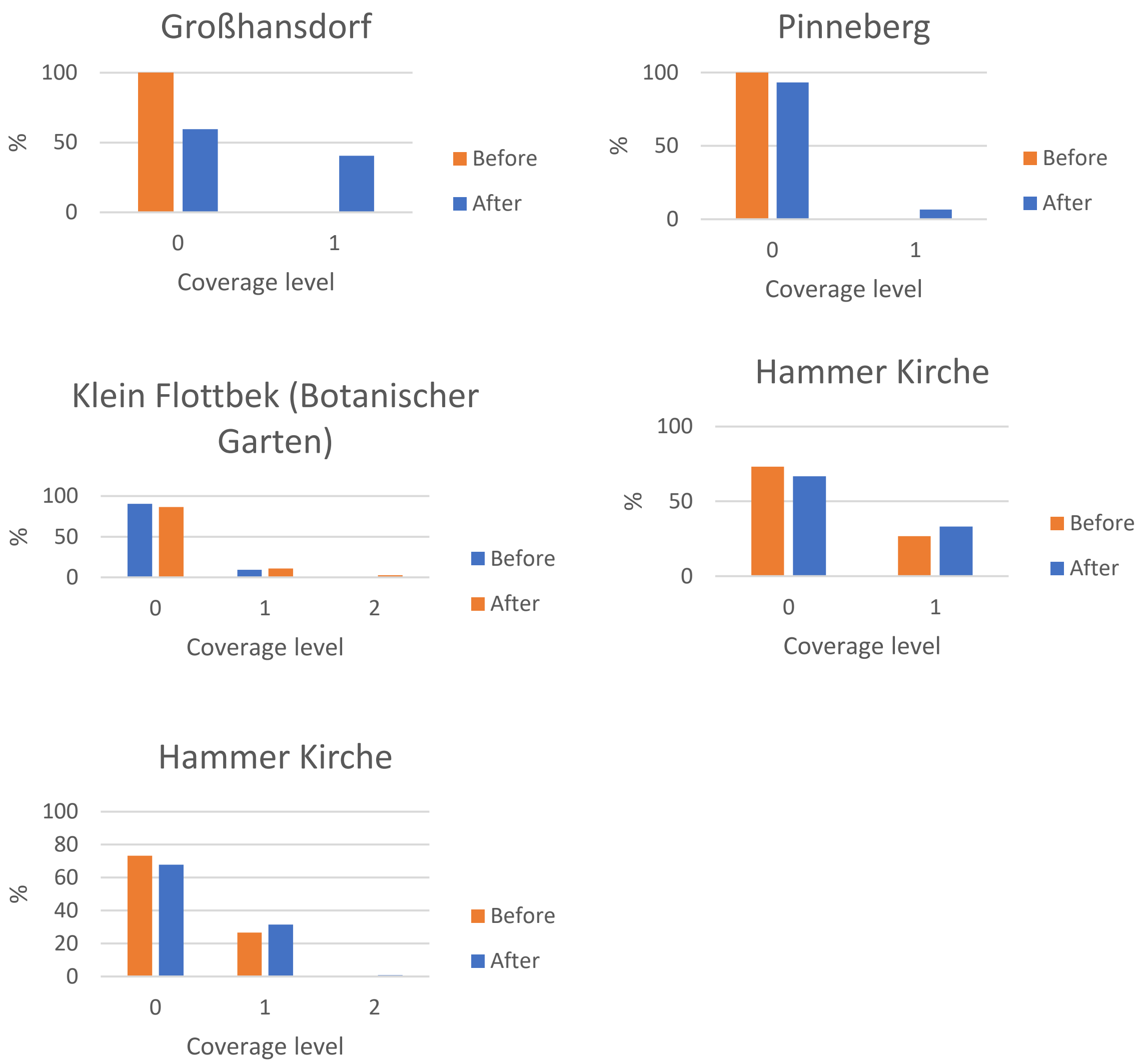


Figure 7.12 Comparing the coverage level before and after the intervention in TYPE III neighborhoods.

2) Spatial distribution of the coverage level of each node before and after the intervention

In order to compare the changes before and after the intervention, nodes are colored by the coverage levels and mapped out in Figure 7.13 to Figure 7.21. It can be observed that after the intervention the darker colored nodes (i.e., nodes with easier access to BRS) become more numerous and their spatial distribution changes quite significantly.

i)TYPE I

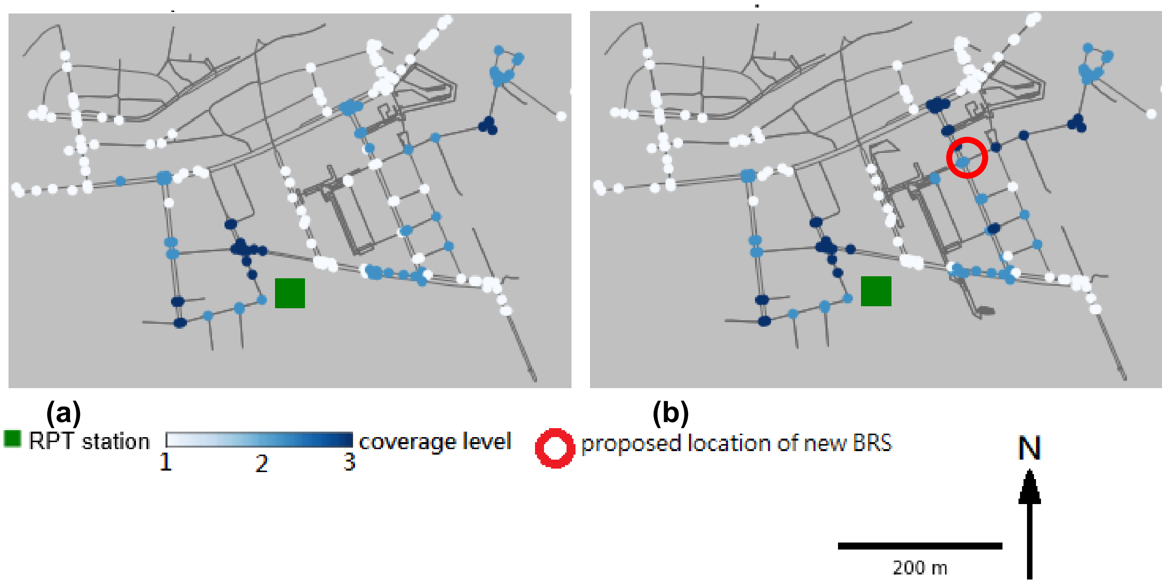


Figure 7.13 Überseequartier: Spatial distribution of the coverage level of each node (a) before and (b) after the intervention.

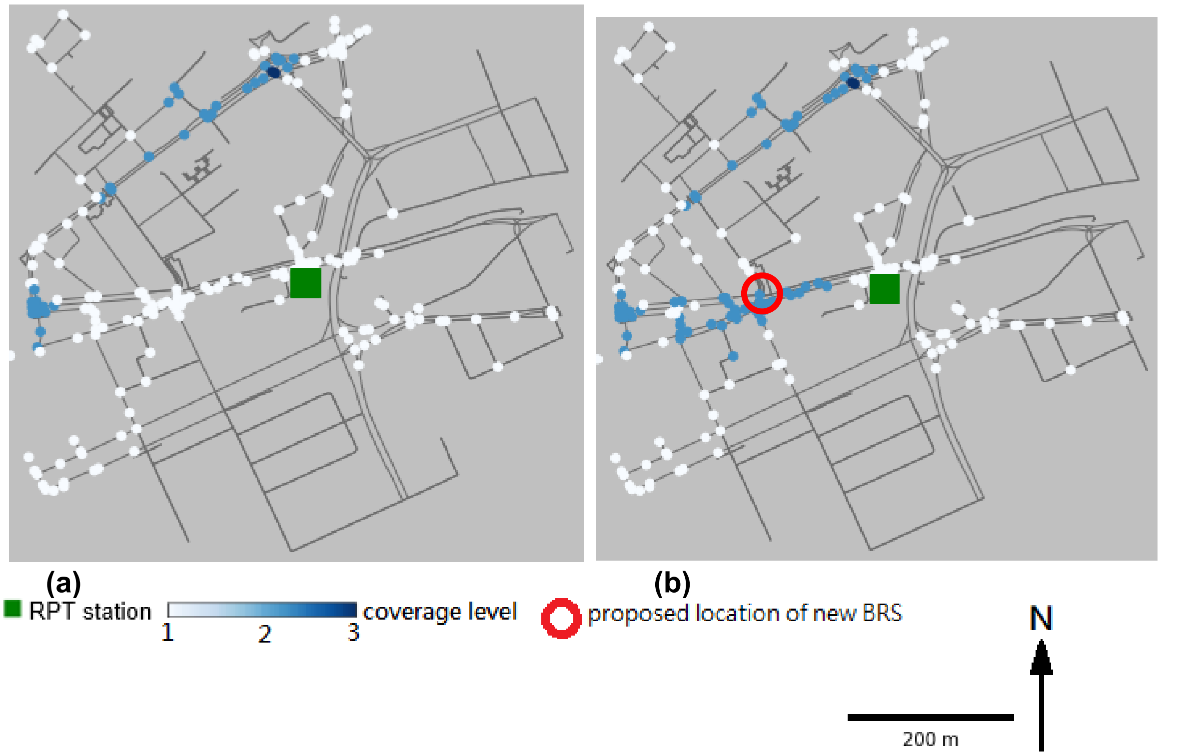


Figure 7.14 Berliner Tor: Spatial distribution of the coverage level of each node (a) before and (b) after the intervention.

