THE AGGLOMERATION ECONOMIES IN THE MEGAREGIONAL CONTEXT: A SPATIAL AND FUNCTIONAL PERSPECTIVE USING GEOSPATIAL ANALYTICS

ZIDONG YU

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University
Department of Land Surveying and Geo-Informatics

The Agglomeration Economies in the Megaregional Context: a Spatial and Functional Perspective Using Geospatial Analytics

Zidong Yu

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

May 2024

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

(Sign)

Zidong Yu

Abstract

The population dynamics in cities have varied significantly in recent years. A large number of cities have experienced a significant increase in population due to urban migration and natural growth. Robust industrial agglomerations exhibit accelerated growth and increased benefits for economic activities and firms. The concept of urban agglomeration economies describing the clustering phenomenon within and cross cities related to economic prosperity, has gained attention in geography studies. As urbanization expands to densely populated megacity regions with abundant services and resources, there is a growing need for comprehensive research emphasizing agglomeration economies in megaregional contexts. The current reliance on economic census and surveys in literature has posed different challenges to capturing and evaluating economic activities, such as costly and undercounting. Therefore, a novel strategy using geospatial analytics should be introduced to bridge the gap between these concepts and depict agglomeration economies in megacity regions.

This thesis presents an innovative data-driven strategy for studying industrial agglomeration economies within megaregions, aiming to achieve four step-by-step objectives: (1) examining urban agglomeration and its correlation with urban environments; (2) exploring the spatial and functional organization of agglomerations within megaregion; (3) analyzing the geographic disparities in agglomeration economies; and (4) delineating the spatial-functional network of regional industrial agglomerations. The first objective involves a thorough reassessment of prevalent official census data, establishing connections between urban environments and socio-demographics within cities through initial geospatial analytical methodologies. Following this, the second objective shifts the research focus towards leveraging advanced geospatial data sources and analytics to investigate the spatial and functional patterns, encompassing organization, disparity, and network, manifested by agglomeration economies within a megaregional context.

The methodology integrates advanced geospatial analytics, including geospatial data science, machine learning, and network analysis, to quantify spatial and functional patterns of agglomeration economies. In the second case study, spatial extent is measured using kernel density functions on points of interest (POIs) data, while natural language processing (NLP) is employed for semantic-based information retrieval to label functional characteristics. In the third case study, thematic topics related to the local industrial sector are identified through topic modeling, and industrial agglomerations are clustered based on topic importance, illustrating spatial and functional variations. The final case study employs network analysis to describe the megaregional agglomeration network, utilizing bipartite network projection and community detection to reveal groups of closely connected agglomerations by their industrial functions.

This thesis furnishes crucial empirical evidence, presenting alternative viewpoints on the geographic and functional intricacies of industrial agglomeration economies within megaregions. Through multiple case studies framed within a geospatial lens, the research provides robust support, showcasing the potent capabilities of utilizing geospatial data sources and analytics. The findings contribute significantly to the understanding of economic geography, regional studies, and urban studies, offering valuable insights and addressing fundamental questions in these domains.

Keywords: agglomeration economies, geospatial analytics, megacity region, spatial and functional patterns, industrial disparity, regional network.

List of Papers

The first-authored research articles from papers 1 to 4 significantly inform the research questions and extended discussions presented in Chapters 4 to 7 of this thesis. These chapters reflect my individual contributions to research design, method development, and initial draft writing. Papers 5 to 8 contribute to the methodology development and broaden the research scope outlined in this thesis. It is important to acknowledge that all co-authors have provided solid support and valuable comments.

Paper 1 **Yu, Z.**, & Liu, X. (2021). Urban agglomeration economies and their relationships to built environment and socio-demographic characteristics in Hong Kong. *Habitat International*, *117*, 102417.

Paper 2 **Yu, Z.**, Zu, J., Xu, Y., Chen, Y., & Liu, X. (2022). Spatial and functional organizations of industrial agglomerations in China's Greater Bay Area. *Environment and Planning B: Urban Analytics and City Science*, *49*(7), 1995-2010.

Paper 3 **Yu, Z.**, Xiao, Z., Yan, Y., Feng, C. C., & Liu, X. (2023). The geographic disparity of agglomeration economies: Evidence from industrial activities in China's emerging Greater Bay Area. *Applied Geography*, *161*, 103128.

Paper 4 **Yu, Z.**, Xiao, Z., & Liu, X. (2023). Characterizing the spatial-functional network of regional industrial agglomerations: A data-driven case study in China's Greater Bay Area. *Applied Geography*, *152*, 102901.

Paper 5 **Yu, Z.**, Zhu, X., & Liu, X. (2022). Characterizing metro stations via urban function: Thematic evidence from transit-oriented development (TOD) in Hong Kong. *Journal of Transport Geography*, *99*, 103299.

Paper 6 **Yu, Z.**, Xiao, Z., & Liu, X. (2022). A data-driven perspective for sensing urban functional images: Place-based evidence in Hong Kong. *Habitat International*, *130*, 102707.

Paper 7 **Yu, Z.**, & Liu, X. (2023). Spatial variations of the third and fourth COVID-19 waves in Hong Kong: A comparative study using built environment and socio-demographic characteristics. *Environment and Planning B: Urban Analytics and City Science*, *50*(5), 1144-1160.

Paper 8 **Yu, Z.**, Zhu, X., Liu, X., Wei, T., Yuan, H. Y., Xu, Y., ... & Chen, W. (2021). Reopening international borders without quarantine: contact tracing integrated policy against COVID-19. *International Journal of Environmental Research and Public Health*, *18*(14), 7494.

Acknowledgement

The three-year odyssey of my PhD study has swiftly passed. This thesis, a significant milestone in this academic journey, would not have been realized without the unwavering support from my supervisor, co-authors, colleagues, and family. I extend my warmest thanks to all who have been instrumental in this endeavor.

Foremost, I would like to express heartfelt appreciation to my supervisor, Dr. Xintao Liu, an inspiring academic mentor whose unwavering support and patience guided me through every stage of my PhD journey. His wealth of knowledge and extensive experience has been a constant source of encouragement for both my academic research and future career. I also acknowledge Dr. Xiaolin Zhu for endorsing my PhD opportunity and providing valuable guidance early in my study, along with Dr. Yang Xu, my MSc supervisor and the one who introduced me to academia.

Moreover, this thesis benefited immensely from the collaboration and support of my co-authors and colleagues. Prof. Wu Chen, Prof. Song Guo, Prof. Angela Yao, Prof. Dr Lilian PUN, Prof. Yimin Chen, Prof. Chen-Chieh Feng, and Dr. Yingwei Yan have provided invaluable comments, inspiration, and lessons from their outstanding works. Special gratitude goes to Prof. Chen-Chieh Feng and Dr. Yingwei Yan, my host supervisors during my visiting program to the National University of Singapore. I am also thankful to the members of GeoX and colleagues in LSGI, including Rui Zhao, Junwei Zhang, Jialyu He, Ahmad M, Jianwei Huang, Zijian Guo, Minglei Liao, Ningyezi Peng, Zhiyang Xiao, Jian Liu, Ketong Shen, Ying Na, Yunjuan Sun, Xinyue Gu, Shuai Xu, Jiaqi Tian, Hengtian Zhang, Zhenhan Peng, Ruoqing Peng, and Shuheng Zhao. Special thanks to Polly Hui, Louis Lau, and Deanna Xu for inviting me to participate in the Proof-Of-Concept (POC) Funding scheme and Ideation program, allowing me to apply innovative research ideas to real-world issues in Hong Kong.

Lastly, my heartfelt gratitude goes to my family, including my dear parents and my fuzzy companion, Lucky, for their warmest support and encouragement throughout my pursuit of a doctoral degree. A special thanks to Swing Huang, my beloved and soulmate, for standing behind me through every critical stage, providing vital support, sharing ideas, and discussing everything with me.

Zidong Yu
Hong Kong, January 2024

Contents

List of Abbreviations

AIC	Akaike Information Criterion
APS	Advanced Producer Service
BMR	Barcelona Metropolitan Region
CBD	Central Business District
DCCA	District Council Constituency Areas
ESS	Error Sum of Squares
FUA	Functional Urban Areas
GaWC	Globalization and World Cities
GBA	Guangdong-Hong Kong-Macao Greater Bay Area
GWR	Geographically Weighted Regression
HHI	Hirschman–Herfindahl Index
ICT	Information And Communications Technology
KDE	Kernel Density Estimation
LDA	Latent Dirichlet Allocation
LISA	Local Indicator of Spatial Association
LQ	Location Quotient
MTR	Mass Transit Railway
NLP	Natural Language Processing
OLS	Ordinary Least Squares
OSM	Openstreetmap
POIs	Points of Interest
PRD	Pearl River Delta
PTN	Public Transit Networks
R&D	Research and Development
TF-IDF	Term Frequency and Inverse Document Frequency
BTHUG	Beijing–Tianjin–Hebei Urban Agglomeration

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Background and Motivation

Over the past few decades, there has been a notable surge in urbanization, witnessing a swift rise in the population settling within most cities and urban regions. Forecasts indicate a sustained trajectory of growth, signaling a significant expansion as the urban population is projected to encompass nearly 70% of the global populace by 2050 (Carley & Spapens, 2017). These urban migrants and newborns have been recognized as the major motivation of the urbanization process and economic development in cities. Cities encompass a diverse range of human activities and functions within a concentrated setting, where economic activities predominantly conglomerate. The notion of urban agglomeration economies is thereby introduced to describe the particular clustering phenomenon seen in cities and explain the close relationship with prosperous economic development (Fang & Yu, 2017). Evidence related to agglomeration economies has supported that firms spatially located in formidable industrial agglomerations often grow faster and benefit more than average. Amidst the panorama of global urbanization, industrial concentrations have expanded across municipal spatial scales, culminating in the creation of megaregions—densely populated areas offering dense services and resources (Camagni et al., 2017). Consequently, there is an increasing imperative to comprehensively investigate agglomeration economies within these megaregional contexts, recognized as pivotal in enhancing local economic vitality. With the emerging geospatial analytics and data, this thesis centrally aims at presenting a systematic strategy to investigate industrial agglomeration economies within a megaregional context. Via the spatial and functional regularities exhibited by agglomeration economies, three specific research topics would be explored and can be associated with organization, disparity, and network.

The stream of agglomeration measurement as a primary research area in understanding economic activities has been characterized as a dual categorization: spatial and functional dimensions, as delineated in relevant literature (O'Donoghue & Gleave, 2004). The spatial dimension examines geographical arrangements and the level of concentration in economic activities, whereas the functional dimension posits that industrial agglomerations comprise interconnected economic activities sharing commonalities and complementarities (Vogiatzoglou & Tsekeris, 2013). The primary purpose of delineating the spatial boundaries of agglomerations lies in quantifying the concentration level of industrial activities. Prior research endeavors have aimed to devise metrics such as Gini coefficient and location quotient (LQ) that can capture this concentration by employing discrete-space methodologies (Gezici et al., 2017; Krugman, 1991). Focusing on the latter, Strange (2008) demonstrated the mutual sharing of resources in specialized manufacturing industries and knowledge exchange among firms spanning diverse sectors. To delineate functional attributes, two fundamental paradigms of agglomeration economies are proposed: diversification and specialization economies. Diversification, as proposed by Jacobs (1969), underscores the correlation between industrial agglomerations and the diversity of industries within larger metropolitan areas. Conversely, specialization is closely linked to the concentration of specific industries or productions (Marshall, 1890). Empirical investigations predominantly highlight the contrasting functional impacts of specialization and diversification, drawing upon regional statistical data encompassing industrial value, patterns, and employment (Guillain & Le Gallo, 2010).

Industrial diversification within agglomeration economies refers to the phenomenon where industrial entities in the same geographic area exhibit a range of functions and engage in various sectors, thus contributing to a diversified economic landscape (Neffke et al., 2018). Jacob (1969) posited that such diversification typically manifests in the form of agglomeration externalities, where a wide variety of industrial sectors within metropolitan areas are closely interlinked. The need for deeper insights into industrial diversification is increasing, particularly in understanding the distinct functional and spatial roles these agglomerations play within cities and urbanized areas. To analyze functional diversity, researchers often utilize metrics based on the presence or proportion of various industrial sectors, usually defined through a digit-based classification of professional sectors (Hausmann & Hidalgo, 2010; O'Donoghue & Gleave, 2004). This classification system may encompass different types of industries or business activities recorded within administrative boundaries such as counties, cities, or provinces. Furthermore, previous studies have employed diversity indices that relate to the distribution across various industrial categories and markets, such as employment and productivity measures, which are typically calculated using probability density functions (Bettencourt et al., 2014). An example of these indices is the Hirschman–Herfindahl index (HHI), widely used to gauge market diversity (Hirschman, 1980). These indices, however, are dependent on specific classification systems or pre-defined taxonomies, which may limit their ability to fully capture the nuances of local industry dynamics, due to the constraints imposed by the taxonomic frameworks established by researchers.

Beyond functional heterogeneity, spatial heterogeneity represents a key dimension examined by researchers to explore the characteristics of industrial diversification, a prevalent topic in regional and urban studies. Various methodologies have been developed to define the spatial boundaries of economic agglomerations. A predominant approach in the literature involves analyzing the concentration and local spatial distribution of economic activities (Parr, 2002; Rosenthal & Strange, 2001). Through the examination of spatial patterns across different industrial sectors, researchers can identify and interpret the spatial variations in economic activities. Due to the availability of data, much of the research relies on economic data aggregated at a broad geographical level, leading to a reliance on a discrete-space approach to analyze spatial variations in agglomerations (Guillain & Le Gallo, 2010). However, some studies have attempted to overcome this limitation by utilizing alternative datasets that provide individual-level information and allow for an analysis based on a continuous geographical space (Duranton & Overman, 2005). These datasets typically include granular details, such as the locations of manufacturing facilities or service providers.

A pivotal discussion in regional studies, particularly within the framework of modern globalization, focuses on the mechanisms that initiate, shape, and enhance economic agglomerations. These urban economies create agglomeration externalities from which firms benefit by co-locating with similar entities within cities and industrial clusters. Early research characterized these externalities as economic activities confined geographically to a specific area, with limited connections to activities outside that area (Burger & Meijers, 2016; Rosenthal & Strange, 2004). However, recent scholarly consensus acknowledges that in the era of globalization and regionalization, agglomeration economies are not isolated entities. Considerable scholarly effort has been directed towards understanding how regional contexts both contribute to and benefit from the dynamics of agglomeration economies (McCann & Acs, 2011).

Network science offers a comprehensive perspective for investigating economic relationships, as the focus of research has increasingly moved from the internal dynamics of cities to their external interactions. The body of literature on network externalities and regional science typically divides into two categories based on the characteristics of network flows. The first category involves tangible physical networks, which primarily concern infrastructure connections such as railways, highways, and telecommunications systems (Keeling, 1995; Rimmer, 1998). These networks facilitate the movement of populations and goods, serving as the physical underpinnings of city or urban area interactions. Conversely, the second category concentrates on intangible networks that establish non-physical links related to socio-economic factors. For instance, economic relationships between cities might be measured through transaction agreements or established routines that indicate economic cooperation or competition (Church & Reid, 1996; Mikhaylov & Bolychev, 2015).

Both categories view cities and their regional relationships as complex network systems, analyzing various structural properties to understand these connections better. However, despite this depth of understanding in how cities interact, there remains a notable gap in detailed knowledge concerning local sectoral agglomeration economies, especially in studies that integrate intra-regional contexts with advanced spatial and network analytics. Enhancing our understanding of the network structures within intra-regional and even intra-city agglomerations could provide significant insights to both academia and practical applications.

The onset of globalization and urbanization has rendered traditional administrative boundaries obsolete, underscoring the growing necessity for alternative spatial analysis scales to delineate economic activities. In response, regional scholars have introduced the concept of megaregions (i.e., megacity regions), defining them as interconnected sets of cities and their surrounding suburban hinterlands. Within megaregions, labour and capital can be easily reallocated at minimal costs, marking a significant shift in economic dynamics. Industrial agglomerations have meanwhile expanded into larger spatial domains, giving rise to megaregions—densely populated areas rich in services and resources. Scholars argue that these megaregions, exemplified by hubs like the San Francisco Bay Area and the Tokaido (Tokyo–Osaka) corridor, function as primary drivers of both global and domestic economies (Angel, 1991; Hui et al., 2018). As these megaregions foster thriving industrial activities, they not only contribute to shaping a highly integrated multi-center structure but also attract population influx from other regions (Walker & Schafran, 2015). Substantial research has been carried out to comprehend the formation and flourishing development of megacity regions, primarily associated with the issues of urban planning and policies (Yeh et al., 2020).

After reviewing the literature on investigating agglomeration economies, it becomes apparent that the conventional approaches commonly used in existing studies might present substantial challenges. Particularly, these approaches might fall short of accurately representing the diverse characteristics of agglomeration within the context of megacity regions. The relevant research gap raised by megaregions can be summarised as follows:

1. Empirical studies have predominantly highlighted the contrasting functional impacts between specialization and diversification using regional statistical data encompassing industrial value, patterns, and employment (Guillain & Le Gallo, 2010). These data

sources operate at an aggregated geographical level, potentially limiting discrete-space methodologies in capturing nuanced functional differences at a finer spatial scale.

2. The rapid urbanization within megacity regions has ushered in highly intricate economic activities, surpassing the comprehensiveness of pre-established digit-based sector classifications. Prevalent standards advise that economic censuses or surveys are mostly conducted at regular intervals, such as every five years in countries like the U.S. and China. Consequently, regional economic data derived from conventional sources often becomes outdated and unreliable (Wardrop et al., 2018). Furthermore, relying solely on diversity indices might overlook crucial locational specifics tied to distinct market segments within local industries.
3. In the current era of globalization, cities, and their intraregional connections are viewed as intricate network systems, offering insights into diverse structural properties that can be studied and analyzed. In this subject, there remains a considerable gap in understanding local sectoral agglomeration economies at both specific and overarching levels, especially in studies aiming to harness intra-regional contexts and incorporate emerging spatial and network analytics.

1.2 Objective and Research Questions

This thesis endeavors to introduce an innovative data-driven approach for studying industrial agglomeration economies within megaregions. The research will integrate a suite of advanced geospatial analytics—such as spatial models and metrics, machine learning, and network analysis—to achieve four step-by-step sub-objectives: (1) examining urban agglomeration and its correlation with built environments; (2) exploring the spatial and functional organization of agglomerations within megaregion; (3) analyzing the geographic disparities in agglomeration economies; and (4) delineating the spatial-functional network of regional industrial agglomerations. To be specific, the first objective would be a critical revisit to current widely used official census data and linked to relationship from urban environments and socio-demographics within cities using preliminary geospatial analytical solutions. Since the second objective, the research focuses would be then shifted to the utilization of geospatial data sources and analytics to explore the spatial and functional patterns (i.e., organization, disparity, and network) exhibited by agglomeration economies in a megaregional context. Four principal research questions are answered by this thesis:

1) Using the conventional census data, what spatial configurations define urban agglomeration economies in cities, and how are these patterns correlated with both the built environment and socio-demographic attributes?
2) Moving to larger scale, what kind of geospatial big data and analytics can be employed for unveiling the spatial and functional organization of agglomeration economies within megaregional contexts?
3) Focusing on the geographic disparity of economic activities, how do spatial and functional heterogeneities within industrial agglomerations in megaregion contribute to geographic disparities, as revealed by geospatial analytical solutions?
4) Last but not least, how do interactions among local industrial agglomerations contribute to forming megaregional networks from a network perspective? What are the structural characteristics and geographic variations of these networks across spatial and functional scales?

1.3 The Subsequent Layout

This thesis consists of eight chapters, presenting a detailed information of my doctoral research, as supported by the first 4 appended publications. The brief introduction of following chapters is provided below:

Chapter 2: The notion of agglomeration and previous studies related to investigating this subject are reviewed in this chapter. Research empirical evidence as well as approaches related to the objectives proposed in this thesis are demonstrated to clarify the current achievement and potential research restrictions. At last, research gaps from literature review are sorted out.

Chapter 3: The methodologies mainly adopted the research framework are presented in this chapter and organized by four domains: urban geospatial data, geospatial model and metrics, machine learning, and network analysis. The concept, formula, applications, and interpretation are meanwhile elaborated.

Chapter 4: This chapter presents a case study in Hong Kong using conventional census data to reveal the spatial distribution patterns of urban agglomerations and their relationship to built environment and socio-demographic characteristics. Having combined official census and geospatial analytics, driven geographical evidence is supplemented by this study to current literature for understanding local agglomeration economies in a metropolitan context.

Chapter 5: This chapter explores the spatial and functional organization of agglomeration economies by utilizing points of interest (POIs) data gathered in the Guangdong-Hong Kong-

Macao Greater Bay Area (GBA). The empirical findings indicate a significant heterogeneity in the concentrations of industrial activities across various economic sectors within the GBA. This study offers novel insights into the spatial and functional organizations of economic activities within a megaregional context.

Chapter 6: This chapter provide a case study to investigate the spatial and functional disparity of industrial agglomerations within GBA, of which semantic information of local manufacturing activities is derived and associated with spatial layouts. Analytical findings offer crucial empirical evidence, fostering in-depth understanding of heterogeneous regularities of industrial activities The implications hold substantial value for theoretical literature and practical policymaking in megaregional economics.

Chapter 7: This chapter presents a data-driven intraregional case study in GBA to investigate the curiosities related to a regional network of sectoral agglomeration economies, with the use of geospatial big data and network analytical approaches. The results confirm substantial geographic heterogeneities of agglomeration networks related to locational advantages. Its associated implications can be of great help to regional economic collaboration and development.

Chapter 8: The conclusion of the entire thesis is shown in the final chapter, summarizing the contributions of each case study, real-world implications, potential limitations, and future research directions.

Chapter 2. Literature Review

2.1 Defining and Understanding Agglomeration Economies

The impacts of agglomeration economies are identified as a form of economic externality characterized by the co-location of economic agents and activities. Regional data commonly supports the premise that businesses situated within strong industrial clusters tend to experience more rapid growth and greater benefits compared to the broader average (Henderson et al., 1995; Porter, 2003).

2.1.1 Spatial and functional aspects of measuring agglomeration

Martin and Sunley (2003) identify two critical dimensions of agglomeration economies: spatial and functional. The spatial aspect focuses on the concentration of activities within a specific geographic area, highlighting how geographical proximity (or spatial co-location) facilitates the establishment and sustenance of physical connections among businesses in industrial clusters. The functional aspect, on the other hand, pertains to the internal operational dynamics within an industrial cluster, irrespective of physical space.

Spatial delimitation of agglomerations serves the primary purpose of quantifying the concentration levels of industrial activities. A large amount of previous studies has endeavored to devise metrics capturing this concentration using discrete-space methodologies given aggregated data sources (Gezici et al., 2017; Krugman, 1991). Krugman (1991) introduced the locational Gini coefficient, in order to evaluate the concentration of specialized employment or patents within local industrial markets. Malmberg (1996) utilized this metric to illustrate the geographic distribution of industries in the U.S at the state level, presenting empirical evidence to support concentrated localized industrial systems. Another measurement, the location quotient (LQ), developed and proposed by Glaeser et al. (1992), gained widespread use in assessing industrial activities and was frequently referenced in subsequent research in the community of economic geography. This metric gauges the ratio between local and national employment percentages within specific industrial sectors. Gezici et al. (2017) applied LQ-based evaluation approach using employment data in Turkey to comprehend the mechanisms of manufacturing clusters and identify areas specializing in high-tech sectors. Conversely, research focusing on distance-based strategies treats locality as continuous geographical space. For instance, Ripley's K function highlights clustering by plotting a distance function against the null hypothesis's reference function, while the Q statistic illustrates co-location associations among multiple entities (Liu et al., 2021b; Scholl & Brenner, 2016).

Two typical categories of agglomeration externalities have been introduced for delineating the functional dynamics within industrial agglomerations: Marshall specialization and Jacobs diversification (Marshall, 1890; Jacobs, 1969; Beaudry & Schiffauerova, 2009; Faggio et al., 2017). Previous studies have aimed to comprehend and outline the unique characteristics inherent in these two types of agglomeration economies. The Marshall model elucidates spatial concentration of specialized economic activities and identifies three external forces prompting firms to locate proximately: input sharing, labor market pooling, and knowledge spillovers (Marshall, 1890). Holmes (1999) provided direct evidence showcasing that more concentrated agglomerations may exhibit higher levels of purchased input intensity, a trend observed in the U.S. manufacturing sector. Considering the influence of human resources, Andini et al. (2013) explored the relationship between specialized agglomerations and labor markets in Italy. In contrast, Jacobs' externalities underscore that diversification of firm types in larger metropolises could lead to the concentration of diverse industries, primarily associated with innovative economics (Jacobs, 1969). Harrison et al. (1996) examined nearly 1000 manufacturing establishments across the U.S, emphasizing that a more diverse industrial agglomeration facilitates the exchange of skills. Additionally, Combes (2000) detailed the nuanced impact of Jacobs diversity across manufacturing and service sectors by investigating employment growth in 341 local areas in France. Overall, these studies treat geographic locations as aspatial units seemingly devoid of spatial correlation (Guillain & Le Gallo, 2010).

2.1.2 Heterogeneity of agglomeration

The concept of industrial variation is crucial in explaining the composition of economic agglomerations, which include a diversity of firms and associated functions. This concept is encapsulated in the widely recognized theory of Jacobs diversification, prevalent in contemporary research. Jacobs diversification posits that a varied local economic framework often results in escalating returns, subsequently fostering urbanization, a phenomenon typically evident in metropolitan settings.

In recent years, there has been an increasing demand for a deeper understanding of industrial diversification, with numerous scholars focusing their research on quantifying this phenomenon. Commonly, the assessment of industrial diversification employs empirical, category-based methodologies that hinge on digit-based industry classifications (Van der Panne, 2004). The Hirschman–Herfindahl index (HHI), a widely utilized metric, calculates the sum of the squared market shares of each industry, with a decrease in HHI indicating greater economic diversity within the specified area. De Lucio et al. (1996) utilized the HHI on employment data from an industrial survey to demonstrate the diversity (Jacobs) and significance of Spanish industrial activities at the provincial level. Variations of the HHI also exist to gauge the diversity in local economies, including those based on industrial patents or value-added data.

Shannon entropy, originally developed to quantify disorder or uncertainty in various scientific fields, is also applied to measure economic diversification. Attaran (1986) employed Shannon's entropy to analyze economic diversity across the US states, using employment data categorized into 56 sectors by the 2-digit Standard Industrial Classification (SIC). Industrial diversification within agglomerations can be further divided into related and unrelated varieties, where related variety encourages knowledge and technology spillovers, and unrelated variety supports the portfolio effect (Cainelli et al., 2019). Frenken et al. (2007) developed an entropy-based method to assess related and unrelated varieties in the Netherlands, analyzing data at the sectoral digit level to explore their impacts within industrial diversification agglomerations. In a detailed study, Nissan and Carter (2010) used Shannon's entropy with a 3-digit SIC classification to compare employment diversity across major industrial sectors and subsectors. These methodologies typically rely on predefined industrial classifications or taxonomies (Bettencourt et al., 2014). Recent advancements in research have introduced more sophisticated systems for measuring functional specialization, which consider specific tasks such as fabrication, management, and R&D within industrial sectors performed by distinct occupational classes (Timmer et al., 2019).

2.1.3 Network science and agglomeration

Network science, a field that leverages graph theory and network models, involves constructing, characterizing, and analyzing network systems along with their behaviors (Brandes et al., 2013). It utilizes nodes (vertices) and links (edges) to form networks that represent observable patterns of interactions among distinct elements. Various network models have been developed to mimic real-world phenomena, with graph theory providing the tools needed for their evaluation and interpretation (Newman, 2003). The fascination with network behaviors, particularly in urban and regional studies, gained significant momentum in the 1990s, propelled by the rapid pace of globalization and regionalization. Neal et al. (2021) highlights the utility of network concepts and methodologies in urban studies, where they are applied to analyze cities and communities across various scales and contexts.

The concept of an urban network is used to describe the interactions among economic actors, typically cities, that are geographically distinct yet capable of generating externalities. Capello (2000) defines urban network externalities as the inter-city complementarities and synergies viewed from a macro-scale intraregional perspective. The scholarly consensus distinguishes between tangible physical networks and intangible non-material networks. The physical networks often encompass infrastructure connections such as highways, railways, and airlines, primarily linked to transportation and mobility (Derudder, 2006; Liu, 2019). These include, for example, flights and passenger flows which are used to establish global and regional urban networks between operating airports, with cities hosting these airports serving as network nodes (Keeling, 1995).

Lee (2009) developed a network analysis model to examine the international air networkability of global cities and their interactions within the airline network. Zhang et al. (2021a) explored the spatiotemporal dynamics of a regional urban network over a period from 1997 to 2015 using air transport data from China. Furthermore, the adoption of urban big data has significantly advanced research into human mobility patterns within the urban network framework. For instance, smart card transaction data from public transit networks (PTN) are extensively used to analyze individual passenger flows (Roth et al., 2011). Beyond traditional networks like PTN, new mobility data sources, including taxi, shared bicycle, and mobile phone location data, are increasingly utilized to develop spatial interaction networks across administrative boundaries (Liu et al., 2021a; Zhang et al., 2021b).

