Urban Transformation Towards Polycentricity
Detecting Functional Urban Changes in Singapore from Transportation Data

Author(s):
Zhong, Chen

Publication Date:
2014

Permanent Link:
https://doi.org/10.3929/ethz-a-010349714

Rights / License:
In Copyright - Non-Commercial Use Permitted
URBAN TRANSFORMATION TOWARDS POLYCENTRICITY
Detecting Functional Urban Changes in Singapore from Transportation Data

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
CHEN ZHONG

MEng, Wuhan University

born on 8 March 1987
citizen of China

accepted on the recommendation of
Prof. Dr. Gerhard Schmitt, examiner
Prof. Dr. Michael Batty, co-examiner
Prof. Dr. Stefan Müller Arisona, co-examiner

2014
Acknowledgements

I would like to take this opportunity to thank Prof. Schmitt for granting me the chance of this PhD study. Thank him for putting his trust on me, even though several setbacks encountered and my performance was not always that good as expected. Thank Prof. Batty for his guidance. If it is not his affirmation and encouragement, I might be still lost in seeking for my research topics, let alone to learn and exchange ideas with people sharing same research interest. I also want to thank Prof. Müller Arisona. I would rather call him a friend than a teacher. I will always remember the “philosophy” he taught me whenever I go too rush “less is more” and “working hard and working smart”.

I feel lucky that I met Dr. Xianfeng Huang, Dr. Stefan Schlapfer, Dr. Jiaqiu Wang, Dr. Matthias Berger at different stages during my PhD study. Through our cooperation and/or discussions, I learned not only techniques, but also different ways of thinking and the right attitude to do a good research. I also gained great help from Prof. Franz Oswald who gave me the first lesson in Architecture and encourages me all the time; Prof. Rudi Stouffs for giving me the chance to be an teaching assistant; and Prof. Ian Smith for answering all kinds of questions patiently.

Special thanks should be given to all my colleagues at FCL who accompanied me for the past three and half years. Thanks to Gideon and Eva who talk with me every day. We had our “happy hours” in and out of our little cubic. Thanks to Maria and Didier for your sweet birthday gift bear Our Module IX is one team. I would also thank IA team in Zurich. Thank Dani and Lukas for helping me with the German Language. Thanks Denise for her warm help always. Same thanks to my CASA friends. London’s winter is cold but the office is warmed by our friendship.

I also owe many thanks to my family for their continuous support and for forgiving me rarely staying at home over the past years. They are my constant source of power and cheerfulness. And of course I thank Felix, who has unbelievable patience to listen to my endless talks about
“science” and fights together with me until the last day of thesis writing. Finally, I cannot help to thank myself for not giving up though that idea appeared so many times during my PhD study. Persistence makes success a truth.

In the end, this work was established at the Singapore-ETH Centre for Global Environmental Sustainability (SEC), co-funded by the Singapore National Research Foundation (NRF) and ETH Zurich. I would like to express my sincere gratitude to the Singapore Land Transport Authority for supporting this research and providing the required data. Thank to Transportation team in FCL for generously sharing the data, resources, and knowledge with me.
Dedication

To those who have helped me along the way
Abstract

This research seeks for a deeper understanding of urban dynamics. The main idea is to integrate urban planning knowledge with methods from geographic science, resulting in a systematic methodology for urban studies. Specifically, advanced spatial analysis methods are highlighted and applied in a study to detect polycentric urban transformation using transportation data.

This research originates from the observation of a gap between available urban data and the information that could be extracted from such data. Information for a better understanding and management of urban change is in high demand, especially in this age of urban transformation. However, the large urban mobility data that is available and contains such information is insufficiently used due to a lack of analysis methods. To help fill this gap, this research proposes integrated spatial analysis methods capable of measuring the changing spatial structure of urban stocks and flows based on multiple years of transportation data.

Particular interest is given to the phenomenon of polycentric urban transformation, which is an ongoing urban process in Singapore as well as many other cities. The conducted research starts from a review of state-of-the-art studies on Polycentricity. The main argument of this research is that Polycentricity is a matter of how people utilize urban space in reality. In other words, beyond physical urban settings, Polycentricity is an emerging spatial structure of urban stocks and flows in socioeconomic urban space. By assessing original plans with reference to measured spatial structures from urban mobility data, we can help to evaluate urban functionality and planning strategies and uncover urban problems. To achieve such a measurement and assessment, this research presents a generic framework explaining how different levels of data services function in urban studies and planning. This is a general framework and is not limited to the issue of Polycentricity. The core elements of this framework include a geospatial pipeline, integrated spatial analysis methods, and a set of visual analytics tools.

To validate the generic framework and implement the theoretical methodology into practice, a case study of Singapore is conducted. Based on the refined definition of Polycentricity, functional changes in Singapore are emphasized and detected from travel survey data and smart card data from multiple years. The latter data is a newly available large dataset generated by an automatic fare collection system. In particular, statistical analysis is performed to extract travel behaviors at the individual level; urban centrality is measured from aggregated urban activity
patterns by a spatial convolution to identify the spatial structure of urban stocks; and a spa-
tial network model is built as an example of analogy models to identify the spatial origination
of urban flows. In these analyses, sets of urban indices of Polycentricity, such as density, en-
tropy, and centrality, are defined and their measures are bound to the proposed spatial analysis
methods. By applying these measures to data from different years, the path of the functional
changes in Singapore can be traced. By referring to a descriptive analysis of physical develop-
ment in Singapore, the driving forces, impacts, successes, and anomalies of polycentric urban
transformation can be identified.

In sum, this work presents a quantitative approach to urban analysis that explicitly identi-
fies ongoing urban transformation. Specifically, the impact of infrastructure development on
peoples lives and, in return, how cities are reshaped by individuals’ needs are examined using
information extracted from mobility data. The urban studies in this dissertation represents a
way to incorporate human behavior into urban and transport design plans, thus leading to more
livable cities. In a broader sense, it presents a systematic framework that facilitates geospatial
techniques for impact assessment using big urban data in urban studies and planning.
ZUSAMMENFASSUNG


Contents

Acknowledgements i
Dedication iii
Abstract iv
List of Tables xi
List of Figures xii

1 Introduction 1
  1.1 Research Background ............................................. 1
  1.2 Dissertation Outline ............................................. 3

2 Literature Review 4
  2.1 The Evolution of Urban Spatial Structures ......................... 5
    2.1.1 The Polycentric Metropolis ................................... 5
    2.1.2 The Fuzzy Concept of Polycentricity ........................... 10
    2.1.3 Defining Functional Polycentric Spatial Structure .......... 12
  2.2 Spatial Interactions in Urban Dynamics .......................... 15
    2.2.1 Urban Structure Models ...................................... 15
    2.2.2 Operational Models for Urban Processes .................... 17
    2.2.3 Land Use and Transportation Interactions .................. 19
  2.3 Advanced Spatial Analysis for Urban Studies ..................... 21
    2.3.1 Spatial Analysis of Urban Structure ........................ 22
2.3.2 Spatiotemporal Analysis ............................................. 24
2.3.3 Spatially Informed Model for Impact Assessment .......... 25
2.4 New Analysis Methods Using Urban Mobility Data ............ 27
   2.4.1 New Concept of Big Urban Data ............................... 28
   2.4.2 The Use of Urban Mobility Data in Urban Studies ....... 32
2.5 Chapter Conclusions .................................................. 36

3 Research Statement ...................................................... 38
   3.1 Problem Statement and Hypothesis .............................. 39
   3.2 Research Aims ...................................................... 41
   3.3 Method: Spatial Analysis and Modeling ....................... 42

4 Framework for Measuring Functional Polycentricity ............ 44
   4.1 Research Design: An Applied Framework ..................... 45
   4.2 Detecting Functional Urban Changes Using Transportation Data ................................. 47
      4.2.1 Spatial Analysis Methods ................................. 47
      4.2.2 Urban Indices for Quantitative Analysis ............... 48
      4.2.3 A Visual Analytics Framework ............................ 50
   4.3 Chapter Conclusions ............................................... 52

5 Functional Changes in Singapore ................................... 54
   5.1 Study Area and Data .............................................. 56
      5.1.1 Case Study Area: Singapore ............................... 56
      5.1.2 Study Materials ........................................... 56
   5.2 Five Decades Fast Development in Singapore .................. 60
      5.2.1 Historical Urban Plans .................................... 61
      5.2.2 Transport Development .................................... 65
      5.2.3 The Geography of Economic Activities .................. 66
      5.2.4 Discussion .................................................. 71
   5.3 Statistical Analysis of Travel Behavior ....................... 71
      5.3.1 Statistical Analysis of Travel Survey Data .............. 72
      5.3.2 Mining User Travel Behavior from Smart Card Data .... 82
      5.3.3 Inferring Activity Types from Travel Behaviors ....... 86
# List of Tables

4.1 Analyses applied to urban transportation data sets. ......................... 49  
5.1 A sample of household travel survey in Singapore with selected information . 59  
5.2 A sample of smart card data in Singapore with selected information ........ 59  
5.3 An overview of travel distance and activity locations. ...................... 74  
5.4 Variable information of smart card data. ..................................... 83  
5.5 Original activity types, aggregated activity types and trip numbers......... 105  
5.7 A comparison of network properties with smart card data in 2010, 2011 and 2012.129  
6.1 A summary of indices used for measuring Polycentricity .................... 158  
6.2 Data innovation applications in this research. ................................ 164  
B.1 Transportation data sets used in this research. ............................ 185
# List of Figures

2.1 World population prospects: urban and rural populations from 1950 to 2050. 6
2.2 Traffic networks between Qingpu industrial town (small circle) and Shanghai (big circle). 7
2.3 Simplified decentralization urban plans in Singapore. 8
2.4 Types of urban spatial structures. 13
2.5 Morphological Polycentricity versus Functional Polycentricity. 14
2.6 Transportation and land use interactions. 15
2.7 Historical urban models. 16
2.8 Bid-rent theory model. 17
2.9 Interdependencies between land-use, transportation and activities. 19
2.10 Land use and transportation models. 20
2.11 Examples of spatial patterns. 23
2.12 An example of spatial interpolation. 24
2.13 The Steinitz model for landscape planning. 26
3.1 The scope of the research topic in this dissertation. 39
3.2 Complete loop of land use and transportation interactions. 40
3.3 A generic framework (bottom) associated with an urban design and planning process (top). 43
4.1 Framework for detecting functional urban changes. 46
4.2 Spatial analysis of urban mobility data. 47
4.3 Mechanism of a visual analytics tool. 50
4.4 Object relations in a prototype system. 52
5.1 Organization of sections in this chapter. 54
5.2 Case study area: Singapore. 57
5.3 Two types of data describing interactions between people and built environment.
5.4 Bus stops and train stations in Singapore.
5.5 The revised Concept Plan in 1971.
5.6 The revised Concept Plan in 1991.
5.7 Historical populations data from national statistics of Singapore.
5.8 Percentage change of private sectors over corresponding period of previous year.
5.9 Share of transport modes in 2004 (top) and 2008 (bottom).
5.10 Shared transport mode of different activities in 2004 (top) and 2008 (bottom).
5.12 Probability distribution of trip starting times for different activities in 2004 and 2008.
5.13 Spatial convex of urban activity locations in 2008.
5.14 Probability distribution of boarding and alighting time in 2004 and 2008.
5.16 Probability distribution of activity frequency in 2008.
5.17 Probability distribution of staying time.
5.18 Probability distribution of walking distance in 2008.
5.19 Probability distribution of trip starting time by age group in 2011.
5.20 Probability distribution of travel distance in 2011.
5.21 OD-matrices of journeys by MRT in 2010, 2011 and 2012.
5.22 Inferring information by “Recombination of Data”.
5.23 A demonstration of the applied Bayesian model.
5.24 Work-flow for inferring travel purpose from travel behaviors.
5.25 Case study area: Jurong East.
5.26 Trip classification.
5.27 An outline of proposed approach for measuring polycentric urban process.
5.28 Grid based data structure.
5.29 A demonstration of mean entropy calculation.
5.30 A demonstration of the misinterpretation of diversity index.
5.31 Spatial convolution with contiguity edges and corners.
5.32 Mapping activity locations in Singapore.
5.33 Density, diversity, centrality and difference between centrality and density.
5.34 Incompatible density and entropy patterns. ........................................ 108

5.35 Empirical probability distributions of the locational centrality, P(CI), for the studied periods. ......................................................... 111

5.36 Centrality map generated from travel survey data in 1997, 2004 and 2008. ... 112

5.37 A Voronoi map defining urban spaces generated from stop locations. ....... 117

5.38 Work-flow of the proposed analysis method. .................................... 119

5.39 Community structure in a network. .................................................. 123

5.40 Communities mapped back to geographical space. ........................... 126

5.41 Two varieties of network mapping. ................................................... 128

5.42 Changing communities and borders detected from daily transportation in Singapore from 2010 to 2012. ............................................. 130

5.43 Degree and average trip strength distribution in 2010, 2011 and 2012. ....... 131

5.44 Changing degree distributions in 2010, 2011 and 2012. ...................... 132

5.45 Changing distributions of Betweenness Centrality in 2010, 2011 and 2012. . 133

5.46 Changing distributions of PageRank in 2010, 2011 and 2012. ............... 133

5.47 Interpolated Betweenness Centrality landscape in 2011. ...................... 134

5.48 Interpolated PageRank landscape of Singapore in 2011. ...................... 134

5.49 Borders defining communities of urban movement in 2012. .................. 136

5.50 Changing communities from 2010 to 2012. .................................... 137

5.51 A visual analytics framework. ......................................................... 142

5.52 Data structure in network space and geographical space. .................... 143

5.53 Three levels of details. ................................................................. 145

5.54 A flow map. ................................................................................... 146

5.55 Three spatial scales: regions, zones and sub-zones. ............................ 147

5.56 Two views in the tool: network view and geographical view. ............... 148

5.57 Visualization of flows at subzone level. .......................................... 148

5.58 Real-time analysis of changing flows. ............................................. 149

6.1 A time-line of study materials used in this research. ............................ 154

6.2 “Analysis” and “Modeling” in the two presented analytic applications. .... 160

6.3 Work-flow based integration. ............................................................ 161

6.4 A generic work-flow for integrating method into geospatial analysis. ....... 162

6.5 Information flow (top) versus conventional planning flow (bottom). ......... 162
Chapter 1

Introduction

Cities are our important living space. A large amount of studies on urban space have been conducted, but they are never redundant as a full understanding of the intricate interactions between urban elements and their unprecedented changes has never been achieved. Progress has been made in every era, addressing new urban issues and exploring the unknown, but there still remains much to learn.

1.1 Research Background

There are currently more than 3 billion people living in urban areas and this number is expected to rise to 5 billion by the year 2050. Thus, cities and future cities are an ever important topic. This rapid population growth results in rapid urban transformation, which is an ongoing process that continuously shapes cities and societies. Many urban issues in rapid urban development have been concerned, such as a shortage of living space, traffic congestion, high energy consumption, and CO$_2$ emissions. To solve such issue in urban transformations requires a comprehensive understanding of the causes of urban issues, the dynamics of urban space, and the complexity of urban systems.

Urban transformation is far more than the simple result of top-down urban planning. Urban space is a natural result of environmental and socioeconomic dynamics. The original planned urban functions are reshaped by the inhabitants’ actual demands and uses, which has been termed as bottom-up self-organization. Due to these multiple driving forces, it is difficult to determine whether a modern city is the same as intended in the original plans, let alone to
manage it. In such a context, a good measure of change is required. Such a measure will help contribute to a better sense of ongoing urban processes and, thus, bring us one step closer to fully understanding, managing, and even predicting changes in cities.

Over recent decades, many urban areas have grown and spread through strong but heterogeneous sprawl. Yet contemporary cities are increasingly polycentric with continuous urban transformation of decentralization. In such context, the following questions are posed and serve as the motivation of this research: How quickly and to what degree are cities being reshaped? Does a polycentric spatial structure exist in socioeconomic space? And is that structure compatible with the planned polycentric physical space? Do people’s activities and intra-city movements follow patterns of polycentric organization in reality? Are there any changes in daily life associated with the sight of Polycentricity? The answers to these open questions would greatly contribute to a better sense of urban changes and, moreover, help to detect urban problems, evaluate planning strategies, and support policy making.

Newly available urban data may have the potential to answer these questions. The advancement of sensor technologies makes it possible to easily and cost-efficiently capture and store massive amounts of urban mobility data in a way that was previously impossible. These spatiotemporal data sets record people’s daily lives and contain rich information about how people are adapting to and exposing individual and collective forces on the shaping of urban space. However, there is a lack of techniques to extract such information from the raw data and interpret the information in certain contexts. For the purpose of this research, contextual information is the spatial structure in functional urban changes.

This research is intended to fill the gap between urban data and information by advancing geospatial techniques. Geospatial techniques have long been used in urban studies, but mainly as data inventory tools and not for data analysis or mining. In fact, the existing spatial analysis methods are already powerful in terms of detecting spatial patterns. They are extensible for enhancing broader data mining methods from computer science and knowledge from urban planning and urban geography. They offer a significant potential to facilitate geospatial techniques to obtain a better understanding of dynamic urban spaces in the era of big spatiotemporal urban data. Along this line, this research explores the potential of spatial analysis of big urban data for urban studies.

This research is conducted with the support of Singapore government agencies who kindly provided the transportation data used in this dissertation. Information about urban activities
and mobility was then extracted by the proposed analysis methods to make insights into polycentric urban transformation in Singapore. The results of conducted analyses were shared with government agencies.

1.2 Dissertation Outline

The rest of the dissertation is organized as follows:

- Chapter 2 examines the existing literature in various fields (urban geography, transportation and land use modeling, spatial analysis, etc.) that this work draws on.

- Chapter 3 raises the research questions from the stated research background and the reviews of state-of-the-art research.

- Chapter 4 describes the methodology that is applied to answer the research question and presents a research design on how to utilize the methodology in a practical case study.

- Chapter 5 implements the proposed methodology in a case study of Singapore, including a short introduction to the data and case study area, a descriptive analysis of physical changes in Singapore, and a set of spatial analyses on functional changes using transportation data.

- Chapter 6 is comprised of two parts. The first part synthesizes all of the insights gained in the analysis and the second part presents the conclusions and future work.
Chapter 2

Literature Review

This work seeks for a better understanding of urban transformation, a real-world issue that is driven by multiple forces from both top-down planning and bottom-up self-organization. It explores the potential of extensive geospatial techniques in the measurement and management of such urban transformation with benefits from readily available big urban data. In addition, integrated knowledge and techniques from diverse domains are required to convey raw data into meaningful information. Consequently, this research is positioned in an interdisciplinary field emerging from the established areas of urban planning, urban geography, geoscience, and data science. The review below covers those four areas. As they are very broad and extensively studied fields, only the most relevant topics are discussed.

In particular, the review starts with the evolution of cities, which is where specific new urban problems come from. A special focus is given to the phenomenon of Polycentricity, which has been described in the literature as a new type of urban form, and is the issue to be examined in the following case study of Singapore. Related work in urban geography regarding its formal definition and quantitative measure has been reviewed and discussed to identify the problem in Section 2.1.

Land use and transportation are two essential elements and their interactions greatly affect the shaping of a spatial structure. Their inter-dependencies and the consequence on urban activity and mobility are discussed from the perspective of urban modeling in Section 2.2.

Spatial analysis is important in measuring spatial structure since urban transformation is a matter of spatial location and their distributions and is the focus of this research. Related
geospatial techniques are discussed in Section 2.3 including three topics: basic spatial analysis methods for identifying spatial patterns, spatiotemporal analysis for understanding urban dynamics and urban applications represented by the new concept of Geodesign.

Emerging big urban data, as mentioned before, is considered as a huge potential for achieving a better understanding of urban dynamics. How newly available urban data could be used in urban modeling and spatial analysis is discussed in Section 2.4. An elaboration of the essentials of the new concept of big data is given along with some recent examples showing its potential for advancing urban studies and planning.

2.1 The Evolution of Urban Spatial Structures

Over the past decades, many urban areas have grown and spread through strong but heterogeneous sprawl. Contemporary cities are increasingly polycentric with a continuous urban transformation of decentralization. Open questions are therefore arise regarding the process of urban changes, for instance, How fast and how far is urban transformation going? Is the spatial structure of today the same as that implied in our plans? And how much are cities being reshaped? To answer such questions, a measurement of Polycentricity is required aiming at better understanding and managing of urban changes. In this section, the emergence of such spatial structure, as well as theories and methods that have been applied to measure such spatial structure, are reviewed.

2.1.1 The Polycentric Metropolis

Discussions about urban spatial structure should start with urbanization, which is one of the dominant trends of social-economic changes in the 20th century. As shown in Figure 2.1, in 1990, only 14% of the world’s population lived in cities. By the end of the 20th century, the proportion had already reached 47%. It is predicted that by 2050 around 70% of the world’s population will live in urban areas according to the data from United Nations. Along with urbanization, the trend is dramatic and ongoing growth of cities. Urban areas are growing with strong but heterogeneous sprawl. Accordingly, urban spatial structure in large cities is becoming even more complex as populations grow in size, engage in more travel, and are enabled to live more diverse lifestyles.
To distract urban flows and reduce density, the Polycentric mega-city region (MCR) emerged as a new phenomenon in the most highly urbanized parts of the world. The polycentric spatial structure is rising through an urban process of decentralization from big central cities to adjacent smaller ones, both old and new. It has been identified as an emerging urban form in [66].

The polycentric form can be identified as a type of ‘space of flow’ [31] in which physically separated areas are connected by a dense flow of people, information, products, etc. According to [6], “The spatial structure of modern cities was shaped, in large measure, by advances in transportation and communication.” The emergence of polycentric urban forms is closely tied to economic change and the rise of the automobile in the 20th century, which greatly improved the efficiency of personal transportation. The result is reclaimed settlements in the area between suburbs and the expansion of residence. The widespread development of intercity highway systems undoubtedly promoted this process, which expanded the persisting dominant concentricity. Together with the dramatic rise in car ownership, these developments caused employment and production to move out from the central area in order to use cheaper land. Consequently, the central area was transformed from manufacturing to service and office centers.

Contemporary mega cities are increasingly polycentric. They formed either by planning decisions that aimed to reduce the pressure of high density and traffic flow or self-organized individual choices to move to new sub-centers for cheaper housing, a relaxing living environment, or so. In any case, these decentralized phenomena are gradually emerging all over the world.
World Polycentricity

Polycentric city was observed in wide areas in Asia, where fast economic growth and development have significantly transformed the built environment. Like that in China, polycentric cities are emerging across the country [156, 155, 158]. As indicated before, one of the significant changes of fast urbanization in recent decades is that more and more peasants are moving from the countryside to cities, which constantly increases urban density. One aspect of growth management in China today is the emerging policy against continuing high density skyscraper-type construction, which is being replaced by new satellite towns. An example is introduced by [168], as shown in Figure 2.2 to demonstrate a simple polycentric structure, of the new satellite town Qingpu located about 35 miles west of the main city of Shanghai. It is expected to attract a population of about 100,000. Though it is a so-called industrial town, public facilities like retail and schools have been built, rather than only residential and industrial areas.

Figure 2.2: Traffic networks between Qingpu industrial town (small circle) and Shanghai (big circle).

Note: Image recreated from official planning website of the traffic network of Qingpu industrial zone


In Europe, polycentric transformation has been observed in many places as well. Hall and Pain studied and compared eight regions South East England, the Randstad (The Netherlands),
Central Belgium, Rhine-Ruhr, Rhine-Main, the European Metropolitan Region (EMR) Northern Switzerland, the Paris Region, and Greater Dublin [66]. Using historical data of these areas, distributions of population as well as communications are mapped. Commuting is used as a quantitative measure of Polycentricity. Taking commuting trends to work in South East England as an example, though the pattern is dominated by strong radial flows into and out of London, other areas located a few miles away from London also have complex cross-links with each other.

In the U.S., “Megalopolitan” areas have been identified that support the idea that “modern cities are better reviewed not in isolation, as centers of a restricted area only, but rather as parts of ‘city system’”[58]. Megapolitan regions is then defined as large areas that are composed of cities and counties connected through commuting patterns and economic exchanges. According to [84], six Megapolitan areas in the eastern US and four in the western US have been investigated. As one of the findings, they state that “as of 2003, Megapolitan Areas contained less than a fifth of all land area in the lower 48 states, but captured more than two-thirds of total US population with almost 200 million people, and are expected to add 83 million people (or the current population of Germany) by 2040”.

Figure 2.3: Simplified decentralization urban plans in Singapore.

This thesis investigates the polycentric phenomenon in Singapore, where decentralization planning was proposed in the early 1990s. As shown in Figure 2.3 is a simplified diagram of the concept plan 1991. The city-state is planned in a polycentric urban form. Decentralized transformation in Singapore has been ongoing for more than a decade. Urban infrastructures have been built in order to form centers distributed in a certain hierarchical structure, which
are visible physical urban changes. This research aims to identify invisible functional urban processes from the way people live and travel in changing urban space.

**Top-Down and Bottom-Up Changes in Urban Transformation**

As shown in the previous examples, spatial decentralization is a planning solution commonly used to redistribute social and economic activities in order to resolve escalating problems in urban areas, such as crowdedness, pollution, and high cost of living. However, its impact on people’s lives is still under investigation and whether new urban issues will emerge is still unclear. Are cities changing and influencing people in the way we expected? What is the real driving force? Is top-down planning or self-organizing dominant in such urban transformation? These are deeper questions needed to be answered not only in the context of decentralization, but in all kinds of urban processes.

Polycentricity can be the result of planning or self-organization. Some polycentric cities are well-planned in advance, like Singapore [47] or Shanghai [168]. Some are formed gradually by changing plans at different historical stages, like the great Jakarta area [70]. Some are the result of special urban policy, like that in Guangzhou and Shenzhen [155]. Some are shaped by both planning and self-development, like London and many other European cities [66].

There is, of course, always problems come in urban development, especially, in a fast urban development with dramatic population growth and GDP increasing. Like that observed by [124] in China, there is over-investment in physical capital, infrastructure, and property. Evidences can be easily found in empty airports and bullet trains highways to nowhere, thousands of colossal new central and provincial government buildings, ghost town. Actually, ghost towns, which are abandoned neighborhoods, villages, towns, or cities, have appeared in many cities worldwide [26, 150, 42].

A town may become a ghost town because of failed economic activity, natural disasters, government actions, and so on. Ghost towns may be an extreme case, but are evidence of how population growth in a specific area does not always meet the original expectations. Even deeper questions regarding the process of urban transformation could include building for what and whom? How much new infrastructures have been used? And how well habitats adapt to the new built environment? These questions can only be answered by measuring large amounts of human activity and mobility data.
Therefore, it is obliged to answer this question by measuring and managing urban transformations from not only physical development but also human lifestyles. The quantitative measurement of urban spatial structure is definitely one of the topics along this line and should be considered as a premise of following research built on certain urban forms and their impacts. A review regarding the definition of Polycentricity and its measures is given below.

2.1.2 The Fuzzy Concept of Polycentricity

The observed successes and failures of polycentric urban transformation can be formalized into a more abstract representation. This representation refers to a formal definition of Polycentricity. Though Polycentricity has drawn an amount of consensus, its definition remains fuzzy and needs further clarification [6, 76, 41, 29]. The main argument concerns distinguishing Polycentricity in physical and socioeconomic space. The review in this section is structured following this debate and is re-summarized based on the author’s understanding.

First of all, Polycentricity is considered a new type of urban form and/or urban spatial structure. In [120], urban form and urban spatial structure has been defined as follows:

“Urban form refers to the spatial imprint of an urban transport system as well as the adjacent physical infrastructures. Jointly, they confer a level of spatial arrangement to cities.

Urban (spatial) structure refers to the set of relationships arising out of the urban form and its underlying interactions of people, freight, and information.”

Polycentricity contains both aspects. On one hand, it is about the distribution of spatial clusters and, on the other hand, it refers to the underlying interactions between them. The related work studying urban spatial structure therefore also falls into two streams: one focuses on the morphological dimensions that donate the size and spatial distributions of centers, and the other looks into the functional dimensions that address the linkages between centers [60, 29].

Morphological analysis identifies centers mostly by measuring the density of urban stocks, such as population and spatial structure, by the number of centers and their spatial distributions. A wide range of methods to identify urban centers from population distributions has been proposed, and some mainstream methods are reviewed here to underline the progress. In [61], a measure is presented using a set of reference thresholds (cut-offs) relying on local knowledge. A disadvantage of this method is that the cut-off points may be arbitrary [6]. Another method
is proposed in [92] using spatial distribution of density functions and considering the peaks as possible sub-centers. With the use of statistics, the parametric method has been proposed using a regression model based on density and distance [93]. Non-parametric methods have been introduced based on a smoothed density function [94, 116]. In [91], a conditional Logit model is used to show how each center is differentiated with regard to establishment size and sector as well as the importance of center characteristics.

This functional aspect of urban spatial structure has been suggested as playing a key role in the overall performance of an urban system [29]. Recent work has shifted the focus to this aspect and highlighted the importance of considering the connectivity between centers [143]. Approaches along such lines that consider functional inter-dependencies between centers include the following: gravity models, which explain the interactions between spatial units based on their size and distance [100, 41]; network models, which use network properties as indices to measure the connectivity and structure of mobility flows in urban areas [138]; and connectivity fields, which quantify the connections of each center to the rest of the urban system [143]. A similar approach was applied by [29] that examined the relation between the morphological and functional aspects of spatial organization. A conceptual framework was developed and applied to reveal the compatibility of two aspects of spatial structure. The results show that functional changes are not necessarily the result of morphological changes.

Building on the work of [29], which differentiated morphological and functional Polycentricity by measuring different components of the intra-flows and inter-flows of an area, this research argues that the observed incompatible patterns might be caused by the predefined boundaries of areas as that has been handled in most of similar studies, since different partitions of space may lead to different divisions of flows. In that sense, it is difficult to draw obsolete parallels between morphological and functional Polycentricity since their measures are somehow interdependent. To avoid this, the present research argues that measures should be based on emerging functional centers from the way urban space is used for human activities, rather than centers with predefined boundaries.

All of these arguments are rooted in crucial questions that have long been placed in the debate of Polycentricity: how do we define a center and how do we measure the centrality/importance of a center? Modern cities have various types of centers, including monofunctional centers like education, work, and sports, and multi-functional centers like central business districts (CBDs). An even stronger argument by [143] states firstly that the concept of
Polycentricity has a weak theoretical foundation, secondly that it is highly scale dependent and shows different levels of spatial clustering at different scales, and, lastly but most importantly, that contrasts exist in morphological and functional context. Summing up the arguments, the next chapter redefines polycentric spatial structure, addressing essential factors that formulate the measure proposed in this dissertation.