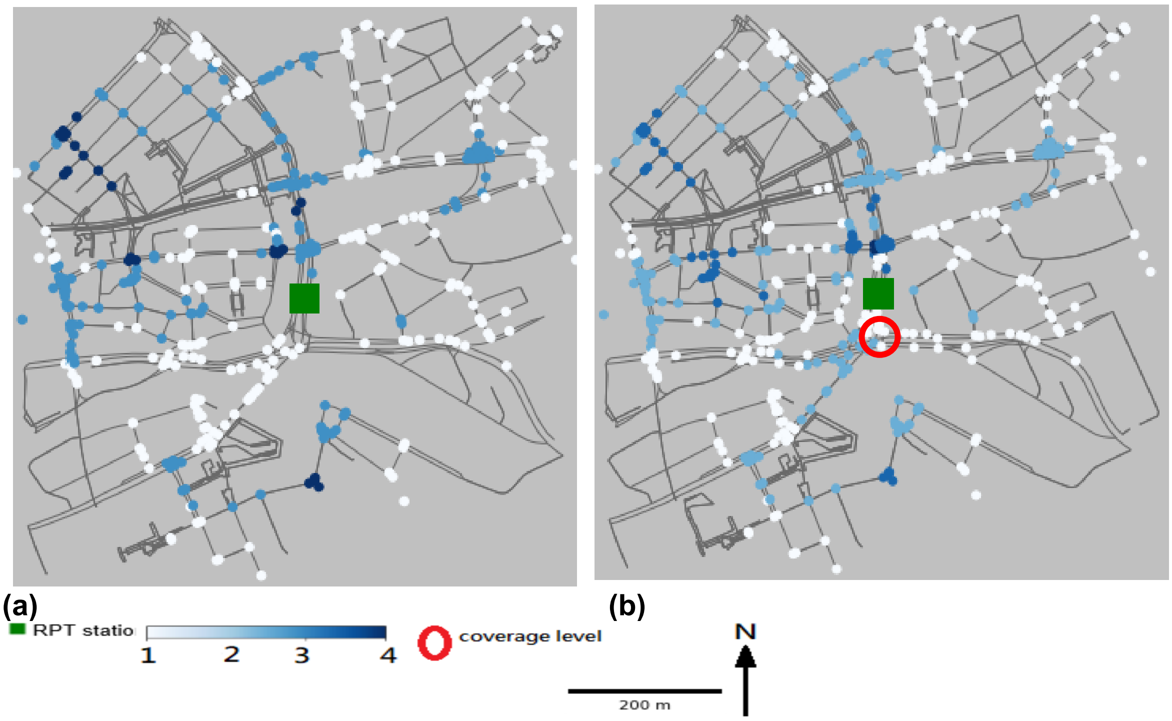


Figure 7.15 Steinstraße: Spatial distribution of the coverage level of each node (a) before and (b) after the intervention.

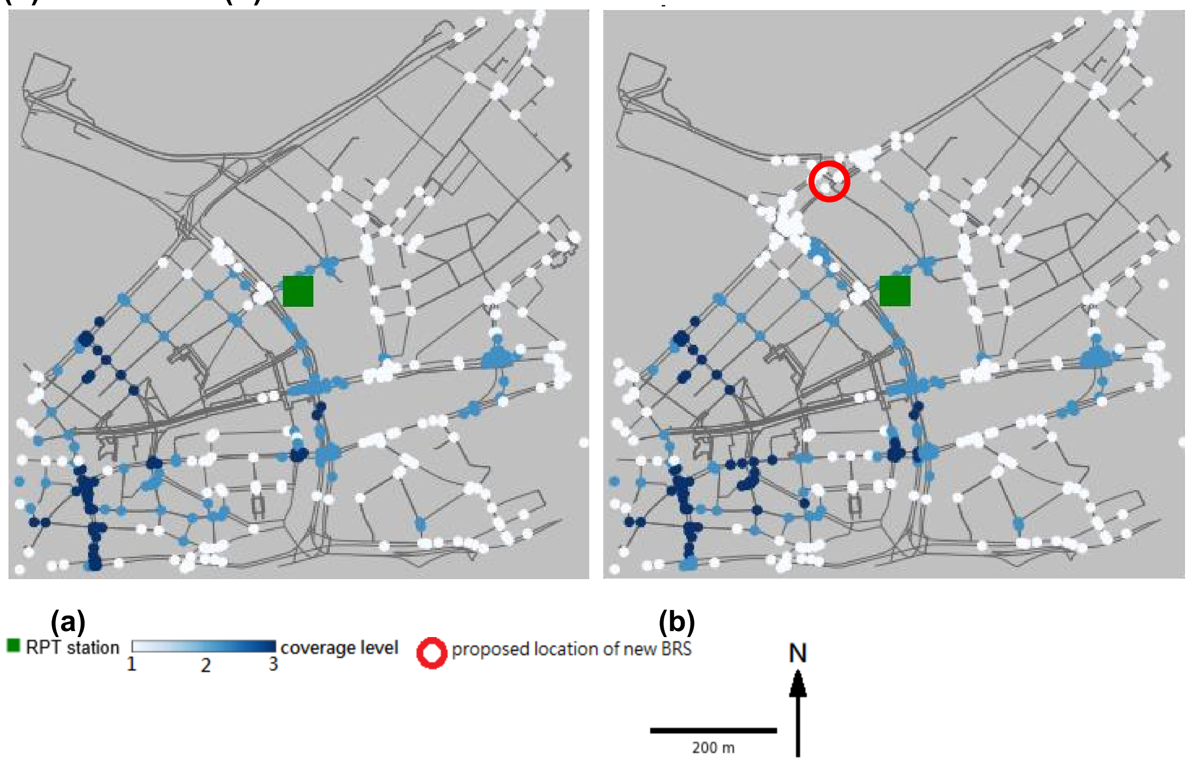


Figure 7.16 Hauptbahnhof Nord: Spatial distribution of the coverage level of each node (a) before and (b) after the intervention.

ii)TYPE II

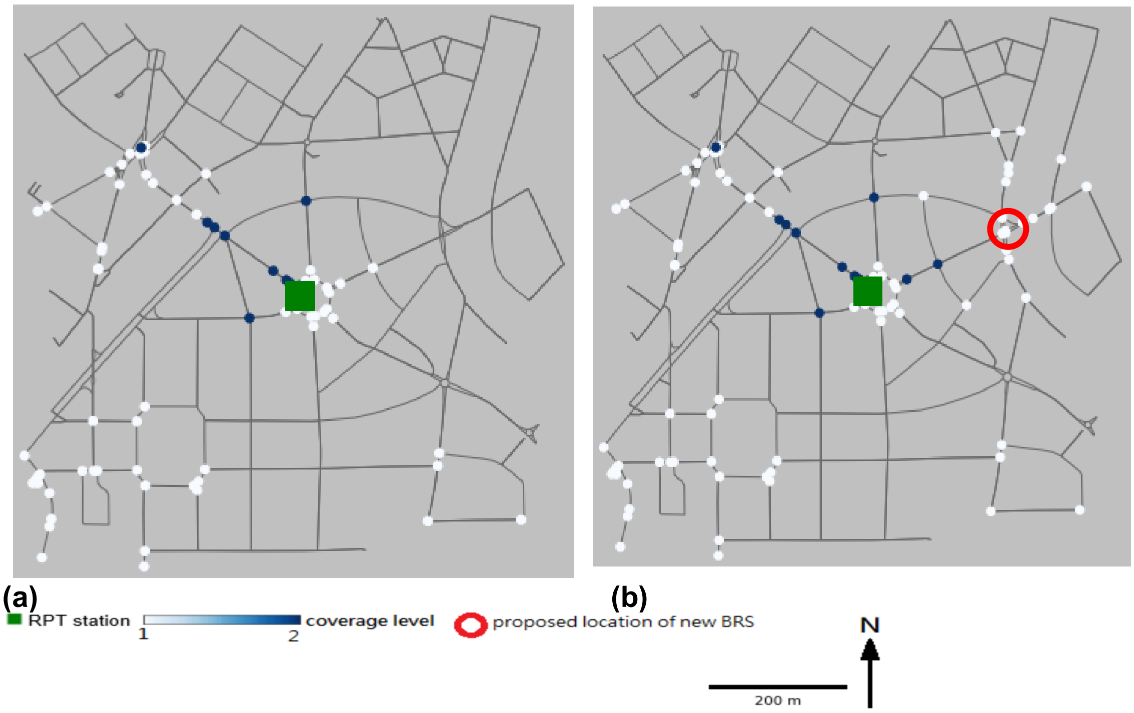


Figure 7.17 Klosterstern: Spatial distribution of the coverage level of each node (a) before and (b) after the intervention.

iii)TYPE III

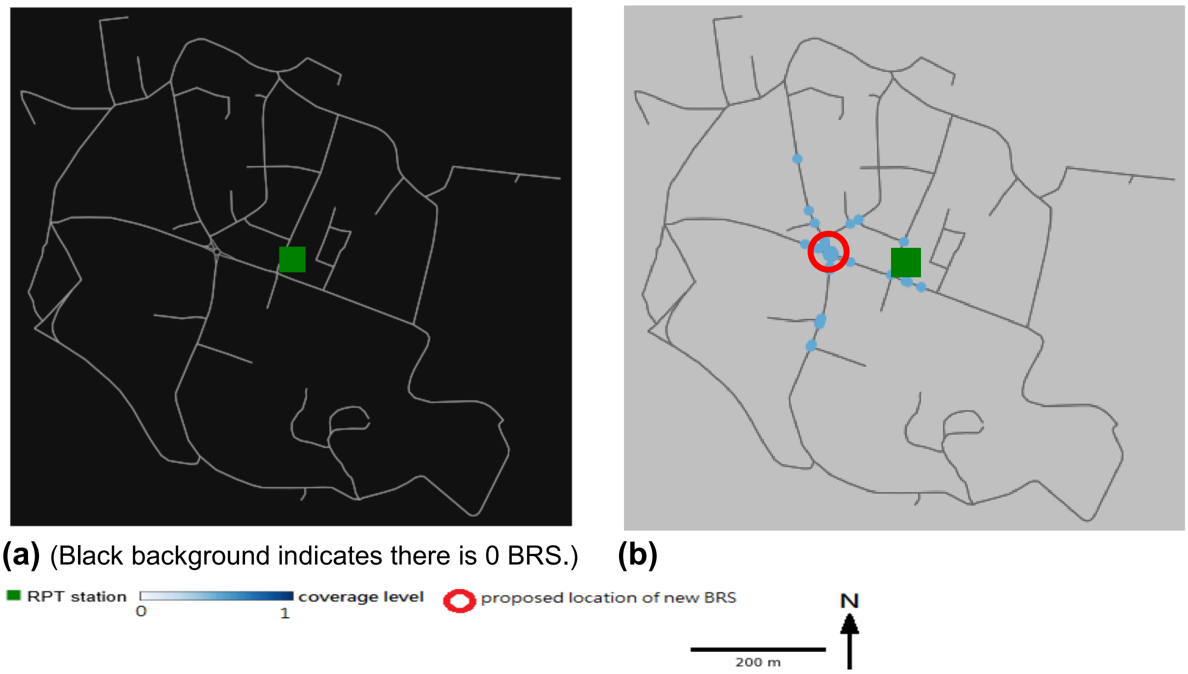


Figure 7.18 Großhansdorf: Spatial distribution of the coverage level of each node (a) before and (b) after the intervention.