In the realm of intangible non-material networks, researchers have treated cities as nodes within regional networks that encompass a variety of socio-economic relationships, such as regional political and economic cooperation (Mikhaylov & Bolychev, 2015). There has been a notable focus in urban studies on networks of cities facilitated by firm-level connections, particularly those involving advanced producer service (APS) firms, to explore urban networks at multiple scales (Sassen, 1991). The presence of producer service firms in a city is often interpreted as a significant indicator of that city's centrality within the urban network. The Globalization and World Cities (GaWC) framework has been instrumental in developing the interlocking network model to assess the connectivity of the global city network (Neal, 2011; Taylor, 2001). According to Taylor (2001), the interlocking network model suggests that cities offering services from the same firms are interconnected, facilitating the exchange of information and influence. Cities such as New York and London, which host numerous APS firms, are viewed as highly central within the network, acting as critical hubs that facilitate extensive connections through the flow of capital, knowledge, and information. Such firm-level networks have proven to be effective tools for illustrating both regional and global urban networks, revealing diverse spatial-functional organizational structures (Dicken, 2007). This approach allows for a nuanced understanding of how cities interact and function within broader economic and informational landscapes.

2.2 Urban Agglomeration in Megaregional Context

Recent decades of economic globalization have significantly altered the geographic and urban landscapes, leading to the rise of massive, urbanized areas that now serve as the foundational regions for economic development (Hui et al., 2020). In this context, urbanists have introduced the concept of the megacity to describe regions that encompass a network of closely situated urban settlements, often anchored by one or more large cities (Hall & Pain, 2006; Scott, 2019). This concept has spurred extensive research exploring the interplay between spatial regularities and industrial clusters within such megacity regions.

For example, Kerr and Kominers (2015) formulated a location choice model that detailed the spatial clusters of economic activities within the San Francisco Bay Area, assessing how agglomeration economies influence the configuration and scale of industrial clusters. Buzard et al. (2017) utilized a multi-scale core-cluster method for a spatial analysis that pinpointed the industrial clusters of over 1700 private Research and Development (R&D) labs across California and the U.S. Northeast Corridor. The topic of industrial diversification, particularly the spatial dimensions of industrial variation in metropolitan regions, has also been a focus of study. Funderburg and Boarnet (2008) investigated the geographic distribution of manufacturing clusters using principal components analysis, revealing diverse spatial concentrations across different sectoral industrial clusters in Southern California. In a study of twenty-six manufacturing and service sectors around Paris, Guillain and Le Gallo (2010) employed the locational Gini coefficient and Moran's I statistics to examine the spatial variations of sectoral agglomeration economies. Additionally, an intra-metropolitan perspective was applied by Garcia-López and Muñiz (2013) to analyze the relationship between urban spatial structure and sectoral agglomeration economies within the Barcelona metropolitan region (BMR). This body of research illustrates the complex and evolving dynamics of spatial and economic patterns within globally integrated urban environments.

2.3 Incorporating the Utilization of Geospatial Analytics

In recent times, spatial big data has emerged as a promising data source, demonstrating considerable potential in consistently and accurately providing large-scale spatial evidence (Elwood et al., 2012; Goodchild, 2007). Comparatively, emerging geospatial datasets present novel prospects for scale-dependent urban studies. Utilization of multisource data, including mobile phone locations, smart card records, POIs, and social media data, has extensively contributed to analyzing spatial structures and human activities within urban contexts (Goodchild, 2007; Pelletier et al., 2011). Point of Interest (POI) data, a subset of spatial big data, stands out due to its vast potential, offering a wealth of valuable information applicable to diverse urban research directions (Elwood et al., 2012; Gao et al., 2017). This dataset represents geographic entities encompassing both spatial and non-spatial attributes, encompassing details like names, addresses, categories, and coordinates. Goodchild (2007) advocates for POI's relevance in landscape research, suggesting it provides additional and complementary perspectives. Unlike traditional datasets utilized for agglomeration measurement, such as household surveys or regional economic statistics, POI data is obtained efficiently and serves as effective proxies for gauging agglomerations. It reflects spatial concentration levels and delineates functional characteristics.

The categorical information within POIs enables scholars to unveil urban functions using machine learning techniques (Yuan et al., 2012; Yu et al., 2022a). Some studies have leveraged POIs to explore urban socioeconomic activities, with Li et al. (2021) using business POI density to illustrate spatial patterns of urban economic activities in Wuhan. Meanwhile, Liu et al. (2020b) employed leisure-related POIs to capture leisure activities in the city and identify urban nighttime leisure spaces. These applications at the city level not only delineate spatial variations in human activities but also reveal diverse functional preferences across agglomerations within a city (Xue et al., 2020).

2.4 Synopsis of Current Research Challenges

A critical assessment of existing research reveals several methodological restrictions that highlight the gaps addressed by our study. Firstly, due to data availability, most current literature relies on economic datasets at an aggregated geographical level, adopting a discrete-space perspective to explore the spatial patterns of agglomeration economies (Guillain & Le Gallo, 2010). This approach often leads to an oversimplified understanding of industrial agglomeration.

Additionally, measurement strategies in the functional dimension predominantly depend on specific classification systems or predetermined taxonomies provided by official censuses. This reliance on human-defined taxonomic resolutions can limit the ability to comprehensively describe the diversity of local industrial sectors. Consequently, many studies focus on sectoral agglomeration economies using economic statistics and predefined taxonomies, often restricting their analyses to micro-scale environments within city regions or metropolitan areas (Yu & Liu, 2021).

These methodological constraints reveal a significant gap in spatial and functional knowledge concerning economic activities within a megaregional context. Our study addresses this gap by proposing more scalable and unbiased strategies, aiming to explore economic dynamics comprehensively. By choosing diverse data sources and employing advanced methods, we can better measure and understand industrial agglomeration, providing a more nuanced and complete picture of economic activities across larger geographical scales.

Chapter 3. Methodology

This chapter presents a systematic description of major methodologies used by the thesis and relevant research papers, with aims of better understanding agglomeration economies in megaregional context. The methodologies are mainly contributed from three knowledge domains: spatial data science, machine learning, and network analytics. Figure 1 shows the major domain knowledge of this research thesis. It should be noted that the machine learning and network analytics proposed in this thesis are both employed in the context of spatial data science, which can be considered as parts of geospatial analytics. Specific methodologies used by affiliated papers are reported in Figure 2, establishing a research framework of comprehensively studying the agglomeration economies. All methodologies shown in this framework can be associated with the domain knowledge presented in Figure 1.

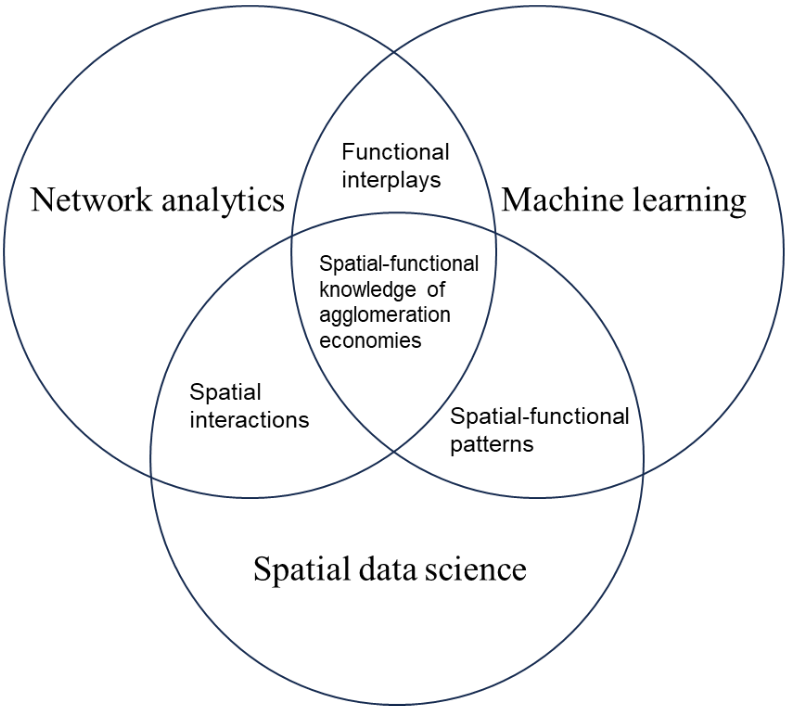


Figure 1. Domain knowledge of this thesis.

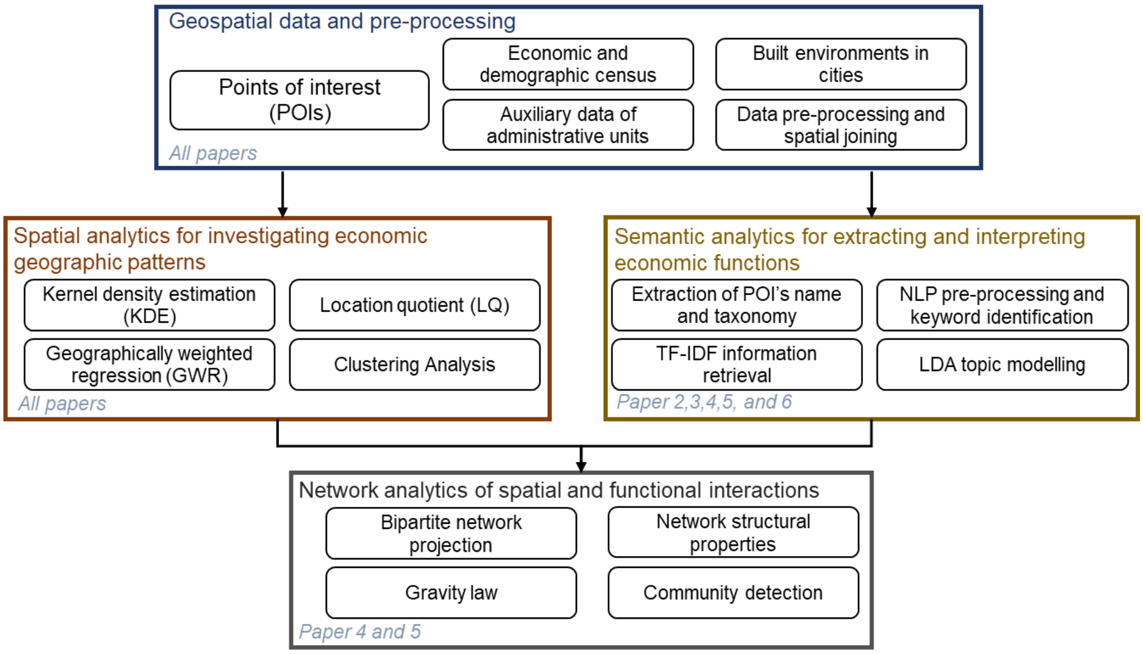


Figure 2. Research framework on the spatial and functional patterns of agglomeration economies. Append papers apply corresponding methods that reported in this Chapter.

The case studies included by relevant publications presented in this thesis surround a core topic that is the spatial and functional knowledge of agglomeration economies in megaregional context. Among these case studies, four are considered as the principal demonstrations related to this thesis and reported subsequently, regarded as individual chapters. To better illustrate the methodology interconnected by these four chapters, Figure 3 briefly presents the chapter-oriented research framework, including the main research objectives and methodologies. The first study shown in Chapter 4 is to evaluate the spatial patterns of sectoral urban agglomeration using conventional census statistics and examine the relationships to built environments and socio-demographics. Having considered the restrictions summarized in Chapter 2 and the context of megacity regions, the second case study contained in Chapter 5 provides a geospatial data-driven perspective into the spatial-functional organizations of sectoral industries in urbanized megaregions. Spatial boundaries of agglomeration economies are delineated, while the functional characteristics are labeled. The third case study in Chapter 6 moves a further step focusing on the geographic disparities of industrial activities using quantitative strategies. Two areas within the megaregion are selected and investigated to understand their spatial-functional characteristics and locational preferences. Lastly, Chapter 7 introduces a case study that network analytics are incorporated into this research stream to explore the spatial-functional roles of megaregional agglomeration and their network behaviors.

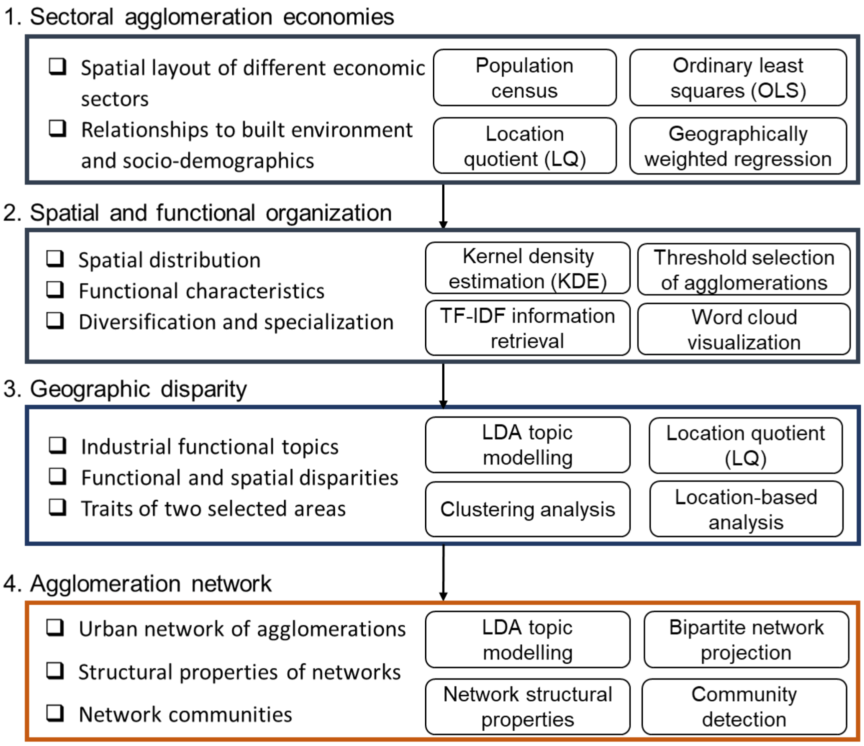


Figure 3. Detailed objectives and methodologies included by the core papers, as introduced in following chapters.

3.1 Spatial Data Science

Spatial data science within the context of studying urban agglomeration has involved leveraging advanced analytical tools to explore the spatial relationships, patterns, and dynamics of urban clusters. It encompasses techniques from data analysis, geospatial model and metrics, to understand the geographical concentration of economic activities, infrastructure, population distribution, and their interdependencies within urban areas. These interdisciplinary approaches help uncover insights into the spatial organization, functional specialization, and development trajectories of urban agglomerations, offering valuable perspectives for urban planning, economic analysis, and policy formulation.

3.1.1 Geospatial data and preprocessing

Geospatial data plays a pivotal role in enhancing our comprehension of agglomeration economies. By providing detailed spatial information on the distribution, density, and interrelations of economic activities within regions, geospatial data enables us to analyze and visualize the intricate patterns of agglomeration. This subsection delineates the primary spatial datasets utilized in this thesis as well as their preprocessing details, encompassing population by-census, point of interests (POIs), built environments, and socio-demographics.

For identifying agglomeration economies, the employment data retrieved from zonal employment statistics in 2016 population by-census is employed in the case study of Chapter 4 that is available at District Council Constituency Areas (DCCA) level. The employment information is industry-specific and covers the majority of various economic activities in Hong Kong, allowing us to evaluate the industrial agglomerations. Industries investigated in this study are classified into three economic sectors: manufacturing, service, and trade and logistics industries.

People frequent diverse Points of Interest (POIs) for varied activities in cities, indicating the potential of POIs in furnishing significant insights into economic activities and detailing the spatial and functional traits of agglomerations. In chapter 5, 6, and 7, we retrieve more than 220,000 POIs within the GBA from AutoNavi, a prominent location-based service provider in China. These POIs encompass industrial and commercial establishments, offering a broad spectrum of economic activities for analysis. Aligned with prevalent economic theories, the post-industrial landscape conventionally segregates industries into primary, secondary, tertiary, and quaternary sectors (Ritzer & Stepnisky, 2007). These sectors primarily focus on raw material procurement, manufacturing, service provision, and innovative industries, respectively (Kenessey, 1987). To classify the POIs into corresponding sectors, we leverage the map provider's three-tiered category system. The first and second-tier categories aid in identifying sectoral industries (e.g., Machinery & Electronics in the secondary sector), while the third-tier category, comprising specific branding details and company characteristics, serves for a robust cross-check through keyword searches and sampling strategies. Consequently, we categorize all POI records into four sectoral industries, with detailed description in Table 1.

Table 1. Classification of POIs.

Sector	Description	POI Category
Primary	Raw materials extraction and harvesting such as agriculture and mining.	Farms, ranches, fishing, horticulture, wood, forest, gardening, agriculture-related, mining, and other industries.
Secondary	Manufacturing and processing that relate to the production of goods.	Automobile, chemical, electronics, energy, food, paper, steel, textile, and other industries.

Tertiary	Services for consumers, containing a wide range of businesses.	Advertising, entertainment, catering, transportation, commercial, financial, and other service industries.
Quaternary	Knowledge-oriented, information-based services and high-tech industries.	Information and communication technology, R&D, and other high-tech industries.

Two groups of aggregated variables are included in the regression analysis reported in Chapter 4: the built environment and socio-demographic characteristics. Referring to the built environment, as proposed by Cervero and Kockelman (1997), we conceptualize built environment indicators into 3Ds including density, diversity, and design. Specifically, density evaluates the activity density of a place that has been often calculated by population per spatial unit. Diversity is a reflection of the existence of various urban functions and services, indicating the vibrancy of an area. Design metric assesses the urban structure such as road networks. Given that residents in Hong Kong highly rely on public transportation for daily commuting, distance to the nearest metro station and the number of bus stops are further included (He et al., 2018). This allows us to quantitatively estimate the effects of public transit on urban agglomerations. Two additional locational variables are the proximities to Central that is the central business district (CBD) in Hong Kong and to the nearest border with mainland China. Socio-demographics are also recommended as essential factors to explain the appearance of urban agglomerations and thus adopted in this study. Retrieved from the 2016 population by-census, four aggregated metrics are considered, namely, education level, population age level, mean income, jobs-housing ratio, and race composition. The summary of built environment and socio-demographic characteristics are presented in Table 2.

Table 2. The built environment and socio-demographic characteristics used in Chapter 4.

Variable	Description	Mean	Std. Dev.
Built environment			
Population density	Population density per sq. km.	70250.75	58111.58
Land use diversity	0 to 1 (single to most diverse land mix)	0.27	0.08
Number of intersection	Road network intersections	265.74	373.18
Road density	Road density in km. per sq. km.	36.60	20.01
Distance to metro	Distance to nearest metro station in km.	0.11	0.14

Number of bus stop	The number of bus stop	10.01	12.50
Distance to Central	Distance to Central in km.	1.16	0.75
Distance to border	Distance to the nearest border with mainland China in km.	1.99	0.86
Socio-demographics			
Education level	% of population with postsecondary degree or above	0.28	0.11
Population age level	Weighted mean age of population	43.17	2.62
Mean income	Mean income	15567.41	3897.92
Jobs-housing ratio	The ratio between working and resident population in one DCCA	0.09	0.04
Race composition			
Percent White	The percentage of White population	0.01	0.03
Percent Filipino	The percentage of Filipino population	0.02	0.03
Percent Indonesian	The percentage of Indonesian population	0.02	0.01

Source: Built environment variables are computed based on the data from Hong Kong Census and Statistics Department (2016), Land Department, Transport Department and OpenStreetMap (OSM). Socio-demographic variables are computed based on the data from Hong Kong Census and Statistics Department (2016).

3.1.2 Spatial model and metrics

Location quotient (LQ)

Urban agglomerations often exhibit a higher concentration of economic activities than other less vibrant areas in cities. In this respect, the location quotient (LQ) is applied in the case study of Chapter 4 and could be the most common approach for spatially delimiting agglomerations regarding urban employment patterns (Beaudry & Schiffauerova, 2009; Nakamura & Paul, 2019). This index measures the ratio between the local and global percentage of sectoral employment. In this study, we calculate LQ for economic sectors at each DCCA. The equation of LQ is shown as follows:

$$LQ_{ij} = \left({}^{e_{ij}}/_{e_i}\right)/\left({}^{E_j}/_{E}\right) \quad (1)$$

where LQ_{ij} represents the location quotient of a specific sector j for DCCA i, e_{ij} and e_i denote the employment of the sector j and total employment for DCCA i, respectively, while E_j and E are the total employment of entire study area in sector j and in all sectors. To interpret the results, a DCCA with an LQ of greater than 1 indicates that it is overrepresented in a particular economic sector in comparison to the entire study area. In contrast, an LQ value lower than 1 denotes an underrepresentation in the employment concentration of the examined economic sector.

However, one major weakness for defining urban agglomeration is that there is no clear agreement regarding the theoretical LQ cut-off values in existing studies. Former investigations have defined their LQ cut-off parameters in an arbitrary manner based on their local situations. For example, Malmberg and Maskell (2002) defined an urban agglomeration as a market region with LQ larger than 3. Thus, quantifying agglomeration can be highly dependent upon the cut-off values, in which the statistical significance cannot be demonstrated.

Local indicator of spatial association (LISA)

The local indicator of spatial association (LISA) is implemented to reveal the spatial pattern of agglomeration economies, which has been presented in relevant case studies (Guillain & Le Gallo, 2010). LISA is a metric for detecting local patterns and the type of geographic cluster founded on the traits of spatial dependence and shown as:

$$I_i = \frac{(x_i - u)}{m_i^2} \sum_{j=1, j \neq i} w_{ij}(x_j - \mu) \tag{2}$$

$$m_i^2 = \frac{\sum_{j=1, j \neq i}(x_j - \mu)^2}{N - 1} \tag{3}$$

here, I_i is the local Moran' I index for DCCA i, x_i and x_j denote the LQ values for DCCA i and j respectively, u represents the average LQ value, and N is the total number of DCCA in our case study. Note that w_{ij} represents the spatial weight matrix, in which $w_{ij} = 1$ means that DCCA i and j are adjacent and vice versa. The results generated from LISA analysis can be classified into four quadrants: High-High (HH) refers to a spatial cluster of DCCA all with high LQ; High-Low (HL) refers to a DCCA with high LQ surrounded by DCCAs with low LQ. Similarly, Low-Low (LL) indicates a spatial cluster of DCCA all with low LQ, and Low-High (LH) indicates a DCCA with low LQ surrounded by DCCAs with high LQ. In this research, DCCAs with a total employment population density below 1000 per square km are excluded for mitigating the sample selection bias. The analysis of LISA is conducted using ArcGIS 10.7.

To evaluate the robustness, an alternative index is introduced to measure the sectoral agglomeration economies. The Herfindahl-Hirschman index (HHI) is a widely applied indicator for measuring the concentration or specialization between companies. It is represented by $\sum_{i=1}^{n} s^2$, where n denotes the total number of companies and s^2 is the square of the share of employment for each company in an area. In our case study, given a particular industrial sector with that of all sectors in an area, the HHI is computed by the formula ${s_j}^2 * \sum_{i=1}^{n} s^2$, where n denotes the total number of economic sectors, ${s_j}^2$ and s^2 are the square of the share of employment for a specific sector j and each sector in a DCCA. The Pearson's correlation between LQ and HHI are 0.96, 0.91, and 0.93, as for manufacturing, service and trade and logistics sectors, respectively, suggesting that LQ can be a reliable indicator for measuring agglomeration economies. LISA is also implemented to identify the spatial clusters of HHI.

Kernel density estimation (KDE)

Kernel Density Estimation (KDE) is extensively employed to assess the distribution of point events and identify hotspot areas, free from reliance on specific areal units. In the case studies of Chapters 5, 6, and 7, Planar KDE with a quartic function is utilized to construct the spatial distribution of Point of Interest (POI) density. The equation of KDE is expressed as:

$$P_i = \sum_{i=1}^{n} \frac{1}{\pi R^2} K(\frac{d_{ij}}{R}) \tag{4}$$

where P_i signify the estimated kernel density value at location i, R denotes the bandwidth, representing the range of the search radius within the examined area. Additionally, d_{ij} represents the distance between research points i and j, and n signifies the total number of sampling points within the range R centred around point i

Researchers have emphasized the significance of the bandwidth as a crucial parameter in Kernel Density Estimation (KDE), asserting its substantial impact on the resulting density surface (Sheather & Jones, 1991). In alignment with prior studies addressing analogous spatial extents, this study considers three bandwidth parameters—namely, 2000, 3000, and 4000 meters (Deng et al., 2019). Figure 4 illustrates the density surface variations with different bandwidths, utilizing manufacturing industries in the Foshan-Guangzhou-Dongguan region as an illustrative example. Notably, when employing search bandwidths of 2000 and 3000 meters, numerous fragmented small-scale hotspots emerge. However, these small-scale hotspots inadequately represent industrial agglomerations, thereby complicating the interpretation of spatial patterns. In contrast, the results obtained with a 4000-meter bandwidth demonstrate that isolated hotspots gradually integrate with their surroundings. The density pattern becomes smoother and more cohesive compared to the outcomes obtained with smaller bandwidths. Considering the objective of delineating regional agglomerations, the choice of a 4000-meter bandwidth is deemed appropriate for subsequent analyses.

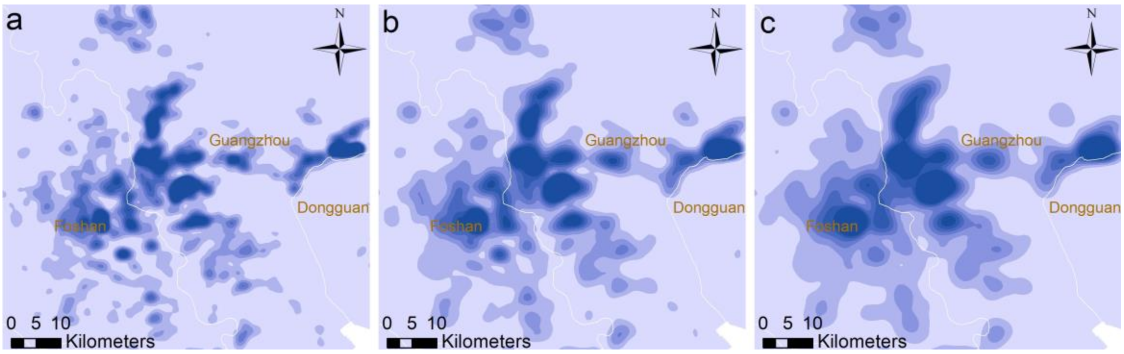


Figure 4. Kernel density results of the secondary industrial activities in Foshan-Guangdong-Dongguan region within the GBA with different search bandwidths. (a) 2000 m. (b) 3000 m (c) 4000 m.

Kernel density is regarded as an indicator with a positive correlation to the concentration of industrial activities. For assessing the concentration level, the use of standard deviation is advocated, as it effectively discerns and emphasizes areas of concentration compared to broader regions (Chainey et al., 2002). In this context, a threshold value for industrial concentrations is established at two standard deviations, derived from the kernel density surfaces generated for various sectoral industries, as depicted in Figure 5.

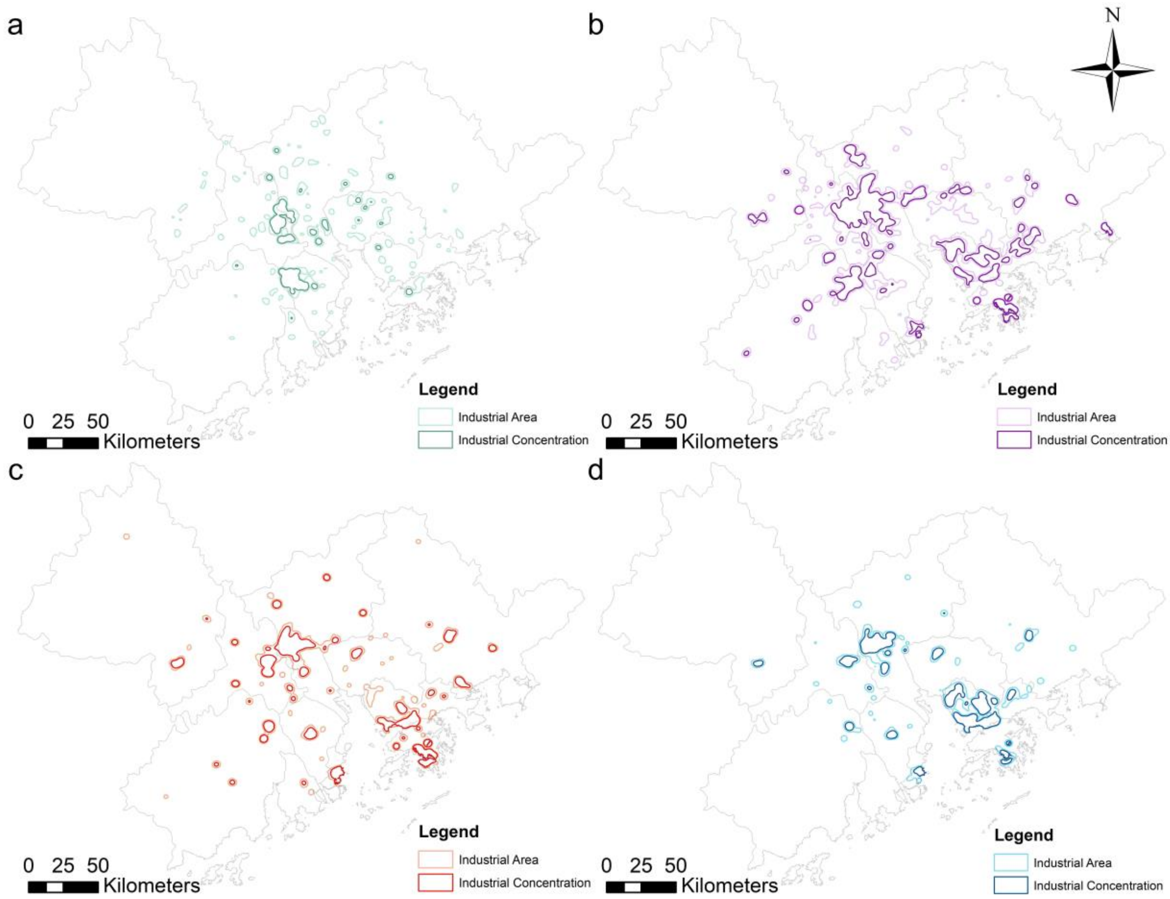


Figure 5. The spatial distribution of industrial concentrations in the GBA. (a) Primary sector. (b) Secondary sector. (c) Tertiary sector. (d) Quaternary sector.