2.1.3 Defining Functional Polycentric Spatial Structure

Beyond all ambiguous definitions, Polycentricity is a matter of number of centers, in other words, multiple centers existing in one area [76] and the kind of spatial organization of centers as shown in Figure 2.4. From the perspective of geography, a polycentric development can be considered a spatial process where urban functions diffuse from major centers to nearby sub-centers [100, 66].

During such spatial processes, the functions of highly centralized areas, such as CBDs, gradually become shared by sub-centers; meanwhile, urban stocks and flows are redistributed. As indicated in the literature such as [76, 95, 29], Polycentricity tends to be more closely associated with a balanced distribution with respect to the importance of these urban centers. Kinds of spatial distributions are demonstrated in Figure 2.4, from the perspective of transport geography, where the urban spatial structure has been categorized by its levels of centralization and clustering [120].

The clusters, which represent centers, refer to all kinds of urban stocks. It is more reasonable to measure the distribution of a population instead of the physical environment since real urban functions can only be found from the way people use urban space. This has been addressed in many comparative urban planning case studies. For example, in [66], it states: “The comparative analysis of cities has to begin by addressing basic problems of definition,..., physical or morphological definitions are better, helping to define limits in terms of urban land uses; but even they fail to unpick the functional relationships that may tie physically separate towns and villages to a central city.” They also identify problems in logic that different systems of national land use regulation result in built-up areas of cities that are related very differently to their functional reality, since increasingly complex cross-commuting patterns mean that built-up areas no longer describe the functional reality. Therefore, the concept Metropolitan statistical area (MSA) was presented and the comparative analysis of cities were done using functional, not morphological criteria.
Moreover, Polycentricity is about the balanced distribution of both stocks and flows. Figure 2.4 shows the distribution of stocks and Figure 2.5 shows the changing distribution of flows in a polycentric urban transformation. The changing spatial organization of these clusters and two kinds of changes have been differentiated and compared by [29] and again named as morphological changes and functional urban changes. They state that Morphological changes refers to changing size and geographical distributions of urban infrastructures, while, functional changes take connections between settlements into account. Together, they are are two kinds of analytical concepts both of Polycentricity. This dissertation agrees with such argument regarding “functional” and “morphological” Polycentricity.

Finally, spatial structure should never be identified by absolute value, but be understood as a relative concept that depends heavily on spatial scale and related subjects. Rather than clarifying the concept to a generally acceptable level, which can barely be achieved, a more reasonable
solution is comparative studies. More specifically, it is a kind of study that comparing the historical spatial structure to manage the progress of polycentric development spatially. Rather than using a binary variable (monocentric or polycentric), it is more rational to measure the degree of Polycentricity or Monocentricity on a comparable spatiotemporal scale so that an urban process can be sensed. In such thinking, this research argues that it is necessary to develop a process-oriented concept, instead of a static target-oriented one. By doing so, overall urban transformation can be measured and, thus, managed. Research in such line is rare but has a huge potential to be further developed at the age of big urban data, which are available and provide unprecedented possibilities to assess functional centers, as well as spatial interactions, on a large scale. Making use of these data sources to push forward the measurement of Polycentricity is one of the motivations of this research. A review related to big data will be given in later sections.
2.2 Spatial Interactions in Urban Dynamics

As discussed before, polycentric urban transformation is a process of relocation and reorganization of urban stocks spatially in a trend that is either more clustered or dispersed (as shown in Figure 2.4). A specific phenomenon associated with such an urban process is the development of new urban infrastructure, followed by changing decisions of activity locations. During this process, two elements that play major roles are transportation and land use. They are part of a retroactive feedback system where they influence one-another [120]. As illustrated in Figure 2.6, the advancement of transportation systems may increase accessibility to certain locations and, in return, increasing travel demands in certain locations leads to higher requirements for the transportation systems.

![Figure 2.6: Transportation and land use interactions.](image)

The following section discusses such urban interactions from a spatial perspective with a specific focus on the means used to interpret and represent such dynamics. In the first section, geographical models representing urban structures are reviewed in chronological order. The second section reviews operational models representing interactions. In the third section, factors that contribute to interactions between land use and transportation are discussed.

2.2.1 Urban Structure Models

The early urban theories and models are conceptual ones. They have been utilized for political planning, agricultural development, industrialization, and ecological development, as well as city and urban land use, and have greatly guided the implementation of real world applications.
They have made significant contributions in representing urban structures and are still referred to in contemporary urban analyses. These classic urban theories and models include posed qualitative theories of urban development, such as the concentric model [108], sector model [68], multiple nuclei model [67] and central place theory by Christaller [37] and Lösch [87], as shown in Figure 2.7.

The following provides a brief explanation of these models:

- The **concentric model** was developed to analyze the distribution of social classes. Each circle represents a specific socioeconomic urban landscape. The impact of such a model still exists in the spatial changes of modern cities such as Chicago, even one century after its development.

- The **sector model** takes into account numerous factors overlooked by the concentric model. In which, different land use and land values are considered.

- The **multiple nuclei land use model** is one step closer to reality. It argues that, even though a city may have begun with a CBD, other smaller CBDs develop on the outskirts of the city near the more valuable housing areas to allow shorter commutes from the outskirts of the city.

- The **central place theory model** determines the number, size, and location of human settlements in an urban system. It makes certain assumptions of spatial structure, which are used in this thesis as fundamental theory for understanding Polycentricity. For instance,
the larger the settlements are in size, the fewer in number they will be, i.e. there are many small villages, but few large cities. As a settlement increases in size, the range and number of its functions will increase accordingly. These assumptions have been implemented in certain applications and, on the other hand, been discussed in the statistical analysis of real data, like in work in [25].

These models are essentially based on social theories, which have no reference to a space or time scale. Only when regional science came along did the models become more preoccupied with space, but less with time. One famous model is known as the Bid-rent theory [5], as shown in Figure 2.8, which refers to the correlations between prices of real estate and distance from the CBD. Lowry’s spatial interaction model is another widely-known model and it combines employment, population, and transportation into one model [88]. It is considered the first land use and transportation model and has been expanded upon by several other models known as “Lowry-type” models.

2.2.2 Operational Models for Urban Processes

Starting with these regional models, operational models were developed and facilitated by the innovation of digital technologies. Forrester’s dynamic urban model [49] contains a temporal dimension represented as attractiveness to model how location choices are gradually changing, but it has been criticized for its lack of spatial dimension compared to micro-simulation models and spatially-informed models like automata-based models [17, 18]. The emergence of geo-information makes these models even more practical and feasible for representing reality. The
history of urban models can be sorted as follows:

- **1960s -1970s**: Spatial interaction models, spatial input-output models, and linear programming models.
- **1980s**: Cellular automata models (CA), agent-based models (ABM), land use models (LUTI).
- **1990s**: Spatial econometric models (SEM) and systems dynamics models (SDM).
- **2000s**: Crowdsourcing multi-agent simulation models.
- **2010s**: Geo-spatial agent-based simulation models.

From the review of developments in urban models and corresponding modeling methods, a clear trend is the involvement of spatiotemporal dimensions, which are indispensable parameters for representing urban dynamics. In actuality, there are more perspectives that can be referred to in the progress of urban modeling. These developments are summarized below:

- From static to dynamic, in other words, adding a temporal dimension.
- From equilibrium to dynamics/evolutionary.
- From aggregation to dis-aggregation.
- From top-down model to bottom-up modeling.
- From single parameters to multi-parameter.
- From sub-models to complex systems interactions between urban elements are considered.

Along with the development of information technology, operational models have been developed for prediction and simulation. Modelers have tried their best to make the models closer to reality from the aspects of appearance and behavior [142]. To do so, Geographic information system (GIS) is brought in as a very powerful tool used for 3D visualization, data management, and spatial analysis. A review on spatial techniques will be given later.
2.2.3 Land Use and Transportation Interactions

“Space shapes transport as much as transport shapes space, which is a salient example of the reciprocity of transport and its geography.” [120]

Though great progress has been achieved by advanced computation power, still, the most essential issue in modeling a more realistic world is a better understanding of the intricate interactions between urban elements. Transportation and land use are vital elements considered indispensable in all operational urban models. For instance, on a small scale, they are studied for traffic optimization and travel behavior; on a medium scale, they matter to population distribution and the location choice model; on a large scale, they are linked to environmental problems and urban economies. The research in this dissertation has a particular focus on activity and mobility patterns because activity patterns reflect how people use urban space and mobility patterns reflect how people move and commute in urban space. Together, they reflect a kind of spatial structure.

Figure 2.9: Interdependencies between land-use, transportation and activities.

Note: Image is recreated from [120] http://people.hofstra.edu/geotrans/eng/ch6en/conc6en/activityuse.html

Land-use, Transportation, Activity, and Mobility

Interdependencies between land use, transportation, and activity are shown in Figure 2.9. There, infrastructure greatly shapes the kinds of urban form: transportation systems serve as a linkage connecting locations and the advancement of transportation systems promotes urban activities and the associated mobility between locations. Urban activities are taken at certain locations
where land use patterns are derived and influenced by the existing urban form and spatial structure. Meanwhile, the spatial separation of human activities associated with certain land uses creates the need for travel and goods transport in reverse. This logic is the underlying principle of transport analysis and forecasting. Urban interactions thus come from such an interdependency/paradox scenario.

Figure 2.10: Land use and transportation models.

Note: Image is recreated from [151]

In particular, the diagrams in Figure 2.10 are a famous land-use transportation model (left) that shows the process how different urban elements trigger changes in other elements at different spatiotemporal scales (middle). It can also be considered an illustration of interactions between humans and the built environment (right). The eight subsystems are ordered by the speed in which they change, from slow to fast (middle), and the land-use transportation feedback cycle in [89].

Factors in Modeling Inter-dependency

Successful models have been developed and utilized for planning decision-making, such as ITLUP [106], TRANUS [43], MEPLAN [45], DELTA [130], UrbanSim [146, 148], and ILUTE [126]. Of these urban models, various categories of land-use and transportation models are summarized in the related literature review, giving an overview of how these models represent the real world. To compare the key features of UrbanSim with other operational urban models, [146] gave a review of their main characteristics, including model structure, indicators, spatial and temporal bases, as well as transport interactions. Instead of a single model, UrbanSim is
a combination of a demographic and economic transit model, household and employment mobility model, location model, real estate development model, and land price model. In [144], computer simulation models are compared on six features: level of analysis, cross-scale dynamics, driving factors, spatial interaction and neighborhood effects, temporal dynamics, and level of integration. These features are actually all indicators that are used to evaluate the process of urban change. In [78], various characteristics of computer simulation models are discussed from the following perspectives: static or dynamic, transformation or allocation, deterministic or probabilistic, sectorial or integral, and zones or grids.

Recently, a comparison of equilibrium and dynamics in urban modeling was addressed in work by [131]. It stated that “it is becoming more and more apparent that without understanding the inherent inertia of different subsystems of cities it is impossible to assess their likely responses to land use or transport policies.” It also called attention to the complexity of urban change. Transport and land use models are mostly static equilibrium models that assume there is certain equilibrium between supply and demand. In contrast, dynamic models consider the different speeds of the processes of urban change and concentrate on their outcomes over time and the path dependence. The evolution of cities is definitely more likely to be represented by a dynamic model. However, the complexity behind urban dynamics is still a wild field that requires new urban science to explore it.

### 2.3 Advanced Spatial Analysis for Urban Studies

A recent expression of the long tradition in urban analysis of rapidly embracing technological developments is the adoption of GIS [106]. A substantial body of research exploring the role and potential of geospatial tools to support various forms of urban analysis and planning has accumulated [89]. Successful examples have been developed for transportation and land use changes [89, 149, 141] and other wide applications like disaster management and environmental resource management.

Although GIS is considered a powerful and successful tool, it has been pointed out in the research agenda for metropolises in the 21st century [125] that GIS is mostly used as a data inventory and information management tool rather than a spatial analysis and modeling tool. Similarly, it was stated in the keynote speech of the 2010 Geodesign Summit that, “In many cases, their beauty is almost literally only skin-deep. Frequently missing is any understanding
of the objects beyond that required to generate the representations themselves.” ([48]). Over the past decade, the GIS community, together with people from various substantive fields, has made great efforts to integrate GIS with comprehensive analytical and modeling techniques.

This section aims to give a discussion about the developments in GIS and how the GIS community facilitates the use of GIS to support urban design and planning. The following review starts with a summary of the spatial analysis toolbox that is widely used as fundamental methods in measuring spatial patterns, and then two trends closely related to this research are discussed as well. The first trend is the concept space and time which is considered as an indispensable element of modeling and simulation [147]. The second trend is to apply GIS in impact assessment, more specifically, it is about how GIS can be integrated and used as a design and planning support tool. Addressing these two aspects shows the potential of GIS in analyzing and modeling urban issues.

2.3.1 Spatial Analysis of Urban Structure

Spatial data refers to urban data that contains location information as well as other attribute information. Spatial analysis as a general term describes a technique that uses location information to better understand the urban process generating the observed attribute values [50]. Spatial analysis is important in this research since urban transformation is a matter of spatial locations and their distributions. To measure the spatial structure of locations, a set of spatial analysis tools is provided. Selected spatial analysis tools of importance are briefly introduced in the following sections to formalize this research as a quantitative analysis problem.

Spatial Autocorrelation

Objects such as houses, trees, and cities are rarely randomly distributed. In fact, there is always a certain degree of patchiness. This is stated in Tobler’s first law of geography: “everything is related to everything else, but near things are more related than distant things” [139]. This observation is embedded in the gravity models and deeply rooted in our understanding of spatial interactions. Spatial auto-correlation is the associated index that describes how similarity varies with distance between locations and how this variation is affected by distance. In other words, spatial autocorrelation represents the spatial patterns of distributions. Examples are shown in Figure 2.11.
Figure 2.11: Examples of spatial patterns.

Note: (a) Spatial randomness. (b) Positive spatial auto-correlation that where objects are clustered.

Indices used for measuring spatial autocorrelation include Moran’s I [97], Geary’s C [52], Ripley’s K [118], and the Getis-Ord G statistic [53], and these are implemented in most of the spatial statistics tool sets.

**Spatial Interpolation**

Although spatial data are abundant nowadays, they are discrete values that cannot be used to represent continuous surfaces in real world. Therefore, spatial interpolation is the process of using points with known values to estimate values at other points based on certain spatial correlations. A basic assumption derived from the first law is that the value to be estimated at a point is more influenced by nearby control points than those that are farther away. An example is given in Figure 2.12 where a digital elevation model (DEM) is interpolated from scattered point data.

A spatial interpolation problem can be simplified as, given a set of spatial data, finding the function that best represents the whole surface and will predict values at other locations [82]. To find the function, a regression model is required.
Figure 2.12: An example of spatial interpolation

**Spatial Regression**

A regression model relates a dependent variable to a number of independent variables in an equation that can then be used for prediction or estimation. It is a core aspect of the spatial methodology [11]. A promise of a successful spatial regression is the existence of a certain spatial autocorrelation. Therefore, a spatial autocorrelation should be simply measured, for instance, by using MoranI. The foundation of a spatial regression is a normal non-spatial linear regression model. Spatial regression methods capture spatial dependency in regression analysis using two major aspects: spatial lag dependence and spatial error dependence. Geographically weighted regression (GWR) is one of the most common models based on a distance function [28]. More complex functions can be estimated using techniques like the Markov Chain Monte Carlo (MCMC) method.

### 2.3.2 Spatiotemporal Analysis

Due to uncertainty and dynamics in the real world, to be successful at characterizing and simulating real world processes, larger dimensions of scale should be considered. These dimensions can be temporal, spatial, or even intuitional. Thus, advanced spatiotemporal analysis and modeling techniques have emerged and are considered the current landmark of GIS [59]. The history of GIS can be summarized as follows:

- **1980s**: The first landmark was the advent of commercial GIS software.

- **1990s**: The second landmark was the application of large database software.
• 2000s: The third landmark was the emergence of car navigation systems and Google Earth.

• Present: The fourth landmark is the rising awareness of the importance of time within the GIS community and the development of models that can be used to represent dynamics.

The initial research on spatiotemporal analysis in geospatial tools dates back to the idea of temporal GIS [85]. In [110], a triad representation framework is proposed known as “when (time) + what (attribute) + where (location)”, which greatly contributed to the later objectification of spatiotemporal data. Next, [79] extended the function of the traditional map, which greatly promoted the usage of the space-time cube (STC) for visualization, advanced object-oriented data queries, and visualization techniques. A review of interactive geo-visualization techniques and tools for spatiotemporal data was made in work by [10]. Due to the availability of spatially informed data sets, such as mobile data and social network data, research in spatiotemporal analysis has been extended to understanding temporal features. Later applications started to focus on enriching semantics into spatiotemporal visualization like in [167] and applications to massive data sets [162]. A more detailed review regarding spatiotemporal visualization can be refer to [166].

Visual Analytics of Movement Data

Nowadays, urban data, especially transportation data all contains rich spatiotemporal information. Corresponding spatial data mining methods are even more urgently needed. It is understandable that in such context, spatiotemporal visual analytics is promoted a lot as an efficient way to explicitly detect movement patterns individually and collectively [9]. Visual analytics is a newly emerged eld that grew from the fields of information visualization, which shifted the research focuses to analytical reasoning operated by interactive visual interfaces [7]. This research considers visualization an important way to convey knowledge and exchange information and a crucial part in geospatially-aided urban design and planning in the era of big urban data.

2.3.3 Spatially Informed Model for Impact Assessment

Integrating expert knowledge into spatial analysis is another issue in GIS application. GIS tools, represented by ArcGIS, provide the “front end” (from data to user), while “back end” (from user
to data) operations allow analysts and planners to better manage, display, and communicate information. However, despite the considerable advantages, what geospatial tools provide is still not sufficient for interpreting urban phenomena such as transitions in economic structures, social settings, and political and legal backgrounds. In most cases, additional expertise is required. As a result, a new concept of “Geodesign” has been proposed and it has drawn much attention. Instead of listing all kinds of urban applications, “Geodesign” is discussed here as a representative way of improving GIS for urban design and planning use. Since the interoperability of the model between different disciplines is the central enabler of the Geodesign concept, it matches well with the core concept of this research.

Figure 2.13: The Steinitz model for landscape planning.

Note: Image is recreated from [134]

In fact, Geodesign has been applied in landscape architecture for almost twenty years. The re-launched discussion gives a new meaning to Geodesign from new perspectives [134, 21]. It emphasizes collaboration and interdisciplinary cooperation to develop the best and most sustainable design that takes into account livability (people), environmental impact (planet), and efficiency (profit) [12]. As introduced in [1], Geodesign influences design tasks throughout the whole process and provides four functions, namely sketching tools, spatially informed models,
fast feedback, and iteration. Of these four functions, spatially informed models are the core element and crucial for knowledge integration. They reduce the complexity of the information and estimate how various systems (social, environmental, economic, etc.) will respond to the plans suggested by the sketches. These models provide information on both the potential impacts (e.g. carbon footprint) and changes (e.g. population growth rates, development patterns).

Figure 2.13 is a general model of landscape change developed by the urban planner Carl Steinitz. The model enables the design of alternative futures, which can then be evaluated in terms of their impact on the natural environment as well as their utility to the human population and the alternative future. It is a general demonstration of the applications of geospatial tools for landscape planning.

Applying geospatial techniques for these kinds of application has long been studied, like using GIS to map and analyze health events [39], for disaster management [81], and for modeling urban processes [106]. In recent years, GIS has been closely linked with smart cities. The essential idea is to build GIS infrastructure to help organize and manage urban resources.

The huge potential of geospatial techniques in urban studies inspired this research. This thesis aims to develop an advanced spatial analysis method for mining implicit patterns from big urban movement data and to build spatially informed models embedded with knowledge of urban form to understand change in cities. Both objectives heavily rely on the availability of data. Fortunately, the advancement of sensor technologies makes it possible for us to easily and cheaply capture and store massive amounts of data in a way that was almost impossible in earlier decades. In the next section, the data issues are discussed.

2.4 New Analysis Methods Using Urban Mobility Data

Big data has received unparalleled attention in both academic research and industry. It is no surprise that the concept of big data has also drawn the attention of computational urban design and planning, because data is so important that no urban design or plan is really made from a sketch, but from analyzing, evaluating, reconstructing, modifying, or expanding existing things. Can this new concept of big data greatly improve urban studies and planning? In particular, is big data a new chance for making geospatial techniques and applications a supporting tool? If yes, in what sense? Such questions lead to the review in this section.

Knowing that “big data” is a comparatively recent research concentration, rather than a
historical review, the following sections discuss big data issues based on the author’s knowledge of the state-of-the-art progress, composed of two parts: (1) a redefinition of “big data”; and (2) existing work done with urban mobility data.

### 2.4.1 New Concept of Big Urban Data

The emerging of big data is considered as a revolution that will transform how we live, work and think as that described in [90], big data is about three major shifts in mindset that are interlinked and, hence, reinforce one another. The first is the ability to analyze vast amounts of data about a topic rather than be forced to settle for smaller sets. The second is a willingness to embrace data’s real-world messiness rather than privileged exactitude. The third is a growing respect for correlations rather than a continuing quest for elusive causality. In particular, this research strongly agrees with the third point and considers it a way to obtain a deeper understanding of urban interactions.

**Definition**

Big data is used to describe a massive volume of data sets that has become too complex to be handled by traditional data processing methods. It has been articulated that the mainstream definition of big data is the three Vs: volume the increasing size of big data, velocity streaming at unprecedented speed, and variety coming in all types of forms. In [23], “big” data is compared to “small” data from 10 aspects, namely goals, location, data structure and content, data presentation, longevity, measurements, reproducibility, stakes, introspection, and analysis. The results can be briefly summarized that “big” data has vague goals, comes with surprises, spreads throughout space virtually and geographically, is unstructured, without predefined users, is stored in perpetuity, and is hard to measure because it is qualified differently and analyzed with incremental steps. In the context of this research, big data refers to urban data, particularly urban mobility data.

Actually, big data is not a new concept but exists in every era where the tools for data processing are always being stretched by increasing size [20]. Even the term “big data” is not new. It was first used by Bill Inmon in the 1990s, then formally used in 2008 in the journal Nature in an article titled - “Big Data: Science in the petabyte era”. Since then, it has become increasingly promoted by the development of the Internet, cloud computing, mobility techniques, and
so on. Foreseeing the huge data market, more and more attention has been given to various “big
data” related topics from academic research to industrial products. Though many attentions
have been drawn, doubts and issues exist meanwhile. The following review summarizes the
current opinions regarding the potentials and challenges of big data.

Potential

Data sets are magic mines. From the first sight of data sets, only explicit values are captured.
They are like icebergs floating in the sea: mores values are hidden below sea level. The kinds
of data innovations are summarized by Mayer-Schönberger. with vivid examples. That review
is reorganized for the context of this research and is as follows:

(1) Data innovation

Data sets are magic mine. From the first sight of the data sets, only explicit values are
captured. It is like the iceberg floating in the sea. More value is hidden below sea level. Kinds
of data’s innovations are summarized in [90]. That review below is reorganized in the context
of this research,

• The recombination of data - dormant values may emerge by combining one data set
  with another. Examples can be easily found, such as fusing two data sets to improve data
  quality or comparing two data sets to draw certain conclusions.

• Extensible data - byproducts may be gained from unspecified analysis. Unlike previous
  purposeful data collection, data are sometimes collected without a specific reason. An
  unexpected reason may arise by post-processing of the data set, such as discovering col-
  lective effects from accumulated individual records or deriving one data set from another.

• The reuse of data - meaningless data can be used for untapped purposes. Since byprod-
  ucts are frequently gained, the same set of data can be used in diverse applications. In
  many urban studies cases, historical data collected during different periods of time can be
  stored, shared, and cross-referred in further studies to avoid repetitive acquisition.

• Open data - more insights can be gained from diverse users. This innovation is especially
  represented by open source data platforms, such as Open Street Map, that sometimes give
even more precise details by contributions made from individuals. In urban studies, it is normal to refer to more than one data resource to reconstruct a complete image, which is in line with the idea of open data.

- **Data exhaust** - even bad or erroneous data can be used for improving intelligent systems.

The above data innovations give an outline of the thoughts on using newly available data for urban studies. In this research, the case study is classified as a “big” data approach since transportation data is analyzed for new usage, which can be classified as “combination of data”, “reuse of data”, or “extension of data”.

(2) Better data quality for an understanding of urban systems

With good enough data that has a time and location tag, our experiments of how cities function is certainly enriched. Urban interactions can be better understood, resulting in more informed decision-making with respect to the knowledge of how better to interact in cities [20].

From the perspective of volume, new data sets with ever higher spatiotemporal resolution allow us to trace the urban process on both the short-term and long-term. From the perspective of data integrity, although there are still some areas of the world excluded from advanced technology, big data is collectable from an ever larger geographical area. Like the famous Facebook had over 845 million users who spent more than 9.7 billion minutes per day on the site [152]. This dissertation conducts experiments using smart card data collected by an automatic fee collection system [109]. In the case of Singapore, more than half of the population uses public transportation, generating more than 5 million records per day, which geographically covers the whole country.

Besides the spatiotemporal feature, these data sets are generated from human activities by human agents. In [56], the concept of “human as sensors” is introduced. Humans are carriers of all kinds of sensors, such as mobile phones. The type of network formed by such sensors consists of the humans themselves; therefore, it contains both spatial and social information. These data sets make it possible for analysis using ground truth and offer a direct look at human behavior. Compared to conventional surveyed data in urban analysis, sensor data has a significant advantage in terms of efficiency and reliability. More state-of-the-art examples on mining urban information from censored data will be reviewed in the next section.
Challenges

There has been much debate on the challenges that have come with the rise of big data. In work by [15], big data is associated with the geographical quantitative revolution that occurred 40 years ago. Problems rooted in the long history have been criticized, mainly regarding the disconnection between data and knowledge. For the context of this research, the challenges in using big data are addressed in this section.

(1) Data management

There is a lack of simple ways to process massive data sets. The volume of automatically collected data sets increases at a dramatic speed. These large data sets come with a higher spatiotemporal resolution but, on the other hand, need to be managed with advanced data management tools, which are too professional and complicated for general use. As a result, many data platforms for integrated techniques are required. One example is “smart cities”, in which GIS plays an ever important role, but is plugged with a simple interface for general users to access.

(2) Knowledge discovering

Stored original big data, unlike traditional data acquisition, is collected mostly without any pre-defined purpose. Therefore, it is meaningless without contextual information. To make use of it, advanced analysis methods are required to find meaning in the random variables and extract potential information. “Urban computing” as a representative new concept was proposed by Microsoft Research in 2009. It is defined as “a process of acquisition, integration, and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces, such as sensors, devices, vehicles, buildings, and humans, to tackle the major issues that cities face, e.g. air pollution, increased energy consumption and traffic congestion, ... , Urban computing also helps us understand the nature of urban phenomena and even predict the future of cities.”

(3) Information representation

As indicated in the definition of urban computing, visualization is an indispensable component that conveys information between different domains. As discussed before, a rising field is the visual analytics of movement data. Visual analytics is defined as “the science of analytical reasoning facilitated by interactive visual interfaces. It combines automated analysis techniques with interactive visualizations so that to support synergetic work of humans and computers” [9]. In many cases, urban designers and planners with expertise may give better perceptions of the hidden patterns. With visual analytics techniques, they can be involved in the data mining process and shorten the pipeline of big data analysis by interacting with the data directly.

(4) Privacy preserving

A negative side-effect of data that cannot be ignored is data privacy. Abuse of data may cause huge losses and harm to individuals and society. In particular, privacy has been highlighted as an important issue regarding smart card data, which is widely used in research - including this dissertation - for understanding travel behavior and improving travel services [2]. The French Council for Computers and Liberty recommends being careful with such data because the personal movement of individuals might be reconstituted. However, smart card data, which is no different from other individual data like credit card or road toll data, can be properly used to avoid the privacy issue [38].

To confront this issue, an emerging field is privacy preserving techniques. In the review by [145], privacy preserving data mining approaches are classified into five dimensions: data distribution, data modification, data mining algorithms, data or rule hiding, and privacy preservation. Regarding spatiotemporal movement data, the work by [54] addressed many data privacy methods and applications from data distribution and sharing to analysis. Privacy preserving has become an important topic at all of the top conferences on data mining and GIS. This research believes that with proper privacy preserving techniques, data can be used in the right way and side-effects can be minimized.

2.4.2 The Use of Urban Mobility Data in Urban Studies

In the age of big data, location data generated by activities that humans are intimately involved with is abundant. Multiple data sources exist everywhere in cities, such as GSM traces on
cars, trains, and taxis; WiFi data collected in shopping malls, auditorium rooms, and other public spaces; social networks such as Twitter and Facebook, which contain indirect location information that can be extracted by text processing; and tagged data such as smart card systems, which have been used in health care, postal services, banking, and transportation. Obtaining a large amount of data is not a key problem anymore; instead, it is more valuable to capture some essential ideas and figure out how they can benefit urban studies.

This research has a special interest in urban dynamics and urban complexity. Centered by this topic, the following review focuses on the use of newly available location data, such as smart card data, for understanding social issues in cities. The review is organized into two parts, namely knowledge discovery and technique improvements gained using big mobility data.

Valuable insights have been provided into social activities and complex urban space through the analysis of big movement data, since urban travel is a good proxy for the transfer of urban flows, such as people, products, and energy, and reflects the dynamics of cities. In particular, the large amount of data makes it possible for us to discover the implicit patterns and regularities of human travel behavior.

**Individual human behavior** can be easily identified. For instance, daily activity patterns have been analyzed using mobile telephone data [114, 111] and the spatiotemporal structure of urban mobility has been studied using travel survey data. In [86], spatiotemporal human mobility patterns were investigated by means of smart card data in Shenzhen, China. In fact, the statistical analysis of human travel behavior using types of transportation data has been conducted in many cities [107, 86, 99]. In particular, as smart card payment systems are rapidly adopted in cities around the world, they have become an important source of large quantities of very detailed data about individuals daily travel [109].