Figure 7.19 Pinneberg: Spatial distribution of the coverage level of each node (a) before and (b) after the intervention.

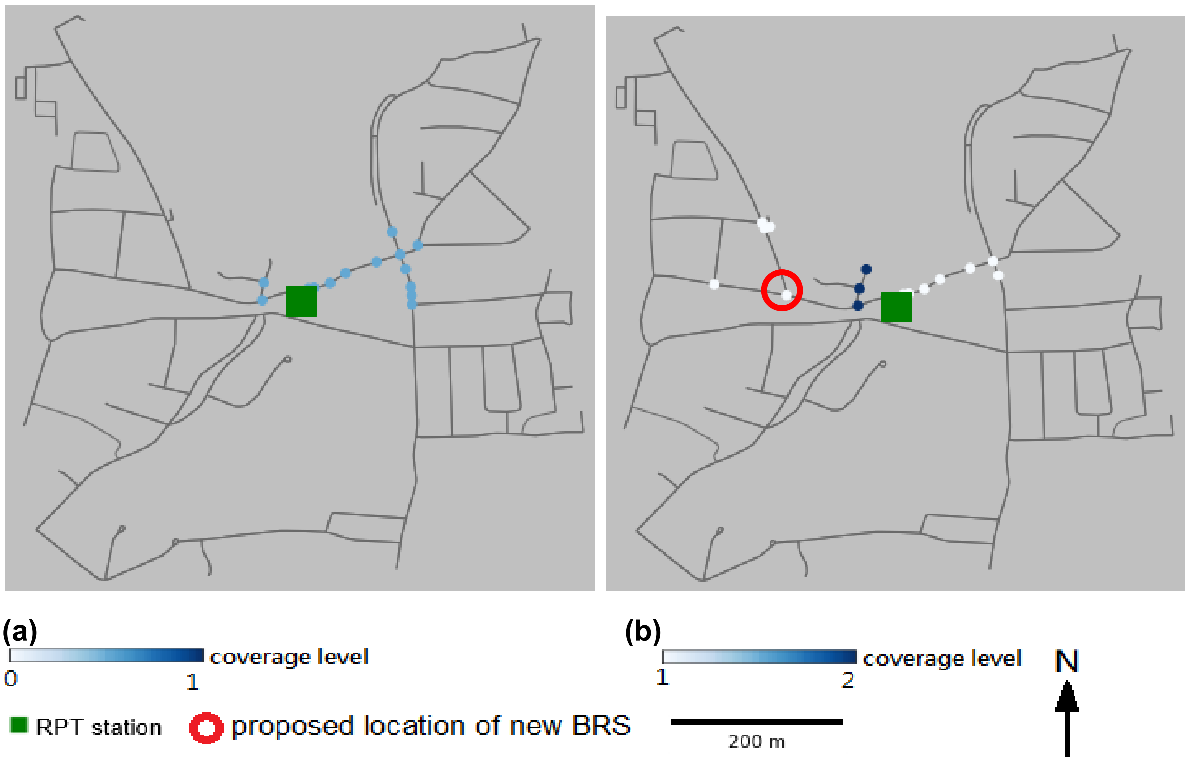


Figure 7.20 Klein Flottbek (Botanischer Garten): Spatial distribution of the coverage level of each node (a) before and (b) after the intervention.

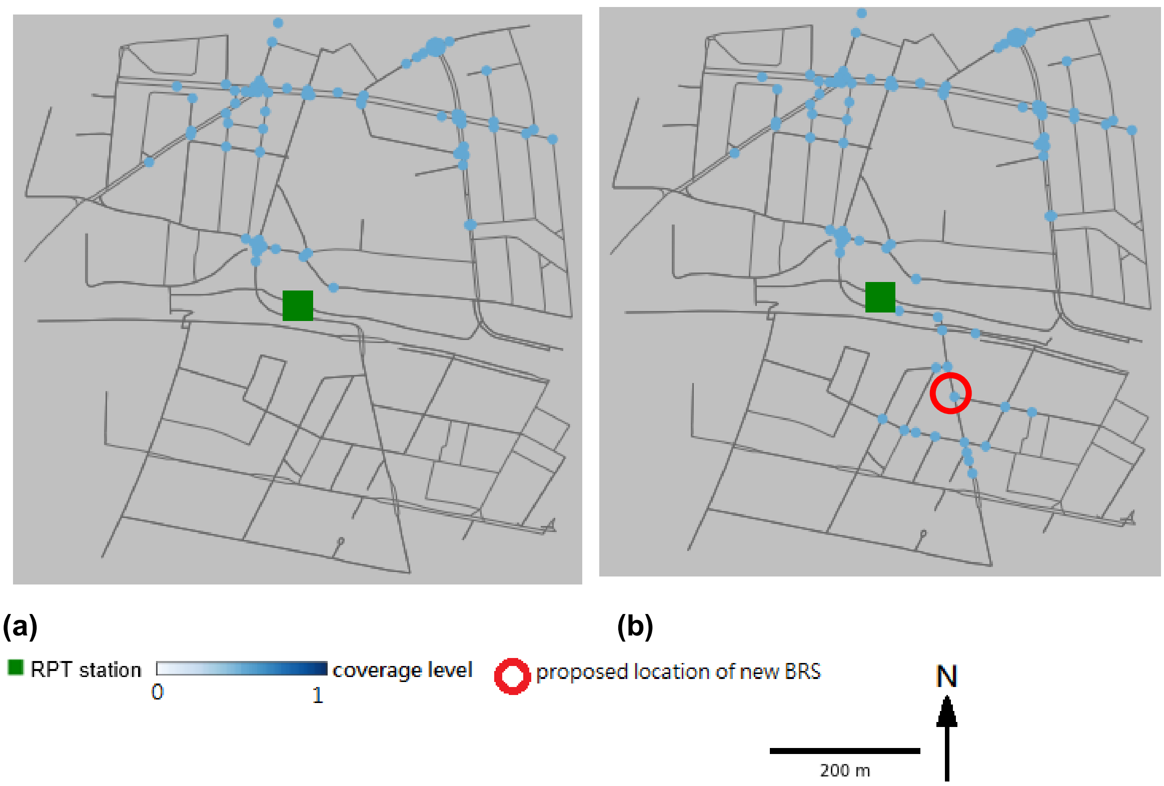


Figure 7.21 Hammer Kirche: Spatial distribution of the coverage level of each node (a) before and (b) after the intervention.

7.4 Conclusion and discussion

This chapter aims at proposing a method to strategically and efficiently select the location for allocating new facilities in order to improve the Place Value of the TOD neighborhood. As shown above, this method allows comparing the coverage level and the number of clusters before and after the intervention as well as comparing different types of interventions. Because this study is just a demonstration of the procedure, we only use BRS as an example. There are, of course, many other types of interventions that can be tested by using the same approach. For example, retail business can use it in order to choose the optimal location for a new shop. Naturally, if the target facility is changed, the problems that need to be solved and the relationship between target facility and the demand point will have to be reconsidered and redefined as well.

Also, in the current study we only increase one new BRS and choose the location based on only three criteria. Depending on the budget, there can be multiple BRS allocated in the neighborhood, instead of just one BRS.

In the current study the relationship between the BRSs is not competition but rather complementarity. However, different types of service, amenity, facility may have different relationships, which affects the allocation strategies. For example, “in facilities such as convenience stores, gas stations, and supermarkets, there are some class groups like brands, chains, affiliates, etc., and facilities in the same group function cooperatively with each other, and facilities in different groups function competitively. When opening a new facility, the location is desirable if it acquires more customers from the facilities of contending groups rather than those of the same group” (Fushimi, Okubo, & Saito, 2020).

Furthermore, <u>the criteria for choosing the location can also be expanded</u>. For example, new BRS can be located near other POI in the neighborhood. If we take mobile clinics as an example, we might discover that schools work well as the locations of mobile clinics because residents can easily find them and because they have been used as mobile clinic sites before. In such a case, this method can be used in order to find out which schools will be the best for meeting the demand.

In addition, this approach can be used for investigating whether the coverage level would be higher if the new BRS is allocated in places with high Closeness Centrality or high Betweenness Centrality. As discussed in previous chapters, each centrality measurement meets different goal:

- Locating new BRS at the node with high Closeness Centrality means that it will be close to the largest number of other nodes and therefore **more accessible**.
- Locating new BRS at the node with high Betweenness Centrality means that it will be placed in the location with **high passing-by frequency**.

It may be interesting to find out which centrality measure would result in the highest coverage level with the smallest number of BRS.

The most obvious weakness of the current study is that we use nodes as the demand points while the distribution of socio-demographic background of the residents is not taken into consideration. It can be argued that being accessible by more street nodes does not necessarily mean that a BRS is more accessible for those users who need it the most. However, the composition of inhabitants may be changing frequently because the residents may move in and out of the neighborhood. Socio-demographic background of the residents is, therefore, a more dynamic indicator than the built environment and the urban components, such as the street network. Therefore, using the centrality measurements of the nodes may not be a bad idea, especially if this approach is used in combination with other types of analysis, e.g. the examination of socio-demographic data where they are available. This research only aims to serve as an experiment for demonstrating how measurements like centrality may contribute to the location-allocation analysis.

Finally, being motivated by the outbreak of COVID-19 in 2020, we hope that this research might be useful for planners by helping them to make more informed and data-driven decisions on the location of new facilities like temporary test stations or mobile immunization centers during the pandemics and other similar situations.

8. Conclusion and future work

This thesis applies network theory to urban planning with a special emphasis on the Transit Oriented Development (TOD) and proposes a parametric method that aims at classifying TOD neighborhoods and determining suitable locations for new facilities. The theories and concepts of Node-Place model and Transit Oriented Development are reviewed and discussed from the perspectives of network science. With the computational application of the network analysis, this study aims at answering the following questions.