It is essential to note that identified concentrations alone may not accurately signify industrial agglomerations, as the accumulation of POIs in small-scale locations can inflate kernel density values. For instance, multiple factories situated within a single industrial building may contribute to a high kernel density without constituting a true agglomeration. In accordance with prevailing literature, it is recommended that the gross employment density of an urban center should represent at least 1% of the overall employment (Huang et al., 2017; Muñiz et al., 2008). Drawing inspiration from these analogous cases, a relative-threshold method is introduced for agglomeration detection in this case study. However, due to the inherent diversity of industrial activities, threshold ratios may vary for specific sectors. For instance, sectors such as agricultural farms and industrial plants may exhibit larger land coverage than others, resulting in a lower concentration degree. Therefore, for primary and secondary sectors, threshold ratios ranging from 0.4% to 1% are considered. Through a comparison of the total number of identified agglomerations and the incorporation of local knowledge, the applied ratios for each economic sector in this study are determined as 0.4%, 0.5%, 1%, and 1%, respectively.

An industrial agglomeration is defined as a geographical concentration that meets two specific criteria. Specifically, we posit that industrial agglomerations should exhibit a density value surpassing two standard deviations, along with a total count of POIs greater than or equal to a specified ratio of the overall quantity. The equations are displayed as follows:

$$D_{t,i} \geq 2\sigma_{GBA} \tag{5}$$

$$P_t \geq R_{sector} \times P_{GBA} \tag{6}$$

Correspondingly, $D_{t,i}$ is the KDE value of each unit within a concentrated area, σ_{GBA} is the standard deviation of the corresponding density surface. As for the equation (X), P_t denotes the gross number of POIs contained in a concentration area, R_{sector} represents the aforesaid ratios for four different sectors, and P_{GBA} is the total amount of sectoral POIs in GBA.

Geographically weighted regression (GWR)

In Chapter 4, since the built environment and socio-demographic characteristics are affected by local context, the spatial non-stationarity of variables cannot be explained by using a global regression. As reported in Table 3, it could be observed that all dependent variables are significantly and spatially clustered across study areas, thus rejecting the null hypothesis and suggesting applicability for using local regression models.

Geographically weighted regression (GWR) is therefore introduced to cope with this issue, namely, the varying association between independent and dependent variables across geographical space (Wang et al., 2018). GWR is considered as a localized regression model and allows the estimated coefficients to be varied, which can mitigate the regression bias resulting from spatial heterogeneity. In our work, the equation of GWR is shown follows:

$$LQ_i = \beta_{0i} + \sum_k \beta_{ki} x_{ki} + \varepsilon_i \tag{7}$$

where LQ_i is the concentration degree of the sectoral agglomeration economies for DCCA i, β_{0i} denotes the local regression intercept, while x_{ki} and β_{ki} represent the kth explanatory variables and its local estimated coefficients of DCCA i, respectively, and ε_i is the error term for DCCA i. Note that β_{ki} is regarded as a spatial function whose values are associated with the two-dimensional specific coordinates of the centroid point of DCCA i. Specifically, estimated coefficient β_{ki} can be expressed as:

$$\beta_{ki} = (X^T W_i X)^{-1} X^T W_i Y \tag{8}$$

where W_i represents a spatial weight matrix and contains a diagonal elements to depict weights for each DCCA used to estimate the local parameter of DCCA i. The determination of the spatial weight matrix W_i should be considered in the calibration of the GWR model. Here, we calculate the weight based on a Gaussian weighting kernel function, which assigns higher weights to the nearby areal units that are closer to the DCCA i. To define the adaptive kernel bandwidth of Gaussian function, Akaike Information Criterion (AIC) is considered that can specify a trade-off value between the prediction accuracy and the complexity of our regression model. The global and local regression analyses are conducted in Python 3.7.

Table 3. The spatial autocorrelation of sectoral LQ by Global Moran'I test.

Dependent variable	Moran's index	Z-score
Manufacturing	0.31	30.69**
Service	0.37	37.16**
Trade and logistics	0.23	22.62**

** Represents statistically significant at the $p < 0.01$ level.

3.2 Machine Learning

3.2.1 Regression analysis

In Chapter 4, global and local models are employed to explore the association between the agglomeration economies and the impacts from built environment and socio-demographic characteristics. Explanatory variables are formed by aforenoted features, and the dependent variable is the sectoral LQ. To address multicollinearity issues, Pearson correlation analysis is performed to evaluate the correlations among all explanatory variables. Given that multicollinearity can lead to bias and inflate standard errors in this regression analysis, variables with correlation coefficients that are larger than 0.7 are removed.

The aim of global regression is to analyze the impacts from built environments and socio-demographics on agglomeration economic activities. The equation of ordinary least squares (OLS) is shown follows:

$$LQ = \beta_0 + \sum_k \beta_k x_k + \varepsilon \tag{9}$$

here, LQ is the concentration degree of the sectoral agglomeration economies, β_0 denotes the global regression intercept, while x_k and β_k represent the kth explanatory variable and its estimated regression coefficients, respectively, and ε denotes the random error. The description of global model is provided in section 3.1.1, as a statistical method of spatial data science.

3.2.2 Natural language processing (NLP)

Term Frequency and Inverse Document Frequency (TF-IDF)

Text mining, within the realm of natural language processing (NLP), pertains to information retrieval techniques that convert textual data into high-quality, interpretable, and easily comprehensible knowledge. In the context of NLP, a widely employed and effective term-weighting scheme is Vectorized Term Frequency and Inverse Document Frequency (TF-IDF). This unsupervised learning technique has been extensively utilized in semantic exploration applications for comprehending urban areas (Berger et al., 2000; Liu et al., 2020c). TF-IDF functions by transforming the text content of a document into a bag of words, assigning weighted values to each term based on their frequency within the document and their significance in comparison to the entire corpus. Given a certain corpus that is a document collection C, a word t, and a test document $d \in C$, the formula of TF-IDF is shown below:

$$w(t,d) = TF(t,d) * \log\left(\frac{N_C}{f_{t,C}}\right) \tag{10}$$

Here, $w(t,d)$ equals the weighted frequency of a word in a document d, $TF(t,d)$ is the frequency of occurrences of word t in a document d, N_C is the number of documents of the corpus, and $f_{t,C}$ denotes the number of documents containing word t in collection C.

In Chapters 5, 6, and 7, the registered name of a Point of Interest (POI) can be tokenized into multiple words, each implying specific industrial functions. Each industrial agglomeration and all agglomerations within the Greater Bay Area (GBA) are treated as a test document and the entire corpus, respectively. Utilizing the TF-IDF model, distinctive semantics related to functional characteristics are identified for each agglomeration. For instance, an urban agglomeration within a metropolitan area might host numerous printing factories catering to the local publishing and press industries. Consequently, it is logical to infer that the keyword "printing" is more frequent and, hence, more significant than other industrial functions.

Latent Dirichlet Allocation (LDA)

The popularization of the natural language process (NLP) has facilitated the applications of various topic modeling techniques in multi-discipline research communities. Among these techniques, the Latent Dirichlet Allocation (LDA) introduced by Blei et al. (2003) has been a prevalent approach in extracting latent thematic topics and widely applied in urban studies to help understand sophisticated human activities (Gao et al., 2017). LDA is an unsupervised model to compute term frequencies in each document by deploying a bag-of-words idea to produce probabilistic topics. The main rationale behind LDA is that two multinomial distributions characterizing two internal relationships, topic-to-word and document-to-word matrices, are generated in LDA. Examined documents are denoted as a joint probability distribution over a predefined number of latent topics, and a joint probability distribution over words can characterize each topic. In other words, LDA is to deem each test document as a collection of topics with corresponding weights. Similarly, each topic is still considered as a collection of keywords with a certain weight, suggesting that the thematic characteristics of each topic can be deduced by words with a higher importance weight. Altogether, LDA can be described as a text-generative process that interactively creates text content to infer the topics in a corpus.

Analogously, the manufacturing POIs can indicate functional semantic information (e.g., fabric and metal tubes). Their registered names are taken as the keywords, while each agglomeration is seen as a document. All the identified agglomerations in GBA form the corpus. On this basis, a generated topic can represent the thematic functions of their associated industrial activities spatially co-exist within the same locations. Such patterns can be considered as the consequence of the agglomeration process. The probabilistic distribution of various keywords in industrial agglomerations thus can be computed and unveiled with a discrete probability distribution for each functional topic. Associations between functional topics and keywords are displayed as a topic-to-word matrix. Hence, each manufacturing area's thematic topics (e.g., clothing, decoration, material) can be figured out with normalized weight values representing the topic significance via a document-to-topic matrix. The analogous process in terms of the functional dimension is shown in Figure 6. A functional topic is a collection of dominant keywords with weights extracted from the POI's name. These related keywords are typical representatives within a topic and can be used to abstract what industrial function is all about. The weights reflect how important a keyword is to that functional topic where the keyword may represent production in a manufacturing context. For instance, a topic mainly contains special keywords with evidently high importance weights like steel, metal, aluminium, or casting. Then, this topic can be interpreted as a functional behavior anticipated to appear in metal- or material-oriented industrial clusters.

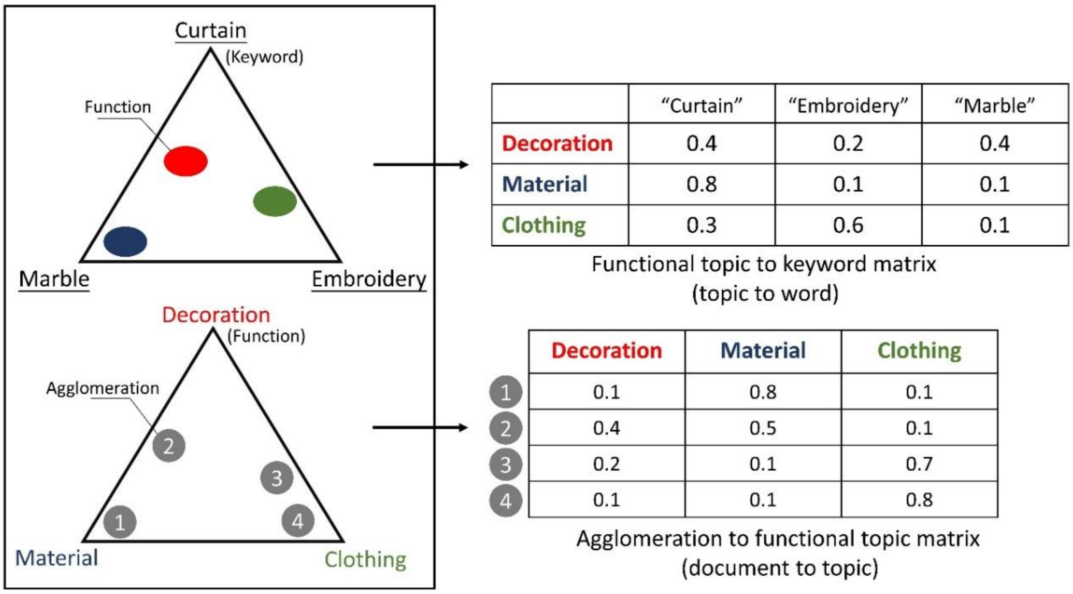


Figure 6. The illustrative process of document-to-topic and topic-to-word matrices generated in the process of LDA that is analogously applied to extracting manufacturing functions from industrial agglomerations in this study.

One key parameter in the LDA process is the number of generated topics that should be pre-defined first. As Bahrehdar and Purves (2018) suggested, the coherence score is the quality indicator to evaluate the number of topics in the LDA model. A higher coherence score hints that a better quality of output results in semantically coherent content within a topic. Besides, knowledge of local manufacturing economies is another crucial reference and should also be considered in determining the number of latent topics. Thus, coherence score and local knowledge of the GBA are simultaneously considered in this study. The LDA topic modelling is conducted using the scikit-learn package that is a machine learning library built on Python.

To figure out the significant industrial functions in agglomerations, the standard deviation index is introduced to compute the importance of thematic topics and highlight the over-representative industrial functions of each agglomeration. The equation is shown as follows:

$$S(k,i) = \frac{p_{k,i} - \mu_k}{\sigma_k} \quad (11)$$

Here, $p_{k,i}$ represent the probability of topic k for agglomeration i, while μ_k and σ_k denote the mean value of and the standard deviation of topic k probabilities over all agglomerations, respectively. This metric can effectively quantify how uniqueness a thematic function is in a place compared to a large area, determining which industrial functions lead this agglomeration economies more distinguishable (de Oliveira Capela & Ramirez-Marquez, 2019).

Hierarchical clustering

As latent topics are extracted from POIs using LDA, various thematic industrial functions are unveiled in the agglomeration-to-function matrix, which can be related to the profile of local manufacturing economies. We aggregate all industrial agglomerations into groups via their probability distribution over latent topics. In this case, hierarchical cluster analysis with an agglomerative (bottom-up) strategy is implemented over different industrial agglomerations. The important weight of extracted topics is used as an input feature. Given an industrial agglomeration with functional topics, the input feature is denoted as a n-dimensional vector:

$$X = [W(1,i), W(2,i), \ldots\ldots, W(n,i)] \tag{12}$$

Here, $W(n,i)$ represents the important weight of n topic in agglomeration i. As for the clustering process, the agglomerative strategy starts by considering each observation as a separate cluster. All separate clusters repeatedly implement following steps: two clusters with most similar attributes are identified and then merged into one cluster. This iterative process ends until all clusters are merged, as they move up to the top of hierarchical structure. To estimate the similarity, Euclidean distance I is computed to measure the distance between two separate clusters i and i' based on their n-dimensional vectors. The equation is represented as follows:

$$I = \sqrt{\sum_{k=1}^{n} \big(W(k,i) - W(k,i')\big)^2} \tag{13}$$

The linkage criterion used to determine similar clusters is the Ward's-linkage method, in which this method computes Euclidean distance as the increase in the error sum of squares (ESS) after two clusters are merged into one single cluster (Ward Jr, 1963). Using the clustering analysis can provide a collective perspective to explore functional patterns.

3.3 Network Analytics

Bipartite Network Projection

As the subdomain of the complex networks, the two-mode bipartite network is a set of nodes with two disjoint and independent partitions. Against the one-mode network, one rule is additionally specified for a bipartite network that nodes within one partition cannot connect to one another. Previous studies have adopted this network approach to construct a world city network via advanced producers service (APS), where two cities are assumed to link if they are home to a branch of the same APS firms (Neal, 2011). The interlocking network strategy based on bipartite network projection is used to transform the bipartite network with two separate node collections into the one-mode network.

Having followed the logic of the interlock network approach, we construct functional network by describing agglomeration-to-function matrix F. For this matrix, agglomerations would be seen as the nodes and $F_{i,j}$ denotes that the presence or absence of a specific industrial function j in agglomeration i. This matrix can be transformed into an agglomeration-by-agglomeration network matrix A, documenting the relationships between each pair of agglomerations. The equation is shown as follows:

$$A_{a,b} = \sum_j F_{a,j} \times F_{b,j} \tag{14}$$

Here, $F_{a,j}$ and $F_{b,j}$ represent the co-presence of functional topic j in agglomeration a and b, while A_{ab} denotes the functional connectivity between these two agglomerations.

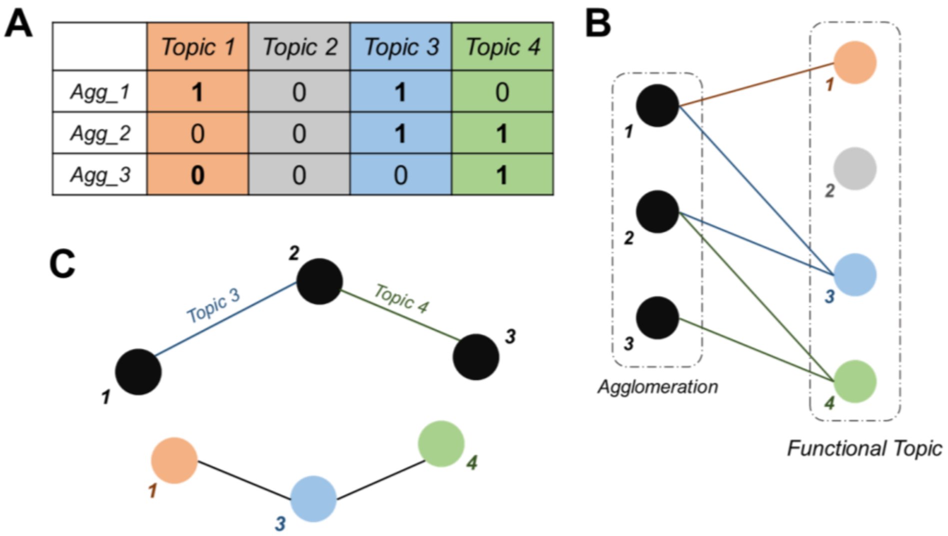


Figure 7. (A) The illustration of the document-to-topic matrix that is analogously applied to extracting manufacturing functions from industrial agglomerations. (B) The illustration of the two-mode bipartite network (graph) based on the relationships between agglomeration and functional topics. (C) The one-mode networks of industrial functions and agglomerations converted from the two-mode bipartite network.

The above process can be regarded as a typical bipartite network projection in graph theories from the two-mode network (i.e., agglomeration-to-function) to the one-mode network (i.e., agglomeration-by-agglomeration). The illustrative process of bipartite network projection is illustrated in Figure 7. The agglomeration-to-function matrix F is seen as bipartite network architecture. Given a bipartite graph $G = (U, V, E)$, U denotes the node set of agglomerations within the GBA and nodes V indicate various industrial topics. E depicts the relationship between agglomerations and their containing functions as an edge. Accordingly, the relationship in terms of industrial functions from agglomerations can be aggregated to build a bipartite projection network $P = (A, E)$. Here, A denoted a set of individual agglomerations in the GBA, and E refers to a set of edges. The edge between two nodes $e_{a,b} \in E$ corresponds to the functional relationships between two agglomerations a and b. Noted that since we define the network as an undirected network, $e_{a,b}$ and $e_{b,a}$ are identical edges. Meanwhile, the one-mode industrial functional network can be also converted from the two-mode network, in which relationships of multiple industrial functions are inferred based on co-location or co-presence patterns within an agglomeration. Having constructed the agglomeration network, four selected structural network properties are employed to characterize the network structure and the relevant descriptions are summarized in Table 4.

Table 4. Network properties used to characterize agglomeration networks.

Network property	Description
Size	The number of nodes and edges in a network
Density	The ratio of observed edges over all potential edges in the network, between 0 to 1
Degree	The number of connections contained by a node that connects to other nodes in a network
Global clustering	A metric to measure the clustering of the entire network that is the ratio of observed closed triplets over all potential triplets in the network

Community detection

Another research aim of Chapter 7 is to inspect the community structure of agglomeration networks, which is a crucial property of real-world networks. The community structure can reflect the partition of networks by various groups of closely connected nodes. Developed by Blondel et al. (2008), the Louvain method is considered to detect community structure in this case. Intuitively, this algorithm assumes that each node from the network is first considered as a separate community. At each step, separate nodes would be merged into the same communities determined by computing the most significant increase in modularity score and terminated until there is no increase in the modularity score. Given an undirected graph G', the modularity Q is defined as follows:

$$Q = \frac{1}{2m}\sum_{i,j}\left[w(e_{i,j}) - \frac{k_i * k_j}{2m}\right]\delta\left(c(v_i), c(v_j)\right) \tag{15}$$

where $w(e_{i,j})$ denotes the sum of weights in this undirected graph, k_i and k_j are the sum of the weights of the edges attached to nodes v_i and v_j, $c(v_i)$ and $c(v_j)$ denote communities of the nodes v_i and v_j, respectively, and $\delta\left(c(v_i), c(v_j)\right)$ is Kronecker delta function used to depict whether two agglomerations belong to the identical communities when it takes the value of 1 and 0 otherwise.

Chapter 4. Revisiting to Urban Agglomeration Economies

4.1 Background and Motivation

Urban scholars have contributed to the knowledge of economic geography by the means of constructing relationships between the types of agglomeration and their relevant determinants. Agglomeration economies are generally characterized as the Marshallian and Jacob agglomeration economies (Beaudry & Schiffauerova, 2009). The former type of economies (i.e., localization) highlights that the advantages of industrial agglomeration are returned from input sharing, labor market pooling, and knowledge spillovers (Delgado et al., 2014). By contrast, the Jacob agglomeration economies (i.e., urbanization) suggest that benefits stem from the concentration of diverse economic activities. In this sense, these findings have provided valuable implications for understanding the external impacts on agglomerations with consideration of productivity functions (Jacobs, 1961).

The specific association between economic activities and local determinants has attracted attention from the research community that largely confirms the impacts of the built environment and socio-demographic characteristics on economic activities. The local factors of the built environment such as population density, transportation infrastructure that are proposed to study commuting behaviors have been discussed, with a special focus on job accessibility to agglomerated workplaces (Chatman, & Noland, 2011; Ewing & Cervero, 2010). Another vein of the research has suggested the connections between economies and socio-demographic characteristics. Aggregated variables such as median income, median age, education level, and race composition often influence industrial local decisions, as they represent essential images for sector-specific labor market pooling (Ellen & Ross, 2018; Overman & Puga, 2010). This information is of great relevance to characterize the sector-specific industries in cities to estimate socioeconomic needs in urban planning and formulate appropriate development strategies. However, featuring local agglomeration economies, the impacts from built environment and socio-demographics has not been properly studied by existing research works.

As abovementioned, this oversight is problematic because it is frequently essential to take into account the built environment and socio-demographics to deliberate policies for economic development in urban contexts. For instance, detailed knowledge on the impacts of the built environment can help policymakers and planners to evaluate the consequence of public facility investment. Moreover, bridging agglomeration economies and socio-demographics can support neighborhood or suburban development projects to offer more appropriate employment opportunities to different resident groups. Accordingly, there is a purposeful question for our work that whether such ideas can propose a unique and vital viewpoint to depict the underlying mechanisms of economic activities in cities.

Using the 2016 population by-census in Hong Kong, this research investigates the relationship between urban agglomeration economies and local context, including a wide range of built environment and socio-demographic factors. With a focus on the spatial varying patterns of such relationships, the sectoral employment at District Council Constituency Area (DCCA) allows us to address three key research questions:

1) What are the key patterns of the spatial distributions of different economic sectors?
2) To what extent can built environment and socio-demographic characteristics are associated with agglomeration economies?
3) And are there notable spatial variations among these associations given local sectoral economic activities across study areas?

We first apply a location quotient for measuring the concentration degree of selected economic sectors including manufacturing, service, and trade and logistics. Then, ordinary least squares (OLS) model and geographically weighted regression (GWR) are implemented to investigate the relationships between agglomeration economies and the influences from built environment and socio-demographic characteristics.

4.2 Main Findings

In this subsection, we report the research findings, which can help us to answer the proposed research questions. Patterns of spatial heterogeneities in terms of three sectoral agglomeration economies are first described. Then, the regression results are presented to explain the association between the agglomeration economies and built environment and socio-demographic characteristics.

4.2.1 Urban agglomeration of sectoral industries

In addressing RQ1, we exhibit the results of identified agglomeration economies using choropleth maps. Figure 8 (A) exhibits the spatial distribution of the manufacturing agglomeration economies. It is obvious that agglomeration economies with higher LQs tend to disperse across the north and west areas in the new towns of New Territories. These new towns established by the government have successfully attracted a large proportion of manufacturing activities from commercial districts in Kowloon and Hong Kong Island that contain extremely low LQs. Meanwhile, our results demonstrated that multiple sub-centers formed by new towns can be defined as a polycentric urban structure that can provide different manufacturing agglomerations for various industries. Further, compared to other suburban areas, most of the identified agglomerations have appeared near the centers of new towns that can provide better transportation accessibility and daily services for workers. An unanticipated finding is that no significant agglomeration is detected in Kwun Tong district that is located in eastern Kowloon and used to be a manufacturing base in Hong Kong. Also, there are a few agglomerations of manufacturing activity in the western Kowloon and eastern Hong Kong Island.

Interestingly, far distinct spatial patterns are indicated in agglomeration economies of the service sector, in comparison with the manufacturing sector. As shown in Figure 8 (B), a large-scale agglomeration has appeared along the northern shore of Hong Kong Island, which is the historical, political, and economic center of Hong Kong. It can thus be suggested that the downtown area with the most vibrant commercial activities has been extended along Victoria Harbor, as opposed to the centric shape of most monocentric cities. As the heart of financial and commercial activities in Hong Kong, Central is situated in the midst of western Hong Kong Island. Another commercial center within this agglomeration is the Causeway Bay located in eastern Hong Kong Island, which is seen as the energetic retail hub of Hong Kong. With respect to the Kowloon and New Territories, agglomeration economies are confirmed in a limited number of areas. Particularly, Tsim Sha Tsui is a famous tourist district and a center for leisure and entertainment activities, and its principal industries relate to commercial and hospitality activities. Besides, we can still notice an agglomeration area located along the coast of eastern Kowloon, as opposed to Quarry Bay, Hong Kong Island. This may be due to the geographic and transportation proximity to eastern Hong Kong Island, which is connected by the Eastern Harbour Crossing, a combined road-rail tunnel crossing beneath Victoria Harbour.

Concerning the trade and logistics sector, Figure 8 (C) suggests agglomeration patterns that are largely concentrated along the western coast or near the border with Shenzhen in mainland

China. The coastal areas of Kwai Chung and Tsing Yi show the largest agglomerations in Hong Kong that are separated by the Rambler Channel. Kwai Chung is the major hub of the commercial cargo handling zone with the Kwai Chung Container Terminal that is one of the largest and busiest port facilities in Asia-Pacific region. Tsing Yi serves as an important transportation center including some dockyards and shipbuilding industries. One unforeseen discovery is that Tuen Mun unveils a strong agglomeration in the trade and logistic sector. The possible inference can be that Tuen Mun owns the only container terminal in Hong Kong for river trade cargoes, explicitly, River Trade Terminal has been a crucial transportation interchange in recent years to tackle the increasing volume of shipments between Hong Kong and ports in cities of the Pearl River Delta (PRD). Besides coastal regions, hinterland areas in New Territories are hotspots for agglomeration economies in the trade and logistics sector. New towns in the northern parts of New Territories, for instance, Sheung Shui and Fanling, have been taken as trade hubs for parallel traders and tourists from mainland China. This is because the geographic proximity to the Shenzhen border via the Mass Transit Railway (MTR) East Railway Line and roads has made these towns become advantageous locations regarding cross-border trade activities. Interestingly, we observe a relatively large agglomeration in Tin Shui Wai that is a conventional residential region in Hong Kong, with a majority of public and private housing estates. This result could be explained by the fact that Tin Shui Wai provides more immediate proximity to Shenzhen Bay port, constituting the predominant part of Hong Kong–Shenzhen Western Corridor. Trade and logistics industries in this area therefore can have more competitive positions for road freight transportation to mainland China.

Figure 8. Sector-specific agglomeration at DCCA level, where the DCCAs with HH and HL are regarded as agglomerations. (A) Manufacturing. (B) Service. (C) Trade and logistics.

4.2.2 Association with built environment and socio-demographic characteristics

In this subsection, we broaden our study scope from the previous part and subsequently focus on the RQ2 and RQ3. Specifically, we assess characteristics from the local context using OLS and GWR models. Table 3 reports the summary statistics of coefficient estimates retrieved from regression results.

OLS regression results

As suggested in Table 5, results from the global model manifest holistic results that each sector indicates a distinct association with relevant explanatory variables. The adjusted R^2 are 0.39 for manufacturing sector, 0.69 for service sector, and 0.42 for trade and logistics sector, revealing that local contexts can explain a substantial portion of the variance regarding sectoral agglomeration economies.

In terms of the manufacturing sector, population density, number of intersections, number of bus stops, distance to Central, mean income, and composition of white people are significantly associated with the agglomeration activities. Taking distance to Central as an illustration, He et al. (2020) suggested that the manufacturing industries have been relocated to new towns due to overcrowding and housing issues in Hong Kong Island and Kowloon. For the service sector, five explanatory variables are closely associated, of which four of them belong to socio-demographics. Among these factors, one unexpected finding is that the race composition of white people is positively correlated to the concentration of service economies. We speculate that this may be due to the relatively higher proportion of white expats working for financial, business, and other professional services. Regarding agglomeration economies in the trade and logistics sector, two built environment and three socio-demographic factors are significantly correlated. Particularly, distance to the border has negative associations with LQ, meaning that a decrease in the distance to the border with Shenzhen, mainland China can attract more trade and logistics activities. This finding ties well with previous implications wherein areas with trade and logistics agglomeration can have higher accessibility to mainland China. Having responded to RQ2, results from the global regression model can be helpful for the identification of the examined variables that are closely related to agglomeration performances in different sectors.