**Collective effects** are the results of crowding activities. Convincing conclusions were previously hard to obtain because of unreliable and limited data sets, but now they can be discovered from abundant big data. For instance, in [135] where a time-resolved in-vehicle social encounter network on the public bus was constructed to discover the hidden encounter small-world in “familiar strangers” daily life. In terms of understanding urban space, relevant work using network analysis to find geographical borders between human movement has used GPS
tracked vehicle data at the regional scale [117], telephone data sets at the national scale [115], and air transportation data at national and global scales [63, 138]. These “border” effects were proved in [136] as a mechanism behind human movement.

**Regularities and laws** Classical theories, such as scaling law and zip law, are supported by much evidence and have been used to explain and predict the growth of cities [80, 25]. The discovery and proof of these universal laws require large sample sets [40]. The availability of large data sets now enables us to discover and verify these various patterns and laws [133, 102, 129]. For instance, a universal rational model has recently been proposed for mobility and immigration patterns and verified by using long-term immigration and communication data between regions [129].

**Heterogeneous local contexts** Though regularities exist, such as scale laws, cities are developed in a heterogeneous way. Local knowledge or local data sets are necessary to calibrate a model in a specific context. Examples can be found in the implementation of spatial interaction models. Real data has been used to calibrate the variables, such as in [159] using taxi data.

A clear trend is exploring the potential of using “big” location data for urban studies, as proposed by [114]. Along with this trend, new urban analysis methods emerged and are summarized as follows.

**Methods for enriching data set:** Data gaps can be filled by the fusion of multiple data sources. For instance, in Singapore, urban activities are identified from a synthesis of smart card data and survey data [33]. In this case, the share of transport modes is analyzed from travel survey data and then, using the known public transportation data sets, a complete data set describing all transport models is generated. Similarly, in a study by [157], taxi data combined with points of interests (POIs) was used to discover regions of different functionalities in Beijing.

**Methods for extracting information:** Inferred techniques, such as machine learning and regression analysis, have been used to generate indirect information for non-transportation planning use. This dissertation considers this a big potential data source for impact assessment of
urban design and planning. A few examples will now be given. In the case of Singapore, a
discrete choice model was used to estimate dynamic workplace capacities [104]. Similarly, the
GPS trajectories of taxi cabs travelling in urban areas provide detailed location information, and
[113] used the getting on/off frequency of taxi passengers in a region to depict social activities.
Machine learning methods are also being introduced to infer land use from mobile phone activ-
ity records and zoning regulations [140]. Differences in temporal patterns of space consumption
have also been compared using mobile data [3] on a large scale. In [159], a theoretical urban
interaction model was calibrated using taxi data.

**Methods for evaluating new proposals:** All kinds of mappings, such as O-D matrices,
have been conducted to identify the influence of local changes on a global system. In a study
by [107], boarding times and alighting times were mapped and analyzed to prove the reliabil-
ity of smart card data in Seoul, South Korea for future use. [99] estimated a public transport
O-D matrix from smart card and GPS data in Santiago, Chile for transport system analysis. So
far, most of the applications are transportation planning oriented, but a few examples for urban
planning exist and are waiting to be explored. For instance, as part of the research work in this
dissertation, a new centrality measurement is proposed to identify functional centers [165]. In
a study by [123], data collected from a smart card system (Oyster card system) was used to
infer the statistical properties of individual movement patterns and to identify polycentric urban
forms in London.

**Methods for predicting and simulating:** Patterns achieved by analysis methods, includ-
ing clustering methods and statistical methods, can be modeled to reconstruct the dynamic pro-
cesses of cities [105]. For instance, for transportation, data mining methods and public transport
planning models can be used to obtain an improved portrait of users’ travel behavior, and this
was tested in Quebec, Canada using twelve one-week records [2]. For land use, the machine
learning classification algorithm has been adopted to identify clusters of locations with similar
zoned uses and mobile phone activity patterns, thereby finding the relationship between land
use and dynamic populations [140].

Assessing the functions of urban space is of significant importance for understanding urban
problems [137] and evaluating planning strategies [73], which are the main concerns in this
dissertation. However, assessing urban functionality requires costly survey methods, such as field investigation and interviewing. Furthermore, the reliability of the information is heavily influenced by subjective factors such as time, place, and the investigators’ personal experience. The advancement of sensor technologies makes it possible to collect large scale and dynamic urban data without the aforementioned challenges. These new data analysis methods inspire us to develop integrated spatial analysis and modeling methods.

2.5 Chapter Conclusions

This chapter started with a discussion of a specific urban phenomenon - Polycentricity - and developed a review of related theories and techniques that contribute to better management of this urban process. There is interdisciplinary research in this dissertation that covers diverse fields, but only the topics closely relevant to the central question and research methodology are reviewed. A summary of the conclusions of the review is as follows:

**Phenomena:** To improve the understanding of cities as complex systems, much attention has been given to analyzing and modeling urban dynamics, urban processes, and interactions between urban elements. This research follows this trend with a special focus on urban transformation of Polycentricity. Since Polycentricity is emerging as a new type of urban form and many issues are raised in the urban process of decentralization. In such context, managing urban transformation has become a priority as one of the central challenges of urban studies and planning.

**Argument:** This research follows the argument that functional changes are not tied to morphological changes. Since cities are shaped by both top-down and bottom-up forces, the real functions of urban space are often redefined by individuals’ actual needs. It is more meaningful to measure the changing polycentric spatial structure that emerges from changing human activities and movement patterns than to investigate urban infrastructure development in purely physical terms.

**Knowledge bases:** Much progress has been made on techniques and/or applications in different fields regarding the analysis, modeling, and representation of urban processes. Although
these fields are reviewed separately, they are actually cross-related to each other. For instance, simulation and GIS both have a software engineering component and some models are used in simulations. The challenges and new trends in these domains that have been identified within the urban realm are briefly summarized as follows:

- Transport geography: interdependency exists between land use and transportation development. Urban activity and mobility are linkages between these two parts. How this kind of interdependency works in urban systems is still unclear.

- Urban modeling: dynamic models that reflect correlations between different elements, such as transportation and land use, are needed to replace equilibrium models.

- Spatial analysis: conventional spatial analysis methods provide a knowledge base for measuring spatial interactions. Spatiotemporal analysis is the new landmark and expertise is needed to build spatially informed models for impact assessments of transportation and land use planning.

**New chance:** Big data comes not as a new term, but as a new way of thinking about massive data sets. It has the potential to fill information gaps, discover hidden correlations, and represent the real world. This newly available human activity and movement data gives us a chance to look at human behavior. Moreover, newly available big mobility data opens a door for us to examine the impact of infrastructure development on peoples’ lives and, in return, how cities have been reshaped by individuals’ travel needs. In other words, it can assess the impact of transportation and land use plans based on what happened in reality.

A conclusion can be drawn from the review is that there is huge potential for geospatial techniques in the integration of data, knowledge, and techniques for a better understanding of urban dynamics. In the next chapter, a refined research question is developed based on these thoughts about the state of the art.
Chapter 3

Research Statement

This is an interdisciplinary study that uses a diverse range of knowledge and approaches to explain the complexity of an urban phenomenon. The specific urban phenomenon addressed in this dissertation is urban transformation. It is not a new phenomenon that appeared recently, yet it remains an increasingly crucial question in urban studies as the 21st century is said to be the century of urban transformation [65]. Unprecedented changes occur in rapid urban processes, but we lack the proper quantitative methods to evaluate and manage such processes. In a much broader sense, this issue is a matter of urban dynamics. The dynamics behind urban processes in terms of interactions between transportation and land use - and between the built environment and people - are most concerned topics, but there is still much knowledge waiting to be discovered.

Urban studies are conducted in diverse fields. This research follows the line of quantitative analysis, focusing on facilitating the use of integrated spatial analysis methods to find extra value in urban data, especially newly available big transportation data. The obligations and potential of such research have already been stated in the literature review. To apply the theoretical approaches to a real-world problem, a case study of Singapore is conducted. The diagram in Figure 3.1 illustrates the position of this research.

Building on this premise and background, the problem statement, research questions, and aims of this research are formalized as follows.
Figure 3.1: The scope of the research topic in this dissertation.

Note: Land use and transportation interactions is the specific topic investigated in this research and a concrete case study is performed in Singapore. From the theoretical perspective, the research is performed under a generic framework of spatiotemporal analysis and modeling approaches, which is one direction of geospatial techniques applied to support urban design and planning. This research aims to propose integrated spatial analysis methods to explore the potential uses of big urban data.

3.1 Problem Statement and Hypothesis

The research question is developed from two points of view:

*Research question 1* focuses on understanding urban dynamics and is formulated as:

Assuming that interactions between land use and transportation is an important factor that shapes the changing urban spatial structure, and the effects of such interactions reflect urban activities and mobility, then:

Can urban changes driven by interactions between land use and transportation be detected from urban activity and mobility data by certain geospatial techniques?
If yes, what is the generic framework and what are the possible spatial analysis methods?
If no, why?

This research aims to improve our understanding of urban dynamics in terms of interactions between urban elements, particularly land use and transportation. This study gives an alternative view of these interactions than the related research typically expresses. Unlike empirical
research focusing on how to develop certain urban forms to constrain or guide urban movement, this study focuses on the spatial structure and functions emerging from how people use and move in real urban space. As shown in Figure 3.2, this research completes the interaction circle between space and people by analyzing urban activities and movement patterns in reality, and the reshaped spatial structure is revealed from these patterns. It represents an important way to examine the impact of infrastructure development on peoples’ lives and, in return, how cities have been reshaped by individuals’ needs to travel.

![Image](image)

**Figure 3.2:** Complete loop of land use and transportation interactions.

Note: The dashed line shows the position of the research in this dissertation - the spatial analysis of urban movement for understanding the dynamic interactions of transportation and land use changes.

In line with such thinking, a few questions can be developed. For instance, what is the spatial structure of urban movement today? Is it the same as in our plan? Have new centers and borders emerged from the way people use the space for their daily activities? Are these borders the same as the administrative borders? Such unanswered questions motivate this study. The answers to these questions will be very valuable to planners for validating their designs and developing a better sense of the implementation process.

Besides insights into urban dynamics, experience can be gained from how geospatial techniques support urban studies. Therefore, from the perspective of computational design, this research tackles the big data challenge, which is as crucial as understanding urban dynamics. Over the last few years, data has become ever cheaper, larger, and in higher spatiotemporal resolution. Urban sensor data provides a direct way of investigating urban issues and will gradually change the ways of urban studies. In line with such thinking, this research asks a more general
research question as follows:

Research question 2 focuses on facilitating geospatial techniques in urban studies and is formulated as:

**How can geospatial techniques be improved to better support urban design and planning tasks in terms of using big urban data?**

In fact, this general question can be reformulated and applied to all information technologies in supporting urban studies. This thesis narrows its scope to geospatial techniques.

### 3.2 Research Aims

To answer the two research questions, this thesis presents a spatiotemporal analysis and modeling approach that makes use of large data sets. Specifically, it develops advanced spatial analysis methods that can be applied to urban transportation data to gain insights into urban phenomena generated by human activity and human mobility. The essential idea embedded in this approach is integration in terms of integrated qualitative and quantitative analysis. Integrated spatial analysis algorithms are explored as a solution for solving interdisciplinary problems. Such an integrated approach to urban analysis can explicitly identify ongoing urban transformations. The aim of this research can be broken down into the following targets:

1. To review the state of the art on related research and methodology to understand urban processes, especially polycentric urban transformation are described in Chapter 2.

2. To propose a generic framework that facilities geospatial techniques to be used as a support tool for urban design and planning processes. Based on this framework, a work flow for detecting urban changes can be derived.

3. To develop advanced geospatial analysis methods to extract changing patterns of urban activity and mobility using transportation data from different years.

4. To define new indices of changing urban functionality, land use mixing, and spatial interaction to measure urban transformation.
5. To develop a framework of visual analytics tools based on the proposed analytic method to support decision making.

6. To conduct practical analysis through a case study of Singapore using real data. The applied methods and results can be used for reference for future research.

In sum, the contributions of this research are two-fold. On one hand, it proposes new analysis and modeling approaches for integrating knowledge and technologies to enhance our understanding of urban dynamics. On the other hand, it develops advanced approaches for urban studies based on the spatial analysis method. The stated objectives will be fulfilled in later chapters.

3.3 Method: Spatial Analysis and Modeling

The elaborated research aims to guide through the practical research task that methods are selected based on these defined aims. In this section, a spatial analysis and modeling approach is presented, highlighting the core idea about providing different levels of urban data analysis to support urban studies.

The method used in this study is inspired by the new concept of ‘Geodesign’, which comes with the main idea of integrating geographic science with design, resulting in a systematic methodology to support urban design decision making. In line with the essential idea and to put the concept into practice, this study expands the role of geospatial technologies in supporting urban design work-flow by making extra value of the data. In particular, three levels of data services are provided: i) reduction: reducing the complexity of the data set by data processing; ii) induction: analyzing the data to produce aggregated information; and iii) deduction: using existing resources to extract information for impact assessment or even prediction.

Figure 3.3 shows how this idea of data service can be plugged into a simplified urban design process. It should be noted that the urban design process is never simple, as it involves different levels of design and needs iterative revisions. The diagram shown here is a generic demonstration of essential features and might differ from real cases since the design details may vary a lot individually. The design and/or planning work-flow could be supported by a geospatial pipeline, based on key concepts in Geodesign, but redefined to present data related functions. At the first stage, this pipeline provides database management functions such as data processing, providing
simple query functions, and sampling and formatting functions. At the second stage, it transforms the original data into meaningful information by the integrated spatial analysis method. Finally, a decision support tool, such as a visual analytics tool, could be implemented based on the conceptual models as well as the extracted information in previous steps. These tools could interactively and operationally support the urban design process by providing real-time analysis results. One may notice that data analysis and data modeling are combined in the second stage. This is because, in this research, the data is analyzed using an analogy model where analysis and modeling are closely combined into one step.

Figure 3.3: A generic framework (bottom) associated with an urban design and planning process (top).

The state-of-the-art GIS tools already meet the requirements for data service at the first level. Quite a few examples reviewed in Chapter 2 belong to the second level; however, they do not analyze the issue of urban transformation using transportation data, let alone use smart card data which has become available only recently. The next chapter presents this dissertation’s research agenda which applies such a framework to the issue of urban processes.
Chapter 4

Framework for Measuring Functional Polycentricity

This chapter details the methodology used to extract value from urban data, especially newly available big transportation data, to give insights into urban change. It addresses research aim 2 stated in the previous chapter - “To propose a generic framework that facilitates geospatial techniques to be used as a support tool for urban design and planning processes.” The framework presented in Section 4.1 follows the general approach of spatial analysis and modeling, which has been adjusted to fit into the context of measuring urban changes. The applied framework also serves as a research design that guides the practical analysis work conducted in the rest of this thesis.

This framework can be broken down into its individual parts. The key innovation in the adapted approach is the advanced spatial analysis methods that provide different levels of data service using historical transportation data. These methods fulfill research aim 3, features of such methods are introduced in Section 4.2.1. Moreover, these methods are quantitative ones that measure the aspects of urban changes with defined indices, which fulfill research aim 4 as introduced in Section 4.2.2. To convey the extracted information to designers, we present a framework of visual analytics tools that embed the analysis methods into interactive visualizations, which allows users to explore data in different level of aggregations. This visual analytics framework fulfills research aim 5, which is presented in Section 4.2.3.

The methodology introduced here only describes the integration of the different parts of the
work contributing to research aims set in previous chapter. The complete methods employed in each part of the work will be introduced in the proceeding sections in Chapter 5 and be illustrated by a case study of Singapore to fulfill research aim 6.

4.1 Research Design: An Applied Framework

Figure 4.1. shows how the presented spatial analysis and modeling method can be applied to a specific issue of urban processes. According to the defined geospatial pipeline in Section 3.3, the data first goes into the data processing section to be cleaned up and reformatted. This step is only able to reduce the size of the data and does not induce any information from the data. Next, the data goes into the core part, which is the data analysis and modeling process. A formal model will be decided according to two criteria: the objective (what kind of urban phenomena) and the availability of the data. The models could be mathematical, formal, or conceptual, all of which are types of representations of urban space. When these models are equipped with real data, the variables will be calibrated and the properties can be computed.

From the applied framework, one can see that urban data is separated into two parts. On the right side is the “conventional” urban thematic data, which is represented by land use plans and transport plans that give insights to the development of the built environment, national statistics data for evaluating changing urban populations, and growth economics indices. This dissertation labels these as “explicit” data because they are collected on purpose and information can be gained in a straightforward way. However, these urban thematic data are mostly associated with the physical development of the built environment. According to the definition in the literature review, they tell morphological Polycentricity of urban stocks, however, does not give little profile of Polycentricity in socioeconomic space.

Therefore, as supplementary information, on the left side is another part of the data set - large urban mobility data - which gives a picture of how people live and travel in cities. More and more urban mobility data are available these days; however, we lack an advanced analysis method to make sense of the data in an urban context. The goal of this research is to detect functional urban changes in terms of travel behavior, urban activity patterns, and urban movement patterns from such data sets. This change explicitly represents how people change their lifestyles to adapt to built environments and, in return, reshape the urban space to meet their individual demands in reality. This represents functional Polycentricity. This research
makes a contribution to the use of urban mobility data.

By linking functional changes detected from urban mobility data and morphological changes detected using thematic data, a complete picture appears. This picture shows us the compatibility of the original plans and reality. Furthermore, it shows us the interactions between the built environment and people and how land use and transportation together exert an influence on urban activity and mobility. In a broader sense, such mobility data is only one type of data set that represents big urban data. Big urban data is either too massive to be managed by data management tools, such as MySQL, and/or does not contain any implicit urban information because it is not meant to be used in certain ways. The work conducted here illustrates the idea of data innovations reviewed in Section 2.4 and shows how geospatial analysis can be advanced for the age of big data to better support urban design and planning tasks.

Figure 4.1: Framework for detecting functional urban changes.

Note: The most essential part is the analysis and modeling method that can be applied to transportation data.
4.2 Detecting Functional Urban Changes Using Transportation Data

As in indicated in the presented research agenda, the main contribution of this dissertation are spatial analysis methods that can be applied to urban transportation data of different years to measure polycentric spatial structure, thus to detect functional urban changes. Specifically, Section 4.2.1 will give more details about the key features of advanced spatial analysis methods. Following that, Section 4.2.2 will show how the method measures urban change using derived urban indices and Section 4.2.3 will show how the analysis method can be further developed as a support tool.

4.2.1 Spatial Analysis Methods

This thesis shows that urban mobility data can be used to analyze travel behavior, activity patterns, and movement patterns as shown in Figure 4.2. Different levels of data service are shown in the defined task. As indicated earlier, this section explains the features of data service. The detailed implementation will be presented later in Chapter 5.

![Figure 4.2: Spatial analysis of urban mobility data.](image)

Deeper information can be extracted by data analysis, mining, aggregation, and modeling. Different levels of data service output different degrees of abstractions of data, which are associated with the following questions:
Q1. To what degree can digital tools help reduce the complexity of data by simplifying and organizing massive data sets?

Q2. To what degree can digital tools help reduce the complexity of urban phenomena by analyzing and reasoning the information?

The answer to the first question can be easily found in database management software, which provides basic functions like indexing, querying, and data editing. Geospatial tools provide additional spatial operations, such as spatial data joining by location. This level of data aggregation can effectively filter unimportant details, reformat data, and sample data sets, but it cannot transform data into information.

The answer to the second question requires data analysis, even data mining methods. New properties that are beyond the original properties in the data set should be defined and computed. For example, counting clusters of people from census data, and then determining where the clusters are and the distance between the clusters. Here, travel behavior is analyzed by simple statistics.

Furthermore, this research aims to find extra value in urban data. Therefore, mining out the implicit patterns is the main task. An analog model will be developed that uses a representational or functional form of certain systems and applies to certain kinds of urban phenomena. In the case of analyzing activity and movement patterns, the models are developed with defined indices to find activity clusters and to measure the spatial distribution of clusters and boundaries of movements. Another definition given here is urban modeling. As discussed earlier, modeling is deeply rooted in all of the analysis methods in this research. Based on [19], this research redefines “urban modeling” based on the specific context of this research as: a spatial analysis and modeling approach used to define a proper formal model, which can be used to represent urban space, and is calibrated by large temporal location data. The properties of the model computed using large data sets can be used to explain urban processes.

4.2.2 Urban Indices for Quantitative Analysis

To compare urban changes over the years, this dissertation defines urban indices for qualitative urban analysis, which results in a better explanation of computed properties.

Table 4.1 shows the main analysis conducted in this study. The second and third analyses
are closely related to the changing urban structures that will be represented in the next chapter. The quantitative approach proposed in this research uses spatial analysis as its base, extending the traditional method with probability statistics, machine learning techniques, and complex network analysis to compute the urban data in different spatiotemporal scales. Enhanced by urban planning knowledge, the outcome parameters are interpreted to identify urban problems, such as traffic congestion, shrinking market areas, and so on.

The approach in this research can be summarized by the following four steps:

1. Set a goal (urban issue), initiate a proper model, and design a data structure.
2. Define indices that are properties of the model and its measurements.
3. Measure the indices using large data sets.
4. Make sense of the measured properties by linking them to facts in reality.

In the next chapter, these four steps are implemented in a case study of Singapore. This dissertation makes insights into the use of urban space by mining transportation data, including surveyed data and smart card data, which reflects people’s daily travel behaviors. These travel behaviors are considered a function of urban functionality, spatial interaction, and spatial structure of centers and borders, which are all elements of land use planning.

Table 4.1: Analyses applied to urban transportation data sets.

<table>
<thead>
<tr>
<th>Data</th>
<th>Integrated Method</th>
<th>Urban Model</th>
<th>Scale</th>
<th>Subject</th>
<th>Index</th>
<th>Impact Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveyed data + Smart-card data</td>
<td>Spatial statistic and probabilistic model</td>
<td>Activity model</td>
<td>Small</td>
<td>Urban function</td>
<td>Urban functionality, Land use mixing</td>
<td>Traffic congestion ...</td>
</tr>
<tr>
<td>Surveyed data</td>
<td>Spatial analysis and clustering method</td>
<td>Central place theory model</td>
<td>Medium</td>
<td>Spatial interaction, spatial structure</td>
<td>Density, Diversity, Centrality, Attractiveness</td>
<td>Market area analysis ...</td>
</tr>
<tr>
<td>Smart-card data</td>
<td>Spatial analysis, complex networks analysis</td>
<td>Network model</td>
<td>Large</td>
<td>Spatial interaction, spatial structure</td>
<td>Connectivity, Closeness, Clustering</td>
<td>Segregation, Census ...</td>
</tr>
</tbody>
</table>
4.2.3 A Visual Analytics Framework

Following up with the two questions given in Section 4.2.1, this chapter poses a third question regarding data use:

**Q3.** To what degree can digital tools help reduce the complexity of the urban design process by using and activating information to generate future scenarios?

The answer to the third question requires real-time feedback tools that offer certain predictive functions indicating the impact of urban design proposals. Simulation tools such as MATSim\(^1\) and UrbanSim\(^2\) are along this line. However, this kind of simulation platform needs costly computing resources and time, and the complex models need massive data sets to calibrate and verify them. Since this dissertation focuses on impact assessment using data analysis instead of simulation, a visual analytics tool is presented as an alternative solution. According to the previous definition of urban modeling, models can be formally structured and developed to relevant computer programs, which, in this dissertation, is a support tool for real-time data analysis.

![Figure 4.3: Mechanism of a visual analytics tool](image)

The two main functions to be provided are interactively visualizing the data and the real-time analysis impacts of modifications on urban plans and/or transport plans. As shown in Figure 4.3, it is quite similar to the analysis and modeling steps given before.

---


\(^2\) UrbanSim. A software-based simulation system [http://www.urbansim.org/Main/WebHome](http://www.urbansim.org/Main/WebHome) accessed in 2013
1. First, the original data set is enriched by semantics defined according to design goal and represented by an urban information model.

2. After computation by the analysis algorithm, the values of the properties come out along with aggregated data sets.

3. The algorithm can be applied to data in different scales.

4. By graphic representation, the visual analytics tools output a context-based visualization.

Linking theory with practice, the following chapter explains how to further implement the analysis into a software tool and its functions. The software implementation is a translation from a theoretical model to a programming language. An object-oriented language has advantages in describing complex objects and processes. The data structure of urban elements and their attributes are essential parts of the proposed visual analytics tool, as shown in Figure 4.4. This data structure focuses on transportation data analysis. There are four elements as follows:

- People: “who” is the object that performs activities and travels.
- Trip: “how” the state of people changes (location change) and is motivated by certain activities.
- Activity: “what” is the event that happens in a spatial location.
- Place: “where” the activity occurs.

Besides “People”, which only has social attributes, the other three objects have spatiotemporal attributes, computed attributes, and geometric attributes for graphic representation. “Place” has four derived classes, which are associated with different spatial scales.

Since the objective of this research is to understand collective effects, such as space use, a person is not considered as an independent element in the analysis model. The other three elements, namely “Trip”, “Activity”, and “Place”, correspond to models A, B, and C. Here, we represent the model in a very generic format. A set of computing methods is defined that feedback the value of the properties, such as indices and aggregation level, to each object.

This visual analytics tool builds a simple work-flow of data processing and makes it possible for general users to explore large data sets and understand the data sets by reading extracted
information. A real-time analysis could also be done on the modified data sets for impact analysis. Based on the model built of whole original data sets, users can partially modify the data to obtain real-time analysis results. For example, a planner may want to know how the global distribution of people changes when accessibility to one area increases. He could modify the traffic flow to one area and the visual analytics tool will automatically re-compute the centrality of all areas.

4.3 Chapter Conclusions

This section presented the research methods used to answer the stated research questions in Chapter 3. Explanations have been given regarding the following subjects:
• Outline of geospatially-aided urban design and planning work-flow, which is developed based on a spatial analysis and modeling approach providing levels of data services. The identity of such an approach is (1) to make full use of large data sets, which contain rich information that is rarely mined out; and (2) to provide urban related information in an explicit way.

• Research design, which applies the generic work flow to the practical study conducted in this research. This research design will guide the analysis of urban changes in Singapore in the next Chapter.

• Key feature of analysis methods applied to mobility data, which is the highlight and main technique contribution of this study.

This section presents the methodology in a very generic form since the framework can be re-formatted and applied to other urban study applications. To further show the feasibility of the proposed methodology and its practical applicability, complete methods employed in the urban study of Singapore is introduced in the proceeding sections in Chapter 5.
Chapter 5

Functional Changes in Singapore

In this chapter, the proposed framework is applied to a case study of Singapore. On one hand, it is intended to implement the proposed methodology into practical to show its feasibility. On the other hand, insights into decentralization development in Singapore will be gained through the analysis. The organization of this chapter follows the research design presented in Chapter 4, including reviews of physical development in Singapore and an analysis of functional changes in different scales using transportation data from Singapore. The conducted analysis covers both physical and functional development in Singapore, from individual to aggregated levels; the logic of the analysis is shown in Figure 5.1.

Figure 5.1: Organization of sections in this chapter.
The structure of this chapter is explained as follows:

Section 5.1 gives a very brief introduction of the case study area Singapore and the study materials.

Section 5.2 reviews physical development in Singapore from the perspective of historical urban plans, transport system development, and growing economic activities. The study materials are related literature and national statistical data, which were defined as urban thematic data in the previous chapter. They give explicit facts of the changes of the built environment, which exert certain influences on urban activities. Thus, they will be used in later sections to explain the possible causes of detected functional changes.

Section 5.3 is the start of the transportation data analysis. Patterns of travel behavior at individual level are investigated by simple statistics and data mining methods. The conducted analysis incorporates human behavior into transport analysis by looking into patterns associated with different types of urban activities, resulting in a better profile of the impact of urban functions on daily traveling. Both travel survey data and smart card data are used, therefore more details of the data can be gained. An application of data fusion is also given, showing the potential of using massive smart card data in an innovative way.

Section 5.4 looks into patterns of urban activities. The conducted analysis shifts from individual to aggregated level. A new measure of urban centrality using travel survey data by integrated analysis method is presented. It follows the arguments in literature review that Polycentricity should be measured from (1) how people use urban space in reality; (2) all types of activities rather than “journal to work”; (3) the degree of spatial clustering and the distributions of clusters. By comparing the analyzed results from three years of data, the path of urban changes can be traced.

Section 5.5 studies patterns of urban movement. It is one step further following the most critical argument of Polycentricity that Polycentricity is not only about urban stocks but also about urban flows. Functional Polycentricity is concerned with how centers are connected and how evenly connected. To measure the spatial structure of urban flow, a spatial network model is constructed from urban travels using smart card data. Human movements are used as a proxy, or physical carriers, of urban flows. Thus, spatial interactions between urban areas can be represented by properties of the spatial network, which are measured and used as urban indices to analyze urban changes.

Section 5.6 introduces a visual analytics framework, which implements the analysis method
used in previous section into a visual analytics tool. A prototype of flow map is implemented as a proof-of-concept tool. It shows that the analogy model used for analysis can be further calibrated and developed by computer programs as defined in urban modeling. This kind of visual analytics tools are also representatives of higher level of data services that make geospatial techniques an impact assessment tool to support urban design and planning.

Section 5.7 is a short discussion about the feasibility of presented methods, their merits, drawbacks, and potentials.

5.1 Study Area and Data

5.1.1 Case Study Area: Singapore

Singapore is an island city-state in Southeast Asia with an area of 710.2 km$^2$ as shown in Figure 5.2. The state as existing today does not have a very long history. Singapore gained independence as the Republic of Singapore on 9 August 1965. Everyone who was present in Singapore on the date of independence was offered Singapore citizenship. The current population of Singapore in 2014, including non-residents, is approximately 5 million. It is expecting to have a population of 5.8 to 6 million by 2020 and 6.5 to 6.9 million by 2030 [112]. In the past decades, life and the living environment in Singapore have changed dramatically. Singapore has transformed itself from a declining trading harbor to a First World economy [69]. And its fast development is still ongoing.

5.1.2 Study Materials

The success of this research depends heavily on the availability of the data sets. Singapore has a well-recorded history, which supplies rich materials for this research. Besides, it is a developed country that applies relative advanced sensor techniques to collect large data sets like smart card data. The main data sets used in this research are provided by Singapore government agencies, including the Urban Redevelopment Agency (URA), Land Transport Authority (LTA), Singapore Land Authority (SLA), and Housing Development Board (HDB). A few data sets are self-collected from open data sources such as Open Street Map, and related literature.