- How to decide the size and location of the catchment area in order to mitigate the size and placement effects on the indicators of network analysis?
- What is the level of connectivity and resilience of the public transit network and the street network in Hamburg?
- How can one quantitatively classify and evaluate the importance of transportation stations in the railway public transport system?
- How to determine a suitable location for a new facility or an intervention that matches the characteristics and challenges of a particular type of neighborhood?

This chapter summarizes the results of the previous chapters, highlights the findings and contributions made in this study, discusses the limitations of the current research and recommends the future research directions.

8.1 Summary of results

8.1.1 Topological characteristics of the public transportation network in Hamburg

Chapter 2 employs network analysis to examine the topological characteristics and properties of Hamburg metropolitan area's public transit network (Hamburger Verkehrsverbund, HVV), including geometric properties, small-world properties and scale-free properties of the transit network. Based on the analysis of the topological properties, that chapter examines the level of robustness and resilience to disruption.

First of all, the examination of the geometric properties, which are measured by the number of nodes, number of links, link-node ratio and the Degree of Connectivity, shows that the HVV network does not facilitate easier movement between stations than does the theoretical reference network or the Soul transportation network.

Secondly, the small-world properties, which are measured by the average path length and Clustering Coefficient, cannot be identified in the HVV network. This indicates that the connectivity and robustness of the HVV network have low level of resilience in case of disruptions. In other words, if one of the stations stops functioning for some reason, the speed and ease of getting to a destination can be severely affected.

Thirdly, in terms of the scale-free properties, which are measured by average and cumulative distribution of the Node Degree, it is revealed that the HVV network is vulnerable to significant disruptions if one of the stations with a large number of links stops operating, e.g. due to construction or maintenance work.

In that chapter, we have shown that the outcome of the analysis of the structural composition can help to evaluate the level of fault tolerance based on the spatial configuration of the public transit network. These results can be useful for the future enhancement of mobility, for example through the strategic creation or relocation of stations or through the construction of more links.

8.1.2 Classifying stations by Node Value

In chapter 3, the goal is to assign each station with its Node Value in order to construct the Node-Place Model. Clustering analysis is used in order to group the railway public transport stations based on the values of their Degree Centrality, Closeness Centrality, Betweenness Centrality and service frequency. Using k-means clustering analysis, three clusters of stations are identified and characterized by mapping their spatial distribution and presenting the statistical results of the indicators' values.

The results show that cluster 2 is predominantly located in the inner city and have the highest Node Value. Cluster 1 is mainly distributed in both inner and middle suburbs. Cluster 1 has the second highest Node Value. Cluster 0 is mainly distributed in the middle and outer suburbs. Cluster 0 has the lowest Node Value. The Node Value defined in this chapter is the first step in constructing the Node-Place Model and it also provides a foundation for the later chapter, where each cluster will be further classified by their Place Value.

8.1.3 Size effect in the evaluation of network characteristics

Before carrying out the empirical street network analysis in the neighborhoods of the individual stations, there are a few challenges that need to be dealt with. Assessment of an urban network requires the determination of the center point and the catchment area around it, and the size of the catchment area exerts significant influence on the values of the network analysis indicators. This influence is referred to as the size effect. Chapter 4 investigates if and to what extent the size of the catchment area influences the indicators' values. This chapter also aims to provide a guideline for determining the appropriate size for the street network analysis.

A methodology for determining the size of the catchment area is proposed, and recommendations for its appropriate size for pedestrian network analysis are offered. For the mathematical deduction of the relationship between the size of the catchment area and the indicator values, an idealized regular network is used as a reference model. The results show that:

- The size effect on the indicator values is prominent.
- The size effect decreases as the size of the catchment area increases.
- The size effect is notable until the side length of the catchment area that equals 50 times the average street length, while the variation is still acceptable if it is 30 to 40 times the average street length.
- The results provide the evidence to support that the lower and upper limits of the size of the catchment area should be 2000x2000m^2 and 4000x4000m^2. Since the typical average street length in cities is about 100 meters, a catchment area larger than 4000 x 4000m^2 is not necessary.
- The reference model proposed in chapter 4 can also be used to evaluate real networks and to compare them with an idealized case.

The results in that chapter contribute to the argument that a catchment area with an area size that is too large would not be practical because the characteristics and patterns of the street networks in a real city may vary significantly from quarter to quarter. For example, the characteristics of the network in the historical center would be different from those in its surrounding areas. Therefore, if the size of the catchment area is too big, the values of the indicators would reflect not the information about one type of network but rather a sum of multiple types of networks in several neighboring and connected quarters. Such mixed information would be less valuable for investigating the relationship between the street network structure and the indicators or for classifying the street networks. Based on these results, we suggest that in all future investigations the size of the catchment area should be defined before carrying out further analysis.

8.1.4 Placement effect and Closeness Centrality of urban street network

The second challenge before carrying out the empirical street network analysis in the neighborhoods of the individual stations is the placement effect, which means that the measurement of the network analysis, such as Closeness Centrality, depend on the placement of the catchment area. In other words, the location of the catchment area, which is a sample of the entire network, exerts significant influence on the value of the indicator. This effect becomes even more significant when multiple catchment areas are sampled to be compared and classified.

In chapter 5, we examine the placement effect on one of the most affected indicators, Closeness Centrality, and propose a method for choosing the location of the catchment area if the Closeness Centrality of different nodes has to be compared.

Chapter 5 proposes to use two kinds of normalized Closeness Centrality in different types of idealized networks as the references for the evaluation of the characteristics of the real network. By comparing Normalized Closeness Centrality in the real network with that in the idealized network, we can 1) evaluate the placement effect on Closeness Centrality and decide whether the real network is better or worse connected than the idealized ones and 2) explore to what extend the nodes in the real network are evenly distributed.

The results show that the Closeness Centrality of the same node varies remarkably depending on its position and how central it is in the chosen catchment area. In other words, the Closeness Centrality of a chosen node may be not necessarily small in the entire city street network, but it may be small in the selected catchment area only because it is not close to the center of the catchment area. Specifically, if the center point of a catchment area is moved by more than 100m away from the original center point, the Closeness Centrality of the same node starts to be significantly influenced by the placement effect.

The implications are, firstly, that because of the placement effect direct comparison of Closeness Centrality between different nodes in the same catchment area is only possible if these nodes are less than 100m away from each other. Secondly, if we want to compare the value of Closeness Centrality of two nodes that are more than 100m away from each other, we need to create two catchment areas with these two nodes being the center in each of these catchment areas.

8.1.5 Classifying stations by Node and Place Value

In chapter 6, we set up two tasks as the goals. The first goal is to apply the findings of chapters 4 and 5 to our empirical study and thus to further divide the stations within each Node-Place cluster into sub-clusters based on their Place Values, which are measured by two categories of indicators: network analysis and the accessibility, diversity and density of the POI. Affinity Propagation is used to cluster the train stations in order to determine their Place Value in the Node-Place Model, and the stations can then be compared and ranked by both the Node- and Place-Values. The results help to identify the neighborhoods that are already successful from the TOD point of view and, on the other hand, the neighborhoods with the potential for further improvement. The values of indicators can be used to determine the direction for improvement in each cluster.

The other goal of this chapter is to select stations that can be considered typical of each sub-cluster so that, in the following chapter, we can zoom into the catchment area around the station and scrutinize the detailed spatial distribution of the street network and POI.

Based on the results of network analysis- and POI-related indicators, the selected examples and the order of their Place Values are identified.

8.1.6 Strategic Allocation of Resources in Transit Oriented Development neighborhood

Chapter 7 aims at proposing a method to strategically and efficiently select the location for setting up new facilities in order to improve the Place Value of the TOD neighborhood. We choose the bike rental station (BRS) as an example of a possible intervention and demonstrate how to choose an appropriate location for such a station by, firstly, evaluating the coverage level and, secondly, clustering the existing BRSs using Density-Based Spatial Clustering of Applications (DBSCAN), which is a non-parametric algorithm for density-based clustering.

As the first step, a new BRS is virtually placed in the location with either the highest Closeness Centrality or Betweenness Centrality. Then the coverage level and the number of clusters before and after this intervention are compared in order to evaluate the improvement of the intervention.

This method can be used for comparing different types of interventions requiring network analysis for determining the optimal location for setting up new services or facilities, such as a mobile clinic, library, test station or a new shop, especially when empirical data (like socio-demographic data, which require significant amounts of time and money for their collection), are not available.

8.2 Contribution of this research

In the last few years, thanks to the rapid development of information technology and the increased availability of data sources, data-driven urban morphology research has vastly advanced. Therefore, through the computational application of network analysis, it has been possible for this research project to achieve the following goals.

- Assessing the existing models in network analysis and dealing with limits of their validity (such as the size effect and the placement effect) on the values of the indicators. The results can be helpful for determining the size and the location of the catchment area in order to mitigate the size and placement effects on indicators of network analysis.
- Evaluating the connectivity and resilience of two types of networks, namely public transit network and street network, by exploring their geometric and topological properties.
- Quantitatively classifying and evaluating the importance of stations in the railway public transport system based on 1) their structural importance in the transportation network, 2) topological characteristics of the street network and 3) accessibility of points of interests in the neighborhoods.
- Determining a suitable location of an intervention that matches the challenges faced by different types of neighborhoods in improving the walkability in the neighborhood.

Furthermore, this project has demonstrated that employing the machine learning approach can be very useful for developing a typology framework and a method that better describe the connectivity of transit networks and the complexity and walkability of the TOD neighborhoods. Therefore, using the approach developed in this thesis can help planners in

- characterizing the street network in each TOD neighborhood;
- providing the necessary means for comparing network systems in multiple neighborhoods;
- defining the catchment areas around the stations in line with the principles of sustainable transport infrastructure and service in order to develop the transit system in combination with spatial development strategies of the TOD neighborhoods.

8.3 Measurements

In the course of this research project, the following limitations with regard to measurements and datasets were identified.

1) Experimenting with more time slots of service frequency

In the current study, we calculate the departures per hour per direction between 6am and 9am. This is clearly not enough to fully explore spatial-temporal interactions on transportation networks. Further studies can carry out a variety of comparisons by differentiating various time slots of service frequency, such as “busy and quiet period”, “busy period before and after working hours”, “weekday and weekend”, “daytime and nighttime”. The results may help to define the guidelines regarding the minimum frequency of a well-used transportation system. Also, this deeper investigation might be necessary if the internal organization, such as routes, transferring stations, etc., of the network are different between the weekday and the weekend. Finally, the operation of the network during special events that attract massive crowds, such as a football match or a concert in a big stadium, can also be examined and compared with the regular routine pattern.