GWR results

Global regression analysis may neglect spatial disparities as to the associations across the study area. In reply to RQ3, we present and discuss these relationships between agglomeration and built environment and socio-demographic characteristics based on the results of GWR model. The adjusted R^2 and Akaike Information Criterion (AIC) of the GWR models shown in Table 3 are obviously higher than and lower than those of the OLS models, respectively. Such improvement provides us a clear connotation that GWR can have a better explanatory capacity than the OLS model.

The spatial distribution in the manufacturing sector presents relatively distinctive results that patterns of max and min values are largely incomparable. For example, the positive coefficients of population density from Figure 9 (A) are mostly located in Fanling, Tai Po, and western Kowloon, while the number of intersections in Figure 9 (B) suggests stronger influences for the entire study area except Sha Tin and Ma On Shan. Correspondingly, former results have implied that manufacturing activities in Fanling and Tai Po tend to be labor-intensive industries that are positively related to population density. The latter distribution in terms of road intersections suggests that transport network connectivity is important for the growth of local manufacturing agglomeration in major parts of Hong Kong. The research finding is also reported by Giuliano et al. (2012) that network accessibility of road design is significantly related to employment center growth. In contrast, significant relationship between road intersection density and the service sector is not reported in our findings. This lack of significance may be attributed to the differing locational preferences and operational requirements of service industries. The distribution of mean income coefficient from socio-demographic characteristics in Figure 9 (E) suggests a positive and considerable association with new towns in New Territories, for instance, Tai Po, Yuen Long, and Tuen Mun. That is to say, an increase of labor income in these districts would attract more manufacturing activities.

Moving to the service sector, we discover strong impacts from socio-demographic factors on Kowloon Peninsula and Hong Kong Island in comparison to the manufacturing sector. These areas have been regarded as the employment center of Hong Kong. Particularly, the coefficient distribution in Figure 10 (B) shows that the mean income has extremely formidable influences on Eastern District, including Causeway Bay, Fortress Hill, and North Point. Our findings hint that in these areas agglomeration economies in the service sector may be positively connected with the local resident's income. Further, evidence indicated in Figure 10 (C) reveals that the jobs-housing ratio is negatively correlated to the service industrial activities across eastern Kowloon and Hong Kong Island. In contrast, new towns in western New Territories are less related to this socio-demographic factor. Figure 10 (D) displays the spatial pattern of race composition for white people. Most of the local coefficients in downtown areas are significantly positive, elucidating the DCCAs with a higher proportion of white people having more concentrated service activities. This finding is not surprising given that financial, business, and other professional services are mainly located in the CBDs of Hong Kong, in which practitioners with western backgrounds can be categorized as advantageous groups.

Coefficient estimates for the trade and logistics sector reveal distinct distribution patterns, suggesting influences of associated characteristics on economies may vary depending on locations. For example, an evidently negative relationship with the number of bus stops is captured in Sheung Shui and Fanling in Figure 11 (A), displaying an exceedingly decreased economic activity with bus usage. A popular explanation is that cross-border trade activities in this region primarily depend on the MTR East Railway Line, which can directly extend to the Lok Ma Chau and Lo Wu immigration control points in connection with Shenzhen mainland China. Again, for mean income, as is shown in Figure 11 (C), estimated coefficients become negatively stronger for Kwai Chung and Tsing Yi that are the commercial cargo handling zones in Hong Kong. These results may be interpreted as indirect evidence of the attractiveness in respect to the labor cost for logistics activities.

Table 5. OLS and GWR regression results of the LQ value in sectoral agglomeration economies.

Sector	Manufacturing				Service				Trade and logistics			
	OLS	GWR			OLS	GWR			OLS	GWR		
Variable	Coef	Max	Min	Mean	Coef	Max	Min	Mean	Coef	Max	Min	Mean
Built environment												
Population density	-0.01**	0.64	-0.16	0.14	0.02	0.34	-0.36	-0.06	-0.02	0.30	-0.32	0.05
Land use diversity	0.03	0.21	-0.39	-0.02	0.05	0.45	0.11	0.05	-0.06	0.06	-0.28	-0.06
Number of intersection	0.14*	0.37	-0.78	0.01	-0.08	0.32	-0.37	-0.04	-0.02	0.35	-0.48	-0.05
Road density	0.06	0.37	-0.36	-0.07	0.04	0.31	-0.26	0.01	-0.08	0.21	-0.37	-0.02
Distance to metro	0.03	0.13	-0.43	0.15	-0.04	0.17	-1.19	-0.25	0.01	0.73	-0.23	0.24
Number of bus stop	-0.13*	0.94	-0.80	-0.01	-0.03	0.23	-0.34	-0.04	0.12*	0.42	-0.35	0.09
Distance to Central	0.26*	6.70	-1.39	1.24	0.04	3.67	-5.09	-0.43	-0.01	2.62	-2.38	0.40
Distance to border	-0.16	5.34	-2.65	0.23	0.44**	5.00	-4.48	0.59	-0.48**	0.54	-3.67	-0.73
Socio-demographics												
Population age level	0.07	0.44	-0.22	0.09	0.03	0.32	-0.25	-0.01	0.07	0.35	-0.19	0.12
Mean income	0.13*	1.38	-0.44	0.18	0.17**	0.41	0.61	0.12	0.33**	0.90	0.13	0.43
Jobs-housing ratio	-0.05	0.64	-0.54	0.01	0.10**	0.30	-0.44	-0.05	-0.14**	0.32	-0.49	-0.07
Percent White	-0.18**	3.82	-1.79	0.21	0.17**	0.87	-2.77	-0.21	-0.12	1.08	-1.08	-0.08
Percent Filipino	-0.11	0.59	-1.61	-0.11	0.41**	1.04	-0.10	0.40	-0.30**	0.17	-0.76	-0.31
Percent Indonesian	0.07	0.69	-0.38	0.14	-0.01	0.39	-0.44	-0.10	0.11	0.53	-0.16	0.15
AIC	892.66	830.99			645.97	562.01			878.29	864.34		
Adjusted R^2	0.39	0.60			0.69	0.80			0.42	0.54		

∗ Represents statistically significant at the $p < 0.05$ level.
∗∗ Represents statistically significant at the $p < 0.01$ level.

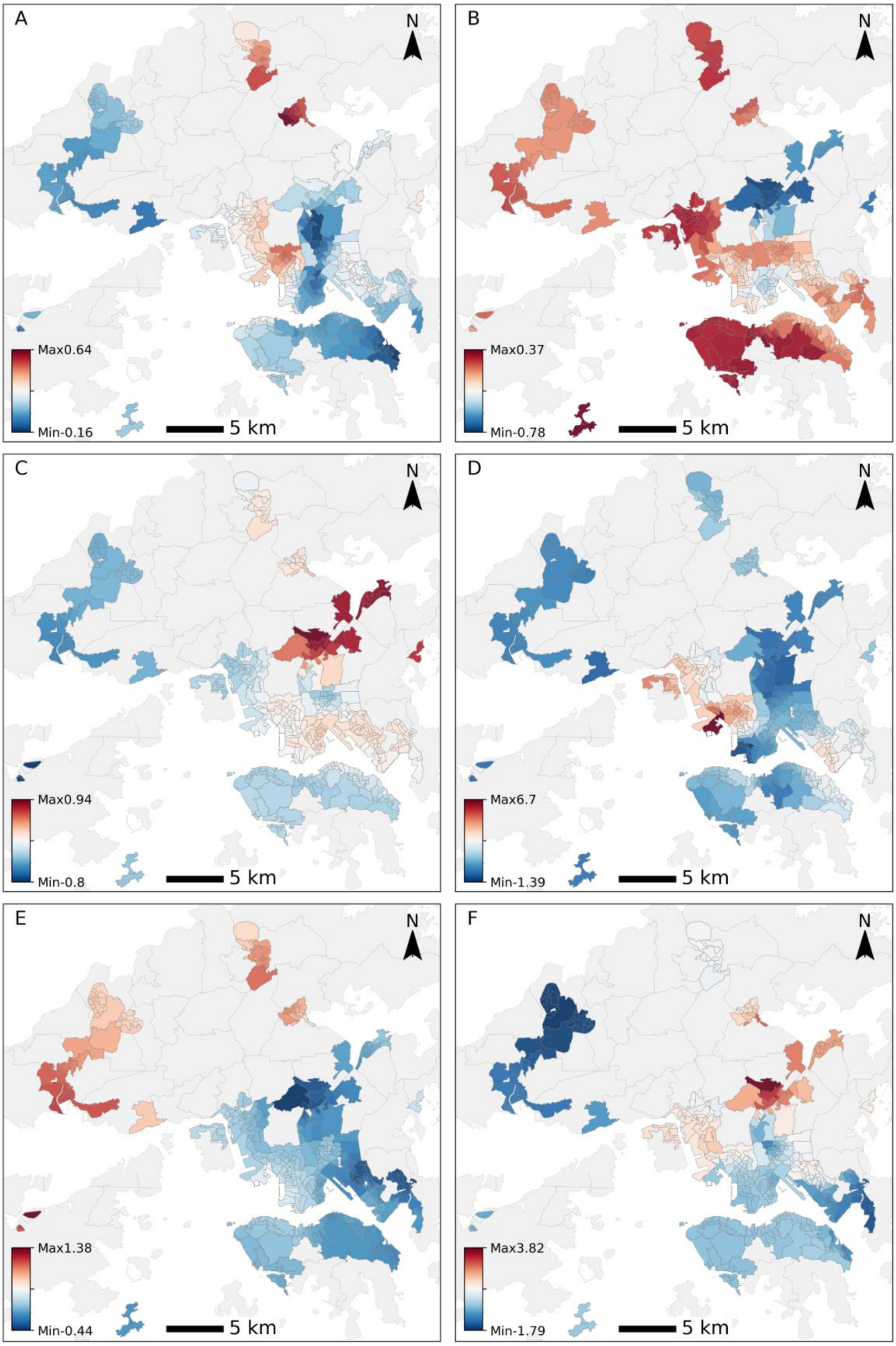


Figure 9. Spatial distribution of estimated coefficients in the manufacturing sector. (A) Population density. (B) Number of intersections. (C) Number of bus stops. (D) Distance to Central. (E) Mean income. (F) Percent White.

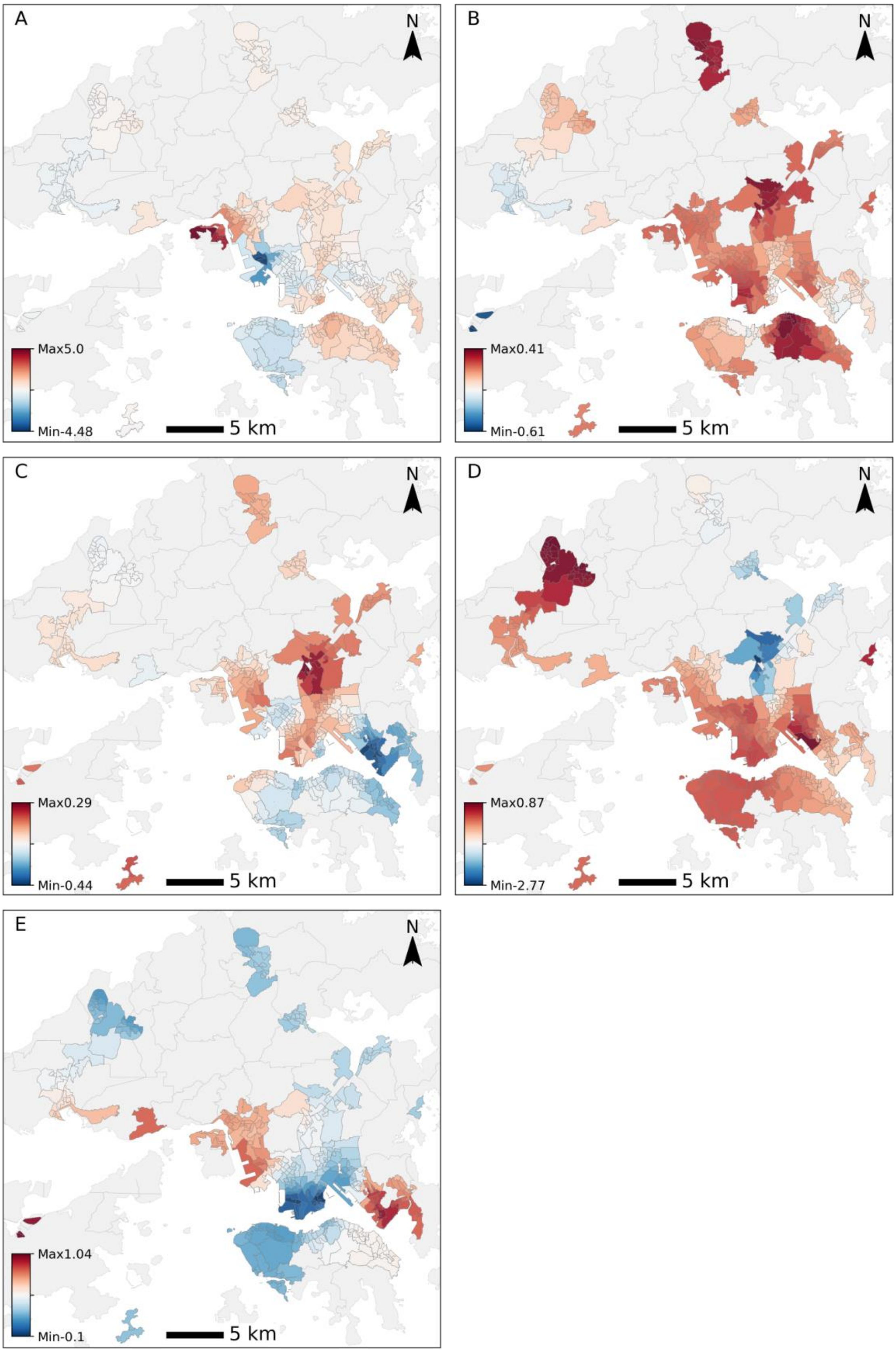


Figure 10. Spatial distribution of estimated coefficients in the service sector. (A) Distance to the border. (B) Mean income. (C) Jobs-housing ratio. (D) Percent White. (E) Percent Filipino.

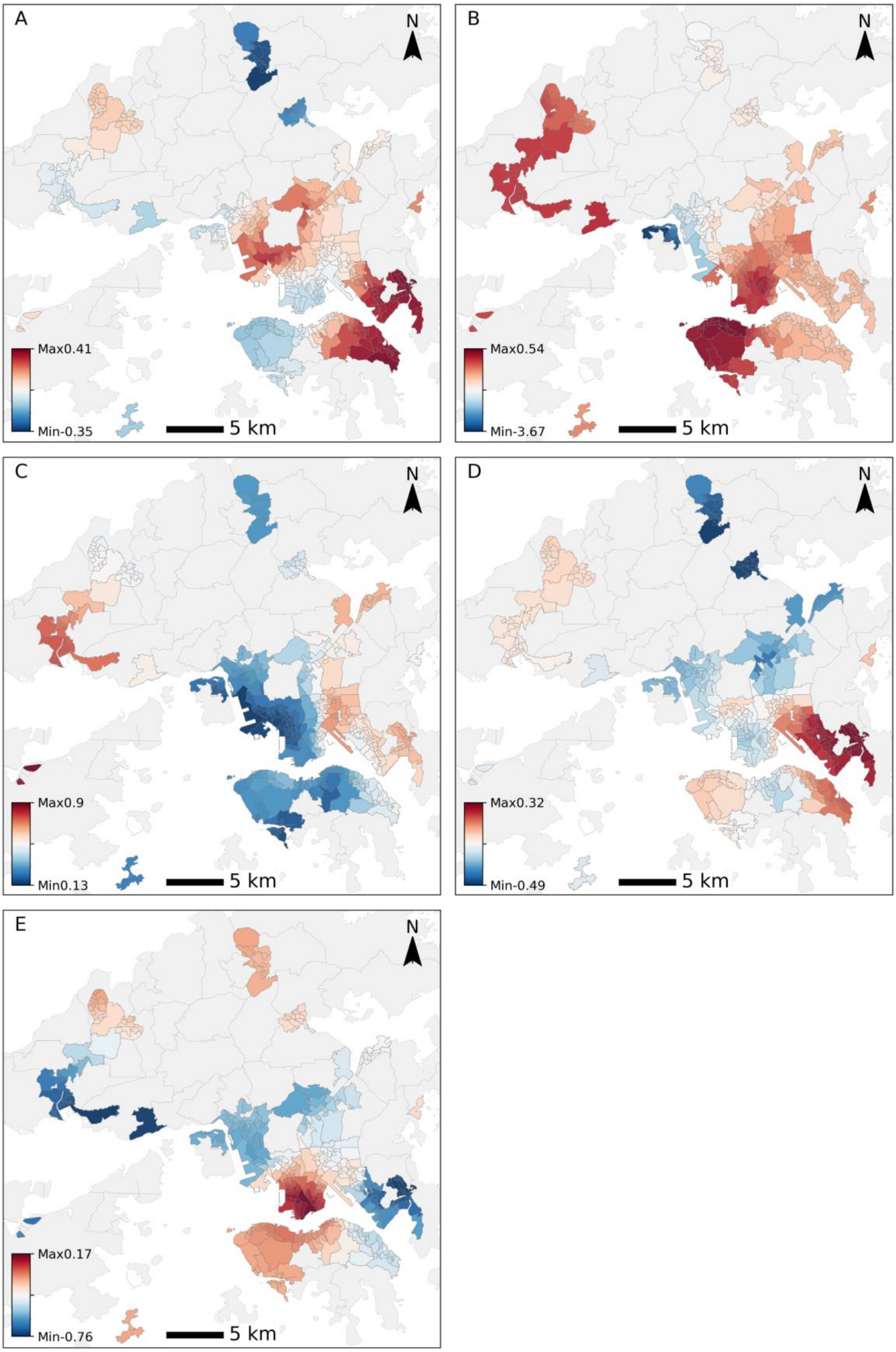


Figure 11. Spatial distribution of estimated coefficients in the trade and logistics sector. (A) Number of bus stops. (B) Distance to the border. (C) Mean income. (D) Jobs-housing ratio. (E) Percent Filipino.

4.3 Discussions

While prior studies have recognized the relationship between local economic conditions and their environments, there has been limited attention given to the spatial diversity of urban agglomeration economies and how they interact with both built environment characteristics and socio-demographic factors. Utilizing data from the 2016 population by-census in Hong Kong, our research delineates the spatial configuration of sector-specific agglomeration economies and examines their association with the importance of local spatial contexts, including attributes of the built environment and socio-demographic elements.

Our analysis employs tools such as the Location Quotient (LQ) and Local Moran's Index to identify local agglomeration economies, uncovering significant spatial variances across different economic sectors. A distinct concentration of economic activities is noticeable in particular areas of the city. For example, manufacturing sectors tend to aggregate in the New Territories, drawn by the advantages of decentralized locations that offer lower land costs. In contrast, key business and commercial centers such as Central, Causeway Bay, and Tsim Sha Tsui are identified as hubs of service agglomerations. The findings highlight the critical influence of locational advantages in determining the spatial distribution of sectoral economic activities. These advantages are influenced by a variety of factors including urban planning policies, land use regulations, the significance of downtown areas, business nature, and geographic proximity to borders, as discussed in research by He et al. (2020). For example, the clustering of trade and logistics in Kwai Chung and Tsing Yi is notably influenced by the natural benefits of the Rambler Channel, which supports the growth of container terminal facilities. In conclusion, our study not only emphasizes the spatial diversity found across different economic sectors in Hong Kong but also the significant role played by a range of locational advantages in shaping these spatial patterns.

Our study employs Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) analyses to investigate the interaction between agglomeration economies and variables from the built environment and socio-demographic domains. The OLS results demonstrate a strong correlation between sector-specific agglomeration economies and several explanatory variables. Regression coefficients suggest that variations in economic activities are explainable through elements of the built environment such as population density, number of bus stops, and proximity to Central and the border. Notably, proximity to the border is a critical factor, showing a positive association with the service sector and a negative link with the trade and logistics sector. These findings highlight the significant impact of locational advantages in determining spatial disparities across economic sectors.

Furthermore, our study uncovers significant links between various socio-demographic characteristics and agglomeration economies. Key metrics like the jobs-housing ratio, mean income, and racial composition display distinct relationships with specific districts in Hong Kong, emphasizing the nuanced interplay of these factors with local economic conditions.

Expanding on this analysis, the GWR approach offers deeper insights into spatial non-stationarity, illustrating how relationships between agglomeration economies and selected variables differ across geographic areas. The GWR model, indicated by superior adjusted R2 values and AIC scores, outperforms the OLS model in capturing local variations. For example, in manufacturing sectors, areas with higher population densities such as Fanling, Tai Po, and western Kowloon attract more activities, whereas the effect is diminished in less dense regions. Also, the relationship between mean income and manufacturing activities varies significantly; industrial hubs like Tai Po, Yuen Long, and Tuen Mun show positive correlations, whereas urban areas such as Hong Kong Island and Kowloon exhibit negative associations (Norton & Rees, 2007; White, 1999).

In the service sector, the influence of mean income differs by location, with areas like the Eastern District on Hong Kong Island benefiting from its proximity to commercial centers which attract high-income populations, reducing commuting times (Yu & Peng, 2020). Additionally, racial composition plays a critical role in service sector agglomeration, with neighborhoods having higher proportions of white and Filipino populations demonstrating a greater concentration of service activities. This suggests that diverse ethnic backgrounds may contribute unique productivity advantages, influencing local economic dynamics. These observations provide a complementary view to findings by Tao et al. (2020), who noted residential segregation among minority groups from Western countries in Hong Kong Island, thus offering a different perspective on sector-specific agglomeration employment. Overall, these patterns emphasize the importance of understanding spatial disparities among economic sectors and their impact on the formation of agglomeration economies.

Chapter 5. The Holistic View

5.1 Background and Motivation

Industrial agglomeration refers to the clustering of economic activities within urban areas, a phenomenon influenced by dynamic shifts in global industrialization and regional urban development. Despite the ongoing transformations in the geographic and functional structures of industries, there is limited research on the spatial-functional organizations of sectoral industries in urbanized megaregions. To address this research void, this investigation focuses on the Greater Bay Area (GBA) as its study domain and aims to respond to the following two research inquiries:

1) What are the spatial and functional characteristics of industrial agglomerations within the economic sectors of the GBA?
2) Considering the identified industrial agglomerations in the GBA, how are the economies of diversification and specialization correlated with their geographic arrangements?

This study aims to elucidate the role of industrial agglomeration by examining the connection between the spatial concentration of industrial activities and their functional attributes in the context of a megaregion. Utilizing points of interest (POIs) data from the GBA, we introduce a data-driven framework to identify and characterize industrial agglomerations. Initially, a kernel density function assesses the spatial extent of these agglomerations, revealing spatial heterogeneity across various economic sectors independent of administrative boundaries. Subsequently, we delve into industrial functions by extracting thematic characteristics through semantic-based information retrieval, utilizing words tokenized from POI registered names. Focusing on manufacturing industries as a case study, we explore the differences between specialization and diversification effects in terms of their local geographic patterns.

5.2 Main Findings

5.2.1 Spatial organization of industrial agglomerations

In this section, we present and discuss the spatial extents of industrial agglomerations for each sector. A comparative analysis of various economic sectors in Figure 12 reveals substantial spatial heterogeneity, unveiling distinct patterns of sectoral industrial concentration. For instance, in the primary sector (Figure 12 (a)), two prominent centers emerge in the western GBA. Surprisingly, a smaller agglomeration is identified in northern Hong Kong, primarily comprising agricultural activities situated in the rural areas of New Territories.

Examining the secondary sector, dominated by manufacturing industries, Figure 12 (b) showcases agglomerations distributed across the central and south-eastern GBA, reflecting a dual-center structure. Notably, the largest agglomeration, spanning the administrative boundary between Guangzhou and Foshan, implies cross-regional industrial collaboration. The second center in Shenzhen exhibits evenly scattered industrial concentrated areas, excluding the southeast coast, a renowned tourist attraction. Limited concentration areas are observed in the hinterland areas of the Pearl River estuary.

In contrast, the tertiary sector displays a distinct pattern, with agglomerations predominantly concentrated in traditional downtown areas surrounding the administrative and economic urban centers of cities in the GBA. Guangzhou, Shenzhen, and Hong Kong house multiple tertiary agglomerations, playing crucial roles in the sector's prosperity and reinforcing the polycentric structure of these megacities.

Figure 12 (d) also illustrates the distribution of quaternary sector agglomerations, notably dense in Shenzhen. This observation can be attributed to a significant workforce from innovative fields and investments attracted by supportive policies from the Shenzhen municipal government. Conversely, other areas, especially the hinterland cities of the GBA, exhibit a notably low presence of quaternary economics. This distribution suggests a spatial clustering of high-tech and knowledge-oriented industries around central cities, distinct from the secondary industrial agglomerations. In summary, these findings affirm distinct geographic configurations among sectoral industries.

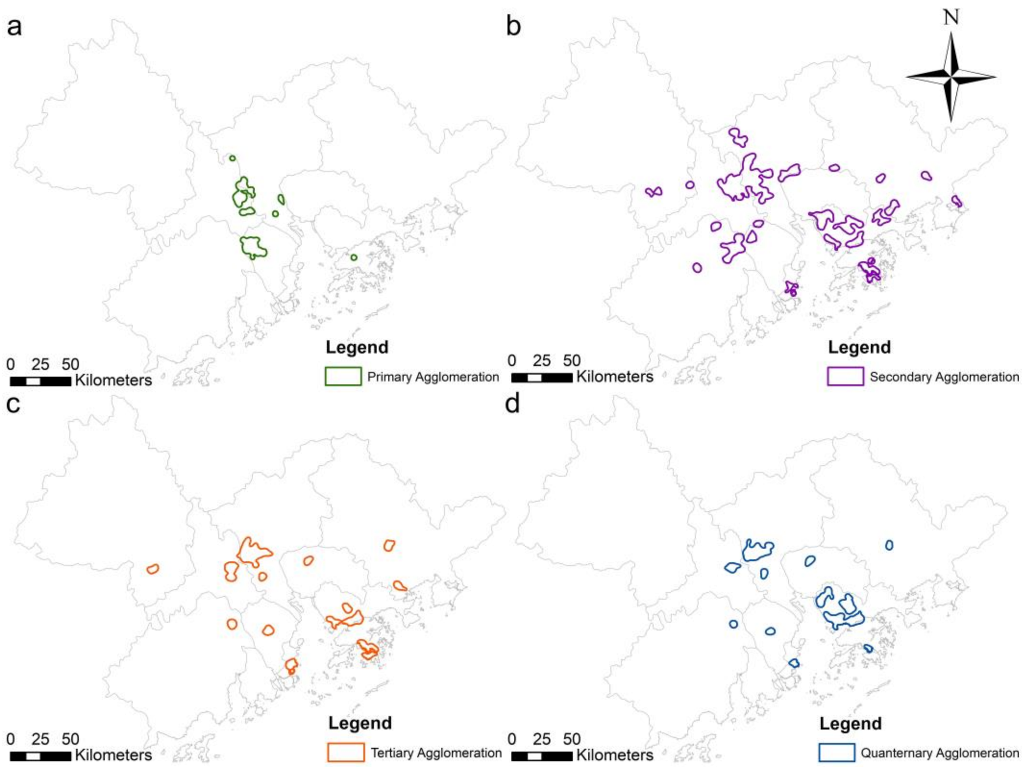


Figure 12. The spatial distribution of different industrial agglomerations in the GBA. (a) Primary sector. (b) Secondary sector. (c) Tertiary sector. (d) Quaternary sector.

5.2.2 Functional characteristics of identified agglomerations

This section elucidates the functional characteristics derived from identified agglomerations for each economic sector. Across specific sectors, diverse industrial functions within various agglomerations suggest the establishment of relatively integrated industrial and supply chains in the GBA for a range of products and services.

In Figure 12, four agglomerations exhibit distinct semantic features related to specialized agriculture productions and services. Figure 13 displays the functional characteristics exhibited by agglomerations in the primary sector. Agglomeration 1 and 2, situated in Zhongshan and spanning the Guangzhou-Foshan boundary, respectively, demonstrate a functional emphasis on horticulture. Horticulture, a subcategory of farming, caters to the significant demand for decorative gardening products, particularly in densely populated urban areas (Pribadi & Pauleit, 2015). Notably, agglomeration 2 offers more nuanced and exclusive horticultural products compared to the general offerings of agglomeration 1. For example, agglomeration 2 features Pachira Aquatica, a tropical tree considered a "money tree" in Cantonese culture symbolizing good luck and commonly used as an indoor decorative plant in Southern China.

Conversely, agglomeration 4 in northern Hong Kong showcases distinctive characteristics, potentially providing high-quality organic foods and a local tourism experience, as evidenced by specific terms such as "experience" and "organic." Additionally, agglomeration 3 appears to be a general agriculture region with traditional farming productions. The outcomes related to the primary sector indicate significant variations in the functional characteristics of agglomerations, highlighting diverse offerings in terms of products and services.

