Temporal Analysis of Accessibility Using Complex Network Theory

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy and Management

by

Emre Özhan Yiğitbaşı

Student number: 5316944

To be defended in public on August 1st, 2022

Graduation committee

Chairperson : Prof.dr. G.P. van Wee, Section Transport and Logistics
First Supervisor : Dr. N.Y. Aydin, Section Systems Engineering
Second Supervisor : Dr. Y. Casali, Section Systems Engineering

PREFACE

This thesis marks the end of my two-year journey at TU Delft, and possibly my student life. Unfortunately, I started this master study amid a pandemic, and I am glad to have made it through without any significant damage, physically or mentally. Though I must admit, it was very difficult at times; moving to a new country and attempting to "remotely" socialize with my peers was a major challenge. One can only wish that things were more pleasant at the start, but I am glad to be able to finalize this chapter of my life with a work that I am proud to present.

Two years ago, I arrived at TU Delft as a mechanical engineer, seeking to expand my vision and skill set and work on projects with a direct value for the society. Complex Systems Engineering and Management program introduced me to a broad range of societal problems that can be solved using a systems engineering perspective. Now I feel confident that I can tackle any real-world challenge through the methods I learned and the way of thinking that I acquired. Though admittedly, I still sometimes struggle to explain the full context of this study to my family and friends.

I was able to deliver this thesis by the help of my dear supervisors Nazlı Aydın and Ylenia Casali. I was fortunate enough to have them ready to guide me and my research even before day one. During our regular meetings, my ideas were always respected and listened to, and I always appreciated their detailed feedback and guidance. I would like to thank Bert van Wee for accepting to be the chair of my graduation committee and providing insightful comments and suggestions to help me structure my research and the report.

Most importantly, I would like to express my gratitude to my family and friends for their support in these last two years. To Mom and Dad, thank you for raising me to be the person I am today, and for your never-ending support throughout my student life. Thank you to Halil Suat for the Netherlands meetups and the overall academic inspiration. To Zeynep and Berk, thank you for the monthly Zoom calls across borders, especially in the times when we all felt lonesome. To Onur, Mert, and Fırat, thank you for the lasting friendship. Finally, to Dilay, thank you for always being there for me and being my muse, no matter the distance. Your encouragement and understanding kept me motivated to complete this challenge.

Lastly, I would like to thank you, the reader, who will invest their time to read this thesis. Hopefully you will gain new knowledge or find an inspiration for yourself.

Emre Yiğitbaşı
Delft, July 2022

EXECUTIVE SUMMARY

Problem Description and Research Gap

Urbanization in many parts of the world brings about the issue of a fair distribution of resources, or social equity [Vecchio et al., 2020]. One of the discussion points for equity is the ability to reach certain areas of interest in an urban environment, or accessibility. Accessibility to critical services such as healthcare, employment, or education are basic determinants of participation in the society and economic activities. The measurement of accessibility enables the assessment of availability of these activities to the population. Acknowledging the need for an accurate measurement of accessibility for social equity assessment, an extensive literature of measurement approaches exist [Geurs and van Wee, 2004]. However, each type of measure assumes a different perspective to accessibility, which makes them useful in certain use cases.

Factors determining the accessibility to opportunities are identified with respect to the interactions between transport and land use systems, as well as external influences of cultural and socioeconomic factors [Bertolini, 2012]. Accessibility to some opportunities are determinants of future economic participation and the individual's stature in the job market. The most prominent opportunities in this aspect are education and job accessibility [Guzman et al., 2017]. While the latter is commonly researched in literature, the former has received comparatively less interest. The identification of factors influencing accessibility to these opportunities determine the aid the formulation of an ad-hoc accessibility measure for their use cases.

Complex network theory based transport network assessments are gaining more attention in urban research and transport studies [Ding et al., 2019]. Network models developed using this approach are used in studies with a focus on topology assessment, community identification, and robustness improvements. A limited number of studies work to expand complex network models' use cases into the accessibility measurement field. Specifically, the data analytical approach based on these models are useful to assess the developments in transport networks (network evolution) and other urban systems (co-evolution). Nonetheless, most of these assessments are limited to network effectiveness and topological developments [Brussel et al., 2019]. A limited number of implementations to the accessibility measurement use case exist in the context of network evolution studies.

Research Approach

The objective of this study is to understand the added value of using complex network models driven accessibility measurement and equity assessment. Therefore the main research question is forumulated is as follows:

How can urban transport equity be temporally analyzed in terms of accessibility to education using complex network theory?

This research is structured upon four subquestions to answer the main research question. Each subquestion either refers to a core concept of this study or develops a methodology to assess the usability of complex network approaches.

- **SQ1:** What are the existing accessibility based transport equity evaluation methods and indicators?

- **SQ2:** What are the factors that influence transport accessibility to education opportunities?

- **SQ3:** How are complex network methods used in the analysis of urban transport systems?

- **SQ4:** To what extent can complex network analysis be used to evaluate and explain the evolution of spatial accessibility distribution?

The selected research approach for this study is a combination of modeling and case study approaches. An urban model is used including transport, land use, and socioeconomic factors based on complex network theory. In combination, an analysis methodology for accessibility and spatial equity assessment is developed. This methodology is then implemented in a case study in the City of Helsinki. The availability of open data in this region makes it easier to collect the necessary data. Furthermore, based on the historical development of its urban form and transport systems, this area provides a suitable case for the developed methodology for temporal accessibility evaluation.

Literature Review

Equity measures are widely used in policy assessments to determine how resources are distributed across populations [Martens et al., 2019]. The case of accessibility is a commonly visited indicator in the assessment of equity [van Wee and Mouter, 2021]. Equity is analyzed in terms of the burdens and benefits, differentiation of population (based on age, income, location, etc.), and guiding equity principles. For an accurate assessment, the measurement of accessibility is an important component of equity assessment. Several measures exist in literature based on infrastructure, location, utility, or personal attributes [Geurs and van Wee, 2004]. To answer to the first subquestion of this study, a review of accessibility-driven equity research is carried out to categorize such studies based on the equity concepts and accessibility methods. Furthermore, common equity indices in the literature are reviewed with their use cases, advantages and disadvantages in their formulation. The most relevant indices for the case of spatial equity analysis of accessibility are determined to be Spatial Gini and Theil indices, whereas Gini index is a suitable metric for overall equity assessment across the population.

For the second subquestion, factors governing accessibility in the transport and land use context are investigated, where a concept framework by Bertolini [2012] is described. The focal case of this study is the accessibility to education opportunities, which has been found to be a factor for participation in higher education and and employability [Dickerson and McIntosh, 2013][Di Paolo et al., 2017]. However, the case of education has not gained as much attention as other cases such as job

accessibility [Sharma and Patil, 2022]. Therefore, a comparative literature review is conducted to point out the common factors related to accessibility to jobs and education. The end results is a conceptualization of these common factors, namely affordability, availability, and proximity, thus answering SQ2.

The last literature review focuses on the complex network theory approaches in transport research, seeking to answer SQ3. A taxonomy of complex network indicators are made, using the categories of centrality, community, topology, and accessibility indicators. These indicators are commonly used in transport studies and aid the understanding of network properties. Thus, in addition to the accessibility indicators, topology indicators can serve to make infrastructure-based inferences and their effect on accessibility. Identified accessibility approaches are space syntax and random walk based access diversity. The former assumes an axial representation of the road network to model human movement, but lacks the land use component. The random-walk approach is a topologically driven approach in the literature, with a potential to be coupled with land use and socioeconomic indicators. Following this, the studies considering the 'temporal' evaluation of networks were studied, namely network evolution and co-evolution studies. The network evolution studies approach historical development from a topological perspective, whereas co-evolution studies investigate the interaction of socioeconomic, demographic, or land use factors and the transport systems. The outcomes of the review identified a research gap as network evolution studies do not consider the accessibility developments.

Methodology

Based on the literature review which identified the equity and accessibility measurement approaches, accessibility factors, and relevant complex network methods, a methodology for accessibility measurement and analysis is developed. This methodology relies on transport infrastructure data to develop a complex network representation, and analyze the accessibility in this model based on the random walk approach. A self-avoiding random walk algorithm specific to school accessibility measurement is developed. In addition to the road network data, the use of this algorithm requires the data of school locations. This algorithm calculates the accessibility metric, visit per walk (vpw) by the average count of schools visited in a random walk simulation per network node. The outputs are descriptively analyzed to examine overall accessibility in the network, and compare across historical timesteps. Then a spatial analysis is conducted using hotspot analysis to identify clusters, and compare with the spatially defined socioeconomic indicators. The last step of the methodology is to evaluate equity based on the selected equity indices of Gini, Spatial Gini, and Theil.

Case Implementation and Results

The selected case area, City of Helsinki, is an urbanizing region that has built an urban core around its historical center near the southern harbor. It has been transforming into a polycentric region, which has lead to a spatial transformation of its transport systems, land use, and demographics.

The historical road network data of this region was collected and transformed into the complex network model. Four separate road networks were built from 1991, 1999, 2007, and 2016. In addition, historical school location data was collected from the Helsinki school register. Another dataset utilized in this implementation is

the socioeconomic and population data based on subdistrict-level aggregation. The self-avoiding random walk algorithm was applied on the network model and the representation of school locations on the network.

The results of the case study show a decrease in the overall accessibility through the historical timesteps. Based on spatial analysis, the high access clusters are mainly located in the historical center of the city. Apart from the numerous small sized clusters around school nodes, no polycentric developments of accessibility are observed. Compared to the population distribution of the city, there are many parallels between accessibility. However, compared to the population of the target group of education (people aged 19 or lower), the high access clusters are not well distributed. Based on the equity, analysis, the overall equity is diminished through time, with a strong spatial correlation observed based on the Spatial Gini calculations. Theil index confirms the high inequity between subdistricts, however it is not possible to estimate the exact contribution due to the developed accessibility metric not being compatible with the formulation of this index.

Reflection on the Method

The proposed methodology based complex network modeling and random walk simulations is a good addition to the existing accessibility measurement approaches. This method considers the transport network topology and the diverse paths and destinations around each node. Thus, any temporal developments in network structure can be captured in the accessibility measurement with the network modeling. Furthermore, the probabilistic approach of random walk relies on network diffusion, thus considering multiple paths between origin destination pairs, rather than the shortest path only. Normally, the outputs of the random walk method would be limited to the identification of topological clusters within the network, however with the addition of school location, a more complete assessment can be done with regard to target activities based on land use. The limitations of this method is the lack of consideration for the capacity of opportunities and potential competition aspects based on attractiveness. Although the competition is not a strong deciding factor in mandatory public education Andersson et al. [2012], it could be valid for other use cases where this methodology is used. The accessibility metric developed for this method needed an adaptation based on spatial distribution of population, and was not compatible with the Theil index. Therefore, the methodology should consider the applicability of the accessibility measurement method to the chosen equity indices.

Conclusion

Based on the assessments of subquestions leading to an answer to the main research question, this study makes conclusions to the usability of complex network theory in accessibility measurement and equity assessment. Using the developed methodology and the findings of the case study the following conclusions are made:

- The complex network model must include land use information regarding locations of relevant opportunities

- The network model and accessibility measurement method should be compatible with the equity index

- Accessibility measurement parameters in the random walk implementation should be representative of real-world circumstances regarding travel behavior and infrastructural changes.

- For an accurate temporal analysis, the transport infrastructure, land use, and socioeconomic information must be accurate and consistent within each timestep.

This study makes a scientific contribution through using a novel measurement approach for accessibility to opportunities based on complex network theory. The developed methodology is useful to identify the transport and land use impacts on historical development of accessibility. This method is repeatable for multiple timesteps, thus contributing to the temporal analysis of accessibility. Furthermore, by choosing the case of school accessibility, a contribution is made to this field of accessibility research through a factor conceptualization and applicable methodology. Empirical contributions are made through the analysis of the case study area, where a reduction in accessibility schools are observed. The societal contributions are made through addressing the challenge of social equity in the case of accessibility. The specific focus on schools is relevant to the sustainable development goals defined by United Nations [2015].

The limitations of this study stem from the assumptions made for the selected methods and the data availability. The case study uses a car-oriented road network in a walking accessibility assessment. The inclusion of pedestrian walking paths could greatly improve the accuracy of results. Furthermore, the inclusion of public transport by using a multi-layer network model could increase the accuracy of this study, yet increase the complexity. Considering the conceptualization of accessibility measurement factors prepared in this study, the methodology fails to consider the capacity effects of opportunities. Lastly, some of the equity indices used in the study were not fully applicable to the developed accessibility metric, which impaired the spatial equity assessments. Future studies should consider developing equity methods suitable with this method or adapt the network model to the existing equity approaches.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 | INTRODUCTION

This chapter introduces the research with a problem definition, knowledge gap explanation, selected research approach, and the research flow with the utilized methods.

1.1 PROBLEM DEFINITION

The world is rapidly urbanizing, as the spatial distribution of its population shifts from rural to urban settings. The world's urban population has quadrupled between 1950-2018, and it is estimated that by 2050 68% of the total population will be living in urban areas [UN-DESA, 2018]. Consequently, there has been growing concerns about the implications of urban growth in terms of social equity [Vecchio et al., 2020]. One perspective to such implications is the mobility activity of individuals. Mobility is a determinant of the extent to which people can reach certain places in the city [Vecchio et al., 2020].

Accessibility refers to the ease of traveling to specific locations, and "good accessibility" enables participation in services and social interaction [Li et al., 2021]. Services which are commonly considered in accessibility studies are education, employment, and healthcare [Curl, 2018]. Accessibility measurement enables the assessment of equity in spatial and social dimensions [Curl, 2018]. Such measurements are used as quantitative indicators of spatial availability of socioeconomic opportunities [Geurs and van Wee, 2004].

Accessibility to activities and services is an important indicator of policies and developments regarding transport and land use systems [Geurs and Ritsema van Eck, 2003]. A continuous cycle defined by [Bertolini, 2012] summarizes how accessibility to activities are affected by the transport and land use infrastructure. Among these activities are economic opportunities, such as job locations, as well as critical services, such as education, healthcare, and other fundamental facilities. The lack of access to these economic opportunities can remarkably affect the quality of life. As an example, accessibility to education opportunities is found to be an important decision factor for the continuation to higher education [Dickerson and McIntosh, 2013]. Thus, accessibility to education impacts the economic participation, as the education level of an individual determines their level of economic stature [Di Paolo et al., 2017]. Despite the identified significance of accessibility to education, this field is not as commonly researched in accessibility research compared to the likes of job accessibility.

The concept of equity and fairness is a visited topic in transport research, although not as much as effectiveness or efficiency [van Wee and Mouter, 2021]. The subjective nature of the ethical principles attached to equity analysis adds a layer of complexity to the equity measurement concepts. Many ambiguous points related to equity are stated in the literature, such as how to measure it, how to present it, and who should

evaluate it. Furthermore, there is a multitude of equity principles to select based on a specified equity measurement use case. These principles are especially important for policy assessments as they determine how a resource should be distributed in the population, whether everyone should get the same amount or if the disadvantaged should be favored. In addition to the guiding principles, the operationalization of equity measurement depends on factors such as the type of benefits and burdens and the ways of differentiating groups of people [Martens et al., 2019]. In transport research, a common use case of equity measurement is based on accessibility. However, there exist a multitude of accessibility measurement approaches, and all have advantages and shortcomings in various use cases.

Accessibility measurement approaches are categorized by Geurs and van Wee [2004] under four main perspectives: accessibility measures based on infrastructure, location, utilities, and people. Most of these perspectives take into account the transport, land-use, temporal and individual-level components. Accessibility measurement outputs can be used both as social and economic indicators in a policy evaluation context [Geurs and van Wee, 2004]. The former use is valuable in the evaluation of availability of opportunities. The latter use is more relevant for measuring the cost and benefits of a specific project or a broader economic impact due to indirect effects. For this study, accessibility as a social indicator is more relevant, specifically for social equity impacts together with the spatial dispersion of measurements.

Complex network theory offers a broad range of possibilities with regard to mapping and spatially analyzing urban systems. Specifically, transport systems benefit from the graph-theory driven approach to mathematically analyze network characteristics, identify communities and groups, improve robustness, and optimize based on performance objectives. In addition, the development of complex network paradigms such as small-world and scale-free network properties transformed the way urban traffic networks are analyzed [Ding et al., 2019]. However, a limited number of studies work on expanding its use case to identify and measure accessibility in transport systems. In contrast, most studies focus on the efficiency of transport networks [Brussel et al., 2019]. To address this shortcoming, novel research should seek to explain the relationship of urban systems with socioeconomic attributes [Venerandi et al., 2018].

Building on the current capabilities of transport research using complex network approaches, the increased availability of historical data makes it possible to study urban systems' development over time [Strano et al., 2012]. There exist studies that develop models for the interdependence of transport network growth, population growth, and urban land use changes within the city through complex network models, also called co-evolution theories [Ding et al., 2019]. Theories for such interdependent growth have been developed and tested in some historical context to analyze the effect of transport systems on population and urban form [Levinson, 2008]. Similar studies investigated the interplay between travel cost and population density [Barthélemy and Flammini, 2009]. However, there is a lack of studies that use historical real-world data to evaluate spatial accessibility development and taking an equity perspective to the co-evolution.

Network evolution studies are often interested in the aspects of network growth and densification [Strano et al., 2012], while ignoring the temporal aspect of socioeconomic indicators. On the other hand, historical maps are getting digitized by the means of geographical information systems (GIS) [Gallotti et al., 2015]. This development makes historical locations of public services, residential- and employment-dense regions available through cities' archives and open data services. The increased data availability provides the necessary information to analyze the development of equity in accessibility distribution over time. Therefore, a potential research field to explore is the application of complex network methods to the historical analysis of social equity indicators such as accessibility.

1.2 KNOWLEDGE GAP AND RESEARCH QUESTIONS

Transport systems are investigated in the literature in terms of their structural growth, socioeconomic variables, and interrelation with other subsystems. Accessibility measurements have been widely used to understand the equity aspects within urban areas. However, when it comes to the topic of urban growth, the societal impact of evolving transport networks is not the focal point of the reviewed accessibility and equity studies. In contrast, studies that actually consider the temporal network development are often focused on the topological growth only. Thus, this research serves to fill the gap regarding the evaluation of complex interactions across urban systems. This study proposes measuring the accessibility based transport equity by adopting a complex network approach to study urban transportation systems' temporal development. The equity analyses will concentrate on education opportunities, wherein spatial accessibility distribution will be investigated. Spatial patterns of school accessibility has been shown limited attention by existing studies, although fair distribution of education is an important dimension of equity and accessibility analysis [Ye et al., 2018].

This research seeks an answer to the following research question:

How can urban transport equity be temporally analyzed in terms of accessibility to education using complex network theory?

The project thereby aims to understand the change in accessibility distribution over time with respect to education locations. In doing so, complex network theory is applied to historical transport networks. The novelty of this study stems from the adopted complex network approach to evaluate accessibility in urban areas. Such methods are not commonly used for accessibility measurement in literature, or the existing ones are not very well-developed to include all relevant accessibility factors. The analysis is based on a case-study, in which the data collected from the Helsinki urban region in Finland. This region has the advantage of providing open-data regarding its sociotechnical systems. The research is structured based on the following sub-questions (SQ) to answer the main research question:

- **SQ1:** What are the existing accessibility-based transport equity evaluation methods and indicators?

- **SQ2:** What are the factors that influence transport accessibility to education opportunities?

- **SQ3:** How are complex network methods used in the analysis of urban transport systems?

- **SQ4:** To what extent can complex network analysis be used to evaluate and explain the evolution of spatial accessibility distribution?

1.3 RESEARCH APPROACH

This study uses a combination of two research approaches, incorporating modeling and case study approaches. The modeling approach is used to develop an urban model including transport, land use, and socioeconomic factors based on complex network theory. Next to this, an analysis methodology for accessibility and spatial equity assessment that is repeatable in various use cases is developed. The methodology is then tested in a case study. These approaches are explained in greater detail in the following subsections, followed by the chosen case study area and the research methods of this study.

1.3.1 Modeling Approach

Numerous conceptualizations have been made by social scientists on the relationship between disparities in socioeconomic variables and the physical world [Tóth et al., 2021]. Physical systems in the urban environment evolve by interacting with each other, making it difficult to formulate every real-world occurrence. Thus, in addition to research that aims to theoretically explain the phenomena that cause inequalities, modeling studies are necessary for a better understanding of the complex dynamics of the urban system. This study aims to bridge this gap, and adopts a modeling approach based on complex network theory to measure accessibility in an urban setting.

Adopting a modeling research approach for this study enables understanding the complex relationship between urban systems over time. According to Bibri [2018] the advantage of computational modeling is the possibility of testing new approaches and methods where traditional theoretical methods are insufficient. The complex behaviors in a system are often difficult to foresee, and modeling is a suitable approach for such cases [Bibri, 2018]. With a well-documented modeling and analysis methodology, the trustworthiness of this study can be established through repeatability.

The limitations of this approach are caused by assumptions made during model construction. Since it is not possible to replicate real world settings one-to-one, there will be certain assumptions that make the modeling and analysis possible. Certain indicators will be chosen to model a complex environment, as well as the limitations caused by the resolution of spatial data. Dawson [2003] maintain that such design decisions are often made too reliant on theory, affecting and constraining the outcome of the model. In the case of complex network theory, the model will unavoidably

simplify the urban system to build a graph representation. This could lead to the loss of some information which would be relevant in different application use cases.

Lastly, adopting a modeling research approach for a complex environment brings about a phenomenon called the Bonini's paradox, where the increasing model complexity makes it more difficult to understand the model [Dawson, 2003]. Therefore, the modeler should include the most relevant and necessary information in the model to keep it understandable. To sum up, the limitations and of the modeling approach needs to be addressed in the research process design through appropriate validation steps. The case study approach is a practical way to validate the model or address the identified shortcomings

1.3.2 Case Study Approach

The second approach selected for this study is a quantitative case study approach. This approach is used when a phenomenon needs to be understood in a real-world setting [Harrison et al., 2017]. Case study approach is often both exploratory and explanatory, therefore seeking to answer how and why questions [Harrison et al., 2017]. As one of the objectives of this study is to assess how accessibility can be measured using real-world data, case study is an appropriate approach. This study uses historical transport network data, as well as socioeconomic indicators in a chosen urban environment, to answer the main RQ. Studies that adopt the case study approach often use qualitative or quantitative methods, as well as mixed methods methodologies. This study adopts a quantitative approach that analyzes a complex network model to explore the urban systems and the road infrastructure network. Thus, the road transport network of the case area is analyzed in terms of accessibility to employment and education opportunities. Consequently, the historical development of transport equity is evaluated and explained. The advantage of adopting a quantitative approach is the ease of analyzing large-scale data in an urban setting. The availability of data in the case area is a significant factor in the realization of this approach, which is explained in the following paragraph.

1.3.3 Case Selection

The selected urban region for this case study is the Helsinki region in Finland. Real-world data is collected from the open-data platform of the city of Helsinki. The service used by the local government, called "Helsinki Region Infoshare (HRI)" makes structured data about urban environment, services, and infrastructure available to use [City of Helsinki, 2021]. Due to the temporal aspect of this study, historical data must be collected for a temporal analysis of accessibility. *Koulurekisteri* is a database provided by the Helsinki City archives where historical data about every school that operated in the City of Helsinki are available, starting from the year 1550 [Koulurekisteri, 2020]. Furthermore, historical data are collected including socioeconomic, built environment and road network data from years 1991, 1999, 2007, and 2016. The open-data availability of the city offers a significant advantage to the selected quantitative approach, due to the great ease of obtaining and processing numerical data. However, the disadvantage of basing the case study on quantitative

data stems from the data source itself. The accuracy of the results of the quantitative analysis is constrained by the amount and quality of information that can be collected regarding the case.

Based on this discussion, the selection criteria for the case area are identified for future research. For the temporal component of this study, it is important that historical data is available with regard to the transport network, and other socioeconomic data. Furthermore, for an effective temporal analysis, a clear development or change in the urban area should be observed in terms of its transport, land use, and overall urban form. Therefore, the criteria for selection are summarized below:

- (Open) data availability of transport systems and socioeconomic and demographic information

- Historical urban development in terms of transport, land use, and urban form

Generalization of the findings of a case study is often a matter of criticism [Crowe et al., 2011]. The results of the analysis for other cases may not be repeatable based on the conditions. Therefore, the Helsinki case is evaluated considering the region's own conditions and historical context. The outcomes of the model analysis with the case data need to be evaluated in consideration when being applied to other regions. Nonetheless, case study approach adds value to this research by providing a use case for the developed network model and analysis methodology.

1.3.4 Research Methods

The first steps of the research require a literature review to answer SQ1 and SQ2. While the first sub-question is focused on transport equity in the context of accessibility, the second one focuses on the relationship between urban transport systems, and accessibility to education by a review of factors. To answer SQ3, another literature review is conducted to understand the added value of complex network theory in temporal analysis of transport systems. Complex network methods and metrics are reviewed, with potential applicability to accessibility-based equity analyses. Finally, SQ4 is answered by developing a methodology based on complex network modeling to measure accessibility and analyze equity. Next step is to apply the methodology to the Helsinki case study. Historical transport network data, school locations and population data are collected to model the City of Helsinki and assess the development of equity based on accessibility to education opportunities. The final step of the study is reflection on the developed method based on the concepts reviewed for the first three subquestions. The assessment of applicability to other use cases are made, and recommendations for future applications are given. The research process is visualized in Figure 1.1.

1.4 STRUCTURE

The structure of this report is as follows: Chapter 2 describes the methodology for the literature review with query keywords, selected key studies, and related concepts
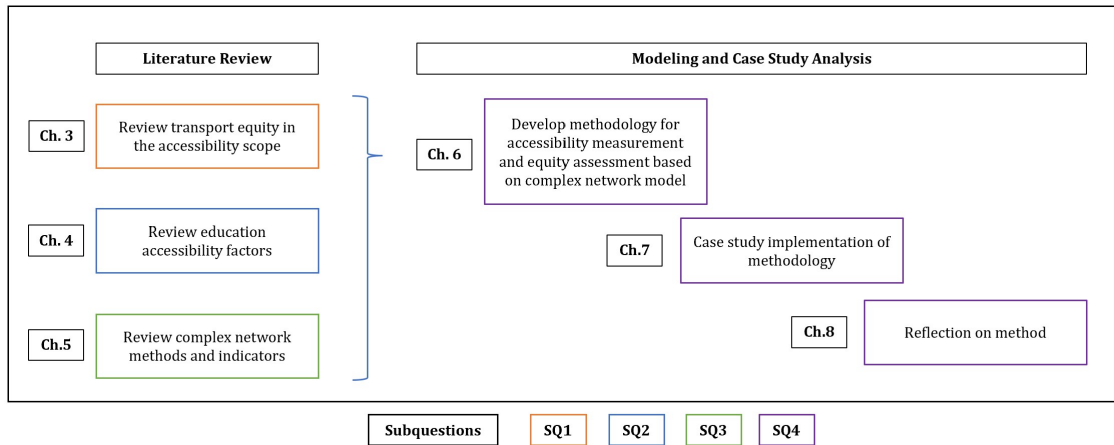
Figure 1.1: Research flow with respect to methodologies, chapters, and subquestions

and academic fields. Then, the key concepts are described. Chapter 3 describes the literature review on transport equity, specifically, equity measurement, and indices found in the literature. Moreover, accessibility-based equity concepts described and the types of accessibility measures are investigated. Chapter 4 expands on the factors that impact accessibility in the context of land use and transport systems. Then, common factors regarding accessibility to jobs and education are conceptualized and inferences are made regarding education accessibility factors. Chapter 5 reviews the complex network theory based approaches in transport research, including metrics and studies of network evolution. Chapter 6 synthesizes the findings of the literature reviews on core concepts to develop a methodology for accessibility measurement and equity assessment based on complex network theory. The application of this methodology to the Helsinki case study is provided in chapter 7, including the background of case study region, data collection, exploration, accessibility measurement, spatial and equity analysis. Chapter 8 reflects on the applied methodology, focusing on the overall applicability of complex network approaches to accessibility-based equity evaluation. This report ends with the conclusion section where main findings are discussed and recommendations for future studies are presented.

2 | LITERATURE REVIEW CONCEPTS

This chapter introduces the methodology of the literature review followed by the core concepts of this study. First, the research keywords and the flow of the literature study are described. Next, the core concepts are defined and introduced, namely spatial inequality, transport equity, accessibility, and complex network theory. This section serves to describe how the research is formulated for the following chapters, which present the results of the literature review on three concepts, equity, accessibility factors, and complex network theory in transport research.

2.1 METHODOLOGY

This sub-section discusses the methodology for resource identification and selection for the literature review. The main databases used for this review were Scopus and Web of Science. This study consists of three distinct literature review sections in chapters 3, 4, and 5. Each of these reviews have their own keywords and research queries as described below:

Chapter 3:

- (*equality OR *equity) AND (transport OR mobility) AND access* AND (indicators OR index)

Chapter 4:

- urban AND transport* AND "land use" AND access* AND services
    - AND (jobs OR employment OR work)
    - AND (school OR education)

Chapter 5:

- ("complex network" OR network OR "graph") AND urban AND (transport OR mobility OR road)
    - AND (metric OR indicator OR index)
    - AND (evolution OR temporal OR historical)

For the first and second queries, equality and equity were included as alternatives, since some studies use them interchangeably when describing the equity measurement methods. For the extensions of some words such as *access, accessibility* or *inequality, equality*, the loose phrase operator "*" was used.

During the initial steps of the review, literature review papers on the core concepts were utilized. This method was specifically useful for identifying relevant studies that did not appear in the specified queries. These review articles and their focus concept are listed in Table 2.1.

Table 2.1: Literature review studies that focus on or touch upon the key concepts of this study

Author(s), year	Concepts	Field
Barthelemy, 2011	Complex networks	Urban research
Ding et al., 2019	Complex networks	Transport, Urban research
Geurs & van Wee, 2004	Accessibility	Transport
Kasraian et al., 2016	Accessibility	Transport, Land use
Litman, 2018	Transport equity	Transport
Martens, 2019	Transport equity	Transport
van Wee & Mouter, 2021	Transport equity, Accessibility	Transport
Vecchio et al., 2020	Transport equity, Accessibility	Transport

Specifically, articles cited For the complex network studies, Barthélemy [2011] and Ding et al. [2019] offer an overview of applications in transport and urban research. In the context of equity in transport, Litman [2018] and Martens et al. [2019] are used to identify equity measurement studies, and van Wee and Mouter [2021] and Vecchio et al. [2020] for accessibility measures in the context of equity. Geurs and van Wee [2004] is one of the most commonly cited review studies on accessibility measurement and Kasraian et al. [2016] reviews the impact of land use and transport systems on several concepts including accessibility. The research queries resulted in research from both urban planning and transport research domains, while some overlapping literature studies were also studied.

Figure 2.1 shows the literature review flow of chapter 3 in the form of PRISMA diagram. Diagrams describing the flows of structured literature reviews of the other two chapters are given in the Appendix A. Publishing year of the papers were not filtered in the queries, nonetheless most papers identified as main resources were published within the last two decades. Snowballing was used to delineate the key literature cited in the reviewed studies, as well as reverse snowballing to identify newer articles citing the key literature.

2.2 CORE CONCEPTS

This section defines the core concepts of the research, such as transport equity, accessibility, and complex network theory.

2.2.1 Transport Equity

The concept of equity is defined by Martens et al. [2019] as the "morally proper distribution of benefits and burdens over members of society". The way in which the distributive aspect of this definition is considered leads to another concept, that is, fairness. An equal distribution to every individual is an egalitarian perspective of equity, yet it is not always fair due to real-world social circumstances [Sun and Zacharias, 2020]. Another approach is a utilitarian case where the benefit to societal welfare is the main goal. This could lead to ignoring individual needs of certain groups. The Rawlsian approach attempts to address this by seeking to improve the
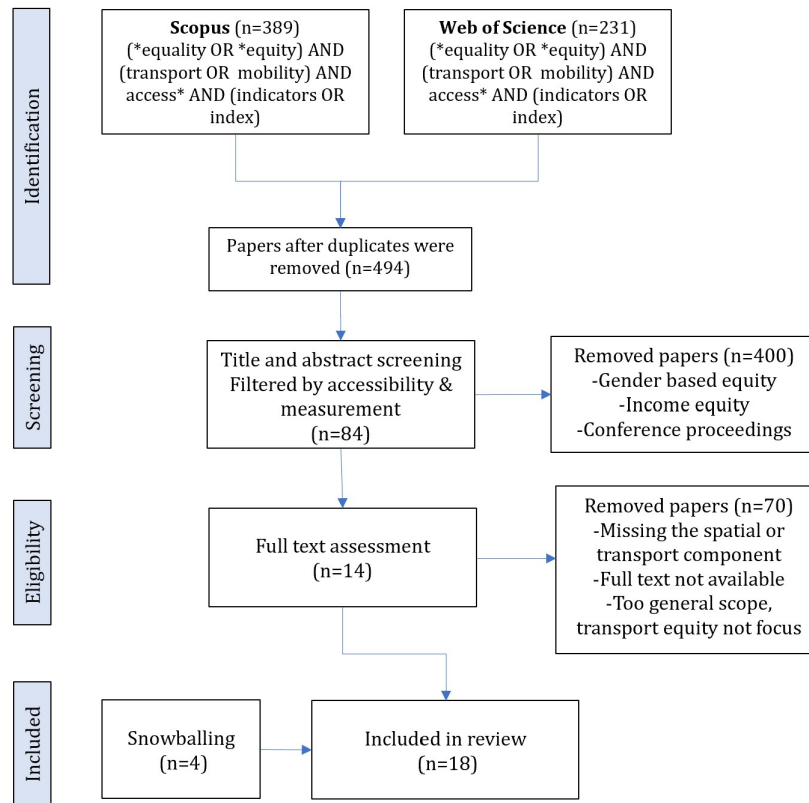
Figure 2.1: PRISMA diagram of the literature review of transport equity (chapter 3)

average benefit while paying more attention to the needs of the disadvantaged [Sun and Zacharias, 2020]. Building on these discussions, van Wee and Mouter [2021] classify different equity types and associate them with three areas of transport policy; namely accessibility, safety, and environment. Among the discussed equity types are vertical, horizontal, territorial, egalitarian, and spatial equity.