The data used for analysis in this research are categorized into two groups as previously defined: thematic data about the physical built environment provides explicit facts regarding
changes in urban space; and urban transportation data provides information about peoples’ daily movements and activities. These two categories of data are analyzed together under the proposed methodology, which was intentionally designed to understand the interactions between built-environment and people as shown in Figure 5.3.

**Urban Thematic Data**

Urban thematic data is used for understanding the physical development of urban space. Diverse data sets are referenced:

1. Geo-referenced data sets, which mainly include master plans over the years. These can be downloaded from the Singapore Urban Redevelopment Authority (URA)’s official website\(^1\); road network data, shown in Figure 5.3 (first layer), building footprints data (second layer), and some point of interest data collected from OneMap\(^2\).

---


\(^2\) OneMap integrated map system of Singapore [http://www.onemap.sg/index.html](http://www.onemap.sg/index.html) accessed in 2014
Note: Urban movement data, mainly transportation data represents human movement and urban thematic data represents physical development of urban space.

2. Post processed geo-referenced data sets, such as census data, which were originally in sheet files, and were later enriched with geo-references in preliminary data processing. Most of the statistical data was collected from Singapore national statistics\(^3\).

3. Non-referenced data including statistics data are mainly obtained from online open resources, media profiles, literature reviews, and reports.

**Urban Transportation Data**

Urban movement data is location data that is used for analyzing travel behavior, human activities, and movement patterns. The focus of this research is to mine the implicit insights of urban changes from such location data. In particular, the data used as inputs are:

1. Surveyed data from three years: A Household Interview Travel Survey (HITS) is conducted by LTA every four to five years to give transport planners and policy makers insights into residential travel behavior. About 1% of households in Singapore are surveyed,

---

Table 5.1: A sample of household travel survey in Singapore with selected information

| id | age | origin postcode | des postcode | start time | arrival time | activity place | activity | travel mode | ...
|----|-----|-----------------|--------------|------------|-------------|----------------|----------|-------------|------
| 1  | 40  | 5****6          | 5****3       | 6:25       | 9:15        | clinic         | work     | bus         | ...
| 2  | 69  | 5****3          | 5****6       | 9:30       | 12:15       | home shops     | go home  | bus          | ...
| 3  | 40  | 5****6          | 5****8       | 12:30      | 14:00       | shops          | shopping | walk         | ...

Table 5.2: A sample of smart card data in Singapore with selected information

<table>
<thead>
<tr>
<th>jour id</th>
<th>card id</th>
<th>card type</th>
<th>mode</th>
<th>boarding stop id</th>
<th>alighting stop id</th>
<th>start time</th>
<th>trip dis (km)</th>
<th>travel time (min)</th>
<th>fare (s$)</th>
<th>transit count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9**1</td>
<td>adult</td>
<td>train</td>
<td>STN Sengkang 64041</td>
<td>STN Hougang 67009</td>
<td>8:30</td>
<td>2.4</td>
<td>6.417</td>
<td>0.23</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>9**2</td>
<td>senior</td>
<td>bus</td>
<td>67009</td>
<td>59079</td>
<td>13:30</td>
<td>4.6</td>
<td>16.667</td>
<td>0.91</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>9**2</td>
<td>senior</td>
<td>bus</td>
<td>59079</td>
<td>67009</td>
<td>14:30</td>
<td>14.2</td>
<td>24.333</td>
<td>0.7</td>
<td>1</td>
</tr>
</tbody>
</table>

with household members answering detailed questions about their trips. The HITS results provide very detailed information including age, occupation, travel purpose, travel destination, walking time, waiting time, traveling time and so on. Table 5.1 is a sample of the surveyed data. Only closely related information is referenced in that table. This paper uses the HITS results of 1997, 2004, and 2008. A report of HITS can be referred [34];

2. Smart card data collected in periods of three years. The smart card data is collected by a fare collection system, which is used in Singapore, and has been gradually adopted by public transit agencies in many countries. While the main purpose of these systems is to collect fares, they also produce large quantities of records on daily traveling [109]. The recorded smart card data contains detailed information on each trip. The data used in this research includes trip id, passenger id, age, boarding and alighting time, boarding and alighting location, distance, fare, and an index associated with transfer trips as shown in Table 5.2. Over half of the population in Singapore are using the public transportation system daily, generating more than 5 million travel records per day. In total there are more than 4700 bus stops and MRT stations covering the whole geographical land area of Singapore as shown in Figure 5.4. This research was conducted using the available smart card records over three sets of workdays in September 2010, April 2011, and September 2012. Some analyzed results from literature are also referenced, like [132], in which, one
week of smart card data in 2008 in Singapore are analyzed.

As presented before in Section 4.1, data processing is the first step in the geospatial pipeline. Therefore, most of the data are geo-referenced and organized in spatial databases. Managing data in one geospatial platform is also an efficient way for data representation and sharing. Extra databases, like MySQL, are used for storing very large data sets such as smart card data. This section only gives a very brief summary of the data sets used in this research. Details about data and techniques for data mapping, structuring, processing as well as analysis will be presented in the following sections.

Figure 5.4: Bus stops and train stations in Singapore.

5.2 Five Decades Fast Development in Singapore

This section gives a more detailed introduction of the physical development of Singapore from three perspectives: its historical urban plans, development of the transport system, and geography of economic activities using urban thematic data. The purposes of the historical review are twofold: (1) a more detailed introduction of Singapore and (2) an analysis of physical urban changes, which will be linked to detected functional changes in later sections.
5.2.1 Historical Urban Plans

“Cities are not designed by making pictures of the way they should look in 20 years from now, they are created by a decision making process that goes continuous day after day.” - Jonathan Barnett

Singapore is claimed as a model city, which successfully transformed itself economically into a first world economy after decades of efforts [69]. Its long-term urban development plans definitely contributed to its success. Since attaining independence in 1965, Singapore has undergone huge changes in its built environment. Many urban development problems often encountered in rapid urbanisation, such as adequate housing and infrastructure, have been solved successfully. As said in [160], “it is a planned city, a result of ‘deliberate urbanisation’ (McGee 1972) where urban growth is managed and made as productive as possible according to its government’s conception of economic, political and social well-being of its inhabitants.” A brief review of the most influential historical urban plans of Singapore is given below.

Phase 1 - Early plans

The Jackson Plan or the Raffles Town Plan, drawn up in the 1820s, could be one of the earliest town plans of Singapore. Its pattern of distinct residential districts for different ethnic groups of settlers became a basis for the later growth of the central area, and its impact is still obvious today. However that plan is just a town plan, since the rest of the territory is simply ignored.

Phase 2 - Starting long-term planning

Throughout most of the 19th century and for the first half of the 20th century, Singapore’s physical growth was haphazard and largely unregulated. It was only in the mid-1950s that Singapore truly began its long-term planning, and the result is that Singapore became the city-state that the whole world sees today. The concept plan, which is the macro-level blueprint, had significant impact on shaping the spatial structure of Singapore.

In 1958, the Master Plan was adopted by Singapore, influenced by a British notion of order and regularity and modern town plans. A sign of decentralization had already appeared there. A green belt was proposed, to arrest the continued expansion of the central areas and to take urban settlement outside the existing city to new towns. However UN consultants and Singapore
Government soon rejected this plan, because the Singapore Government wanted to pursue a drastic transformation of the city-state rather than a slow and steady rate of social and economic changes [160].

**Phase 3 - Structuring the space**

Though the urban plan was rejected in 1958, its essential idea about new towns exerted influence in later plans [47]. The Concept Plan of 1971 adopted the “Ring Concept Plan” as shown in Figure 5.5. It is outlined to functionally link the whole island by a dense network of communication lines between new towns, as well as other active sectors. Meanwhile, a detailed plan was made for central areas to enhance their function as financial districts. This plan produced longstanding impacts on land use development in Singapore. From 1971 to 1990, the plan was implemented. During that period, land use share was dramatically changed, such that land use for residential and industrial purposes, and especially transportation, were all increased. Many large scale residential houses as well as retail units and offices were built. The population of the central area declined as well. Decentralization steadily emerged.

![Figure 5.5: The revised Concept Plan in 1971.](image)

Note: Image is recreated from [47].
Phase 4 - Decentralization urban planning

In 1991, a revised version of the concept plan was released, which is also the most referenced one that significantly shaped the structure of Singapore and projected into the future beyond 2010 (shown in Figure 5.6). It proposed a development strategy involving the decentralization of the present central area to regional centers and other functional centers. The idea was to reduce the space demand on the central area and reduce commuting time, and in the long run, to achieve a balanced distribution of industry for further growth. A city was planned, with a hierarchy of functional centers. The old central areas were to be surrounded by 4 regional centers, 5 sub-centers and 6 fringe centers. A later concept plan gave more detailed guidelines to each specific space and promoted the development of sub-centers.

In subsequent years, after decades of development, Singapore was awarded a high rank (4th) among world cities. Greater competitions came along. “The first level of competition is from outside cities and countries. Second level is from inner towns, which is between central area and the outlying new towns since higher level retailers which used to be in the central areas were moved out to the Orchard tourist zone or new regional centers.” [154] A polycentric urban form started. This concept plan is considered to have greatly influenced the spatial structure of Singapore, and is linked to the enhancement of quality of life.

Phase 5 - From developing infrastructure to improving life quality

The revised Concept Plan in 2001 was intended to develop Singapore towards being a thriving world-class city in the 21st century. They sought to transform Singapore into a global financial hub by setting aside land in the city center to support the growth of the financial and service sectors. One new focus was to enhance Singapore’s natural and built identity so as to create a distinctive city with rich heritage, character, diversity and identity. In early 2000, the Urban Development Authority re-designed Singapore as a City-in-a-Garden. The heritage and nature resources, such as parks and water bodies, became the focus of this plan.

The Master Plan of 2008 which followed converted the strategies of the Concept Plan into detailed plans to guide Singapore’s physical development. There were four key thrusts that aimed to make Singapore a more livable city - “A Home of Choice, A Magnet for Business, an Exciting Playground, and a Place to Cherish”\(^4\). The latest review of the concept plan was

\(^4\) Master plan  http://www.ura.gov.sg/uol/master-plan/View-Master-Plan/
in 2011. The Realized Land Use Plan 2013 focuses on strategies to support population and economy growth, while ensuring a high quality living environment for all Singaporeans.

This historical review of urban plans is highly important to the rest of the research, since the Polycentricity examined in this research was largely shaped by those early plans, together with bottom up changes that gradually emerged in later periods. The urban plans of Singapore are done at different levels of details, and projected into different spatial scales. The research puts more focus on concept plans, because concept plans are strategic plans which give flexible frameworks for action including vision, goal, and objectives in a long run. The concept plans provide very strong driving forces for urban transformation, which continue for decades, while the master plan provides the framework for the regulation and coordination of physical development, including detailed land use that currently exist and that which will be changed in the future. In the case of Singapore, it is required to be reviewed every 5 years. Detailed design plans give very detailed physical instructions at district level, including spatial configurations, mix-used in complex, faade design, and pedestrian ways and so on. This kind of design contains so many aesthetic factors that regularity can barely be extracted.

Note: Image is recreated from [47].
5.2.2 Transport Development

Transportation has a strong influence on the spatial structure at the local, regional, and global levels [120]. Cities have traditionally responded to growth in mobility by expanding the transportation supply, by building new highways and/or transit lines. In the case of Singapore, this strong influence is very obvious in urban development. As said in [160], enhancing mobility and accessibility is considered as one of the key issues in Singapore’s sustainable planning. The early development of the transport system has been identified in [36] and summarized in three periods of planning: early 1960s - little or no systemic planning; 1960s to 1980s - early period of planning but mostly problem-driven; since 1990s - vision-driven planning. This review discusses the role of transportation systems in the later period, focusing on its role in shaping urban structure.

Phase 1 - Linking the city hubs

Urban planning and transportation planning have a strong influence on each other, and visibly impact Singapore’s urban development through a tight planning system that is closely linked to the location choices of housing and industry. From the 1970s, transportation has been prominently considered in shaping the structure of the city. According to the concept plan, high-density public housing areas are arranged along proposed high-capacity public transportation lines while low- and medium-density housing is next to the corridors and served by a road-based transportation system. Industrial areas and other employment centers are located close to public transport.

Phase 2 - Facilitating urban mobility

The development of a public transportation system has undoubtedly increased the accessibility of Singapore. In 1987, first line of the Mass Rapid Transit (MRT) system in Singapore was initiated. The system now covers 102 subway stations, with particularly fast development of the system during the last 5 years with several new lines opening. Today, the land-based public transportation system in Singapore comprises two networks: the MRT system and the bus system. More than half the population is now using public transportation as their main transport mode [34].
Phase 3 - Integrated plans for a more livable city

A clear trend that can be seen from the development of land use and transportation in Singapore is that, livability became more and more important. After meeting the basic demand, pursuing a higher quality of life and more people-centered plans becomes the next goal. To achieve this goal, new challenges have been identified for future development. In LTA's vision of a people-centered land transport system, there are three key strategies, namely, making public transport a choice mode, managing road usage, and meeting diverse needs. In these strategies, integrating transport and land use planning has been emphasized in terms of integrating transport facilities with building developments, working closely with other agencies to integrate transport with land use planning.

5.2.3 The Geography of Economic Activities

Another important factor that led to Singapore’s success is its economic development strategies. These strategies guided the development of functional zones such as industrial zones, commercial centers, financial centers and mixed use areas across the island, which exerted a significant long-term impact on the geography of economic activity in Singapore. Population growth, public housing program and development of urban infrastructure are the three features reviewed here.

![Figure 5.7: Historical populations data from national statistics of Singapore.](image)

Note: Data source is from Singapore Department of Statistics
Population and Economic Development

Singapore’s first population census after independence started in 1970, and was conducted every ten years. The first register-based approach started from 2000. Beyond 2000, the Singapore Department of Statistics established a system of continuous measurement of the population.

According to the annual report from Singapore Department of Statistics, there were 3.31 million Singapore citizens at end-June 2013. Together with 0.53 million permanent residents, there was a total of 3.84 million residents. As shown in Figure 5.7 is the historical population data in the last 12 years. The total population in 2013 registered a 1.6 percent annual growth, while the population of permanent residents had a slightly lower 0.9 percent annual growth. The difference between growing speeds is due to a policy welcoming “foreign talent”, from which a path of economic development can be traced. The word “foreign talent” is used to denote an aggressive immigration program which was intended to attract high-end educated workers to Singapore. It was instituted as a consequence of an inadequate labor supply in economic development of Singapore, during the most recent decades [69].

![Figure 5.8: Percentage change of private sectors over corresponding period of previous year.](image)

Note: Data source is from Singapore Department of Statistics
Dating back to early 1960s, foreign capital came in and changed Singapore’s original trading economy to one that focused on low-end industrial manufacturing. Thus the rate of unemployment was decreased. The issue of labor shortage emerged in the early 1970s. In the beginning, this issue could be managed by importing labor from neighboring countries. But by the 1980s, it became clear that Singapore could not keep its high competitive force due to its small population. An steady evolutionary trend started, to transform the major economy from a low-end industrial one to that of higher technology. This trend became clearer in the 1990s, especially after the 1997s Asian Financial Crisis, that the core of the economy had been shifted to knowledge-based industries such as finance, bio-science, and electronics. Even from the most recent statistics in Figure 5.8, changes can be read showing that manufactures have a decreasing share, and the service industries keep on growing. A consequence of this immigration program is that low-end foreign workers became abundant, and were even perceived as a threat to local people. To respond to the unhappiness of Singaporeans, the intake of permanent residents has been reduced since 2010.

The phase of economic development coming along with different strategies also reflects on the development of urban infrastructure and the location choice of housing and industry, such as the industrial parks built in 1960s and 1970s. Thus, the following two sections will discuss the public housing program which solved the housing problems for the increasing population and also notes some highlights of urban infrastructural development that attracted urban flows in past decades. The current spatial structure of Singapore was greatly shaped by all these aspects.

**Public Housing Program**

The Housing and Development Board (HDB) was established in 1960 to solve Singapore’s housing shortage. At that time, many people were living in unhygienic slums and crowded squatter settlements. Only 9% of Singaporeans lived in government flats. The HDB started by building very simple rental flats to meet basic needs. After five decades of efforts, the HDB has built more than 800,000 flats, which houses about 85% of Singapore’s population. The development of Singapore’s public housing program has gone through many phases to confront the challenges in different eras. The historical materials were collected from the URA annual report, which can be retrieved from the official website, as well as a literature review [35, 46, 22, 69, 44], and briefly summarized as follows:
Phase 1 - Meeting basic needs. The provision of basic, low-cost rental accommodation for the poor was the original concern of HDB. In the first 20 years of the public housing program, HDB aimed to provide new public housing units in the shortest possible time to relieve the issue of over-crowding and poor hygiene in the post-independence period [153]. Some of the buildings, such as the ones in the Tiong Bahru area, still exist today. Launched in 1964, the Home Ownership for the People Scheme gives home-owning citizens a tangible asset and stake in the country, and promotes rootedness and a sense of belonging among Singaporeans, thus contributing to the overall economic, social, and political stability of Singapore.

Phase 2 - Shaping urban space. Coordinating with the urban plans, and also because of the rising affluence, greater social aspirations, and higher expectations for public housing in the 1980s, stimulated the new strategy. Town planning began to consider more factors such as urban form, town structure, and the provision of regional facilities such as parks and open spaces to improve community interactions.

Phase 3 - Upgrading program. Since the 1990s, HDB has adopted a comprehensive estate renewal strategy. Various upgrading programs have been carried out with the aim of improving the living environment of its residents. Smaller-scale programs have also been developed from 1990 to bring the benefits of upgrading programs to more residents. These include the home improvement program, which was launched in 2007, targeting common maintenance problems within the flat such as spilling concrete and ceiling leaks.

Phase 4 - Livable space. The strategy of HDB keeps on changing over time, to adapt to new circumstances. Nowadays, greater emphasis is placed on creating a high quality living environment and building up the identities of precincts, neighborhoods, and towns. New residential concepts such as the “Punggol 21” waterfront town were developed in response to changing lifestyles. Some concepts are highlighted here, all targeted at creating more livable cities. These include visual identity (landmark buildings, landscaping, open spaces and special architectural features were incorporated to achieve a strong identity), more flat types (to provide different age groups alternative housing options), and accessibility (to meet accessibility needs, particularly the older members of the aging population).
The Development of Urban Infrastructure

Singapore’s fast development has been explained as a result of a comprehensive package of strategies [69]. Besides, long-term urban plans and public housing program introduced, a series of economic practices is are indispensable factors. The development of commercial zones, industrial zones, and financial centers exert great influence on the location choice and structure of urban flows. Only a brief review is given, including selected developments. The historical materials were collected from the URA annual report, which can be retrieved from the official website, literature review [160, 69], and online resources. These have been selected for highlighting significant developments in terms of attracting urban flows and re-summarized by the author.

- 1960s - JTC (Jurong Town Corporation) was established. Jurong industrial estate became a self-contained satellite town. The Jurong Industrial Park is the first industrial zone in Singapore.
- 1970s - The waterfront district, which was always a commercial area was expanded, adding a banking and financial district. This waterfront district was originally located at the famous Golden Shoe area.
- 1990s - A set of seven small islets to the south of the main island was reclaimed to constitute Jurong Island, and dedicated to petrochemical industries.
- Early 2000s - Three landmark projects were launched: Singapore flyer, Marina Bay Financial Central and the Marina Bay Sands Integrated resort, which all of which were completed between 2008 and 2010. The development of Marina bay area continues until today.
- 2005 - Orchard Road has been gradually transformed into a street-like shopping environment. Entertainment and art were sited and developed in Bras Bash area; more than 203 unites were approved for conservation.
- 2007 /2008 - The blueprint for the Jurong Lake district was unveiled to transform the area into unique lakeside destinations for business and leisure in the next 10-15 years. Big shopping malls have been built or upgraded to make Jurong area another sub-center of the city.
2010s - Proposals for Punggol Point and the Woodlands Waterfront were made, to enhance the development in northern part of Singapore. Two pedestrian bridges were opened - Henderson Waves and Alexandra Arch, linking up the three hill parks at the Southern Ridges, enabling the public to walk from Kent Ridge, Telok Blangah Hill to Mount Faber. It is another implementation of the “city in garden” concept.

5.2.4 Discussion

This section reviewed the morphological changes of Singapore from the angle of the physical development of Singapore, focusing on three aspects: historical urban plans, transport development and economic geography. They are clear evidence of urban changes, however, don not tell so much about impacts of these physical developments on people’s life styles. Previous work attempts to estimate such impacts in terms of assumptions, modeling, or predictions. The result is hard to be validated. This research argues that human senor data is gradually available nowadays and offers us a straightforward view of life styles which are ground truth.

Therefore, from the next chapter, functional changes in terms of human behaviors are analyzed using transportation data. It traces the urban changes from another angle. When linking these two viewpoints - physical development and human mobility and/or activity together, urban plans can be evaluated, impacts can be assessed, and knowledge about human behaviors can be gained.

5.3 Statistical Analysis of Travel Behavior

This section investigates travel behavior at the individual level using both surveyed data and transportation data. A set of statistical analyses are conducted for three main purposes. Firstly, more details about data can be gained before diving deeper into the more complex analysis in the later sections. Secondly, the most straightforward way of using data is to find changing patterns of individual travel behavior by statistical analysis. The changes reflect the impact of changing urban infrastructure on people’s daily activities. Finally, both types of urban mobility data - travel survey data and smart card data are analyzed and the results are compared. An application about data fusion is given at the end of the chapter as an example of a data innovation. The goal is to explore the potential of easily collected smart card data used to analyze travel behavior, furthermore, for urban studies.
By incorporating human behavior and social impacts into the transport and urban analysis, three questions relating to people’s daily lives are raised and discussed. These questions form the basis of the analysis: (1) Travel behavior: how do people travel? (2) Travel purpose and urban activities: why do people travel? which is about travel purpose/urban activities; and (3) Location choice: where do people go? These three questions are answered by separate analyses of travel survey data and smart card data. Both analyses have their own strengths and weaknesses. To highlight the idea of data innovation promoted in this dissertation, an extensive application fusing two data sets to enrich information by an inferred method is presented. Discussions regarding the analysis method and findings are given at the end of this section.

5.3.1 Statistical Analysis of Travel Survey Data

As introduced, the household interview travel survey (HITS) offers insights into residents’ travel behavior. In Singapore, HITS is conducted every four to five years. About one percent of all households in Singapore are surveyed. A more complete introduction can be found in the official reports (e.g., [34]). The official reports focus on travel mode and total travel demand, which are also included as part of the analysis in Section 5.3.1. Beyond that, a further analysis is done to examine the varieties of travel behavior corresponding to different activities. The reason for such an analysis of different activities lies in the new definition given to Polycentricity, as discussed in Chapter 2. Previous related work measured the spatial structures and spatial interactions mainly in home to work journeys. However, nowadays, the development of the built environment and increasing amounts of disposable income enable people to have more diverse lifestyles. The “Journey to work” is no longer the dominant motivation of travel. Other activities, such as traveling for education or entertainment, are playing the same important role as going to work. Therefore, the travel behaviors for different activities are compared to create a more detailed profile of the impact of transportation on people’s daily lives.

Travel behavior

Travel survey data from 1997, 2004, and 2008 is used in this section. Since the data from 1997 is not complete, only a partial analysis can be conducted. As indicated earlier, travel survey data contains a lot of social information such as income level, occupation, and education. However, this social information is completely absent from smart card data, therefore, some analyses is
not applicable. In this section, five types of analyses are conducted and the results are compared.

(1) An overview of trip generation

The surveyed data from 1997 contains the addresses of the trip destinations, which can be geo-coded, and the travel survey data from 2004 and 2008 provides the postcodes of the trip origin and destination. Therefore, it is possible to create a geographical map of all the activity locations. Table 5.3 shows an overview of the travel distances and activity locations. As introduced earlier, the idea of “satellite towns” in the “Ring Concept Plan” of 1997 was to develop self-sufficient towns. Similarly, decentralization in the concept plan of 1991 was to build hierarchical urban centers that reduce the demands on central areas so that Singaporeans spend less time commuting. The set of maps shown in Table 5.3 provides a visual impression of the distributions of trips as well as a rough idea of spatial clusters in Singapore.

The Euclidean distance (point to point distance) follows almost the same distribution over one decade (three surveys), although the average travel distance increases evenly. When dividing the trips into multiple groups by distance range, clusters at different spatial scales emerge. Since travel distance likely follows an exponential distribution, different intervals are assigned to get bins of trips as [0,1000m], [1000,3000], [3000,6000], [6000,10000], and [10000, -]. Map views of each bin of the trips are shown along with the trip counts. From these maps, local clusters can be easily spotted. For instance, on the map [0-1000m], the trips turn out to be clustered at many local centers.

However, this overview cannot tell us what kind of functions are provided by the centers or what kinds of trips are taken. Thus, a more detailed analysis will be given as follows. In particular, an analysis of travel behaviors associated with different activities is conducted.

(2) Share of transport mode

As mentioned in the review of earlier plans, the public transportation system was built to help shape the spatial structure of Singapore. Many policies, such as travel prices, have been carried out to promote the usage of public transportation. The figure below compares the share of transport modes in 2004 and 2008 (complete data set in 1997 is not available).

The share of public transportation including the MRT (Mass Rapid Transport), LRT (Light
Table 5.3: An overview of travel distance and activity locations.

<table>
<thead>
<tr>
<th></th>
<th>Year 1997</th>
<th>Year 2004</th>
<th>Year 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Distance(m)</td>
<td>6679.024795</td>
<td>6025.103026</td>
<td>7198.035154</td>
</tr>
<tr>
<td>Trip Counts</td>
<td>49026</td>
<td>50909</td>
<td>76923</td>
</tr>
<tr>
<td>(0, 1000m]</td>
<td>11520</td>
<td>8904</td>
<td>12184</td>
</tr>
<tr>
<td>Counts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1000, 3000]</td>
<td>8791</td>
<td>11884</td>
<td>13711</td>
</tr>
<tr>
<td>Counts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3000, 6000]</td>
<td>8378</td>
<td>9758</td>
<td>13771</td>
</tr>
<tr>
<td>Counts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6000, 10000]</td>
<td>7942</td>
<td>9039</td>
<td>14269</td>
</tr>
<tr>
<td>Counts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10000, −]</td>
<td>12395</td>
<td>11324</td>
<td>22988</td>
</tr>
</tbody>
</table>
Rail Transit), and public bus in total travel modes (vehicle only) is about 50%. From 2004 to 2008, the number of trips taken by MRT and LRT increased. After public transportation, travel by private car ranks second. Since the influence of mode share may vary for different urban journeys, Figure 5.9 looks at mode share from the angle of urban activities.

Figure 5.9: Share of transport modes in 2004 (top) and 2008 (bottom).

Five kinds of activities are selected because they occur regularly with different patterns. As shown in Figure 5.10, the surveys from 2004 and 2008 give slightly different options for transport mode. For a fair comparison, nine transport modes are selected. As indicated in the comparison, the MRT has an increasing share for all kinds of journeys. The public bus shows a significant increase for both journey to home and journey to work that replaces the share of
travel by private car. There are some other interesting trends. For instance, the taxi share is decreasing while the cycling share is increasing, which indicates a greener and healthier lifestyle and the effect of recent cycling routes.

(3) Trip starting time and ending time

Peak travel time is always of interest in transport planning. The temporal distribution in Figure 5.11 shows that the morning peak is shifting earlier and lasting for a longer time. Figure 5.12 shows the travel time for different kinds of journeys. Travel to study has a longer morning
Figure 5.11: Probability distribution of trip starting times in 2004 and 2008.

Figure 5.12: Probability distribution of trip starting times for different activities in 2004 and 2008.
peak as same as that of journey to work.

(4) Coverage of travel

According to the previous comparison, public transport has more than a 50% share of travel modes. This means that, in the case of Singapore, public transportation travel behaviors may well represent overall travel behaviors. Figure 5.13 shows the convex of activity locations. A convex measures the coverage of activity locations. It also shows that public transportation has almost the same coverage as that of all transport modes. Considering the demographics and geographic coverage of public transportation systems, smart card data is used as a proxy of overall urban movements. This statement is very important because it is a premise of later analysis conducted in this dissertation using smart card data. In the next section, a special focus will be placed on the way that people use public transportation systems.

![Figure 5.13: Spatial convex of urban activity locations in 2008.](image)

Note: Activity locations reached via all transport modes (red) and public transport modes only (green).

**Travel Behavior using Public Transportation**

Five patterns of travel behavior are analyzed from the surveyed data and used to build clustering prototypes for urban activities. These five features were chosen because they show the most remarkable differences between travels for different urban activities. Similar reasons apply to
the selection of activity types. HITS 2004 and HITS 2008 are analyzed for a comparison.

(1) Boarding and alighting time

Boarding time indicates when the peak hour is and alighting time indicates when people start their activities. The peak hour of all activities using public transportation is shown in Figure 5.14. The result is quite similar as that gained by using all transport modes. The morning peak is mainly caused by journey to school and journey to work, which start early and last long. The alighting time shows different travel times for different travel purposes. For instance, going to work and going to school mostly happen in the morning; going home normally happens in the evening; eating happens at lunch and dinner times; and social visiting and shopping are evenly distributed throughout the whole day. Identifying the temporal patterns of urban activities is of great importance for urban modeling and transport simulations.

Figure 5.14: Probability distribution of boarding and alighting time in 2004 and 2008.

(2) Age group

In this analysis, trips are divided into different groups according to the age of the travelers.
Figure 5.15 shows the major groups traveling for certain purposes. Different age groups have very distinct patterns. As shown, going to school occurs mainly among teenagers, while journey to work occurs in all age groups, but is concentrated in young people.

Comparing the results of the two surveys, major changes occur in the age group 25-29. As shown, young people generate more and more working trips, while the number of shopping trips is reduced. It might be caused by changes of age distribution in all occupations. The other activities have comparatively similar distributions.

(3) Travel frequency

Using one week as the temporal unit, the frequency of activities indicates how often people carry out the same activity. It is reported in the survey data as how many times people performed the same activity in the past seven days. This data is only available for the 2008 survey, so no comparisons by year can be made. As shown in Figure 5.16, going to work, going home, and going to school occur regularly, while the other activities occur more occasionally.
Figure 5.16: Probability distribution of activity frequency in 2008.