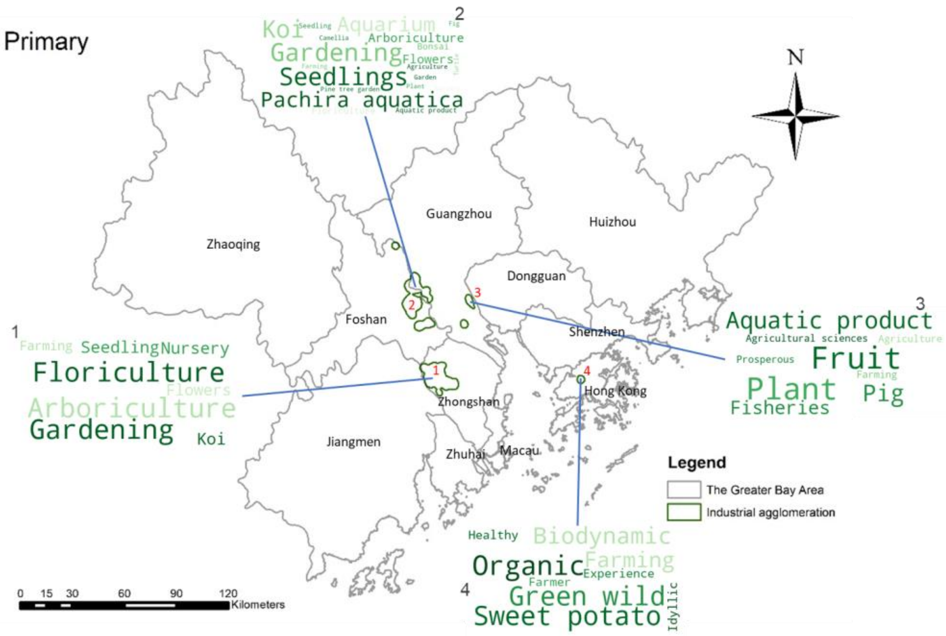


Figure 13. Functional characteristics shown by agglomerations from the primary sector.

Shifting focus to the secondary sector, Figure 14 reveals a pronounced presence of specialization effects in the peripheral GBA. These word clouds illustrate manufacturing activities aligned with specific industrial chain systems, indicating distinct preferences for specialized production. Notably, peripheral agglomerations 1 and 5 highlight the production of bathroom accessories and footwear products, respectively, primarily associated with low-tech industries.

In contrast, central GBA agglomerations (6, 7, and 8) present diversified production structures, showcasing keywords related to largely unrelated industrial functions. This pattern provides robust evidence for the existence of diversification economies. For instance, agglomeration 8 in Hong Kong features simultaneous displays of different industrial productions such as food, plastic, offset printing, and toys within a shared industrial space. Collectively, the findings from the secondary sector indicate the coexistence of specialized and diversified economies in the GBA, with close associations with their respective geographic locations.

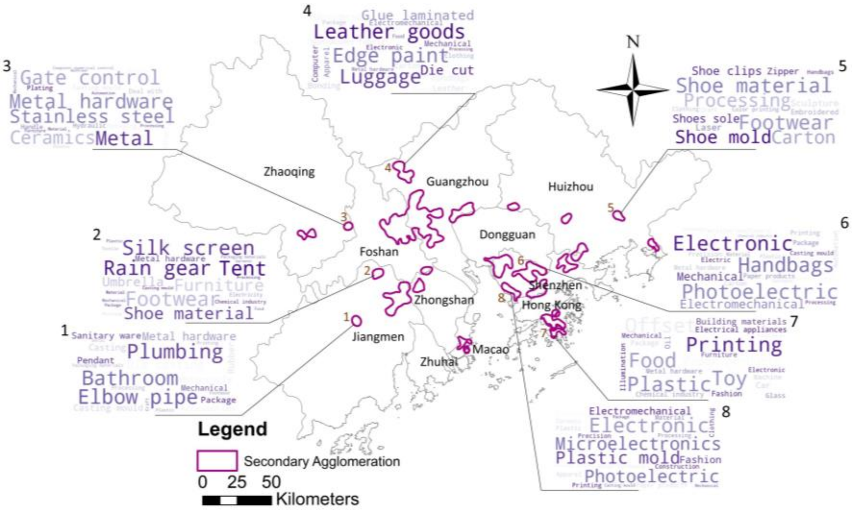


Figure 14. Functional characteristics shown by agglomerations from the secondary sector.

Figure 15 provides insights into the functional characteristics of the tertiary sector, primarily composed of business and commercial services. A notable discovery is the remarkable abundance of functional characteristics in Hong Kong, a renowned global financial and commercial hub. Keywords from agglomeration 5 predominantly align with financial and commercial services.

Distinct semantic variations emerge between agglomeration 2 in Guangzhou and agglomeration 4 in Shenzhen, two prominent megacities in the region. The former exhibits a pronounced focus on the storage and exhibition of products, exemplified by events like the annual Canton Fair held in Guangzhou. Conversely, our results indicate that Shenzhen places greater emphasis on international trade and E-commerce, leveraging its strategic location bordering Hong Kong to the south.

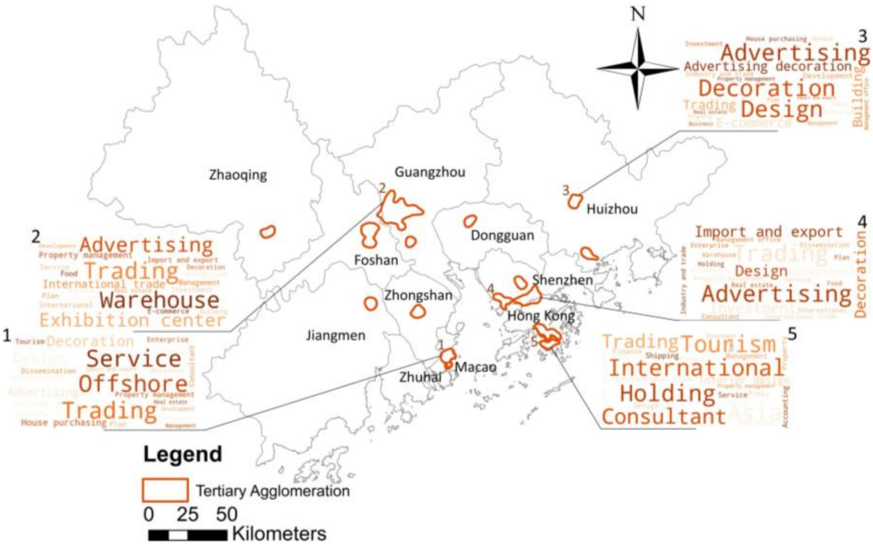


Figure 15. Functional characteristics shown by agglomerations from the tertiary sector.

Figure 16 illustrates and discusses the findings of the quaternary sector. While agglomerations generally share similar semantic implications with a distinct focus on high-tech manufacturing processes, there are discernible variations. Shenzhen stands out for its prominent role in innovative industries, specifically showcasing functional characteristics related to information and communications technology (ICT). This distinction can be attributed to a higher concentration of localized firms specializing in diverse ICT services and digital products.

Additionally, other cities exhibit unique functional trends in high-tech industries. Agglomerations 3 and 4 in Guangzhou and Huizhou, respectively, offer distinctive high-tech content, including pharmaceutical and aerospace-related industries. The semantic content derived from the quaternary sector implies a strong connection between high-tech and knowledge-oriented functions and local manufacturing productions.

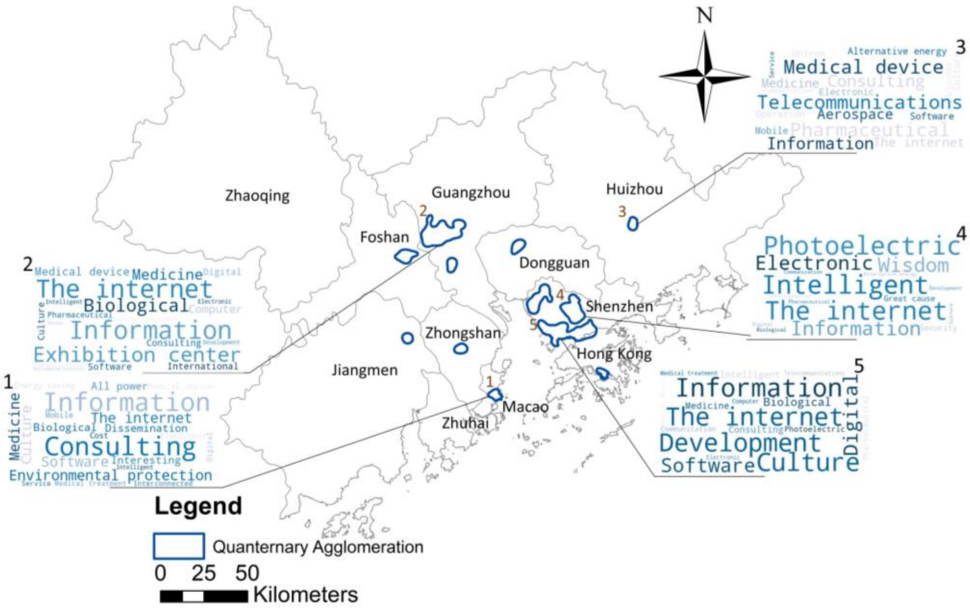


Figure 16. Functional characteristics shown by agglomerations from the quaternary sector.

5.2.3 Spatial-functional patterns of secondary agglomeration economies

In this section, we delve deeper to gain a comprehensive understanding of the discernible impacts of specialization and diversification economies on regional industrial performance within the GBA. Given the prolonged implementation of manufacturing-oriented policies in this region, we conduct a more detailed exploration of agglomeration economies based on the secondary sector to uncover spatial variations in different agglomeration economies.

Figure 17 shows the functional characteristics derived from industrial agglomerations that host specialized manufacturing activities. Notably, these productions are predominantly linked to low-tech industries, such as footwear, leather, clothing, and bathroom hardware. The semantic results reveal that, for each agglomeration, the majority of keywords are closely associated with specific manufacturing productions. For instance, leather production takes precedence in agglomeration 4 located in northern Guangzhou. Simultaneously, we observe that a significant portion of the remaining keywords pertains to leather byproducts, including luggage, handbag, and leather edge painting within the same agglomeration. It is noteworthy that certain agglomerations specialize in overlapping industries, as seen in the case of footwear industries in agglomerations 2, 8, and 9.

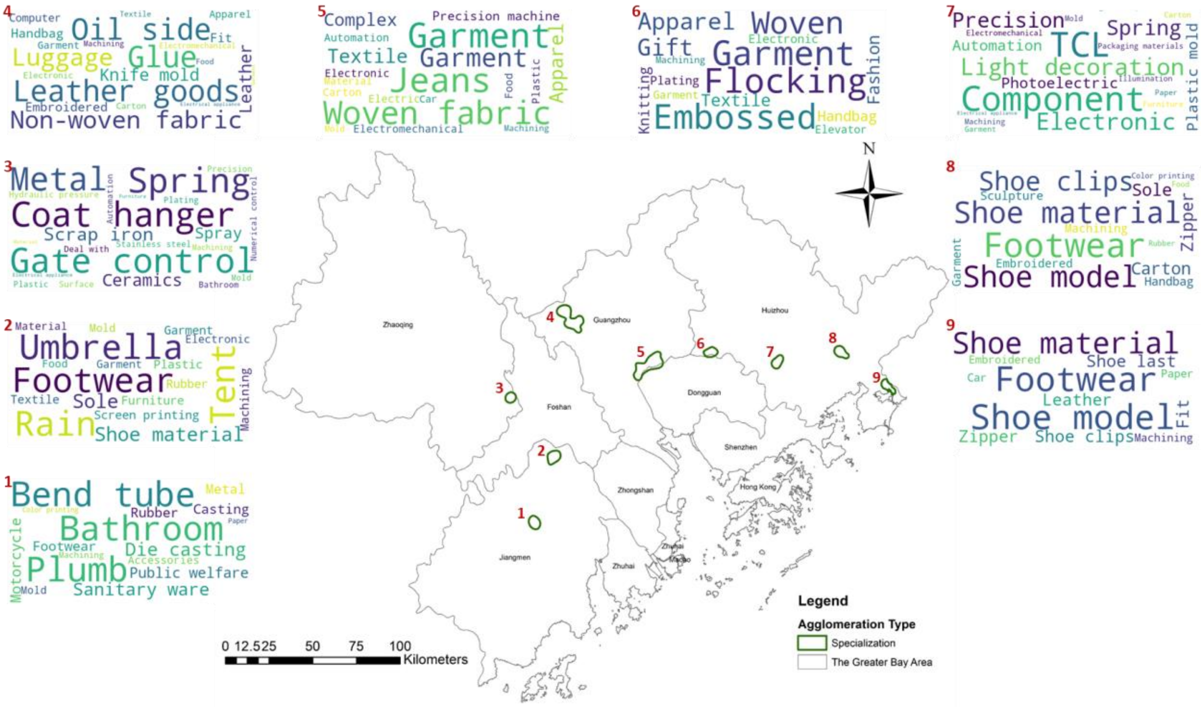


Figure 17. Functional characteristics shown by specialization agglomerations from the manufacturing sector in the GBA.

Conversely, Figure 18 illustrates the semantic functions across all agglomerations classified under diversification economies. These agglomerations display various semantics related to unrelated industrial divisions, primarily comprising printing, electronics, appliances, garments, metal hardware, and material industries. In contrast to word clouds associated with specialization economies, those representing diversification agglomerations present a more extensive and diverse array of industrial functions. These findings further bolster the idea that diversified effects inherently contribute to a broader supply of goods and services, fostering a more robust interplay across different industrial sectors (Capello, 2015).

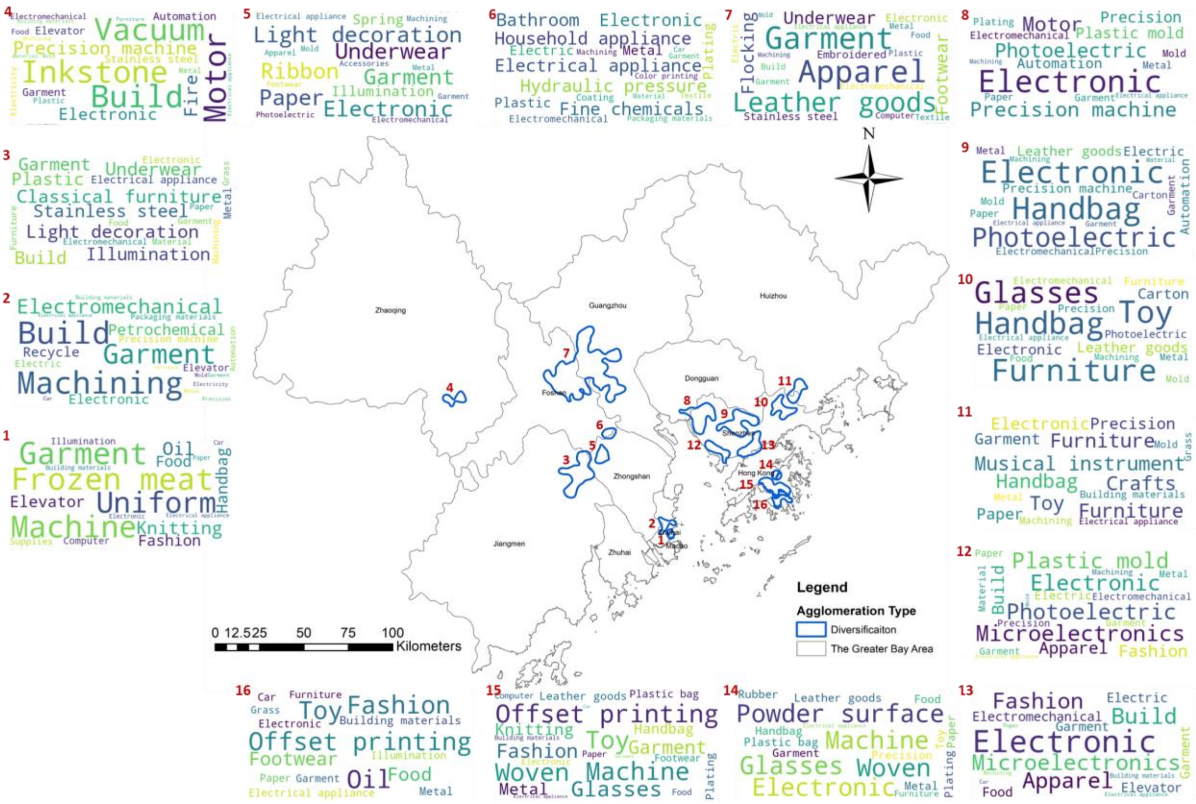


Figure 18. Functional characteristics shown by diversification agglomerations from the manufacturing sector in the GBA.

It is crucial to highlight the discernible local spatial patterns observed in various agglomeration economies. The geospatial distribution of specialization and diversification economies reveals a distinct core-periphery pattern, illustrated in Figure 19 (a). Specifically, specialization agglomerations are predominantly situated in the hinterland areas of the peripheral GBA, while diversification industries concentrate in the central cities. In Figure 19 (b), a population density map for the year 2018 at a resolution of approximately 1 km², based on census data from Worldpop, further emphasizes this distinction. Regions characterized by diversification economies exhibit high population density, contrasting with specialization agglomerations where population density is relatively lower. Nevertheless, both types of agglomeration economies are identified in urbanized regions like city centers and sub-centers, particularly considering the remaining areas with extremely low population density (Van der Panne, 2004; Strange, 2008). In summary, our empirical findings underscore the strong connection between agglomeration economies, their geographic location, and population density. This discovery significantly contributes to understanding the local spatial disparities between specialization and diversification economies.

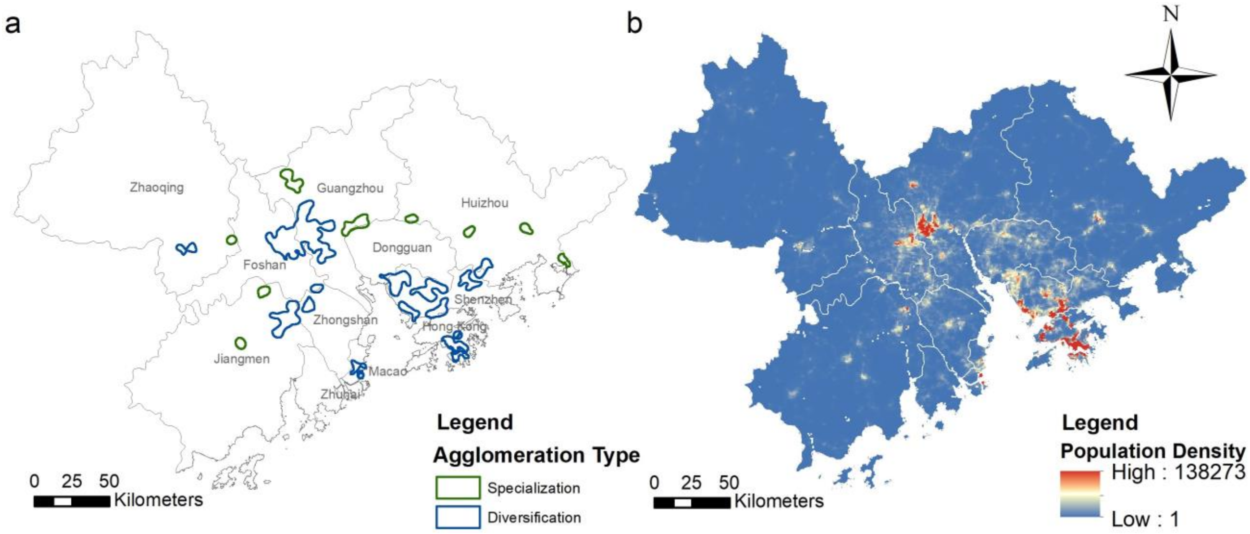


Figure 19. (a) Spatial distribution of diversification and specialization agglomerations. (b) Spatial distribution of population density in GBA.

5.3 Discussions

The measurement of industrial agglomeration has been extensively examined in prior studies. However, our understanding of the local spatial and functional patterns of sectoral industries has been limited in both theoretical and empirical explorations, largely due to challenges in accessing relevant data. This study addresses this gap by utilizing POIs data to measure sectoral industrial agglomerations in the GBA, aiming to comprehend the roles played by spatial and functional organizations.

The empirical findings on the spatial aspect indicate that different sectoral industries display notably distinct geographic configurations. Specifically, the primary sector demonstrates two large-scale concentrations located in the western GBA. In contrast, the manufacturing industries reveal a dual-center configuration in the central megacities, diverging in part from findings from Liu et al (2019) based on remote sensing images, which suggested a limited number of manufacturing spaces in the metropolitan areas of Shenzhen and Guangzhou. This inconsistency could be attributed to the limitations of remote sensing images, which may not capture details on the vertical concentration of industrial space, such as small workshops and factories operating within downtown industrial buildings.

Unlike the primary and secondary sectors, the results for the tertiary sector suggest a dispersed spatial layout surrounding the traditional administrative and economic centers in the cities of the GBA. This dual implication provides insights into the spatial organization of economic activities. Firstly, the spatial heterogeneity observed in sectoral industries can be considered a geographical proxy for economic activities, reflecting an imbalanced regional development across various economic sectors in the GBA. Secondly, the discussion of such imbalanced activities revealed by our analytical findings is warranted. External determinants, including input resources, the local workforce, land-use considerations, intercity transport, and regional policies, are to some extent linked to geographic proximity, shaping industrial distribution patterns (Faggio et al., 2017).

In addition to the spatial aspect, we present variations in functional characteristics across agglomerations based on the semantic visualization of POIs. For each specified economic sector, each agglomeration encompasses a range of distinct industrial productions and services that set it apart from others. The industrial agglomerations identified in our results can be classified into two types of agglomeration economies. Specialization economies are clearly evident as predominant, particularly in the primary sector, while diversification economies emerge prominently in industries from the tertiary and quaternary sectors. An implication drawn from these empirical findings is the establishment of relatively integrated industrial and supply chains in the GBA for diverse products and services.

More significantly, our analysis reveals the presence of both specialization and diversification economies in the functional semantics of secondary sectoral industries in the GBA. Previous research by Caragliu et al. (2016) suggested that specialization agglomerations tend to be more pronounced in less populated areas, while diversity effects have a more substantial impact in densely populated urban regions. Our case study aligns with this conclusion, indicating that specialization agglomerations are more concentrated in the peripheral GBA with lower population density. These agglomerations typically encompass a higher proportion of keywords related to low-tech specialized products or manufacturing activities. Our observation is consistent with the findings of previous work by Greunz (2004), which suggests that low-tech industries with more conventional and standardized production can derive greater benefits from Marshall specialization economies. For instance, industrial spaces labelled as garment and footwear productions likely involve interconnected manufacturers and firms, fostering commonalities and complementarities.

Conversely, diversification agglomerations are identified in the densely populated areas surrounding the central megacities of the GBA, covering a diverse range of industrial divisions. This enables the provision of various productions and services, contributing to the growth of metropolitan economies. Moreover, the industrial functions within diversification agglomerations can play a crucial role in facilitating knowledge spillovers. In this context, knowledge spillovers allow individuals in one industry to gain cross-disciplinary knowledge from practitioners in other industries (Strange, 2008). Compared to the utilization of predefined industrial categories, our analytical results surpass previous studies by illustrating that semantics can unveil richer and more nuanced information about industrial functions.

Chapter 6. Geographic Disparity

6.1 Background and Motivation

The prevailing methods frequently utilized in past research may not adequately capture the professional diversity within megacity regions. This inadequacy arises due to the rapid urbanization and complex economic activities characterizing such regions, which may render the traditional digit-based taxonomy of professional sectors insufficient to encompass all local industries. Moreover, the use of standard economic censuses, which are typically conducted at intervals—such as every five years in the U.S. and China—can lead to outdated and potentially unreliable data on regional economic activities (Wardrop et al., 2018). Another significant limitation of conventional approaches is that they may overlook specific locational characteristics and market segments unique to distinct local industries by relying solely on diversity indices, thus failing to adequately interpret their thematic functional attributes.

In response to these challenges, our study adopts a novel methodological approach based on semantic extraction. This approach aims to bridge the existing research gaps by providing a more nuanced understanding of professional variation, particularly within the context of the Guangdong-Hong Kong-Macao Great Bay Area (GBA) in China. Through this alternative methodology, we seek to capture a more detailed and timelier picture of the economic landscape in rapidly urbanizing megacity regions. Three key questions are proposed that seek to be answered:

1) Given the industrial agglomerations identified in the GBA, what are the prototypical topics in terms of industrial functions excavated from manufacturing-related POIs?
2) Empowered by the clustering analysis based on topic importance, are there notable functional and spatial disparities exhibited by agglomerations with distinctive thematic patterns?
3) In view of the two selected subregions, what are the spatial-functional traits exhibited by local industrial agglomerations?

This study seeks to enrich the understanding of economic activities within a megacity region, with a particular focus on their spatial and functional variations. Utilizing points of interest

(POIs) within the Guangdong-Hong Kong-Macao Great Bay Area (GBA), we initially apply Latent Dirichlet Allocation (LDA) to extract semantic information concerning local manufacturing activities. This approach helps us to delineate various functional topics that serve as crucial indicators reflecting the diverse thematic characteristics of economic agglomerations.

Subsequent clustering analysis enables the identification of collective patterns of these functional topics within the GBA and across individual cities, revealing spatial regularities and functional variations of agglomerations. To further explore these dynamics, we analyze agglomerations in two specific subregions that exhibit unique operational modes by extracting high-frequency keywords associated with each. This analysis provides insights into the spatial-functional patterns and distinct locational preferences within the region, offering a detailed view of the economic landscape across different urban settings.

6.2 Main Findings

6.2.1 Functional topics from agglomerations

The semantic representation of manufacturing activities' functional topics is constructed through probabilistic word weights. Figure 20 (A) displays these topics using a word cloud where different colours distinguish nine functional themes, with the size of the font reflecting the relevance of keywords within the functional topic-to-keywords matrix. This visual arrangement facilitates an intuitive understanding of the most prevalent and specialized functions within each topic.

Within this framework, certain topics are closely associated with specific industrial functions or productions. For example, Topic 2 focuses on the furniture industry, incorporating related keywords such as "furniture," "mahogany," "couch," and "sofa." Topics 3 and 8 relate to home decoration, encompassing terms like "stainless steel," "ceramics," "bathroom," "illumination," and "metal hardware." This indicates a thematic concentration on elements integral to home furnishing and design. Conversely, Topics 5 and 7 highlight terms specific to the clothing and leather industries respectively. Topic 5 includes words such as "garment," "clothing," and "apparel," whereas Topic 7 features leather-related terms like "leather," "footwear," "handbag," "embroidered," and "edge paint." These topics suggest a focused exploration of specialized industrial activities relevant to fashion and accessories manufacturing. In contrast to these more specialized themes, other topics present a broader array of industrial interests. For instance, Topic 6 spans a diverse set of manufacturing sectors, represented by keywords related to the garment, printing, food, and toy industries. This broad scope suggests a thematic versatility, encapsulating a wide range of production activities across different industrial domains.

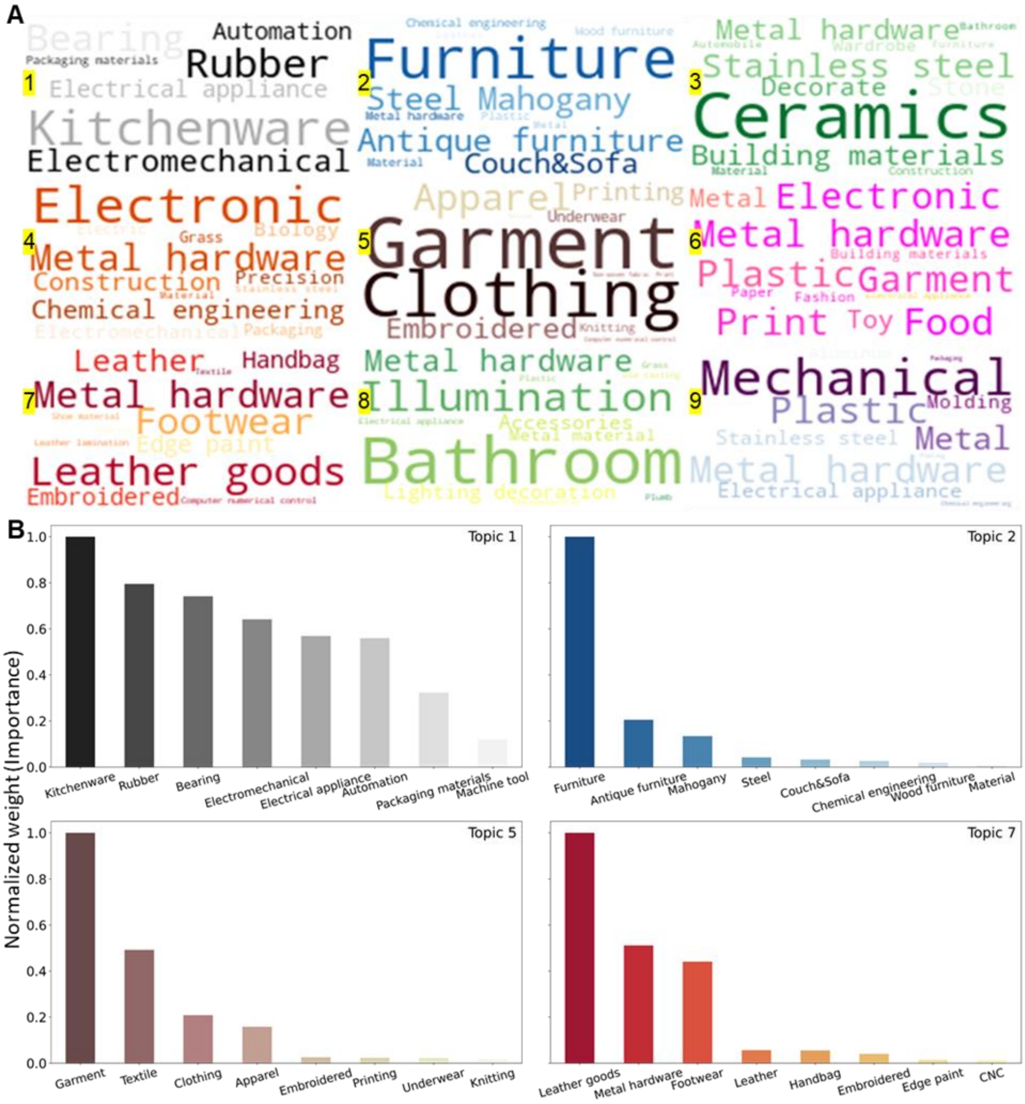


Figure 20. (A) Word cloud visualization of 9 functional topics related to manufacturing industries, each color represents key words related to an identified topic. (B) The importance weight of the top-8 ranked keywords in 4 selected topics. Note that the values of importance weights are normalized between 0 and 1.