2.2.2 Accessibility

Geurs and van Wee [2004] define accessibility as the "extent to which land-use and transport systems enable individuals to reach activities or destinations by means of a (combination of) transport mode(s)". Vecchio et al. [2020] develop a framework which is grounded on the chosen social issue, the ethical goal and theories. The underlying concepts develop into the accessibility components adapted from Geurs and van Wee [2004]. These include the land use, transport, temporal, and individual components. The land-use component of accessibility refers to the location of opportunities, whereas transport refers to infrastructure. The individual component is the socioeconomic and demographic information, and the temporal component refers to the time availability of individuals [Geurs and van Wee, 2004]. Bertolini [2012] define a feedback cycle to conceptualize the relationship of accessibility with transportation and land use.

2.2.3 Complex Network Theory

Complex network theory is an analysis approach that aids the understanding of complex systems that are difficult to envision solely based on the behavior of their individual components. It relies on a mathematical notation where individual elements are represented as nodes, and their connections or interactions are represented as edges in a comprehensive graph [Mata, 2020]. *Connectivity* of a graph is defined on the basis of the existence of a connection between every pair of nodes. On the local (node) level, *centrality metrics* (e.g.betweenness centrality, closeness centrality) consider the number of direct connections to other nodes and the node's role in connecting other node pairs [Barthélemy, 2011]. Community indicators are used to identify clusters within the network, and topological indicators investigate the development structure based on the size of the network, the number of connections and the density of network components in an area of interest [Cats, 2017]. Lastly, accessibility indicators seek to understand how human movements would be influenced by network topology and evaluate the level of reach of sets of nodes within a network [Lee and Kim, 2021].

Transport networks are analyzed through to complex network metrics understand network characteristics, dynamic processes, communities within systems, resilience, and robustness properties [Ding et al., 2019]. Adopting a temporal approach, network evolution studies investigate developments in network characteristics in terms of topology or community structures. Furthermore, co-evolution studies investigate the interplay between transport systems and other urban systems such as land use, or spatial distribution of economic and demographic attributes.

2.3 CHAPTER SUMMARY AND DISCUSSION

This chapter described the systematic literature review methodology that forms the basis of the literature reviews carried out in the chapters 3, 4, 5. The core concepts of this study were defined primarily using the review studies found during the initial investigation. The takeaway from these definitions is that from an ethical perspective, equity is not a definitive concept when distribution of a resource in the society is concerned. The definition of equity depends on the perspective taken by the researcher, as well as how diversity within the society is approached. As the measurement of equity is more specified, accessibility emerges as a resource for evaluating the fairness of distribution among groups of people. There exist approaches in the literature for the measurement of accessibility, referring to most, if not all, of the components (land use, transport, temporal, individual) defined by Geurs and van Wee [2004]. An emerging method for accessibility measurement is complex network theory which is used in transport literature, though commonly for the purpose of network efficiency, topology, or community identification. The later sections of this report will therefore focus on the applicability of the complex network approaches to the accessibility use case and, ultimately, transport equity evaluation.

3 | TRANSPORT EQUITY

This chapter discusses the topic of equity in the context of transport research. First section of the chapter explains the components of an equity measure, such as the benefits and burdens, differentiation factors, and the equity principles. Next section gives an overview of the equity indices used in the transport research literature. The final section discusses the accessibility use case of equity measurement based on the concepts in the first section, followed by an overview of the types of accessibility measures used in the literature. The chapter ends with the conclusions of the literature review relevant to this study.

3.1 EQUITY MEASUREMENT

Choosing and implementing appropriate measures and indicators for a robust equity assessment remains to be a challenge in transport studies [Martens et al., 2019]. Three components are defined by Martens et al. [2019] relevant for the formation of equity measures. First, is the definition of properties that are regarded as *benefits* and *burdens*. Second component is the selection of the *differentiation factor* to investigate the effect on groups of people. An example of this factor is income, which is widely used in assessment of equity. Third, is the *equity principle*, which is the discussion of how burdens and benefits should be distributed (e.g. egalitarian, utilitarian).

Benefits and Burdens

For the selection of benefit or burden indicators, Martens et al. [2019] defines four local variables: resources, opportunities and risks, outcomes, and well-being. Possessions of individuals and conditions they are subjected to are investigated under resources, such as car ownership, proximity to public transport, or local air quality. Opportunities and risks are induced by resources, which depend on individual cases, such as increased life quality by driving a car to some, or the reverse effect when roads are heavily congested. Outcomes signify the results of resources and opportunities together, which can be objectively measured. For example, the time spent traveling each day per individual. Well-being is an individual's evaluation of their mental state based on their experience. Martens et al. [2019] investigates these variables in different use cases, accessibility/mobility being the relevant one for this study, described in section 3.3.

Differentiation Factors

The topic of how individuals are differentiated in equity research is important to ensure the distribution of a property considers everyone. Studies focusing in transport research on the distribution of benefits identify the disadvantaged groups more often as low-income persons and non-car-owners. In various research, age, ethnicity, gender, and disabilities are also regarded as indicators of disadvantaged groups. The evaluation of distribution of burdens, on the other hand, seeks to find vulnerable groups that are more affected. For example, in the case of air pollution, people with respiratory diseases, children, or the elderly are more vulnerable.

Among the factors of population distinction is the residential location of people. It is often the case that disadvantaged groups reside in neighborhoods that are less attractive based on their subpar proximity to activities, and limited supply of public services [Martens et al., 2019]. Nonetheless, the heterogeneity of zones and neighborhoods should also be considered, meaning that not all individuals experience an equal disadvantage. Differences may be related to other factors such as car ownership or the effectiveness of public transport services.

While car ownership often is often associated with higher accessibility in many regions, the actual effect of car use on accessibility depends on the region El-Geneidy et al. [2016]. Nonetheless, it is a way of distinguishing groups of people which can be effective when coupled with other factors. For instance, low-income car owners still bear the burdens of car use financially. Thus, owning a car does not always translate into being able to use it when coupled with financial factors. Furthermore, a household owning a car does not mean it will be available to all members. Considering all of these limitations, car ownership needs to be evaluated with caution, especially when disaggregating groups of people based on this indicator.

Equity Principles

The principles which form a guide for equity principles are examined under two use cases. One is the measurement of equity in the current state, and the other is the equity of an intervention. The intervention is a set of actions that aim to move the current state to the most equitable state [Martens et al., 2019]. This study is interested in the former, that is the evaluation of the current (and past) state. There are several types of equity principles in literature which are relevant in the transport systems' assessment. Litman [2018] defines three categories of transport equity. *Horizontal equity* refers to the egalitarian approach where groups of individuals are treated equally and no distinction is made among groups based on their characteristics, stature, or needs. *Vertical equity* on the other hand, makes a distinction among groups which have differing economic or demographic properties. Thus, an vertically equitable distribution favors disadvantaged groups to offset their lower stature in the distribution of amenities. The distinction may be made in terms of *income and social class*, or *mobility need and ability*. The latter considers individuals' abilities and needs, and groups with special needs are favored. Nonetheless, the types of equity principles are not limited to the categories presented by Litman [2018]. In their review, van Wee and Mouter [2021] compile several more types of equity commonly referred to in transport research. Among these types, *Territorial equity* considers the

equity among regions or districts, where deprived regions would be prioritized in accessibility improvement. Another type which applies to all other types would be *spatial equity*, which refers to the geographical location of differentiated groups in equity assessment.

3.2 EQUITY INDICES

As discussed in the previous section, the measurement of equity involves layers of factors making it a complex procedure to determine the optimal method in a use case. There are several equity indices in literature used in transport research to accurately identify distributions accurately. In this section, the identified indices in transport and urban research are discussed with their use cases, advantages or disadvantages compared to other common indices. The indicators to be discussed are namely, Gini index, Spatial Gini index, Theil index, Atkinson index, and Kolm-Pollak.

Gini Index

Gini index or coefficient is a value used to evaluate how a resource is distributed in the overall population. It is obtained from the assessment of the Lorenz curve, which graphically indicates the distribution of a resource. Gini index is derived from the area between the line representing the 100% equal distribution and the actual distribution (Lorenz curve). Therefore, Lorenz curve is a way of visually representing the equity, whereas Gini index is used to numerically express the overall equity [Delbosc and Currie, 2011]. The index is often displayed in a two-axis graph, where the X-axis displays the unit of measurement (regional, individual, etc.), and the Y-axis is the cumulative sum of the chosen distribution [van Wee and Mouter, 2021]. The a larger Gini index indicates a more unequal distribution. The formulation of the Gini index approximation is given in Equation 3.1, adapted from Delbosc and Currie [2011].

$$G = 1 - \sum_{k=1}^{n} (X_k - X_{k-1})(Y_k + Y_{k+1}) \tag{3.1}$$

where,

X_k is the cumulative proportion of the variable in population

Y_k is the cumulative proportion of the accessibility

In the transport equity context, the resource which is evaluated in terms of its distribution is likely to be accessibility.The disaggregation metric used in such equity analyses depend on aim of the research, as previously discussed in 3.1. Most common metrics used are income, age, and location-based variables. Lucas et al. [2016] emphasize one advantage of Gini index as the scale-independence, which suggests that when the measurement scale is altered (change in currency or aggregation period), the outcome is unaffected. However, it should be noted that for distance- or contour-based accessibility measures, selected thresholds can influence the results. For example, the selection of a value a time contour could influence the results sig-

nificantly, considering the average travel times differ significantly in urban and rural contexts, and Gini index would be greatly affected this selection [Lucas et al., 2016].
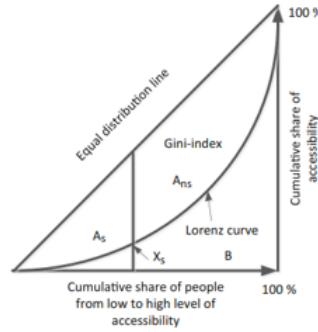
Figure 3.1: Visualization of accessibility-based equity with Lorenz curve [Lucas et al., 2016]

Spatial Gini Index

Spatial Gini is a decomposition of the original Gini index in terms of the equity with respect to near locations and far locations. Developed by Rey and Smith [2013], this index relies on an alternative approach which considers spatial autocorrelation among distant locations. Ultimately, a new formulation is developed still respecting the original Gini index equation, shown in 3.2. The first term represents the near differences whereas the second term refers to the far differences in the selected attribute. As the near difference approaches zero, meaning that pairwise differences are non-existent, the far differences contribute to most of the inequity in the study area. Therefore, a spatial dependence is observed in such a case.

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij}|x_i - x_j|)}{2n^2 \bar{x}} + \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} ((1-w_{ij})|x_i - x_j|)}{2n^2 \bar{x}} \tag{3.2}$$

where,

w_{ij} is the binary spatial weights (or adjacency matrix)

representing the neighbor relationships

n is the number of observations

x is the observed attribute for node i and its pair j

For the verification of outputs in terms of inference, a permutation based approach is used, as described by [Rey and Smith, 2013]. Based on this approach, observed attributes are randomly distributed to different locations, therefore testing the null hypothesis if similar results can be obtained in a random distribution. The resulting equity outputs for multiple iterations are compared with the original results for the analysis of statistical significance, obtaining a p-value.

Spatial Gini index is proposed as a solution to the lack of the spatial component of the conventional Gini index. Panzera and Postiglione [2020] apply a variation of this method on the spatial income equity in Italy. Their results indicate a successful separation of spatial and non-spatial component of equity. Rey and Smith [2013]

maintain that this index can be used to measure the *spatial mismatch* phenomenon in the case of housing and job locations.

Theil Index

Theil index [Theil, 1967] was developed to evaluate the level of inequity of a distribution. One part of this index indicates the differences between groups of people, whereas the other part indicates the inequity within groups. It is a measure that can be used for the assessment of accessibility inequality across geographical units. [Liu et al., 2021] measure the inequality of regional development with respect to an output (e.g. gross regional product) using Theil index. However, one shortcoming of this index is the limited ease of understandability for policymakers [van Wee and Mouter, 2021].

Essentially, this index relies on the entropy approach, seeking to understand the disorder in a region of interest. It measures how much the current state deviates from the most ordered state, which is the egalitarian case of perfect equality.

For this estimation, the population is divided into mutually exclusive groups. Thus, equity is assessed with respect to the distribution of the selected value (e.g. accessibility) across these groups. These groups can be defined as geographical units (GU), aggregating the values in a selected area within the region. Examples of such could be state-defined districts, or regions of equal population, defined by the analyst.

The entropy is measured with respect to the mean value in the whole region. Thus, if the aggregate value in a GU is equal to the regional average, the contribution to entropy (i.e. inequity) is zero. As the value in the GU deviates away from the mean, the inequity increases. Thus, the zero Theil index value represents perfect equity, whereas values of higher magnitude represent higher inequity.

The formulation of the Theil index calculation, adapted from Liu et al. [2021] is provided as follows:

$$Th = \sum_{i=1}^{m} \left(n_i \frac{y_i}{\bar{y}} \ln \left(\frac{y_i}{\bar{y}} \right) \right) \tag{3.3}$$

where,

m is the total geographical units (GU) in the region

y_i is the average measure being analyzed in the GU , e.g. income

$\bar{y}$ is the average measure being analyzed in the region, e.g. income

n_i is the ratio of population in GU over total population in the region

The perfect equity case would be represented with a Theil value of zero, whereas the least equitable case would result in a value of $ln(m)$. Developing on the original formulation, Camporeale et al. [2019] decomposes the Theil index into two components: between and within Theil. Two terms of the decomposed Theil equation can be either positive or negative. When positive, the component contributes to the inequity, whereas when negative, a contribution is made towards equity. By the nature of the equation, the positive term is always larger than the other, thus achieving a

positive Theil value. Their study utilizes this method in an accessibility use case, comparing the effect of a pricing scheme on the accessibility of road network users [Camporeale et al., 2019]. It is noteworthy that this type of decomposition is similar to the Spatial Gini, however it does not consider the neighborhood effects. In other words, the spatial configuration of geographical units do not have an impact on the results of a Theil index calculation.

$$Th = WITHIN + BETWEEN$$

$$Th = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{1}{P_T} \frac{y_{ij}}{y_i} \ln \left(\frac{y_{ij}}{y_i} \right) \right) + \sum_{i=1}^{m} \left(\frac{p_i}{P_T} \frac{y_i}{\bar{y}} \ln \left(\frac{y_i}{\bar{y}} \right) \right) \tag{3.4}$$

where,

m is the total geographical units (GU) in the region

y_i is the average measure being analyzed in the GU , e.g. income

$\bar{y}$ is the average measure being analyzed in the region, e.g. income

p_i is the number of people in the GU

P_T is the number of people in the region

Atkinson Index – Equally Distributed Equivalent

Originally intended to measure income inequality, Atkinson index proposes an equally distributed equivalent (EDE) measure. This measure expresses a value for the case where everyone would have equal distribution and the total welfare would not change [Logan et al., 2021]. Thus, inequality is penalized and general welfare is prioritized in this index, yet the judgments regarding the concept of general welfare is subjective. Furthermore, this index is not suitable for the evaluation of distributions of undesirable properties (e.g. pollution) [Logan et al., 2021]. Taking an egalitarian approach based on subjective definitions of a greater good, this equity also lacks a multivariate assessment for combinations of benefits and burdens. Thus, it is not applicable for an equity assessment based on the multiple components of accessibility such as transport and land use factors.

Kolm–Pollak

Kolm-Pollak approach uses the EDE together with an inequality index in a novel form. The EDE represents the case where the existing distribution would not be favored over individuals having the same value. The inequality aversion parameter depends on the historical measure of society's indisposition to inequality. If the aversion parameter is zero, the EDE is equal to the mean of the distributed property, whereas when the aversion is very high, the EDE approaches the value of the most disadvantaged groups. Therefore, the inequality index is the difference between the mean of the distribution and the EDE. This index only measures a single variable, and is prone to under-estimate inequalities. However, it is suitable for measures of distance-based accessibility [Logan et al., 2021]. One major shortcoming is the

need for historical inequality data to asses the inequality aversion parameter. In the reviewed studies, this parameter is often used as a sensitivity factor of the model to determine the extent the outputs are affected by this value. Therefore, this index is particularly useful for comparisons of different case studies where this parameter would be differing.

3.3 ACCESSIBILITY MEASURES

Accessibility measures are of great significance to understanding the effectiveness of transport system to enable the reach of activities. Using the approach of Martens et al. [2019], the benefits and burdens of an accessibility based equity would enable a better understanding of the type of measures necessary for assessment. Firstly, resources which are relevant for accessibility are often connected to the transport means available to a person. The most straightforward example of a resource would be owning a car, however other types of accessibility resources also exist, although less intuitive to measure. When public transport is considered as a resource, factors such as number of stops, proximity, or the frequency of trips can be listed. Furthermore, walkability or cyclability of the road network could also be regarded as a resource. Opportunities related to accessibility are the movement potential of a person, determining the area which they can traverse through using travel modes. Another potential that is brought about by movement is the participation to activities and services. Outcomes of accessibility analysis are used to indicate differences among population. It is possible to collect data of outcomes using traditional methods such as surveys, and big data based on ICT systems. Examples of such data are travel times, distances, or expenses in the population. Lastly, the well-being component refers to the level of satisfaction due to their mobility experiences.

Table 3.1: Reviewed accessibility-based equity studies including th etype of accessibilty measurement, differentiation factors, equity principles and indices used

Author(s), Year	Accessibility measure	Equity differentiation factor	Equity principle	Equity index
Cui et al., 2019	Location-based (Distance)	Income	Vertical equity	
Dixit & Sivakumar, 2020	Utility-based (logsum)	Income, Location, Age	Spatial equity, Horizontal equity	Gini index
Camporeale et al., 2019	Infrastructure-based	Location, Income	Spatial equity, Vertical equity	Theil index
Guzman et al., 2017	Location-based (Potential)	Income	Horizontal equity, Vertical Equity	Gini index
Järv et al., 2017	Location-based (Dynamic)	Location	Spatial equity	Gini index
Liu & Duan, 2020	Location-based (Distance)	Location	Spatial equity	Theil index, Gini index
Sharma et al., 2022	Location-based (Distance)	Location	Spatial equity	Gini index

All of the categories of benefits and burdens associated with accessibility need to be operationalized in order to be measured. Investigating how such operationalizations occur in the transport and land use context Geurs and van Wee [2004] cate-

gorize the accessibility measures under four main categories: infrastructure-based, location-based, utility-based, and person-based measures. Accessibility-based equity studies in the literature are analyzed under these four categories in the following paragraphs. The overview of the reviewed studies can be found in Table 3.1.

Infrastructure–based

Among the accessibility measures used for transport policies and transport system evaluation, infrastructure-based measures have an important role. Metrics specific to transport systems such as travel times, operating speed and congestion are used to evaluate policy decisions [Geurs and van Wee, 2004]. One shortcoming of such metrics is the omission of land use influence, as well as limited use for individual case attributes. Camporeale et al. [2019] uses this type of measure in their study to assess accessibility-based spatial equity. Their measurement is based on the transport network, considering infrastructural factors such as congestion and travel costs to determine how accessibility across traffic zones are affected [Geurs and van Wee, 2004].

Location–based

The simplest form of distance measures are used to measure accessibility based on the straight line distance between two points. When more than two points are analyzed, contour measures (isochronic measures) are utilized. These measure the aggregate proximity to opportunities, or the count of activities that can be reached [Geurs and van Wee, 2004]. The advantages of such measures are the ease of interpretation and the low need for data. Despite considering the land-use and transport factors, these measures fail to factor their combined effect. Furthermore, competition and limited capacities are not taken into account, specifically for opportunities such as jobs, schools, and hospitals [Geurs and van Wee, 2004].

Distance measures

Distance measures are a subcategory in location-based accessibility measures, also called *connectivity measures*. An example of distance measures would be relative accessibility, which was developed to measure the accessibility between two points. Some measures in this subcategory can also fall under infrastructure-based measures, if for instance average speed between two points using the road network are used to measure accessibility. If location-based measures are considered, a straight line between two points is more commonly the case. Such measures are applied in the area of land use to set the maximum travel distance (or time) between two locations [Geurs and van Wee, 2004].

Contour measures

Contour measures are similar to the distance measures, however an analysis of multiple destinations, rather than only two, are considered in their use cases. Counts

of a number of opportunities to be reached in a time threshold can be an example of a contour measure. Urban planning and geographical studies make use of this measure [Geurs and van Wee, 2004].

Potential accessibility – Gravity method

This type of accessibility measure considers the distance of opportunities and assigns a lower likelihood to more distant ones. Generally this approach uses a negative exponential cost function, however several other forms of impedance functions exist such as Gaussian or logistic functions [Geurs and van Wee, 2004]. The most general form of such functions can be expressed as the following based on Hansen [1959]'s formula:

$$S_i = \sum_j O_j f(C_{ij}) \tag{3.5}$$

Where S_i is the accessibility of location i and O_j is the number of opportunities at location j, $f(C_{ij})$ is the impedance function based on the cost (i.e. travel time). This formula is only valid under two assumptions. First, the demand is uniformly distributed. Second, the available opportunities are not limited by capacity. Shen [1998] maintains that first assumption is not valid in urban environments as activities, firms, and people are not evenly distributed across the city. The second assumption is not valid for rival goods unless major facilities such as national parks are considered. In the case of employment, job districts and residential areas of laborers are often unequally distributed in the city. Therefore Shen [1998] offer another form of the potential accessibility function which includes the demand potential for every location. This function also includes separate weights for travel modes, therefore considering the effect of mode choice on accessibility.

Dynamic accessibility

Most location-based accessibility measures fail to consider the temporal component of accessibility, with an assumption that people move from their homes to relevant opportunities. Järv et al. [2018] approach this limitation where location-based accessibility measurements are formulated as a function of time. Their methods incorporate spatial patterns of travel behavior in different time-steps with the temporal sensitivity of transport modes. Travel duration throughout the day differs whether it is a factor of public transport timetables or travel times in private cars with respect to congestion. Another factor is the temporal availability of activities and potential opportunities, considering the day-night, and weekday-weekend variations [Järv et al., 2018]. Their analysis is done on a case study where accessibility to food stores are analyzed.

Person–based

These measures approach accessibility in terms of space and time constraints imposed on individuals' traveling patterns. Space-time approaches consider areas of

opportunities and the extent they can be reached within specific time-constraints [Geurs and van Wee, 2004]. The advantage of these approaches is that individual differences in accessibility can be observed based on individual background (i.e. ethnicity, gender). Nonetheless, person-based approaches fail to consider competition effects on the supply side, focusing on the demand. As a result, it is not ideal to use these measures on accessibility to opportunities where capacity is the limiting factor such as the job vacancies [Geurs and van Wee, 2004].

Utility-based

Utility-based accessibility is based on the choice behavior modelling approach, where accessibility is explained through individiual choices. Random utility maximization theory developed by Ben-Akiva and Bierlaire [1999], also known as the multinomial logit or logsum measure, assumes that individual choices serve to maximize a utility function of travelers. Another approach is the doubly constrained entropy model, which includes additional consideration of the competition effects within balancing factors [Geurs and van Wee, 2004]. These measures make it possible to evaluate user benefits of land-use and transport projects, therefore addressing the drawback of infrastructure-based measures. Moreover, utility measures assume a non-linear relationship between accessibility and user utility, therefore low-accessibility regions would benefit from accessibility improvements to a higher magnitude than high-accessibility regions. This property is especially relevant from the vertical equity perspective. Dixit and Sivakumar [2020] uses a utility-based logsum measure with land-use and transport attributes as well as individual characteristics to determine accessibility patterns in transport equity analysis. Nonetheless, this approach does not offer an additional insight to the spatial distribution of accessibility geographical zones [Dixit and Sivakumar, 2020].

3.4 CHAPTER SUMMARY AND DISCUSSION

This chapter presented the results of the literature review seeking to answer SQ1. The first section investigated the components of an equity measure, which was then exemplified in the last section in the accessibility use case. The takeaway from this section is that during equity measurement, identification of the impacts of the measured phenomenon is as important as the ways of disintegrating the population. Furthermore the way equity is assessed is predominantly based on the governing equity principle. Transport research commonly utilizes vertical, horizontal, and territorial equity types. What is most interesting for this study is the measurement of spatial equity. Therefore the developed equity measurement methodology should not only consider the distribution in the population, but also the geographic locations of the accessibility indicators. The second section of this chapter investigated the commonly used equity indices in literature. The most common index is the Gini coefficient which measures the distribution of a resource among the population. Gini index is often used in accessibility equity measurement, however lacks a spatial component. To address this shortcoming, Spatial Gini can be used to assess the

neighbor relationship effects on equity. Theil index is another common metric that also considers the equity across geographical units such as subdistricts, therefore falling under the scope of territorial equity. It is potentially useful for the equity assessment of spatial accessibility distribution, but its complexity makes it hard to communicate to policy makers. Thus, more common and easy to understand index such as Gini/Spatial Gini can be coupled with Theil index.

Lastly, the types of accessibility measures are compiled with their use cases in literature. For a comprehensive assessment of accessibility in the urban environment infrastructure- or location-based measures are found to be more applicable. However, an appropriate accessibility measure should include both transport and land use components, therefore analyzing accessibility solely within the transport network would not yield an outcome that can be used as a social indicator due the lack of information of where activities are located.

4 | ACCESSIBILITY

In this chapter, the outcomes of the literature review regarding accessibility in transport studies are presented. The first section describes the role of accessibility in the transport and land-use context. The next two sections review the accessibility studies in the context of job and school opportunities to identify common factors. The review is concluded by the conceptualization of common factors and findings relevant to this study to answer the SQ2.

The structured literature review process for this chapter is visualized in Figure A.1. In addition to the search queries defined in chapter 2, an additional query filter was added to review the studies with a discourse on equity. Among the other accessibility studies, these were assessed to be more relevant for this research. Throughout the review process, it was identified that accessibility to services and opportunities n the context of education had connections with the job accessibility studies, yet received less specific attention. Therefore, a comparative literature review approach was preferred to identify common factors influencing accessibility.

4.1 ACCESSIBILITY IN TRANSPORT AND LAND–USE CONTEXT

Accessing a desired destination is contingent upon factors governed by systems and external conditions. It is important to define the context in which accessibility is defined. After all, a person might have very high mobility, yet their desired location may not be within reach. On the contrary, an individual might not have the means to move across great distances, yet their desired location could be in their proximity, leaving their disadvantage irrelevant. Therefore, accessibility differs on a case by case basis depending on the individual preferences, but also locations of opportunities and the adequacy of the transport system.

Accessibility is widely used by transport researchers to understand the effect of policy and infrastructure developments on urban systems and their users. Geurs and van Wee [2004] define four components of accessibility, two of them naturally being land-use and transport. The former involves the supply and demand of spatial distribution of socioeconomic opportunities, as well as the potential competition on both sides due to capacity limitations. The transport component consists of the transport system, including the demand by passengers traveling within, and the infrastructure as in supply. Other two components are namely temporal and individual components. Temporal component deals with the time based availability of supply and demand. Lastly, the individual component signifies the characteristics of individuals in the context of demographics, economics, or abilities [Geurs and van Wee, 2004]. In addition to these components, Vecchio et al. [2020] develop their framework based on social-spatial inequalities. Their approach involves an ethical stance

which defines a social issue to be faced. This stance defines how the components of accessibility are assessed.
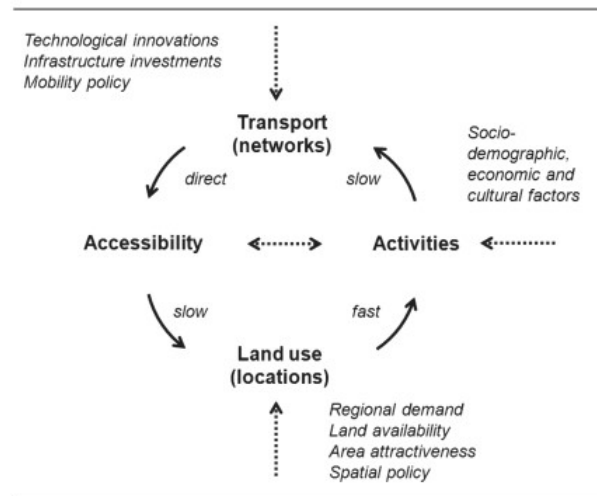
Figure 4.1: Transport land-use feedback cycle [Bertolini, 2012]

Bertolini [2012] explains the nature of transport land use interaction in a feedback cycle where exogenous factors such as innovations, policy, and land availability are also concerned. In this systematic view, location of activities carried out by people are driven by land use policies. Transport system serves the purpose of moving people from and to activities. The developments in the transport system are driven by the demand for these activities, which in turn determine the accessibility of a location. External factors such as land availability and attractiveness of a location drive land use developments together with the accessibility. The cycle is demonstrated in Figure 4.1.

4.2 ACCESSIBILITY TO JOB AND EDUCATION OPPORTUNITIES

This section provides an insight into the underlying factors related to accessibility to opportunities. Based on the SQ2, the main field of consideration is the education accessibility. Despite being an important topic for sustainable development, education accessibility has received less attention compared to the likes of accessibility to employment [Sharma and Patil, 2022]. Therefore, a limited number of studies exist compared to the likes of job accessibility. Considering the previously discussed mechanism of accessibility in the transport and land use context, the accessibility to job and education opportunities are expected to share some common factors. Furthermore, both are important enablers of economic participation. The "spatial mismatch" hypothesis developed by Kain [1992] suggest a direct relationship exists between urban structure and the performance of disadvantaged communities in the labor market. The studies of Jin and Paulsen [2018] and Bastiaanssen et al. [2022] confirmed that improved job opportunity accessibility of disadvantaged groups has an impact on reduced unemployment rates. When the case of education is consid-

ered, a direct relationship between level of education and employment likelihood is observed Di Paolo et al. [2017]. In several studies, a significant relationship between school accessibility and continuing to higher levels of education has been identified [Dickerson and McIntosh, 2013][Sá et al., 2006]. Therefore, accessibility to both jobs and education opportunities are deemed to have a great impact on employability, and ultimately, economic participation. The factors related to accessibility to these opportunities are discussed in the following subsections. This section benefits from the extensive studies in the job accessibility literature, and associates them with the education accessibility studies to form a common conceptualization. The reviewed studies to identify common accessibility factors are tabulated in Table 4.1.

Table 4.1: Studies used to identify accessibility factors of education and jobs opportunities

	Opportunity	Affordability	Availability	Proximity
Andersson et al., 2012	Education	X		X
Cui et al., 2019	Jobs		X	
Dickerson et al., 2013	Education			X
El-Geneidy et al., 2016	Jobs	X		
Foth et al. 2013	Jobs	X		
Guzman, 2017	Jobs & Education	X	X	X
Kain, 1992	Jobs		X	X
Kawabata & Shen, 2007	Jobs	X		X
Levinson, 1998	Jobs		X	X
Sá et al., 2006	Education			X
Salonen & Toivonen, 2013	Jobs			X
Sharma & Patil, 2022	Education	X	X	
Shen, 2001	Jobs	X		
Xu et al., 2018	Education	X	X	X

4.2.1 Job Accessibility Factors

Transport research has been interested in the travelers' journey to work, the trends of movement, and the factors influencing its progression [Cui et al., 2019]. Urban developments as well as societal shifts and technological innovations have influenced the commute times of individuals. Specifically, in the second half of 20th century, a global increase in the commute distances have been observed [Banister, 2012]. Transport studies have been aiming to address this increase through the perspective of accessibility. The most commonly acknowledged factors influencing job accessibility are job and house locations (distance to jobs), level of income, and transport availability. Guzman et al. [2017] explains the dynamic relationship of these three common factors as enablers of individual accessibility, mobility, and job security.

Numerous accessibility studies seek to approach commute behavior from the perspective based on the comparison of housing and job opportunities in a region. In their study, Levinson [1998] investigate the question posed by previous studies: whether urban structure can be used to explain the commute behavior. Their approach considers the municipal region of Washington DC in the years 1987 and 1988. An accessibility analysis is developed based on job availability and housing avail-

ability of traffic zones for both origin and destination of a travel path [Levinson, 1998]. This approach therefore considers the competition effects at both ends of the journey. Accessibility depends on the *availability*, based on the number of jobs and houses, as well as the travel time as an inversely weighing factor. The result of this approach shows that from a worker's perspective, *proximity*, which refers to having job locations closer to their homes, leads to shorter commute times. In addition, increasing the number of houses in the region leads to increased difficulty in finding jobs, assuming that the job supply if fixed [Levinson, 1998]. As a result of their study, commute behavior can be explained through urban structure to a significant degree. Furthermore, it is suggested that commute times are leveling off due to jobs moving to suburbs in a polycentric urban area. This conforms to a hypothesis developed by Levinson and Ajay [1994] that workers tend to adapt to the changes in urban land use changes by either moving their homes or changing jobs. Yet these conclusions are also rebated due to the generalization not applying to all income groups [Cui et al., 2019].

Criticizing the approach of solely counting the number of jobs and workers available in a region, Shen [1998] maintains that a better approach would consider spatial interactions and whether jobs occupationally match workers. Furthermore, an inclusive approach would be needed to evaluate accessibility of disadvantaged groups. Kain [1992] introduced and developed the 'spatial mismatch hypothesis' to describe the risk of unemployment imposed on disadvantaged groups due to jobs moving to suburban areas. According to the study of Shen [2001], in US cities, low-income groups have higher commute times compared to the city average. These results are related to the mode choice of low-income travelers, which is mainly public transport. While the affordability of car ownership has an important role in this outcome, a study by Foth et al. [2013] showed that in the city of Toronto, lower-income areas benefit from higher accessibility to jobs by using public transport. Another study by El-Geneidy et al. [2016] substantiates this conclusion in another Canadian city, Montreal. Their research builds on top of the common accessibility approach of using travel time as the cost variable, as they include travel fares into their methodology. Such an approach creates a new window of *affordability* in accessibility research. The results of this study show that low-income residents are burdened by transit fees, and affordability of transit plays an important role in accessibility [El-Geneidy et al., 2016].