(4) Staying time

Staying time (shown in Figure 5.17) is estimated as the period between two trips used to perform the activities. There is no direct information in the survey, so it is estimated from the literature including statistical data of working hours obtained from the official Singapore Department of Statistics website. Other surveyed data about time use, such as from U.S. Statistics(2011), is taken into account as well.

Figure 5.17: Probability distribution of staying time.
Walking distance is how long it takes to travel from the bus stop to the destination by walking. It is reported in survey as the distance from the bus stop to the destination. In some aspects, it measures how convenient it is to use the public transportation system to reach the activity locations. A goal of public transportation planning in Singapore is to bring services closer so people can easily reach them by public transport. As shown in Figure 5.18, in most cases, the walking time is within five minutes.

Figure 5.18: Probability distribution of walking distance in 2008.

5.3.2 Mining User Travel Behavior from Smart Card Data

Analyzing travel survey data is a straightforward way to extract travel behavior. However, travel surveys are a costly exercise in terms of time and manpower, and conducted only every five years in the case of Singapore. This research suggests an alternative solution, which is to use cheaply and constantly collected smart card data. In fact, some research already demonstrates the possibility of making insights into urban problems by analyzing other sources of urban mobility data such as [109, 2]. In Singapore, payment for public transportation is mainly done using an automatic smart card fare collection system. These tap-in and tap-out records collected by smart card systems provide millions of observations on individual urban movements and have almost the same geographical coverage as that of all travel modes as shown in Section 5.3.1. This section presents some analysis and data mining work done using smart card data from 2011.
and 2012. Changes at the individual level are rare; therefore, the focus is on comparing the use of smart card data with surveyed data.

**Data Structure**

It is necessary to give more information about the data structure because smart card data in different countries records trips in different formats. In addition, the format is not as simple as that of travel survey data. In smart card data from Singapore, a journey is defined as a set of rides/trips on a bus and/or train from the origin to the destination. A journey may involve more than one ride if a transfer occurs (within 40 minutes). In the provided data set, one record is considered one ride with variables shown in Table 5.4.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journey ID</td>
<td>The unique number for a journey.</td>
</tr>
<tr>
<td>Card ID</td>
<td>The unique number of a stored value card.</td>
</tr>
<tr>
<td>Card Type</td>
<td>There are mainly three kinds of card divided by age group: adult, senior citizen and child (including student).</td>
</tr>
<tr>
<td>Travel Mode</td>
<td>Refer to transport mode of the ride  Bus or RTS.</td>
</tr>
<tr>
<td>Service number</td>
<td>Bus service number if it is a bus ride; NULL for RTS ride.</td>
</tr>
<tr>
<td>Bus number</td>
<td>Bus number if it is a bus service.</td>
</tr>
<tr>
<td>Bus Direction</td>
<td>Direction of bus route if it is a bus ride; NULL for RTS ride.</td>
</tr>
<tr>
<td>Boarding point</td>
<td>Id of Boarding bus stop / station.</td>
</tr>
<tr>
<td>Alighting point</td>
<td>Id of Boarding bus stop / station.</td>
</tr>
<tr>
<td>Ride start date</td>
<td>The date of a ride started. NULL if no tapping.</td>
</tr>
<tr>
<td>Ride start time</td>
<td>The time of a ride started. NULL if no tapping.</td>
</tr>
<tr>
<td>Ride distance</td>
<td>The ride distance in km. NULL if not tapping.</td>
</tr>
<tr>
<td>Ride time</td>
<td>The time interval (minutes) of a ride. NULL if not tapping.</td>
</tr>
<tr>
<td>Transfer number</td>
<td>The transfer sequence (ride) number of a journey.</td>
</tr>
</tbody>
</table>

This data enables us to carry out two kinds of analysis: one about general travel behavior using rides/trips and another about spatial interactions using journeys. Both analyses are presented below.
Spatiotemporal Patterns

(1) Temporal patterns

The set of plots in Figure 5.19 indicates the travel time for different age groups. It clearly shows that the morning peak starts at about 6:30 am and lasts for about three hours, which are the same insights as obtained from the statistical analysis of survey data. It can be easily concluded that the morning peak is mainly caused by adults rushing to work and students heading to school. The long-lasting morning peak is a consequence of the wider temporal choice of journey to work. As pointed out in [86, 132], there are indeed differences in travel patterns on different days of the week. These differences can also be spotted from the plots, e.g. the morning peak disappears on Saturday; adults and students have different travel schedules on weekdays.
(2) Travel distance

Since the distributions of trips from Monday to Thursday are almost the same, trips on Monday are picked as a representative of the other weekdays (except Friday). Figure 5.20 shows the travel distance on Monday, Friday, and Saturday for the different age groups. The distribution of travel distance shows the same patterns as the distribution of travel time, indicating that travel distance is closely correlated with urban activities. The travel distance for adults and students changes similarly between weekdays and weekends. On weekdays, they regularly travel to workplaces and schools. This reveals that their workplaces or schools are mostly far away from their living places.

Figure 5.20: Probability distribution of travel distance in 2011.

Note: Number of trips counted using five minutes as the time interval.

OD-Matrix of Journeys

An origin-destination (OD) matrix is a useful and powerful tool for transport planning, urban modeling, and simulation. OD-matrices are generated to represent the travel flows between different transportation zones at a specific time. OD-matrices can be generated/estimated from smart card data such as that in [99]. In the case of Singapore, an OD-matrix can be easily generated by linking all the rides/trips together by the transfer number given in the data or by the geographical locations of two trips. This dissertation looks at spatial interactions as a more
meaningful way to analyze journeys instead of trips. The OD-matrix below was generated from the three years of data. Not all of the available data sets cover the whole week, therefore, for a fair comparison, only weekday data is used.

Figure 5.21 shows the distributions of journey destinations in 2010, 2011 and 2012. For a clearer view, only MRT journeys are shown. Barely any change can be found from a visual comparison. Based on the number of trips at each individual bus stop and MRT station, some changes can be discovered. However, this does not mean there is no change in the intrinsic structure of flows because the raw number of journeys is not tied to the spatial structure of urban movement. This point will be addressed again in later analysis using spatial networks instead of direct mapping and statistics.

5.3.3 Inferring Activity Types from Travel Behaviors

This section explores the extensive use of the two types of data by fusing them to produce new information. It is an example of “re-combination of data”, which was mentioned as the first data innovation in previous discussion on big data in Chapter 2. By combining data, potential values may emerge. There are a few examples of fusing surveyed data and smart card data or another urban mobility data set. But most of them are used for enriching data sets instead of making extra value. As shown in Figure 5.22, the objective here is to infer people’s activity type/ travel purpose by synchronizing travel survey data and smart card data with an inferring method. Smart card data contain trips records with much higher spatiotemporal resolution than that in travel survey data. If travel purpose of trips can be retrieved, urban activities and dynamic
urban functions can be more precisely represented. Eventually, a better understanding of urban functionality can be gained.

Simply put, given a set of travel behaviors and their corresponding urban activities obtained from surveyed data as prior knowledge, the problem here is to deduce the most likely travel purposes of trips recorded in smart card. After investigating several possible methods, using prior knowledge to classify new data sets is identified as a very typical application of the Bayesian classifier. Moreover, Bayesian models are considered as the most fundamental and important method for data mining and information retrieval [98]. It is already a mature technique in data mining applications [71] and can process events with multiple variables and known prior probabilities. This characteristic makes it powerful for dealing with sequential events in cities or events with complex network relations [74, 91, 72, 77] In this specific case, a Bayesian model is used to retrieve travel purpose from travel behavior of daily activities as shown in Figure 5.23.

The related work has been conducted by the author and published in [164]. In which, a complete application has been done to infer activity type from travel behavior, moreover, to detect building functions from aggregated urban activities. This section re-organized the materials and use them to demonstrate the idea of data innovation - extracting information by fusing two data sets.

**Basic Concepts**

This section gives detailed information about this application starting with key concepts that are used to formally describe the research problem of this section. Urban function, daily activity, and travel behaviors are three basic concepts used throughout this application. Briefly stated,
in reality, functions of a building or an area is a compromised decision by both top-down land use planning and bottom up changes raised by individual’s actual needs. One way to find out the actual functions of a building is to observe what kind of urban activities are performed inside. Instead of costly fieldwork and survey, an alternative method is to infer the activity types (equal to travel purposes) from travel behaviors. These three concepts are generally used with ambiguous meaning, so it is necessary to redefine them in the context of this application.

*Urban function* refers to the actual use of a spatial unit. This application takes a building as a basic unit to describe the function. And the function is determined by what kind of daily activities happen inside the building in reality. In contrast to land use plans, it is how a building is used in reality. For instance, a residential area is planned to use as living places for people. However, sometimes a restaurant may locate on top of the buildings because of the actual demands.

*Urban activity* refers to the kind of daily activities like working, shopping, and eating which are all common social activities done by everyone. This kind of activity happens regularly, as been reported in introduced travel survey data and is able to be predicted [120]. To be noticed, travel purpose and daily activity are used interchangeable.

*Travel behavior* refers to the kind of travel behaviors analyzed in previous sections such as alighting time and activity frequency. The research shows that an individual has very stable mobility patterns that can be analyzed and used as travel behavior to make predictions [14, 2, 108, 86].
Figure 5.24: Work-flow for inferring travel purpose from travel behaviors.

Framework

Basing on these definitions, a framework of the proposed method is introduced here. A framework is proposed embedding a Bayesian model as shown in Figure 5.24. The first step is preliminary data processing, which contains two parts. One part is to extract travel behavior from travel survey data of typical travel purposes like travel time, activity time, and travel frequency, which has already been done in previous sections; and the other is to clean up and format smart card data. The second step is to deduce information about the daily activities that motivate the trips using statistical travel behaviors. This is done basing on a Bayesian classifier. And the result is probability distributions of daily activities.

A Bayesian Probability Model

As indicated before, a probabilistic model is the core of the framework. The Bayesian model is introduced in this section and is redefined in the context of the specific problem handled here.

A Naive Bayes classifier is a probabilistic classifier basing on Bayes’ theorem. Bayes’ theorem expresses the relationship between conditional probabilities when some events are contingent on others [30]. Given input sampled data, the Bayesian classifier assigns the most likely class label to a sample by evaluating its feature vector and prior probability. The Naive Bayes model has been shown to be effective in many practical applications [119].

Since the events of trips and their feature attributes satisfy conditioned independence, inferred information about daily activities can be formulated as an application of the Bayesian classifier. Simply put, given selected features (travel behaviors) of a trip, what is the probability of certain travel purpose of this trip? In the following part, parameters used in the Bayesian classifier will be defined formally. In this research, most distinguishable travel behaviors are selected as age groups, arrival time, duration and activity frequency, while activity types are
working, going home, shopping, studying, eating, and all other activities aggregated as social visiting.

**Definition 1** Trip T: A trip is a generated record. A record is generated by a set of time-ordered points recording how a passenger arrives and leaves one place to do a certain urban activity. Each trip reveals mobility patterns, which are expressed by multiple attributes. For instance, trip \( t = [a_a, a_t, a_d, a_f] \) where the attribute \( a_a \) stands for passenger age, \( a_t \) for arrival time, \( a_d \) for duration, and \( a_f \) for frequency. These attributes are mobility patterns that reveal people’s travel purposes, linking to a certain activity created by a passenger after traveling.

**Definition 2** Activity class \( C \): This is the set of possible urban activities that motivate a trip. It is also the information need to be deduced. In our case study, six activity classes are used, i.e. \( C = \{c_{home}, c_{working}, c_{studying}, c_{shopping}, c_{eating}, c_{social-related}\} \).

For each activity candidate \( c \), there is a prior probability \( P(c) \). For each attribution \( a_i(a_a, a_t, a_d, a_f) \) of a trip instance \( t = [a_a, a_t, a_d, a_f] \) belonging to activity class \( c \in C \), there is a prior probability \( P(a_i|c) \). This prior probability is our prior knowledge that was learned from statistical analysis of the surveyed data. As shown in Formula (1), given a new trip instance \( t = [a_a, a_t, a_d, a_f] \), the question can be formulated as: what is the most likely activity \( c \) that motivates the travel basing on the prior known probability? The answer is to calculate the maximum \( P((a_a, a_t, a_d, a_f)|c) \). Therefore, the likelihood of trip \( t = [a_a, a_t, a_d, a_f] \) belonging to \( c \in C \) is,

\[
p(c|(a_a, a_t, a_d, a_f)) = \frac{p(c)p((a_a, a_t, a_d, a_f)|c)}{p(a_a, a_t, a_d, a_f)} = \frac{p(c)p(a_a|c)p(a_t|c)p(a_d|c)p(a_f|c)}{p(a_a, a_t, a_d, a_f)} \tag{5.1}
\]

\( t = [a_a, a_t, a_d, a_f] \) belongs to activity class \( c_{map} \) which has a maximum likelihood as shown in (2):

\[
c_{map} = \max_{c_j \in C} P(c_j) \prod P(a_i|c_j) \tag{5.2}
\]

The result of this step is a probability distribution of travel purpose of each trip.
Experiment: A Case Study of Jurong East Area

(1) Case study area - Jurong East

As a tentative work, the proposed Bayesian model is applied to a case study area in Singapore. The case study is in Jurong East, Singapore (shown in Figure 5.25). Jurong East is part of the largest town in Singapore. Jurong has the second largest resident population and contains multiple land uses such as education, commercial, residential, and industrial. Its dimension has an area roughly 1500m*2000m, totally around 3,214,650.00 square meters. The statistical data cover trips in seven days from 136 bus stops located inside and on the border of the selected area. After preliminary data processing, there is an average of 128,000 valid trip records per day.

Figure 5.25: Case study area: Jurong East.

Note: Green dots denote bus stops.

(2) Preliminary data processing

Three types of input data are used: surveyed data which are used for statistical analysis of travel behavior as shown in Section 5.3.1; smart card data which contains only travel records. Travel purpose of these travel records will be inferred from a Bayesian model. Bus stops and
building footprints are stored in Shapefile format, which are imported into ArcGIS and manipulated by ArcGIS functions such as redefining projections and calculating distances.

Preliminary data processing of these data sets are conducted. First, statistical analysis is applied to the surveyed data to find out travel behaviors. The results of statistical analysis are used as prior knowledge of peoples travel behavior. Figure 5.26 table B (top right) is an example showing how one of the attributes frequency is used in the Bayesian model.

Smart card data is processed to extract the same attributes. The original smart card data provide information about trip ID, passenger ID, boarding bus stop ID, alighting bus stop ID, trip transfer time, starting time, traveling time, fare, and distance. A generated new record consists of six parameters, namely passenger id, passenger age, arrival time, staying time, frequency, and id of the arrival stop. In particular, passenger id, age, arrival time and stop can be read directly from the original data. Staying time/duration of activities is estimated by calculating the interval time between two trips, starting from tap in, ending with tap out, from a select area with the same passenger ID. Frequency is a statistic of how many times a passenger ID appears on different dates. Statistical results as well as the processed data structures are demonstrated with real sampled data. A sample of generated records is shown in Figure 5.26 Table A (top left).

(3) Results

As shown in the framework, after the preliminary data processing, a trip classification is performed using the Bayesian classifier with input from the analyzed results. Figure 5.26 shows example tables including the generated trip records shown in Table A (top left), and table B (top right) in an example of prior probability and table C (bottom left), which are the results of the classification showing the inferred probability distributions of daily activities linking to each bus stop.

In the first step, the value of prior probability $P(a_i|c)$ is read from the prior probability table. Different frequency refers to a different value of prior probability. As such, there are tables of prior probability distributions for the other attributes. In the second step, after checking all the individual attributes’ prior probability, Formula (1) is applied to calculate the probability of activities, thus finding the most likely activity that motivates a trip. Table C is the posterior probability distribution of the six daily activities linking to trips arriving at one stop, e.g. bus stop “284**” has the highest probability of education, abbreviated as “e” in the table. It means
that the majority of people alighting at this bus stop are traveling for studying, which implies that there might be an educational institute nearby. The chart figure (bottom right) in Table C shows the probability distributions of the six activities at 136 bus stops in Jurong East.

The probability distributions of the six daily activities are labeled in six different colors. The x-axis shows the bus stop id, while the y-axis shows the proportion of activities at each stop. An intermediate evaluation of the results is done to check the general effectiveness of estimated results. Buildings surrounding bus stop “284**” are checked on Google Maps. The closest building is a school, which explains why the main activity of going to bus stop “284**” is studying. It is also a rough validation of inferred results.
5.3.4 Discussion

As a first step of data analysis, this section conducted statistical analysis to travel survey data and smart card data to detect the changes of travel behaviors for different urban activities over years. Besides that, it is a comparison of usage of travel survey data and smart card data. Actually, several advantages of using smart card data have already been identified in related works [13], such as:

- Access to larger sets of individual data.
- Possibilities of links between users and card information.
- Continuous data available for long periods of time.
- Better knowledge of a large part of transit users

The work in this section gives additional evidence of such advantages by analyzing travel behavior using two types of data sets.

Some trends in urban transportation can be drawn from the analysis of both data sets. However, the surveys are conducted every four to five years and only cover 1% of households in Singapore, providing about 100,000 records. These household surveys are also a costly process in terms of time, money, and manpower. In comparison, smart card data is much cheaper and, according to the statistics, more than 2 million people use the two transit systems and generate about 5 million records each day. This means that smart card data can easily provide a large quantity of information with respect to extracting travel behavior more efficiently. It is undeniable that surveyed data contains richer information than smart card data. However, our extensive study using a Bayesian model shows an example of inferring extra information by combining two data sets. It is a typical example of data innovation and points out a potential way to support urban planning processes by providing advanced data services. These inferring techniques may radically change the conventional method of data acquisition in urban analysis. The inferred data may be of higher quality and better able to represent urban dynamics. For instance, the presented inferred application achieves information about urban activities that reflects how people use urban space in reality. These urban functions were originally defined by urban planning and then redefined by individuals’ actual needs through bottom-up changes. As defined in [120], land use has two aspects: formal land use refers to its form, pattern, and aspect; while
functional land use refers to a socioeconomic description of space. The latter aspect may have a higher dynamic level than the former. As discussed in work by [60], functional changes in cities are not tied to morphological changes. It is crucial to understand urban functions and their compatibility with the original plans. This leads to the work in the next sections, which uses information about urban activity instead of urban infrastructure to measure polycentric spatial structure.

5.4 Detecting Changing Spatial Structure from Urban Activity Patterns

Travel behaviors at individual level are easy to extract like what has been shown in previous sections. These individual changes can even be observed directly from people’s daily life. Comparatively, it is a more difficult task to identify spatial structure, because spatial structure requires an overview of the global spatial organization. It is a result of collective effects in an aggregated level at larger spatial scale. Therefore, a more advanced spatial analysis is needed firstly, to identify the activity center, secondly, to measure how central a center is comparing to the other centers, and finally to detect how much the overall spatial structure is changing over years. This section analyses aggregated activity patterns using travel survey data in different years and detect emerging spatial structure.

To do it, a new measure of urban centrality is introduced to identify activity centers and the degree of polycentric distribution in the urban process of decentralization. A Centrality index is defined based on a combination of density and entropy of urban activities with a spatial convolution. With this centrality index, we are able to build a relationship between the activity patterns and urban form. Moreover, changing distributions of activity centers can be detected and compared quantitatively using centrality values. Consequently, the urban process can be detected and expressed explicitly.

A detailed literature review regarding measuring Polycentricity has already been given in Chapter 2. Here, only highlights of the proposed measure are emphasized:
1. The proposed centrality index takes various urban activities into account and differentiates mono-functional centers with multi-functional centers by types of activities performed in the centers. Previous related work measuring spatial structure and spatial interactions are mainly based on commuting patterns of “journey to work” [66, 143] which, however, is no longer the only dominant travel purpose as that observed in even earlier studies [57, 55]. Evidence could also be found from the statistics in this dissertation, non-work trips such as to school and to go shopping also plays important roles in today’s city life.

2. The proposed method measures functional centers. This dissertation studies Polycentricity as kind of spatial distributions of clusters. The clusters measured are human activity gathering areas, which are called functional centers reflecting the function of a place in reality.

3. The proposed method measures the degree of Polycentricity and reconstructs the process of decentralization through years of development. Since Polycentricity is highly context and scale dependent that cannot be associated with an absolute value of urban elements, it is more reasonable to consider it as a relative value about spatial distributions of centers and sub-centers.

4. Beyond the specific urban phenomenon - polycentricity and in a broader sense, this method is also an example of data innovation introduced previously in section 2.4 - “extensive data”. Travel survey data is used for extracting spatial structure instead of its original usage for estimating travel demand.

Note that both Section 5.4 and Section 5.5 detects emerging spatial structure from urban mobility, but from different perspectives of view. In both analyses, the centrality index as well as other related indices are defined and measured with different methods. For a better understanding, the two sections are structured in a unified structure:

- **Definition of indices** used to quantitatively describe the urban transformation, and the origination of the derived indices.

- **Measures** used to compute indices using given data set.

- **Experiments** demonstrate the implementation of the measure with real data set.
Insights of the decentralization urban process are gained from interpretation of the calculated indices. The computed value of indices are analyzed and linked it to morphological changes to find out the driven force and impacts of urban changes.

Discussion is given mainly regarding the feasibility of applicability of the method.

5.4.1 Definition of Indices

In this scenario, there are three key concepts defined as follows:

- **Functional centers** are places where people are accumulated to perform certain activities.

- **Centrality** is an index that measures the degree of clustering of activities in a same places (functional centers). The two key characters - density and diversity are used to quantitatively measure the aggregated patterns of urban activities. In particular, **Diversity** index measures how mixed the distribution the activities is, and **Density** index measures how dense the distribution of activities is.

- **Polycentricity** is a set of indices computed based on the centrality index, mainly including (1) statistical distribution of centrality values, (2) geographical distributions of centers and (3) spatial influence areas of centers defined by a relative centrality level.

This definition of Centrality is based on the central place theory (CPT) which has been claimed as the original foundation theory about the organization of an urban system, and extensively used in many disciplines like urban geography, spatial planning and urban economics [29].

As previously reviewed, CPT is first introduced by W.Christaller [37] and A.Lösch [87]. It tells the number, size, and location of human settlements in an urban system. It has been later developed for more general and realistic models by [24]. There, a scenario is constructed as a distribution center of goods and services to a scattered population, which was simply formatted as N types of central goods selling at centers to reveal a hierarchical spatial structure. A center is the place of a supply of goods and services, and a periphery (regions complementing the center) where demand, i.e. the population using them, resides. Centrality then measures clustering in a
Applying these basic concepts into the context of urban activity, the two fundamental attributes - size and order that determine the importance, or centrality, of an area within a given city are replaced by (1) density of the visits which tell the number of people attracted to one area, and (2) diversity of their activities, which tells how many different functions an area provides. Intra-urban centres can then be identified as spatial clusters of activity locations by their centrality value.

Figure 5.27: An outline of proposed approach for measuring polycentric urban process.

5.4.2 Measure: A Spatial Convolution Method

The proposed calculation combines two functions into one. First, it reduces two-dimensional information - diversity and density to one-dimension - centrality. Second, as an essential function of all local spatial analysis, it is a smoothed density function that detects clusters and outliers. These clusters form the defined functional centers. Consequently, the outline of the presented approach can be sorted as shown in Figure 5.27.

In this measure, urban space is partitioned by grids in unified size as shown in Figure 5.28. Urban activities which are represented as points are aggregated into a grid that they fall into. Each grid cell is considered as one smallest spatial unit. Spatial structure is identified in three steps: (1) Calculating basic indices, namely density, diversity/entropy of each grid cell. (2)
Calculating centrality index, which is a combination of density and diversity value of each grid cell. Functional centers are identified as clustered contiguity grids that have comparatively high centrality values. (3) Obtain the spatial structure from a global view. There, a set of indices that are frequently used in convention spatial analysis is introduced to assess the spatial distribution of centrality values as well as identified functional centers. Changes of spatial structure is then analyzed visually and described quantitatively by classical indicators such as Moran’s I index. The detailed measurement is explained as follows.

**Step 1:** Calculating density and diversity index

In fact, density and diversity index have long been used in land use and transportation planning [32, 96]. However, they were always used for land use data, not for mobility data. Here, they are modified to measure the pattern of urban activities using travel survey data.

The *Density* index is measured as the proportion of people accumulated in one unit area \((x, y)\) in \((m \times n)\) units space \(S\) in a given period of time, defined as

\[
D(x, y) = \frac{N(x, y)}{\sum_{i=1,j=1}^{i=m,j=n} N(x, y)} \quad (5.3)
\]

where \(N(x, y)\) is the proportion of accumulated number of people arriving at a unit area \((x, y)\).
The diversity index is replaced by entropy, as entropy is a more quantitative index that describes not only number of types of activities, but also the disorder of activity types. The concept of entropy index was originally proposed in information theory by C.E. Shannon [127]. In general, the smaller the entropy, the lower the disorders of the land use. Derived formula are used for measuring the disorder and/or evenness of land use arrangement.

The calculation in this section is developed based on the more generic definition of land use entropy. A regular grid is employed to split the whole data set into cells, according to their geographical coordinates of X and Y directions. Given a geographic space \( S \) split into \( m \times n \) cells. For a cell \((x, y)\) with \( J \) types of land use, its land use Entropy index is defined as

\[
E(x, y) = -K \sum_{j=1}^{J} P_j(x, y) \ln(P_j(x, y))
\]  

(5.4)

where \( P_j \) is the proportion of land in the use type \( J \) within a cell, \( K \) is the number of neighborhood cells, which is used to smooth the entropy value [32]. A single land use in a cell results in an entropy value of 0. An example demonstrating the calculation of land use entropy is shown in Figure 5.29.

![Figure 5.29: A demonstration of mean entropy calculation.](image)

Note: (a) is the original land use, different land uses such as road, park, business are marked with different colors. (b) calculated the local entropy value of each grid cell, (c) calculated the mean entropy.

The measure is reformulated to mix of activity types instead of land uses. In such context, diversity index of urban activity measures how mixed the activity types in one unit area, where
$P_j$ is the proportion of travels to cell $(x, y)$ for the activity type $j$ during a period of time. $J$ is the number of different activity types considered.

However, density and diversity are two quantitative values of different dimensions and physical meanings. They can be directly manipulated together. Normalizing the two parameters into the same range could be one option which is normally used in many high dimensional clustering applications. Considering the spatial convolution functions in next step, a simple ranking mechanism is used for normalizing.

Given the rank of the largest density and/or entropy value as 1, and then the rank of the other areas depends on the comparative scale between them and the largest ranking area. Another explanation could be given from the perspective of probability theory. Assuming that the area with higher density and/or entropy value implies a higher probability to be a multi-functional center, which is in line with our intuition. Given the highest probability as 1, the probabilities of the other cell are calculated by related scales. A formal definition is given as follows.

In a two dimensional $m \times n$ space $S$, denote the density function $D = D(x, y)$, with $x = 1, \ldots, m$ and $y = 1, \ldots, n$. $d_{x,y}$ is the density of cell $(x, y)$ in $S$. For each cell, there will be a function $P_d = f(x, y, d_{x,y})$ to denote the probability of a cell to be a city center based on its density only. Thus, the density ranking function is

$$R_d(x, y) = f_d(x, y, d_{x,y}) = d_{xy}/\text{Max}(D)$$ (5.5)

Similarly, $P_e = g(x,y,e_{x,y})$ is a probability density function related to diversity $E_{x,y}$ at cell $(x,y)$. The probability density function of diversity is

$$R_e(x, y) = f_e(x, y, e_{x,y}) = e_{xy}/\text{Max}(E)$$ (5.6)

Diversity and entropy are indispensable attributes to identify a center, however, none of them can represent central areas individually, especially in the context of modern cities, where monofunctional areas exist. For instance, a residential town might have very high density but limited type of activities there, which should be differentiated with multi-functional centers. More complicated situations should be considered and typical examples are given to demonstrate the possible misinterpretation by a single parameter in Figure 5.30 (top), there could be two areas
Figure 5.30: A demonstration of the misinterpretation of diversity index.

Note: The example shows that density and diversity are two independent indices. Circles and triangles represent different type of activities.

having same level of diversity but very different densities. Figure 5.30 (bottom) shows that there could be some non-central areas with high diversity of activity types and less visiting people. A centrality index is, thus, developed to integrate these two indices into one.

Step 2: Centrality index: a convolution-based smooth function

The centrality index, $C_{x,y}$, measures the centrality of an area $(x, y)$ in a city. It is the possibility of one area to be a center, being derived from both the density of people and the diversity of their activities by a spatial convolution operation.

\[
C(x, y) = R_D(x, y) \otimes R_E(x, y) \tag{5.7}
\]

Convolution is a fundamental concept in signal processing and analysis. It is a combination of two functions $f$ and $g$, which produces a third function that can be interpreted as a modified version of $c$.

Given two time sequential functions $f(t)$ and $g(t)$, as the signal energy at time sequence $t$. 

A new time-energy function $c$ will be the convolution of $f(t)$ and $g(t)$, as shown here:

\[
c(t) = f(t) * g(t) = \int_{-\infty}^{+\infty} f(x)g(t - x) \, dx.
\]  

(5.8)

Figure 5.31: Spatial convolution with contiguity edges and corners.

If $f$ and $g$ is defined on a spatial variable like $x,y$ rather than a time variable like $t$, the it is called spatial convolution. In this paper, a discrete 2D spatial convolution is applied to “add” the $R_D$ and $R_E$. At each cell $(x,y)$ in the output function, place a window centered at $R_E$, with continues cells as shown in Figure 5.31, and scaled up or down according to the value of window centered at $R_D$. After adding the nine values (center and surrounding cells) all together as $C(x,y)$.

**Step 3:** Measuring Polycentricity by quantifying spatial distribution of functional centers.

Polycentricity indices are a set of indicators that give more details comparisons of spatial distributions of functional centers.

- **Number of centers**: is a simple indicator. Decreasing number of centers indicates a monocentric urban process while increasing number of centers indicates a more polycentric city.

- **Size of a center**: measures the area of a center which is formed by contiguity grids with centrality value higher than certain standard.
• **Variance of Centrality value:** in probability theory and statistics, variance measures how far a set of numbers is spread out. Here, it indicates the evenness of centrality value among all areas in a city. Lower variance indicate higher degree polycentric since Polycentricity tends to be more closely associated with a balanced distribution with respect to the importance of these urban centers as indicated in [76, 95, 29].