Figure 20 (B) presents the top-8 ranked keywords across four selected functional topics, highlighting their importance weights. Notably, the height of the bars illustrates significant variations among these topics. In Topic 1, which encompasses a range of industrial activities, the keywords are relatively evenly distributed, suggesting a topic characterized by a diversified effect capable of providing a variety of manufacturing services and goods. In contrast, Topics 2, 5, and 7, which are associated with more specialized industries, show a pronounced emphasis on the initial keywords within each topic. These leading keywords typically represent the final products of specific supply chains, whereas subsequent terms are more likely connected to raw materials, components, or steps in the manufacturing process and carry lesser weights.

Table 6 compiles the top-3 ranked keywords for all topics discussed, providing a clear summary of their thematic focus. Additionally, based on the analysis of keywords within each topic, we categorize them into two types of agglomeration externalities: specialized or diversified functions. The specialized function highlights a concentration on specific industries or manufacturing processes, indicating a strong focus within particular sectors. Conversely, the diversified function encompasses a broader range of industrial productions, indicating a topic's versatility in accommodating multiple industrial domains.

Table 6. The functional topics extracted from industrial agglomerations in the GBA with functional types and top-3 ranked keywords.

Topic No	Functional Types	Top-3 Ranked Keywords
1	Diversified	Kitchenware, Rubber, Bearing
2	Specialized	Furniture, Antique furniture, Mahogany
3	Specialized	Ceramics, Stainless steel, Building materials
4	Diversified	Electronic, Metal hardware, Chemical engineering
5	Specialized	Garment, Textile, Clothing
6	Diversified	Metal hardware, Print, Plastic
7	Specialized	Leather goods, Metal hardware, Footwear
8	Specialized	Bathroom, Illumination, Metal hardware
9	Diversified	Mechanical, Metal hardware, Plastic

6.2.2 Collective patterns of aggregating functional topics

Upon aggregating industrial agglomerations based on their functional topic weights, several collective patterns emerge, as depicted in Figure 21's clustering heatmap. This heatmap categorizes all observations into seven distinct groups, each represented by different color labels. Notably, the majority of observations (over 80% within the Greater Bay Area) fall into groups 4, 5, 6, and 7, which primarily focus on topics 3, 4, and 9. Within these groups, group 6 displays a pronounced focus on topic 4, whereas group 5 exhibits interests in multiple functional topics, often related to building materials and mechanical productions. This suggests a significant engagement in these industries, contributing substantially to local economic development. In the lower section of the heatmap, 24 agglomerations from group 1 demonstrate a strong emphasis on topics 2 and 8, which relate to home decoration and furniture manufacturing, respectively. Meanwhile, groups 2 and 3 are primarily concerned with topics 5 and 6, which are linked to the clothing and leather industries. Specifically, group 2 shows a focused interest in topic 5, indicating a specialized economic role in supplying garment-related products.

Overall, the clustering analysis reveals significant functional similarities and differences among the agglomerations, highlighting a diverse range of local industrial activities. This analysis confirms not only the variation in industrial focus among groups but also underscores the complex economic landscape within the region.

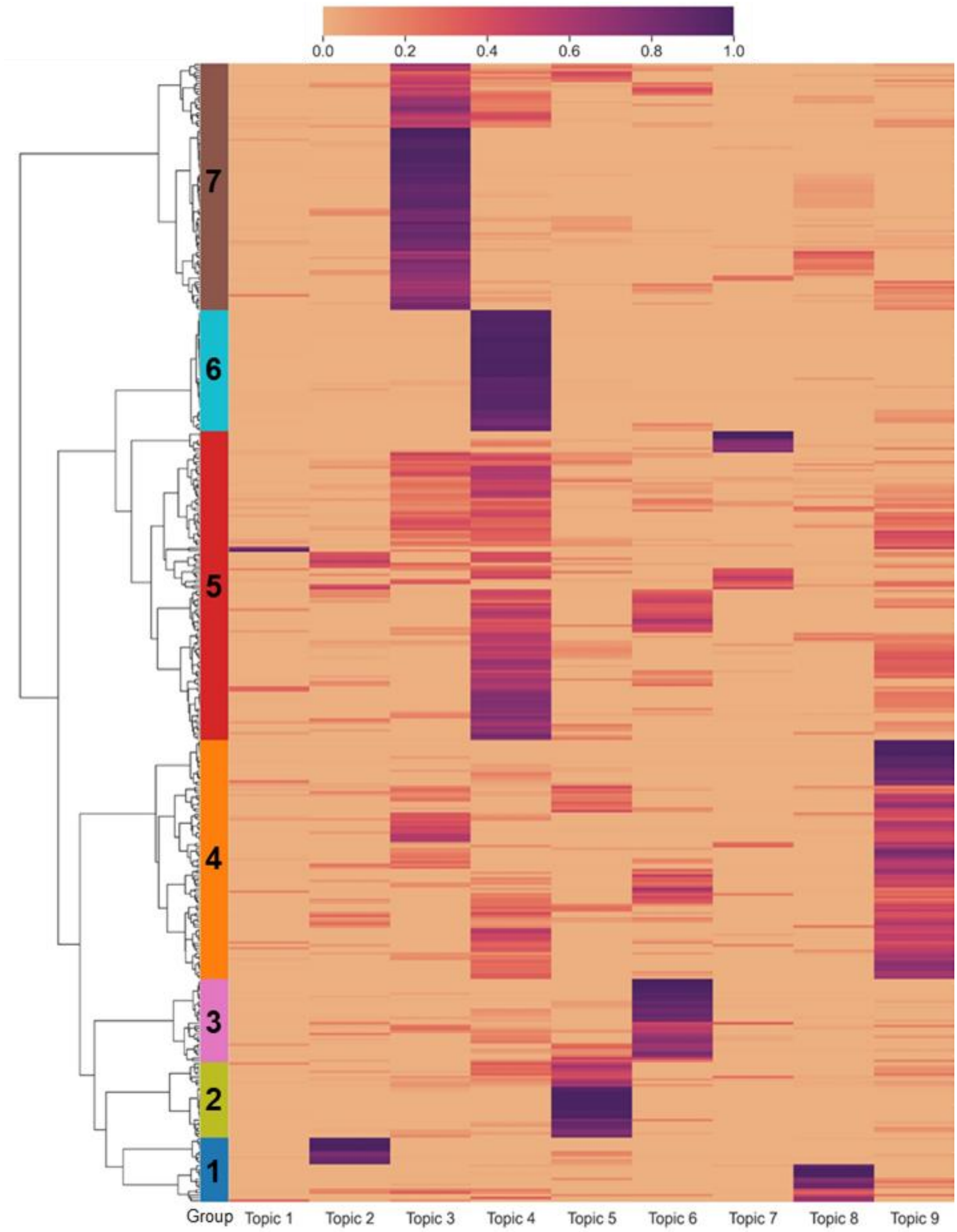


Figure 21. Clustering results of functional topics across industrial agglomerations in the GBA. Colors labelled by number in the first column represent different groups, and the remaining columns represent the normalized weights of topic importance.

Figure 22 and subsequent visualizations explore the spatial disparities among various industrial agglomerations within the Greater Bay Area (GBA), categorized according to their respective groups as indicated by the color-coded clustering heatmap. Specifically, Figure 22 (A) identifies two primary clusters in the central and southeastern parts of the GBA, indicative of a dual-center structure in these regions. In Figure 22 (B), a notable spatial cluster straddling the border between Guangzhou and Foshan highlights a cross-regional industrial collaboration primarily involving agglomerations from groups 2, 4, and 5. This cluster predominantly supports local economic activities in garment and mechanical production, sectors for which Guangzhou and Foshan are traditional hubs within southern China, known for their apparel, machinery equipment, and household appliances markets.

Figure 22 (C) illustrates a second major cluster in Shenzhen, where industrial agglomerations are more evenly distributed compared to the cluster between Guangzhou and Foshan. A significant portion of these agglomerations falls within group 6, which is closely associated with the electronics production industry. Additionally, a few agglomeration clusters are scattered across the hinterlands of the Pearl River estuary, with a majority of group 7 agglomerations—linked to building material productions—located in the peripheral areas of the GBA. Moreover, a substantial industrial cluster is observed in the southwestern GBA, particularly at the boundary between Jiangmen and Foshan. Another notable finding is the geographic proximity of group 3 agglomerations in Hong Kong, which predominantly focus on light and manufacturing industries. These spatial mappings and the clustering heatmap collectively offer a more detailed perspective on the geographical variations and proximities of industrial activities within the GBA, enhancing our understanding of the regional economic landscape and its underlying industrial structures.

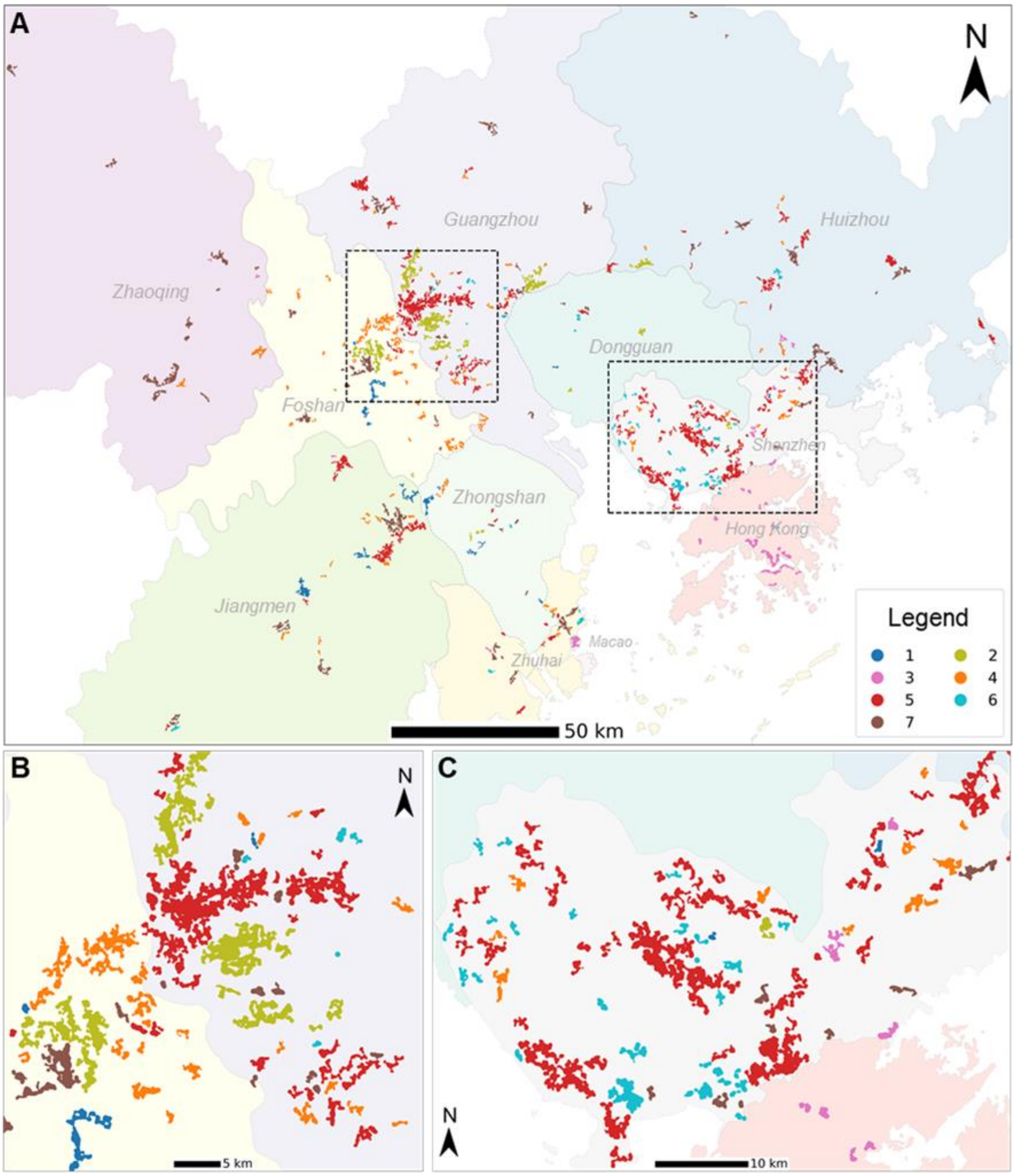


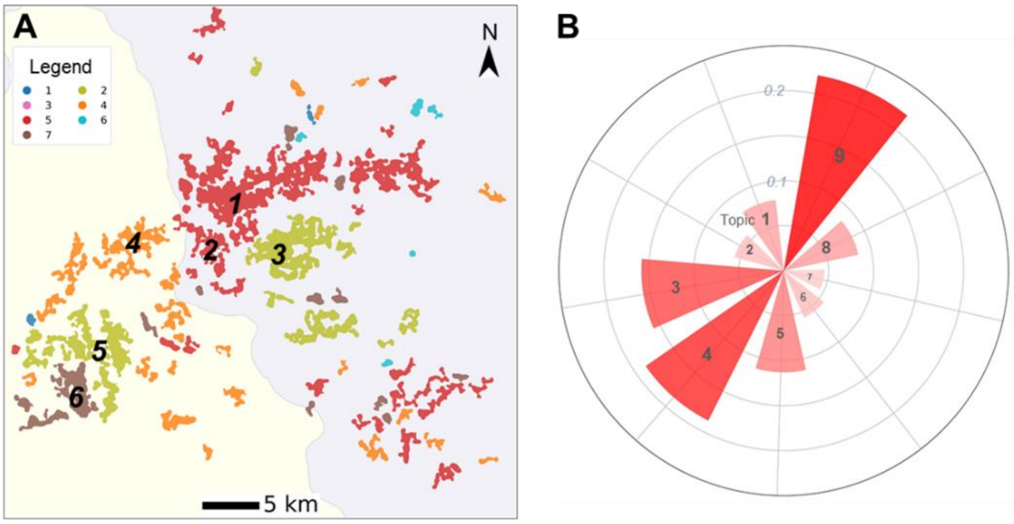
Figure 22. The spatial distribution of clustered industrial agglomerations based on their topic importance weights. Color shown in Figure 4 (i.e., clustering heatmap) are used to represent different groups. (A) subregion across the border between Guangzhou and Foshan, and (B) subregion situated in Shenzhen.

6.2.3 Spatial-functional disparities of representative subregions

This subsection analyzes spatial-functional disparities in the Greater Bay Area (GBA), focusing on two key subregions identified in prior analyses. We utilize a polar chart to display the distribution of each functional topic and select six agglomerations per subregion to highlight their specific functional keywords. The term frequency-inverse document frequency (TF-IDF) method is used to pinpoint representative keywords within these industrial agglomerations by assessing the occurrence of each term relative to the broader study area (Yu et al., 2022b).

As depicted in Figure 23, we examine the spatial-functional dynamics along the border between Foshan and Guangzhou. The polar plot in Figure 23 (B) reveals that topics 9, 4, and 3 are predominant, significantly influencing local manufacturing outputs. Specifically, topic 9 is strongly associated with mechanical equipment production, whereas topics 4 and 3 relate primarily to the electronics and building materials sectors, respectively. This distribution indicates a diverse range of manufacturing activities within the area. Further analysis is conducted to identify the top three ranked keywords within these topics from the selected agglomeration samples, summarized in Figure 23 (C). Notably, keywords related to the garment industry, such as "footwear," "apparel," and "textile," suggest a focus on low-tech manufacturing processes. This observation is consistent with the prominence of topic 5 as identified through the LDA method. However, it is noted that certain significant topics, such as topic 4 involving "Electronic," "Metal hardware," and "Chemical engineering," are not adequately captured by the TF-IDF keywords in agglomeration samples 1 and 2. This discrepancy may occur because the TF-IDF method prioritizes keywords that are most descriptive and unique to the given agglomerations, indicating that apparel-related manufacturing activities are more distinct in this subregion compared to others.

This analysis confirms that the region centered around Guangzhou is a traditional hub for fabric trading and garment production, predominantly featuring medium- and small-scale enterprises clustered in urban villages. These enterprises are closely linked to rural migrants from northern provinces, contributing to the local economic landscape (Liu et al., 2015).



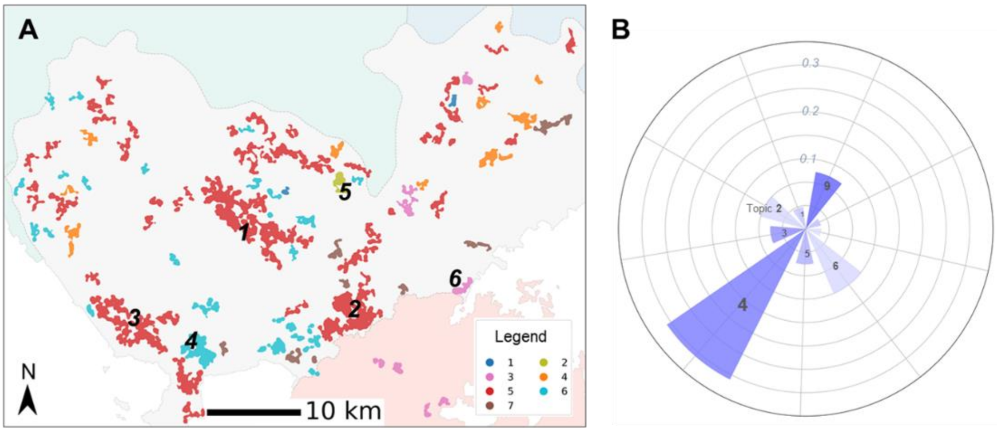
C

	Topic	N_{POIs}	TF-IDF keywords (Occurrence)
1	4,7	6620	Footwear(129), Computer(62), Biology(155)
2	2,3,4	843	Tea(23), Apparel(24), Medical device(15)
3	5	4901	Textile(927), Garment(1040), Embroidery(196)
4	1,5,8,9	1350	Textile(367), Undergarment(1040), Aluminium(196)
5	1,5	3192	Garment(1076), Textile(115), Apparel(99)
6	3,8	1316	Ceramic(459), Bathroom(338), Wardrobe(26)

Figure 23. The spatial-functional details of the subregion across the border between Guangzhou and Foshan. (A) Spatial distribution of clustered industrial agglomerations (B) Polar plot for representing the proportion of significant functional topics (C) Significant topic, number of POIs, and top-3 ranked TF-IDF keywords with their occurrence of six selected agglomerations.

In contrast, the second subregion, primarily located in Shenzhen, exhibits distinct spatial-functional patterns compared to the cross-regional industrial collaboration observed between Guangzhou and Foshan. As illustrated in Figure 24 (A), industrial agglomerations in Shenzhen are more uniformly distributed across the city, with the notable exception of the southeast coastal area, which is a well-known tourist destination featuring beaches. Figure 24 (B) reveals a pronounced concentration of activities within a specific functional topic, demonstrating a particularly strong influence from topic 4, which is associated with electronics production. This pattern underscores the robust local economic activities centered around the advanced electronics sector, supported by Shenzhen's well-developed electronic markets and manufacturing capabilities (Fu et al., 2012). Detailed examination of selected industrial agglomerations in Shenzhen shows that the majority highlight thematic keywords related to electronics production, such as "electronic" and "microelectronics," occurring with considerable frequency. This indicates that electronics manufacturing is a pivotal industry within these agglomerations, significantly contributing to the local economy.

Interestingly, one agglomeration located near the Hong Kong-Shenzhen border, classified into group 3, diverges from the dominant electronics theme. The TF-IDF keywords for this agglomeration suggest a focus on accessory and apparel industries instead. This notable difference may be influenced by its geographic proximity to Hong Kong, where many agglomerations are similarly categorized into group 3. This proximity likely fosters a different industrial focus, reflecting the unique economic interactions and market dynamics at the border.



C

	Topic	N_{POIs}	TF-IDF keywords (Occurrence)
1	3,4	3582	Electronic(656), Stainless Steel(179), Aluminium alloy(32)
2	3,4	3821	Electronic(780), Digit(45), Biology(48)
3	1,3,4	1707	Electronic(335), Plastic Moulding(15), Building materials(69)
4	4	1177	Microelectronics(35), Cyber(11), Intelligent(22)
5	5,6,7	290	Textile(112), Flocking(5), Leather(5)
6	3,8	255	Accessory(12), Apparel(32), Jewellery(6)

Figure 24. The spatial-functional details of the subregion situated in Shenzhen. (A) Spatial distribution of clustered industrial agglomerations (B) Polar plot for representing the proportion of significant functional topics (C) Significant topic, number of POIs, and top-3 ranked TF-IDF keywords with their occurrence of six selected agglomerations.

6.3 Discussions

While prior research has shed light on the diversity of geographical aggregation, it has generally not addressed the spatial-functional variations within agglomeration economies, especially in megacity regions. To explore these dynamics, our study employs Latent Dirichlet Allocation (LDA) topic modeling, which has helped to uncover a total of nine distinct functional topics that characterize various aspects of local manufacturing activities. For example, topic 2 focuses specifically on furniture-making, whereas topics 3 and 8 relate to home decoration and construction, indicating areas of specialized production within the manufacturing landscape.

Our findings also underscore the nuanced specialization evident through the keywords identified in topic modeling. These keywords offer a more detailed description of local industries than broader categories typically found in national economic activity classifications, such as China's industrial classification system. For instance, terms like "edge painting" provide a specific insight into processes within the leather industry. The distribution of keyword importance within these specialized topics often shows a concentration effect around primary terms, which usually represent the end products of a supply chain. This aligns with existing literature that notes the benefits of input sharing within specialized agglomeration economies, potentially leading to centralized procurement practices among end-product manufacturers (Rosenthal & Strange, 2001). Conversely, topics that do not focus on specific industrial specializations tend to show a more gradual decrease in keyword importance, covering a broad range of sectors, particularly those in the mid- and high-tech areas. This breadth suggests the potential for greater knowledge and skill transfer across industries, enhancing the innovative capacity within the region (Jacobs, 1969; Van Soest et al., 2002).

This research introduces a novel, data-driven framework that utilizes geospatial data and analytical techniques to pinpoint significant industrial functions, providing a deeper understanding of industrial dynamics (Yu & Liu, 2021). Our approach, which diverges from traditional methods dependent on predefined sector classifications in economic censuses, represents a more dynamic method akin to an augmentative social sensing perspective. This methodology not only captures the current industrial landscape more accurately but also offers insights that are crucial for policy and strategic planning in megacity regions.

Our analysis now shifts to examining the similarities and differences among agglomerations using hierarchical clustering to address Research Question 2 (RQ 2). The clustering heatmap provides a visual representation and clarification of functional disparities among the agglomerations. Predominantly, agglomerations are grouped into categories 4, 5, 6, and 7, which share interests in topics 3, 4, and 9. These topics correspond to the primary industrial functions in the Greater Bay Area (GBA)—electronic manufacturing, building materials, and mechanical equipment—highlighting their crucial roles in regional economic development. This finding aligns with Meyer et al. (2012), who highlighted the significant presence of the electronics industry in the Pearl River Delta (PRD) over recent decades. Spatial diversity is evident in the geographical distribution of industrial agglomerations, with each showcasing distinct functional traits. Notably, the central and southeastern areas of the GBA display a dual-center structure, characterized by two prominent spatial clusters encompassing various clustering groups. Additionally, agglomerations in group 3, predominantly located in Hong Kong and focused on light industry and manufacturing, are notably proximate. This concentration is likely driven by the local demand for fast-moving consumer goods, such as food and beverages, as suggested by He et al. (2020). The absence of certain industries like leather or metal hardware in Hong Kong can be attributed to several factors including industrial relocation to the PRD, constrained by land use limitations in Hong Kong. This shift has been further facilitated by mainland China's liberal economic policies and the broader forces of globalization since the 1980s (Hayter & Han, 1998; Ng, 2008). This migration reflects broader economic and policy dynamics influencing industrial patterns in the region.

Within the Greater Bay Area (GBA), our investigation focuses on the dual-center industrial structure, specifically analyzing spatial-functional disparities in two key subregions: the area between Foshan and Guangzhou, and Shenzhen. In the former, our findings indicate a multi-topic dominance in local manufacturing activities, particularly involving mechanical equipment, electronics, and building materials. This reflects insights from Hu and Lin (2011), who noted the vital role of mechanical and electrical industries in Guangzhou in bolstering regional manufacturing economies. Additionally, our analysis of agglomeration samples in Foshan highlights a manufacturing landscape characterized by ceramic and apparel production, as evidenced by TF-IDF keyword analysis. In contrast, Shenzhen, a rapidly evolving megacity known for its innovative industries, displays different spatial-functional trends. Our study highlights a strong focus on electronic production within its industrial agglomerations, aligning with Enright et al. (2005), who documented significant industrial clusters specializing in electronics along the east bank of the Pearl River Delta (PRD).

Despite these regional differences, a commonality emerges in the prominence of the electronics sector across the GBA. This sector stands out as a pivotal element of the regional economy, with industries gaining competitive advantages through geographical clustering within agglomeration economies. This convergence suggests a significant strategic emphasis on electronics manufacturing throughout the GBA, underscoring its central role in the region's industrial fabric.

The incorporation of geospatial technologies into urban settings has significantly enhanced the ability to explore, understand, operationalize, and visually communicate extensive datasets across a broad spectrum of stakeholders, from government bodies to the general public. The proliferation of geospatial tools in urban areas is designed to bolster the capabilities of policymakers and other relevant parties in handling and showcasing the growing amounts of geospatial data. This research contributes to the evolving methodological conversation on leveraging modern geospatial analytics to grasp regional economic dynamics.

Employing up-to-date geospatial data, particularly Points of Interest (POIs), this study showcases the capacity to identify a variety of industrial activities that traditional sectoral classifications might overlook. This methodology is not only effective for analyzing industrial sectors but also holds potential for exploring other economic sectors, including newly emerging service industries like bubble tea shops and cafes. Additionally, the use of visualization techniques to display industrial agglomerations, highlighted with key frequently occurring keywords, offers detailed and nuanced perspectives on the spatial-functional complexities within agglomeration economies. This approach enriches our understanding by providing a more detailed and contextual view of how industries cluster and interact within urban environments.

Chapter 7. A Network Perspective

7.1 Background and Motivation

While Chapters 1 and 2 have detailed interactions among cities, there remains a notable gap in understanding local sectoral agglomeration economies at both detailed and broad levels, particularly concerning how these economies integrate intra-regional contexts with modern spatial and network analytics. Enhancing our grasp of network organizations within intra-regional and intracity agglomerations is vital for both scholarly discourse and practical applications.

The lack of detailed exploration into these dynamics poses significant issues. On one hand, intraregional economies, especially in megacity regions, are becoming increasingly crucial in the global economic landscape. Small and medium-sized cities, by leveraging their unique locational advantages, can both contribute to and benefit from regional economic dynamics. The concept of "borrowed size," introduced by Alonso (1973) and further discussed by Burger & Meijers (2016), suggests that smaller cities can enhance their urban functions by tapping into the agglomeration benefits of larger neighbouring cities. This interconnection between cities of varying sizes can lead to reductions in production and transportation costs, where functional urban areas (FUAs) might extend beyond or contract within traditional city boundaries (Meijers & Burger, 2017).

On the other hand, the reliance on economic censuses and surveys, typically recommended to occur every five years as in the U.S. and China, presents a challenge. These methods are not only costly and time-consuming but also potentially yield data that quickly becomes outdated and unreliable, as noted by Wardrop et al. (2018). Additionally, the static nature of this data, often constrained by limited sample sizes and typically confined to administrative units, does not adequately capture the dynamic and heterogeneous nature of urban agglomeration economies across different sectors (Yu & Liu, 2021). Given these challenges, there is a pressing need for a data-driven approach to understand and map the complex regional organization of sectoral agglomeration economies more effectively. This approach would offer new insights into the spatial-functional patterns of economic interactions within regions, addressing a critical gap in current research and policy-making.

This study adopts a data-driven approach to bridge the research gaps, utilizing geospatial big data and network analytics to delve into the industrial dynamics of China's Greater Bay Area (GBA). We begin by extracting functional traits of industrial activities from Points of Interest (POIs) to construct a network that elucidates the spatial-functional roles of these agglomerations. Through the use of Latent Dirichlet Allocation (LDA) topic modeling, we extract thematic information that represents the various functional topics associated with each agglomeration. To further analyze these relationships, we employ a bipartite network projection technique to transform the initial two-mode network of agglomerations and functions into a one-mode network. This transformation allows for a clear depiction of interactions either among the agglomerations themselves or between the different industrial functions they encompass. The structural properties of these networks are then assessed using specific network metrics to gain insights into the nature of these connections. In the final phase of our analysis, we apply a community detection algorithm to the agglomeration networks. This method helps identify clusters of agglomerations that are tightly interconnected through similar industrial functions, providing a nuanced understanding of the regional industrial landscape and highlighting key functional linkages within the GBA. This study is proposed to answer three main research questions:

1) Given the bipartite network projection, how can underlying industrial networks among local agglomerations be constructed via POIs collected in the GBA?
2) By cooperating with multiple network properties, what are the structural characteristics of the agglomeration networks? In this sense, is there a significant heterogeneity of network structures across different topics regarding their geographical layouts?
3) Which agglomerations are closely connected by their industrial functions and spatial proximity, and are there diverging spatial-functional patterns across different communities?