In many US and European urban areas, accessibility of car owners are higher than of transit commuters, however the underlying model assumptions for comparing travel modes often make these comparisons unreliable [Salonen and Toivonen, 2013]. Zooming into the effect of travel mode on accessibility, Kawabata and Shen [2007] compare car and public transit commute times in San Francisco Bay Area between 1990-2000. In this study, accessibility estimation is based on by a specified travel time threshold of 30 minutes for both car and public transport. Their findings not only find considerable differences in the accessibility of public transport and car users in favor of cars, but also a trend of increasing accessibility to jobs for public transport [Kawabata and Shen, 2007].

4.2.2 Education Accessibility Factors

Primary education is a basic human right in 135 countries, however sustaining the access to this right is not always very straightforward. It is a challenge that is being addressed globally. Specifically stated in UNESCO's sustainable development goals, the objective is to "ensure inclusive and equitable quality education and promote life-long learning opportunities for all" [United Nations, 2015]. Access to education is investigated under four categories, namely, spatial accessibility, affordability, acceptability (quality and satisfaction), and appropriateness which considers the provision of facilities for special needs [Sharma and Patil, 2022].

Studies focusing on accessibility often evaluate the *proximity* to primary education opportunities. Dickerson and McIntosh [2013] showed that being closer to primary (or compulsory) education increases the participation to post-compulsory education of higher levels. According to this study, while not the most significant deciding factor, distance is a determinant for young persons who are marginal about continuing their education. A similar study by Sá et al. [2006] conducted on Dutch high school students showed that geographical proximity to professional education increases the probability of continuing their education. Zooming into the impact of *proximity*, Andersson et al. [2012] conducted a comparative analysis of the effect of distance to schools from 2000 to 2006. Their findings indicate that low-income groups are often constrained to schools closer to their homes due to not owning personal vehicles. Therefore, the *affordability* has an indirect impact in the access range of individuals, and an overall increase in the distances to reach schools impairs the accessibility or disadvantaged groups. Mei et al. [2019] approach the school access case from an urban planning case, focusing on the Shenyang area in China. Their results indicate a spatial cluster of basic education is in the center of the city. The accumulation of *availability* at the city center, increases the travel distance and time from the peripheral zones. Their recommendation is to adapt the school supply according to the spatial distribution of demand, which is coincident with a polycentric urban form.

Xu et al. [2018] conduct a historical analysis on the socio-spatial accessibility to urban education in a case study in Nanjing. Their method involves three distinct accessibility indices: geographic accessibility, opportunity availability, and economic affordability. These indices reflect the three main factors they identify with regard to education accessibility, which are the proximity to schools, supply of schools in comparison to housing, and the affordability of access to school districts. These factors overlap with job accessibility indicators explained by [Guzman et al., 2017]. Therefore, a conceptualization of accessibility factors based on job and education opportunities are proposed in the following subsection.

4.2.3 Conceptualization of Accessibility Factors

Based on the literature review conducted on accessibility to job and education opportunities, a conceptualization of the influencing factors is proposed. As shown in Figure 4.2, this conceptualization is based on Bertolini [2012]'s feedback cycle involving transport, land use and activities. Three main concepts are identified as common factors in the accessibility to job and education context. *Proximity* refers
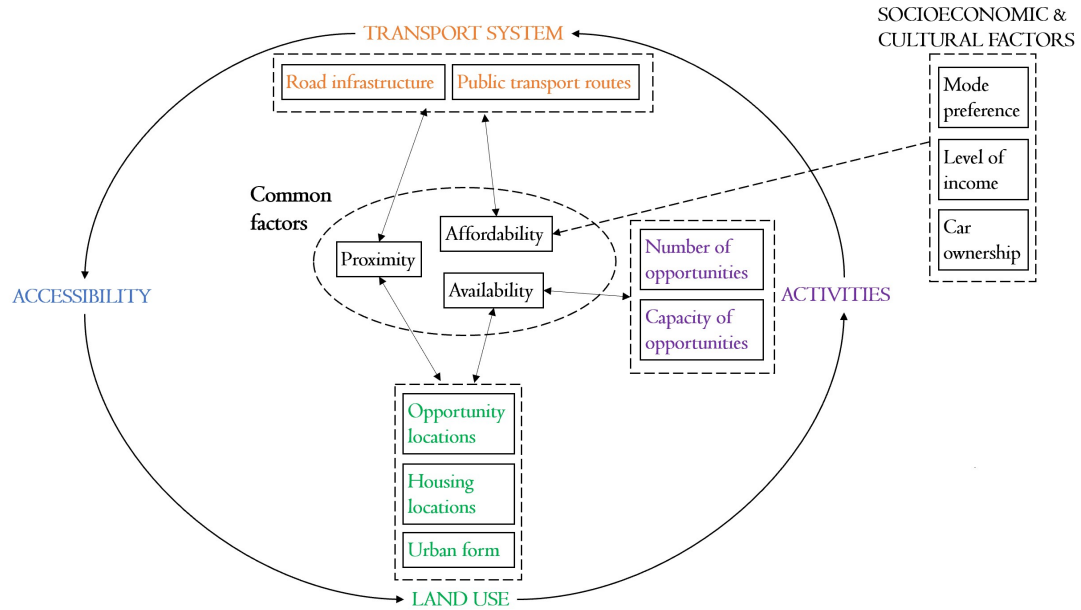
Figure 4.2: Accessibility factors in job and education context

to the distance to the opportunities, which is governed by land use and transport systems' components. Therefore, the distance to an opportunity is defined by the location of where the individual lives, and where the opportunity is located, as well as the impedance defined by the transport infrastructure. *Availability* of opportunities is based on the number of reachable opportunities as well as the capacity of these opportunities, which refers to the activity component of the feedback cycle. The availability of an opportunity is also determined by the land use component, as the spatial distribution is based on the supply of opportunities and the respective demand. Urban form also has a role in the availability, as the historical developments and urban planning determines the overall spatial distribution of activities. Lastly, the *affordability* factor is partly determined by exogenous variables which originate from socioeconomic and cultural factors. The transport system also has a role as in the provision of an infrastructure or public transport service determines the cost related to reaching an opportunity.

4.3 CHAPTER SUMMARY AND DISCUSSION

This chapter presented the outcomes of the literature review of factors influencing the accessibility to education opportunities, therefore investigating SQ2. The literature review on education accessibility benefited from a comparative approach, using the more extensive literature on job accessibility to identify common factors on accessibility. The importance of jobs and education is their direct impact on economic participation, specifically the likelihood of employment of an individual. Accessibility has a significant role in both contexts as an enabler of both indirect and direct participation in the labor market.

In the education accessibility context, reaching higher levels of education enables a better placement in the job market, which is a factor contributing to the employment likelihood indirectly. On the other hand, good access to jobs is a direct factor influencing the chances of being employed. These two types of opportunities are affected by similar factors with regard to accessibility. It is deduced from the literature that three main components of accessibility are often considered: proximity, availability, and affordability. These factors often have shared impact on the level of access of individuals. For example it is evident that the level of income has an impact on car ownership which influences the dominant travel mode by specific groups. Furthermore, it is often the case that proximity to opportunities also influences the mode choice. In addition, the affordability component often determines the travel range of individuals, therefore limiting them to opportunities at a certain level of proximity. Examples of such limitations of these factors are found in both education and job accessibility literature. Disadvantaged groups are often limited to job and education opportunities close to their homes, and by the same token, constrained by the accessibility of the mode choice (often public transport) based on the transport system. If the availability of these opportunities are limited, they are discouraged from economic participation. Therefore, there is a feedback cycle which governs the factors influencing accessibility for both use cases. Consequently, accessibility measurement in these use cases should consider both land use and transport components as well as the economic indicators such as the level of income. From the equity perspective, numerous studies focus on the disadvantaged groups' accessibility to opportunities, hence adopting a vertical equity viewpoint.

This study benefits from this literature review by making use of the conceptualized accessibility factors. The factors in the education accessibility context need to be treated as guiding principles for a comprehensive measurement of accessibility. Consequently, consideration of transport infrastructure, land use, and socioeconomic factors should all be ensured when a modeling and analysis methodology is proposed. Otherwise, the approach will be constrained by the impact of infrastructural effects, given that the complex network model implemented in this study is focuses on the infrastructure topology by principle. The review of applications using the complex network approaches are described in the following chapter.

5 | COMPLEX NETWORKS & URBAN TRANSPORT SYSTEMS

In this chapter, the complex network approaches in the context of transport research are reviewed. The first section provides an overview of complex network metrics under four main categories, namely, distance and centrality, community, topological, and accessibility indicators. Next, the state of art of transport network evolution studies are discussed.

The flow diagram of this structure literature review is provided in Figure A.2. This review was analyzed under two categories, the first part investigated the common complex network metrics and indicators used in transport and urban studies. The second part focused on the "evolution" studies which involve the temporal aspect of network models. Table 5.1 shows an overview of the studies used in this literature review. In addition to these studies Barthélemy [2011]'s extensive literature review of spatial networks was consulted. This review includes almost all of the indicators describes in this chapter, and serves as an extensive guide to the application of complex network theory on urban spatial systems.

5.1 COMPLEX NETWORK METRICS

5.1.1 Distance and Centrality Indicators

A network graph consists of N nodes and E edges which connect certain subsets of nodes. Such connections are represented by an adjacency matrix A of size N x N. The elements of this matrix are shown in 1s and 0s, so if two nodes are connected to each other with an edge, the element of the adjacency matrix corresponding to such relationship is equal to 1. When a weighted graph is considered, an additional matrix W is used where edge weights of node connections are provided. In the case of urban networks, the edge weight is often the length of the edge, which is a road segment or a street.

The *shortest paths* in a network graph is the smallest sum of edge lengths among the possible paths between two selected nodes [Porta et al., 2006]. This metric is calculated using various algorithms which are designed for calculation speed and efficiency, a commonly used one is Dijkstra's algorithm [Levinson et al., 2007] [Volchenkov and Blanchard, 2007]. Watts and Strogatz [1998] define a characteristic path length metric, *L*, which is the average of *shortest paths* in a network. This metric is designed to evaluate the connectivity of a global network, with the assumption that the subject network is connected. To assess whether a network is connected, it is determined whether every pair of nodes can reach each other via paths of one or more edges.

Table 5.1: Studies using the complex network approach in transport and urban research

Author (Year)	Type of Data	Assessment	Metrics
Cats (2017)	Rail network	Topological evolution	Topology indicators: - Diameter - Connectivity (gamma index) - Meshedness (alpha index) - Directness Centrality indicators: - Degree, closeness, betweenness, - Shortest path
Crucitti et al. (2006)	Road network	Centrality	Centrality indicators: -Closeness, betweenness, straightness Space syntax
Hanna (2021)	Road network	Accessibilty	Random walk
Hillier et al. (1993)	Road network	Accessibility	Space syntax: -Connectivity, control value, integration index
Lee & Kim. (2021)	Street network	Accessibility	Accessibility indicators: - Random walk based access diversity - Spatial correlation index
Liu, et al. (2017)	Road, air and rail networks (Intercity)	Accessibility Community structure	Community indicators: - Community structure modularity Centrality indicators: - Degree, closeness, betweenness
Porta, et al. (2012)	Road network	Economic activity correlation with centrality	Centrality indicators: - Multiple centrality index - Kernel density estimation
Strano, et al. (2012)	Road network	Topological evolution	Centrality indicators: - Betweenness centrality Topological indicators: - Cell shape - Edge length
Xie & Levinson (2009)	Road network	Topological evolution	Topological indicators -Density, Gamma, Alpha

The degree of a node is the number of its neighbors. In order to use this metric as an indicator of the network, the *average degree* is used, shown in the equation below [Barthélemy, 2011]:

$$\langle k \rangle = \frac{\sum k_i}{N} = \frac{2E}{N} \tag{5.1}$$

where k_i is the degree of node i

Multiple Centrality Assessment

Multiple centrality assessment (MCA) is a model used in urban spatial network analysis to evaluate street centrality [Porta et al., 2012]. The network representation depicts streets as edges and intersections of streets as nodes. It is often compared to *space syntax*, where MCA stands out by actual geographic distance estimation rather than the topological estimations adopted in space syntax. This model uses three commonly used centrality indices: closeness centrality, betweenness centrality, and straightness centrality.

Centrality metrics define a category of indices which are used to evaluate the importance of a node within a network. The simplest form of centrality is *degree centrality*, which assumes the more connections a node has, the more important it is within a network. However, the use of this metric in urban networks is limited due to the geographical constraints limiting the node degrees [Crucitti et al., 2006].

$$C_i^D = \frac{k_i}{N-1} \tag{5.2}$$

Another centrality metric, *closeness centrality* uses shortest path lengths originating from a node to all the other nodes in the network. When the sum of shortest paths from the selected node is low, this metric assigns a higher centrality value to the node. The equation for this metric is formulated as follows [Crucitti et al., 2006]:

$$C_i^C = \frac{N-1}{\sum_{j=1}^{N} d_{ij}} \tag{5.3}$$

where d_{ij} is the shortest path length between nodes i and j

Betweenness centrality is a metric which considers the number of shortest paths between every node pair in the network. When a node is located on many shortest paths the centrality value is higher. This metric is relevant in urban street networks as it is used to identify most commonly traversed road segments. The equation for this metric is formulated as follows [Crucitti et al., 2006]:

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{\substack{j,k=1 \\ j \neq k \neq i}}^{N} \frac{n_{jk}(i)}{n_{jk}} \tag{5.4}$$

where $n_{jk}(i)$ is the number of shortest paths passing through i

Straightness centrality is used to measure the reachability of a part of the network from other locations on a straight path. The assumption that drives the use of this metric is that the efficiency of connections between nodes is higher when the shortest path is closest to a hypothetical straight line between them [Porta et al., 2012].

$$C_i^S = \frac{1}{N-1} \sum_{\substack{j=1, \, j \neq i}}^{N} \frac{d_{ij}^{eucl}}{d_{ij}} \tag{5.5}$$

where d_{ij}^{eucl} is the Euclidean distance between nodes i and j

5.1.2 Community Indicators

Clustering coefficient is a significant indicator of spatial networks. When urban networks are considered, densely connected nodes form clusters, which suggests that nodes within clusters are better connected to each other than any other node in the network [Liu et al., 2017]. The clustering coefficient is given by the ratio of number of connected edges to the maximum number of edges possible in a graph based on the number of nodes, as formulated in equation (5.6) adapted from Liu et al. [2017].

$$CC_i = \frac{2M_i}{k_i(k_i - 1)} \tag{5.6}$$

where k_i is the number of nodes which are directly connected to node i

M_i is the actual number of edges between k_i nodes

Average clustering coefficient $\langle CC \rangle$ is a global metric that which indicates the completeness of the network. As this metric approaches 0, a greater share of nodes become isolated with fewer connections. When the $\langle CC \rangle$ is 1, the graph is complete with every node pair having direct connections [Wang et al., 2017].

5.1.3 Topological Indicators

The evaluation of the topology of a transport network gives an insight into its structure and layout. The planar graph representation makes it possible to analyze the properties of transport systems in the spatial-geographical context. This approach is interested in the taxonomy of network structures, such as scale-free networks or small-world networks. The former signifies networks in which there are many nodes having few connections, and few nodes having many connections [Cats, 2017]. Real-world geographical networks, such as surface transport networks are found to display scale-free properties [Xie and Levinson, 2009]. Small-world networks, on the other hand, are highly clustered and more than often have short path lengths. Watts and Strogatz [1998] define these networks by having a characteristic path length almost as low as a connected random network with the same number of nodes and edges. Concurrently, the clustering coefficients of such networks are higher than the specified random network [Watts and Strogatz, 1998]. To analyze such network structures and identify their characteristics, topological measures are used based on complex network theory.

Network density is a measure used to evaluate how developed a network is. This measure is calculated by dividing the length of edges in a region by the surface area of the region [Xie and Levinson, 2009], expressed as the following:

$$density = \frac{L}{B} \tag{5.7}$$

where,

L is the length of observed links

B is the area of inspected region

Diameter of a network is used to measure its size. This measure can be defined as the maximum topological distance between any node pairs, as shown in Equation (5.8). The shortest paths between every node pair is calculated using Dijkstra's algorithm, and the maximum shortest path of any node pair is selected [Cats, 2017].

$$diameter = max_{i,j \epsilon N}(d_{ij}) \tag{5.8}$$

where d_{ij} is the hortest path length between i and j

Connectivity indices were introduced to the transport planning literature by Garrison and Marble [1962] and Kansky [1963] are used to determine the characteristics of a planar graph. These indices are valid in a two-dimensional network with no crossing edges Casali and Heinimann [2019]. They are explained in further detail in the following paragraphs.

Alpha index (5.9) is used to measure get an insight of the existence of redundancy in the network. When calculated, this metric indicates the probability of moving through the network and coming back to the same location without passing through an edge more than once.

$$\alpha = \frac{E - N + 1}{2N - 5} \tag{5.9}$$

Beta index is the ratio of number of edges to the number of nodes in the network. This index signifies the complexity of the network in terms of the existence of connections with respect to the number of nodes.

$$\beta = \frac{E}{N} \tag{5.10}$$

Gamma index is a measure of connectivity (5.11), which measures the number of edges in a network and compares this number with the number of possible edges in the network. On a scale of 0 to 1, a higher gamma value indicates better connectivity.

$$\gamma = \frac{E}{3(N - 2)} \tag{5.11}$$

A metric relevant for public transport systems is the *directness*. Cats [2017] define this measure as the disparity of network distance and geographical distance, given in Equation (7.1). The metric compares the network distance with a hypothetical shortest-path, and a higher directness value indicates a greater impedance.

$$q = \frac{\sum\sum I_{ij}}{N(N-1)\sum\sum d_{ij}} \quad i, j \in N, \quad i \neq j \tag{5.12}$$

where

I is the path length of node pair i, j

d_{ij} is the hortest path length between i and j

5.1.4 Accessibility Indicators

Random Walk Access Diversity

Random walk is a topologically induced movement mechanism through networks. The dynamics of random walk have been used in physics and the research of linear dynamics of diffusion Travençolo and da [2008]. It has been used in complex network analysis, and is gaining attention in road transport network studies Lee and Kim

[2021]. The principal mechanism of random walk is the probabilistic diversity of how many nodes can be reached within a specified cost limit. Traditionally, the cost limit is defined through the number of steps in the network topology. However, Lee and Kim [2021] propose an accessibility metric based on the geometric distance, which is calculated by the summation of edge weights (lengths) in the network. Hanna [2020] suggest that human movements can be modeled by random walk to predict the movement of agents and network centrality measures. Their model has few assumptions with regard to movement, which means that the movement is random with no memory, no goals or intentions and direction changes based on the angle of the intersection. Lee and Kim [2021] use random walk to model access diversity in a road network, that is, the number of distinct nodes reached in 1000 different walk iterations. Their approach relies on the sequential draws of road segments (edges) on each intersection (node). In random walk path generation, each edge is given the same probability of selection. Thus, at each step, the probability of the walk passing through an edge is inversely proportional to the degree of the node. The sequential node selection of the random walk approach is shown in Figure 5.1. Next, Lee and Kim [2021] conduct a spatial correlation analysis to identify clusters of nodes based on the metric of access diversity. Consequently, they are able to determine the regions of the network which can access more diverse locations within a high statistical significance. The random walk method is particularly useful for including the network topology effects on accessibility measurement. However, the reviewed literature often lack the land use component and focus on the network structure to measure diversity of reachable locations. Lee and Kim [2021] identify the shortcoming of their approach and propose the inclusion of origin-destination pairs based on real activity data. Therefore, although useful to assess accessibility based on network topology, this method could benefit from the inclusion of land use and activity components as described previously in chapter 4.2.
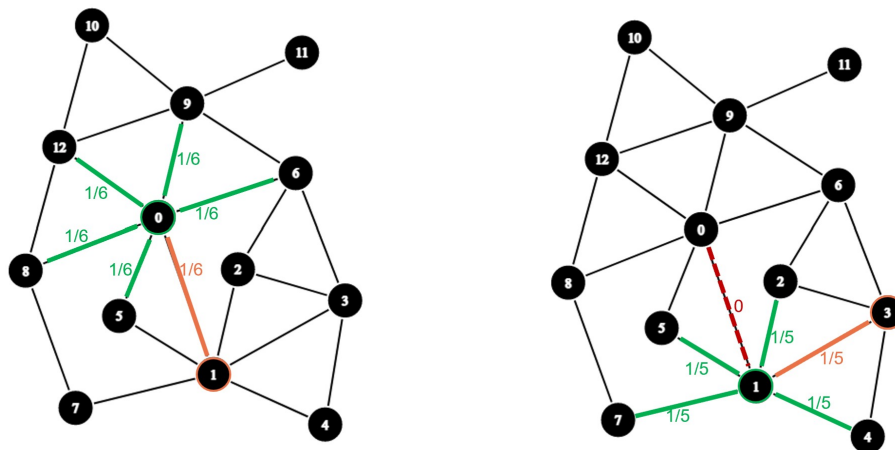
Figure 5.1: Visualization of self-avoiding random walk for a simple two-step simulation traversing nodes 0, 1, 3. The probabilities for route selection are shown on the edges.

Space syntax

Space syntax (SS) is an analysis method developed to interpret the morphology and growth of urban grids [Hillier et al., 1993]. Despite the initial purpose of its development, in recent studies SS has been used to estimate the relationship between street infrastructure and human activity, accessibility, safety, and pollution [Crucitti et al., 2006]. In contrast to common network analysis approaches in the urban context, SS adopts an axial graph representation in which nodes signify streets and edges signify the intersections of axial lines, the visualization of SS and MCA representations are explained further in Appendix B.1. Steadman [2004] maintains that, while it is expected that street configuration influences the movement patterns, a simple urban model would include functions and services such as job locations, housing would give more accurate outcomes of accessibility. Therefore, the main criticism towards SS is the reduction of urban accessibility patterns to the topology of the network. The failure to incorporate land use component makes it an unsuitable for a comprehensive assessment of accessibility based on the factors described in 4.2 and the components outlined by Geurs and van Wee [2004].

5.2 TRANSPORT NETWORK EVOLUTION

Urban systems can be described as complex systems comprising subsystems such as transportation, land use, and population [Ding, 2019]. As far as urban expansion is concerned, these subsystems develop interdependently. To investigate the dynamic relationship between these subsystems, a range of research and modeling approaches have been developed, often referred to as "co-evolution" models. Furthermore, there are studies which focus on the topological evolution of networks, which refer to the graph theory more extensively. Both of these approaches are explained in the following paragraphs.

5.2.1 Topological Evolution

The topic of how graphs evolve has been researched under various fields of technological, sociological and scientific research [Leskovec et al., 2005]. These studies often consider certain properties of graphs, such as the node degrees and distances between nodes (diameter). Real-world networks' evolution can be simply explained through the emergence of new nodes or disappearance of existing ones. The work of Leskovec et al. [2005] emphasize two empirical observations about real-world networks. Firstly, networks become denser over time, leading to a higher average node degree. Secondly, the diameter of real-world networks often decrease as the network grows. These observations are analyzed through real-world networks such as patents citations or co-authorship affiliations. Strano et al. [2012] study road networks' evolution through two fundamental processes identified through centrality metrics. *Exploration*, refers to the spatial growth of the network caused by additions of roads, whereas *densification* is the increase in the number of roads within the urban center. The former is more prevalent in the early stages of evolution, whereas a more

established network often experiences the latter [Strano et al., 2012]. Building on this theory, Cats [2017] analyze the topological evolution of a rail network, focusing on network structure, and centrality indicators. Such studies distinguish different periods of network growth and seek to make a more graphical analysis. There also exist studies which consider interdependent evolution in complex networks, described in the following paragraph.

5.2.2 Co-Evolution Models

Considering the cooperative evolution of urban networks, Levinson et al. [2007] describe a dynamic process involving land use and transport systems. Initially, travel demand is driven by changes in land use, which in turn influence traffic flows. The traffic flows stimulate infrastructural investments, which influence the accessibility patterns within cities. Ultimately, land use within the city are updated to adapt to these patterns. This perpetual process describes the evolution of urban areas together with land use dynamics and transport networks. Levinson et al. [2007], using a bottom-up approach to investigate the interrelations of decisions made by businesses, travelers, and transportation agents. The application of a co-evolution model in a case study of London rail systems in the 19th and 20th century showed that a correlation exists between population and network density. Furthermore, their findings showed that rail systems transformed London's city center. This transformation resulted in a decrease in residential density and an increase in the commercial density [Levinson, 2008]. Another model by Barthélemy and Flammini [2009] investigated the co-evolution with respect to rent prices and accessibility demand. They used the transport costs as an independent variable to simulate the spatial evolution of population density in zones with higher accessibility to economic centers. Lastly, a relevant study by Ding et al. [2021] developed a complex-network based framework to analyze the relationships in multi-layer urban systems. Such systems consider multiple transport network topologies (i.e., traffic and rail), integrated with land use and population growth. Their approach is based on simulated city data rather than empirical observations, thus lacking the real life validation.

5.3 CHAPTER SUMMARY AND DISCUSSION

This chapter investigated the complex network theory applications in transport research. First, the basic concepts of this theory were described based on graph properties. Next, complex network metrics used in literature were categorized based on their use cases, namely, centrality calculation, community detection, topological assessment, and finally accessibility evaluation. Lastly, the network evolution studies were investigated to compile the common temporal network analysis approaches. Complex network metrics offer a useful means of analysis on the structure and the development of a transport network. Representation of the network is commonly done through two distinct approaches. Using a direct transformation of roads to edges and intersections to nodes, is commonly used in topological assessment and centrality assessment.

Another approach, space syntax, adopts an axial approach with intersections as edges, and road segments as nodes. In the context of mobility, this approach assumes visibility and following a straight path is the main driver of movement decisions. However, this approach lacks the land use component and focuses on the network topology, or infrastructure. Thus, an accessibility measure using space syntax can be described as infrastructure-based measure described by Geurs and van Wee [2004], as it is lacking the land-use component. This approach would not fully encapsulate the accessibility factors described in Figure 4.2.

The other accessibility indicator discussed in this chapter, random walk access diversity, is a potentially useful approach to accessibility measurement. Although not yet implemented in the literature, the land use component neglected by other graph theoretical approaches can be included in this approach. Thus, it will be possible to analyze the impact of transport system in terms of network topology, and land use in terms of opportunity locations (as mentioned in chapter 4) using this method.

Topological indicators are potentially useful for an exploratory analysis of the network, and make comparisons across historical timesteps. For an accessibility measurement approach based on network topology, these indicators are useful to delineate the impact of transport infrastructure on the output. Specifically, measures related to connectivity such as *gamma* index are useful to understand the network properties in a region of interest.

Based on this literature review, complex network theory is assessed to be a modeling method which could add value to the accessibility measurement. The complex network indicators offer an extensive investigation means for the infrastructure-based measurements. Furthermore, the inclusion of land use and demographic indicators in a network model enable a more comprehensive approach to the accessibility measurement. The most important benefit of this approach is being applicable to different urban settings, given that sufficient data quality is provided. The most important added value of this structured method becomes apparent in the use case of temporal analysis. The complex network model enables a fairer measurement of location-based accessibility by accounting for the changes in the transport network topology. Thus, when historical timesteps are compared, a more accurate assessment of accessibility can be achieved by considering the changes in the network. Furthermore, this approach offers versatility as different accessibility use cases can be analyzed by including their respective location in the model. For example, the approach can be extended to measure accessibility to job locations. Consequently, this study will benefit from the repeatability of the analysis in temporal analysis by using a measurement methodology for different timesteps.

6 | ANALYSIS METHODOLOGY

In this chapter, the methodology for the analysis part of the study is explained. First, the data collection methods and pre-processing is discussed. Next, an initial overview of the data with descriptive network metrics and properties of socioeconomic indicators are given. Next the accessibility measurement methodology is described, as in the chosen complex network method, application to school accessibility cases, spatial analysis methods, and finally the equity estimation methodology.

6.1 ANALYSIS PROCESS

This section describes the chosen research approach to answer the final subquestion (SQ4), and ultimately the main research question of this study. Based on the literature review findings of three main concepts of equity, accessibility, and complex network theory, an analysis methodology is designed to be tested on a case study.

This study seeks to incorporate complex network theory into accessibility measurement and use spatial analysis together with equity analysis to make a historical comparison of accessibility in the selected case study. The focal activity of the accessibility will be education. Therefore, accessibility to school locations will be analyzed. As discussed in the section 4.2, education access is an important indicator of economic participation, and compared to job accessibility, less attention is given to this field in literature. Based on the previously identified factors, an approach which encapsulates common accessibility factors is developed. Complex network approaches are at the core of this accessibility measurement approach, as the random walk based accessibility measurement method is adopted. Thus, the research question of this study will be addressed at the end of the application of this methodology. An assessment will be made regarding the usability of this approach in the selected case and other use cases. The following paragraphs explain the flow of the analysis methodology.

The case study begins with collection of the necessary transport network data as well as the school locations and socioeconomic indicators. The case specific data collection, pre-processing, and complex network set up processes are discussed in chapter 7. Following these processes, an exploratory analysis of the transport network and the spatial distribution of school and socioeconomic data is carried out. This step serves to gain a better understanding of the case specific characteristics and their projections to the complex network model. Specifically, transport network topology attributes discussed in chapter 5 are measured.

The following part of the methodology is divided into three sections: accessibility measurement, spatial analysis, and equity analysis. Accessibility measurement is based on the selected complex network model of the road network. The measurement is carried using a modified *Random Walk-Based Access Diversity* method. As

previously discussed in section 5.1.4, this complex network method measures the probabilistic diffusion based on a network topology. However, based on the literature review on accessibility factors, this measure lacks the land use context, as it is closer to an infrastructure-based measure than a location-based measure as defined by Geurs and van Wee [2004]. In consideration to the selected opportunities in the urban context, the education locations are included in this method. Thus, access diversity calculated by the random walk is adapted to the transport land-use context by incorporating the locations of schools. Following the accessibility measurement, descriptive statistics of the measurement results are presented, where the mean, standard deviation, and median of the accessibility metric is provided.

After the accessibility measurement is completed, a spatial analysis is conducted using Hotspot Analysis for the identification of high accessibility clusters. Hotspot analysis is a widely used method for the visualization of statistically significant clusters of high and low values of an attribute, which in this case is accessibility. Using Hotspot outputs enable the qualitative comparisons across historical timesteps.

Lastly, a series of equity assessments is carried out. These assessments are based on the outputs of the literature review on transport equity and the most appropriate indices based on their application areas. As described in 6.5, an equity measure needs to carefully select the type of disaggregation of the population. For this research approach, the population is distinguished by their location, for this purpose, population is transformed into the network nodes. Gini, Spatial Gini and Theil indices are used for the assessment of equity for each timestep. These indices all offer a different insight into equity, such as the overall equity, spatial correlation with neighbors, and within regions, respectively.

The last component of this methodology is the temporal analysis of outputs and collected data. This methodology is followed for several timesteps based on the collected historical data. For every step of the analysis, the results are compared across timesteps to examine the evolution of accessibility metrics and the related equity indices.

The full scope analysis methodology flow is visualized in Figure 6.1.

6.2 EXPLORATORY ANALYSIS – NETWORK TOPOLOGY

In this part of the methodology, an initial network analysis is conducted. using the topological indicators described in Chapter 5. Among these indicators, the node degree and the gamma index are utilized for this study. The node degree signifies the number of neighbors each node has. Thus, it gives an indication of how many potential random walk paths can be generated.

The gamma index uses a similar approach where the ratio of number of connections to the number of maximum connections are investigated. Therefore, this indicator reveals how developed one part of a network is in terms of connectivity. Similar topology indicators such as the alpha and beta are not used in this study. Alpha index considers the number of cycles in the network paths. These cycles are already disregarded using the self-avoidance parameter of the random walk, explained in the following paragraphs of this chapter. The Beta index describes the complexity of
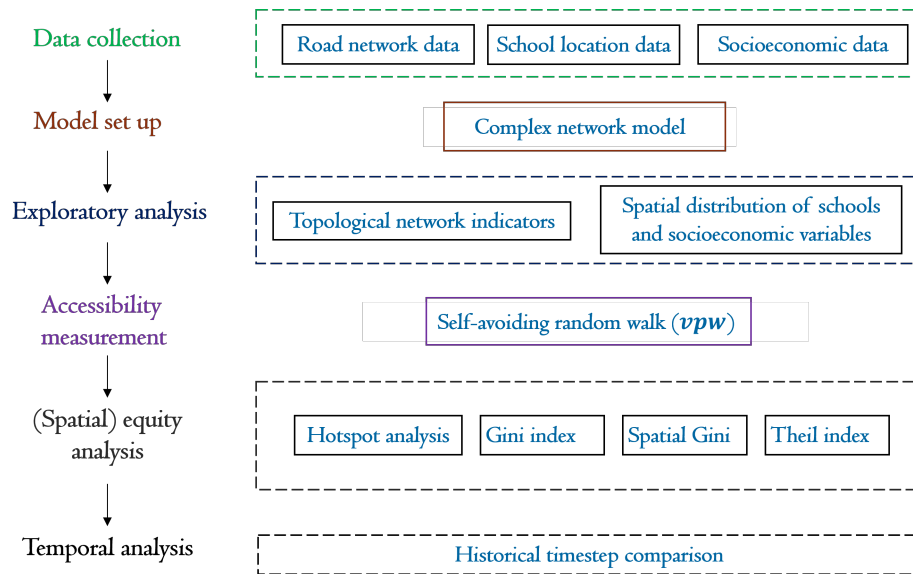
Figure 6.1: Process flow of the developed methodology to be applied in the case study

the graph based on the number of edges and nodes, however a more in-depth insight is already generated by the gamma index with regard to the maximum connected case. Therefore, this index is not included.