• **level of clustering:** is a measure of spatial auto-correlation developed by Patrick Alfred Pierce Moran [97]. It is used here to quantify the spatial distribution of centers. A more clustered spatial distribution of centers indicates a mono-centric urban process, and vice verse. Similar to Variance, lower variance indicate higher degree polycentric. The Moran’s I statistic can be easily computed using ArcGIS.

• **Global mean center:** is measured using centrality as weight. The moving of global mean center indicates a fast local development.

### 5.4.3 Experiment: Analysis of Travel Survey Data in 1997, 2004 and 2008

(1) Preliminary data processing

In this experiment, travel survey data - the so called Household Interview Travel Survey (HITS), is used as input. In order to track the changes, three years’ HITS data are used, including HITS 1997, which contain 48,881 validated records after data processing and HITS 2004, which contain 51,000 validated records and HITS 2008, which contain 76,923 validate records. As shown in Figure 5.32, the activities locations cover almost all the areas.

These surveyed data of three years originally have different classifications of activity types. For a fair comparison, certain data aggregation is conducted to get a unified base of classification and the number of trips for each aggregated activities are given in Table 5.5.

(2) Density, diversity and centrality

This experiment set 24 hour as a temporal unit since the survey is a report of people’s activity in one day. 500m × 500m is the size of the grid which is used to partition the whole city space. 500 meters is an approximate average walking distance to transportation infrastructure according to statistical results of travel survey data. In the end, the whole area is partitioned
Table 5.5: Original activity types, aggregated activity types and trip numbers.

<table>
<thead>
<tr>
<th>Aggregated Categories</th>
<th>Year 1997 trip number</th>
<th>Year 2004 trip number</th>
<th>Year 2008 trip number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Go home</td>
<td>21100</td>
<td>23543</td>
<td>34314</td>
</tr>
<tr>
<td>2 Go to school</td>
<td>3177</td>
<td>7498</td>
<td>9757</td>
</tr>
<tr>
<td>3 Go to workplace</td>
<td>8407</td>
<td>10425</td>
<td>18310</td>
</tr>
<tr>
<td>4 Part of work</td>
<td>1166</td>
<td>736</td>
<td>1453</td>
</tr>
<tr>
<td>5 Shopping</td>
<td>3511</td>
<td>2372</td>
<td>2239</td>
</tr>
<tr>
<td>6 Eating</td>
<td>2966</td>
<td>830</td>
<td>1634</td>
</tr>
<tr>
<td>7 Social</td>
<td>1768</td>
<td>1115</td>
<td>2108</td>
</tr>
<tr>
<td>8 Recreation</td>
<td>1212</td>
<td>267</td>
<td>767</td>
</tr>
<tr>
<td>9 Others</td>
<td>5574</td>
<td>4123</td>
<td>6341</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serve Passenger</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal business</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>valid record total</td>
<td>48881</td>
<td>50909</td>
<td>76923</td>
</tr>
<tr>
<td>record</td>
<td>52801</td>
<td>60917</td>
<td>88601</td>
</tr>
</tbody>
</table>
Figure 5.32: Mapping activity locations in Singapore.

Note: Activity locations are arrival locations of trips in HITS 2008. The areas which barely have any activity points are mainly open space, port, reserve site, special used areas, and water body according to the master plan 2008.

into 3578 grids. Activity points are aggregated to grids by joined spatial locations. As indicated in previous discussion about big data, aggregation is also a way to safeguard individual privacy. To avoid small errors of geo-coding, a mean entropy and density is used to smooth the value of one grid with its eight neighborhood grids defined by contiguity edges and corners. To evaluate the influence of smooth function, experiments have been conduct. The results generated with and without smooth function show qualitatively similar patterns.

The results of diversity, density, and centrality maps are shown in Figure 5.33. There are incompatible diversity and density patterns of activities clearly shown in some of the areas, like the one marked by rectangles is Jurong West area, which is most occupied by residential blocks with some schools. In (a) has a peak point, while in (b) contains comparatively low values, thus in (c) centrality value of that areas has been scaled down after a spatial convolution.
Figure 5.33: Density, diversity, centrality and difference between centrality and density.

Note: X,Y axis represent the index of geographical coordinates system of Singapore. The four maps are density of urban activities (a) entropy of urban activities (b) result of convolutions the centrality map (c) a difference map of centrality and density (d) to assess the functionality of convolution. From (d), you can see that central areas are enhanced while other areas filtered out.

Another example is given in Figure 5.34 with more details about the distribution of density, entropy, and centrality. The density and diversity value of each cell are plotted out, X-axis is the density value, Y-axis is the entropy value, and each dot represents a cell. Though the correlation between the two dimensions is very high as shown in there are some clear exceptions. As demonstrated, the selected dots are corresponding to areas in north-east of Singapore (Hougang area) with comparatively higher density but lower entropy. Because residential building has a dominate number in that area. After a spatial convolution, the centrality values dropped into lower level bins as shown in the histogram view.

Moreover, process of urban development in Hougang area can already be spotted from the changing values from 1997, 2004 to 2008. The rise and down of centrality value in that area before and after 2004 might cause by the continuous development of new neighborhood in that area in 1990s, but the opening of a rapid train line in 2000s led the flow of people to go outside.
Figure 5.34: Incompatible density and entropy patterns.

Note: Density and entropy value of 2004 are plotted (left), x-axis denotes density, y-axis denotes entropy. A selection is made to get the dots with comparatively high density but low entropy. Since each dot denotes a cell in geographical space (right). As demonstrated, the selected dots are corresponding to areas in north-east of Singapore. Number of this kind dot decreased in the result of 2008.

To emphasize, the centers are the areas with high density and high diversity, while filtering out the others. A question is that there is no standard level to classify and divide the areas into different groups. This is also part of the reason for using convolution to combine the two indices into a simple centrality index.

Besides comparing the density, entropy and centrality value from map views, their statistical distributions are also plot out. The centrality values which are achieved by a spatial convolution show a very typical cut-off power-low distribution. It can be taken as another evidence of universal scaling low in urban system. And on the other sides, proof the meaning of the calculated centrality value. More discussion about the meaning of this distribution in the context of
understanding urban process is given in next sections.

5.4.4 Insights of Polycentric Urban Transformation

Last section is a demonstration of the presented measure. This section interprets the results in the context of urban process. From reading the changing values over years, a dynamic urban process can be reconstructed. Linking to the physical urban changes reviewed in previous sections, the cause and sequence of observed phenomena can be explained. When comparing the results with original urban plans that have been introduced before, the actual effects of urban plans can be evaluated. In particular, three aspects are address in this section: (1) overall value of centrality - how is the general development of Singapore; (2) balance of the distribution - where are the centers (3) anomalous - any incompatible that against original motivations of Polycentricity.

The insights are made based on the following results: a statistical mapping of accumulative probability distributions of centrality given in Figure 5.35; and a geographical mapping of the centrality distribution in 1997, 2004 and 2008 shown in Figure 5.36. As indicated before, big centers and small centers are relative concept. Therefore, centers are identified by a ranking mechanism according to their centrality value. Different intervals can be customized for ranking. As an example, nine levels are given here using 0.1 as the interval value. A color scale is for graphic mapping. A more detailed comparison is shown in Table 5.6 giving more statistics about the distribution of values.

(1) Overall increasing of centrality

As previously mentioned, in the short about five decades as an independent city-state-nation, Singapore have gone through fast urban development and transformed itself from a declining trading post to a First World economy [69]. It cannot be a surprising result that the average centrality value increased continuously. The increasing centrality value means that the whole city became more ‘active’ in general. This change can be clearly captured from the geographical map that areas with medium centrality value dispersed (centrality < 0.3 and centrality < 0.1). The number of the cell with comparatively higher centrality value (centrality > 0.3) are increasing significantly. An agglomerated central area is defined as a group of adjacent cells that have centrality values higher than certain standard. The geographical distribution then can be told by

<table>
<thead>
<tr>
<th>Indices</th>
<th>Year 1997</th>
<th>Year 2004</th>
<th>Year 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Centrality</td>
<td>0.024611</td>
<td>0.039981</td>
<td>0.04654</td>
</tr>
<tr>
<td>Max. Centrality</td>
<td>0.54349</td>
<td>0.7083</td>
<td>0.83775</td>
</tr>
<tr>
<td>Standard deviation centrality</td>
<td>0.056668</td>
<td>0.090856</td>
<td>0.095621</td>
</tr>
<tr>
<td>Moran’s I index</td>
<td>0.739429</td>
<td>0.744428</td>
<td>0.776470</td>
</tr>
<tr>
<td>Max. density</td>
<td>0.0185</td>
<td>0.0085</td>
<td>0.0091</td>
</tr>
<tr>
<td>Standard deviation density</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0006</td>
</tr>
<tr>
<td>Moran’s I index</td>
<td>0.725631</td>
<td>0.76267</td>
<td>0.759729</td>
</tr>
<tr>
<td>Avg. entropy</td>
<td>0.2912</td>
<td>0.2133</td>
<td>0.2763</td>
</tr>
<tr>
<td>Max. entropy</td>
<td>2.0347</td>
<td>2.1274</td>
<td>1.9653</td>
</tr>
<tr>
<td>Standard deviation entropy</td>
<td>0.5160</td>
<td>0.4243</td>
<td>0.4798</td>
</tr>
<tr>
<td>Moran’s I index</td>
<td>0.856524</td>
<td>0.838281</td>
<td>0.859153</td>
</tr>
<tr>
<td>Density &amp; entropy</td>
<td>0.5668</td>
<td>0.6925</td>
<td>0.6280</td>
</tr>
<tr>
<td>correlation coefficients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of grids with centrality &gt; 0.3</td>
<td>23</td>
<td>94</td>
<td>104</td>
</tr>
<tr>
<td>Number of centres &gt;0.3</td>
<td>5</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Number of grids with centrality &gt; 0.7</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Number of centres &gt;0.7</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Avg.travel distance (meters)</td>
<td>6679.024795</td>
<td>6025.103026</td>
<td>7198.035154</td>
</tr>
<tr>
<td>(point to point distance)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg.in vehicle Time (walking excluded)</td>
<td>?</td>
<td>20.5173</td>
<td>21.2826</td>
</tr>
</tbody>
</table>
the number of centers tells which is increasing.

Figure 5.35 gives another perspective from the empirical probability distributions of the locational centrality, P(CI), for the studied periods. Note that the power laws are marked by a sharp exponential cutoff that appears at a lower value for the year 1997 (CI ≈ 0.1) than for the other two years (CI ≈ 0.2). This indicates a significant increase in the number of central hubs between 1997 and 2004. The distributions are remarkably stable over the different years and follow a truncated power law with $P(CI) \propto C^{-\alpha}$, being valid over several orders of magnitude. This heavy-tailed distribution shows evidence for a high heterogeneity of locations with respect to their centrality. Simply put, most locations are visited by just a few people and for similar reasons, while a few central ‘hubs’ attract a huge part of Singapore’s population for many different reasons. Yet all intermediate centrality values are present. Hence, the average centrality does not represent any typical value of the distribution such as, for instance, the most probable value for a Gaussian distribution. Also notice that though the centrality value of three years follows the same overall distribution, the geographic locations of the ‘hubs’ are changing and are discussed in next section.
Figure 5.36: Centrality map generated from travel survey data in 1997, 2004 and 2008.

(2) More evenly distributed of centers

The geographic mapping in Figure 5.36 shows that in 2008, the three significant sub centers: Jurong area in the east region, Tampines in the west region, Woodlands in the north region, were emerging and gradually growing to be regional centers with similar centrality value, except Seletar in the north-east region having comparatively lower centrality. If you compare the centrality map in 2004 and 2008, it is obvious that the centrality value in Hougang areas decreased, while the centers in western part of Singapore are having increasing centrality values. It means the urban development tends to be more even distributed. To prove this intuitive observation, the global mean center of Singapore using centrality value as weight is calculated. The
center point is actually gradually moving towards the western part of Singapore.

The result is quite in line with Singapore’s essential planning concept in general. However, what has also been found are other emerging sub centers like that in Yishun and Bedok having higher centrality values than the planned sub centers in some years. To some aspect, this abnormal phenomenon is an evidence of the unpredictable bottom-up changes which reshaped the urban structure in reality beyond that in our plans. Besides detecting the spatial structure of today and using it to evaluate urban plans, the changing path can also be read from analyzed results. Both standard deviation of density and entropy increased in 2004 and decreased in 2008, indicates that distribution of activity becomes unevenness in 2004 and back to evenness in 2008. The western region of Singapore - Jrong East area was mainly occupied by industrial, and the blueprint to transform Jurong Lake district into unique lakeside destinations for business and leisure was unveiled in recent years. It is promising to see even higher centrality value using upcoming new surveyed data 2013 in future analysis. In sum, our finding proved the kind of urban process not from the aspect of physical changes that can be easily gained from land use data, but from the aspect of urban activity and movements.

(3) Anomalous increasing high centrality in central area

As indicated previous, global autocorrelation - Moran’s I index of centrality value is calculated to evaluate the spatial distribution. The value is increasing throughout three years. It indicates (1) a very significant spatial auto-correlation that high centrality areas are well clustered. (2) The difference of centrality between areas is increasing.

The second point is in line with the clear evidence from the statistical result in Table 5.6 that standard deviation of overall centrality values are increasing. The numbers of cells that form the biggest agglomeration areas are increasing. From geographical mapping, as shown in Figure 5.36, the biggest center in the southern part of Singapore has increasing high value of centrality. It was developed as a CBD even in the earliest urban plans. The impact of that plan is still obvious today. The big center keeps on growing with a reason. Since the development of this area and the neighborhood area are always high priorities in urban plans, with heritage protect attracting more tourists, trading markets are building to promote economy. Another reason could be the development of transit system. Rapid transit system is built to shorten the travel time from everywhere to the big CBD. It functions against the idea of decentralization
that urban stocks are flowing into one center instead of being distracted to the other centers. The increasing travel distance but slighting changed travel time of all kind activities is a reasonable result.

5.4.5 Discussion

This section proposes a centrality index for detecting functional urban centers from urban activity patterns using travel survey data of different years. With a simple density and entropy index, multiple types of urban activities are integrated. A spatial convolution is used as a smooth function and a function that combines two indices into one. With the centrality index, functional centers are identified and spatial distributions of these functional centers are compared through spatial analysis. Taking Singapore as an example, surveyed data of different years are used to reconstruct the urban process over one decade. The quantitative approach and the results can be used as references for explicitly interpreting and representing urban changes to support urban plan applications. On one hand, it is a way to measure spatial structures that are shaped by the way that people are effectively using urban space emerging from people’s daily activities. On the other hand, it is an example of the presented data innovation that travel survey data which are original used for estimating travel demands, are used for detecting spatial structure.

This presented method can be easily adapted to other case study areas which have available travel surveyed data since the inputs of the method are rather simple that without any specific requirements. The presented method should be considered as a basic framework that still remains potentials to be further extended. Firstly, the usage of these indices are not limited to surveyed data, they can be expanded to apply to the other mobility data set like smart card data which have higher spatiotemporal resolution. A way to adding extracted activity information into smart card data was presented in last section. Secondly, the indices can be used not only for detecting urban activity centers but can also be further derived for detecting other functional centers, like education centers, shopping centers that a more detailed market area analysis can be made. Finally, it should be noted that the ranking/probability functions defined here are in simple forms. Those functions are based on the hypothesis that functional activity centers have high density and high entropy. These functions could be changed according to a further refined hypothesis within the proposed framework.
5.5 Detecting Changing Spatial Structure from Urban Movement Patterns

This section measures Polycentricity from urban flows. The spatial structure revealed in distribution of urban flows tells not only the distribution of stocks in centers, but also the connections between centers. Materials in [163] which is published by the author are organized and used to demonstrate the proposed measurement of functional Polycentricity.

Unlike urban stocks that can be represented by limited number of samples, to represent all kind of possible links between all spatial units, the number of required samples are increasing exponentially. Smart card data is therefore a better choice. Besides the advantage of data volume, smart card data also contains rich information about urban mobility. As that proved in previous analysis in Section 5.3, in the case of Singapore, public transportation data gives almost equal representation of urban mobility as that given by all travel modes. However, as indicated, smart card data have less demographic information than travel survey data. Considering the advantages and disadvantages of such data sets, a spatial network model is proposed and further questions about urban changes are to be answered. First of all, is that a polycentric urban transformation of human movement in Singapore? Secondly, are the functional centers formed by people in surrounding areas or people who live far away but are used to travel long distance? Third, at which spatial level, people’s movement follows the polycentric structure? Similar as that in last section, these questions will be answered by tracing changes over years. A changing path which reflects how people adapted to and reshape the use of urban space can be identified, in particular, by identifying the hubs, centers and borders of urban movement landscape. Innovations of the proposed analysis are emphasized as follows:

1. This method measures polycentricity from urban flows, which follows the argument in [29] that: “Morphological changes addresses changing size and geographical distributions urban infrastructures, and functional changes take connections between settlements into accounts, which are two kinds of analytical concepts both of polycentricity”.

2. The proposed spatial network analysis is new. Actually, research using network and flow theory with smart card data analysis does not have a very long history, largely because network science has only very recently been extended to deal with spatial networks [16] and smart card data pertaining to travel on such networks has only just become available.
Besides, it is also new from the perspective of spatial analysis and modeling approach, since data are analyzed with an analogy model that takes a representational or functional form of network and applies it to urban stocks and flows.

3. Similar as that in last sections. Polycentricity is measures the degree of Polycentricity through years of development. Quantitative indices are proposed.

4. This method is another example of data innovation - “Extensive data” and “open data” - that data can be used for untapped purpose. Smart card data are not intentionally collected for urban planning, but now it is used in this study for extracting spatial structure. Besides smart card data, there are various new available data sets in high spatiotemporal resolution such as mobile phone data, taxi data, which undoubtedly provide unprecedented possibilities to develop this type of data innovation.

5.5.1 Definition of Indices

As mentioned before, the spatial structure of modern cities was shaped, in large measure, by advances in transport and communications [6]. The complexity of human movements has redefined the usage of urban space and the arrangement of resources. People, as physical carriers, motivate the transfer of materials, money, and information and so on between areas in urban space. Therefore, taking travel as a proxy for spatial interaction, an illustration of the basic idea behind the analysis in this section is shown in Figure 5.37.

Stops are representatives of surrounding areas. Trips between stops are aggregated to represent flows between areas. By measure the structure of flows, different characteristics of areas reveal. Moreover, areas with similar features are grouped and forms neighborhoods. With the partition of neighborhood, new borders are emerging, which represent how the urban space has been re-partition by social-economic features in reality. In total, this provides us with proxies for the physical urban flows between places and although these are a crude simplification of the homogeneity and heterogeneity of well-defined urban spaces, this model represents a first shot at defining such places with respect to flow networks, linking ideas about regionalization from the 1960s to contemporary network approaches.

It is necessary to describe the scenario of the construction of urban spatial network before moving on to a formal definition of urban centrality index. There are three essential elements
Figure 5.37: A Voronoi map defining urban spaces generated from stop locations.

Note: People traveling between stops create the physical interactions between any two areas, and this human movement is a proxy for the transfer of urban stocks such as materials, products, money, information, diseases and so on.

for representing an urban spatial structure:

**Hubs** refer to the most significant areas that connect spaces between which urban stocks are transferred. These act within the urban structure as spatial bridges between different neighborhoods.

**Centers** refer to the most relevant areas that accumulate urban stocks, which can differ from hubs but are very often the same.

**Borders** refer to socio-economic boundaries, which are generated by aggregated travel location choices that subdivide a city into small neighborhoods.

Network structure affects function, and vice-verse. Network anatomy is crucial since they tell the structure of a network. Based on the three defined basic elements, a spatial network model can be built that takes a representational or functional form of network and applies it to urban stocks and flows. Consequently, network properties are used for analyzing functions of such network in terms of promoting urban movement from three perspectives of view:
(1) Global properties to gain an overview of urban mobility.

The basic topological and planar properties of a network gives us an overall view of changing travel demands, in particular,

- **Number of nodes** indicates how many areas are accessible in total.

- **Number of edges** indicates how many areas are directly connected to each other.

- **Degree of each network node** denotes how many areas are directly connected to an area from any other, in terms of their in-degrees - those which contain trip volumes that are destined for that area, and out-degrees - those that originate from that area.

- **Strength** is the weighted degree that indicates intensity of travel - trip volumes - to and from one area.

- **The shortest path** refers the minimum network distance possible from one area to another area.

- **Clustering centrality** is an index that measures how ‘close’/cohesive’ the areas are to one another in terms of their accessibility to shared neighbors.

- **Closeness centrality** is an index that evaluates how fast information spreads in the whole area.

(2) Local information pertaining to city hubs and centers.

- **The Hub Index**: Betweenness centrality is an index which measures how well-connected an area is and is key to identifying city hubs [51].

- **The Center Index**: PageRank measures the role of a node or local area in attracting flows from all nodes in the network.

(3) Community detection to identify neighborhoods and their borders.
The borders index, which subdivide the whole land area which is covered by the network into smaller neighborhoods, are obtained by detecting what is called community structure in network science.

Spatial structure emerged from urban movement can be then detected and compared quantitatively with these indices. A more detail introduction about calculations of such indices are given in next section.

5.5.2 Measure: A Spatial Network Analysis Method

Previous work in line with the proposed analysis either ignored the network information or geographic information. The proposed method here combines network and spatial analysis through a spatial network modeling and analysis. Similar as that in measuring urban activity patterns, a work flow is given in Figure 5.38 with three main steps.

Figure 5.38: Work-flow of the proposed analysis method.

The first step is to convert the raw trip records into a network. The process starts out with the smart card data obtained from automatic fare collection systems as the input dataset. From these data sets, a weighted directed network is constructed as input to the network analysis in the next step.
In the second step, three kinds of indices are calculated through a network analysis. As indicated, the global properties provide an overall view of travel demand and interactions in the city. They are basic properties in any kind of network analysis therefore no complete details will be provided; Centrality indices are used to identify the hubs and centers in the spatial structure defining by ‘Betweenness centrality’ and ‘PageRank’. Partially based on the PageRank value, ‘community detection’ of network clusters is achieved and used to identify borders. Until now, the identified hubs, centers and borders are still abstract without any intuitive representations.

In the third step, they are mapped into a geographic space, not only to provide and immediate intuitive visualization, but also for further analysis of the spatial impacts using various spatial statistics. There are two major operations in this step, which are frequently used in geographical analysis. Spatial interpolation is applied to generate human movement landscapes. Summary statistics are finally used to group spatial units of any one community into neighborhoods, from which new borders defining the partition into a contiguous landscape of social-economic spaces are generated. Details of the three steps as well as calculations of indices are explained as follows:

**Step 1: Network construction and representation**

In this step, the recorded smart cared data is converted to an OD-matrix and then to a weighted directed network. The recorded smart card data contains detailed information for each ride as shown before in Table 5.2, and as introduced in Table 5.4, the information including ride id, passenger id, age, boarding and alighting time, boarding and alighting location, distance, fare, and an index associated with transfer trips. It is important to note that the OD-matrix is constructed from trips instead of rides. (A trip is composed of several transferred rides). The weight of the OD-matrix is the number of people traveling between two areas during a weekday. From this OD-matrix, a weighted directed network is constructed which fully captures the richness of the information contained in the data [103]. The weight of the network is the volume of travel (actual human flow) from one area to another.

Formally, a directed weighted graph is formatted as $G \equiv (N, L, W)$ that represents the overall travel on every pair of links in the city during an average workday. It consists of a set $N$ of stops or nodes denoting areas around locations, a set $L$ denoting travel between any two areas, such that $L$ is a set of ordered pairs of elements of $N$ and a set $W$ denoting the volume of
travel between any two areas. Hence \( N = n_1, n_2, n_3, \ldots, n_i \) are the nodes of the graph \( G \), and \( L = l_1, l_2, l_3, \ldots, l_i \) are the \( J \) edges of graph \( G \) with associated weights \( W = w_1, w_2, w_3, \ldots, w_i \).

**Step 2: Extracting network structure**

With the constructed network, analysis can be performed. According to defined network indices, number of edges measures how many connections between different areas exist; number of in-degree equals to the number of connection into one area and similarity to out-degree; strength is the weighted degree equals to the number of trips in reality that travels from and to an area.

Clustering and Closeness Centrality are not used for detecting hubs or centers, but they are important indicators telling the structure of a network. Therefore, they are included in the global properties and a very brief introduction is given as follows.

**Clustering Centrality** is a measurement of cohesiveness around a given node \( n \), which quantifies the local cliquishness of a network. It is defined as the probability that all possible triangles going through a node is connected. The clustering coefficient \( C_{\text{clustering}} \) of a node \( n \) is defined as

\[
C_{\text{clustering}}(K) = \frac{2 E_n}{K_n (K_n - 1)}
\]  

(5.9)

where \( K_n \) is the number of neighbors of \( n \) and \( E_n \) is the number of connected pairs between all neighbors of \( n \).

**Closeness Centrality** is a measurement of how fast information spreads from a given node to other reachable nodes in the network, which quantifies the affinity of a network. The closeness centrality of isolated nodes is equal to 0.

\[
C_{\text{closeness}}(K) = \frac{1}{\text{avg}(L(k, m))}
\]  

(5.10)

where \( L(k, m) \) is the length of the shortest path between two nodes \( k \) and \( m \). The closeness centrality of each node is a number between 0 and 1.

Beyond global properties, two kinds of centrality are used to identify the hubs and centers of a network. The first one is the well-known measure - Betweenness Centrality, which is use
for our definition of a hub. The second one is PageRank, which is a measure of accessibility in the network taking account of all direct and indirect links, their weights and their directions. This is another measure of the degree of urban centrality.

The Hub Index: Betweenness Centrality is an index which measures how well-connected an area is and is key to identifying city hubs [51]. The Betweenness Centrality of a node $k$ is the number of shortest paths connecting any two areas (nodes) $i$ and $j$ in the graph that pass through the node $k$. A node has a higher centrality $C_{\text{betweenness}}$ the greater the number of shortest paths that traverse it, and it is defined as:

$$C_{\text{betweenness}} = \sum_{ij} \frac{\delta_{ij}(k)}{\delta_{ij}}$$

(5.11)

where $\delta_{ij}(K)$ is the number of shortest paths between any two nodes $i$ and $j$ that pass through $K$, and $\delta_{ij}$ is the total number of such paths between $i$ and $j$. Sometimes this measure is normalized with respect to the total number of nodes $N$ but here it is used in this basic form.

The Center Index: PageRank measures the role of a node (a local area) in attracting flows from all nodes in the network (the whole region). The measure is a generic representation of the probability of any random walker on a network visiting a particular node. Its calculation relates directly to a first order (Markov) probability process that is used as foundation of many processes of social interaction. The basic form of calculations of PageRank was originally used for extracting information about Internet link structures. The measure used here is based on an applied method proposed in [121], in which they determine the importance of nodes in a network in analogy to Google’s PageRank [27].

In fact, this measure is implicit in the community detection algorithm, which is used below to determine community structures. The probability $r_j$ of visiting any node $j$ (or in Google’s term, the ‘page rank’ which is represented as a probability between 0 and 1) is defined as:

$$r_j = \left(\frac{(1 - \rho)}{N}\right) + \rho \sum_i r_i P_{ij}$$

(5.12)

where $(1 - \rho)$ is the probability of the walker $j$ making a random switch to any other node in the network, and $P_{ij}$ is the probability of making a switch from node $i$ to $j$ which is proportional to the trip weight on the link $i$ to $j$, that is:
The steady state probability \( r_j \) is computed by solving the linear simultaneous equations in equation 5.12 using iteration, the power method, or the appropriate matrix inversion method. The parameter \( \rho \) is a damping factor which can be set between 0 and 1, but usually is set to 0.85 used in this application. If \( \rho = 1 \), then for all nodes to have a positive probability (for all pages to have a rank), the matrix \( P_{ij} \) must be strongly connected.

Besides local information - all kinds of centralities, the organization of components of the network is also crucial for understanding spatial structures. The borders, which subdivide the whole land area, which is covered by the network into smaller neighborhoods, are obtained by detecting what is called community structure in network science.

The Border Index is generated by partitioning the network into two levels where the nodes form modules, which are communities, and the divisions between the modules are the borders. In the case of constructed spatial network in this research, communities are identified communities based on the density and interactions of flows that within each community are stronger and in volume terms greater than those between communities as shown in Figure 5.39. Therefore, the network can be partitioned into mutually exclusive clusters that are communities.

![Community structure in a network.](image)

Community detection has always been a fundamental problem in complex network analysis. According to the comparative analysis in [83], the map equation approach called Infomap developed by [121] is one of the recent algorithms that has shown excellent performance. Moreover, it is also one of the few algorithms suitable for weighted and directed networks. Essentially, Infomap considers not only pairwise-relationships, which most partitioning algorithms work
with, but also flows between pairs of nodes. It uses the probability flows created from random walks on the graph and the probabilities of visiting a node at random (which is the same as the PageRank above) as a proxy for information flows in a real system. It then decomposes the network into clusters by compressing a description of the probability flow in such a way that the average description posed by the probabilities associated with each community and those of the nodes within each community are the most dense and have minimum entropy. In short, the algorithm divides the nodes of the graph into modules or communities that are highly structured, which implies a minimum in the entropy of the partitioned graph.

This entropy is essentially a subdivision of the total entropy of the system into entropy between the modules and a weighted entropy between the modules, these weights being related to the probabilities of the occurrence of each module. Rosvall and Bergstrom [121] define this entropy as:

$$Lg(M) = H(P) + \sum_{i=1}^{m} P_i H(P)_i = -p \sum_{i} P_i log P_i - \sum_{i} P_i \sum_{k=1}^{M_i} \frac{P_k}{P_i} log \frac{P_k}{P_i}$$

(5.14)

where $P_i$ is the probability of the module $m$ being visited, and $P_k/P_i$ is the probability of the node $k$ which is part of module $M_i$ being visited. These probabilities are not the actual page ranks but the page ranks modified by appropriate exit probabilities as defined in detail by Rosvall and Bergstrom [121]. The way the algorithm works is by first setting each node in its own module and then at each step identifying the node that can be added to a module that decreases the overall entropy in equation 5.13. This process continues until no further reduction in entropy can take place and at this point, the number of modules provides a distribution of nodes within communities that is the most organized. Note that $M_i$ is a module, which contains a series of nodes $k \in M_i$ that become stable when the algorithm has converged to minimum entropy. Like all such iterative optimization procedures, simulated annealing or a related procedure is used to ensure that the likelihood that the true optimum has been reached is maximized. This then gives the distribution of nodes, or stops in this case, within each community and this distribution is then mapped to geographical locations.