7.2 Main Findings

This section presents research findings by considering the industrial agglomerations identified in the GBA as a case study, helping us answer the proposed research questions.

7.2.1 Industrial functions identified by topic modeling

The results from our LDA topic modeling of industrial functional topics in the Greater Bay Area (GBA) are outlined in Table 7, which displays the top-6 ranked keywords for each thematic topic based on their probabilistic weights. The table highlights several functional topics that are closely linked to specific industries or productions. For instance, topic 2 is marked by keywords like "furniture," "antique furniture," and "mahogany," suggesting a strong association with the furniture industry. Topics 3 and 8 indicate connections to industries involved in building and construction as well as home decoration, evidenced by keywords such as "ceramics," "stainless steel," "building materials," "bathroom," and "illumination."

Furthermore, topics 5 and 7 appear to be linked to the garment and leather industries, featuring terms such as "garment," "textile," "clothing," "footwear," "handbag," and "embroidered." These topics demonstrate a focus on specialized industrial functions that are predominantly related to specific manufacturing processes. In contrast, other topics are characterized by a more diverse range of industrial functions. For example, topic 1 encompasses a wide array of manufacturing activities, including keywords related to "kitchen equipment," "mechanical parts," and "appliances," indicating a diversified industrial function. This differentiation in topic content illustrates the varied nature of industrial activities within the region, showcasing both specialized and diversified functional roles.

Table 7. The top-6 ranked keywords of each identified industrial functional topic.

Topic	**Identified top-6 ranked keywords**

1	Kitchenware, Rubber, Bearing, Electromechanical, Electrical appliance, Automation
2	Furniture, Antique furniture, Mahogany, Steel, Couch & Sofa, Wood furniture
3	Ceramics, Stainless steel, Building materials, Metal hardware, Decorate, Stone
4	Electronic, Electromechanical, Metal hardware, Chemical engineering, Construction, Precision
5	Garment, Textile, Clothing, Apparel, Embroidered, Knitting
6	Metal hardware, Print, Plastic, Food, Electronic, Garment
7	Leather goods, Metal hardware, Footwear, Handbag, Embroidered, Edge paint
8	Bathroom, Illumination, Metal hardware, Lighting decoration, Accessories, Metal material, Plumb
9	Mechanical, Metal hardware, Plastic, Electrical appliance, Stainless steel, Molding

Our analysis of industrial functional topics within the Greater Bay Area employs a bipartite network approach, the findings of which are detailed in our report. Figure 25 (A) showcases the node degree for all industrial topics, a metric that represents the number of connections between two distinct sets—agglomerations and industrial topics. In practical terms, node degree reflects the count of agglomerations that incorporate a significant topic. We note that topics 3, 4, and 9, which are primarily associated with material, electronic, and mechanical productions, exhibit higher node degrees. This suggests these topics are more prevalent across multiple agglomerations. Conversely, other functional topics display comparatively lower degrees, indicating they play a less dominant role in the network of agglomerations. Subsequently, a bipartite projection is performed to consolidate every pair of functional topics that appear together within the same manufacturing agglomeration. To visualize these relationships, both a Circos plot and a heat map are utilized, as depicted in Figures 25 (B) and (C). The Circos plot illustrates the connections between topics, with the width of the edges indicating the intensity of the relationships. Notably, connections such as those between topics 4 and 9 are highlighted by significantly wider edges, suggesting strong inter-topic relationships.

The heat map further quantifies the intensity between pairs of topics, with darker colors within the squares indicating higher frequencies of co-occurrence in the same manufacturing agglomerations. This visualization technique helps to identify potential commonalities and complementarities between topics, indicating where functional synergies may exist within the industrial landscape. These visual tools collectively provide a deeper understanding of how different industrial topics interconnect and the potential for collaborative interactions within the agglomeration network.

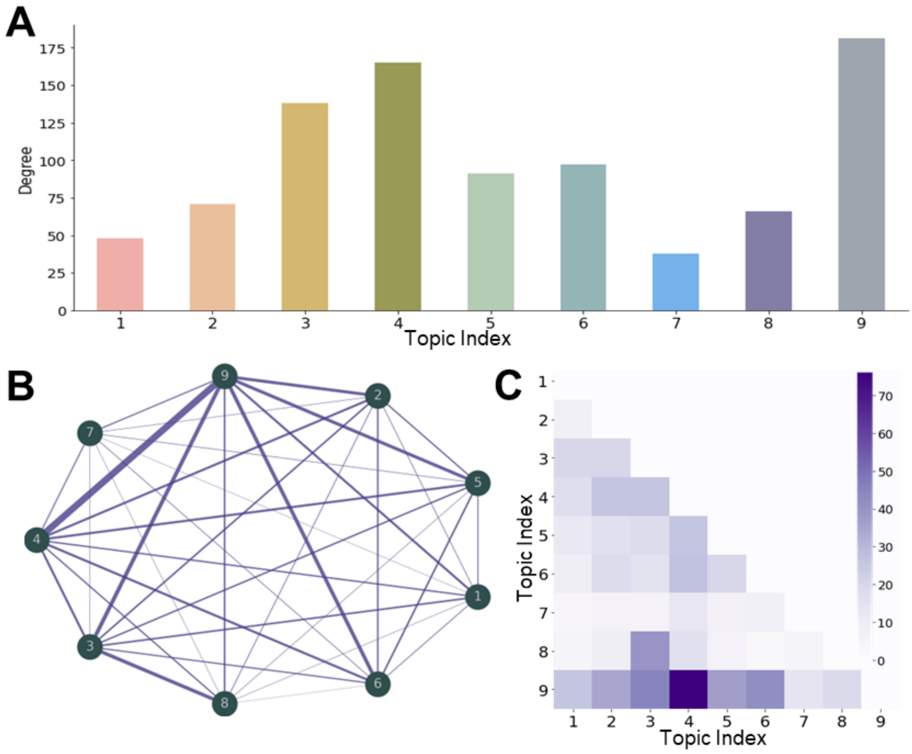


Figure 25. (A) The node degree distribution of industrial functions. (B) The Circos plot to indicate the co-occurrence relationships among industrial functions (C) The heat map used to quantify the co-occurrence intensity between each pair of topics.

7.2.2 The network properties and spatial structures of industrial topics

This subsection delves into the network properties and spatial structure across various industrial topics, highlighting how different industrial functions manifest distinct patterns of spatial co-occurrence within manufacturing agglomerations. Table 8 presents an overview of the network properties, aggregating data for all topics as well as for individual topics identified through LDA topic modeling.

Notable patterns emerge from the analysis. Specifically, the network characteristics of topics 4 and 9 stand out as relatively unique among the functional topics examined. These topics demonstrate agglomeration networks with notably high density and clustering coefficients. High network density indicates that these networks possess more intensive connections, significantly surpassing the connectivity observed in the broader network and in networks related to other topics. This suggests that agglomerations linked to topics 4 and 9 are more interconnected, facilitating dense clusters of industrial activity.

The global clustering coefficient further corroborates these findings, suggesting that networks related to topics 4 and 9 are more likely to form tightly-knit clusters characterized by robust functional connections. This degree of connectivity implies a strong collaborative framework within these specific industrial sectors. Additionally, the analysis of degree metrics sheds light on structural variances across different topic networks. For instance, the average degree within the network of topic 5 is notably lower compared to topics with comparable network sizes, such as topics 2 and 6. This indicates that agglomerations associated with the garment industry may have fewer connections to other industrial sectors, suggesting a lesser degree of interdependence on regional industrial networks.

In contrast, agglomerations associated with topics that exhibit a higher average degree demonstrate the potential for greater regional production complementarity. These agglomerations likely benefit from enhanced interactions and dependencies within the same regional network, suggesting a more integrated industrial ecosystem. This analysis not only highlights the unique network properties of each topic but also underscores the varied degrees of interconnectivity and potential for collaboration across different industrial functions.

Table 8. The network properties of overall and single industrial topics.

Topic	**Size**		**Density**	**Degree**			**Global clustering**
	Node	**Edge**		**Average**	**25%**	**75%**	
Overall	424	21368	0.24	100.79	56	114	0.62
Topic 1	47	451	0.42	19.19	2	29	0.71
Topic 2	71	1121	0.45	31.58	6	48	0.70
Topic 3	138	3882	0.41	56.26	7	69	0.71
Topic 4	165	6274	0.46	76.05	6	103	0.79
Topic 5	85	765	0.21	18.00	2	32	0.71
Topic 6	89	953	0.24	21.41	1	36	0.82
Topic 7	34	282	0.50	16.59	3	22	0.74
Topic 8	55	510	0.34	18.55	2	32	0.70
Topic 9	181	7478	0.46	82.63	15	103	0.75

To deepen our understanding of the distinct characteristics among industrial functional topics, we have mapped out the agglomeration networks for each topic, analyzing their structure from a geospatial perspective within the Greater Bay Area (GBA). Figure 26 illustrates the spatial configurations of these networks, both overall and for individual industrial topics. The general observation from Figure 26 (A) reveals clustering patterns that appear to correlate with the geographic proximity of industrial activities. Notably, agglomerations are primarily concentrated in a dual-center structure encompassing the major cities of Guangzhou and Shenzhen in the central and southeastern parts of the GBA. Additionally, there are dense networks isolated from this main layout, such as those in Hong Kong and the southwestern GBA, near the Jiangmen and Zhongshan areas. Given that the overall network aggregates data from all topics and may obscure specific details, our analysis focuses on the networks associated with individual industrial functions. Figure 26 (B) showcases these networks, underscoring significant spatial heterogeneities that suggest manufacturing activities related to different industrial functions exhibit distinctly varied patterns.

A key observation is the pronounced clustering within networks for topics 4 and 9, which align closely with the dual-center structure previously noted. This clustering indicates a high density of industrial functions within these areas, affirming earlier findings about their dense and cohesive network structures. Moreover, the networks for topics 6 and 8 reveal less pronounced clustering patterns, particularly noted in Hong Kong and the southwestern GBA—areas highlighted within the broader network context. This spatial differentiation provides insight into the functional segregation between Shenzhen and Hong Kong, particularly between topics 4 and 6. Such findings offer a nuanced understanding of the regional industrial landscape, highlighting how different industrial functions are geographically and functionally segmented within the GBA.

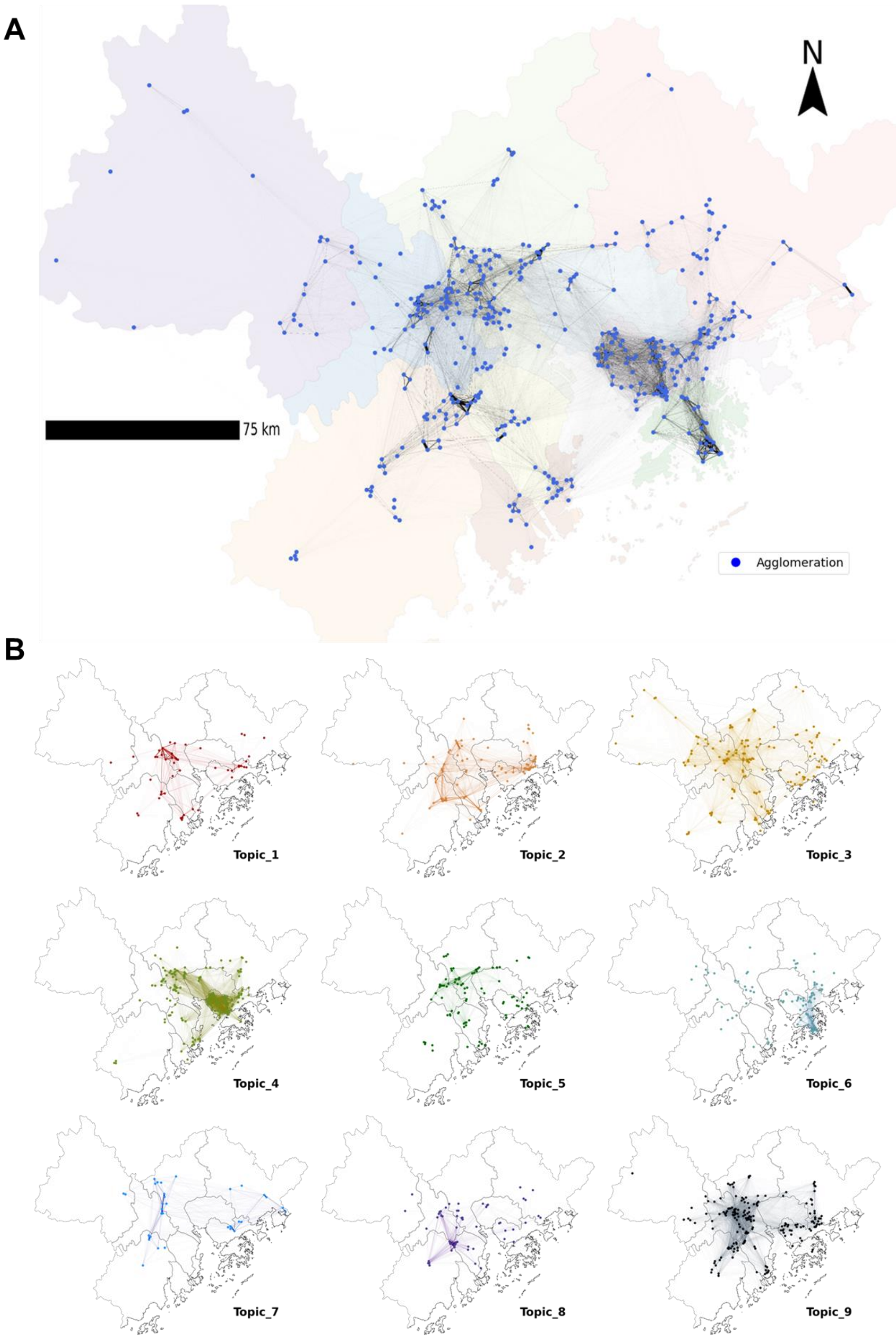


Figure 26. (A) The agglomeration network of overall industrial functions. (B) The agglomeration networks of single industrial functions.

7.2.3 Community structure of the agglomeration network

In this subsection, we report the result of applying community detection to the agglomeration network. Figure 27 (A) shows the visualization of communities extracted from the agglomeration network, where different colors labeled on nodes represent generated communities. Eight communities are identified via a community detection algorithm. As can be seen, the bulk of the communities covers a geographically cohesive area, suggesting that spatially adjacent agglomerations tend to contain stronger connections. Moreover, the modularity score of 0.79 delivers solid proof to demonstrate the community structure of this network, in which the functional connections are not randomly distributed amongst industrial agglomerations.

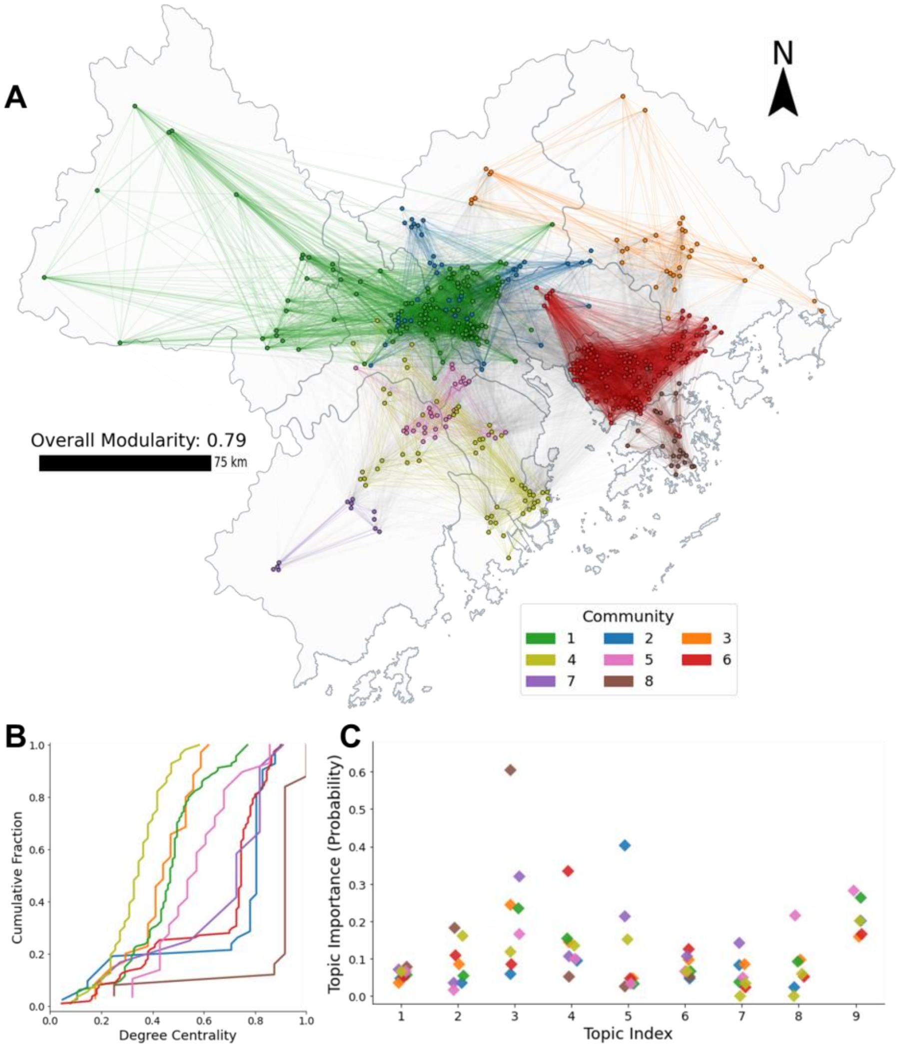


Figure 27. (A) The partition result of community detection for the overall agglomeration network, of which various colors represent the identified communities. (B) The cumulative probability distribution of degree centrality in terms of different communities. The different colors match the legend from the partition result of community detection. (C) The categorical strip plot for showing the importance of industrial functions. The different colors match the legend from the partition result of community detection.

Community 1 in the Greater Bay Area (GBA) is predominantly composed of industrial agglomerations located in the central and north-western regions, including Guangzhou and Foshan—cities recognized for their traditional industrial bases in southern China. Additionally, peripheral agglomerations in Zhaoqing are integrated into Community 1, suggesting a notable influence exerted by the central cities on their surrounding areas. According to the analysis of degree centrality distribution and topic importance weights depicted in Figure 27 (B) and (C), Community 1 demonstrates weaker internal connections and shows a pronounced focus on heavy industries, specifically building material and mechanical productions. Conversely, Community 6 features distinct spatial-functional attributes, encompassing the entire region of Shenzhen and extending into adjacent areas of southern Huizhou near Shenzhen and Dongguan. This community is characterized by tightly-knit interactions within industrial agglomerations, predominantly focused on electronic productions, as indicated by detailed metrics of degree centrality and topic importance.

Other communities within the GBA also display unique spatial-functional patterns. Community 2, for instance, is distributed around the central GBA, particularly in the outskirts of Guangzhou, and is heavily oriented towards garment production. Moreover, Hong Kong predominantly houses Community 8, which maintains minimal interactions with neighboring Shenzhen (Community 6). This aligns with previous findings and underscores the industrial segregation between Shenzhen and Hong Kong, influenced by the broader industrial transformations and public transportation configurations in the region (Yu et al., 2022c). Yeh (2012) noted a significant shift in manufacturing activities from Hong Kong towards cities along the Pearl River estuary, such as Dongguan and Shenzhen.

These observations reinforce the idea that regional agglomeration economies are shaped not only by spatial proximity but also by specific industrial sectors, offering insights into how regional dynamics influence industrial distribution and interactions.

7.3 Discussions

An examination of the current literature reveals that while there is substantial empirical data concerning the dynamics of interaction among cities, there is a noticeable lack of detailed focus on the urban network of regional sectoral agglomeration economies. In response to this deficit, this study selects the Greater Bay Area (GBA) as a focal point and utilizes geospatial big data along with network analysis techniques. This approach is aimed at exploring three specific research questions that will underpin the discussions that follow.

To identify specific industrial functions within the Greater Bay Area (GBA), we utilize Latent Dirichlet Allocation (LDA) topic modeling to analyze industrial agglomerations. Certain keywords within these topics suggest strong associations with particular industries. For example, topic two predominantly relates to furniture production, and topic three is tied to the building and construction materials sector. Rosenthal and Strange (2001) noted that specialized agglomeration economies often allow manufacturing industries to gain substantial advantages by geographically concentrating, which likely contributes to the observed patterns of industrial co-location.

To further elucidate the connections between the identified industrial topics and their respective agglomerations, we introduce an interlocking network model. In this model, a higher node degree among topic nodes signifies a strong presence within manufacturing agglomerations, indicating their prominence and influence within the GBA. Our analysis reveals that the most connected industrial functions generally belong to medium-technology sectors, such as material, electronic, and mechanical productions, reflecting a trend towards medium-tech industries (Eurostat, 2018). Additionally, we employ a bipartite projection technique to combine pairs of functional topics that frequently co-occur within the same manufacturing agglomerations. This method helps in measuring the interdependencies between different industrial functions. For instance, the association between topics three and eight suggests the formation of a composite supply chain system. Through addressing the first research question, this study employs a granular topic-learning approach that facilitates the development of interlocking network models, providing a nuanced understanding of how various industrial topics interconnect within the GBA.

We explore the network properties and spatial configurations of various industrial functions to reveal their structural and geographic diversities. Specifically, topics 4 and 9 exhibit distinct network characteristics that set them apart from other topics, likely due to their significant roles in material, electronic, and mechanical productions within the Greater Bay Area (GBA). This finding corroborates with Hu and Lin (2011), who noted the profound influence of mechanical and electrical sectors on the manufacturing economy of the Pearl River Delta (PRD). In examining the spatial distribution of agglomeration networks, we identify dense interaction zones particularly in the central and southeastern GBA, forming a dual-center structure around the major cities of Guangzhou and Shenzhen. Additionally, peripheral areas show signs of economic radiance from these central hubs, aligning with observations by Liu et al. (2020a) that manufacturing industries are gradually relocating to the PRD's periphery while high-tech industries concentrate in the central and sub-central urban areas.

By mapping the spatial layouts of topic networks, we uncover pronounced spatial heterogeneities. For instance, topic 4, which is heavily concentrated in Shenzhen and its adjacent areas, underscores Shenzhen's emergence as a leading hub for information and communications technology (ICT) industries in southern China. This pattern reflects broader network dynamics observed by Storper et al. (2015) in the San Francisco Bay Area, where distinct network configurations were identified across different interconnected industrial sectors. This network-based methodology, employed here for the GBA, offers valuable insights into the spatial structures of interrelated industrial sectors and can similarly be applied to other regions, including other megacities or bay areas. Such analyses are instrumental in understanding regional dynamics and supporting growth and development strategies (Funderburg & Boarnet, 2008).

This research further examines the community structure within the agglomeration network, a crucial feature of real-world networks. The analysis identifies eight communities, each displaying significant geographical cohesion and dense functional interactions. For example, community 6, primarily located in Shenzhen with extensions into western Huizhou and Dongguan, showcases specific spatial preferences. Analyzing its centrality and thematic importance reveals that these spatial patterns are intricately connected to sectoral economies, enhancing our understanding of urban network dynamics.

Theoretical contributions by Meijers (2005) and Capello (2000) describe urban networks as nonlinear systems that foster complementary relationships among cities through cooperative activities to capitalize on scale economies. This study's findings align with their theories, presenting empirical support for the existence of complex and diverse agglomeration communities within a megaregional framework, observed through the prism of industrial functions. This aligns with observations from other megacity regions, such as the Beijing–Tianjin–Hebei urban agglomeration (BTHUG). Liu et al. (2021a) delineated the urban structure of the BTHUG by developing a weight-directed spatial network, which identified varied spatial interaction patterns. However, a notable difference in the BTHUG is its reliance on a single global center, contrasting with the dual-center configuration seen in the GBA. This disparity highlights the crucial role of regional spatial proximity and industrial functions in defining the interactions and network dynamics of urban agglomerations.

Chapter 8. Conclusion

8.1 Theoretical Implications for Real-world Issues

Given the research focus of this thesis on economic topics, practical implications regarding geospatial perspectives on economic activities within the megaregional context can be drawn from the present case studies. Aligning with the specific background of the chosen study area, the following implications for local and regional policy formulation and implementation are introduced based on analytical empirical evidence:

In Chapter 4, the study presents a number of policy implications derived from its key findings, which offer empirical support for the economic transformations observed in various Asian cities. Specifically, the results confirm the effects of economic transformation in Hong Kong from a spatial perspective, providing a comprehensive view of job creation and employment distribution patterns. As noted by Yeh in 2011, Hong Kong has undergone a significant economic shift since the late 1990s, transitioning primarily from manufacturing to service-oriented industries. The research highlights substantial spatial variability among different economic sectors, attributing this to the economic transformation within the region. Additionally, by analyzing the locational preferences of these sectors, the study equips policymakers in other regions with insights necessary for developing future strategies in urban planning and economic development.

Furthermore, our study contributes to the ongoing debates concerning the Hong Kong government's new town program. The regional urban planning and land use strategies are principally influenced by local economic developments that enhance tax revenues and employment opportunities. Recent discussions have raised important questions about the role of new towns in decentralizing the population from urban centers, which are crucial for evaluating policies and shaping future economic strategies (He et al., 2020). Our findings indicate that the manufacturing and trade and logistics sectors have shown significant agglomeration economies in new towns, primarily located in the New Territories. However, the service sector has not demonstrated a comparable capacity to attract the working population away from traditional commercial and business hubs (Hui & Lam, 2005).

Significantly, our study underscores the importance of considering local contexts when formulating policies, as highlighted by the spatial heterogeneity of agglomeration economies and their correlation with built environment factors and socio-demographic variables. Our regression analyses reveal that agglomeration economies are sector-specific and influenced by locational advantages, which include natural features, market accessibility, and land use patterns. For instance, the trade and logistics sector often locates near borders, a strategic choice that facilitates thriving cross-border trade with cities in mainland China. This finding supports the strategy of enhancing regional integration and improving spatial connections through transportation networks, which can boost the growth and effectiveness of regional collaborations among different local agglomeration economies (Hui et al., 2020). In the context of the Greater Bay Area (GBA), this raises a critical consideration for regional and local stakeholders on how to augment local economic activities by fostering collaboration with neighboring cities or regions.

Chapter 5 of our analysis offers significant contributions to both the theoretical literature and practical policymaking. Our empirical findings provide essential insights that aid in policy formulation and implementation, with an emphasis on the spatial heterogeneity of human activities across economic sectors influenced by locational advantages. Such advantages include specific location characteristics, regional policies, and land-use factors (Faggio et al., 2017). For instance, the manufacturing sector shows a preference for rural locations such as northern Shenzhen, western Guangzhou, and the peripheries of the Greater Bay Area (GBA) due to factors like lower land costs, abundant labor, and proximity to thriving markets. These insights offer regional policymakers focused on economic development and secondary industries substantial reasons to promote manufacturing activities.

Moreover, our research delineates functional heterogeneity within identified agglomerations. The primary sector, for example, is inclined towards horticulture in urbanized areas, whereas the quaternary industries in Shenzhen lean heavily towards Information and Communications Technology (ICT) production. Interestingly, manufacturing activities display varying functional tendencies, with specialization in less populated regions and diversified industrial agglomerations playing a more significant role in densely populated urban areas. This suggests that policymakers and urban planners should consider more nuanced development plans to foster industrial agglomerations within specific areas and sectors, such as devising tailored strategies for manufacturing agglomerations to facilitate the transition from traditional manufacturing methods.

Lastly, at the mega-regional level, our findings underscore the importance of in-depth collaboration among cities within the GBA to foster regional economic development. The analysis reveals patterns of homogeneous competition in industries such as clothing and electronics across different city tiers within the GBA. Addressing these competitive overlaps, avoiding resource misallocation, and enhancing both spatial and functional collaborations among cities are critical for future regional strategies, underscoring the need for concerted efforts to mitigate competition in conventional industries and promote efficient regional integration.

In Chapter 6, our study offers valuable contributions to the practical implications concerning regional economic issues and the policy-making process, particularly in the cities surrounding the Pearl River Delta (PRD). The manufacturing industries, widely acknowledged as the backbone of local and regional economies, have been the focus of policy efforts by regional policymakers in the Greater Bay Area (GBA) aimed at transitioning from traditional manufacturing to mid- and high-tech industries (Hui et al., 2020). Despite the shift towards electronic and mechanical productions in the central GBA, empirical evidence from this study identifies persistent traditional labour-intensive industries in certain areas, primarily engaged in low-tech and specialized production such as textiles, leather, and building materials. This observation provides critical insights for refining industrial restructuring strategies and assessing policy scalability.