6.3 ACCESSIBILITY MEASUREMENT

This section discusses the methodology for accessibility measurement methodology using complex network metrics, specifically the *Random Walk* method. The random walk algorithm is explained and the method of application to two distinct cases are described: school and job accessibility. The explained methodology is applied in Chapter 7 where model parameters and results are presented.

6.3.1 Random Walk–Based Access Diversity

The network analysis methods selected for accessibility evaluation in the Helsinki case study is the random walk approach. This approach simulates a diffusion process through the road network multiple times to assess how many diverse nodes can be reached with a given threshold. The algorithm of this complex network method is presented in pseudo-code form in Algorithm 6.1.

Throughout a single walk, the path selection is completely randomized. This suggests the probability assigned to any road segment is equal to all other available road segments when the random walker makes a decision. Self-avoidance property is imposed on the random walker, meaning that in a single walk, an already visited node cannot be visited again. This property eliminates the possibility of the random walker being stuck in a certain area, or even going back and forth between two same nodes. At the end of every walk, the list of visited nodes are reset, therefore it is possible to visit the same nodes in different walks. The random walk finalizes when a

---

Algorithm 6.1: SELF-AVOIDING RANDOM WALK

Input: weighted network $wG(N, E)$, distance threshold D
Output: list of visited nodes V

---

1 for $s \in N$ **do**
2   *initialize V;*
3   *initialize total distance of walk W;*
4   **while** $W \leq D$ **do**
5   *add s to V;*
6   *get list of neighbor nodes (L);*
7   *remove visited nodes V from L;*
8   *choose random neighbor r from L;*
9   *collect distance w between r and s add to W;*
10   *r is the new s;*

---

predefined distance threshold is reached. The distance traveled is calculated by summing the weights of each edge traveled during the walk. However, the termination of the walk is not only subject to this threshold. Due to the self-avoidance property, when the walker reaches a node with a degree of 1 (meaning only one neighbor, which was already visited), the walk ends. Similarly, if the walker reaches a node whose all neighbors were already visited, no available options remain, hence the end of the walk. To sum up, three conditions could potentially stop a self-avoiding random walk:

- The distance threshold D is reached

- A node with a degree of 1, in other terms, an extremity node is reached

- All the neighbors of the last visited node have been previously visited

In layman's terms, this algorithm can be explained as dropping a number of people (i.e. number of walks) at the same intersection of a city and recording their individual movements in the city with a certain distance range. The road segments (edges) and intersections (nodes) which they visit are tracked, not allowing them to traverse through the same locations again. By the end, the visited locations are collected to analyze the diversity of nodes they reach, as well as count any target locations they pass through. This model provides a very simplistic movement simulation within the network. No prior information is provided regarding the network, and the results give an indication of how many diverse locations can be reached with this movement pattern. For a real life case, it is not very realistic that an individual would take random trips through the network. However, this model gives an insight into the topological characteristics around a location. Thus, connectivity with the other locations in the network can be observed by simulation. An analyst using this model has to make several parameter decisions based on the use case it is being implemented in. In order to make the correct decisions, a background research on the modeled case needs to be made. Also important, is the impact of the change in

parameters in the model, independent of the use case. The next paragraph explains the impact mechanism of these parameters.

In this algorithm, the diversity of visited nodes are directly affected by the selected distance threshold. The distance threshold defines how much distance a walker can cover, giving them an approximate movement area. Based on this explanation, it might be expected that increasing the threshold would make it likely that the random walker reaches more destinations. Yet, this assumption would only hold in the absence of the self-avoidance property. As the path is restricted by the already visited locations, in certain cases the random path selection is actually limited to only a few choices, if any. Based on the network characteristics, or the orientation of edges in a certain part of the network, the walker might be limited to a shorter range of motion than expected. Therefore, it is often the case that as the distance threshold increases, the diversity of newly reached nodes initially increases, then decreases, showing a parabolic trend. Based on this evaluation, this method is potentially useful to observe clusters within the network topology.

The other parameter to consider in this approach is the number of walks. When a simulation with few number of walks is considered, it is likely that the model output will not be as diverse as expected. Thus, increasing the number of walks is useful to reach the expected diverse distribution of accessibility. Nevertheless, the number of walks can only be increased to a certain value before the output diversity stops increasing significantly. This behavior can be identified as the diminishing returns phenomenon. After all, the number of distinct locations reached through N number of random walks can only be as high as the network topology and the specified distance threshold allows. Moreover, increasing the number of walks generously could also impact the computational cost of the model. As the random walk model is essentially a sequence of probabilistic draws made N times, the value of N proportionally impacts the computation time. Thus, a trade-off is observed in the number of walks and the time spent simulating the model. It is the modelers' duty to find the optimum point where an accurate result is obtained in an appropriate processing time.

6.3.2 School Accessibility Using Random Walk

For the case of accessibility to schools, a methodology is proposed using the self-avoiding random walk algorithm. The measurement method relies on the node-specific analysis of visited nodes and counts the number of times a school node is visited. The number of school visits are summed across walks and divided by the number of walks to obtain the metric called *visit per walk (vpw)*. This metric describes the number of schools within the reach of a starting node.

$$vpw = \frac{\sum_{w=1}^{N_w} |\{S \cap V_w\}|}{N_w} \tag{6.1}$$

where S is the set of school nodes

V_w is the set of visited nodes in walk w

N_w is the number of walks in the simulation

The threshold of this reach is dependent on the distance threshold set for the use case. In order to define the travel range of the random walk, a travel mode of interest is selected, which determines the travel speed. When the speed is multiplied with an average travel time metric, a distance threshold is obtained. For every application case of this method, the dominant travel mode and the average travel time need to be considered when travel threshold is determined. According to the literature review (section 4.2) on school accessibility factors, it was observed that different dominant travel modes exist for specific travel purposes. Some regions offer higher accessibility by public transport than personal cars, whereas some urban regions excel in their walkability to opportunities. Furthermore, affordability component has a significant influence, as in the context of trips to schools disadvantaged groups are more likely to choose affordable travel modes such as walking or public transport.

Accessibility measurement outputs are presented in two methods. First, a descriptive statistics overview is provided, where statistics such as the range, mean, median, and standard deviation of *vpw* is provided. These statistics provide an overview on the development of the measured accessibility in the whole network. With the mean value, the average accessibility in the network can be tracked, standard deviation indicates the extent measured accessibility values are spread out. The median gives an indication of which end of the range most values are accumulated. For the school *vpw* measurement, another significant indicator is added to the descriptive statistics, which is the percentage of zero school visit nodes. This value indicates the ratio of nodes which never visit a school location through any of the walks to the total number of nodes in the network.

6.4 SPATIAL ANALYSIS

To analyze the spatial relationship of measured accessibility per network node this study uses a specified spatial analysis and visualization method. Initially, two separate methods are considered which are the *Kernel Density Estimation (KDE)* and the *Hotspot Analysis* methods. KDE assigns a value to each spatial unit based on the cumulative proximity of opportunities to it. The density component of KDE is connected to a specified window, wherein objects are inversely weighed by their Euclidean distance [Porta et al., 2012]. The other method, Hotspot Analysis, is commonly used to identify spatial clusters which are statistically significant. With this method, statistically significant clusters with high values are identified as hot spots, whereas clusters with low values are identified as cold spots.

KDE is a useful method to identify hotspots in a spatially distributed dataset. The estimations made in KDE are based on the pattern point densities. High and low points are detected within a specified search radius (bandwidth). However, the search radius is a user specified parameter which makes the KDE output subjective in terms of the parameter choices made. When the search radius is set to be very large, information of finer scale might be lost. On the other hand, a very small search radius, data on a smaller scale has more impact on the results, yielding a more granular output which could make it difficult to distinguish actual clusters and patterns. Furthermore, KDE is useful for identification but does not give any insight into the

statistical significance of clusters, that is, whether clusters are formed randomly or based on an actual spatial pattern [Kalinic and Krisp, 2018]. Hotspot Analysis, which is selected for this study, addresses these issues as a more methodical approach to the statistical significance.

6.4.1 Hotspot Analysis

Hotspot analysis relies on the statistical assessment of statistical significance based on p-values and z-scores. Getis-Ord Gi* statistic is used to calculated these values. The calculation of Gi* is based on the neighboring attributes. If a specific node has a high attribute, and is surrounded by neighbors with high attribute values, it is potentially a hotspot node. In order to calculate the z-score statistics, attribute values of a location and its neighbors are summed. This local sum is then compared with the sum of all values in the area of interest. If the difference is significant large, it potentially means that the chances of such a high difference occurring at random is very low. This probability is approximated using the z-score based p-value, which is the significance [Kalinic and Krisp, 2018].

The utilized Hotspot Analysis method relies on a software tool developed by ESRI. The *Optimized Hotspot Analysis* tool selects the most suitable analysis parameters based on a set of conditions. The parameter decisions made are related to how spatial relationships are defined, which often a fixed distance threshold. This value defines the search radius around the location of interest, and must contain at least one neighbor. Another option for relationship definition is the inverse distance method, where all locations are considered, and the ones with the lowest distance weights are selected as neighbors. Optimized Hotspot Analysis tool uses the fixed distance threshold by default, and optimizes this value based on either Incremental Spatial Autocorrelation, or average distance calculation based on spatial distribution of data. The methodological principles of these optimization methods are beyond the scope of this study.

The outputs of hotspot analysis are compared for each historical timestep, identifying clusters in the case area and their development through time. For the school accessibility analysis two major factors are relevant for cluster analysis, which are network topology indicators and the school locations. Based on the comparisons with these indicators, the objective of the analysis is to distinguish the impact of transport network and the land use.

6.5 EQUITY ANALYSIS

This section describes the equity analysis methodology, with the selected indices to be applied in chapter 7. These indices are Gini index and Theil index. The aim of equity analysis is to determine whether an equitable distribution of accessibility is observed, using the output from the self-avoiding random walk in job and education use cases. The equity measurement is carried out across the whole Helsinki urban area in terms of population and income with the Gini index. Theil index enables equity assessment in a regional context, comparing the accessibility across subdistricts.

6.5.1 Gini Index

Gini index will be used to assess the distribution of accessibility across the population. The calculation method, as described in section 3.1, requires the accessibility measure to be displayed on the Y-axis as a cumulative proportion to plot the Lorenz curve. On the X-axis lies the cumulative proportion of population. Accessibility measurement is available through the preceding part of the analysis methodology. Population distribution, on the other hand, is obtained from the transformation of socioeconomic data to the network nodes. This process is explained in 7.2.2. Ultimately, for every network node an accessibility and population value is assigned. The accessibility value is repeated as many times as the population assigned to a node (see section 7.2.2), which results in a frequency based accessibility field. This field is then plotted on the two-dimensional coordinate system to obtain the Lorenz curve as previously discussed. Numerical integration of the Lorenz curve is subtracted from the line of equal distribution to obtain the Gini index.

6.5.2 Spatial Gini Index

As discussed in section 3.2, Gini index can be decomposed into two terms to understand the contribution of near locations and far locations to the overall output. For the measurement of this index, the accessibility metric per node is used as well as a adjacency matrix to define neighbor relationships between nodes. For near difference calculation, only adjacent nodes to the subject node is considered, the value of which corresponds to the node degree. If the near difference value is low, this indicates a strong spatial correlation, therefore most of the Gini value comes from the inequity of distant locations.

6.5.3 Theil Index

Theil index is used in the analysis methodology as an equity indicator on a regional aggregation. As described in section 3.3, this method is useful to evaluate equity within and across specific zones. As was the case with Gini index, Theil index requires population data to compare zone to the greater region. This analysis is therefore carried out in subdistrict level, as these regions are already used in the breakdown of socioeconomic indicators and demographic data. Theil index is therefore relevant to the specific policy assessments by comparing the equity across spatial units defined by decision-makers. The contribution of within region equity and between region equity to the overall measure can be indicated based on the decomposition of this index as formulated in section 3.3.

7 | APPLICATION

7.1 CASE INTRODUCTION

This section provides the background for the selected case area in terms of its historical development, urban form, and changing transport systems. Additionally, the modal split of average daily trips in the case region are provided with inferences for the analysis.

Background

The selected case study area for the this study is the City of Helsinki, Finland. The City of Helsinki is located in the Greater Helsinki metropolitan area. It is the largest city and the capital of Finland with a population of 658 thousand people in 2021. Starting from 1970s the city has gone through a rapid urbanization involving depopulation of rural regions and construction of suburban regions [Nevanlinna, 2016]. Söderström et al. [2015] assess Helsinki's city center to have a dominant role in number of jobs, services, and housing. The scope area of this study is the urban core of Helsinki, which consists of areas within the approximate range of 10 km from the historical city center. The Helsinki city area is divided into 142 subdistricts. The historic center of the city, as well as the south harbor are located in the southwest of the urban core, as shown in Figure 7.1.
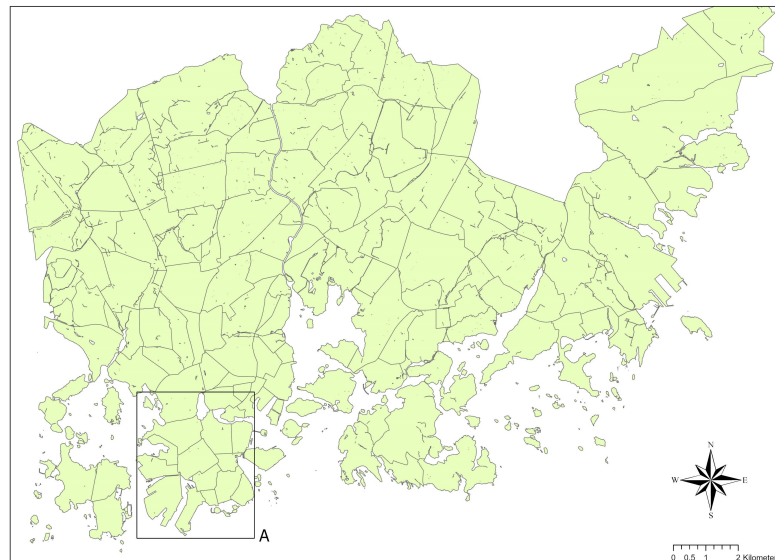

Figure 7.1: Helsinki's subdistrict divison map, box A showing the historic center

The city underwent several transformations in the 20th century in terms of urban planning and transport systems. The city expanded its boundaries beyond the historic center after 1950s, annexing neighboring municipalities [Nevanlinna, 2016]. Furthermore, the south harbor was closed down in 1970, moving the freight port to the east of the city. As a result of such transformations, the urban core started to transform to a polycentric form [Söderström et al., 2015]. Consequently, transport systems and travel behavior also changed significantly throughout this transformation. The construction of suburbs around the urban core increased the travel distances, which lead to an increase in car dependency starting from 1960s [METREX, 2020]. Concurrently, the road network investments were made including highways connecting peripheral towns. Car ownership has steadily increased since 1980s in Finland, however the Helsinki region is consistently below the national average [Liljamo et al., 2021]. City planning has played a role in this, as many pedestrian zones in and around the city center were built starting from 1989 [City of Helsinki, 2020]. In a relevant study about travel-related urban zones in Helsinki, Mäki-Opas et al. [2016] identify zones based on urban characteristics. Their criteria for the identification of zone sis the distance to business locations, public transport, and location of subcenters. Their zoning classification of Helsinki is shown in Figure 7.2. Based on this study, historical center of Helsinki as well as a few subcenters are pedestrian zones as they are at most 2 km away from central business districts. The surrounding areas in the urban core of the city are identified to be public transport regions, and the outskirts of the city are often car oriented with occasional pedestrian and public transport zones in subcenters. It is noteworthy that this assessment is mostly infrastructure-based, and requires validation based on the individual travel behavior of transport users. Therefore, the next part of this section provides a background on the share of travel modes used in the case area.
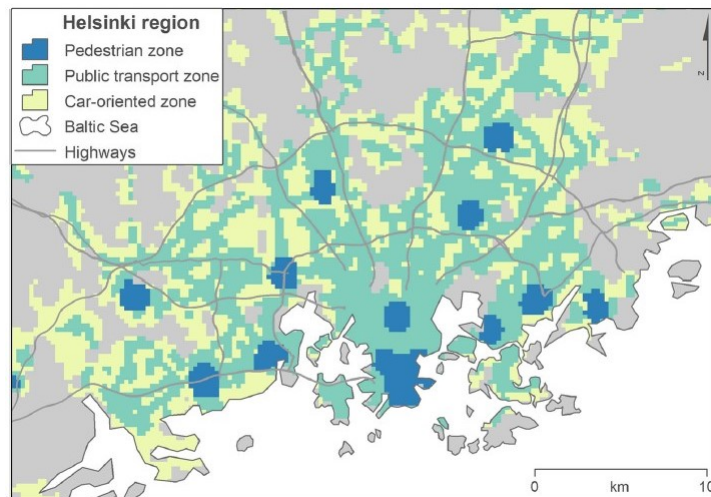

Figure 7.2: Travel-based zones in Helsinki region [Mäki-Opas et al., 2016]

Travel Modal Split

On a nationwide travel survey conducted in 2016, modal split of daily trips were determined. The trips in Helsinki region was observed to be predominantly car and walking, despite having the highest public transport mode share in Finland, as shown in Figure 7.3.
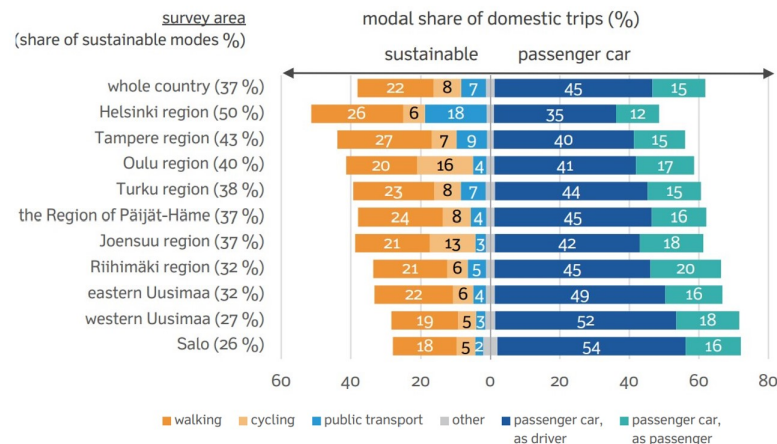
Figure 7.3: Modal split in regions of Finland Finnish Transport Agency [2018]

To understand the situation in City of Helsinki, another study focusing on the Helsinki region was consulted. This regional travel survey conducted on the same year zoomed into the modal split, based on the travel distance and the sub-areas of the city [Helsingin Kaupunki, 2016]. The results show that (see Figure 7.4) walking is the dominant mode of travel in the city center, with 38% share of travel. When the suburbs are concerned, this value drops to 26% with a car dominance of 42% (both driver and passenger use), which validates the analysis of Mäki-Opas et al. [2016] in Figure 7.2. When travel distances are concerned (Figure 7.4), the modal split shows a changing trend. Very short travel distances (less than 1 km) are heavily walking dominated with 82% of shares. When distances greater than 2 km are concerned, highest modal share shifts initially to public transport and eventually to personal cars for distances greater than 7 km.

When the travel purpose of daily trips within the region are concerned, there is a considerable difference in the modal split, as well as the average travel times. Average annual trip share of a Helsinki resident is split as shown in Figure 7.5. For work related travel, the highest share of transport is of public transport, followed by the personal car. For school trips, on the other hand, there is a shared dominance of walking and public transport.

Lastly, the travel time of daily trips in the Helsinki region are presented in Table 7.1. According to the Helsinki travel survey, the average travel time is 26 minutes according to 2016 data. When a differentiation with respect to travel purposed are made, the results show that the travel time associated to travel-to-school trips are 20 minutes, whereas for work related trips this number is as high as 29 minutes. This difference in travel purpose is also reflected in the average travel distance, where
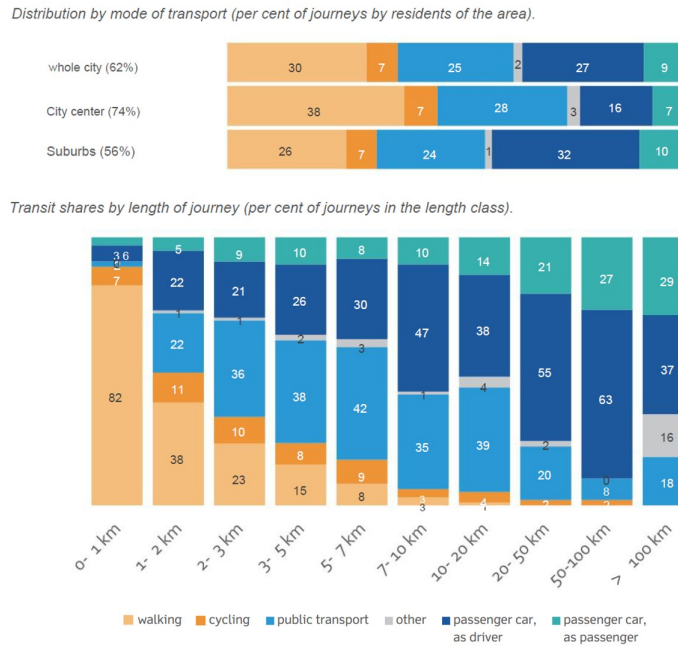
Figure 7.4: Modal split in Helsinki region with respect to on sub-area and travel distance [Helsingin Kaupunki, 2016]
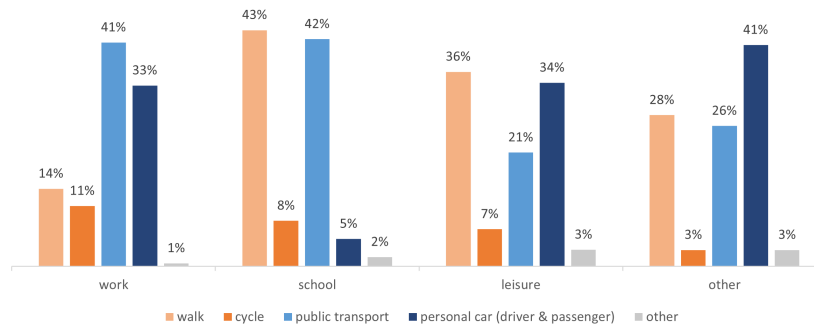
Figure 7.5: Modal split in Helsinki region with respect to travel purpose [Helsingin Kaupunki, 2016]

work related travel amounts to 10 km per trip, and for school trips this value is 5 km.

Table 7.1: Travel times and distances based on travel purpose in the Helsinki region [Helsingin Kaupunki, 2016]

	average distance (km/trip)	travel time (min/trip)
work	11	29
study	5	20
leisure	17	26
all	11	26

Summarizing the findings from national and regional travel surveys, the most common travel mode in the entire City of Helsinki is car travel with walking being the second most common mode. However, when the city center is concerned walking is clearly the most preferred. Furthermore, trips with the purpose of school are highly walking dominant. This can also be inferred by the average travel distance of study related trips being approximately 5 km. The modal split based on travel distances up to 5 km are highly walking and public transport dominant based on Figure 7.4. Therefore, it can be inferred that for a school accessibility case, walking needs to be the focal travel mode for analysis.

7.2 DATA COLLECTION AND PRE-PROCESSING

This section explains the steps carried out to collect and process the necessary data for model implementation. The road network, socioeconomic data, and school location data sources are described with the steps to bring the raw data to a format that is usable in the proposed methodology.

7.2.1 Road Network Data

Helsinki Region Infoshare (HRI) platform contains open data regarding the urban transport systems and built environment [City of Helsinki, 2021]. Some of these datasets offer historical information of the urban environment, unfortunately, not in consistent file formats. Road infrastructure data for 2016 was collected from this source in the form of shapefiles, with the name *liikennevaylat*. For the other timesteps used in this study (1991, 1999, and 2007), guide maps from the Helsinki region were collected. The shapefile was built using ArcGIS, using a process which used the shapefile from 2016 and laid it onto the orthophoto of Helsinki from 2016. Also utilizing Google Maps for visual inspection, the shape file of roads were built. A sample visualization of Helsinki road network in 2016 can be found in Figure 7.6, as well as the other timesteps in higher resolution in Appendix C.1.

7.2.2 District Socioeconomic Data

The demographic data provided by PxWeb database of Helsinki was utilized to collect the socioeconomic data [Til, 2022]. Included in this data is the income, population, housing units, labor force and job count on a district level aggregation. These data were collected in *csv* format, which were then transformed into shapefiles. For this, sub-district shapefiles of the Helsinki Map Service were used. After a data manipulation work in Python to sort the data, ArcGIS was used to join district names and socioeconomic variables. As a result, for each time period, a shapefile was obtained which contains sub-district based socioeconomic indicators.
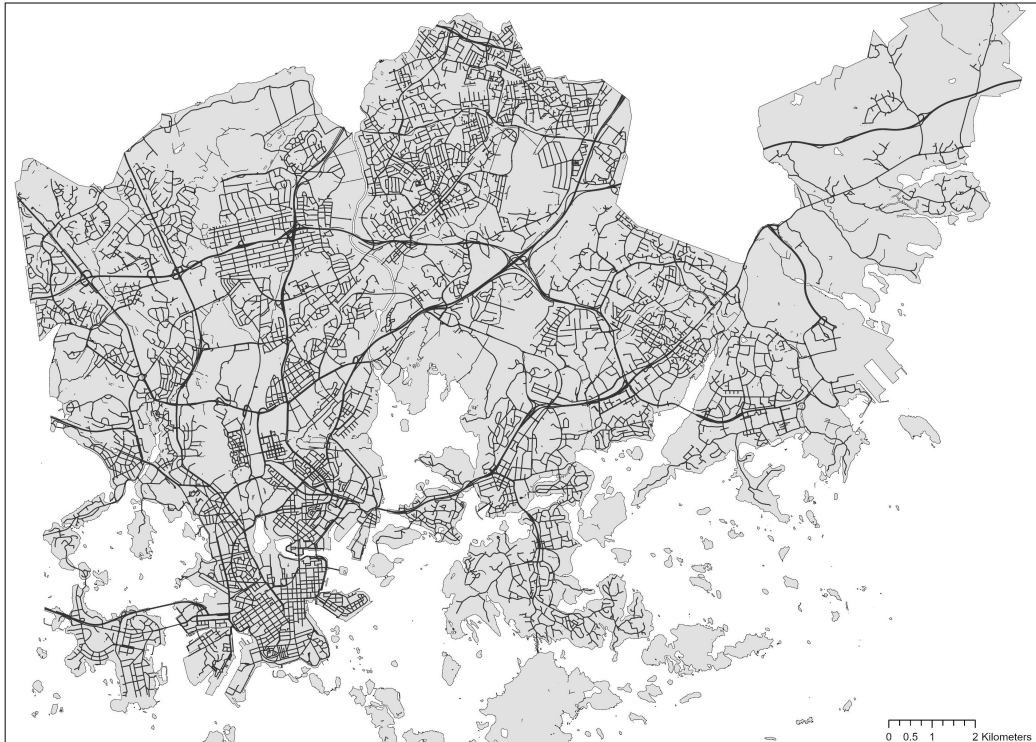
Figure 7.6: Helsinki Road Network in 2016

Voronoi Tessellation

In order to transform district properties to network nodes, the Voronoi tessellation method is used. Specifically, transformation using Voronoi tessellation was applied to the population dataset. The population per node estimates were used in the equity calculations described in 6.5. This method of data processing creates a direct relationship with the size of the land area around a node. Therefore, it creates a spatial connection with the network representation and the geographic properties of the region.

The tessellation consists of Voronoi cells which fill the chosen district based on the spatial distribution of nodes [Barthélemy, 2011]. Each cell is a polygon where the edges are equidistant lines to node pairs of the network. Therefore the node at the middle of every cell is the closest network node to every spatial unit inside the cell. A district is divided into subareas which define the influence area of nodes within the district as shown in Figure 7.7.

Consequently, these areas are used to distribute the properties of a district to each node proportional to their Voronoi cell area. The equation for the calculation of an example property of population P of node i in district d is given as follows [Casali and Heinimann, 2019]:
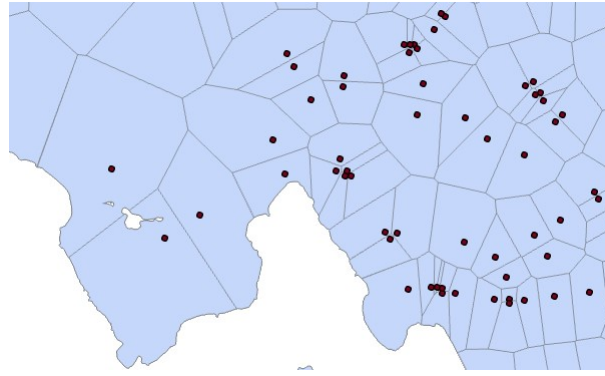
Figure 7.7: Sample Voronoi cell visualization

$$P_i = \frac{P_d \times A_i}{A_d} \tag{7.1}$$

where

A_i is the area of Voronoi cell i

A_d is the area of the whole district

7.2.3   Historical School Data

Helsinki school register, *Koulurekisteri* was used to collect historical school location data [Koulurekisteri, 2020]. This database contains information about all levels of schools and buildings from the year 1550 onwards. The data is supplied through a REST API, where data is collectible in *json* format. This data was structured using Python pandas library. The columns of the structured data are the name, address, start and end years of school buildings. Based on the start and end year, the data was sorted and divided into the timesteps (1991, 1999, 2007, 2016). For example, if a school has a start year of 1985 and end date of 2005, it can be found on both 1991 and 1999 datasets. Using the provided address, the geographical coordinates were obtained using the geopy library. Latitude and longitude values were then projected to the Helsinki land map as points using the ArcGIS environment, and stored as shapefiles. The map of school locations through the selected historical timesteps is provided in 7.8. An exploratory analysis of the school locations' development through time is made in the subsequent section (7.3.2).

Projection to Road Network

The school locations stored in shapefiles were often not coincident on the road network. In other words, schools were not represented as streets or intersections in the model, but as standalone points. The reason was that the overall land area of the school was represented by a point location in the center of the area. For these cases, a processing tool was used to select the closest nodes in the road network. A number of methods for translating school locations to the network were considered. The initial approach to be used was the *Buffer* method. This method creates a circle of
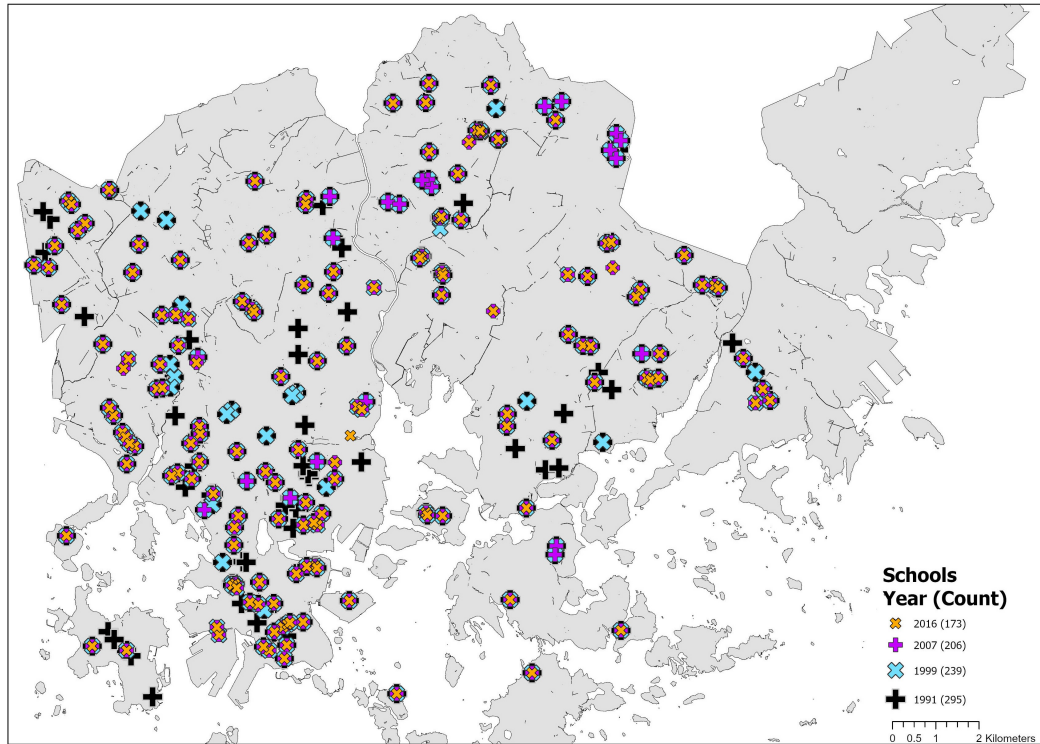
Figure 7.8: School locations in the Helsinki region through 1991-1999-2007-2016

a specified radius around each input node, ensuring all buffer areas would include at least one neighbor node for each school. However, this would mean some school nodes' buffer areas would include a high number of neighbor nodes, 5 to 6 at times, as shown in Figure 7.9. Therefore, another approach was used to overcome this issue. Using the *Near* tool of the ArcGIS environment, the closest edge (road segment) to each school node was selected. These edges were then translated to the two nodes at the end of each road segment, using the NetworkX library. Thus, each school node were represented by two nodes of the road network. This enables better integration to the network due to every school having the same number of nodes representing them on the network.