2) POI can be further differentiated by daily, weekly and monthly needs.

Further differentiation of POI based on the residents' frequency of visiting them (i.e., daily, weekly or monthly destinations) will provide an alternative perspective on spatial pattern of the importance of street nodes. It is quite possible that the street segments most frequently used to reach a daily amenity may be different from the ones for reaching a monthly destination.

3) Changing parameters in location-allocation analysis

With regard to location-allocation analysis, the following experiments can be conducted by changing the values of the parameters.

1. **Multiple target facilities:** In this thesis, only one target facility is used because the main goal is to demonstrate the methodology. In future research, the number of target facilities can be increased and the research question can then be, for example, "how does the increase of target facilities affect the average Closeness Centrality and Betweenness Centrality?"
2. **Differentiate target facilities by usage frequency:** Analysis of target facilities of different usage frequency, (i.e., the daily, weekly and monthly usage) can be compared.
3. **Alternative target problem:** Target problem to be solved in location-allocation analysis can be changed. In chapter 7, the target problems are increasing the coverage level of the bike rental station, avoiding the overlap of the service areas, minimizing the distance between facilities, increasing the level of movement opportunities. However, depending on the purpose of the project or research, the target problems can be changed to, for example, minimizing impediments, minimizing the number of facilities, maximizing usage or attendance, or maximizing market share.

4. Compare the effectiveness of solving the problem between different centrality indices

It may be interesting to find out which centrality measures would be more effective in solving the problem, such as reaching the highest coverage level with the least number of target facilities. Potential questions to be answered would be, for example, "does locating facilities in high Closeness Centrality increase the coverage level more than locating facilities in high Betweenness Centrality?"

8.4 Datasets

Considering that the complex network theory is a continuously developing subject, there are many aspects that are absolutely worthy of more attention. The network analysis in this thesis is used from a static and aggregate perspective and ignores other aspects like traffic flows, people's behavior or residents' socio-economic background. As a relatively new subject, many issues in this field remain to be resolved. However, one of the potential limitations for such a research is the availability of datasets.

1) Data of traffic flow and human dynamics for exploring the geometric and topological properties of network.

There is a growing concern with human dynamics on transportation systems, which is about how humans behave on transportation network systems and what influence human mobility patterns have on the general energy consumption. By including a dataset of traffic flow characteristics, human dynamic, travel demand distribution and passengers' behavior that arise in a realistic urban traffic network, some regional hubs may be uncovered and thus the analysis of depicting geometric and topological properties of the whole network can be improved.

2) Data on subjective preference and objective quality of the road infrastructure for the analysis of accessibility

With regard to accessibility of POI, this thesis focuses on the contribution of the "physical street configuration" to the accessibility of local amenities and destinations. In reality, human behavior is much more complex, and it is not taken into consideration at this stage. In future research, the index for accessibility analysis should be improved and extended to include indicators that can better describe the subjective willingness and the preferences of the pedestrians and the objective quality of the road infrastructure.

3) Data of socio-economic features for location allocation analysis

More socio-economic features should be considered in determining the accessibility of POI and in the location-allocation analysis of target facilities. However, since the composition of inhabitants may be changing frequently because the residents may move in and out of the neighborhood, socio-demographic background of the residents is a more dynamic indicator, and it is much more cost- and time-consuming to collect such data in comparison with the data on the built environment and urban component, such as the street network. By considering the socio-demographic background of the residents, the target facilities can be located in the place that is accessible not only for more street nodes but also for the users (especially the users who need to use these facilities the most).

4) Data on different cities, cultures or climates zones

Further research can use complex network theory to explore urban networks from different countries or regions and discover their unique characteristics and quantitative differences from other networks. Also, it could be interesting to discover what are the common quantities that are similar or identical among different cities, cultures, climates zones. For example, with the comparison among cities, we can compare the effect of size or configuration on the measurements and investigate whether a larger street network has more communities and hierarchical layers.

8.5 Further work based on the results of the current study

1) Future enhancement of mobility based on the results of the topological analysis of the transit network

In chapter 2, we examined the topological properties of Hamburg metropolitan area's public transportation network by using the network topology measures. To be more specific, the primary purpose of that chapter is to **examine whether the networks being studied show small-world and scale-free features**, which are the two unique features required for efficient operations and network resilience. Through analyses and comparisons between the PTN in Hamburg and Seoul, it is shown that the Hamburg case does not show strong evidence of small-world and scale-free features. The outcomes of that chapter are important for the future enhancement of mobility, e.g. through the strategic creation or relocation of stations or through the construction of new links.

2) Comparison of multiple scales of complex networks

Also, one can compare the networks of different scales. For example, one can compare the small-world properties of the entire network with that of the network in the inner city, where the network density is high. By identifying the clusters of sub-networks within the network, the researchers can further distinguish the variety of the characteristics of these sub-networks and examine how these characteristics affect the connectivity or resilience of the entire network.

3) Testing the scalability of the network shape and structure

Further investigation can also consider the effects of other factors (such as the size and shape of the catchment area) on the measurements. For example, a city in a circular shape may have a network with many links between nodes to cover the area. An interesting investigation would be the testing of scalability of the network shape and structural pattern. This means examining whether the existing network can scale both in size and connectivity. "A network's shape dictates its scalability and connectivity. A good shape and structure allow a network to scale with growth and still maintain the focus of its vision and purpose through connectivity. If the topology of the network discourages scalability and connectivity or has an unhealthy obsession of one over the other, then it may not be shaped for movement …. It's crucial to evaluate whether your network can scale both in size and connectivity" (Yang, 2018). Hamburg has recently constructed the new route, U5 (https://www.hamburg.de/u5/). Further research can examine to what extend this additional route improves the connectivity. If adding one additional route largely improves the connectivity, the shape and structure of HVV network can be considered to have good scalability.

4) Further investigations base on the results of the TOD classification

Chapter 3 demonstrates how to classify stations for TOD based on their centrality indices and service frequency. Based on the evaluation of the TOD and the importance of the stations, the following targets can be continued in future research projects.

- Designing the system in order to distribute the flow of passengers more evenly.
- Improving network connectivity and robustness by adjusting the spatial configurations of the network.

- A method for identifying the importance of stations can be developed as a tool to forecast the increase in ridership expected from the opening of a new line (Derrible, 2012)

References

Ahuja, N. (1983). On approaches to polygonal decomposition for hierarchical image representation. *Computer Vision, Graphics and Image Processing, 24*, 200-214.

Akbarzadeh, M., Memarmontazerin, S., Derrible, S., & Salehi Reihani, S. (2019). The role of travel demand and network centrality on the connectivity and resilience of an urban street system. *Transportation, 46*(11), 1127–1141. doi:https://doi.org/10.1007/s11116-017-9814-y

Austin, M., Belzer, D., Benedict, A., Esling, P., Haas, P., Miknaitis, G., . . . Zimbabwe, S. (2010). *Performance-based transit-oriented development typology guidebook.* CA: Center for Transit-Oriented Development.

Aydin, N., Duzgun, H., Wenzel, F., & Heinimann, H. (2018). Integration of stress testing with graph theory to assess the resilience of urban road networks under seismic hazards. *Natural Hazards, 9*(1), 37-68. doi:https://doi.org/10.1007/s11069-017-3112-z

Badlanda, H., Mavoa, S., Boulangé, C., Eagleson, S., Gunn, L., Stewart, J., . . . Giles-Cortia, B. (2017). Identifying, creating, and testing urban planning measures for transport walking: Findings from the Australian national liveability study. *Journal of Transport & Health, 5*, 151-162. doi:https://doi.org/10.1016/j.jth.2016.08.010

Barrington-Leigh, C., & Millard-Ball, A. (2019). A global assessment of street-network sprawl. *PLoS ONE , 14*(11), e0223078. doi:https://doi.org/10.1371/journal.pone.0223078

Batty, M. (2005). *Cities and complexity - understanding cities with cellular automata, agent-based models, and fractals.* MIT Press.

Batty, M. (2013). Resilient cities, networks, and disruption. *Environment and Planning B: Planning and Design, 40*(4), 571-573. doi:https://doi.org/10.1068/b4004ed

Batty, M., & Longley, P. (1994). *Fractal cities: a geometry of form and function.* London, UK: Academic Press.

Bell, S., Diaz, B., Holroyd, F., & Jackson, M. (1983). Spatially referenced methods of processing raster and vector data. *Image and Vision Computing, 1*, 211–220.

Bernick, M., & Freilich, A. (1998). Transit villages and transit-based development: the Rules are cecoming more flexible — how government can work with the private sector to make It happen. *The Urban Lawyer, 30*(1), 1-31.

Bertolini, L. (1999). Spatial development patterns and public transport: the application of an analytical model in the netherlands. *Planning Practice and Research, 14*(2), 199-210. doi:https://doi.org/10.1080/02697459915724

Bertolini, L., & Spit, T. (1998). *Cities on rails: the redevelopment of railway station areas and their surroundings.* London: Routledge.

Bissonnette, L., Wilson, K., Bell, S., & Shah, T. (2012). Neighbourhoods and potential access to health care: the role of spatial and aspatial factors. *Health Place, 18*(4), 841–853. doi:https://doi.org/10.1016/j.healthplace.2012.03.007

Blumenfeld-Lieberthal, E., & Portugali, J. (2010). Network cities: a complexity-network approach to urban dynamics and development. In X. Y. Bin Jiang, *Geospatial Analysis and Modelling of Urban Structure and Dynamics* (pp. 77-90). Dordrecht: Springer.