Additionally, the study highlights the issue of spatial-functional inequality. It identifies a dual-center cluster structure within the GBA, characterized by a broad range of professional sectors concentrated in densely populated areas. The cluster effect, which facilitates geographic concentration of manufacturing industries, has been well recognized for its efficiency benefits, including optimized supply chains that reduce production and transportation costs (Enright et al., 2005; Guo et al., 2023). However, peripheral areas of the GBA continue to experience a scarcity of industrial agglomerations, particularly in sectors labelled as low-tech. This disparity underscores the importance of addressing spatial and functional inequalities in regional economic policy-making, particularly in rural subregions, to ensure sustainable development.

Chapter 7 further builds on these insights, providing valuable contributions to both theoretical literature and practical applications. The study observes significant structural and geographic heterogeneities in agglomeration networks that are influenced by locational advantages. Industries such as home decoration, garments, and leather products tend to locate in peripheral areas of core cities, driven by factors like land cost, labour availability, and proximity to markets (Faggio et al., 2017). This differential spatial pattern suggests the need for targeted urban planning and economic development strategies that address the unique needs of local industrial agglomerations and enhance their integration with broader regional economies.

Moreover, the study expands on regional scale considerations, discussing the strategic planning necessary for coordinated regional development across various city tiers. It highlights an emerging agglomeration network as an alternative framework to analyse intraregional city relationships. Empirical evidence points to a robust radiation effect from industries like building materials and mechanical productions on the north-western hinterland, indicating the development of synergistic relationships that reduce production and logistical costs through intensive agglomeration networks. However, significant functional disparities, such as between the electronics industries in Shenzhen and the lighter industries in Hong Kong, call for greater efforts in policy innovation to ensure sustainable integration within the GBA.

8.2 Potential Drawbacks

It should be acknowledged that some limitations can be observed in the chapters of case studies used and presented in this thesis. We summarize and discuss the main potential drawbacks as follows:

In the Chapter 4, a few caveats should be mentioned related to the methodology. Initially, we assess broad employment trends using aggregate-level employment data, which may overlook certain individual nuances, such as self-employed individuals or small-scale enterprises employing part-time workers. To address this limitation, supplementary questionnaire data is recommended to provide a more comprehensive understanding of employment dynamics. Secondly, our analysis is confined to the 2016 population by-census in Hong Kong and does not encompass longitudinal dynamics. For instance, areas like Kwun Tong and Kowloon Bay have historically been recognized as hubs for manufacturing employment, based on local knowledge. However, our findings do not indicate significant clustering effects in these regions. This discrepancy could be attributed to recent governmental initiatives transforming these districts into new central business districts (CBDs), with a focus on commercial services, by repurposing old industrial infrastructure into office spaces. To capture such evolving spatiotemporal dynamics, a longitudinal analysis is recommended, as it provides a suitable framework for tracing changes in agglomeration economies over time (Garcia-López & Muñiz, 2013).

Regarding Chapter 5, a notable concern pertaining to the spatial organization findings is the potential inadequacy of capturing primary industries like forestry, plantation, and mining activities due to their land-intensive nature. This challenge arises from the limitations of accurately delineating their spatial extents solely through point events via kernel density estimation (Sheather & Jones, 1991). Moreover, relying solely on Points of Interest (POIs) might result in overlooking crucial details regarding the intensity of industrial activities, such as local companies' transactions, employment rates, or generated profits. To address these limitations, it is essential to complement POI data with regional statistical information encompassing industrial value, patterns, and employment data. This integrated approach would enrich the exploration of industrial intensity and provide a more comprehensive understanding.

Another limitation of this case study lies in its exclusive focus on the spatial and functional characteristics of sectoral agglomerations. However, within a specific industrial sector, there may exist latent relationships among identified agglomerations that warrant further investigation. Thus, future research should aim for a more focused exploration of the interplay among agglomerations to uncover potential implications that were not addressed in our study.

The case study outlined in Chapter 6 faces several limitations. Firstly, it relies on Latent Dirichlet allocation (LDA), an unsupervised algorithm utilized to extract functional topics across industrial agglomerations. While LDA effectively captures overarching semantic patterns and spatial co-occurrence of keywords, the specific functional topics it generates may not accurately reflect reality, as noted in prior literature (Gao et al., 2017). Additionally, there is a concern regarding the sole reliance on Points of Interest (POIs) data, which may not comprehensively reflect industrial activities and the local economy, thus limiting the capture of human activities within these areas. To address these limitations, one potential approach is to integrate more detailed ground truth information, such as regional economic statistics, as supplementary datasets. Furthermore, recent advancements in language models, like BERT, have shown promising capabilities in enhancing topic extraction (Balsebre et al., 2023). Therefore, future research could explore the integration of BERT to gain a more nuanced understanding of the relationships among various elements within industrial agglomerations.

Chapter 7 highlights several important limitations worth noting. Firstly, the information derived from Points of Interest (POIs) regarding spatial and functional attributes may not fully represent real-world manufacturing industries, potentially introducing biases. One approach to mitigate this is cross-validation with alternative datasets, such as firm-level economic statistics and remote sensing imagery, which provide more accurate ground truth information (Seto & Kaufmann, 2003). Another limitation is the requirement for long-term observation datasets. Expanding the scope of research to include time-series analysis allows for a more comprehensive understanding, including the temporal evolution of agglomeration networks and the diversity of behaviors over time. For instance, exploring the spatiotemporal variations of specific industrial functions within the Greater Bay Area (GBA) presents intriguing avenues for future research.

8.3 Future Direction and Outlook

The utilization of advanced geospatial analytics, encompassing diverse data sources, spatial models, metrics, and methodologies from machine learning and network analysis, has enabled us to gain alternative and comprehensive insights into agglomeration economic activities, as demonstrated from Chapter 4 to 7. My doctoral research commenced with a re-examination of conventional data sources to explore urban agglomeration economies in cities. Over the course of my three-year study, several empirical case studies were introduced to support the integration of geospatial analytics in the field of megaregional agglomeration economies.

However, given the rapid development of diverse machine learning technologies and interdisciplinary knowledge, my current research work serves as an initial point. Beyond the proposed methodologies in this thesis, numerous questions remain, necessitating exploration through more innovative solutions. Three major aspects of future directions are highlighted and are worthy of further pursuit.

a) Utilizing AI-driven strategies to simulate the spatial-functional change of agglomeration economies

   AI-driven strategies are increasingly employed in the research community to analyze and model complex systems within cities, especially in projecting the impacts of diverse factors on various human activities (Crooks & Heppenstall, 2011; Macal & North, 2005). A promising research avenue involves simulating the spatial-functional patterns of agglomeration economies in the future. This simulation should consider externalities arising from built environments and socioeconomics, as highlighted in previous evidence in Chapter 4, where economic activities were found to be closely linked to these factors. The scenario-based outcomes derived from such simulations are crucial for comprehending and assessing the direct impacts of urban environments. This, in turn, provides valuable insights into future planning strategies aimed at enhancing urban economic development.

b) Incorporating population flows network to better understand agglomeration economies

   The urban agglomeration expansion has facilitated substantial population, goods, capital, and services movements between cities, fostering connectivity and vitality beyond city borders and giving rise to megaregional agglomerations. Existing literature has explored spatial structure dynamics within cities through network analysis rooted in spatial interactions. However, a fundamental inquiry remains: how do spatial interactions of population flows contribute to the megaregional network, and what potential impacts do they entail? Investigating the population flow network among diverse agglomeration areas within a megacity region becomes crucial. Spatial network analysis emerges as a vital tool for comprehensively understanding the spatial structure of population flows, leveraging the increasing availability of mobility data, including migration information and mobile phone trajectories (Liu et al., 2021a).

c) Extending the research scope from physical economic to virtual economic activities

Incorporating urban agglomeration into virtual spaces presents a challenge as the existing understanding of urban vitality lacks explicit consideration for patterns in the virtual realm. The pervasive influence of information and communication technology (ICT) has propelled a surge in human activities occurring in virtual spaces, reshaping geospatial knowledge with distinct logic and morphology (Ash et al., 2018). The term "virtual space," recently introduced in geography communities, not only encompasses online activities and interactions but also extends to include human dynamics, social relationships, and various relational constructs within non-digital physical spaces. Diverse activities, such as browsing, commenting, shopping, and trading within a virtual space, correspond to geographic locations in a physical space (Lloyd & Cheshire, 2017). This prompts pertinent questions regarding the integration of virtual activities into the current understanding of urban agglomeration, introducing a hybrid physical-virtual perspective.

8.4 Closing Words

In recent decades, urbanization has experienced a significant surge, with the urban population projected to reach nearly 70% of the global populace by 2050. Urban migrants and newborns play pivotal roles in this growth, contributing substantially to economic development in cities. The concept of urban agglomeration economies, describing the clustering phenomenon in cities and its close link to economic prosperity, has gained prominence. Evidence indicates that firms in robust industrial agglomerations tend to experience accelerated growth and greater benefits. As urbanization extends to megaregions—densely populated areas with abundant services and resources—there is an escalating need for comprehensive studies on agglomeration economies within these contexts.

This thesis employs geospatial analytics to systematically investigate industrial agglomeration economies in megaregional settings. The subsequent chapters comprehensively explore and discuss topics related to organization, disparity, and network dynamics, substantiated by empirical evidence. The analytical results make a substantial contribution to both theoretical understanding and practical applications in the field of regional economic studies and policy formulation. This thesis provides crucial empirical evidence that offers new insights into the geographical and functional dynamics of agglomeration economies in the megaregional context.

Furthermore, the methodologies derived from multidisciplinary domain knowledge demonstrated in this thesis underscore the recent advancements in geospatial analytic solutions and their potential applications in subsequent geospatial studies. It is suggested that geospatial data science and analytics could serve as powerful tools for addressing research questions related to socioeconomic issues in cities. The research communities are urged to continue their efforts in contributing to the understanding of urban agglomeration economies through the lens of geospatial analytics.

References

Alonso, W. (1973). Urban zero population growth. *Daedalus*, 191-206.

Angel, D. P. (1991). High-technology agglomeration and the labor market: the case of Silicon Valley. *Environment and Planning A*, *23*(10), 1501-1516.

Ash, J., Kitchin, R., & Leszczynski, A. (2018). Digital turn, digital geographies?. *Progress in Human Geography*, *42*(1), 25-43.

Attaran, M. (1986). Industrial diversity and economic performance in US areas. *The Annals of Regional Science*, *20*(2), 44-54.

Beaudry, C., & Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research policy*, *38*(2), 318-337.

Bahrehdar, A. R., & Purves, R. S. (2018). Description and characterization of place properties using topic modelling on georeferenced tags. *Geo-spatial Information Science*, *21*(3), 173-184.

Balsebre, P., Yao, D., Cong, G., Huang, W., & Hai, Z. (2023). Mining Geospatial Relationships from Text. *Proceedings of the ACM on Management of Data*, *1*(1), 1-26.

Berger, A., Caruana, R., Cohn, D., Freitag, D., & Mittal, V. (2000, July). Bridging the lexical chasm: statistical approaches to answer-finding. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 192-199).

Bettencourt, L., Samaniego, H., & Youn, H. (2014). Professional diversity and the productivity of cities. *Scientific reports*, *4*(1), 1-6.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Brandes, U., Robins, G., McCranie, A., & Wasserman, S. (2013). What is network science?. *Network science*, *1*(1), 1-15.

Burger, M. J., & Meijers, E. J. (2016). Agglomerations and the rise of urban network externalities. *Papers in Regional Science*, *95*(1), 5-15.

Buzard, K., Carlino, G. A., Hunt, R. M., Carr, J. K., & Smith, T. E. (2017). The agglomeration of American R&D labs. *Journal of Urban Economics*, *101*, 14-26.

Cainelli, G., Ganau, R., & Modica, M. (2019). Industrial relatedness and regional resilience in the European Union. *Papers in Regional Science*, *98*(2), 755-778.

Camagni, R., Capello, R., & Caragliu, A. (2017). Static vs. dynamic agglomeration economies: Spatial context and structural evolution behind urban growth. *Seminal Studies in Regional and Urban Economics: Contributions from an Impressive Mind*, 227-259.

Capello, R. (2000). The city network paradigm: measuring urban network externalities. *Urban Studies*, *37*(11), 1925-1945.

Capello, R. (2015). *Regional economics*. Routledge.

Caragliu, A., de Dominicis, L., & de Groot, H. L. (2016). Both Marshall and Jacobs were right!. *Economic Geography*, *92*(1), 87-111.

Carley, M., & Spapens, P. (2017). *Sharing the world: sustainable living and global equity in the 21st century*. Routledge.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation research. Part D, Transport and environment*, *2*(3), 199-219.

Chainey, S., Reid, S., & Stuart, N. (2002). *When is a hotspot a hotspot? A procedure for creating statistically robust hotspot maps of crime* (pp. 21-36). Taylor & Francis, London, England.

Chatman, D. G., & Noland, R. B. (2011). Do public transport improvements increase agglomeration economies? A review of literature and an agenda for research. *Transport Reviews*, *31*(6), 725-742.

Church, A., & Reid, P. (1996). Urban power, international networks and competition: the example of cross-border cooperation. *Urban studies*, *33*(8), 1297-1318.

Combes, P. P. (2000). Economic structure and local growth: France, 1984–1993. *Journal of urban economics*, *47*(3), 329-355.

Crooks, A. T., & Heppenstall, A. J. (2011). Introduction to agent-based modelling. In *Agent-based models of geographical systems* (pp. 85-105). Dordrecht: Springer Netherlands.

de Oliveira Capela, F., & Ramirez-Marquez, J. E. (2019). Detecting urban identity perception via newspaper topic modeling. *Cities*, *93*, 72-83.

de Lucio, J. J., Herce, J. A., & Goicolea, A. (1996). *Externalities and industrial growth: Spain 1978-1992* (No. 96-14). FEDEA.

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research policy*, *43*(10), 1785-1799.

Deng, Y., Liu, J., Liu, Y., & Luo, A. (2019). Detecting urban polycentric structure from POI data. *ISPRS International Journal of Geo-Information*, *8*(6), 283.

Derudder, B. (2006). On conceptual confusion in empirical analyses of a transnational urban network. *Urban Studies*, *43*(11), 2027-2046.

Dicken, P. (2007). *Global shift: Mapping the changing contours of the world economy*. SAGE Publications Ltd.

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, *72*(4), 1077-1106.

Ellen, I. G., & Ross, S. L. (2018). Race and the City.

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the association of American geographers*, *102*(3), 571-590.

Enright, M. J., Scott, E. E., & Chang, K. M. (2005). *Regional powerhouse: the greater Pearl River Delta and the rise of China*. Wiley.

Eurostat. (2018). Glossary: High-tech classification of manufacturing industries.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American planning association*, *76*(3), 265-294.

Fang, C., & Yu, D. (2017). Urban agglomeration: An evolving concept of an emerging phenomenon. *Landscape and urban planning*, *162*, 126-136.

Faggio, G., Silva, O., & Strange, W. C. (2017). Heterogeneous agglomeration. *Review of Economics and Statistics*, *99*(1), 80-94.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional studies*, *41*(5), 685-697.

Funderburg, R. G., & Boarnet, M. G. (2008). Agglomeration potential: the spatial scale of industry linkages in the Southern California economy. *Growth and Change*, *39*(1), 24-57.

Fu, W., Diez, J. R., & Schiller, D. (2012). Regional innovation systems within a transitional context: evolutionary comparison of the electronics industry in Shenzhen and Dongguan since the opening of China. *Journal of Economic Surveys*, *26*(3), 534-550.

Garcia-López, M. À., & Muñiz, I. (2013). Urban spatial structure, agglomeration economies, and economic growth in Barcelona: An intra-metropolitan perspective. *Papers in Regional Science*, *92*(3), 515-534.

Gao, S., Janowicz, K., & Couclelis, H. (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, *21*(3), 446-467.

Gezici, F., Walsh, B. Y., & Kacar, S. M. (2017). Regional and structural analysis of the manufacturing industry in Turkey. *The Annals of Regional Science*, *59*(1), 209-230.

Giuliano, G., Redfearn, C., Agarwal, A., & He, S. (2012). Network accessibility and employment centres. *Urban Studies*, *49*(1), 77-95.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of political economy*, *100*(6), 1126-1152.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*, 211-221.

Greunz, L. (2004). Industrial structure and innovation-evidence from European regions. *Journal of evolutionary economics*, *14*(5), 563-592.

Guillain, R., & Le Gallo, J. (2010). Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis. *Environment and Planning B: Planning and Design*, *37*(6), 961-981.

Guo, Z., Zhao, P., Senousi, A. M., Liu, X., & Mansourian, A. (2023). Exploring the structural characteristics of intra-urban shared freight network and their associations with socioeconomic status. *Travel Behaviour and Society*, *32*, 100576.

Harrison, B., Kelley, M. R., & Gant, J. (1996). Specialization versus diversity in local economies: The implications for innovative private-sector behavior. *Cityscape*, 61-93.

Hall, P. G., & Pain, K. (Eds.). (2006). *The polycentric metropolis: Learning from megacity regions in Europe*. Routledge.

Hausmann, R., & Hidalgo, C. (2010). Country diversification, product ubiquity, and economic divergence.

Hayter, R., & Han, S. S. (1998). Reflections on China's open policy towards foreign direct investment. *Regional Studies*, *32*(1), 1-16.

He, S. Y., Tao, S., Hou, Y., & Jiang, W. (2018). Mass Transit Railway, transit-oriented development and spatial justice: the competition for prime residential locations in Hong Kong since the 1980s. *Town Planning Review*, *89*(5), 467-494.

He, S. Y., Wu, D., Chen, H., Hou, Y., & Ng, M. K. (2020). New towns and the local agglomeration economy. *Habitat International*, *98*, 102153.

Henderson, V., Kuncoro, A., & Turner, M. (1995). Industrial development in cities. *Journal of political economy*, *103*(5), 1067-1090.

Hirschman, A. O. (1980). *National power and the structure of foreign trade* (Vol. 105). Univ of California Press.

Holmes, T. J. (1999). Localization of industry and vertical disintegration. *Review of Economics and Statistics*, *81*(2), 314-325.

Hu, F. Z., & Lin, G. C. (2011). Situating regional advantage in geographical political economy: Transformation of the state-owned enterprises in Guangzhou, China. *Geoforum*, *42*(6), 696-707.

Huang, D., Liu, Z., Zhao, X., & Zhao, P. (2017). Emerging polycentric megacity in China: An examination of employment subcenters and their influence on population distribution in Beijing. *Cities*, *69*, 36-45.

Hui, E. C., & Lam, M. C. (2005). A study of commuting patterns of new town residents in Hong Kong. *Habitat international*, *29*(3), 421-437.

Hui, E. C., Li, X., Chen, T., & Lang, W. (2018). Deciphering the spatial structure of China's megacity region: A new bay area—The Guangdong-Hong Kong-Macao Greater Bay Area in the making. *Cities*.

Hui, E. C., Li, X., Chen, T., & Lang, W. (2020). Deciphering the spatial structure of China's megacity region: A new bay area—The Guangdong-Hong Kong-Macao Greater Bay Area in the making. *Cities*, *105*, 102168.

Jacobs, J. (2016). *The economy of cities*. Vintage.

Keeling, D. J. (1995). Transport and the world city paradigm. *World cities in a world-system*, 115-131.

Kenessey, Z. (1987). The primary, secondary, tertiary and quaternary sectors of the economy. *Review of Income and Wealth*, *33*(4), 359-385.

Kerr, W. R., & Kominers, S. D. (2015). Agglomerative forces and cluster shapes. *Review of Economics and Statistics*, *97*(4), 877-899.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, *99*(3), 483-499.

Lee, H. S. (2009). The networkability of cities in the international air passenger flows 1992–2004. *Journal of Transport Geography*, *17*(3), 166-175.

Li, Z., Jiao, L., Zhang, B., Xu, G., & Liu, J. (2021). Understanding the pattern and mechanism of spatial concentration of urban land use, population and economic activities: A case study in Wuhan, China. *Geo-spatial Information Science*, *24*(4), 678-694.

Liu, B., Xue, D., & Tan, Y. (2019). Deciphering the Manufacturing Production Space in Global City-Regions of Developing Countries—a Case of Pearl River Delta, China. *Sustainability*, *11*(23), 6850.

Liu, H., Fan, J., Zhang, H., & Wang, F. (2020a). Dynamics of manufacturing industry and change of its spatial pattern in the Pearl River Delta urban agglomeration. *Prog. Geogr.*, *39*(2), 195-206.

Liu, J., Deng, Y., Wang, Y., Huang, H., Du, Q., & Ren, F. (2020b). Urban nighttime leisure space mapping with nighttime light images and POI data. *Remote Sensing*, *12*(3), 541.

Liu, K., Qiu, P., Gao, S., Lu, F., Jiang, J., & Yin, L. (2020c). Investigating urban metro stations as cognitive places in cities using points of interest. *Cities*, *97*, 102561.

Liu, X. (2019). Characterizing broken links on national and local expressways in Chinese city-regions. *Regional Studies*, *53*(8), 1137-1148.

Liu, X., Huang, J., Lai, J., Zhang, J., Senousi, A. M., & Zhao, P. (2021a). Analysis of urban agglomeration structure through spatial network and mobile phone data. *Transactions in GIS*, *25*(4), 1949-1969.

Liu, Y., Li, Z., Liu, Y., & Chen, H. (2015). Growth of rural migrant enclaves in Guangzhou, China: Agency, everyday practice and social mobility. *Urban Studies*, *52*(16), 3086-3105.

Liu, Z., Chen, X., Xu, W., Chen, Y., & Li, X. (2021b). Detecting industry clusters from the bottom up based on co-location patterns mining: A case study in Dongguan, China. *Environment and Planning B: Urban Analytics and City Science*, 2399808321991542.

Lloyd, A., & Cheshire, J. (2017). Deriving retail centre locations and catchments from geo-tagged Twitter data. *Computers, Environment and Urban Systems*, *61*, 108-118.

Nakamura, R., & Paul, C. J. M. (2019). Measuring agglomeration. In *Handbook of regional growth and development theories*. Edward Elgar Publishing.

Neal, Z., Derudder, B., & Liu, X. (2021). Using urban networks to gain new insight into old questions: Community, economy, bureaucracy. *Journal of Urban Affairs*, *43*(1), 2-15.

Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2018). Agents of structural change: The role of firms and entrepreneurs in regional diversification. *Economic Geography*, *94*(1), 23-48.

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

Ng, M. (2008). From government to governance? Politics of planning in the first decade of the Hong Kong Special Administrative Region. *Planning Theory & Practice*, *9*(2), 165-185.

Nissan, E., & Carter, G. (2010). Shannon Measure of Industrial Specialization at the State Level. *Southwestern Economic Review*, *37*, 29-45.

Norton, R. D., & Rees, J. (2007). The product cycle and the spatial decentralization of American manufacturing. *Regional Studies*, *41*(S1), S61-S71.

Macal, C. M., & North, M. J. (2005, December). Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005.* (pp. 14-pp). IEEE.

Malmberg, A. (1996). Industrial geography: agglomeration and local milieu. *Progress in Human Geography*, *20*(3), 392-403.

Malmberg, A., & Maskell, P. (2002). The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering. *Environment and Planning A: Economy and Space*, *34*(3), 429-449.

Marshall, A. (1890). Principals of economics Macmillan. *London (8th ed. Published in 1920)*.

McCann, P., & Acs, Z. J. (2011). Globalization: countries, cities and multinationals. *Regional studies*, *45*(1), 17-32.

Meijers, E. J., & Burger, M. J. (2017). Stretching the concept of 'borrowed size'. *Urban Studies*, *54*(1), 269-291.

Meijers, E. (2005). Polycentric urban regions and the quest for synergy: Is a network of cities more than the sum of the parts?. *Urban studies*, *42*(4), 765-781.

Meyer, S., Schiller, D., & Diez, J. R. (2012). The localization of electronics manufacturing in the Greater Pearl River Delta, China: Do global implants put down local roots?. *Applied Geography*, *32*(1), 119-129.

Mikhaylov, A. S., & Bolychev, O. N. (2015). Forms of transnational economic cooperation and integration in the Baltic Region. *International journal of economics and financial Issues*, *5*(2), 55-64.

Muñiz, I., Garcia-López, M. À., & Galindo, A. (2008). The effect of employment sub-centres on population density in Barcelona. *Urban Studies*, *45*(3), 627-649.

O'Donoghue, D., & Gleave, B. (2004). A note on methods for measuring industrial agglomeration. *Regional studies*, *38*(4), 419-427.

Overman, H. G., & Puga, D. (2010). *4. Labor Pooling as a Source of Agglomeration: An Empirical Investigation* (pp. 133-150). University of Chicago Press.

Parr, J. B. (2002). Missing elements in the analysis of agglomeration economies. *International Regional Science Review*, *25*(2), 151-168.

Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, *19*(4), 557-568.

Porter, M. (2003). The economic performance of regions. *Regional studies*, *37*(6-7), 549-578.

Rimmer, P. J. (1998). Transport and telecommunications among world cities. *Globalization and the world of large cities*, 433-470.

Ritzer, G., & Stepnisky, J. (2017). *Modern sociological theory*. Sage publications.

Rosenthal, S. S., & Strange, W. C. (2001). The determinants of agglomeration. *Journal of urban economics*, *50*(2), 191-229.

Roth, C., Kang, S. M., Batty, M., & Barthélemy, M. (2011). Structure of urban movements: polycentric activity and entangled hierarchical flows. *PloS one*, *6*(1), e15923.

Sassen, S. (1991). TheGlobalCity: New York, London, Tokyo. *Princeton, NJ*.

Scholl, T., & Brenner, T. (2016). Detecting spatial clustering using a firm-level cluster index. *Regional Studies*, *50*(6), 1054-1068.

Scott, A. J. (2019). City-regions reconsidered. *Environment and Planning A: Economy and Space*, *51*(3), 554-580.

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(3), 683-690.

Seto, K. C., & Kaufmann, R. K. (2003). Modeling the drivers of urban land use change in the Pearl River Delta, China: Integrating remote sensing with socioeconomic data. *Land Economics*, *79*(1), 106-121.

Strange, W. C. (2008). Urban agglomeration. *The new Palgrave dictionary of economics*, 1-5.

Storper, M., Kemeny, T., Makarem, N., & Osman, T. (2015). *The rise and fall of urban economies: Lessons from San Francisco and Los Angeles*. Stanford University Press.

Taylor, P. J. (2001). Specification of the world city network. *Geographical analysis*, *33*(2), 181-194.

Timmer, M. P., Miroudot, S., & de Vries, G. J. (2019). Functional specialisation in trade. *Journal of Economic Geography*, *19*(1), 1-30.

Walker, R., & Schafran, A. (2015). The strange case of the Bay Area. *Environment and Planning A*, *47*(1), 10-29.

Wang, J., Lee, K., & Kwan, M. P. (2018). Environmental influences on leisure-time physical inactivity in the US: An exploration of spatial non-stationarity. *ISPRS International Journal of Geo-Information*, *7*(4), 143.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236-244.

Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., ... & Tatem, A. J. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences*, *115*(14), 3529-3537.

White, M. J. (1999). Urban areas with decentralized employment: Theory and empirical work. *Handbook of regional and urban economics*, *3*, 1375-1412.

Van der Panne, G. (2004). Agglomeration externalities: Marshall versus jacobs. *Journal of evolutionary economics*, *14*(5), 593-604.

Van Soest, D. P., Gerking, S. D., & van Oort, F. G. (2002). *Knowledge externalities, agglomeration economies, and employment growth in Dutch cities*. Tilburg: Tilburg University.

Vogiatzoglou, K., & Tsekeris, T. (2013). Spatial agglomeration of manufacturing in Greece: sectoral c and determinants. *European Planning Studies*, *21*(12), 1853-1872.

Yeh, A. G. O. (2011). Hong Kong: the turning of the dragon head. *Planning Asian Cities: Risks and Resilience*, 180-200.

Yeh, A. (2012). Hong Kong: the turning of the dragon head. In *Planning Asian Cities* (pp. 192-212). Routledge.

Yeh, A. G., Liu, X., Xu, J., & Wu, M. (2020). The emergence and economic restructuring of two global super megacity-regions in China: comparing the Pearl River and Yangtze River Deltas. In *Handbook of Megacities and Megacity-Regions*. Edward Elgar Publishing.

Yuan, J., Zheng, Y., & Xie, X. (2012, August). Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 186-194).

Yu, H., & Peng, Z. R. (2020). The impacts of built environment on ridesourcing demand: A neighbourhood level analysis in Austin, Texas. *Urban Studies*, *57*(1), 152-175.

Yu, Z., & Liu, X. (2021). Urban agglomeration economies and their relationships to built environment and socio-demographic characteristics in Hong Kong. *Habitat International*, *117*, 102417.

Yu, Z., Xiao, Z., & Liu, X. (2022a). A data-driven perspective for sensing urban functional images: Place-based evidence in Hong Kong. *Habitat International*, *130*, 102707.

Yu, Z., Zu, J., Xu, Y., Chen, Y., & Liu, X. (2022b). Spatial and functional organizations of industrial agglomerations in China's Greater Bay Area. *Environment and Planning B: Urban Analytics and City Science*, *49*(7), 1995-2010.

Yu, Z., Zhu, X., & Liu, X. (2022c). Characterizing metro stations via urban function: Thematic evidence from transit-oriented development (TOD) in Hong Kong. *Journal of Transport Geography*, *99*, 103299.

Zhang, F., Ning, Y., & Lou, X. (2021a). The evolutionary mechanism of China's urban network from 1997 to 2015: An analysis of air passenger flows. *Cities*, *109*, 103005.

Zhang, J., Liu, X., & Senousi, A. M. (2021b). A multilayer mobility network approach to inferring urban structures using shared mobility and taxi data. *Transactions in GIS*, *25*(6), 2840-2865.