7.3 EXPLORATORY ANALYSIS

This section provides an exploratory analysis of the collected data within the case study. First, a complex network analysis of the road network topology is conducted, and the results are compared across the historical timesteps. The basic topological indices described in the literature review (section 5.1.3) are used. Then a quantitative analysis based on the average values, as well as the qualitative comparison using the spatially mapped insights are performed. The next subsection explores the schools in terms of their spatial distribution in Helsinki and its subdistricts. Additionally, a comparison with the demographic data in terms of population is made.
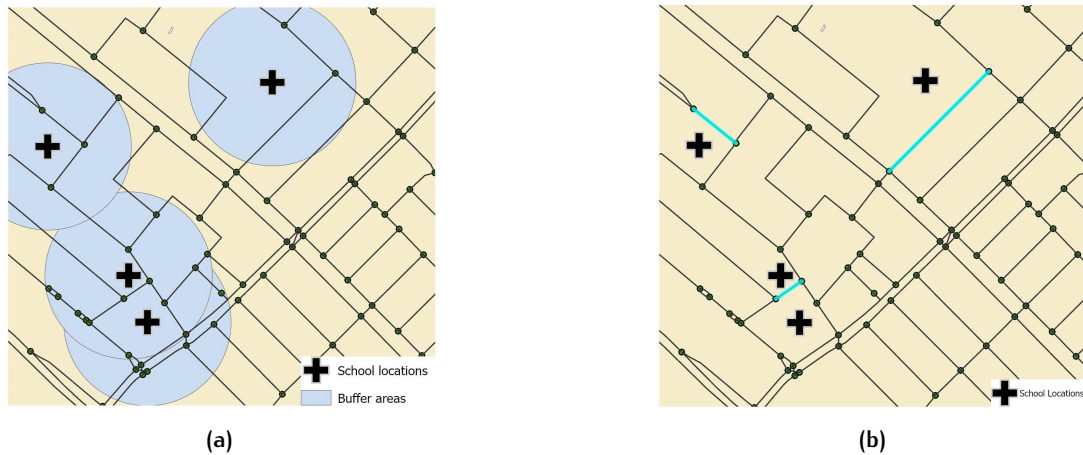
Figure 7.9: (a) Buffer method (b) Nearest edge method for network projection of school locations

7.3.1 Topological Network Analysis

The Helsinki road network was analyzed using the NetworkX package to collect basic topological information such as the development of number of nodes and edges, and the average degree of the network, as displayed in Table 7.2. The initial investigation showed that the number of nodes and edges increased over time. Based on qualitative inspection, new edges were formed, either connecting already existing edges or branching out of existing edges to new areas. This observation coincides with Strano et al. [2012]'s classification of road network growth, separating the process into the exploration and densification stages. Their study quantifies such observations using betweenness centrality of new edges formed, however due to the scope of this study, a qualitative observation and the comparison of topological network metrics were preferred to confirm the existence of topological differences for every time step. Indeed, the road network has evolved between 1991 and 2016, and the historical analysis is possible based on complex network analysis. The visualization of Helsinki road network for every time step is given in Appendix C.1.

Based on the initial complex network analysis of road networks for each historical timestep, an initial insight is obtained about the topological network evolution. As tabulated in Table 7.2, the number of nodes and edges increase through time in the road network. This suggests that new road segments are formed, potentially branching out to new zones or developing new alternative roads in the already established regions. In order to investigate how new roads develop, the average node degree metric is used. According to this metric, the average number of neighbors of each network location is calculated. Throughout time, this value decreased, specifically from 1991 to 1999. This indicates a lower connectivity of nodes in general.

For further investigation of connectivity, another topological metric, *gamma index* is calculated to investigate how exhaustive are the connections between nodes. The gamma index indicates how much the network fulfills its potential with regards to connections formed. Casali and Heinimann [2019] define a theoretical proportionality between the average node degree and the gamma index. Their derivation shows that as the number of nodes approaches infinity, the gamma index approaches 1/6

of average node degree. The findings confirm this trend, especially in the last two timesteps. Furthermore, the resulting average gamma index values show a decreasing trend, although not to a great extent. Therefore, a slight drop in connectivity is observed overall.

Table 7.2: Basic topological metrics of the Helsinki road network in the selected timesteps

	Nodes	Edges	Average Degree	Gamma
1991	9574	12962	3.1	0.451
1999	10856	14279	2.73	0.438
2007	11222	14722	2.64	0.437
2016	11330	14831	2.62	0.436

For a more in-depth analysis of the gamma index, a subdistrict based analysis is conducted, as shown in Figure 7.10. In this analysis, gamma indices are individually calculated for each subdistrict, displaying how well the nodes are connected within each subdistrict. Through a visual inspection, the high gamma districts in the western part of Helsinki do not change in distribution, whereas the number of subdistricts in the eastern part have a gradual decline in their gamma index. This indicates a decreased connectivity in these regions. The historical inspection of gamma index is relevant for the random walk results, as the number of edges within a subdistrict increases the total access diversity through walks. When more edges connect the nodes, more access routes are formed. Therefore, it is an important index to evaluate the infrastructural impact on the random walk outputs.

Figure 7.10: Calculated gamma values for Helsinki subdistricts through time

Another index that was spatially analyzed was the node degrees within the network. The node degree indicates how many possible paths can be formed from a selected node. This has a direct influence on the random walk outputs, since regions with higher average degree offer more diversity in the paths. Therefore, there are more ways to reach a nearby school node, given that such a school node is located within the distance threshold. The output of the node degree estimation is provided in the Appendix D.1.

The hotspot analysis for node degrees in the City of Helsinki network model is visualized in 7.11. The historical development of node degrees show that clusters of high degree nodes do not change over time significantly. What this indicates is that the number of neighbors per network node has not changed in spatial distribution to a great extent. As previously discussed in 7.2, the average node degree has decreased through time, and this decrease did not change locations and size of the high and low degree node clusters. The only conclusion can be made for the historic city center where the statistical significance of the node cluster has decreased, which may be explained by the expansion of the network in that region to branch out rather than densify further.
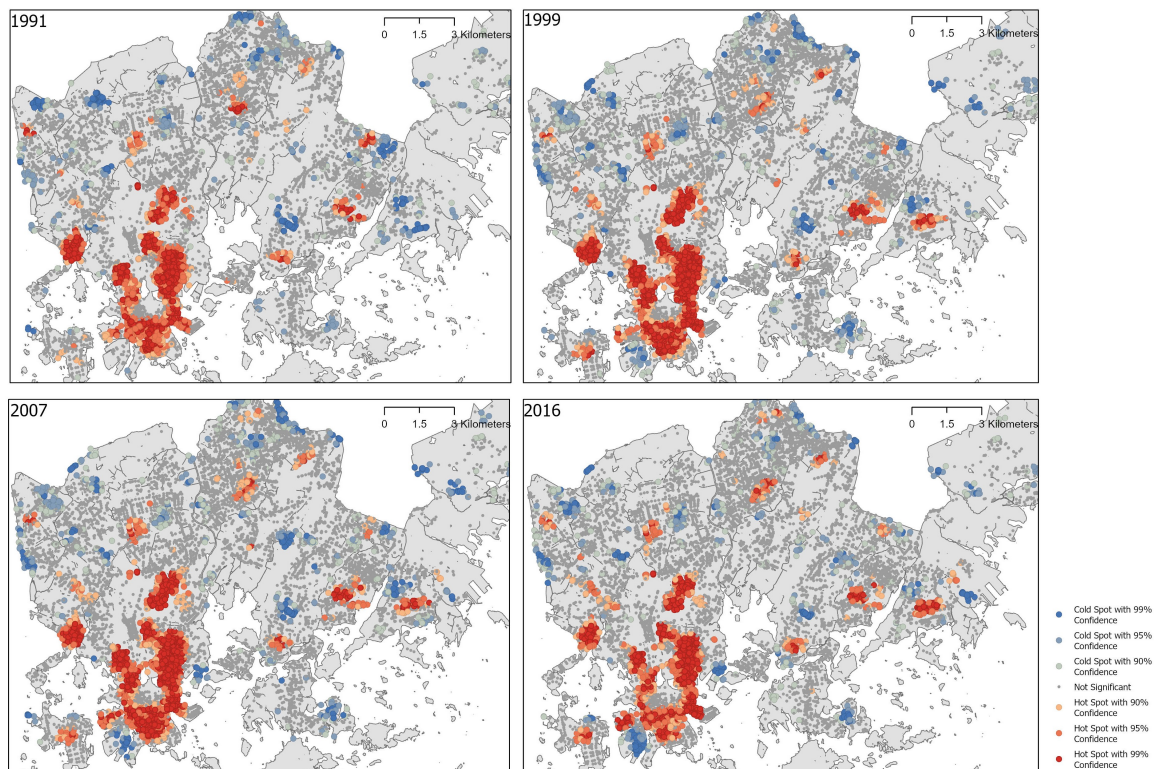
Figure 7.11: Node degree distribution over time based on hotspot analyses based on K-nearest neighbor approach (K=30)

7.3.2 Spatial Analysis of Schools and Districts

This subsection explores the spatial distribution of schools over the timesteps, and makes a subdistrict based assessment based on the number of schools and the population in each subdistrict.

Figure 7.8 illustrates the spatial locations of schools, and their evolution through time. The initial inspection of the collected school locations data shows that the number of primary, secondary, and high schools decrease over time. However, one can also observe an improved spatial distribution of schools, based on Figure 7.12. There is a shift from the accumulation near the historical center to a more dispersed distribution of schools from 1991 to 2016.

Figure 7.12: Number of schools per subdistrict

In order to understand where people reside, an spatial map of population is used, shown in Figure 7.13. It is clear that most accumulation of population is still in the historical center in all historical timesteps. High population districts do not change over time to s a significant degree. Some population increase is observed near the south harbor next to the historical district. As previously discussed, this area has been continuously transforming since the closure of the freight port in the 1970s.
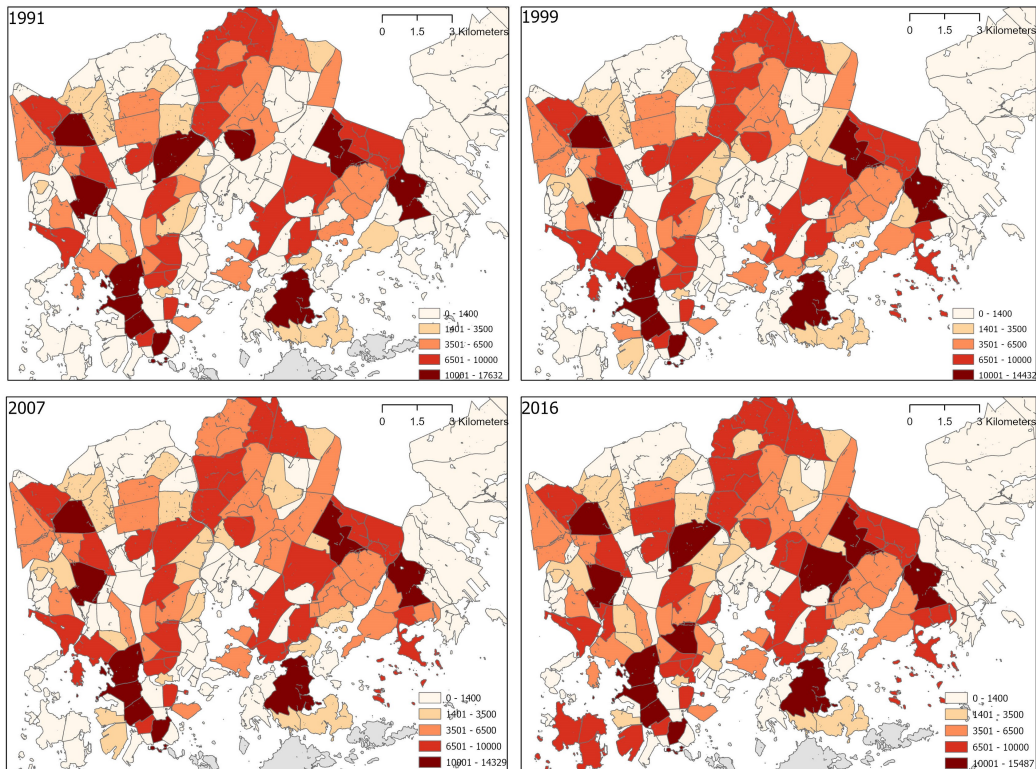
Figure 7.13: Population visualized on subdistrict level

7.4 MODEL IMPLEMENTATION

This section explains the selected parameters for the random walk model, and the reasoning behind the selections in the case study context, and the model characteristics.

7.4.1 Parameter Selection

In order to analyze the collected and processed data as accurately as possible, the case study context was investigated with regard to regional statistics. Specifically, travel speeds of the transport modes, and the average travel time of users needed to be collected for the accurate random walk threshold selection.

First, the dominant travel mode for each accessibility case, were studied to determine the scope of the random walk model. As discussed in section 7.1, according to the Travel Survey conducted in 2016, walking is the most common travel mode for reaching school and study destinations, where 26 out of 60 journeys to school are on foot. Walking is closely followed by public transport use (25 out of 60). From these two most dominant modes, historical data of public transport routes in Helsinki are not provided in an open database. As the collected data includes the road network, the random walk model uses the walking distance threshold. It is noteworthy that the road network data includes the automobile roads, and a there is lack of information regarding the distinction of pedestrian paths. For this study, the pedestrian road network is assumed to be identical to the car road network.

The two parameters which determines the travel threshold of the random walk algorithm are the travel time and speed of transport users. For the latter, the Helsinki Journey Planner was consulted, wherein the walking speed of a pedestrian is determined as 70 meters per minute. This is an average speed of a pedestrian considering all the impedances related to traffic lights and crosswalks [Tenkanen and Toivonen, 2019], and assumed to be constant in this analysis. For the travel time, the Helsinki Travel survey is used. Based on the data shown in Table 7.1, the travel times for trips to study locations are lower than the average, with 20 minutes per trip per individual. This value, together with the walking speed, is used for the determination of the distance threshold by the simple multiplication of speed and time. Thus, the distance threshold for school accessibility via walking is determined to be 1400 meters.

This study compares four different historical data regarding changing road networks, land uses, and socioeconomic indicators. As expected, the travel behavior of individuals would be subject to change over the historical timesteps. Nonetheless, for the sake of consistency through historical analysis using a specified quantitative analysis, the parameters are kept constant, based on the estimations made based on 2016 data as previously discussed.

7.4.2 Verification

This section outlines the parameter verification process for the random walk analysis in the Helsinki region. While the walk distance threshold depends on the real world context of mode selection and the corresponding movement speed, there are also certain parameters to the model which are defined by the analyst. The most significant parameter in the case of random walk is the total number of walks, or iterations, which will be used in the analysis. This parameter plays an important role in the aggregation of data and the consistency of the output. Due to the random nature of the algorithm, having a significantly low number of walks would result in unrealistic outputs, jeopardizing the integrity of the study. Specifically considering the number of nodes, and the average degree of nodes in the network, every starting node offers a variety of paths that can be taken across iterations. On the other hand, increasing the number of walks could mean observing more variety in distinct paths. However, it should be noted that constantly increasing this number would lead to a diminishing returns effect after some point. Therefore, selecting the number of walks for a more consistent outcome is an important part of this study.

For the verification of the number of walks, a series of sample analyses were conducted. For this purpose, a limited area was selected from the sub-districts of the Helsinki urban area. The chosen sub-district, Etela-Haaga, has an area of 2.3 square kilometers and 187 nodes. The district includes 5 school nodes in 2016. The sample run was formulated as follows:
1. Starting nodes must be in the selected region
2. Walks are not restricted to the selected region, free to continue on the edges outside
3. Walk threshold is 1400 meters based on the pedestrian walk case
4. The analysis is conducted using 7 separate number of walks: 50, 100, 250, 500,
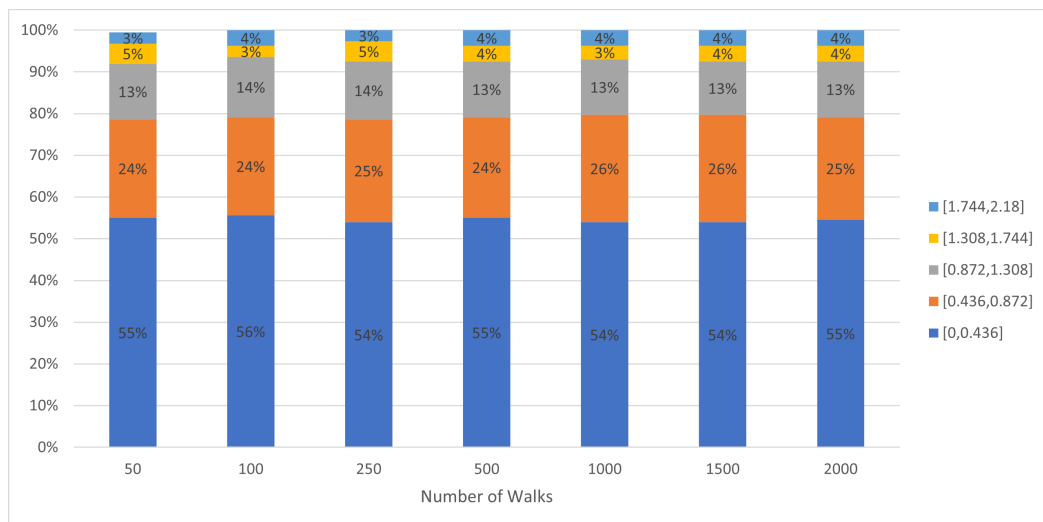
Figure 7.14: Distribution of *vpw* values for sample analyses of 50, 100, 250, 500, 1000, 1500, and 2000 walks

1000,1500, 2000

The random walk algorithm was run with the seven distinct number of walks. Based on 7.14, the number of walks did not have an impact on the distribution of calculated *vpw* values. According to the distribution of 5 equal sized value bins, the share of nodes in each bin was steady for every instance.

Furthermore, for a more in-depth understanding of the effect of number of walks, three distinct nodes were selected within the sub-district. Comparing the mean and standard deviation of average school visits across walks, the sensitivity of the metric to the number of walks were analyzed. As expected from the previous assessment, the mean value of the *vpw* showed a relatively steady trend irrespective of the number of walks. However, the standard deviation displayed higher values for 50 and 100 walks, which started to level only after 250 walks. Therefore, 500 walks or more would be needed for the accessibility metric to converge on its variance. However, considering the computational load directly proportional to the number of walks parameter, going beyond 500 walks was not preferred. The results of this verification analysis are shown in Figure 7.16.

7.5 RESULTS

This section provides the results of accessibility analysis using the self-avoiding random walk for both school and job accessibility scenarios. Comparison of every historical timestep is made by descriptive statistics of the outcomes and spatial analysis. Using the accessibility outputs, the results of the equity analyses using Gini, Spatial Gini, and Theil indices are provided.

As an initial overview of the random walk results, the access diversity of models for each timestep simulation are given in Figure 7.17. This figure shows the
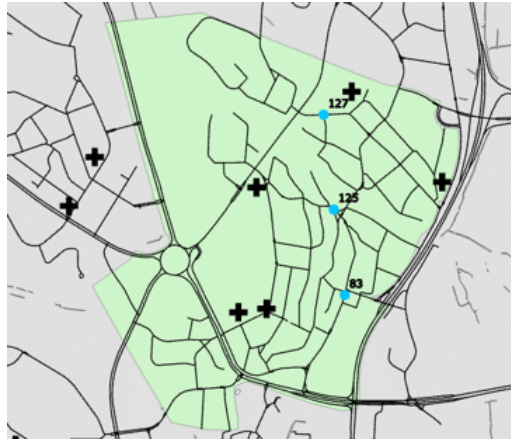
Figure 7.15: Three selected nodes in the Etela-Haaga district for verification of number of walks parameter
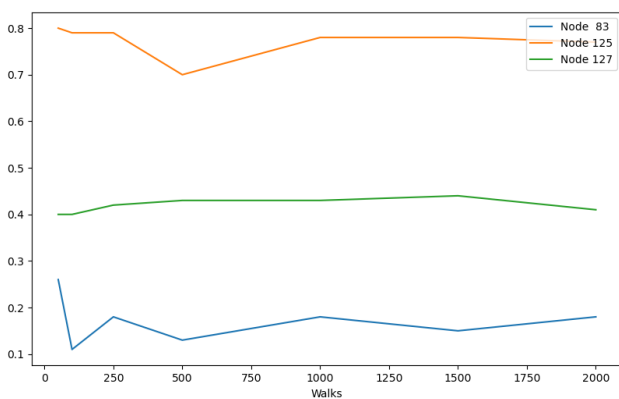
(a) Mean *vpw*
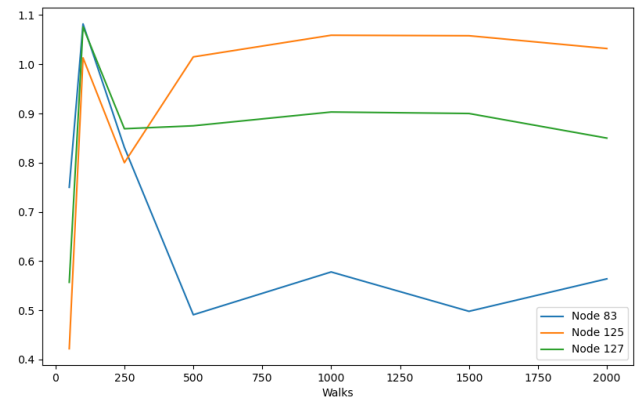
(b) Standard deviation *vpw*

Figure 7.16: Mean and standard deviation visit per walk metric for three selected nodes in 50, 100, 250, 500, 1000, 1500, and 2000 walk models

probability density of the frequency of visited unique nodes in every random walk simulation, without considering the school locations. Based on this output, it can be suggested that with the same parameters, 1991 road network resulted in a flatter probability density compared to the other years, meaning that an average node had access to more diverse nodes. This result gives an indication of network topology and how the road infrastructure influences the diversity of locations every starting node can reach. Thus, based on this output, it is inferred that road network developments after 1991 have led to a diminished access diversity.

The school accessibility measurement incorporates the school locations into this approach to get a view of accessibility to opportunities. The following paragraphs explain the results of school accessibility and analyze them based on spatial and socioeconomic factors, and eventually the application to a series of equity assessments.
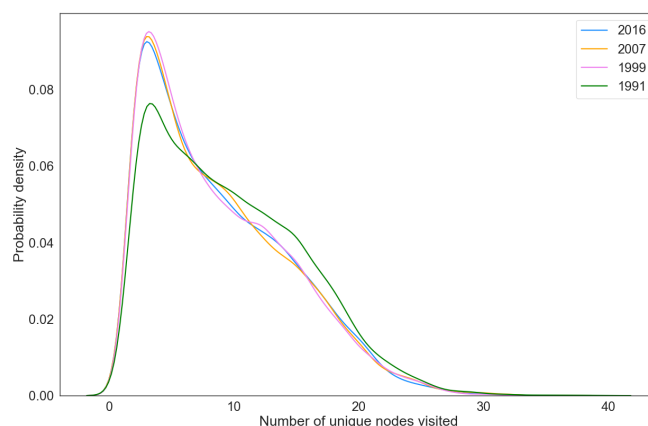
Figure 7.17: Probability density of number of unique node visits per starting node

7.5.1 School Accessibility Measurement

Based on the methodology provided in section 6.3.2, and the parameters described in section 7.4, accessibility measurements for schools are carried out for the selected years of 19991, 1999, 2007, and 2016. The general descriptive statistics of the results of the analysis are shown in the Table 7.3, and the spatial mapping of school access outputs are provided in Appendix E.1. According to this overview, the aggregate accessibility to school metric displays a decreasing trend. This would be expected considering the increase in the number of nodes in the network and the reduced number of schools, as described in 7.3. In addition, the share of nodes which never visit a school node across 500 walks also increases in time. Starting from 12.7% of non-visiting nodes in 1991, this value reaches 24.3% in 2016. As expected, this translates into a decrease in the median value of school visits across all nodes in the network, which potentially means lower school accessibility is more commonly observed. Overall, it can be observed that the overall accessibility decreases from the network perspective. However, it should be considered that network nodes are heterogeneous, meaning that they represent different share of the population. To account for these differences, the following paragraphs propose an in depth analysis considering the population.

Table 7.3: Descriptive statistics of accessibility based on school visits per walk in the whole network

School Visit Per Walk Statistics\Year	1991	1999	2007	2016
Non-school visiting node percentage	12.7%	16.5%	18.6%	24.3%
Mean	0.446	0.323	0.285	0.241
Standard deviation	0.554	0.461	0.427	0.402
Median	0.234	0.126	0.098	0.060

It is important to note that the overall statistics do not provide a comprehensive insight into accessibility, as all locations in the urban environment are included in this analysis. In order to get a more in-depth view, spatial analysis based on the distribution of accessibility metrics and the land-use data is considered. Firstly, the spatial distribution of low-access nodes is analyzed, specifically nodes which never
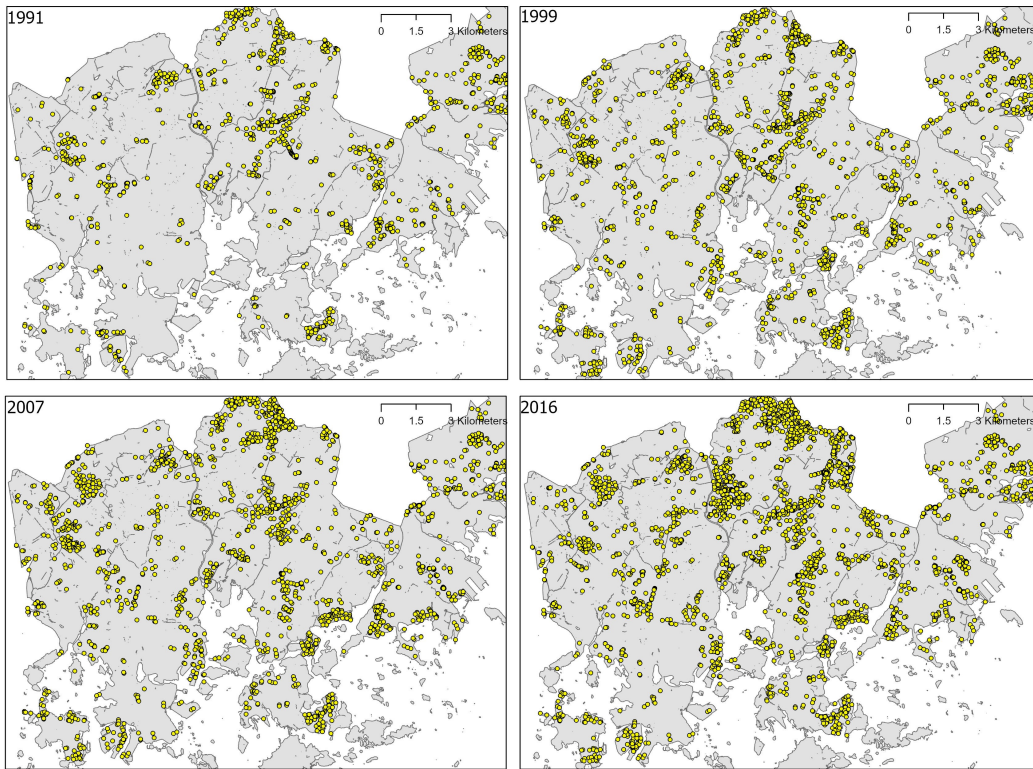
Figure 7.18: Spatial distribution of nodes with zero school visits across 500 random walks

visit a school node in any of the 500 walks are investigated. Figure 7.18 shows the distribution of such nodes in every historical timestep. It can be observed that the the nodes with low numbers of school visits are accumulated in the northern region in 1991, but get more dispersed through time.

Followed by the analysis of accessibility from the perspective of starting nodes, another important indicator is the number of times each school is visited. In this study, the capacities of schools are not included, due to the lack of data. Thus, the supply of activities is only considered through the number of opportunities. However, with the estimation of this indicator, it is possible to deduce what should be the approximate capacity for each school based on the random walk, and which schools attract more demand based on their locations. As illustrated in Figure 7.19, the most visits are accumulated in the historical center of the city, similar to the distribution of school locations.

Another determinant of the school demand in this model is the number of nodes and edges in the periphery of the school locations. As each node is a starting point for the random walk, the number of times a school node is visited is directly influenced by the average node degree. Compared to the node degree cluster analysis, the road network around the historical center is better connected with higher number of neighbors, which influences the school accessibility in this region. Therefore, if a region of the network is denser in terms of number of connections, the random walk based model measures higher number of school visits consistently.
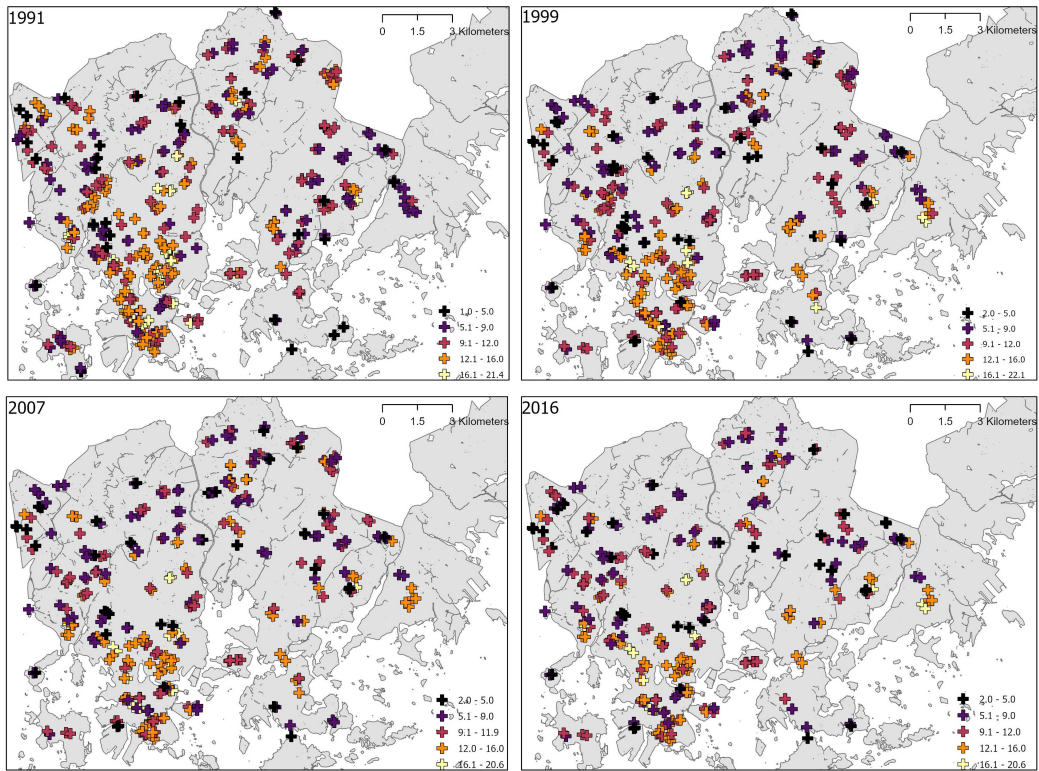
Figure 7.19: Average number of visit per school per walk

Socioeconomic Assessment

District based socioeconomic statistics enable a more informed analysis based on the residential hotspots in the network. Therefore, a specified analysis would focus on the accessibility of regions where residential population is the highest. For this purpose, 10 sub-districts are selected based on the population. A comprehensive overview of the school visit statistics are included in Appendix E.1.

Zooming in on these 10 districts, it is observed that there is higher accessibility to schools than the overall picture. As the mean school visit per walk metric is higher than the urban average (see Table 7.3. However, investigating the trend in the mean and median school visit metrics, most districts display a drop similar to the urban average. This could be explained by the drop in the number of schools.