Boeing, G. (2017). OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems, 65*, 126-139. doi:https://doi.org/10.1016/j.compenvurbsys.2017.05.004

Boeing, G. (2018). *update examples/gboeing/network-clustering*. Retrieved from https://github.com/gboeing/network-clustering/commit/5ec44a18707bb4d3a2d13ecda28af7be954a8b93

Boeing, G. (2019). Urban spatial order: street network orientation, configuration, and entropy. *Applied Network Science, 4*(67), 1-19. doi:https://doi.org/10.1007/s41109-019-0189-1

Boeing, G. (2020). Planarity and street network representation in urban form analysis. *Environment and Planning B: Urban Analytics and City Science, 47*(5), 855-869. doi:http://dx.doi.org/10.2139/ssrn.3191236

Boots, B. (1980). Packing polygons: Some empirical evidence. *Canadian Geographer, XXIV*(24), 406–411. doi:https://doi.org/10.1111/j.1541-0064.1980.tb00264.x

Bourne, L. (1982). *Internal Structure of the City: Readings on Urban form, Growth, and Policy.* New York, NY, USA: Oxford University Press.

Buhl, J., Gautrais, J., Reeves, N., Sole, R., Valverde, S., Kuntz, P., & Theraulaz, G. (2006). Topological patterns in street networks of self-organized urban settlements. *The European Physical Journal B-Condensed Matter and Complex Systems, 49*(4), 513-522. doi:https://doi.org/10.1140/epjb/e2006-00085-1

Cao, G., Shi, Q., & Liu, T. (2016). An integrated model of urban spatial structure: Insights from the distribution of floor area ratio in a Chinese city. *Applied Geography, 75*, 116–126. doi:https://doi.org/10.1016/j.apgeog.2016.08.010

Cerin, E., Conway, T., Cain, K., Kerr, J., Bourdeaudhuij, I., Owen, N., . . . al, e. (2013). Sharing good NEWS across the world: Developing comparable scores across 12 countries for the neighborhood environment walkability scale (NEWS). *BMC Public Health 2013, 13*(1), 309. doi:https://doi.org/10.1186/1471-2458-13-309

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research, 2*(3), 199–219. doi:https://doi.org/10.1016/S1361-9209(97)00009-6

Chen, T., Hui, E., Wu, J., Lang, W., & Li, X. (2019). Identifying urban spatial structure and urban vibrancy in highly dense cities using georeferenced social media data. *Habitat International, 89*, 102005. doi:https://doi.org/10.1016/j.habitatint.2019.102005

Chopra, S., Dillon, T., Bilec, M., & Khanna, V. (2016). A network-based framework for assessing infrastructure resilience: A case study of the London metro system. *Journal of The Royal Society Interface, 13*(118), 20160113.

Czerkauer-Yamu, C. (2012). *Strategic Planning for the Development of Sustainable Metropolitan Areas using a Multi-Scale Decision Support System – The Vienna Case.* Geography. Université de FrancheComté.

Dadashpoor, H., & Yousefi, Z. (2018). Centralization or decentralization? A review on the effects of information and communication technology on urban spatial structure. *Cities, 78*, 194–205. doi:https://doi.org/10.1016/j.cities.2018.02.013

Derrible, S. (2012). Network centrality of metro systems. *PLoS ONE, 7*(7), e40575. doi:https://doi.org/10.1371/journal.pone.0040575

Donohoe, J., Marshall, V., Tan, X., Camacho, F., Anderson, R., & Balkrishnan, R. (2016). Evaluating and comparing methods for measuring spatial access to mammography centers in Appalachia. *Health Services and Outcomes Research Methodology, 16*(1), 22-4. doi:https://doi.org/10.1007/s10742-016-0143-y

European Commission. (2007). *Towards a new culture for urban mobility (Green Paper).* Brussels: COM 551 final.

Ewing, R. (1996). *Best Development Practices: Doing the Right Thing and Making Money at the Same Time.* Chicago, IL: American Planning Association.

Ewing, R., & Cervero, R. (2010). Travel and the built environment. *Journal of the American Planning Association, 76*(3), 265-294. doi:https://doi.org/10.1080/01944361003766766

Finkenbinder, A., Britt, K., & Blair, C. (2010). *Transit-Oriented Development Tools for Metropolitan Planning Organizations.* Metropolitan Planning Organizationspp. Retrieved from http://ctod.org/pdfs/2010TODToolsMPOs.pdf

Fortin, P., Morency, C., & Trépanie, M. (2016). Innovative GTFS data application for trainsit netowrk analysis using a graph-oriented method. *Journal of Public Transportation, 19*(4), 18-37. doi:https://doi.org/10.5038/2375-0901.19.4.2

Fortney, P., Rost, J., & Warren, J. (2000). Comparing alternative methods of measuring geographic access to health services. *Health Services & Outcomes Research Methodology, 1*, 173–184. doi:https://doi.org/10.1023/A:1012545106828

Frey, B. J., & Dueck, D. (2007). Clustering by Passing Messages Between Data Points. *Science, 315*(5184), 972-976. doi:https://doi.org/10.1126/science.1136800

Fushimi, T., Okubo, S., & Saito, K. (2020). Multiple perspective centrality measures based on facility location problem under inter-group competitive environment. *Applied Network Science, 5*. doi:https://doi.org/10.1007/s41109-020-00326-7

Gao, F., Kihal, W., Le Meur, N., Le Meur, N., Souris, M., & Deguen, S. (2017). Does the edge effect impact on the measure of spatial accessibility to healthcare providers? *International Journal of Health Geographics, 16*(1), 46. doi:https://doi.org/10.1186/

Gil, J. (2017). Street network analysis "edge effects": Examining the sensitivity of centrality measures to boundary conditions. *Environment and Planning B: Urban Analytics and City, 44*(5), 819-836. doi:https://doi.org/10.1177/0265813516650678

Giles-Corti, B., Foster, S., Koohsari, M., Francis, J., & Hooper, P. (2015). The influence of urban design and planning on physical activity. In S. T. H.E. Barton, *The Routledge Handbook of Planning for Health and Well-Being: Shaping a Sustainable and Healthy Future.* Abingdon, Oxon New York, NY: Taylor and Francis, Hoboken.

Gomez-Ibanez, G. (1996). Big-city transit, ridership, deficits, and politics. *Journal of the American Planning Association, 62*(1), 30-50. doi:https://doi.org/10.1080/01944369608975669

Greenberg, E. N.-G. (2020). A physical effort-based model for pedestrian movement in topographic urban environments. *Journal of Urban Design, 25*(5), 1-22. doi:https://doi.org/10.1080/13574809.2019.1632178

Gunn, L. D., King, T. L., Mavoa, S., Lamb, K., Giles-Corti, B., & Kavanagh, A. (2017). Identifying destination distances that support walking trips in local neighborhoods. *Journal of Transport & Health, 5*, 133-141. doi:https://doi.org/10.1016/j.jth.2016.08.009

Hamaina, R., Leduc, T., & Moreau, G. (2011). A structural analysis of the streets' network to urban fabric characterization. *ICC 2011 - 25th International Cartographic Conference.* Paris, France.

Hespe, D., & Sanders, P. (2019). More hierarchy in route planning using edge hierarchies. *19th Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2019)*, (pp. 10:1–10:14). München. doi: https://doi.org/10.4230/OASIcs.ATMOS.2019.10

Higgins, C., & Kanaroglou, P. (2016). A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region.

Journal of Transport Geography, 52, 61-72. doi:https://doi.org/10.1016/j.jtrangeo.2016.02.012

Hillier, B., Penn, A., Hanson, J., Grajewski, T., & Xu., J. (1993). Natural Movement: Or, Configuration and Attraction in Urban Pedestrian Movement. *Environment and Planning B: Planning and Design, 20*, 29–66. doi:https://doi.org/10.1068/b200029

Hong, J., Tamakloe, R., Lee, S., & Park, D. (2019). Exploring the topological characteristics of complex public transportation networks: focus on variations in both single and integrated systems in the seoul metropolitan area. *Sustainability, 11*(19), 5404. doi:https://doi.org/10.3390/su11195404

Huang, R., Grigolon, A. B., Madureira, A. M., & Brussel, M. (2018). Measuring transit-oriented development (TOD) network complementarity based on TOD node typology. *Journal of Transport and Land Use, 11*(1). doi:https://doi.org/10.5198/jtlu.2018.1110

Humphries, M., & Gurney, K. (2008). Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PLoS ONE, 3*(4), e0002051. doi:https://doi.org/10.1371/journal.pone.0002051

Iredale, R., Jones, L., Gray, J., & Deaville, J. (2005). 'The edge effect': an exploratory study of some factors affecting referrals to cancer genetic services in rural Wales. *Health Place, 11*(3), 197–204. doi:https://doi.org/10.1016/j. healthplace.2004

Jacob, J. (1961). *The Death and Life of Great American Cities.* New York: Random House.

Jacobson, J., & Forsyth, A. (2008). Seven american TODs: Good practices for urban design in Transit-Oriented Development projects. *Journal of Transport and Land Use, 1*(2), 51-88. doi:https://doi.org/10.5198/jtlu.v1i2.67

Jayasinghe, A., & Munshi, T. (2014). 'Centrality Measures' as a tool to identify the transit demand at public transit Stops: a case of Ahmedabad City, India. *International Journal of Advanced Research, 2*(7), 1063-1074.

Jayasinghe, A., Kasemsri, R., Abenayake, C. C., & Mahanama, P. (2019). Network Centrality Analysis of Public Transport Systems: Developing a Strategic Planning Tool to Assess Passenger Attraction. *International Journal of Innovative Technology and Exploring Engineering, 8*(10).

Jeffrey, D., Boulangé, C., Giles-Corti, B., Washington, S., & Gunn, L. (2019). Using walkability measures to identify train stations with the potential to become transit oriented developments located in walkable neighbourhoods. *Journal of Transportation Geography, 76*, 221-231. doi:https://doi.org/10.1016/j.jtrangeo.2019.03.009

Jordan, H. R. (2004). Distance, rurality and the need for care: access to health services in South West England. *International Journal of Health Geographics, 3*(1), 21. doi:https://doi.org/10.1186/1476-072X-3-21

Kabra, A., E., B., Girotra, K. .., & Girotra, K. (2019). Bike-Share Systems: Accessibility and Availability. *Management Science, 66*(9). doi:https://doi.org/10.1287/mnsc.2019.3407

Kamruzzaman, M., Baker, D., Washington, S., & Turrell, G. (2014). Advance transit-oriented development typology: Case study in Brisbane, Australia. *Journal of Transport Geography, 34*, 54–70. doi:https://doi.org/10.1016/j.jtrangeo.2013.11.002

Kazerani, A., & Stephanr, W. (2010). Modified Betweenness Centrality for Predicting Traffic Flow. *Available from: URL:http://www.biodiverse.unsw.edu.au/.../PDF/Kazerani_ and_Winter.pdf* .