This research introduces a general framework of the approach. The specific network analysis algorithm can be replaced depending on certain context. The conventional community detection methods are mostly node-based. The research here chose node algorithm based on the
knowledge generated from previous works such as [62, 138, 136] to some aspects, all proved the possibility to find geographical partitions using node-based community detection method. Only very recently, link-based methods have been proposed [4], and were later improved by [128], mainly based on a criterion-partition density $D$. To provide more complete information, this research also suggests that it is necessary to explore the possibility of edge-based community detection, which has an advantage in finding overlapped hierarchical community structure.

**Step 3: Enrich spatial information**

So far, the extracted information is only about network characters without any spatial information. The third step enriches spatial information by projecting the nodes in the network back into the geographical space. With geo-references of each node, the network is converted back to a set of spatial units. However, discrete points represent the projected geographical space, spatial interpolation is thus applied to generate a continuous movement landscape. In the context of analysis, such a landscape portrays the properties of each area. While discrete points are the stops are surrounding the area in question where assuming that people choose the nearest stop to their destinations.

A spatial interpolation is applied to the nearest neighbors of each stop. Although there are many variants of interpolation, inverse distance weighting (IDW) is used as a simple demonstration. To be noted that the proposed framework contains individual algorithms such as IDW that can be replaced and improved individually. The IDW assumes that each measured point has a local influence that diminishes with distance. The method weights the points closer to the particular location more highly than those further away, and the weights are defined generically for each point as:

$$W_i(x, y) = 1/d_{ij}(x, y)^\lambda$$

(5.15)

where $W_i(x, y)$ is the weight of the location around the point $i$ at coordinates $(x, y)$ which are nearest neighbor points to $j$ and $1/d_{ij}(x, y)$ is the distance at $(x, y)$ from point $i$ towards the nearest neighbor point $j$. Note that the weights are normalized around a particular point to sum to 1, that is $\sum_{x,y} W_i(x, y) = 1$, and $\lambda$ is a parameter which is set here as 2, which implies an inverse square law.
Summary statistics are used to assign a community to individual spatial units based on the sampled points. The main problem here is to deal with noisy points, which refer to points that belong to a community in network space but are not geographically adjacent to the main cluster defining that community as that shown in Figure 5.40.

This situation happens because the community detection algorithm is not constrained to achieve geographically contiguous areas and thus the communities that are initially detected in network space may have non-contiguous parts in the 2-dimensional space. This situation does not occur very often but when it does, it typically occurs in boundary areas where people have different travel preference to nearby centers. To remove these noisy points, summed PageRank value are computed. The points dropped on the boundary areas are the assigned to the nearest communities with the highest PageRank values. By this summary statistics, compact and geographically intact communities are produced which are geographically contiguous and exhaust the whole space.

5.5.3 Experiment: Analysis of Smart Card Data in 2010, 2011 and 2012

In this experiment, smart card data is used as input. In order to compare the changes of spatial structure, data collected in three years are used. The collected tap-in/tap-out events offer a huge data set, with around 5 million daily travel records. Noted that data set is chosen due to its availability. In September 2010, only data of one day is available. In April 2011 and September 2012, data of one week is available. For a fair comparison, data of one average weekday is used to evaluate the feasibility of proposed method in exploring emerging spatial structure in Singapore.
(1) Preliminary data processing

As indicated previously, an OD trip volume matrix is constructed from the original smart card data. Each node in this network denotes an area with one stop inside. The network does not to be node-based strictly, other partitions such as grid-based partition as that used in last chapter may also work here. The purpose of node-based partition is to divide the space into smaller spatial units.

Through a first glance of the OD-matrix, it is easy to find that the overall travel activity in Singapore using public transportation system reveals a very regular pattern with the usual morning and evening peaks. The peak hours appear almost exactly at the same time every day in the same areas and the overall distribution curves are similarly shaped to one another. This very regular travel behavior also reveals in previous Section 5.3.2 - mining travel behavior from smart card data. The regular pattern proved that the constructed network using an average working day is reasonable. Besides, as indicated previous, public transportation has a big share in transport mode in Singapore and the share keeps on growing through years. From the geographical mapping, the origin and destination of travels through public transportation and all transport modes has almost the same coverage geographically as that shown in Figure 5.13. Since the destination points form convex having almost the same size. It means that, in the case of Singapore, public transportation can be used to represent overall mobility, and covers the whole array of daily activity types.

Figure 5.41 illustrates of two types of mapping. The top image shows the network mapping at an early stage in the work-flow and highlights structure but neglects geographical information so that local changes cannot be detected. On the other hand, the image at the bottom shows a traditional geographical mapping from which structures can be barely identified, but local relevance is clearly visible. Thus, the proposed approach is attempted to combine the two representations in order to obtain the missing information.

(2) An overview of urban movement

After data processing, there are 621731 edges linking 4638 nodes from the 2010 data, 702803 edges linking 4716 nodes from 2011 data, and 730885 edges linking 4727 nodes from the 2012 data. Network properties and indices were computed using the i-graph package on
Figure 5.41: Two varieties of network mapping.

Note: Top: The weighted directed graph constructed from smart card data; nodes represent the module it belongs to and the larger the nodes, the higher the total PageRank of its module. Bottom: nodes mapped into geographical space in proportion to analyzed property values, in this case by node degree, which is mapped to node size.

the R platform (http://igraph.sourceforge.net/). Community structure was generated using the tool Map Equation(http://www.mapequation.org/). Spatial analysis was conducted on the ArcGIS platform (http://www.arcgis.com/).

Table 5.7 shows the global network properties for the years 2010, 2011 and 2012, and from the table, some explicit changes can be read from the numbers:

- The number of edges has increased, which means more areas in Singapore are connected and the whole city becomes more accessible in general.

- Strength in terms of trip volume has increased in total and on average, which means there are more and more people are using public transportation. It could because of the
Table 5.7: A comparison of network properties with smart card data in 2010, 2011 and 2012.

<table>
<thead>
<tr>
<th>Indices</th>
<th>Year 1997</th>
<th>Year 2004</th>
<th>Year 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>BUS: 4599</td>
<td>BUS: 4599</td>
<td>BUS: 4599</td>
</tr>
<tr>
<td></td>
<td>MRT: 107</td>
<td>MRT: 107</td>
<td>MRT: 117</td>
</tr>
<tr>
<td>Number of edges</td>
<td>621730</td>
<td>702052</td>
<td>725046</td>
</tr>
<tr>
<td>Average degree</td>
<td>131.8342</td>
<td>148.866</td>
<td>153.4164</td>
</tr>
<tr>
<td>Average trip volume by link</td>
<td>645.5789</td>
<td>788.577</td>
<td>801.2078</td>
</tr>
<tr>
<td>Average shortest path length in kms</td>
<td>2.229015</td>
<td>2.196655</td>
<td>2.185142</td>
</tr>
<tr>
<td>Clustering centrality</td>
<td>0.2116035</td>
<td>0.2238426</td>
<td>0.2268748</td>
</tr>
<tr>
<td>Closeness centrality</td>
<td>1.161199e-06</td>
<td>1.170022e-06</td>
<td>1.085218e-06</td>
</tr>
</tbody>
</table>

increasing share of public transportation in all transport modes and also the increasing population.

- The length of shortest paths has decreased slightly indicating everywhere in Singapore, which means areas in Singapore are connected to each other more tightly. Information can be easier transmitting across the city.

- The increasing average degree means that each BUS/MRT stop has more connections to other stops/stations, though the total number of stops/stations did not increase from 2010 to 2011. Possible reasons led to this increase might be the newly added bus lines or more active human behavior due to an increase in economic and related demand.

- Though traffic jams still exist, increasing clustering centrality and decreasing closeness centrality shows that transferring between lines and modes in Singapore has gradually become more convenient and efficient.

5.5.4 Insights of Polycentric Urban Transformation

To anticipate the ultimate outcomes of conducted analysis, the emergence of sub centers and communities for Singapore based on the data for 2010, 2011 and 2012 is shown in Figure 5.42. In this figure, three regionalizations or partitions of the Singapore are taken from network analysis of communities based on Rosvall and Bergstrom’s method.

From the view of geographical mapping (top), it is clear that at an emerging neighborhood in Toapayoh area. In overall, Singapore has been partitioned into smaller neighborhoods emerging
Figure 5.42: Changing communities and borders detected from daily transportation in Singapore from 2010 to 2012.

from urban movements (top row). A representative emerging new neighborhood is highlighted in the center row. The overall partition of the space and the emerging new neighborhoods over the 3-year time series reveals rapidly changing polycentric urban transformations.

From the view of flow diagram (bottom), the ranking of importance/urban centrality of the partitioned areas remain stable in overall. But locally, there are flows exchanging between the partitioned neighborhood indicating growth and shrinking in the urban process. The alluvial diagram (bottom) shows the changing values of network attributes in terms of significant communities with highest PageRank (values shown in rectangles), as well as the changing organization among these communities (interchanging flows). All this is explained in detail in the sequel. The rest of this section gives more details about the insights gained from the analyzed results.

(1) City hubs and centers anomalous centrality
Figure 5.43 shows a plot of the degree and average trip strength for the years 2010, 2011 and 2012. In the constructed network of human movement, there are a limited number of areas that have very high and intense connections to the other areas. Together with the relative short length of the shortest paths in the network, this is indicative of the ‘small world’ phenomena in the network over each of the years. However, a strong conclusion cannot be drawn from this result, since the constructed spatial networks tend to be planar and in their pure form, do not demonstrate small worlds.

In Figure 5.44, the distribution of degrees in 2010, 2011, and 2012 are compared. It shows that this distribution is becoming slightly more even over time. In other words, it appears that travelers have more diverse location choices for their activities, and their average activity spaces are becoming larger.

Figure 5.45 is a plot of Betweenness Centrality. Similarity, centrality of different year is plot in different color for comparing the changes. It shows that the number of areas with
Figure 5.44: Changing degree distributions in 2010, 2011 and 2012.

Note: There are few nodes with a very high degree, which results in a very broad tail of the degree distribution. For a better view, degrees < 1200 is shown in a magnified figure (top right).

lower betweenness centrality have slightly decreased, while the number of areas with higher betweenness centrality have increased.

Figure 5.46 is a plot of PageRank. Only slight changes can be found from comparing the PageRank distribution in three years. In general, if the number of highly centered areas has deceased while the number of secondary centered areas has increased, this implies a polycentric urban transformation where the influence of strong center areas has gradually reduced, their centrality increasingly shared with emerging sub centers.

The calculated network properties were then projected into geographical space to generate urban movement landscapes, from which, the locations of hubs and centers can be identified. As shown in Figure 5.47 and Figure 5.48 are two interpolated maps of computed centrality index, namely Betweenness centrality and PageRank. There is barely any changes in geographical distribution, therefore, only centrality of 2011 is shown here as a demonstration of the proposed method.

By comparing these two maps, an anomalous distribution appears. Those city hubs that are most efficiently connected are not necessarily the most central areas. This is a finding
Figure 5.45: Changing distributions of Betweenness Centrality in 2010, 2011 and 2012.

Note: The overall distribution becomes more concentrated. Higher Betweenness centrality is associated with fewer areas.

Figure 5.46: Changing distributions of PageRank in 2010, 2011 and 2012.

Note: The overall distribution shows slight changes while the number of highly centered areas slightly decreases.
Figure 5.47: Interpolated Betweenness Centrality landscape in 2011.

Note: The areas in red are detected hubs that are consistent with locations of the MRT stations.

Figure 5.48: Interpolated PageRank landscape of Singapore in 2011.

Note: The areas in red are detected centers.
that is implicit in our observations even though it tends to fight against our intuition about the role of centrality and accessibility in cities, which traditionally have been monocentric. More specifically in Figure 5.48, the PageRank map shows that the central area is one of the most visited and most significant places, but also shows that the most efficiently connected areas are not only found in the city center, but in many other areas across the whole island. Indeed, these hub locations are almost perfect matches with key points defined by the MRT lines. This means that the MRT lines have a significant position and serve as the wider skeleton linking all regions of the city state together. In fact, this finding is consistent with Singapore’s physical concept plans. Back in the 1970s, transportation was prominently considered in shaping the structure of the city. According to the various concept plans, high-density public housing areas were planned along high-capacity public transportation lines, near to industrial areas and to other employment. And to an extent, this is now borne out in the patterns of accessibility and transport usage revealed from the smart card data.

The network landscapes are also changing like natural landscapes but these are driven by multiple forces, including new development in the city, advances in the infrastructure of the transportation system, and the way peoples’ individual choices have been augmented. Combining the maps with the plots, some trends can be seen. The changing Betweenness centrality indicates that the most connected areas (the city hubs) largely coincide with MRT stations and these are likely to function more intensively. It also means that the development of the MRT promotes longer distance travel because the population can easily travel to areas that are more central from anywhere in the system.

However, the slight changes in Figure 5.45 as well as Figure 5.46 does not provide us with very strong evidence of urban transformation. As a supplementary analysis, this interpretation is reinforced from the generated borders of urban movement within different communities described as follows.

(2) Borders and new neighborhoods - entangled community structure

Borders are important elements that subdivide the entire space into smaller communities. These serve as an important reference for measuring and analyzing the urban data in terms of the original urban structure, the administrative borders, which were planned throughout the 20th century. They are historical markers that represent past human interactions during the last 100
years.

The generated borders, which are emerging border from daily movement will be mapped and compared to administrative boundaries. The changing communities in terms of volume of flows, number of communities, and their sequences were previously shown in Figure 5.42 using the concept of the alluvial diagram according to [122] based on data taken from the different community clusters at the three points in time 2010, 2011 and 2012.

![Figure 5.49: Borders defining communities of urban movement in 2012.](image)

Note: Community structure detected from smart card data using Infomap marked in different colors. The black boundaries indicate the original administrative borders. In the right corner, planned decentralization of urban form is drawn based on the 1991 concept plan, which is quite in line with the overall structure of urban movements.

Only first layer of community clusters is used for generating borders. A hierarchical structure is failed to be generated, because in the case of Singapore, only this layer of communities generates clear geographical partitions of neighborhoods. At lower spatial levels, the neighborhoods are entangled, which indicates a random distribution of peoples’ activities in smaller spatial areas.

Figure 5.49 is the results for 2012. The figure is enlarged for a better comparison with the original urban plan. In the figure, Singapore has been subdivided into nine small regions
that are the most significant communities detected from the network analysis. As introduced in the measure, to clean up the noise in these results, data aggregated is conducted to sum points into subzones which are equivalent to the smallest levels of geographical subdivision used in Singapore’s national statistics. Summing the PageRanks determines the most significant community. The original results before data cleaning can be found in Figure 5.50.

Another insight found from the results, which could be applied to a much boarder cases. As introduced earlier, the actual network contains no geographic information per se. The community structure is generated from the natural patterns within the network itself. Communities forms by something in common among all the community members [101]. The common characters of urban space could be economics, land use, people, and so on. However, after several iterations of the detection algorithm, a clear territorial subdivision emerges. These results show that spatial impact is the most prominent factor that influences people movement in cities and their interaction. When comparing the generated borders of human movement in 2012, it is clear that these borders have shifted a little bit west because of the development of new centers.
such the Jurong East area in the west. This conclusion is in line with what has been found in last section.

At a larger scale, this phenomenon also matches the planned “decentralization of urban form” which was part of the revised concept plan of 1991 where the emphasis was on facilitating sustainable economic growth through the idea of decentralization. The city was then planned to be surrounded by four regional centers, located in the west, north, northeast, and east, several sub centers and fringe centers, as shown in the inset in Figure 5.6. This decentralization is part of a top-down panning process that will likely take decades to realize as some sub-centers are still under development. Detecting these trends of change does indeed provide deeper information for planners and designers to evaluate their plans or to link these plans to their actual realization on the ground.

This research attempts here to track the path of changes by comparing the analyzed results of the data in 2010, 2011 and 2012 as shown originally in Figure 5.50. It shows that though there are some significant changes in flows between communities, the most important communities remain the same, with only a few changes in their sequence with respect to their summed PageRanks.

An obvious and gradual change from 2010 to 2011 shows there is an emerging new community. When mapping the nodes as shown in Figure 5.50, all the nodes in this new community are falling into one area, the Bishan, Toa Payoh and east Novena area. If compare it to the concept plan of new centers shown in Figure 5.49, the emerging sub community consists of one of the sub centers and this suggests that Singapore is slowly becoming more polycentric. Moreover, the emergence of this new community has occurred within only one year, illustrating the rapidity of the urban development process in Singapore. But, this results can not be taken as very strong facts implying the ultimate outcome of these development processes in Singapore since this is only a snapshot of change.

When comparing these results from 2010 and 2011, certain differences with respect to the flows can be found. The difference of the PageRank among communities even out a little, which means, the share of flows to each community becomes more balanced. From the geographic perspective, the results show that the areal sizes of communities also becomes even. In addition, an interesting finding is that the south-west area, which is an isolated area in 2011, disappears and is dissolved in adjacent neighborhoods in 2012. The reason for this change is likely to be because of the extension of the MRT lines, which started operation across this area in early
2012, making this region much more accessible to the rest of the network. Even over this short period of time, our results show how quickly and how strong the transit system influences the pattern of urban movement and the communities that define it. In summary, all these insights from the analysis reveal that the Singapore urban system is becoming ever more polycentric and diverse as developments spread throughout the city-state.

5.5.5 Discussion

This section presented a spatial network analysis, which is considered as a novel and useful approach in the following sense. Firstly, it is a quantitative method for detecting urban hubs, centers, and borders as well as changes in the overall spatial structure of urban movement using daily transportation data. An appropriate work-flow is presented. Secondly, a systematic analysis is given linking measured parameters with real urban phenomena, which is applicable to new methods of identifying communities based on mobility; and thirdly, the proposed method is validated from novel insights into the actual development of Singapore. By comparing the results from data from three years of big data associated with smart card data set, besides the similar insights of polycentric urban transformation as that found in last section, the results shows a very fast development of Singapore. Even from such a short time series, Singapore is changing rapidly. To summarize, this approach yields important insights into urban phenomena generated by human movement. It represents a quantitative approach to urban analysis, which explicitly identifies on-going urban transformations.

A comparison should be made to the measure presented in last Section. This section presented a spatial network model; the spatial network analysis presented in this section is another measure of urban centrality index but not an exclusive one. The centrality index measured in last section measured from urban activity patterns, and thus identifies spatial structure from the spatial distribution of urban stocks. This section develops another form of centrality, which detects emerging spatial structure from urban flows. Both of these two measures detect functional changes.

Moreover, it is another example of the presented data innovation, which makes use of newly available big mobility data. It represents an important way to examine the impact of infrastructure development on peoples’ lives and in reverse how cities have been reshaped by individuals’ needs to travel. There is still much to do by focusing on extra information from this kind of high spatiotemporal resolution, and cheaply collected urban data. This will undoubtedly contribute
to a better understanding of urban dynamics, in terms of human behavior, movements and urban processes, and the template established here shows the direction in which the future research should go.

Finally, it is important to reemphasize that the presented spatial network analysis belongs to the family of spatial analysis and modeling approach. The spatial network model is an analogy mode that takes a representational or functional form of network and applies it to urban stocks and flows. Properties of the network are redefined in the context of urban structure and used to interpret the characteristics of areas within the total urban system. In fact, many more properties can be defined and the presented analysis only indicated the richness of this approach in so doing. As defined previous about urban modeling, this model can be further implemented as programming models that embed in a software tool to support urban analysis. As a proof of concept, a prototype is development in next section.

### 5.6 A Visual Analytics Framework for Spatial Analysis and Modeling

In most of the cases, the massive urban data cannot be properly and/or efficiently used by urban planner and designers mainly due to the unmanageable data volume and the lack of analysis method to convey raw data to meaningful information. The analyses in previous sections provide means to extract meaning information from available data sets, but not yet address the problem of data or/and information management. One possible solution might be a communication tool that brings the extract information back to designers or planners and visualization is the tool. This section addresses such issue which is also the last component of the designed data supported design process proposed in Chapter 3.3.Here, a visual analytics framework is presented to combine propose analysis method with interactive visualization resulting in a computational design tool.

As introduced before, visual analytics is more than customized visualization. Visual analytics aims at multiplying the analytics power of both human and computer by finding effective ways to integrate interactive visual techniques with algorithms for computational data analysis. Therefore, visualization and computation can interplay and complement each other [75, 8]. Though a visual analytics framework could be generic, the one presented here focus on integrating geospatial techniques and utilizing mobility data. The mobility data used here can be
categorized as spatiotemporal data, which are data in numeric time referring to same location in space. However, integrated analysis method is the emphasis, modeling on temporal dimensional is seldom included in the previous analysis and will be research in the future work.

A review about geo-visualization and visual analytics has been given before, the author only re-emphasizes aims of the propose framework here: firstly, it is designed to help users to better explore urban data. Secondly, it explores the way to make the analysis process transparent to users therefore a better understanding of data and the analysis can be gained. Thirdly, it is used as a way here to make analytics model alive that real-time analysis can be gained. If so, the tool can be used as a data service based decision supporting tools that quick feedback can be provided when making new design or planning proposals. There are two major contributions of this tentative work:

1. A generic visual analytics framework is presented, which is based on a geospatial pipeline. It integrates the proposed analysis method into geo-techniques supported work-flow, which allows users to explore and manipulate the data interactively.

2. An example is given which applies the presented framework to build a flow map. The spatial network model introduced in last section is embedded in the framework. A prototype of a flow map is development. Sampled transportation data is used as input to illustrate the feasibility of the proposed framework and analysis method.

5.6.1 A Visual Analytics Framework

The proposed of visual analytics framework can be decomposed into three components as shown in Figure 5.51, which are explained as follows:

A GIS-based data processing pipeline serves as a basic collected and sensed data processing engine. The input is a raw data set. The outputs are data views at different levels of detail (LODs). LOD1 outputs cleaned up data sets; LOD2 outputs aggregated data, generated by simple database queries. LOD3 outputs aggregated data, which is generated by certain spatial analysis method.

An analytics method is the core part of the presented framework. Spatial data will be input into analytical models, and formatted in required structure. For instance, a grid based structure in the analysis of urban activity patterns in Section 5.4, a spatial network structure in Section
5.5. Urban indices are calculated in the analytics model and output as properties of spatial objects such as stops, buildings, areas and mapped to a geographical view as a view in LOD3. Interactive operations are used to explore the data sets and to interact with them. A graphical user interface is implemented that allows users to select add travel records to perform a real time analysis. These user interactions are supported to facilitate the understanding of the process of data analysis and the analyzed results.

5.6.2 Application: A Flow Mapping Tool

Geo-visualization is a powerful tool that can convey information to different domains and therefore improve communication. The task here is to visualize massive urban flow, in particular, traffic flows. For visualizing kinds of flows, traditional flow maps simply use arrowed curves with various sizes and colors to represent information in terms of direction, volume, and speed. Many examples can be found such as the famous flow map by Charles Minard showing of Napoleon’s march; computational wind maps (http://hint.fm/wind/), interchange flow charts [161] or spatiotemporal visualization of trajectories [167]. These examples provide intuitive views of flows. However, analysis of important properties and structures of flow networks are missing. It has been explored in [64] that using network analysis methods to aggregate areas, which are nodes in the context of a flow map. Similar work has been done in other applications using network analysis [115, 138]. Based on these previous achievements, research in this dissertation continues exploring the use of network analysis methods and integrate it into a GIS-based framework, explained as follows:
(1) An integrated data processing pipeline

As mentioned previous, the overall framework is built on top of a traditional GIS data processing pipeline. Therefore, GIS is used in this work as base for data management and processing that multiple data sources will be reformatted into uniform structures as inputs of the pipeline.

![Figure 5.52: Data structure in network space and geographical space.](image)

Note: The red line with two arrows shows the correspondence between elements in two spaces.

A mechanism is used to integrated analytics model into the data processing pipeline. More specifically, data are modeled and represented in two data structures. In the case of the presented flow map, a polygon is a spatial object - an area in geographical space and a network object - a node in network space. The first one is the traditional geographical data structure. The second one is a network data structure. Elements in these two structures are corresponding to each other. As shown in Figure 5.52, a node denotes a spatial unit (area) and edges denote traffic flows between two areas. The objects in two spaces refer to the same data sets but enrich the data with different semantics. By such way, an analytics model is integrated into the data processing pipeline.

(2) Spatial network analysis
In the network space, nodes and edges together construct a network representing spatial interactions. As that already introduced in Section 5.5, a spatial network analysis is conducted to uncover the hidden information of urban movements. Network properties such as PageRank and Community, can be used as urban indices which are used here as demonstrations to show the proposed visual analytics tool can be used.

This section focuses on presenting the kinds of mechanism to implement a visual analytics tool. Mathematics behind these measures can be further referred to related works in complex network analysis like [101] or previous introduction in section 5.5. Only a very brief review is given here.

PageRank measures the role of a node in attracting flows from all nodes in the network. The measure is a generic representation of the probability of any random walker on a network visiting a particular node. In reality, it measures the role of an area in attracting urban flows such as people, information and so on. The areas with higher value of PageRank are important urban hubs for transferring and exchanging urban stocks and flows.

Communities in a network are generally defined as groups of nodes with dense connections internally and sparser connections between groups. In reality, communities refer to neighborhoods in which, people have more internal movements than that going outside. The community’s structure is generated by the nature patterns of the network itself. It is matters of common experience that people do divide into groups along lines of interest, occupation, ago, and so on. In the case presented here, urban traffic flow is a proxy of interactions between spaces. The nodes denote areas in reality. The areas which have more internally interactions are closely connected and clustered into one community.

(3) Interactive geo-visualization

In the third component of the framework, properties of original data and extracted information from the network analysis are mapped back to geographical space, and queried and explored by interactive operations. Two main feature should be addressed here, namely data aggregation and linkage operation.

Data aggregation for Information query: Multiple levels of detail data views are achieved by data aggregation techniques, which are widely used in many visualization applications to
Figure 5.53: Three levels of details.

Note: Area is referring to different elements in reality. Trips are aggregated by stops, subzones, neighborhoods that defined by community detection.

simplify flooding information and give clear views. In our case, the first level of data view is the row data set as shown in Figure 5.53. The second and third level data views are achieved by basic and advanced aggregation methods that can be explain as follows: A commonly used method is to aggregate data by certain attributes. For instance, when do data aggregating, parameters such starting time, ending time or even personal information can be used as conditions. This aggregation can be easily done by standard database query functions. Geospatial statistics provides a spatial joint which aggregates data by location information. As a higher level of data aggregation, analyzed results are used. Here results from the network analysis are mapped. Subzones are aggregated into big neighborhoods corresponding to the communities detected from network analysis, which reflects the spatial structure emerging from urban movements.

Linkage function supports interactive operations between geographical and network window: Besides basic operations, like zoom in and zoom out to get views with different levels of details, click to query the information of lands and flows, a linkage operation is the highlights of proposed flow map. Since the objects in two data structures are corresponding to each other, computed results in the network view will be directly mapped to the geographical view. Vice verse, changes in the geographical view are reflected in the network view. Instead of a black box, this linkage operation between two views provides a transparent way to users for a better
understanding and controlling analysis processes.

(4) Implementation and results

A prototype of flow map is development and sample data are used as input to provide an interactive visualization of Singapore. This prototype of flow visualization is implemented in Java, using the third party dynamic graph library GraphStream\(^5\). The preliminary data processing is done with ArcGIS. Input data is a sample set from one-day public transport smart card data in April 2011.

As show in Figure 5.54 is a first sight of flow map. Curved links shows flows between different areas. Areas which have flows in or out will be lighted with colors. The color is assigned according to calculated PageRank value - in other words, the comparatively attractiveness of an area in the global urban space.

![Figure 5.54: A flow map.](image)

Four spatial scales - region, zones and subzones as shown in Figure 5.55 and stops shown in Figure 5.56. These are be switched automatically when a user zoom in and out the view. These three spatial scales are corresponding to different levels of data aggregation by simple spatial joint.

Figure 5.56 shows the two types of views provided by the presented tool. A network view is given on the left side, while a geographical view is on the right side. Dish lines are added indicating the green dots in both views are referring to the same data. Green dots denote nodes in a spatial network and stops/stations in urban space.

Figure 5.57 shows a simple query function at a subzone scale. By clicking one zone, all

\(^5\) GraphStream, [http://graphstream-project.org/](http://graphstream-project.org/), accessed in 2014
connections between this zone to the others are shown as curve-lines. By this, flow volumes and connections of zones can be visually compared.

In Figure 5.58, a real-time analysis is demonstrated. Besides analyzing collected data sets, users can also add data by themselves. When data is changing, PageRank will be re-computed in the network space and results are shown in the geographical view simultaneously. Shown in the figure are two views before and after change the flow data. In figure (top), selected 25 subzones are selected as a test case. The traffic flows between 25 subzones are shown in a network view (top left), the calculated centrality value (PageRank) is mapped with colors (red to blue, high to low) in the geographical view (top right). In the figure (bottom), flows are added between a subzone in the middle part and towards the other subzones. Local centrality values across the whole space are changing meanwhile. This is a typical application that this research wants to illustrate. Most of the state-of the art analysis tools perform well in terms of identifying local impact, however without a global view of the other areas. Spatial structure comes out of kinds of global distribution of urban stock and flows. It is an even more extreme case that needs both local and global analysis.
Figure 5.56: Two views in the tool: network view and geographical view.

Note: Elements in each view are corresponding to each other. This example shows the linkage functions. When the tools started to load trip data, the geographical view is adding links between stops, and on the other side, the network view is add nodes and links.

Figure 5.57: Visualization of flows at subzone level.

Note: Trips are aggregated by subzone. By selecting in visual zones, you can get detailed information. By a visual comparison, you can see that subzone in left figure has less interactions than that in right figure.
Figure 5.58: Real-time analysis of changing flows.

Note: When Add flows from and to one area in the geographical view, with linkage functions, nodes will be added in network view immediately and PageRank will be recalculated.
This tool could be used by different groups of peoples. Planning decision makers, who are mostly concerned about the global distribution of people, can map and obtain insights of spatial structures and urban movements. Urban designers who want to use big data for urban studies, such tools are a way to convey the massive data into readable views. Transportation planners, who are mostly concerned about traffic conditions, can have a better idea of the impacts of their decisions on transportation planning on the distributions of urban resources.