Another real-world factor to consider is the population of the specific group of interest in the case of education. An age based assessment of population distribution enables a specialized analysis in the population group that is the most affected. A subdistrict based visualization of population below the age of 19 is provided in Figure 7.20. According to this map, most student population is located in districts outside the historical center. Therefore, the development of accessibility in these regions is an important consideration for this analysis. The next section analyzing the clusters of high and low accessibility will provide a comparative analysis based on these socioeconomic indicators.
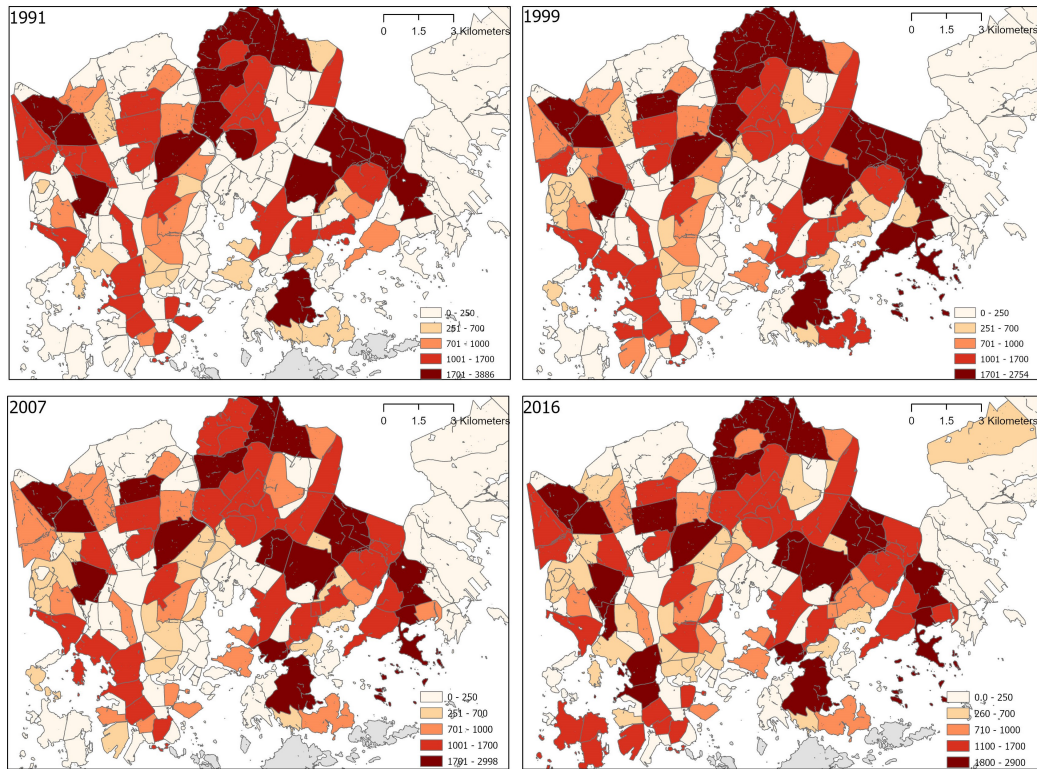
Figure 7.20: Population aged 19 or lower by subsdistrict

7.5.2 Hotspot Analysis

For the assessment of statistically significant clusters, hotspot analysis was conducted on the accessibility metric for each timestep. As described in the *methodology* (6.4) the Optimized Hotspot Analysis tool is used for this purpose. Based on the optimized parameters of this tool, the most appropriate distance band was selected for the K-nearest neighbors approach. The parameters for each timestep are given in Table 7.4

Table 7.4: Optimized hotspot parameters for every timestep

year	1991	1999	2007	2016
K-nearest neighbors	30	30	30	30
distance band (meters)	374	366	520	361

The results of the hotspot analysis are mapped in Figure 7.21. The overall statistics of the output results are provided in Appendix E.2. The percentage of hot and coldspots show no consistent trend, however it is noteworthy that the outputs of 2007 displayed the highest level of hotspots and coldspots which are in the 99% confidence interval of statistical significance. After 2007, the number of nodes in the 99% confidence dropped on both sides of the spectrum. This could indicate a lower level of clustering, however the overall share of nodes above the 90% has increased, still showing clustering to a high level of confidence. Lastly, compared to the inital two timesteps, the share of statistically significant cluster nodes is slightly higher in

2007, and 2016, which could be inferred as a higher level of clustering and spatial dependence.
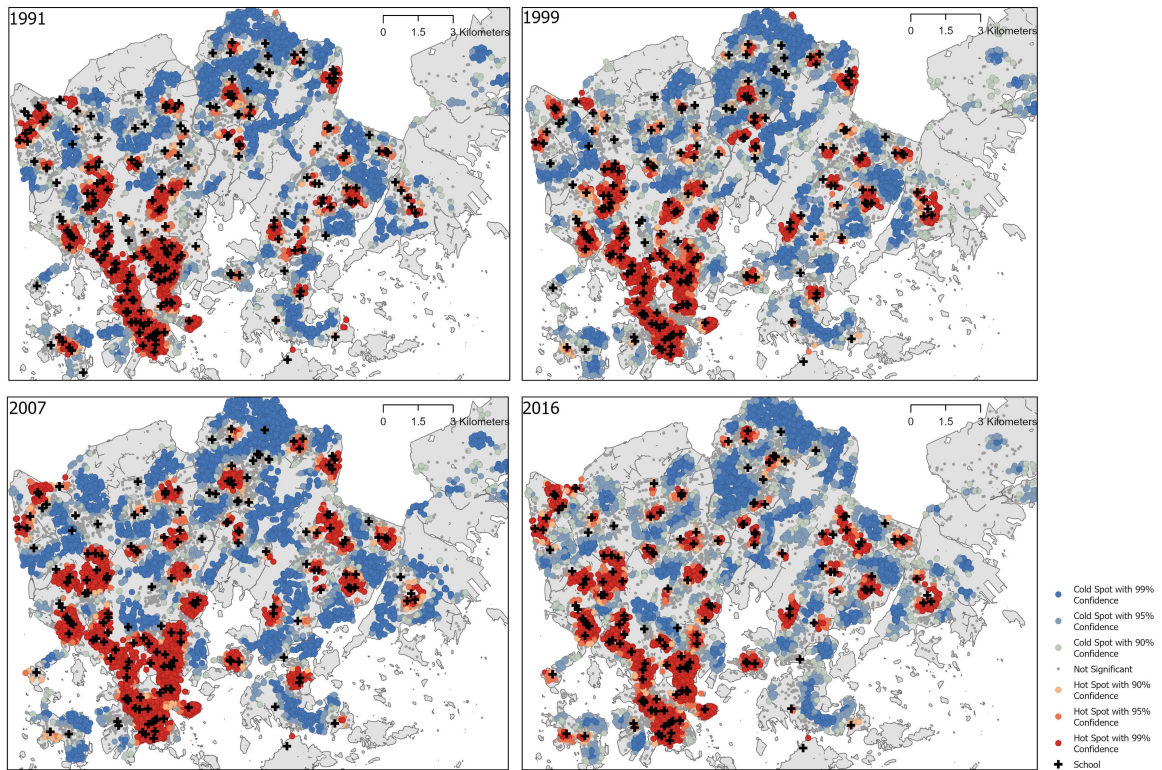
Figure 7.21: Hotspot analysis of school accessibility

According to the spatial mapping, the distribution of hot- and cold-spots can be compared across timesteps. When the 1991 outputs are investigated, it is clear that a large cluster around the historical center is observed, which is expected from the high density of schools in these districts (see Figure 7.12), as well as the clusters of high node degree (see Figure 7.11). Nonetheless, there are also smaller statistically significant clusters which form around schools in the northern subareas of the city. A striking observation is the lack of high accessibility clusters around certain schools, or even the existence of cold spots. Some of these clusters can be explained by the network characteristics, such as the low network density around those schools. Nonetheless, this observation shows that the accessibility metric is not fully correlated with the school locations, but transport network characteristics also play a role.

Looking at the node degree hotspots and the gamma index findings, the lack of high access clusters around certain schools can be justified. Figure 7.22 shows a focus area where accessibility to schools is analyzed in comparison to network topology indicators. Region C shows a small high access cluster where both number of schools and the network node degree is high. Region A on the other hand, benefits mostly from the spatial density of schools in the region. Region B displays an example where the lack of high node degree clusters influence the accessibility in the region negatively. This finding validates the feedback cycle defined in section 4.2, as the measured accessibility is influenced both by the transport network and the school locations.
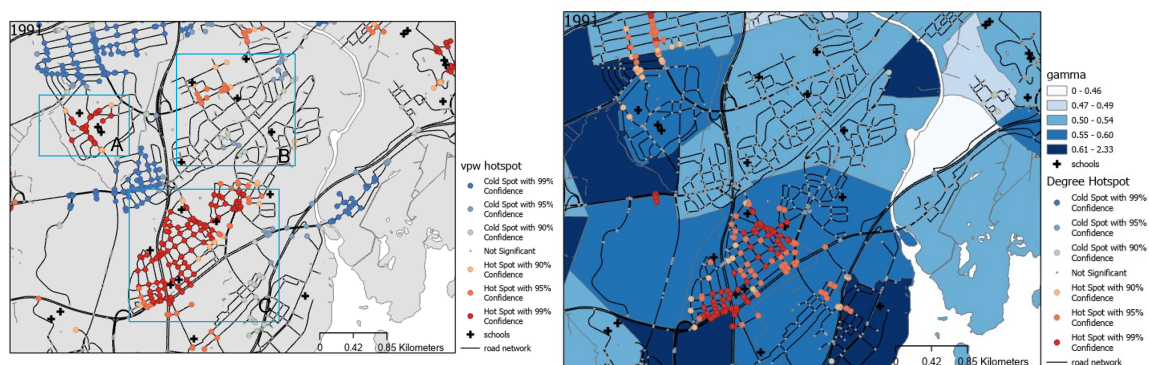
Figure 7.22: A sample study area in City of Helsinki, left: accessibility hotspots right: node degree hotspots and gamma values in the subdistricts

Another significant assessment to be made is the comparison of identified clusters with the socioeconomic factors discussed in the previous subsections. The spatial comparison of subdistricts with population below the age of 19 are made with the hot and coldspots. The qualitative comparison shows a significant mismatch of high accessibility regions and the high population subdistricts for all timesteps. Most significant clusters are not located in high student population districts, but mostly around the historical center.

7.5.3 Equity Analysis

This subsection presents the results of the equity analysis based on Gini, Spatial Gini, and Theil indices.

Figure 7.23 shows the Lorenz curves for school accessibility for all four years considered in this study. This visualization considers the population corresponding to each node, therefore every node accessibility metric is repeated n times, n being the population assigned to the node through Voronoi tessellation. According to the output of this analysis, the area between the equal distribution and the Lorenz curve increases every time step, as quantified by the Gini index. Therefore the equity of school accessibility distribution decreases within the Helsinki urban area.

In order to understand the equity on a regional level, Theil index is utilized on the school accessibility case. The results for every historical timestep is shown in Table 7.5. The results of Theil index calculation shows a similar trend to the Gini indices. The overall inequity rises through each time step, with a significant jump from 2007 to 2016. On the other hand, Theil index offers further insight into the regional equity. Here, the within Theil indices have negative values in high magnitude. This means that the equity inside subdistricts are very high, therefore the accessibility outputs are equitable within their subareas. However, the between Theil is comparatively high on the positive side, contributing to the inequity overall. Therefore, the spatial correlation is very high when nodes are aggregated in their subdistricts. It is noteworthy that between and within Theil values show a very sharp difference due their large magnitudes on the opposite sides of the spectrum. These values having positive and negative values are expected as discussed in section 3.2. However, the
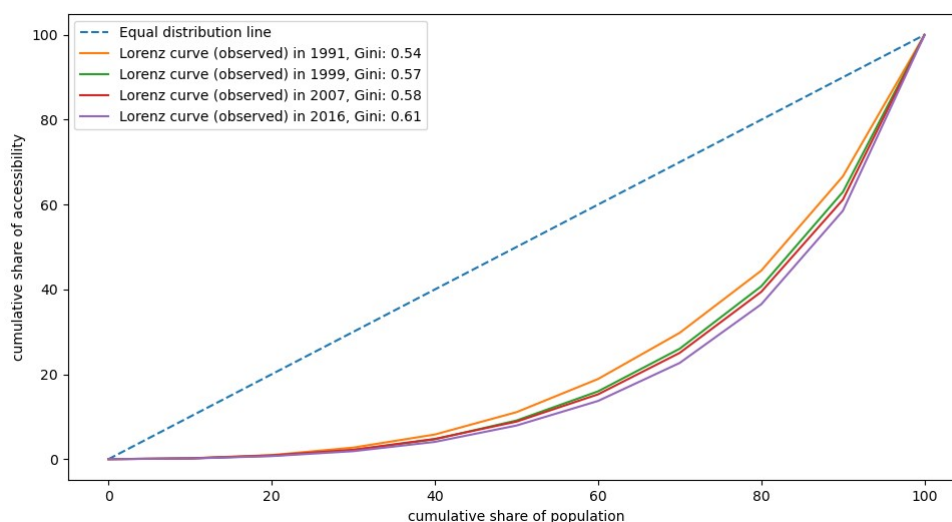
Figure 7.23: Lorenz curves and calculated Gini coefficients based on school accessibility for all timesteps

large magnitudes are caused by the formulation of Theil index, where very small attributes, close to zero, contribute to a very high output due to the logarithmic formulation. This is a shortcoming of the proposed accessibility measurement, and discussed further in chapter 8.

Table 7.5: Theil index decomposition for school accessibility

	1991	1999	2007	2016
Theil index	0.662	0.7998	0.8502	0.9802
between Theil	84.09	101.02	84.01	82.47
within Theil	-83.43	-100.22	-83.16	-81.48

Lastly, Spatial Gini index is measured to evaluate the neighborhood effects on equity. The results from this analysis indicates that for all timesteps, the values contributing to inequity are the differences between far node pairs. For this reason the near difference value is small. Based on this, it can be deduced that the spatial characteristics play an important role in the accessibility values, complementing the output of the Theil index. The pairwise relationships disregarded by Theil are considered in these measurements, thus completing the spatial equity analysis. Thus, the inequity among distant node pairs is higher compared to the neighboring nodes based on the Spatial Gini. The p-value assessment based on z-scores show that the calculated Spatial Gini indicators are statistically significant. This means that, the null hypothesis that randomly distributed accessibility values would yield the same outputs is rejected.

The results of the equity analysis shows a decreasing trend in the equity of school accessibility. Furthermore, spatial equity analyses using Spatial Gini and Theil both confirm a spatial correlation in the accessibility distribution. Based on Spatial Gini, the inequity contribution of distant node pairs is very high. The same assessment cannot be made for Theil based on this accessibility metric.

Table 7.6: Spatial Gini index results for school accessiblity

	1991	1999	2007	2016
near difference	0.0003	0.0003	0.00003	0.00003
far difference	0.6139	0.6654	0.6822	0.7209
p-value	0.01	0.01	0.01	0.01

7.6 CASE STUDY CONCLUSIONS

The subsections of this chapter initially introduced the case study area of City of Helsinki, explained the data collection processes, conducted an exploratory analysis. These steps were followed by the implementation of the self-avoiding random walk model and the analysis of the results. Historical school accessibility analysis showed that the overall accessibility has diminished through time. This is partly because of the reduction in the number of schools, while an increase in the number of network nodes was observed. Comparing the resulting spatial clusters from the Hotspot Analysis with the exploratory analysis results showed that the historical center of the city is the largest cluster of high accessibility nodes. Thus, the dominance of the urban center still remains, although many small clusters around school locations exist. These small clusters are expected in the areas where high values of node degree are observed, or the high connectivity based on gamma values. Nonetheless, it was also observed that network nodes around some school nodes were not estimated to be statistically significant clusters. The lack of high node degree in the region contribute to this outcome. The average node degree and gamma index estimations around those locations showed that network topology is an important factor in these low confidence areas. Therefore, the school accessibility is not only explained through proximity to schools, but also network characteristics.

The results of equity analysis showed a negative trend in the equity for all three equity indices used. The general evaluation of equity with respect to population using Gini index resulted in a gradual increase in inequity. When the neighbor relationships in the equity assessment are considered through Spatial Gini, high spatial dependence was observed. This means that the neighboring nodes contribute to accessibility results significantly. For the Theil index, which analyzed the regional equity within and between subdistricts of Helsinki, an overall reduction in equity was observed in general. The contribution of within and between region equities could not be measured due to the divergent results of the logarithmic function. This is caused by the high number of zero, or near zero accessibility nodes. Thus, Theil index is not suitable to use for this accessibility metric.

8 | REFLECTION ON THE METHOD

This chapter makes an assessment of the developed methodology and its applicability to other use cases. In the previous chapter, the results of the case study were discussed, whereas this chapter adopts a broader perspective to the use of complex network theory in accessibility measurement, and its usability in the equity context.

DATA QUALITY

The analysis methodology used in this study relies on the open data of City of Helsinki. Specifically, road network data, subdistrict based socioeconomic and demographic data, and school location data were collected for this study. The collected road network data only includes roads suitable for car trips, thus missing the pedestrian and cycling routes. As a result, the use of this network is restrained to certain assumptions when modes other than personal car are considered. As the main transport mode in school accessibility is walking, a comprehensive road data is required for every timestep, including the walking paths.

The school location data is comprehensive in terms of historical timesteps, however lacks the capacity data of schools. Consequently, the capacity constraints could not be implemented for the supply of schools, making it less accurate for the real-life analysis of accessibility. Shen [1998] maintain that the capacity and competitions effects offer significant value to accessibility measurement, which had to be neglected for this study.

Lastly, subdistrict based historical socioeconomic data made it possible to aggregate the accessibility outputs based on the zones defined by policymakers. The quality of this data was appropriate for this study.

EXPLORATION USING COMPLEX NETWORK METRICS

The first use of the complex network theory was the exploratory analysis of network evolution in the case area. Topological analysis of the network is a frequently visited application of the complex network theory, as discussed previously in section 5.1.3. For the case of accessibility measurement, the use of complex network metrics makes it possible to analyze the impact of transport systems on accessibility. Thus, a road network analysis through time is used to understand how the network spatially develops, as in whether it is densifying or still branching out. To understand the overall changes in topology, basic values such as the number of nodes and edges were referred to. While these numbers indicate whether a network is growing or not, they do not give the whole picture of how node connections are formed. Thus,

two other metrics, node degree and gamma index, were used to assess how well nodes are connected. This analysis was also carried out with respect to subareas in the case study, and has potential to be developed into more specified geographical units based on the use case. The results of this exploratory analysis proved useful to distinguish the effect of network structure and the school locations. As a result, it was possible differentiating transport and land use factors.

ACCESSIBILITY MEASUREMENT USING SELF–AVOIDING RANDOM WALK MODEL

The self avoiding random walk model was used for the assessment of the school accessibility case. Accessibility measurement based on this model can be classified as an infrastructure-based measurement approach, based on the classification made by Geurs and van Wee [2004]. The advantage of this method is the use of road network to determine the range of movement and introduce the concept of randomness to achieve a probability distribution of reaching diverse nodes. These individual movements are modeled with the assumption of no prior information. Practically, this leads to non-targeted movements without the consideration of other factors such as locations of activities or land use patterns. When not addressed, this lack of contextual information in the model would make it not very useful modeling a real world case. However, with the inclusion of activity location such as school nodes, the model is made more coherent with the transport land use cycle proposed by Bertolini [2012] and used in the accessibility factor conceptualization in Figure 4.2.

While the movements are still only dependent on the road infrastructure, and influenced heavily by the topological factors, inclusion of activity locations in the assessment makes this model more applicable to the accessibility measurement use case. Referring back to the accessibility factor conceptualization, the model has several shortcomings. First, the capacity of opportunities are not considered in the model adapted to the case study. Therefore, the supply of opportunities which leads to the accessibility factor of availability is only limited to the number of schools. This could be misleading in an example case where the number of schools are diminishing, but the capacity stays the same. When considering the school accessibility, it is important to consider the capacity of schools and limit the random walk visits to these school based on this capacity value. However, this would increase the computational load on the model, and would require a simultaneous simulation of random walks from all the nodes. The current model takes a sequential approach, simulating random walks one starting node at a time. Furthermore, the available case data lacked the school capacity information, therefore the model implementation did not consider implementing this constraint.

Another factor about the usability of this model is the choice of the distance threshold. The parameters selected in this study were based on the walking speed and distance range, however for a study of car-based accessibility, the distance threshold is inevitably going to be higher. When such a case is analyzed, there are two main points of concern regarding the self-avoiding random walk model. First, as previously discussed in section 6.3.1, increased travel threshold causes a downward

facing parabolic trend as the amount of newly visited nodes are concerned. Due to the self-avoidance property, it is often the case that increasing the threshold beyond a certain point does not lead to a proportionately increased area of reach. This is due to the increased likelihood of encountering the stopping conditions of the random walk as the distance threshold increases, as described in section 6.3.1. Thus, for a car-based accessibility case, the actual maximum reach of the random walks need to be analyzed to understand whether the distance threshold reflects the output results. Secondly, as longer travel distances are considered, the effect of competition in the transport system become more pronounced. Therefore, the model choice of not including public transport becomes more questionable when an accurate model is the objective. However, for studies evaluating car accessibility specifically, this method is still applicable. For the inclusion of public transport, a multi-layer approach can be used, which has applications in literature [Ding et al., 2021], but not considered in this study.

An important point of discussion would be about the random walk model's computational performance. As described in the parameter verification section (7.4.2), increasing the number of walks and the distance threshold proportionately increases the model's computation time. Therefore, the computation time is one of the constraining factors of this model in a larger area of implementation, or in cases where a higher number of walks would be necessary. In the selected use case, the number of walk were limited to 500 as no significant change was observed in the consistency of the output with higher number of walks. In addition, the distance threshold was selected in consideration of the real-world situation where schools are often accessed by walking. These parameters did not lead to extremely long computational times. However, for implementations of a larger scope, high performance computing approaches could be needed. The reason stems from the way the random walk approach is designed, that is, compensating the randomness factor by a high number of repetitions to obtain a well-rounded probability distribution.

EQUITY ANALYSIS

The accessibility measurement results of the self-avoiding random walk were analyzed in terms of equity. Due to the network driven analysis, the output of this model were disaggregated in terms of the network node. However, each network node corresponds to a certain number of population, as estimated by Voronoi tessellation. Therefore, when making an equity assessment based on the population network nodes do not directly correspond to the population, some nodes even located in areas with zero population. Therefore, the before conducting an equity assessment, the network outputs require processing based on population distribution. For the Gini index calculation, such an approach was assumed, contributing to the increased accuracy of the equity assessment.

Spatial Gini index was used to measure the neighborhood effects on the overall Gini measure and compare with the equity among distant locations. The effect of population is again relevant in this case, as the nodes with zero population offer no insight into the equity across groups of individuals. However, due to the prede-

fined neighbor relationships in the network representation, removing zero population nodes could potentially impact the near differences in the spatial Gini calculation. As one or more neighbors are removed around a subject node, the composition of their surroundings' accessibility values also change, which is not desirable for measuring the pairwise effects on the equity measurement.

For the case of Theil index, the accessibility metrics resulting from random walk could be analyzed in terms of the historical development, and the results were coherent with the other two equity indices. Yet, this index is not particularly applicable to the accessibility measure when the within region and between region effects are considered. As there are several zero accessibility nodes in the case area, the logarithmic formulation of this index fails to calculate their contribution to the equity accurately. To account for this issue, zero values can be replaced with very small values close to zero, yet this does not entirely fix the issue, as the sums of equity have very high magnitudes. Therefore, it is impossible to measure the share of contribution of equity of between and within equity. Instead, the only conclusion can be made in therms of the positive or negative effect of equity of these components. When the within Theil index is negative, an equitable distribution within the region is observed, therefore spatial correlation is very high. Hence, the observation is limited to whether spatial dependence exists or not, but it is not possible to measure the extent of contribution. To address, this issue, the accessibility metric should be reformulated to not have zero values in its range, or the Theil index should not be used for other accessibility use cases.

9 | CONCLUSION AND RECOMMENDATIONS

The final chapter of this report presents to conclusions of the study and recommendations for future research.

9.1 CONCLUSIONS TO RESEARCH QUESTIONS

This study was carried out to answer the main research question:

How can urban transport equity be temporally analyzed in terms of accessibility to education using complex network theory?

In order to answer this question, four separate subquestions were formed. Each question seeks to either understand the theoretical background behind a concept, or assess the applicability of a method.

SQ1: What are the existing accessibility based transport equity evaluation methods and indicators?

This subquestion was investigated by the means of literature review on transport equity measurement. As a result of the review, the components of an equity measure were identified, and were exemplified in the accessibility use case. Based on the study of Martens et al. [2019], three main components of equity measures were identified, benefits and burdens, differentiation factors, and the governing equity principle. In addition, the commonly used equity measurement methods were investigated based on their use cases and associated equity principles. Based on this, the case of accessibility based equity evaluation was investigated in terms of the benefits and burdens. It was found that the operationalization or the method of measurement of accessibility has a significant role in the identification of benefits and burdens. Thus, the review also included the types of accessibility measures based on the extensive review of Geurs and van Wee [2004]. Based on this study, and other studies in literature who work on accessibility based equity approaches, the use case of every accessibility measure were identified. Thus, this subquestion was answered by a compilation of equity measurement concepts, and the exploration of accessibility case based on the measurement concepts. Reviewed studies were categorized based on the type of accessiblity measure, euqity principle, and the usee equity index. The review was concluded with the selection of suitable equity indices for the developed methodology, which are namely Gini, Spatial Gini, and Theil indices.

SQ2: What are the factors that influence transport accessibility to education opportunities?

For this subquestion, initially the literature regarding accessibility in the land use and transport context was consulted. Based on the feedback cycle defined by Bertolini [2012], the main components of the accessibility cycle were identified (i.e. transport system, land use, activities, accessibility) as well as exogenous factors such

as a socioeconomic and cultural factors, or policy and infrastructure investments. Using this background, accessibility to education opportunities were analyzed. A rationale for the specific focus on education was provided based on the argument that accessibility to schools increases the chances of continuation to higher education, which in turn increases the level of economic participation. However, a limited number of studies exist evaluating the factors related to accessibility to education opportunities. Therefore a similar but more commonly visited topic of employment accessibility was reviewed to identify common factors which influence accessibility to education. By the end of this review, common factors of education and job accessibility were conceptualized. Three main factors were identified, namely, affordability, availability, and proximity. Using the feedback cycle of Bertolini [2012], these factors were associated with the components of transport system, land use, or activities.

SQ3: How are complex network methods used in the analysis of urban transport systems?

This subquestion was answered by a literature review of complex network theory applications in the context of transport research. Four categories of complex network indicators were identified, namely, centrality indicators, community indicators, and accessibility indicators. To answer the 'temporal' part of the main research question, the network evolution studies were investigated. The results of the review showed that network evolution is often analyzed in terms of topology, using the relevant indicators such as network density or gamma index. While co-evolution studies exist which consider the land use and transport systems' interaction, these models are often predictive rather than driven by historical data. Furthermore, accessibility evolution is not commonly researched in the co-evolution literature. When complex network theory driven studies were examined in terms of accessibility measurement, two methods were identified, namely space syntax, and random walk based accessibility. The former uses an axial representation of the road network, and fails to consider land use effects [Crucitti et al., 2006]. Considering the accessibility factors identified in the previous subquestion, the more practical method was assessed to be random walk based access diversity method. This method is commonly used in applied physics, with an increased attention from transport research [Travençolo and da, 2008] [Lee and Kim, 2021]. This network analysis method is commonly used to identify network topology driven clustering, but several studies focusing on transport network adapted it for measuring the access diversity [Lee and Kim, 2021].

SQ4: To what extent can complex network analysis be used to evaluate and explain the evolution of spatial accessibility distribution?

Based on the findings of the previous subquestion, a random walk based school accessibility measurement methodology was developed. This methodology relies on complex network models of the road network as well as the locations of schools and spatially defined socioeconomic data. Therefore, a mostly infrastructure-based approach was incorporated with land use concepts and socioeconomic factors. This methodology was then adopted to the selected case study of this study, which is the City of Helsinki. The model parameters were selected based on case specific circumstances such as the identified preferred travel mode in the city (walking) and the travel times of individuals (20 minutes). Other parameters were based on the physical constraints of the algorithm, i.e. self-avoidance and the number of walks. These parameters were verified on a test area, the Etela-Haaga subdistrict before

full implementation. The full implementation of the model on the case area was analyzed in terms of the overall accessibility evolution in the city, as well as the spatial distribution. Lastly, the distribution was evaluated in terms of horizontal and spatial equity. The main findings to answer the SQ4 is that complex network representation aids a spatially driven accessibility measurement method, incorporating transport infrastructure and land use information. The use of this analysis method depends on the availability of extensive spatial data such as socioeconomic and demographic data, as well as locations of opportunities. One added value of the complex network model is to ability account for the existence of multiple paths and the diffusion across the infrastructure. Therefore, rather than using just the total distance between specific origin and destination pairs for accessibility assessment, the network topology and alternative paths are also embedded in the accessibility measure. This enables differentiation of infrastructural effects from the location of opportunities. In terms of temporal analysis, this approach is repeatable for historical timesteps given that the data availability and quality is sufficient to construct and analyze the network model. Therefore, the same analysis steps can be implemented on different timesteps, and the changes in transport infrastructure as well as activity locations can be modeled for every timestep. Thus, the most important added value of the complex network theory is the possibility of measuring accessibility within the temporal context. Rather than assessing location based accessibility based on fixed distances for every timestep, contextual infrastructure and land use data can be utilized. Lastly, the defined methodology is versatile in terms of addition of other opportunities in future research. Using a similar data processing and modeling approach, accessibility to various socioeconomic opportunities and critical services can be measured.

The shortcoming of this approach is the in the equity measurement context where the developed accessibility metric always suitable with the common equity indices such as the Theil index. Therefore, the developed accessibility metric should be compatible with the adopted equity measures.

Compiling all the outcomes of the subquestions, the main research question is revisited. A complex network theory driven methodology is presented to answer the research question. Using historical data regarding transport systems, land use, and socioeconomic data, the complex network representation is a suitable approach to measure accessibility and make equity assessments. The greatest advantage of this methodology is its repeatability across timesteps with the corresponding historical data. The application of this methodology in other use cases of accessibility and equity measurement is subject to certain conditions:

- The complex network model must include land use information regarding locations of relevant opportunities

- The network model and accessibility measurement method should be compatible with the equity index

- Accessibility measurement parameters in the network model should be representative of real-world circumstances regarding travel behavior and infrastructural changes.

- For an accurate temporal analysis, the transport infrastructure, land use, and socioeconomic information must be accurate and consistent within each timestep.

The first condition refers to the necessity of including the opportunity locations on the network model and including them in the accessibility measure. Without the target locations in the random walk model, the output only results in the number of diverse nodes reached, without any land use information, which provides limited information into accessibility. The second condition refers to the a priori adjustments to the network model for spatial equity assessment, such as the inclusion or exclusion of zero population nodes. Third, the accessibility measurement based on the network model needs to consider the travel behavior and preferences, to include the external socioeconomic factors, defined in the accessibility conceptualization (see 4.2). Last condition refers to the temporal analysis. For an accurate assessment within each timestep, each collected dataset should be specific to the correct timestep. Doing so, a consistent analysis with respect to the transport and land use data is ensured, coherent with the accessibility conceptualization shown in Figure 4.2.

9.2 SCIENTIFIC AND SOCIAL CONTRIBUTIONS

This study has made several scientific contributions. First, with it proposes a new application use case of complex network theory, which is accessibility and equity measurement. Complex network models are readily used for the assessment of transport infrastructure. Yet, most studies using such models are limited to the use cases of topological evolution and its effect on network efficiency. This study developed an accessibility measurement method and an analysis methodology to be used in equity assessment, thus expanding the scope of complex network approaches in the literature. Furthermore, the methodology proposed uses a random walk based approach, which is widely used in fields such as applied physics, yet lacks diverse applications in transport network studies. This approach is able to measure and differentiate the effects of transport and land use on the accessibility. Second, the methodology developed in this study is a suitable approach for temporal assessment of accessibility due to its repeatability. When historical transport and land use data is made available, this method offers a time-consistent analysis of accessibility and equity. Furthermore, the utilized method is quite versatile as different opportunity locations (e.g. other critical services) can be included in the network model using similar approaches. Lastly, the adopted case study investigates the case of accessibility development to schools. Prior to application, this study makes the contribution of identifying the education accessibility factors, making connections with job accessibility through a comparative literature review. Based on this review, the common factors of accessibility to these two different opportunities are identified, making way for future shared applications of job and education accessibility.

In addition to the method driven scientific contributions, this study makes an empirical contribution through the case of school accessibility in the City of Helsinki. Based on the analysis, the spatial accessibility patterns in the region were identified and compared across timesteps. The results were discussed considering the population distribution of subdistricts as well as the historical developments in urban form,

shifting to a polycentric urban form. The decreasing accessibility and equity were the main findings of this case study. Comparisons with network topology and distribution of school locations made it possible to delineate which developments were related to the transport or the land use. Although school capacity information could not be collected for this region, the random walk results offered a means for estimating the capacity of schools based on the network topology. Therefore, schools that are more accessible (i.e. higher number of random walk visits) based on the road network were identified. Future research could combine this approach with other transport network layers to identify the areas where the supply of schools do not match the transport accessibility based demand.

This study explored the concept of equity from the perspective of accessibility, and developed a methodology to make a temporal analysis. The developed methodology offers a novel approach to determine the impact of transport, land use and socioeconomic indicators' impact on spatial accessibility. Thus, this study adds value to the existing assessment methods of accessibility and equity, and could be used to make historical assessments of the impacts of policy changes or infrastructural investments. Another societal contribution of this study is the predictive ability of the method to make future assessments, as compared to the historical nature of this study. A similar methodology for accessibility measurement and equity assessment can be repeated for a predictive analysis based on different policy scenarios. Lastly, this study contributes to the literature regarding accessibility to education opportunities. The importance of accessibility to mandatory education for the continuation of studies and employment were identified in this study [Dickerson and McIntosh, 2013]. In addition, the equity of education provision is stated as a sustainable development goal by United Nations [2015]. This study strives to contribute to this goal by developing a method for accessibility and equity measurement that is applicable in a range of case studies.

9.3 LIMITATIONS AND FUTURE RESEARCH

This study has numerous limitations that need to be addresses with regard to its adopted approach and developed methodology.

The first limitation is regarding the selected transport network of this study. The transport network used in this study is the road network, which is a complete network for personal cars and some pedestrian movement. However, the city also has a public transport infrastructure that needs to be considered for a realistic assessment of accessibility. In the case study, this fact was mitigated by the focus on walking accessibility. However, for other analyses using longer travel distances, the public transport options would need to be integrated to the model. The case study analysis focused on accessibility with parameters based on walking as a travel mode. Yet, the collected road network only includes a car-based road network, therefore the pedestrian walking paths are not included. The assumption that pedestrians use only car-oriented roads could be valid in a car-dominated city, however for the case of Helsinki, the urban core is regarded as the pedestrian zone with several pedestrian

streets. Therefore, the network driven accessibility measure requires an ad-hoc data collection with regard to the selected travel modes in the study.

The second limitation relates to the random walk approach. This approach is a good indicator of how network topology influences the reach to diverse locations when limited information is available in terms of actual travel paths. One shortcoming of this method is embedded in its theoretical basis, as the completely random behavior does not represent real life mobility patterns. Therefore, it cannot be used to simulate human behavior to a great extent, but it is still useful as a measure of network topology.

Third, from an accessibility measurement perspective, the applied methodology can be regarded as a location-based measure due to the inclusion of school locations and population based residential zone information. However, not all parts of the accessibility measurement factors are complete, such as the lack of consideration of attractiveness of opportunities and competition factors. A count of school visits does not consider the full picture with respect to the capacity of schools, or the attractiveness of certain schools based on their facilities. While competition effects are less considered in critical service accessibility, it could still be valid where opportunities are not perfect substitutes for each other. Furthermore, this study included data of primary, secondary and high schools. However, a more detailed analysis could have been done with each type of school and compare with the population characteristics of target age groups.

The fourth limitation is regarding the chosen indices in the equity assessment part of the analysis. Spatial equity was measured using Spatial Gini and Theil indices, which made it possible to capture the contribution of spatial relationships in the measured equity. While these indices are commonly used in literature for this purpose, in the case of Theil index, the measure is not designed for the accessibility metric of this study, due to the abundance of zero values. The logarithmic formulation fails to indicate the exact contribution of such nodes due to mathematical constraints.

This study makes way to future research directions, specifically with regard to the use of random-walk driven methodology in accessibility assessment of complex network models. Based on the limitations discussed in the previous section, this research can be complemented in several ways. This study focused on the short distance accessibility based on walkability of the network, a future direction could be to expand this model to longer distance thresholds and different travel modes such as cycling, car, and public transport. For the public transport case, a multilayer network approach can be implemented to include historical development of public transport networks.