Kermanshah, A., & Derrible, S. (2017). Robustness of road systems to extreme flooding: using elements of GIS, travel demand , and network science. *Natural Hazards, 86*(1), 151-164. doi:https://doi.org/10.1007/s11069-016-2678-1

Ketchen, D., & Shook, C. (1996). The application of cluster analysis in Strategic Management Research: An analysis and critique. *Strategic Management Journal, 17*(6), 441–458.

Krafta, R. (1994). Modelling intraurban configurational development. *Environment and Planning B: Planning and Design, 21*(1), 67 – 82. doi:https://doi.org/10.1068/b210067

Krehl, A., Siedentop, S., Taubenböck, H., & Wurm, M. (2016). A comprehensive view on urban spatial structure: Urban density patterns of German city regions. *ISPRS International Journal of Geo-Information, 5*, 76. doi:https://doi.org/10.3390/ijgi5060076

Larco, N. (2016). Sustainable urban design – a (draft) framework. *Journal of Urban Design, 21*(1), 1-29. doi:https://doi.org/10.1080/13574809.2015.1071649

Lau, B. (2020). *Human Centric Routing Algorithm for Urban Cyclists and the Influence of Street Network Spatial Configuration.* Lisbon, Portugal: NIMS - MSc Dissertations Geospatial Technologies (Erasmus-Mundus). Retrieved from http://hdl.handle.net/10362/95144

Levinson, D., & El-Geneidy, A. (2007). Network circuity and the location of home and work. *University Transport Study Group Conference .* Harrowgate, England.

Litman, T. (2020). *Evaluating Accessibility for Transport Planning:Measuring People's Ability to Reach Desired Goods and Activities.* Victoria Transport Policy Institute. Retrieved from https://www.vtpi.org/access.pdf

Liu, Y., Wei, X., Jiao, L., & Wang, H. (2016). Relationships between street centrality and land use intensity in Wuhan, China. *Journal of Urban Planning and Development, 142*(1), 05015001. doi:https://doi.org/10.1061/(ASCE)UP.1943-5444.0000274

Luo, J., Tian, L., Luo, L., Yi, H., & Wang, F. (2017). Two-step optimization for spatial accessibility improvement: a case study of health care planning in rural China. BioMed Research International: 1-12. https://doi.org/10.1155/2017/2094654. *BioMed Research International*, 1-12. doi:https://doi.org/10.1155/2017/2094654

Lyu, G., Bertolini, L., & Pfeffer, K. (2016). Developing a TOD typology for Beijing metro station areas. *Journal of Transport Geography, 55*, 40-50. doi:https://doi.org/10.1016/j.jtrangeo.2016.07.002

Marshall, S., Gil, J., Kropf, K., Tomko, M., & Figueiredo, L. (2018). Street network studies: from networks to models and their representations. *Networks and Spatial Economics, 18*, 735-749. doi:https://doi.org/10.1007/s11067-018-9427-9

Masucci, A. P., Stanilov, K., & Batty, M. (2013). Limited urban growth: London's street network dynamics since the 18th century. *PLOS ONE, 8*(8), e69469. doi:https://doi.org/10.1371/journal.pone.0069469

Masucci, A., Smith, D., Crooks, A., & Batty, M. (2009). Random planar graphs and the London street network. *The European Physical Journal B, 71*(2), 259–271. doi:https://doi.org/10.1140/epjb/e2009-00290-4

Mattsson, L., & Jenelius, E. (2015). Vulnerability and resilience of transport systems - A discussion of recent research. *Transportation Research Part A: Policy and Practice, 8*, 16-34. doi:https://doi.org/10.1016/j.tra.2015.06.002

McHugh, B. (2013). Pioneering Open Data Standards: The GTFS Story. In B. a. Goldstein, *Beyond Transparency: Open Data and the Future of Civic Innovation.* San Francisco:: Code for America Press.

Mishra, S., Welch, T., & Jha, M. (2012). Performance indicators for public transit connectivity in multi-modal transportation networks. *Transportation Research Part A: Policy and Practice, 46*(7), 1066–1085. doi:https://doi.org/10.1016/j.tra.2012.04.006

Mohmand, Y. T., & Wang, A. (2014). Complex network analysis of pakistan railways. *Discrete Dynamics in Nature and Society, 4*, 1-5. doi:https://doi.org/10.1155/2014/126261

Mohmand, Y. T., Mehmood, F., Amjad, F., & Makarevic, N. (2015). Complex network analysis of Guangzhou metro. *European Researcher, 100*(11), 711-718. doi:https://doi.org/10.13187/er.2015.100.711

Molaei, P., Tang, L., & Hardie, M. (2021). Measuring Walkability with Street Connectivity and Physical Activity: A Case Study in Iran. *World, 2*(1), 49-61. doi:https://doi.org/10.3390/world2010004

Monajem, S., & Nosratian, F. E. (2015). The evaluation of the spatial integration of station areas via the node place model; an application to subway station areas in Tehran. *Transportation Research Part D: Transport and Environment, 40*, 14-27. doi:https://doi.org/10.1016/j.trd.2015.07.009

Newman, L., Baum, F., Javanparast, S., O'Rourke, K., & Carlon, L. (2015). Addressing social determinants of health inequities through settings: a rapid review. *Health Promotion International, 30*(suppl_2), ii126-ii143. doi:https://doi.org/10.1093/heapro/dav054

Ngui, A., & Apparicio, P. (2011). Optimizing the two-step floating catchment area method for measuring spatial accessibility to medical clinics in Montreal. *BMC Health Service Research, 11*(166). doi:https://doi. org/10.1186/1472-6963-11-166

Okabe, A., & Sugihara, K. (2012). *Spatial Analysis Along Networks: Statistical and Computational Methods.* John Wiley & Sons.

Omer, I., & Jiang, B. (2008). Topological qualities of urban streets and the image of the city: A multi-perspective approach. *11th AGILE International Conference on Geographic Information Science*, (pp. 1-11).

OpenStreetMap contributors. (2017). Planet dump retrieved from https://planet.osm.org.

Ozbil, A., Peponis, J., & Bafna, S. (2009). The effects of street configuration on transit ridership. In L. M. Daniel Koch (Ed.), *Proceedings of the 7th International Space Syntax Symposium.* Stockholm: KTH.

Papa, E., Moccia, F., Angiello, G., & Inglese, P. (2013). An accessibility planning tool for network transit oriented development: SNAP. *Planum, 2*(27), 1-9.

Park, H.-T. (2009). Boundary effects on the intelligibility and predictability of spatial systems. *Proceedings of the Seventh International Space Syntax Symposium* (pp. 086:1–086:12). Stockholm: Royal Institute of Technology.

Pedregosa, F., Varoquaux, G., Gramfort, A., & Michel, V. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830. Retrieved from https://scikit-learn.org/stable/about.html#citing-scikit-learn

Penn, A., Hillier, B., Banister, D., & Xu, J. (1998). Configurational modelling of urban movement networks. *Environment and Planning B: Planning and Design, 25*, 59–84.

Peponis, J., Allen, D., French, S., Scoppa, M., & Brown, J. (2007). Street connectivity and urban density. *6th International Space Syntax Symposium*, (pp. 004-01--004-12).

Pezzica, C., Cutini, V., & De Souza, C. B. (2019). Rapid configurational analysis using OSM data: towards the use of Space Syntax to orient post-disaster decision making. *12th International Space Syntax Symposium (12SSS)* (pp. 147.1–147.18). Beijing, China: Space Syntax Network / Sejong University Press. Retrieved from http://orca.cardiff.ac.uk/id/eprint/120103

Pipley, B. (1981). *Spatial statistics.* New York: Wiley.

Pooley, C., Jones, T., Tight, M., Horton, D., Scheldeman, G., Mullen, C., . . . Strano, E. (2013). *Promoting walking and cycling: New perspectives on sustainable travel.* Bristol University Press. doi:https://doi.org/10.2307/j.ctt1t6p71q

Porta, S., Crucitti, P., & Latora, V. (2006). The network analysis of urban streets: a primal approach. *Environment and Planning B: Urban Analytics and City Science, 33*(5), 705–725. doi:https://doi.org/10.1068/b32045

Porta, S., Crucitti, P., & Latora, V. (2008). Multiple centrality assessment in Parma: A network analysis of paths and open spaces. *URBAN DESIGN International 13(1), 13*(1), 41-50. doi:https://doi.org/10.1057/udi.2008.1

Porta, S., Latora, V., Wang, F. H., Rueda, S., Strano, E., Scellato, S., . . . Latora, L. (2012). Street centrality and the location of economic activities in Barcelona. *Urban Studies, 49*(7), 1471-1488. doi:https://doi.org/10.1177/0042098011422570

Porta, S., Strano, E., Lacoviello, 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. doi:https://doi.org/10.1068/b34098

Portugali, J. (2000). *Self-Organization and the City.* Heidelberg: Springer.

Ratti, C. (2004). ‘Space syntax: some inconsistencies. *Environment and Planning B: Planning and Design, 31*(4), 487 – 499. doi:http://doi.org/10.1068/b3019

Remali, A., Porta, S., Romice, O., & Abudib, H. (2015). Street quality, street life, street centrality. In (. L. Vaughan, *Suburban Urbanities: Suburbs and the Life of the High* (pp. 104-129). UCL Press.

Renne, J. L. (2009). Measuring the Success of Transit-oriented Development. In J. L. Renne, & C. Curtis, *Transit-oriented Development: Making It Happen* (pp. 241-255). Routledge.

Renne, J. L. (2009). Measuring the Success of Transit-oriented Development. . In J. L. Renne, & C. Curtis, *Transit-oriented Development: Making It Happen* (pp. pp. 241-255). Routledge.

Reusser, D. E., Loukopoulos, P., Stauffacher, M., & Scholz, R. W. (2008). Classifying railway stations for sustainable transitions – balancing node and place functions. *Journal of Transport Geography, 16*(3), 191-202. doi:https://doi.org/10.1016/j.jtrangeo.2007.05.004

Rommel, J. (2014). *Topological Analysis of the Evolution of Public Transport Networks.Master’s Thesis.* Stockholm, Sweden: KTH Royal Institute of Technology.