5.6.3 Discussion

In this section, a generic framework of visual analytics tool is presented as an effective communication tool to convey extract information to designers and planner. With this framework the proposed analysis can be further developed as planning decision supporting tools. The implemented prototype as well as case study shows the feasibility of the proposed framework and method. With this framework, the proposed strategy in section 3.3 which uses data service to supported urban design process becomes complete.

In general, this approach makes the big data usable and computable to non-technique users. It is not a kind of data innovations *perse*. But it undoubtedly facilitates the data innovations by converting theories and techniques to practical tools.

As follow up work of this research, there are still much potential to explore. In this tentative work, primary data processing is done separately in ArcGIS, parts of network analysis is also pre-calculated due to limitations of computing power, while visual analytics is done with a self-developed tool. To integrate these parts into one platform to achieve real-time big data processing and analysis is one direction. Other improvements will be made to make the framework more adaptive to integrate various analysis and modeling methods.

5.7 Chapter Conclusions

This chapter presents a case study of Singapore’s polycentric urban processes, including a historical review of morphological changes and a set of analyses of functional changes using transportation data. The work in this chapter is a practical implementation of the theoretical research design introduced in Chapter 4. The organization of the analysis can also serve as a template for the analysis of other urban processes, and is not limited to Polycentricity. In particular, there are five aspects to conclude:
(1) New definition and measures of Polycentricity.

The measures of Polycentricity correspond to previous arguments about its fuzzy concept. Polycentricity has been examined in this chapter from individual to aggregated levels, combining morphological changes of physical urban space and functional changes of socioeconomic space, and quantitatively measured from both urban stocks and urban flows.

(2) Analyzing functional urban changes from human behavior.

This research looks into the aspect of human behavior in urban transformation. Three different levels are investigated. On a small scale, individual travel behaviors are analyzed; on a medium scale, regional centers and urban activities clustered in the centers are compared; on a large scale, emerging center, hubs, and borders are detected. Together, both the individual and collective effects are examined.

(3) Linking functional changes to morphological changes of Polycentricity.

Functional changes reflect how people use urban space in reality. These functional changes are consequences - as well as causes - of changes in the built environment. On one hand, linking physical changes and functional changes is an evaluation of the original plans by making comparisons with reality; on the other hand, it results in a better understanding of the interactions between people and space. These kinds of studies are important for evaluating urban plans and uncovering urban problems.

(4) Measuring changes quantitatively through an advanced spatial analysis method.

Different formats of urban centrality indices are defined to measure urban stocks and flows using transportation data. A qualitative interpretation of the various quantitative indices is also given and it enriches the analysis with a semantic interpretation that is meaningful to urban planning applications.
(5) Using urban data in more innovative ways.

The analyses in this chapter have revealed an alternative approach to the study of urban dynamics than the traditional macro-analysis of urban structure. This is primarily due to the availability of new data sources and techniques. Some examples of data innovations are demonstrated, such as extensively using data by fusing two data sets, reusing travel survey data for other purposes, and using open data for urban studies.

In the future, more work could be done along these lines. The methods used here could be applied to other forms of urban location data, such as food chain analysis, package delivery, and other systems that involve flow data such as migration, trade, various materials, and, of course, information between different spatial locations. Moreover, further analysis could be done, for instance, using a node-based community detection method to uncover overlapping and hierarchical neighborhoods; comparing differences in movements between weekdays and weekends; or finding out the causes and consequences of changes by adding other thematic data sets with proper statistical methods. More advanced methods are waiting to be developed. As new data becomes available each year, this type of analysis should be updated and deepened.

The work here is just the first step toward a better understanding of urban complexity. More details about the causes and consequences of changes should be examined, which need to be interpolated by mining information from other data sources, such as GDP, population censuses, and housing markets. There is still much to be achieved by focusing on integrated techniques using multiple data sources for studying urban processes.

In sum, this work contributes to a better understanding of urban dynamics in terms of morphological and functional urban changes. The methodology can be applied not only to the case of Singapore or a unique phenomenon of Polycentricity, but also to other case studies and other urban processes. The template established here shows the direction for future research.
Chapter 6

Synthesis and Conclusions

There are two parts in this Chapter. Section 6.1 presents a synthesis of the results and a comparative discussion of different analyses in the conducted case study. The aim is to sum up the insights made to the urban transformation of Singapore; and in a broader sense, the phenomenon of Polycentricity. Beyond that, a methodology that can be applied to urban studies on urban processes using urban data is inducted. Section 6.2 concludes the accomplishments achieved in this research and posits future research directions.

6.1 Synthesis: An Overview of Findings

This section synthesizes the findings of this research into four aspects organized from phenomenon to essence as follows:

1. Insights into the development of Singapore with a focus on urban decentralization. The three most significant conclusions are highlighted, based on comparing and linking results generated from measures and reviews in Chapter 5.

2. The measures of Polycentricity using dynamic data sets. Five major characteristics of the redefined Polycentricity are summarized. Based on these definitions, key indices used in this research for measuring Polycentricity are listed.

3. Integrated spatial analysis and modeling approach that proposed and tested in this dissertation. This section aims to do an inverse study that abstracts research methodology
from the applied case study. The methodology and methods used in this research are considered generic, and can be applied to a broader range of similar research.

4. The potential use of large data sets in supporting urban design and planning. In view of the larger debate on the practical value of “big data”, this thesis shares experiences gained from the conducted data applications.

6.1.1 Insights into the Development of Singapore

The case study in this dissertation examines both the physical and functional changes of Singapore. Due to the data availability, data sets used for detecting changes cannot be synchronized over the entire period as shown in Figure 6.1, and they do not have the same temporal resolution. However, some inter-dependencies between long-term and short-term changes are already revealed through analysis of the results such as the changing speed and changing path.

**Figure 6.1: A time-line of study materials used in this research.**

**Fast Development**

As reviewed, the first Master Plan in Singapore was developed in the 1950s, influenced by a British notion of order, regularity, and modern town planning. However, the plan was quickly rejected, because the Singapore Government wanted to pursue a drastic transformation of the city-state rather than have it undergo social and economic changes at a slow and steady rate [160]. The expectation of fast development is not just a imagined plan. Looking back at Singapore’s history, it can be seen that Singapore has gone through a very swift transformation that is still ongoing in many aspects, including population, economy, urban infrastructure. In
particular, the impact of such changes on urban activities, and mobility revealed from the analyzed result of smart card data. Even though the analysis is limited by the availability of data sets from only three years, it can be seen even from such a short time series that Singapore is being developed towards a polycentric urban form, where new sub-centers and communities are emerging and growing to a balanced size that is largely in line with the city’s master plan. Moreover, it also shows the high speed of the development of Singapore since the large scale changes are visible in a matter of a few years.

A Top-Down Planned Polycentricity

Though Singapore represents a model for changes in many urban settings, its success can hardly be copied. The same conclusion has been drawn in other studies about the urban morphology of Singapore such as [47, 69]. Many problems usually encountered in fast development are overcome in the case of Singapore, mostly because its development is driven by well-organized plans.

In short terms, the shifting of human activity clusters matches very well with trend of physical development of Singapore. For instance, from the analysis in Section 5.4, the rising and falling of centrality value in Hougang area before and after 2004 might be caused by the continuous development of new neighborhoods in that area in the 1990s, but the opening of a rapid train line in the 2000s led the flow of people to go outside; from the analysis in Section 5.5, the merging of west coast areas to the big west region after opening of a part of yellow MRT line in 2011.

In long terms, the polycentric urban form is greatly shaped by urban plans, especially the ring plan in the 1970s and decentralization plan in the 1990s. Transport planning also contributed to this urban process. Especially in the early date like 1970s, high-density public housing area was arranged along proposed high-capacity public transportation lines; low and medium housing area was beside the corridors and served by road based transport system; industrial areas and other employment centers were located close to public transport. These urban settings initiated the early structure of Singapore. From the analyzed result of transportation data, the consistency of land use and activity patterns reveals the compatibility of transportation planning and land use planning. Especially, public transit system has particularly significant influence on shaping both physical and functional spatial structure. From analysis of both surveyed data and smart card data, an increasing importance of MRT lines in daily transportation
is clearly shown. As you may see from the analyzed result of smart card data in Section 5.5, the most connected nodes in the spatial network are quite overlapped with MRT lines, which acts as hubs contributed greatly to the overall transportation in Singapore. The detected changes over years also show that the opening of MRT systems reveals its impacts on urban movement in very short time.

**Emerging Bottom-Up Changes**

Besides top-down planning, there are also bottom-up changes ongoing at the same time. The urban development of Singapore in previous years was mostly carried out to meet basic living demands of inhabitants. Once this basic requirement had been fulfilled, people started to seek more diverse lifestyles. The result of this might be a loss of control of a planned urban process, because more options with equal costs are offered and more factors are taken into account when making a location choice. Consequently, the uncertainty in urban development increased.

Taking the regional development in Singapore as an example, the initial purpose of new developed centers such as Jurong East was to distract flows from the old CBD area to sub-centers. However, the analysis results show that the centrality in the CBD area is continuously increasing instead of decreasing. One possible reason is the advance of a long distance massive transport system that encourages people to travel long distances from everywhere in Singapore to the biggest and oldest center (the CBD). The result is a negative impact on distracting flows that go somewhat against the original idea of Polycentricity. How the city will be shaped by these two contrasting forces is still unknown. A second piece of evidence attesting to the bottom up changes is that some other emerging sub-centers, such as the Yishun area, have an even higher centrality than the planned sub-centers. Finally, the increased travel distance and comparatively stable travel time also presented a convincing explanation. People have more location choices over a wider range of traveling distances that lie within an acceptable travel time. In that sense, figuring out how to evaluate and predict the outcomes from multiple driving forces, and how to manage them will be another challenging task that requires cooperation between different government agencies. In a broader sense, integration on many levels are required to understand urban complexity, urban dynamics, and bottom-up changes.
### 6.1.2 Defining and Measuring Polycentricity

The conducted case study detects urban changes in Singapore, and focuses on tracing its polycentric urban transformation. The case study is an interpretation of presented definition in practical contexts and an evaluation of corresponding indices proposed for measuring polycentricity. The key concepts of the presented definitions and measured indices are summarized as follows.

**Definition of Polycentricity**

The presented definition of Polycentricity is made on two bases: the debate of the fuzzy concepts and its measurement. The improvement of our understanding of Polycentricity can be gained from the newly available human mobility data. Five major points are addressed below:

1. **Polycentricity is a specific type of spatial organization of clusters.** Therefore, spatial distribution matters as much as statistical distribution.

2. **A successful Polycentricity should be achieved based on compatibility of urban form and urban spatial structure.** Urban form refers to physical clusters of urban infrastructures, and spatial structure refers to functional clusters represented by urban activity and urban mobility. In other words, Polycentricity depends on both socioeconomic and physical urban space.

3. **Polycentricity is not only defined by the quantity of clusters, but also the balanced distribution of clusters, which is a matter of connections between urban flows.** Therefore, the structure of urban flows is as important as the structure of urban stocks.

4. **Urban flows have more diverse content than ever before.** Single journey types, such as “journey to work”, cannot represent overall urban mobility. More types of journeys should be taken into account considering the circumstances of today’s lifestyles.

5. **Urban processes are driven by multiple forces from both top-down planning and self-organized changes.** The original planned cities are already reshaped by individual needs in reality. Urban space is re-partitioned, redefined, and reorganized. Therefore, it is more reasonable to use emerging centers, instead of pre-defined administrative centers, in measuring Polycentricity.
Measuring Polycentric Urban Transformation

Polycentricity is a matter of both physical urban space and socioeconomic space. Previous research made much progress on measuring physical urban space, while this research focuses more on measuring socioeconomic space from human behavior using newly available urban mobility data. Both urban stocks (activities) and flows (movement) are measured by a two-step approach: (1) identify centers/neighborhoods using defined urban indices (2) identify certain spatial structures from the spatial distribution of indices’ values. Sets of indices used in these two steps are summarized in Table 6.1.

It should be noted that only directly related indices are summarized in the table. An extensive summary could include more indices such as those used to compare individual travel behavior in Chapter 5.3; the spatial interaction index used in many gravity model-based analyses; Ripley’s K index, or joint accounts that can be used to replace some of the global spatial statistical indices used in this research.

Table 6.1: A summary of indices used for measuring Polycentricity

<table>
<thead>
<tr>
<th>Index</th>
<th>Description in urban context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>Measured as the proportion of people accumulated in one unit area</td>
</tr>
<tr>
<td>Diversity/Entropy</td>
<td>Equal to entropy. Measures how mixed the activity types in one unit area</td>
</tr>
<tr>
<td>Evenness (extensive)</td>
<td>A modified entropy index. Entropy/number of existing types of stocks.</td>
</tr>
<tr>
<td>Centrality of centers</td>
<td>Area with both high density and high diversity</td>
</tr>
<tr>
<td>Degree</td>
<td>How many areas are directly connected to an area from any other area</td>
</tr>
<tr>
<td>Strength</td>
<td>Intensity of connection to and from one area</td>
</tr>
<tr>
<td>shortest path</td>
<td>How fast is the transfer between two areas</td>
</tr>
<tr>
<td>Clustering centrality</td>
<td>How ‘close’/‘cohesive’ the areas are to one another in terms of their accessibility to shared neighbors</td>
</tr>
<tr>
<td>Closeness centrality</td>
<td>How fast a kind of stocks could spread in the whole area</td>
</tr>
<tr>
<td>Betweenness centrality</td>
<td>How well-connected an area is and is key to identifying city hubs</td>
</tr>
<tr>
<td>PageRank</td>
<td>the role of a node or local area in attracting flows from all nodes in the network.</td>
</tr>
<tr>
<td>Community detection</td>
<td>identify neighborhoods and their borders</td>
</tr>
<tr>
<td>Variance of Centrality</td>
<td>How balanced is the statistical distribution of urban centrality</td>
</tr>
<tr>
<td>Size of Convex geometry</td>
<td>The minimum size of convex covering all spatial objects</td>
</tr>
<tr>
<td>Variance of size</td>
<td>How balance is the geographical size of centers/neighborhoods</td>
</tr>
<tr>
<td>Morans I</td>
<td>Spatial autocorrelation measures how well clusters of individual centers</td>
</tr>
<tr>
<td>Global mean center</td>
<td>Identifies the global center using centrality as a weight</td>
</tr>
</tbody>
</table>
6.1.3 Integrated Spatial Analysis and Modeling Approach

All of the analysis adopted indices from other domains such as complex network and signal processing, applied them to transportation data, and finally explained them in the context of urban studies. This kind of interdisciplinary approach is called integration in this research. In this section, we provide a deeper interpretation to the methodology used in this research.

Definition of Spatial Analysis and Modeling Approaches

The definition of analysis and modeling is given in Chapter 3 in a more general sense. This discussion elaborates the meaning of “analysis” and “modeling” in the context of presented methods.

Spatial analysis is explained in [50] as a general term for a kind of technique that utilizes location information to better understand the processes of generating the observed attributes’ values. Nowadays, spatial analysis covers wider topics. Besides conventional research in geography like statistics, aggregation, and spatial interpolation, there are also many inputs from other domains, especially computer science, like data mining, information visualization, all of which this research benefits significantly from.

Modeling has many various meanings in different contexts. It mostly equates to data modeling in this research. The result of data modeling is a conceptual model, which represents objects and their relations in a system with formal data structure. The data structures in the conceptual model can then be implemented using programming language and computed with methods of analysis.

Data analysis and data modeling are interdependent. As shown in Figure 6.2, it is a simplified work flow, extracted from an analysis of activity patterns and movement patterns. As one can see, the centrality value is not measured directly from the original data sets, but from a conceptual model, which is built to make the data meaningful in the context of certain urban phenomena. In particular, in Case 1, a central place theory model is adopted, which is very classical in urban geography. The theoretical model is reformatted to describe urban activities, and loaded with travel survey data. In Case 2, a network model is built, giving a representation of urban flows. Both the central place theory model and the network model are existing concepts, but primarily theoretical ones. They cannot get closer to reality unless they are put into certain contexts and calibrated by real data. The case study in this research implements such
models in a practical mode to deal with an issue of Polycentricity using transportation data. These models can be further implemented into interactive tools, which allow users to get inputs and give real-time analyzed feedback such as an implemented prototype of flow maps.

In sum, the research derived the definition of analysis and modeling from (Batty, 2009): Spatial analysis and modeling is “the process of identifying appropriate theory, translating this into a mathematical or formal model, developing relevant computer programs and then confronting the model with data so that it might be calibrated, validated and verified prior to its use in prediction”.

The Mechanism of Integration

Cities are complex systems that contain interdependent urban elements intricately interacting with each other. To understand the city as a system, interdisciplinary research is obliged to link all parts of an analysis together to result in a more comprehensive conception of an urban system. Integration of knowledge and techniques become crucial for doing this.

The research in the thesis follows the trends of integration. From the perspective of subjects, two main urban elements have been addressed, which are transportation and urban form in terms of land use and urban infrastructure. The method combines conventional geospatial analysis from GIS, data mining from computer science, and qualitative analysis in urban design and planning. Accordingly, there are two kinds of integration in this research namely, knowledge integration, which fills the gaps and exchanges information between different domains; and technique integration, which takes advantage of different methods for a better one. Simple diagrams are drawn to show the underlying mechanism embedded in the case studies. This subsection attempts to extract the general mechanism from studies conducted in this dissertation.
(1) Integrating knowledge

The purpose of integrating knowledge is to make sense out of random variables using contextual information. In data modeling, an appropriate theory is chosen and translated into a formal model. This step is actually a process of knowledge integration. To represent this process in a more formal and systematic way, the study abstracts two ways of knowledge integrations. (1) Model-based integration, where one object has a corresponding identity with an attributed model space and geographical space. These two spaces serve as two facets, which provide different angles to look into the subjects. The result is a more comprehensive understanding because hidden information is mined from more aspects. The conducted spatial network analysis is a good example. (2) Work-flow based integration, which is shown in Figure 6.3. Although the emphasis of this research is to trace the hidden functional urban changes using transportation data, thematic data was also studied to trace physical changes. These two aspects of change are then integrated in a descriptive analysis of the driven force and impacts in urban changes.

![Figure 6.3: Work-flow based integration.](image)

Note: Results of data mining are interpreted in joint with conventional urban study, in reverse, used as complementary materials of a more comprehensive urban study.

(2) Integrating spatial techniques

This study facilitates the use of spatial techniques to support urban design and planning. The main approach uses integrated geospatial techniques. A generic workflow can be extracted as shown in Figure 6.4. A GIS-based pipeline provides almost the same function in both cases. In the first step, data processing is used to clean up data sets and reformat the multiple data
sources into a unified structure. In the second step, analytical methods are applied. Examples of these include the Bayesian inferring method, spatial convolution operation, and spatial network analysis, which are used in this research. In the third step, geographical analysis is carried out to conduct basic spatial analyses like spatial joint, spatial interpolation, and geo-visualization. In general, this research shows a general method of building an integrated infrastructure which brings techniques together.

![Figure 6.4: A generic work-flow for integrating method into geospatial analysis.](image)

**An Emerging Field Bridging Urban Planning and Transportation Planning**

A long trend in urban studies is to understand the interactions between different urban elements, as shown in Figure 6.5 (top). Transportation data is used as input, but the difference from conventional research on routing, transport system planning, bus planning, the output of the analysis in this thesis lies in changes in the landscape of human activity and mobility, which are used to evaluate the original land use plans and as evidence of the interactions between transport planning and urban planning.

![Figure 6.5: Information flow (top) versus conventional planning flow (bottom).](image)

Most related research on land use and transportation interactions follows the logic shown in Figure 6.5 (bottom). The goal is to develop land use plans to restrict or predict the expected
impact on transportation, inspired by the statement that “Space shapes transportation as much as transportation shapes space” [120]. The study in this thesis investigates the interactions in the opposite direction. Urban functions and spatial structure in reality are extracted from transportation data. With this reversed direction, a loop between transportation and land use is goal.

This research is considered an emerging field of study that bridges transportation and land use research. The original land use which was defined by planning in a top-down urban process is reshaped by the practical needs of urban activities in reality, through a bottom-up urban process. The reality is the result of both forces. Investigation should be conducted to know the real situation and compare this to the original plans.

Applicability of Proposed Methods

The frameworks as well as method proposed in this thesis are generic. Whether or not the urban transformation of cities can be detected and measured by the proposed approach is mainly constrained by the availability of suitable data, not the methods. From perspective of technique, reasons are given in two points: simple input data format and repeatable algorithms. For instance, the network model using smart card data sets uses very limited information, counts of travels and location information, which can be retrieved from many resources, including direct resources such as GPS-traced cars and mobile phones. The algorithm presented is written in a generic format and can be easily adapted to other cases. Though this thesis focuses on one urban phenomenon, which is the polycentric urban transformation, the proposed integrated methods and approach can be further extended to other urban dynamic phenomena.

6.1.4 The Use of Big Location Data for Urban Studies

Data analysis is always important in urban studies, therefore it is not surprising that the new concept of “big data” gained so much popularity. However, “big data” is as fuzzy as Polycentricity, as it depends greatly on the related scale and context. Therefore, it is necessary to give a scope or context to the conducted research before a deeper discussion.

*Big data*, mostly referring to *location data* in this study, is defined as (1)Data so massive that it has to be managed by data management tools; (2) does not contain any straightforward social information of studied subjects; (3) raw data sets that are not presented in an intuitive
and very comprehensible way. With this definition, many data sets will be excluded, such as literature studies that give direct information or questionnaires that mostly have limited data size. The research is targeting the data sets that are hardly or seldom used by designers and planners or non data experts, and the extensive reuse of long-used data sets.

**Data Innovation Used in the Case Study**

Two kinds of transportation are used in these case studies. The first is travel survey data, which does not match the previously defined criteria (1). However, the conducted analysis gives an illustration of how this study deals with criteria (2) and (3). Travel survey data is used for detecting urban changes instead of its original application of estimating travel demand. The second case used smart card data, which is collected by an automatic fare collection system with very limited information about individual trips. It fits very well with the given definition of “big” data, as it is massive, provides location data without contextual information or intuitive view. Beyond individual trip information, information about individual and collective activity and movement patterns are extracted in the conducted analysis. These examples demonstrate a method of deriving extra value from data for urban studies and to support urban design and planning. A summary is given in Table 6.2 linking practical data applications (in Chapter 5) to the presented data innovations (in Chapter 2).

<table>
<thead>
<tr>
<th>Applications</th>
<th>Recombination of data</th>
<th>Extensible data</th>
<th>Data reuse</th>
<th>Open data</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining travel behaviors from smart card data</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>5.3.2</td>
</tr>
<tr>
<td>Fusing surveyed data and smart card to infer activity type</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>5.3.3</td>
</tr>
<tr>
<td>Identify functional centers from travel survey data</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>5.4</td>
</tr>
<tr>
<td>Extracting spatial structure from smart card</td>
<td>-</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Table 6.2: Data innovation applications in this research.
Alternative Information Resources for Urban Analysis

Both design or planning are rarely made from scratch; most of the time, it is based on investigations of current and past situations. Extracting information from abundant urban data may be an alternative way of collecting direct data via surveys.

Taking the urban studies related to this dissertation as examples, assessing the functions of urban space is of significant importance for understanding urban problems and evaluating planning strategies. However, conventional ways of data acquisition for urban functions in urban analysis are manual work, which consumes huge amounts of manpower and time to do field work to get direct information. Besides, the reliability of information is heavily influenced by subjective factors (such as time, place, investigators’ personal experience), since it is a qualitative estimation. Using automatically collected sensor data may not give direct information, but with the analysis and modeling method, required information can be extracted or inferred from it. As shown in this dissertation, the use of urban space is inferred from how people travel in a city. Nowadays, there are all kinds of sensor locations in the real world and virtual world, like social media, both generating data at a dramatic speed. These sensors and large data sets make it possible for us to observe and examine urban phenomena on a very high spatiotemporal scale that was almost impossible before. Applying a proper method of analysis to nearly all of the data sets with spatiotemporal labels can result in the extraction of rich information about the dynamic spaces.

6.2 Conclusion: Critiques and Outlook

In the context of ever faster urban transformation, urban designers and planners around the world are subject to very high expectations. Being able to manage and control the urban changes is a prerequisite for the development and validation of adequate planning strategies. This research studied the issue of polycentric urban transformation, which is considered a new type of urban form in many related urban studies. By reviewing the related work and analyzing the main debate about measuring Polycentricity, a refined definition has been presented, with a focus on measuring emerging functional Polycentricity from urban stock and flows. Correspondingly, a set of measures are presented based on spatial analysis methods making use of new available big urban mobility data. A case study of Singapore is conducted, implementing the presented methodology into practical application.
As a proof of concept, the implemented analysis methods presented in this work is successful. The results based on the case study of Singapore showed that urban transformation can be identified and measured quantitatively using an advanced spatial analysis and modeling approach using big transportation data. This study proposed a method of looking into urban functional changes and shows that increasing human movement data is a good resource for evaluating urban functionality and the impact of urban plans. Moreover, in terms of achieving better life qualities, to investigating urban activities and mobility is an emerging field that is related social science, human geography, transportation planning and urban planning. In fact, cities as complex systems raise issues which always demand expertise spanning across disciplinary boundaries, involving social, economic, and environmental studies, among others. Making use of new resources and developing advanced methods based on previous achievements open a path for such complex issues and constitutes an interesting agenda for further research about cities. In this respect, this research developed a holistic framework of integrated geospatial techniques applying to large urban mobility data for urban studies and planning.

Though this dissertation states that insights from spatial data analysis and modeling of urban large transportation data improve the understanding of urban transformation, there is much more potential that could be explored along this line of thought.

1. Rather than trying to exert increasing control over urban forms using the top-down approach of planning cities, it is more important for the planners to understand the mechanics that underlie urban dynamics, particularly the bottom-up changes which are driven by the actual needs of inhabitants. This work provides a view of human activities and mobility patterns as well as the final results of shaping landscape of urban functions; however, little work has been done to theoretically or quantitatively link the phenomena with driven forces. Urban planning decisions by the government constitute only one factor. More indicators should be found from urban economies, and societies that need more data and deeper investigation.

2. This study placed more emphasis on developing methods to detect functional urban changes from urban activities and movement. Analysis is given via linking physical changes with functional changes, but on very sparse spatial and temporal scales. To understand the cause and impact of asynchronous changes between physical space, built environment and socioeconomic space is the final objective that will lead future studies.
To achieve this, data in higher spatiotemporal resolution about both functional and morphological aspects are needed. Undoubtedly, in the age of big data, a study in this direct has unlimited potential.

3. As indicated, solving urban issues primarily requires interdisciplinary expertise, therefore another important point is that the different professionals engaged in the process of urban development should share a common language in talking about the built environment. The effectiveness of visual language has been proven useful in many cases, and was also implemented in the visual analytics tools in this study. It is necessary to expand the visual analytics tools towards a collaborative design environment, which provides geospatial infrastructure for real time analysis and visualization. For urban designers, such an environment is essential to be able to demonstrate and communicate the value of spatial planning to representatives of different professions.

Beyond the scope of this dissertation, the mechanism of the integrated analysis method can be developed to solve other urban issues other than managing urban transformation. In sum, there is still much to do by focusing on integrated techniques using multiple data sources for studying urban processes. This will contribute to a better understanding of urban dynamics, in terms of human behavior, movements, and urban processes, and the author believes the approach and results presented in this thesis show the direction in which the future work should go.
References


[40] Matthieu Cristelli, Michael Batty, and Luciano Pietronero. There is more than a power law in Zipf. Scientific Reports, 2, 2012.


[68] Homer Hoyt. The structure and growth of residential neighborhoods in american cities. 1939.


Information Science (ICCIS), 2012 International Conference on, volume 1, pages 143–147. IEEE.


Appendix A

Glossary

- **Urban Form** “refers to the spatial imprint of an urban transport system as well as the adjacent physical infrastructures. Jointly, they confer a level of spatial arrangement to cities” [120].

- **Urban Spatial Structure** “refers to the set of relationships arising out of the urban form and its underlying interactions of people, freight, and information” [120].

- **Spatial Interaction** is a realized transfer of people, freight, or information between areas. It is express a demand / supply relationship over a geographical space. Examples can be given as journeys to work in small scale, migrations in big scale, and the transmission of information or capital.

- **Urban Dynamics** representations of changes in urban spatial structures over time that embody a myriad of processes at work in cities on different, but often interlocking, time scales. These range from life cycle effects in buildings and populations to movements over space and time as reflected in spatial interactions [19].

- **Urban Model** representations of functions and processes in urban space. These are usually embodied in computer programs that enable location theories to be tested against data and predictions of future location patterns to be generated.

- **Urban Modeling** is redefined based on [19] as: a spatial analysis and modeling approach used to define a proper formal model, which can be used to represent urban space, and is
calibrated by large temporal location data. The properties of the model computed using large data sets can be used to explain urban processes.

- **Spatial Analysis** is explained in [50] as a general term of a kind of technique that utilizes location information to better understand the processes of generating the observed attributes values.

- **Integrated Spatial Analysis** covers wider topics. Besides conventional research in geography like statistics, aggregation, and spatial interpolation, there are inputs from other domains, especially, from computer science, like data mining, information visualization.

- **Spatiotemporal Analysis** incorporates time into geographical information systems. It raises awareness of the importance of time within the GIS community and the development of models that can be used to represent dynamics.

- **Visual Analytics** is a techniques aiming at multiplying the analytics power of both human and computer by finding effective ways to integrate interactive visual techniques with algorithms for computational data analysis. Therefore, visualization and computation can interplay and complement each other [8].
Appendix B

Data Inventory

Table B.1: Transportation data sets used in this research.

<table>
<thead>
<tr>
<th>Data sets</th>
<th>Year</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household Travel Survey</td>
<td>1997</td>
<td>SENSEable City Lab</td>
</tr>
<tr>
<td>Household Travel Survey</td>
<td>2004</td>
<td>Future Cities Lab</td>
</tr>
<tr>
<td>Household Travel Survey</td>
<td>2008</td>
<td>Future Cities Lab</td>
</tr>
<tr>
<td>Smart-card Data</td>
<td>September 2010</td>
<td>Future Cities Lab</td>
</tr>
<tr>
<td>Smart-card Data</td>
<td>April 2011</td>
<td>Future Cities Lab</td>
</tr>
<tr>
<td>Smart-card Data</td>
<td>September 2012</td>
<td>Future Cities Lab</td>
</tr>
</tbody>
</table>