As a similar study focusing on the walking case, the effect of pedestrian paths and no-car zones can be investigated with respect to accessibility to opportunities. Based on this, similar approaches can be taken to assess the effect of specific urban transport policies on accessibility using the network model, such as changes in speed limits.

Another future direction would be the use of this methodology in different accessibility contexts. This study readily proposed a conceptualization of common factors between education and job accessibility. The job accessibility direction can also be analyzed. The difference in these approaches would be the type of location data to

be collected, as a comprehensive location of distinct job opportunities are often not available. Alternatively, subdistrict based data collected in this study includes the number of jobs per subdistrict. Using the method of transforming the population data to the complex network model, Voronoi tessellations can be used for a comparable transformation to the network model. One significant factor to consider would be the effect of spatial distribution of network nodes on the job location transformation.

Lastly, the concept of equity can be revisited to design a tailor-made equity measurement framework for network theory driven accessibility measures. The developed accessibility metric in this study lacked compatibility with some indices such as the Theil index, therefore another approach to territorial equity can be considered for future studies.

9.4 COSEM RELEVANCE

This research contributes to the Complex Systems Engineering and Management (CoSEM) programme through several factors. First, this study investigates a problem that has a societal relevance, as the main premise is to evaluate and explain the evolution of accessibility-based equity over time. While doing so, sociotechnical systems in an urban case study are analyzed. As discussed in the literature review regarding accessibility factors, the interplay between transport and land use determine the development of accessibility in a urban system. Connecting to this, the second connection of this research with CoSEM programme is based on the investigation of the development of the complex interconnected systems. It is aimed to understand the evolution of transport systems and socioeconomic development in an urban area. Therefore, both the social and technical components of the problem are to be researched. Third, this research has a multidisciplinary scope, bridging urban studies and transport systems studies. Fourth, this research utilizes complex network analysis to approach the defined problem, which is an analysis method embedded in the CoSEM programme.

BIBLIOGRAPHY

(2022). PxWeb - Tilastokeskus.

Andersson, E., Malmberg, B., and Östh, J. (2012). Travel-to-school distances in Sweden 2000-2006: Changing school geography with equality implications. *Journal of Transport Geography*, 23:35–43.

Banister, D. (2012). Assessing the reality-Transport and land use planning to achieve sustainability. *Journal of Transport and Land Use*, 5(3):1–14.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(1-3):1–101.

Barthélemy, M. and Flammini, A. (2009). Co-evolution of density and topology in a simple model of city formation. *Networks and Spatial Economics*, 9(3):401–425.

Bastiaanssen, J., Johnson, D., and Lucas, K. (2022). Does better job accessibility help people gain employment? The role of public transport in Great Britain. *Urban Studies*, 59(2):301–322.

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete Choice Methods and Their Applications to Short Term Travel Decisions. In *Handbook of Transportation Science*, number 1985, pages 5–33.

Bertolini, L. (2012). disP - The Planning Review Integrating Mobility and Urban Development Agendas: a Manifesto. *The Planning Review*, 48(October):16–26.

Bibri, S. E. (2018). Systems Thinking and Complexity Science and the Relevance of Big Data Analytics, Intelligence Functions, and Simulation Models. In *Smart Sustainable Cities of the Future*, chapter 6, pages 297–369. Springer International Publishing.

Brussel, M., Zuidgeest, M., van den Bosch, F., Munshi, T., and van Maarseveen, M. (2019). *Can the urban poor reach their jobs? Evaluating equity effects of relocation and public transport projects in Ahmedabad, India*. Elsevier Inc.

Camporeale, R., Caggiani, L., Fonzone, A., and Ottomanelli, M. (2019). Study of the accessibility inequalities of cordon-based pricing strategies using a multimodal Theil index. *Transportation Planning and Technology*, 42(5):498–514.

Casali, Y. and Heinimann, H. R. (2019). A topological analysis of growth in the Zurich road network. *Computers, Environment and Urban Systems*, 75(March):244–253.

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

City of Helsinki (2020). Walking — City of Helsinki.

City of Helsinki (2021). Helsinki Region Infoshare.

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., and Sheikh, A. (2011). The case study approach. *BMC Medical Research Methodology*, 11(1):100.

Crucitti, P., Latora, V., and Porta, S. (2006). Centrality in networks of urban streets. *Chaos*, 16(1).

Cui, B., Boisjoly, G., El-Geneidy, A., and Levinson, D. (2019). Accessibility and the journey to work through the lens of equity. *Journal of Transport Geography*, 74(December 2018):269–277.

Curl, A. (2018). The importance of understanding perceptions of accessibility when addressing transport equity: A case study in Greater Nottingham, UK. *Journal of Transport and Land Use*, 11(1):1147–1162.

Dawson, M. R. (2003). Advantages and Disadvantages of Modeling. In *Minds and Machines: Connectionism and Psychological Modeling*, chapter 2, pages 5–25. Wiley-Blackwell.

Delbosc, A. and Currie, G. (2011). Using Lorenz curves to assess public transport equity. *Journal of Transport Geography*, 19(6):1252–1259.

Di Paolo, A., Matas, A., and Raymond, J. L. (2017). Job accessibility and job-education mismatch in the metropolitan area of Barcelona. *Papers in Regional Science*, 96:S91–S112.

Dickerson, A. and McIntosh, S. (2013). The Impact of Distance to Nearest Education Institution on the Post-compulsory Education Participation Decision. *Urban Studies*, 50(4):742–758.

Ding, R. (2019). The complex network theory-based urban land-use and transport interaction studies. *Complexity*, 2019.

Ding, R., Ujang, N., Hamid, H. B., Manan, M. S. A., Li, R., Albadareen, S. S. M., Nochian, A., and Wu, J. (2019). *Application of Complex Networks Theory in Urban Traffic Network Researches*, volume 19. Networks and Spatial Economics.

Ding, R., Zhang, Y., Zhang, T., and Ma, C. (2021). Development of a Complex Network-Based Integrated Multilayer Urban Growth and Optimisation Model for an Efficient Urban Traffic Network. *Complexity*, 2021.

Dixit, M. and Sivakumar, A. (2020). Capturing the impact of individual characteristics on transport accessibility and equity analysis. *Transportation Research Part D: Transport and Environment*, 87(July):102473.

El-Geneidy, A., Levinson, D., Diab, E., Boisjoly, G., Verbich, D., and Loong, C. (2016). The cost of equity: Assessing transit accessibility and social disparity using total travel cost. *Transportation Research Part A: Policy and Practice*, 91:302–316.

Finnish Transport Agency (2018). *National Travel Survey 2016*.

Foth, N., Manaugh, K., and El-Geneidy, A. M. (2013). Towards equitable transit: Examining transit accessibility and social need in Toronto, Canada, 1996-2006. *Journal of Transport Geography*, 29:1–10.

Gallotti, R., Louail, T., Louf, R., and Barthelemy, M. (2015). Big data: a new perspective on cities. *Big Data over Networks*, pages 247–277.

Garrison, W. L. and Marble, D. F. (1962). The structure of transportation networks. Technical report.

Geurs, K. T. and Ritsema van Eck, J. R. (2003). Evaluation of accessibility impacts of land-use scenarios: The implications of job competition, land-use, and infrastructure developments for the Netherlands. *Environment and Planning B: Planning and Design*, 30(1):69–87.

Geurs, K. T. and van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport Geography*, 12(2):127–140.

Guzman, L. A., Oviedo, D., and Rivera, C. (2017). Assessing equity in transport accessibility to work and study: The Bogotá region. *Journal of Transport Geography*, 58:236–246.

Hanna, S. (2020). Random walks in urban graphs: A minimal model of movement. *EPB: Urban Analytics and City Science*, 48(6):1967–1711.

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, 25(2):73–76.

Harrison, H., Birks, M., Franklin, R., and Mills, J. (2017). Case study research: Foundations and methodological orientations. *Forum Qualitative Sozialforschung*, 18(1).

Helsingin Kaupunki (2016). HLT 16 - kaupunkijulkaisu. Technical report, City of Helsinki, Helsinki.

Hillier, B., Penn, A., Hanson, J., Grajewski, T., and Xu, J. (1993). Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B: Planning and Design*, 20(1):29–66.

Järv, O., Tenkanen, H., Salonen, M., Ahas, R., and Toivonen, T. (2018). Dynamic cities: Location-based accessibility modelling as a function of time. *Applied Geography*, 95(November 2017):101–110.

Jin, J. and Paulsen, K. (2018). Does accessibility matter? Understanding the effect of job accessibility on labour market outcomes. *Urban Studies*, 55(1):91–115.

Kain, J. F. (1992). The Spatial Mismatch Hypothesis: Three Decades Later. *Housing Policy Debate*, 3(2):371–460.

Kalinic, M. and Krisp, J. M. (2018). Kernel density sstimation (KDE) vs. hot-spot analysis - detecting criminal hot spots in the city of San Francisco. *21st International Conference on Geographic Information Science (AGILE)*, (June):1–5.

Kansky, K. J. (1963). Structure of Transportation Networks: Relationships between Network Geometry and Regional Characteristics.

Kasraian, D., Maat, K., Stead, D., and van Wee, B. (2016). Long-term impacts of transport infrastructure networks on land-use change: an international review of empirical studies. *Transport Reviews*, 36(6):772–792.

Kawabata, M. and Shen, Q. (2007). Commuting inequality between cars and public transit: The case of the San Francisco Bay Area, 1990-2000. *Urban Studies*, 44(9):1759–1780.

Koulurekisteri (2020). Koulurekisteri.

Lee, M. and Kim, J. B. (2021). Spatial distribution of access diversity on urban road networks. *Journal of the Korean Physical Society*, 79:504–511.

Leskovec, J., Kleinberg, J., and Faloutsos, C. (2005). Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. Technical report.

Levinson, D. (2008). Density and dispersion: The co-development of land use and rail in London. *Journal of Economic Geography*, 8(1):55–77.

Levinson, D., Xie, F., and Zhu, S. (2007). The Co-Evolution of Land Use and Road Networks. *Transportation and Traffic Theory*, (February):1–12.

Levinson, D. M. (1998). Accessibility and the journey to work. *Journal of Transport Geography*, 6(1):11–21.

Levinson, D. M. and Ajay, K. (1994). Operational Evidence of Changing Travel Patterns: A Case Study of Montgomery County, Maryland. *ITE Journal*, April:36–44.

Li, Y., Lin, Y., Geertman, S., Hooimeijer, P., and Xu, W. T. A. (2021). Accessibility-Based Equity of Public Facilities: a Case Study in Xiamen, China. *Applied Spatial Analysis and Policy*, 14(4):947–968.

Liljamo, T., Liimatainen, H., Pöllänen, M., and Viri, R. (2021). The effects of mobility as a service and autonomous vehicles on people's willingness to own a car in the future. *Sustainability (Switzerland)*, 13(4):1–28.

Litman, T. (2018). Evaluating Transport Equity. Technical Report 2.

Liu, X., Dai, L., and Derudder, B. (2017). Spatial Inequality in the Southeast Asian Intercity Transport Network. *Geographical Review*, 107(2):317–335.

Liu, Z., Zhang, M., and Liu, L. (2021). Benchmark of the trends of spatial inequality in world megaregions. *Sustainability (Switzerland)*, 13(11).

Logan, T. M., Anderson, M. J., Williams, T. G., and Conrow, L. (2021). Measuring inequalities in urban systems: An approach for evaluating the distribution of amenities and burdens. *Computers, Environment and Urban Systems*, 86(January):101590.

Lucas, K., van Wee, B., and Maat, K. (2016). A method to evaluate equitable accessibility: combining ethical theories and accessibility-based approaches. *Transportation*, 43(3):473–490.

Mäki-Opas, T. E., Borodulin, K., Valkeinen, H., Stenholm, S., Kunst, A. E., Abel, T., Härkänen, T., Kopperoinen, L., Itkonen, P., Prättälä, R., Karvonen, S., and Koskinen, S. (2016). The contribution of travel-related urban zones, cycling and pedestrian networks and green space to commuting physical activity among adults - A cross-sectional population-based study using geographical information systems. *BMC Public Health*, 16(1).

Martens, K., Bastiaanssen, J., and Lucas, K. (2019). *Measuring transport equity: Key components, framings and metrics*. Elsevier Inc.

Mata, S. (2020). Complex Networks : a Mini-review. pages 658–672.

Mei, D., Xiu, C., Feng, X., and Wei, Y. (2019). Study of the school-residence spatial relationship and the characteristics of travel-to-school distance in Shenyang. *Sustainability (Switzerland)*, 11(16):1–15.

METREX (2020). Helsinki City Boulevards Strategy and Projects. Technical Report December.

Morales, J., Flacke, J., Morales, J., and Zevenbergen, J. (2019). Mapping Urban Accessibility in Data Scarce Contexts Using Space Syntax and Location-Based Methods. *Applied Spatial Analysis and Policy*, 12(2):205–228.

Nevanlinna, A. K. (2016). Ports, industries and economics in the urban history of Helsinki.

Panzera, D. and Postiglione, P. (2020). Measuring the Spatial Dimension of Regional Inequality: An Approach Based on the Gini Correlation Measure. *Social Indicators Research*, 148(2):379–394.

Porta, S., Crucitti, P., and Latora, V. (2006). The network analysis of urban streets: A primal approach. *Environment and Planning B: Planning and Design*, 33(5):705–725.

Porta, S., Latora, V., Wang, F., Rueda, S., Strano, E., Scellato, S., Cardillo, A., Belli, E., Càrdenas, F., Cormenzana, B., and Latora, L. (2012). Street Centrality and the Location of Economic Activities in Barcelona. *Urban Studies*, 49(7):1471–1488.

Rey, S. J. and Smith, R. J. (2013). A spatial decomposition of the Gini coefficient. *Letters in Spatial and Resource Sciences*, 6(2):55–70.

Sá, C., Florax, R. J., and Rietveld, P. (2006). L'accessibilité à l'enseignement supérieur importe-t-elle? Comportement du choix des diplômés de l'enseignement secondaire au pays-bas. *Spatial Economic Analysis*, 1(2):155–174.

Salonen, M. and Toivonen, T. (2013). Modelling travel time in urban networks: Comparable measures for private car and public transport. *Journal of Transport Geography*, 31:143–153.

Sharma, G. and Patil, G. R. (2022). Spatial and social inequities for educational services accessibility - A case study for schools in Greater Mumbai. *Cities*, 122(January 2021):103543.

Shen, Q. (1998). Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers. *Environment and Planning B: Planning and Design*, 25(1):345–365.

Shen, Q. (2001). A spatial analysis of job openings and access in a U.S. metropolitan area. *Journal of the American Planning Association*, 67(1):53–68.

Söderström, P., Schulman, H., and Ristimäki, M. (2015). *Urban Form in the Helsinki and Stockholm City Regions*.

Steadman, P. (2004). Developments in Space Syntax. *Environment and Planning B: Planning and Design*, 31(4):483–486.

Strano, E., Nicosia, V., Latora, V., Porta, S., and Barthélemy, M. (2012). Elementary processes governing the evolution of road networks. *Scientific Reports*, 2:1–8.

Sun, Z. and Zacharias, J. (2020). Transport equity as relative accessibility in a megacity: Beijing. *Transport Policy*, 92(April):8–19.

Tenkanen, H. and Toivonen, T. (2019). Helsinki Region Travel Time Matrix. Technical report.

Theil, H. (1967). *Economics and information theory*. North-Holland, Amsterdam.

Tóth, G., Wachs, J., Di Clemente, R., Jakobi, [U+FFFD], Ságvári, B., Kertész, J., and Lengyel, B. (2021). Inequality is rising where social network segregation interacts with urban topology. *Nature Communications*, 12(1).

Travençolo, B. A. and da, L. (2008). Accessibility in complex networks. *Physics Letters, Section A: General, Atomic and Solid State Physics*, 373(1):89–95.

UN-DESA (2018). *World Urbanization Prospects*.

United Nations (2015). Education. Technical report.

van Wee, B. and Mouter, N. (2021). *Evaluating transport equity*, volume 7. Elsevier Inc., 1 edition.

Vecchio, G., Tiznado-Aitken, I., and Hurtubia, R. (2020). Transport and equity in Latin America: a critical review of socially oriented accessibility assessments*. *Transport Reviews*, 40(3):354–381.

Venerandi, A., Quattrone, G., and Capra, L. (2018). A scalable method to quantify the relationship between urban form and socio-economic indexes. *EPJ Data Science*, 7(1):1–21.

Volchenkov, D. and Blanchard, P. (2007). City Space Syntax as a Complex Network.

Wang, S., Yu, D., Kwan, M.-p., Zhou, H., and Li, Y. (2017). The Evolution and Growth Patterns of the Road Network in a Medium-Sized Developing City : A Historical Investigation of Changchun , China , from 1912 to 2017.

Watts, D. J. and Strogatz, S. H. (1998). Strogatz - small world network Nature. *Nature*, 393(June):440–442.

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

Xu, Y., Song, W., and Liu, C. (2018). Social-spatial accessibility to urban educational resources under the school district system: A case study of public primary schools in Nanjing, China. *Sustainability (Switzerland)*, 10(7).

Ye, C., Zhu, Y., Yang, J., and Fu, Q. (2018). Spatial equity in accessing secondary education: Evidence from a gravity-based model. *Canadian Geographer*, 62(4):452–469.

Appendices

A | LITERATURE REVIEW
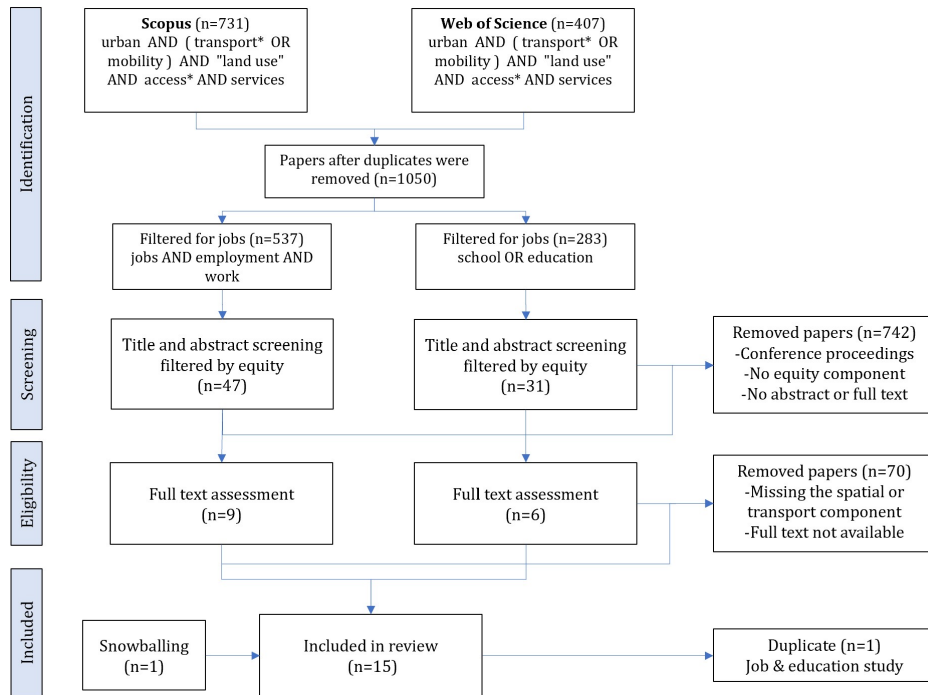
A.1 PRISMA DIAGRAMS

Figure A.1: PRISMA diagram of the structured literature review carried out for chapter 4
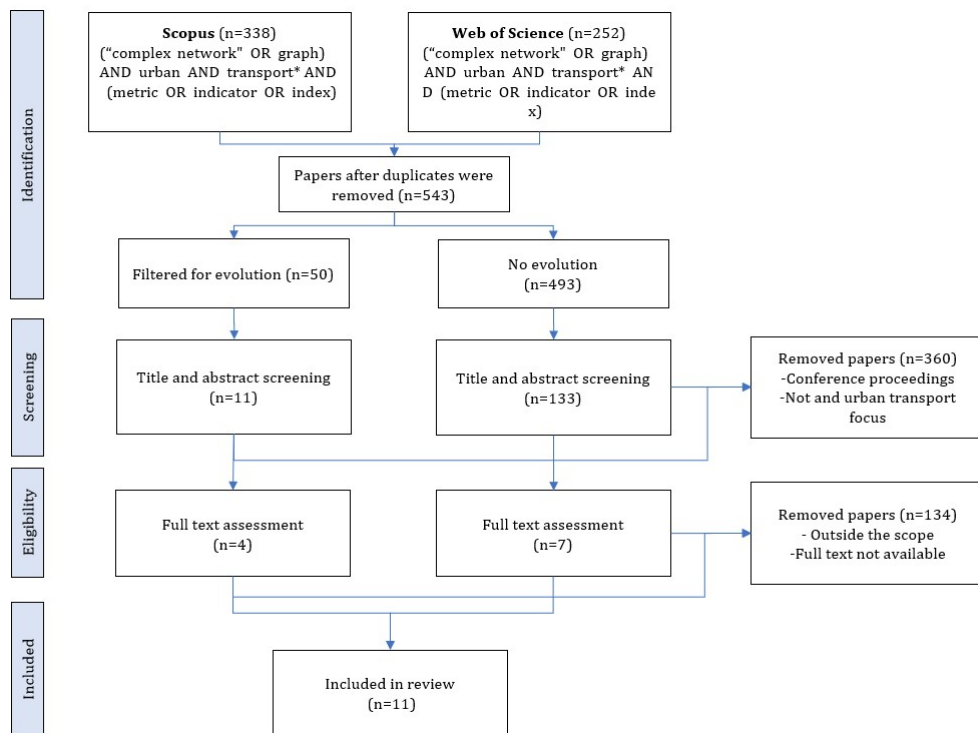
Figure A.2: PRISMA diagram of the structured literature review carried out for chapter 5

B | COMPLEX NETWORK THEORY

B.1 SPACE SYNTAX

In the context of space syntax, *connectivity* (degree in graph theory) of a node can be measured by the count of other nodes directly accessible, as well as the *control value* which the reciprocal of connectivities between neighboring nodes, which measures the extent which a node controls access to adjacent nodes [Hillier et al., 1993]. The integration index, which is the 'shortest journey routes between each link [or space] and all of the others in the network (defining 'shortest' in terms of fewest changes in direction)' [Hillier et al., 1993], is by definition very similar, if not identical, to the closeness centrality approach. Choice is another variable used in SS analysis, which is equivalent to betweenness centrality. Areas with high integration are expected to have denser network topology, whereas high choice values are typically observed in areas which connect neighborhoods to high-hierarchy zones [Morales et al., 2019].
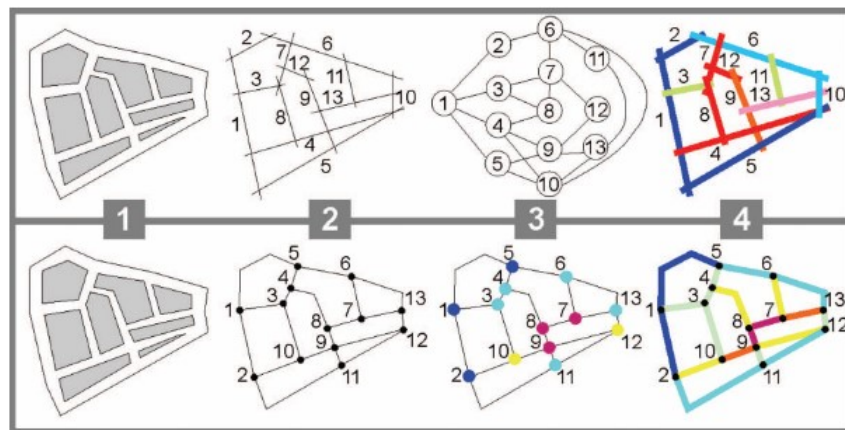
Figure B.1: Graph representation of a street network comparison of space-syntax (top) and MCA (bottom) [Crucitti et al., 2006]

With this particular approach, another contrast emerges regarding the behavioral aspect of urban travel behavior. While the traditional complex network metrics consider shortest paths in terms of metric distances, the axial graphs of SS deal with axial lines. Therefore 'visibility' becomes the main driver of traveler decisions. Furtermore, SS makes the assumption that trip in an urban area (origin destination pairs) are uniformly distributed, and therefore not effected by the land use.

C | CASE STUDY DATA

C.1   ROAD NETWORK

Figure C.1: Helsinki Road Network 1991

Figure C.2: Helsinki Road Network 1999

Figure C.3: Helsinki Road Network 2007

Figure C.4: Helsinki Road Network 2016

C.2 SCHOOLS

Table C.1: School count per subdistrict in Helsinki and the change of the count between timesteps

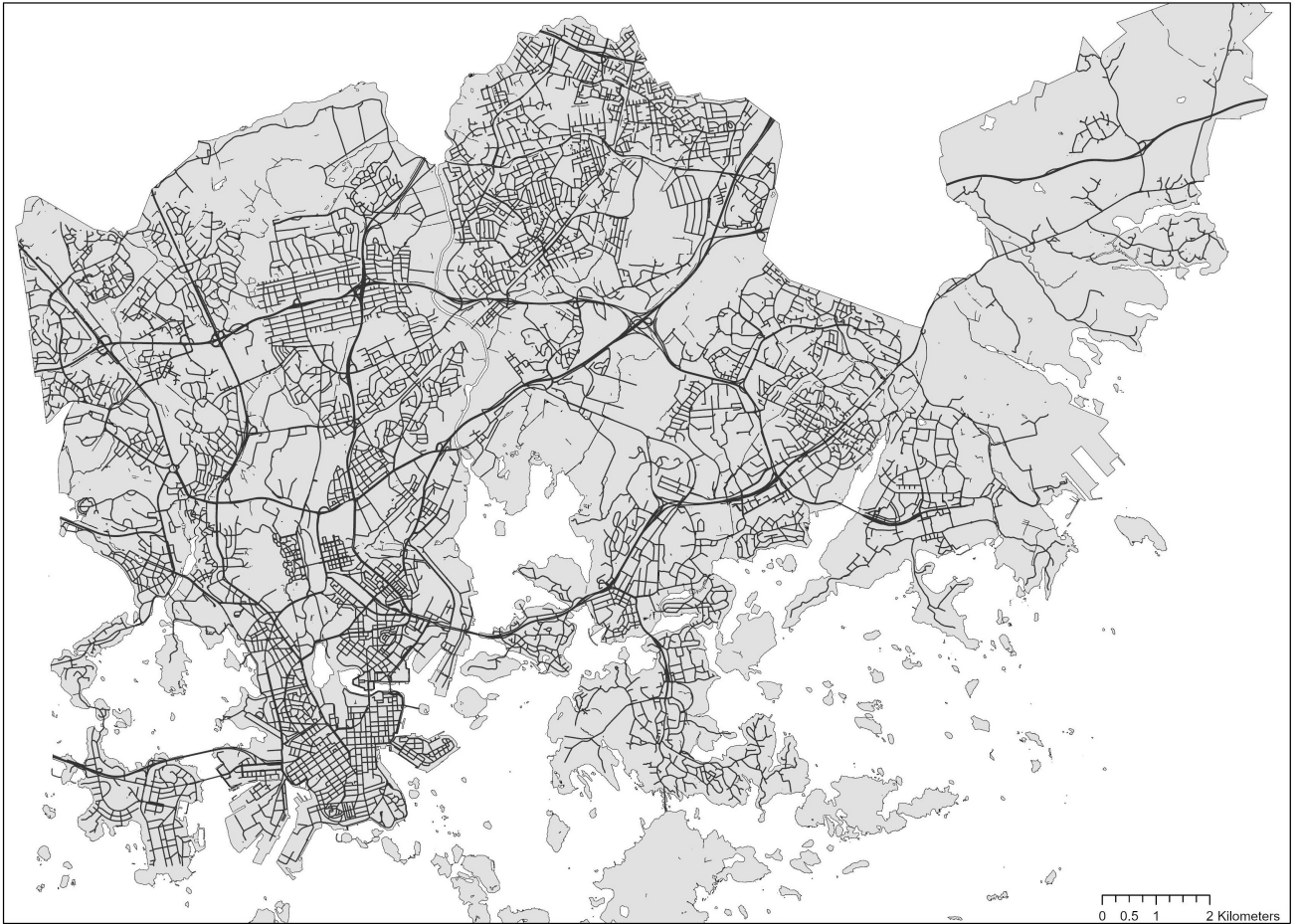| Name of subdistrict	Count of schools				Percentage change of school count (%)		
	1991	1999	2007	2016	1991-1999	1999-2007	2007-2016
Tapanila	7	7	8	8	0.0	14.3	0.0
Harju	2	1	1	1	-50.0	0.0	0.0
Alppila	9	6	6	5	-33.3	0.0	-16.7
Etu-Töölö	11	8	5	4	-27.3	-37.5	-20.0
Puistola	4	4	4	1	0.0	0.0	-75.0
Taka-Töölö	10	10	4	4	0.0	-60.0	0.0
Meilahti	10	8	6	4	-20.0	-25.0	-33.3
Länsi-Pasila	5	5	3	1	0.0	-40.0	-66.7
Itä-Pasila	1	1	0	0	0.0	-100.0	
Laakso	4	1	1	1	-75.0	0.0	0.0
Ruoholahti	0	1	1	1		0.0	0.0
Jätkäsaari	0	1	1	1		0.0	0.0
Vallila	8	5	4	2	-37.5	-20.0	-50.0
Lehtisaari	1	1	1	1	0.0	0.0	0.0
Munkkivuori	3	2	2	2	-33.3	0.0	0.0
Konala	7	3	3	3	-57.1	0.0	0.0
Toukola	0	1	1	1		0.0	0.0
Arabianranta	1	2	2	2	100.0	0.0	0.0
Pohjois-Haaga	4	4	3	3	0.0	-25.0	0.0
Kumpula	1	0	0	0	-100.0		
Käpylä	10	7	3	3	-30.0	-57.1	0.0
Etelä-Haaga	13	10	8	6	-23.1	-20.0	-25.0
Torkkelinmäki	3	4	4	4	33.3	0.0	0.0
Kannelmäki	3	3	2	2	0.0	-33.3	0.0
Maununneva	2	2	1	1	0.0	-50.0	0.0
Veräjälaakso	0	1	1	1		0.0	0.0
Malminkartano	4	4	4	4	0.0	0.0	0.0
Koskela	1	1	1	1	0.0	0.0	0.0
Maunula	3	4	3	3	33.3	-25.0	0.0
Vanha Munkkiniemi	5	5	5	5	0.0	0.0	0.0
Hakuninmaa	5	6	3	3	20.0	-50.0	0.0
Länsi-Pakila	2	2	2	2	0.0	0.0	0.0
Patola	4	3	3	3	-25.0	0.0	0.0
Itä-Pakila	6	3	3	2	-50.0	0.0	-33.3
Veräjämäki	1	0	0	0	-100.0		
Itäkeskus	2	0	0	0	-100.0		
Kruununhaka	4	3	3	3	-25.0	0.0	0.0
Kluuvi	1	1	1	1	0.0	0.0	0.0

Kaartinkaupunki	2	2	2	2	0.0	0.0	0.0
Linjat	3	2	2	2	-33.3	0.0	0.0
Kamppi	11	5	3	3	-54.5	-40.0	0.0
Punavuori	4	6	6	4	50.0	0.0	-33.3
Eira	0	1	1	1		0.0	0.0
Ullanlinna	6	4	5	4	-33.3	25.0	-20.0
Katajanokka	1	1	1	1	0.0	0.0	0.0
Latokartano	0	0	1	1			0.0
Pihlajamäki	5	5	5	5	0.0	0.0	0.0
Pikku Huopalahti	1	0	0	0	-100.0		
Kotkavuori	3	1	1	1	-66.7	0.0	0.0
Vattuniemi	1	0	0	0	-100.0		
Tapaninvainio	1	1	1	1	0.0	0.0	0.0
Siltamäki	1	1	1	1	0.0	0.0	0.0
Vanha Ruskeasuo	5	5	4	4	0.0	-20.0	0.0
Tapulikaupunki	3	3	1	1	0.0	-66.7	0.0
Töyrynummi	3	3	2	2	0.0	-33.3	0.0
Myllykallio	3	1	1	1	-66.7	0.0	0.0
Jakomäki	5	4	4	0	-20.0	0.0	-100.0
Kulosaari	5	6	6	6	20.0	0.0	0.0
Länsi-Herttoniemi	4	4	2	2	0.0	-50.0	0.0
Roihuvuori	4	1	1	1	-75.0	0.0	0.0
Herttoniemen yritysalue	1	0	0	0	-100.0		
Herttoniemenranta	1	1	2	2	0.0	100.0	0.0
Vartioharju	1	1	1	1	0.0	0.0	0.0
Puotila	4	3	3	3	-25.0	0.0	0.0
Puotinharju	2	1	1	0	-50.0	0.0	-100.0
Reimarla	3	1	1	1	-66.7	0.0	0.0
Myllypuro	9	5	5	4	-44.4	0.0	-20.0
Marjaniemi	1	1	0	0	0.0	-100.0	
Pitäjänmäen yritysalue	1	2	3	3	100.0	50.0	0.0
Mellunmäki	4	4	4	4	0.0	0.0	0.0
Kontula	6	5	6	6	-16.7	20.0	0.0
Kivikko	0	1	1	1		0.0	0.0
Hermanninmäki	0	0	1	1			0.0
Yliskylä	3	3	3	1	0.0	0.0	-66.7
Jollas	1	1	1	1	0.0	0.0	0.0
Paloheinä	1	1	1	1	0.0	0.0	0.0
Kyläsaari	1	0	0	0	-100.0		
Tuomarinkartano	1	1	1	0	0.0	0.0	-100.0
Keski-Vuosaari	9	5	4	4	-44.4	-20.0	0.0
Pukinmäki	2	2	2	0	0.0	0.0	-100.0
Ylä-Malmi	4	3	3	0	-25.0	0.0	-100.0
Ala-Malmi	10	10	8	6	0.0	-20.0	-25.0