Sadayuki, T. (2018). Measuring the spatial effect of multiple sites: An application to housing rent and public transportation in Tokyo, Japan. *Regional Science and Urban Economics, 70*, 155–173. doi:https://doi.org/10.1016/j.regsciurbeco.2018.03.002

Sadler, R. C., Gilliland, J. A., & Arku, G. (2011). An application of the edge effect in measuring accessibility to multiple food retailer types in Southwestern Ontario, Canada. *International Journal of Health Geographics, 10*(1), 34. doi:https://doi.org/10.1186/1476-072X-10-34

Salze, P., Banos, A., Oppert, J. M., Charreire, H., Casey, R., Simon, C., . . . Weber, C. (2011). Estimating spatial accessibility to facilities on the regional scale: an extended commuting-based interaction potential model. *10*(2). doi:https://doi.org/10.1186/1476-072X-10-2

Sasaki, K. (1989). Transportation system change and urban structure in two-transport mode setting. *Journal of Urban Economics, 25*, 346–367. doi:https://doi.org/10.1016/0094-1190(89)90055-7

Scheurer, J., Porta, S., & Curtis, C. (2008). *Spatial Network Analysis of Public Transport Systems: Developing a Strategic Planning Tool to Assess the Congruence of Movement and Urban Structure in Australian Cities.* Victoria., 2-5: The University of Melbourne.

Schlossberg, M., & Brown, N. (2004). Comparing transit-oriented development sites by walkability indicators. *Transportation Research Record Journal of the Transportation Research Board, 1887*(1), 34-42. doi:https://doi.org/10.3141/1887-05

Sharifi, A. (2019). Resilient urban forms: A review of literature on streets and street networks. *Buiding and Environment, 147*, 171-187. doi:https://doi.org/10.1016/j.buildenv.2018.09.040

Sharkey, J., & Horel, S. (2008). Neighborhood socioeconomic deprivation and minority composition are associated with better potential spatial access to the ground-truthed food environment in a large rural area. *Journal of Nutrition, 138*(3), 620–627. doi:https://doi.org/10.1093/jn/138.3.620

Shi, J., Wen, S., Zhao, X., & Wu, G. (2019). Sustainable development of urban rail transit networks: A vulnerability. *Sustainability, 11*(5), 1-24. doi:https://doi.org/10.3390/su11051335

Singh, Y. J., Fard, P., Zuidgeest, M., Brussel, M., & van Maarseveen, M. F. (2014). Measuring transit-oriented development: a spatial multi-criteria assessment approach for the City Region Arnhem and Nijmegen. *Journal of Transport Geography, 35*, 130–143. doi:https://doi.org/10.1016/j.jtrangeo.2014.01.014

Soczówka, P., Żochowska, R., & Karoń, G. (2020). Method of the analysis of the connectivity of road. *Computation, 8*(2), 54. doi:https://doi.org/10.3390/computation8020054

Soh, H., Lim, S., Zhang, T., Fu, X., Lee, G., Hung, T., . . . 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. doi:https://doi.org/10.1016/j.physa.2010.08.015

Spadon, G., Gimenes, G., & Rodrigues-Jr., J. (2017). Identifying urban inconsistencies via street networks. *Procedia Computer Science, 108*, 18–27. doi:https://doi.org/10.1016/j.procs.2017.05.103

Strano, E., Giometto, A., Shai, S., Bertuzzo, E., Mucha, P., & Rinaldo, A. (2017). The scaling structure of the global road network. *Royal Society Open Science, 4*(10). doi:https://doi.org/10.1098/rsos.170590

Strano, G., Hao, L., Everson, R. M., & Evans, K. E. (2013). A new approach to the design and optimisation of support structures in additive manufacturing. *The International Journal of Advanced Manufacturing Technology, 66*, 1247–1254. doi:https://doi.org/10.1007/s00170-012-4403-x

Stutzer, A., & Frey, B. S. (2008). Stress that Doesn't pay: the commuting paradox. *Scandinavian Journal of Economics, 110*(2), 339-366. doi:https://doi.org/10.1111/j.1467-9442.2008.00542.x

Tang, J. (2003). Evaluating the relationship between urban road pattern and population using fractal geometry. *UCGIS: University Consortium for Geographic Information Science Summer Assembly*, (pp. 1-13).

Telesford, Q., Joyce, K., Hayasaka, S., Burdette, J., & Laurienti, P. (2011). The ubiquity of small-world networks. *Brain Connect, 1*(5), 367–375. doi:https://doi.org/10.1089/brain.2011.0038

Tenkanen, H. (2017). *GeoPython - AutoGIS: Pandas and Geopandas Modules*. Retrieved from https://automating-gis-processes.github.io/2016/Lesson2-overview-pandas-geopandas.html

Thavikulwat, P. (2008). Affinity propagation: a clustering algorithm for computer assisted business simulation and experimental exercise. *Developments in Business Simulation and Experiential Learning, 35*, 220-224.

Val, D. S. (2015). Transit-oriented development, integration of land use and transport, and pedestrian accessibility: Combining node-place model with pedestrian shed ratio to evaluate and classify station areas in Lisbon. *Journal of Transport Geography, 45*, 70-80. doi:https://doi.org/10.1016/j.jtrangeo.2015.04.009

Van Meter, E., Lawson, A., Colabianchi, N., Nichols, M., Hibbert, J., Porter, D., & Liese, A. (2010). An evaluation of edge effects in nutritional accessibility and availability measures: a simulation study. *International Journal of Health Geographics, 9*(40). doi:https://doi.org/10.1186/1476-072X-9-40

Van Nes, A., & ZhaoHui, S. (2009). Network typology, junction typology and spatial configuration and their impacts on street vitality in Singapore. *Proceedings of the 7th International Space Syntax Symposium*, (pp. 124-1--124-11).

Vidal Rodeiro, C., & Lawson, A. (2005). An evaluation of the edge effects in disease map modelling. *Computational Statistics & Data Analysis, 49*(1), 45-62. doi:https://doi.org/10.1016/j.csda.2004.05.012

Wan, N., Zhan, F., Zou, B., & Chow, E. (2012). A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in Texas. *Applied Geography, 32*(2), 291–299. doi:https://doi.org/10.1016/j.apgeog.2011.05.001

Wang, F., & Luo, W. (2005). Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health Place, 11*(2), 131–46. doi:https://doi.org/10.1016/j. healthplace.2004.02.003

Wang, K., & Fu, X. (2017). Research on centrality of urban transport network nodes. *AIP Conference Proceedings*, *1839*, p. 020181. doi:https://doi.org/10.1063/1.4982546

Wang, X., & Chen, G. (2003). Complex networks: Small-world, scale-free and beyond. *IEEE Circuits Syst. Mag., 3*(1), 6–20. doi:https://doi.org/10.1109/MCAS.2003.1228503

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small world networks. *Nature, 393*(6684), 440–442. doi:https://doi.org/10.1038/30918

Woodcock, J., Edwards, P., Tonne, C., Armstrong, B., Ashiru, O., Banister, D., . . . I., R. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport. *Lancet, 374*(9705), 1930-1943. doi:https://doi.org/10.1016/S0140-6736(09)61714-1

Wu, J., Gao, Z., Sun, H., & Huang, H. (2004). Urban transit system as a scale-free network. *Modern Physics Letters B, 18*(19n20), 1043–1049. doi:https://doi.org/10.1142/S021798490400758X

Yamaoka, K., Kumakoshi, Y., & Yoshimura, Y. (2021). Local betweenness centrality analysis of 30 european cities. In S. Geertman, C. Pettit, R. Goodspee, & A. Staffans, *Urban Informatics and Future Cities* (pp. 527-547). Springer International Publishing. doi:DOI: 10.1007/978-3-030-76059-5_26

Yang, D. (2018). *What shap is your network?* Retrieved from SEND Institute: https://www.sendinstitute.org/network-shape/

Yoshimura, Y., Santi, P., Arias, J. M., Zheng, S., & Ratti, C. (2020). Spatial clustering: Influence of urban street networks on retail sales volumes. *Environment and*

Planning B: Urban Analytics and City Science. doi:https://doi.org/10.1177/2399808320954210

Zahedi, Z., Mawengkang, H., Masri, M., Ramon, H., & Putri, Y. (2019). Mathematical fallacies and applications of graph theory in electrical engineering. *Journal of Physics: Conference Series, 1255*, 012085. doi:https://dio.org/10.1088/1742-6596/1255/1/012085

Zemp, S., Stauffacher, M., Lang, D. J., & Scholz, R. W. (2011). Classifying railway stations for strategic transport and land use planning: context matters! *Journal of Transport Geography, 19*(4), 670-679. doi:https://doi.org/10.1016/j.jtrangeo.2010.08.008

Zhang, L., Lu, J., Fu, B., & Li, S. (2019). Dynamics analysis for the hour-scale based time-varying characteristic of topology complexity in a weighted urban rail transit network. *Physica, 527*(6825), 121280. doi:https:doi.org/10.1016/j.physa.2019.121280

Zhang, X., Li, W., Deng, J., & Wang, T. (2014). Research on hub node identification of the public transport network of Guilin based on complex network theory. In M. J., P. D., H. H., & Y. Y., *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation* (pp. 1302–1309). Reston, VA, USA: American Society of Civil Engineers.

Zhang, X., Lu, H., & Holt, J. (2011). Modeling spatial accessibility to parks: a national study. . *International Journal of Health Geographics, 10*(31). doi:https://doi. org/10.1186/1476-072X-10-31

Zhang, Y., Song, R., Nes, v. R., He, S., & Yin, W. (2019). Identifying Urban Structure Based on Transit-Oriented Development. *Sustainability, 11*(24), 7241. doi:https://doi.org/10.3390/su11247241

Zhu, J., & Sun, Y. (2017). Building an urban spatial structure from urban land use data: An example using automated recognition of the city centre. *ISPRS International Journal of Geo-Information, 6*(5), 122. doi:http://doi.org/10.3390/ijgi6040122

Żochowska, R., & Soczówka, P. (2018). Analysis of selected structures of transportation network based on graph measures. *Scientific Journal of Silesian University of Technology. Series Transport, 98*, 223–233. doi:https://doi.org/10.20858/sjsutst.2018.98.21