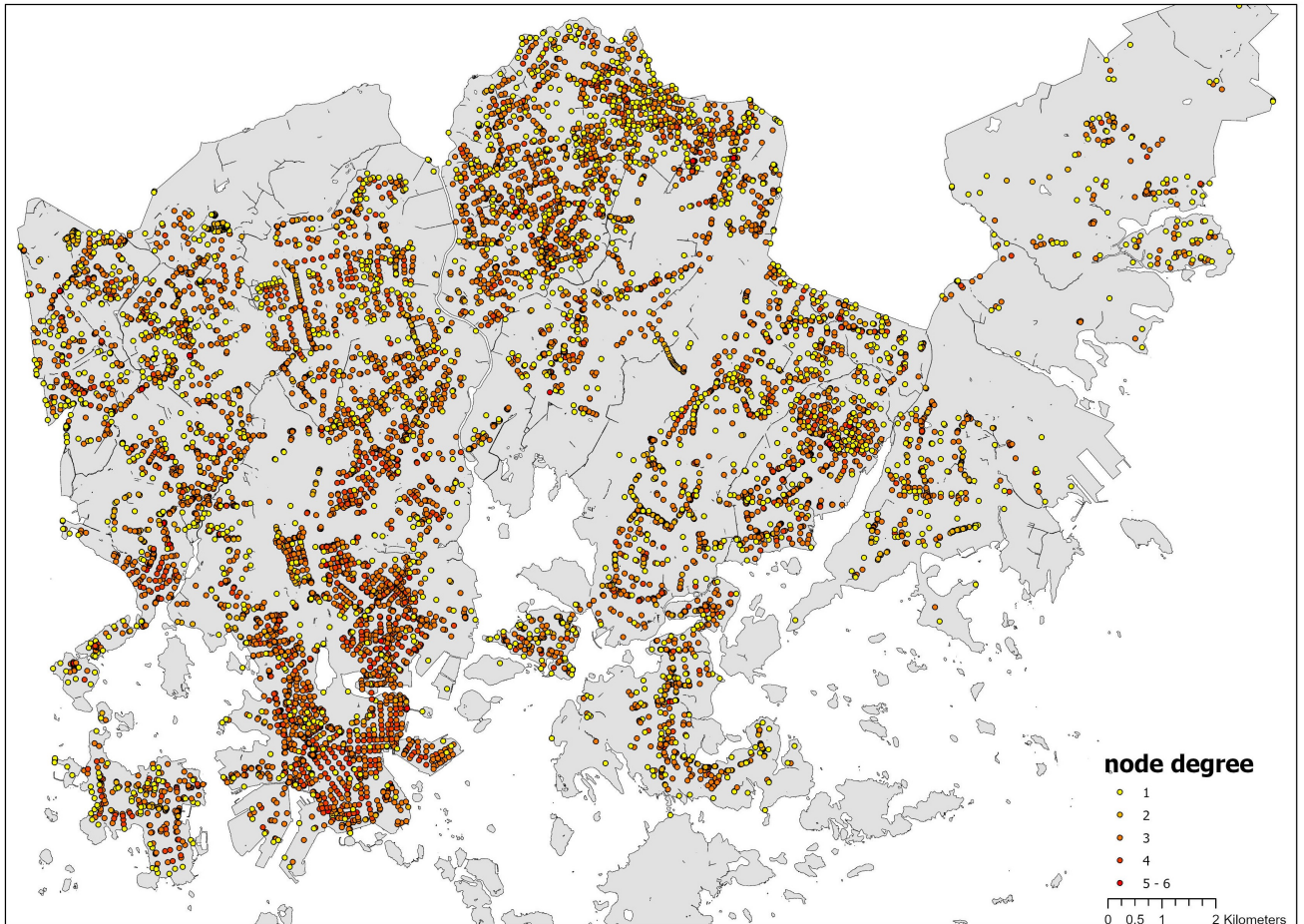# D | EXPLORATORY ANALYSIS

D.1 NODE DEGREE MAP

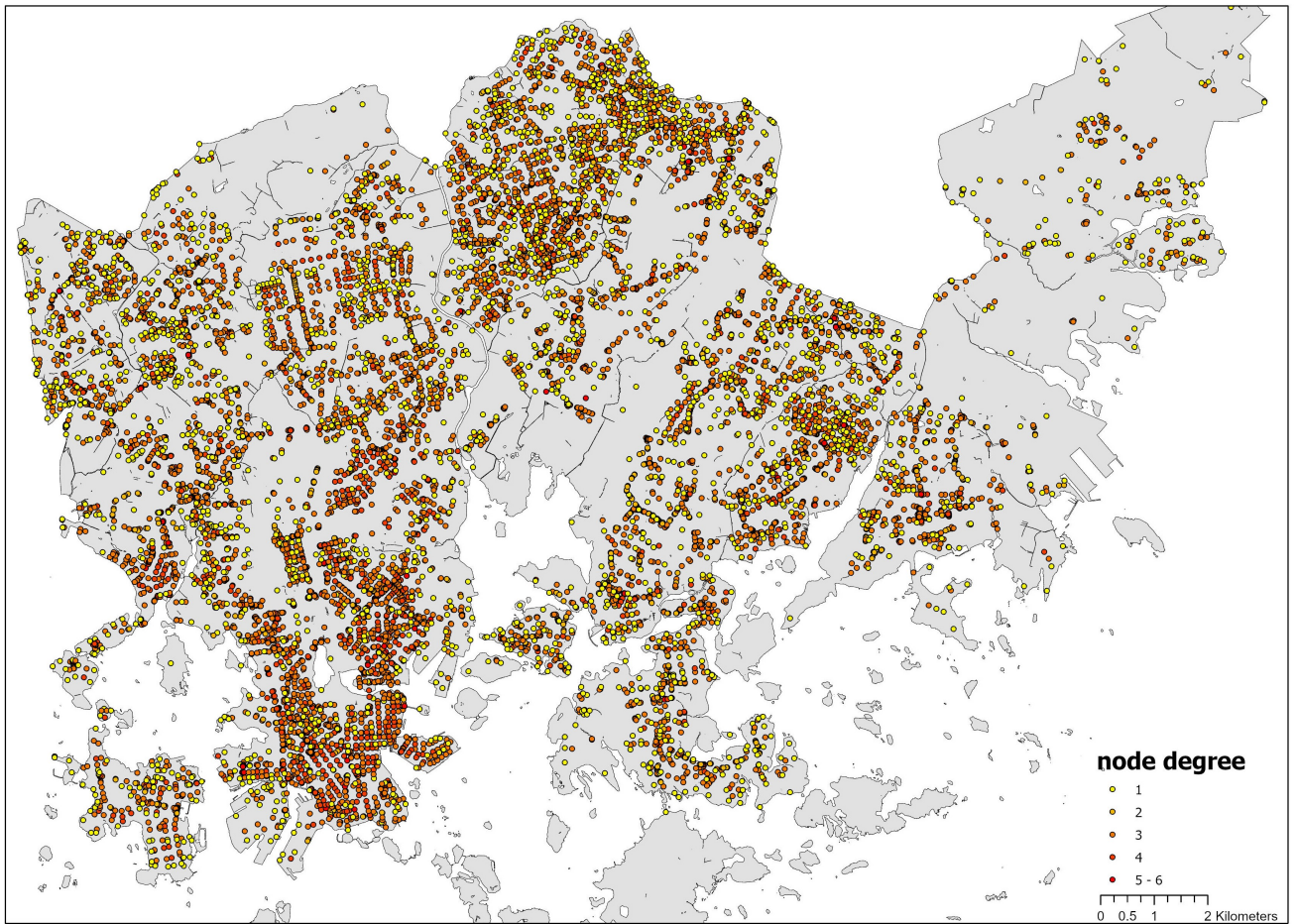**Figure D.1:** Node degree distribution in 1991 network model

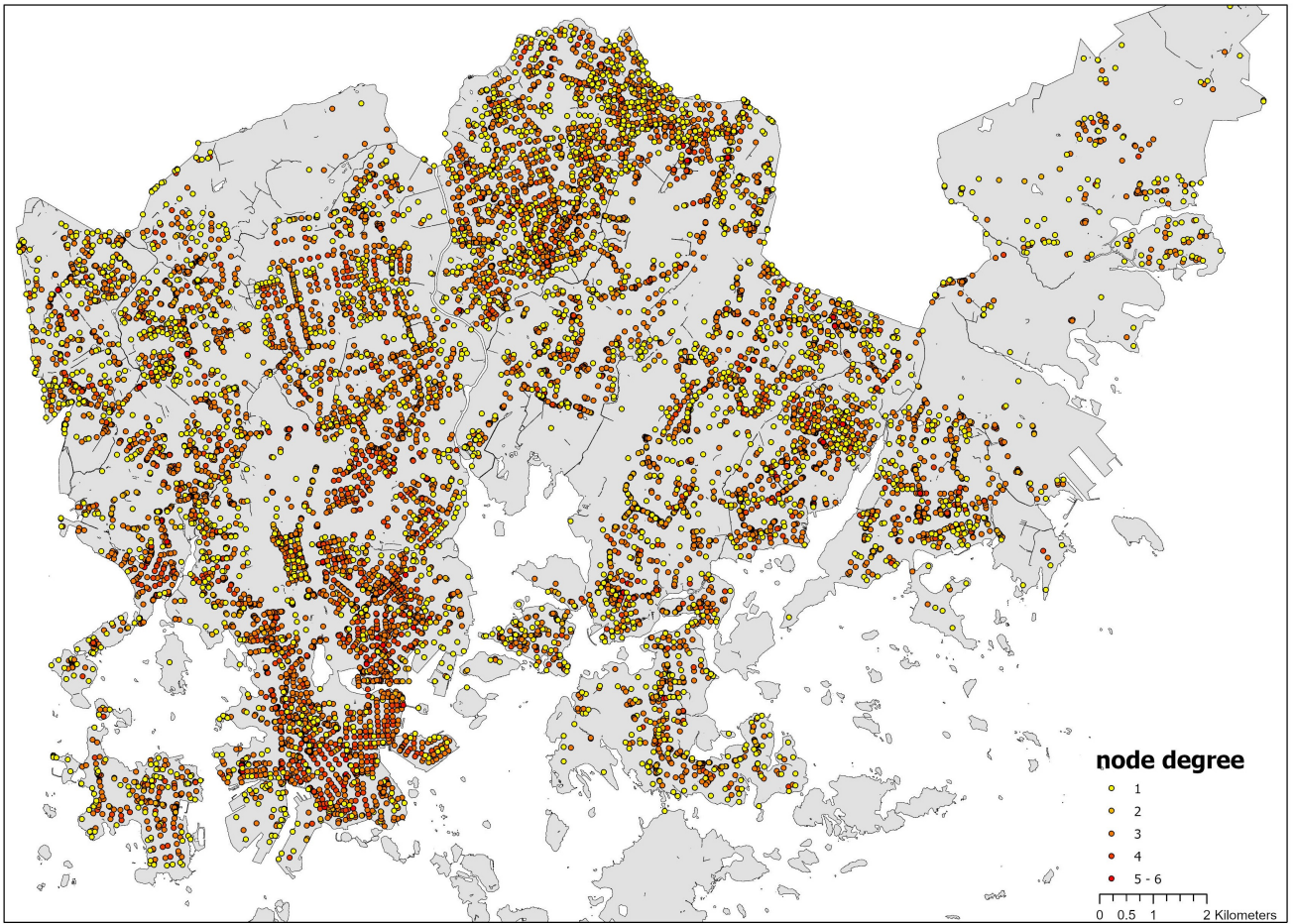**Figure D.2:** Node degree distribution in 1999 network model
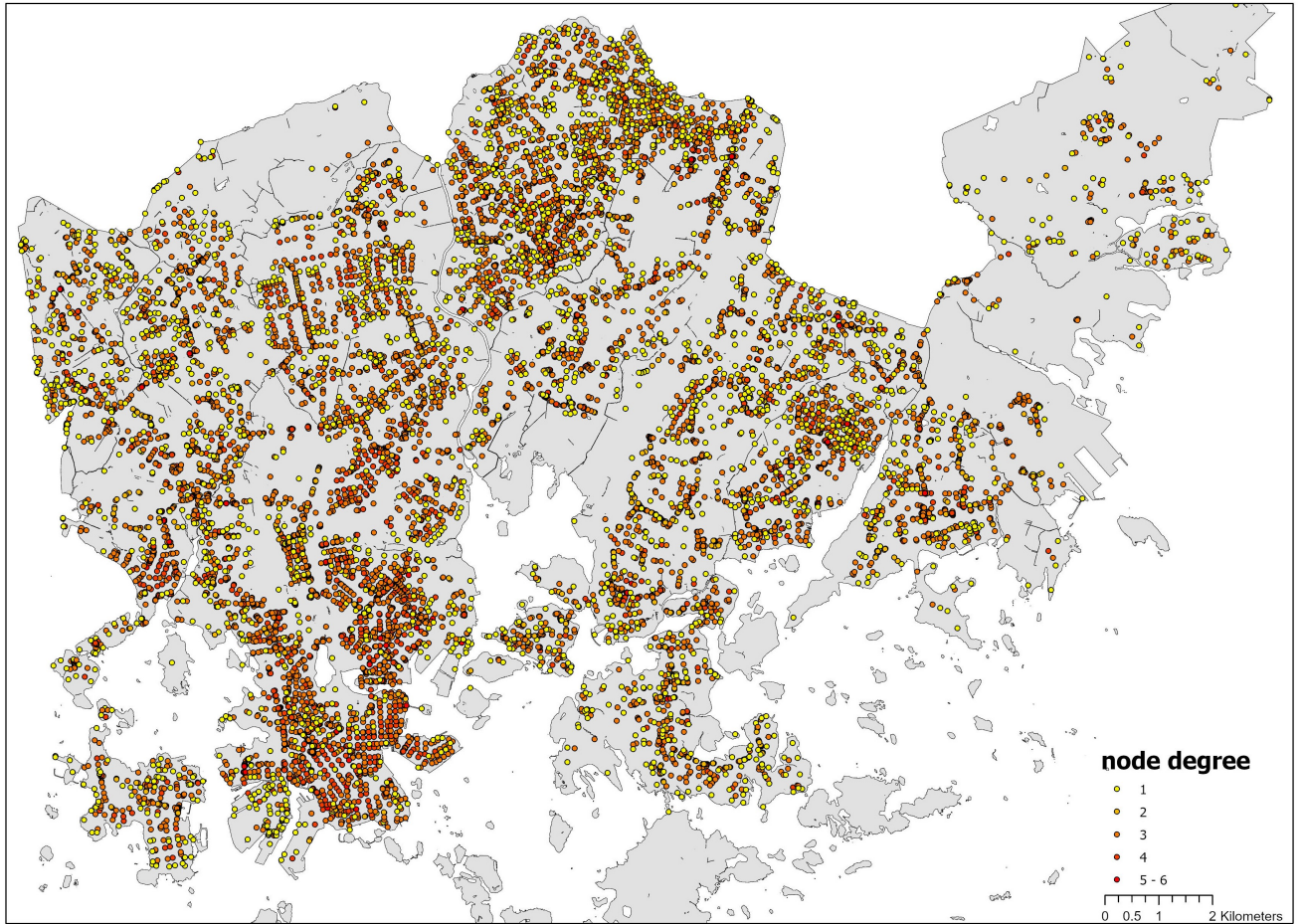
Figure D.3: Node degree distribution in 2007 network model

Figure D.4: Node degree distribution in 2016 network model

E | ANALYSIS OUTPUTS

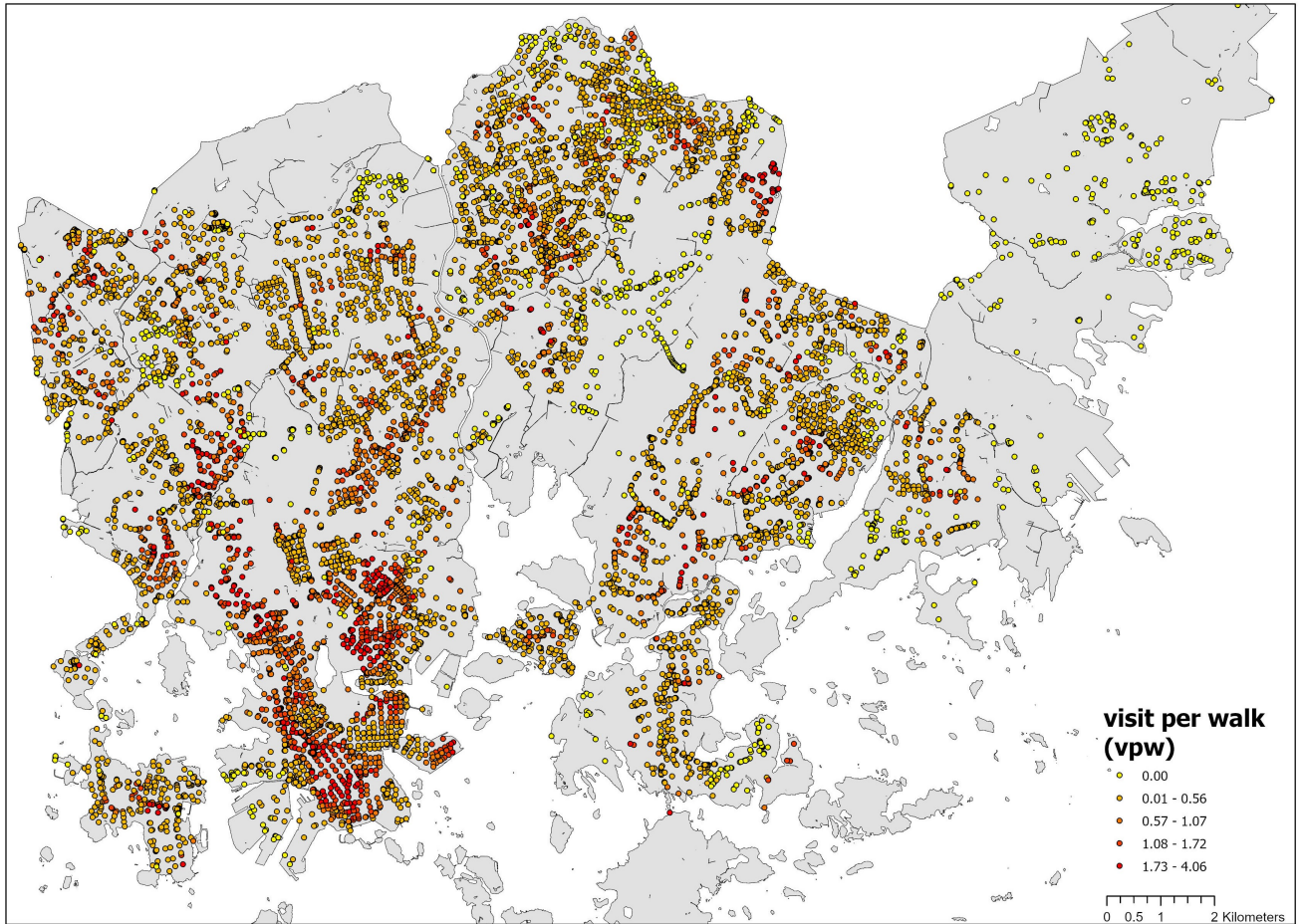## E.1   ACCESSIBILITY MEASUREMENT

Figure E.1: School accessibility results showing visit per walk 1991
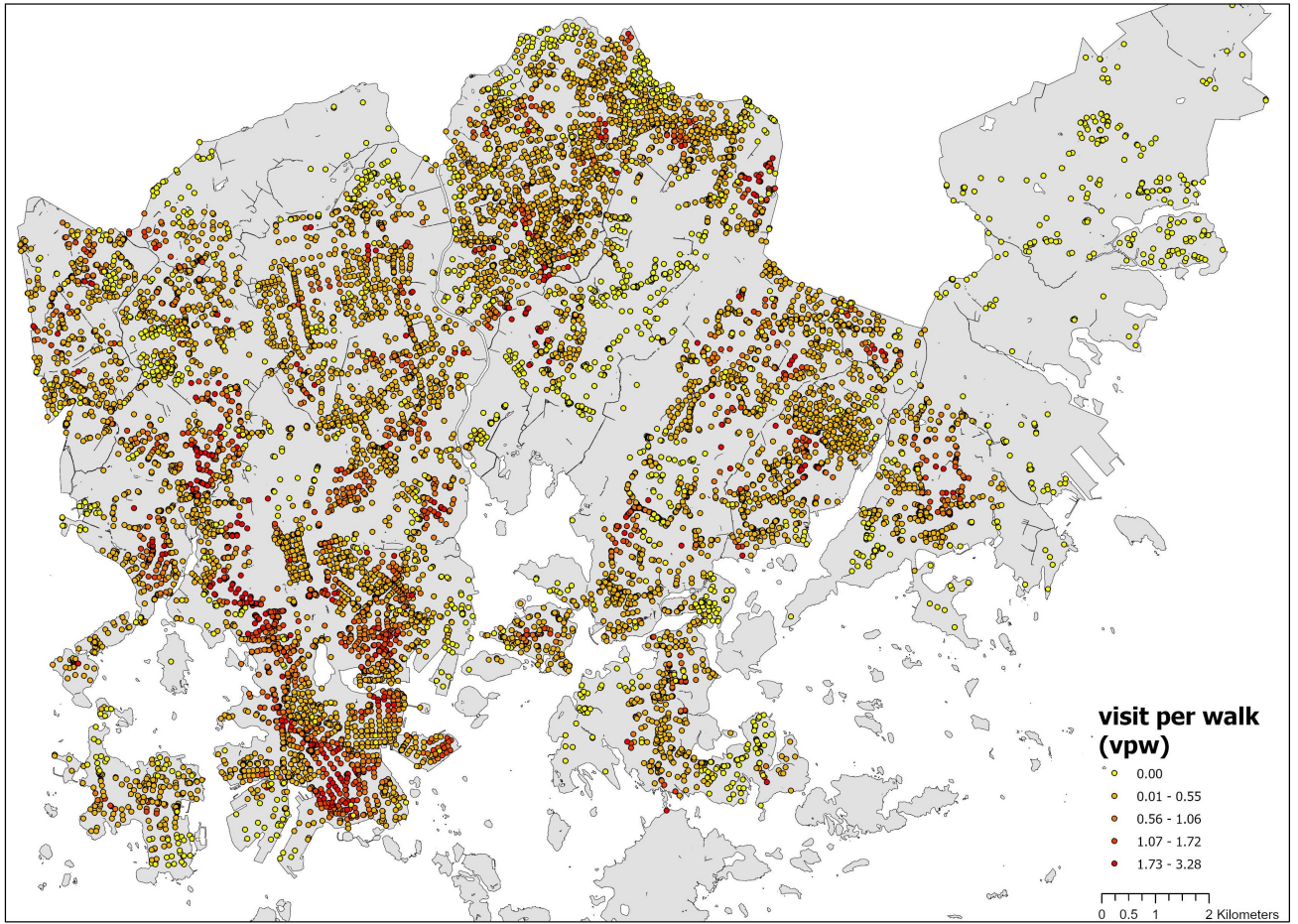
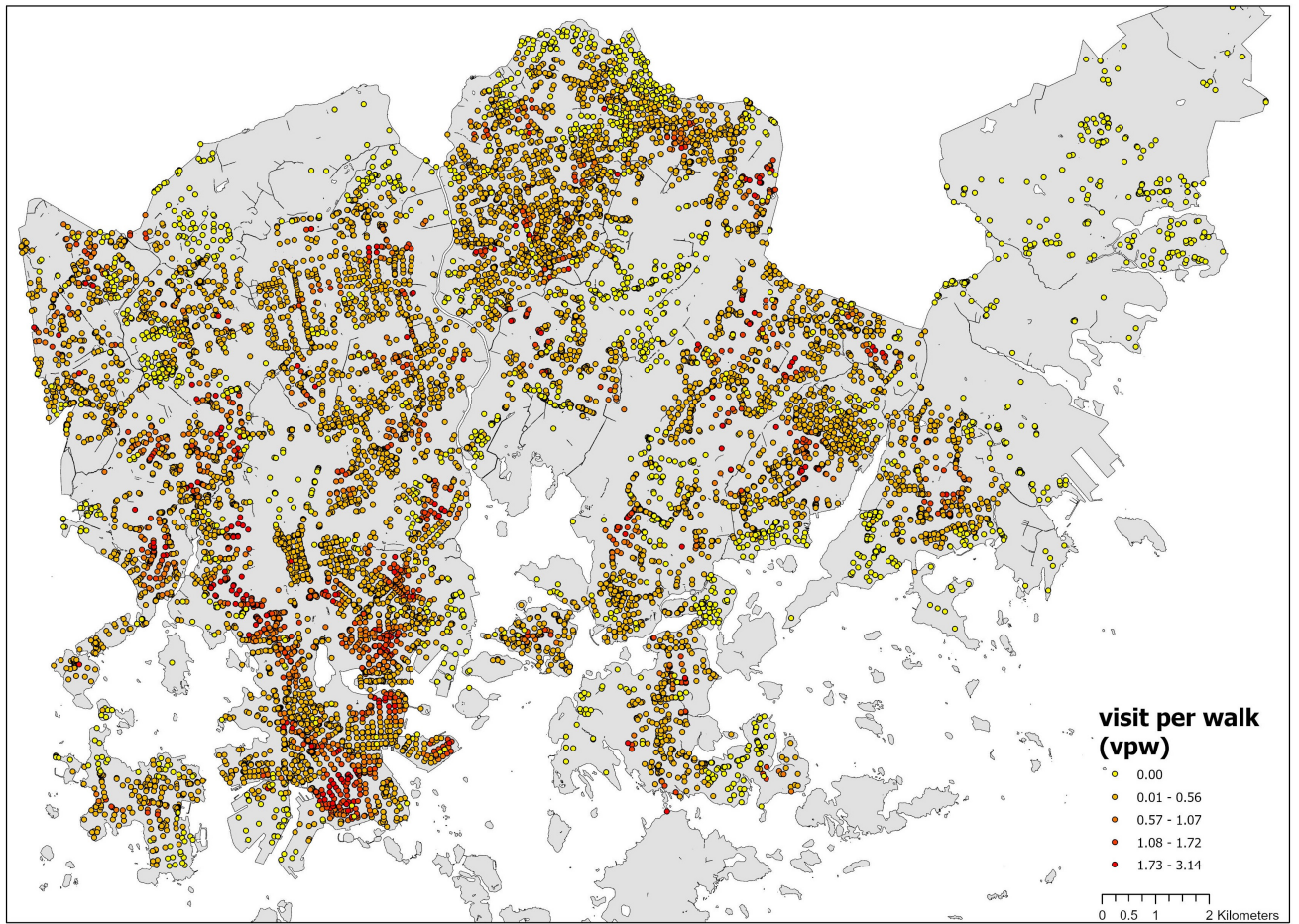**Figure** E.2: School accessibility results showing visit per walk 1999

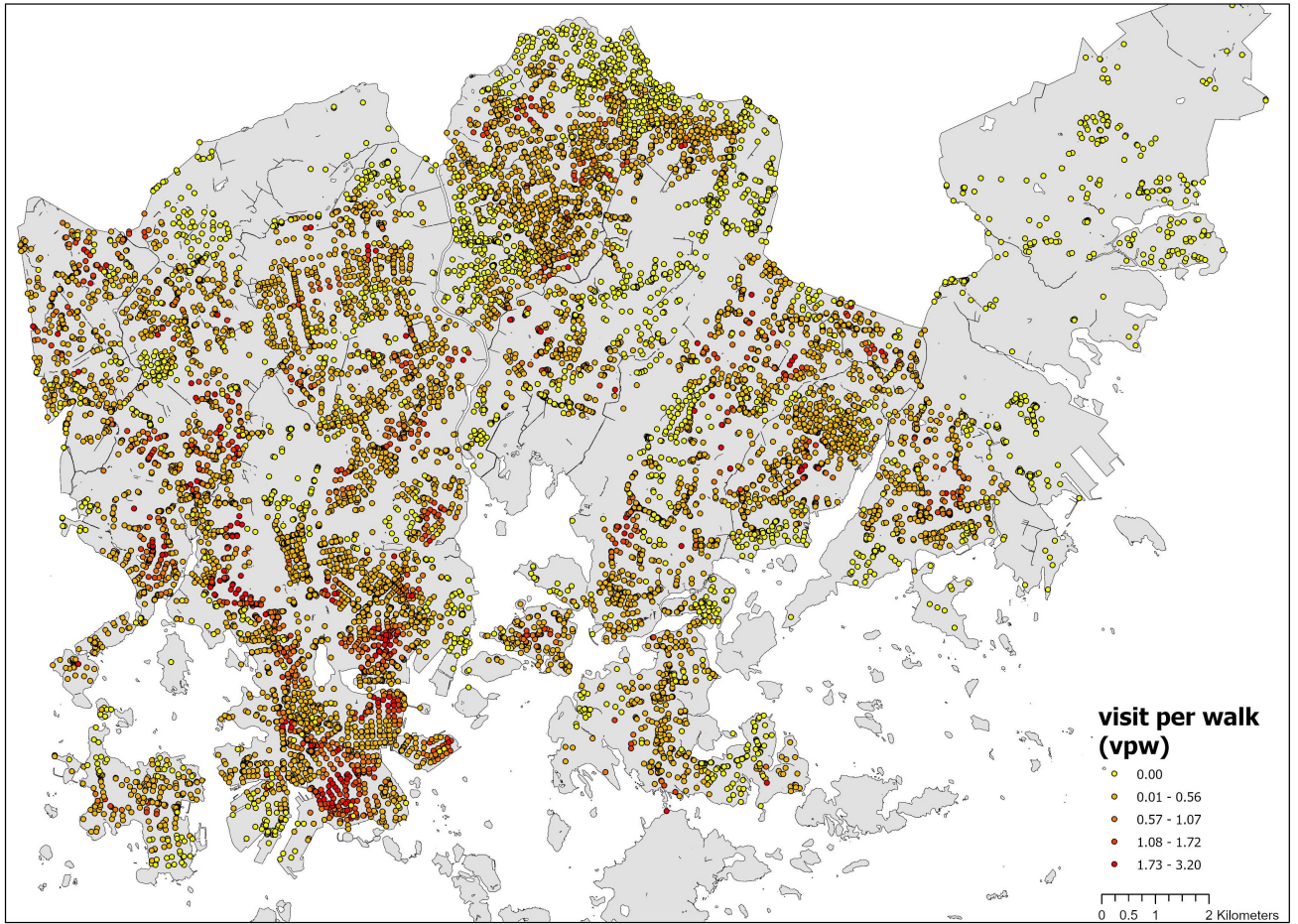Figure E.3: School accessibility results showing visit per walk 2007

Figure E.4: School accessibility results showing visit per walk 2016

Table E.1: Descriptive statistics of self-avoiding random walk output in highest 10 populated subdistricts

		Etelä-Haaga	Etu-Töölö	Kamppi	Kannelmäki	Keski-Vuosaari	Kontula	Linjat	Taka-Töölö	Ullanlinna	Yliskylä
1991	number of nodes	176	177	145	160	130	130	48	168	39	201
	non-school visiting nodes	4	0	0	27	7	0	0	2	0	4
	mean visit per walk	0.89	0.88	0.96	0.29	0.51	0.50	1.65	0.95	1.42	0.31
	median visit per walk	0.86	0.80	0.93	0.24	0.42	0.28	1.68	0.88	1.39	0.15
	standard deviation of visits	0.73	0.55	0.59	0.29	0.50	0.52	0.56	0.35	0.56	0.45
1999	number of nodes	178	176	146	174	176	149	49	175	41	223
	non-school visiting nodes	2	4	2	42	4	6	0	4	2	14
	mean visit per walk	0.81	0.56	0.70	0.23	0.46	0.44	0.96	0.84	1.09	0.31
	median visit per walk	0.63	0.47	0.62	0.13	0.31	0.24	0.90	0.80	1.05	0.13
	standard deviation of visits	0.71	0.49	0.52	0.28	0.45	0.48	0.30	0.39	0.55	0.40
2007	number of nodes	189	172	147	175	199	149	49	175	43	222
	non-school visiting nodes	2	5	2	40	19	6	0	4	2	14
	mean visit per walk	0.53	0.36	0.53	0.19	0.31	0.57	0.95	0.70	1.07	0.32
	median visit per walk	0.36	0.30	0.47	0.08	0.21	0.38	0.89	0.65	1.00	0.14
	standard deviation of visits	0.50	0.38	0.41	0.25	0.34	0.55	0.30	0.33	0.54	0.40
2016	number of nodes	187	174	147	153	194	142	49	175	40	222
	non-school visiting nodes	2	4	4	10	19	6	0	4	2	27
	mean visit per walk	0.53	0.38	0.55	0.23	0.32	0.47	1.05	0.55	1.10	0.11
	median visit per walk	0.37	0.31	0.51	0.11	0.22	0.31	0.90	0.48	1.07	0.05
	standard deviation of visits	0.48	0.37	0.42	0.32	0.36	0.51	0.45	0.35	0.56	0.19

E.2 SPATIAL ANALYSIS

Table E.2: Output statistics of the hotspot analysis for each timestep displaying the numbers of nodes for each hot/coldspot and their share in all nodes

	number of nodes				share of nodes (%)			
year	1991	1999	2007	2016	1991	1999	2007	2016
statistically significant	6435	7234	7956	7932	67%	67%	71%	70%
number of hotspots (99%)	2148	2313	2895	2287	22%	21%	26%	20%
number of hotspots (95%)	315	351	321	447	3%	3%	3%	4%
number of hotspots (90%)	193	232	207	253	2%	2%	2%	2%
number of coldspots (99%)	2699	2699	3499	2733	28%	25%	31%	24%
number of coldspots (95%)	726	1053	717	1600	8%	10%	6%	14%
number of coldspots (90%)	354	586	317	612	4%	5%	3%	5%
total nodes	9574	10856	11222	11330				