FLOW MATTERS: USING SPACE OF FLOWS TO UNDERSTAND URBAN DYNAMICS IN COMPLEX CITY

JUNWEI ZHANG

PhD

The Hong Kong Polytechnic University

2023

The Hong Kong Polytechnic University

Department of Land Surveying and Geo-Informatics

Flow Matters: Using Space of Flows to Understand Urban Dynamics in Complex City

Junwei Zhang

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

November 2022

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)

Junwei Zhang

Abstract

As technology has become an indispensable part of city, human and land have been connected with unprecedented frequency and resolution. The connectiveness captured by urban flows fuels new perspectives of quantifying how people behave and how city work. The present thesis investigates the modelling and quantification of multi-dimensional urban flows through numerical and multilayer network approach, and demonstrates the extracted knowledge can be used for enhancing comprehension of urban dynamics through flow-based paradigm.

Leveraging rich data captured from mobile devices, traditional public transport, shared mobility services, and volunteer geographic information, the framework of defining, extracting, integrating, and modelling urban flows in unified model is demonstrated. Bridging the latest theory and methods of network science, a shared mobility multiplex network and a temporal multiplex network are constructed, from which multilayer statistical feature and community structure results constitute to the limited knowledge on how new transport mode influence flow patterns of the traditional, and how multi-flow-induced urban structure may change over transport modes and time.

Along with chapters, a set of explicit metrics are developed and discussed for quantifying flow patterns. Some of these, such as multilayer degree and multiplex PageRank, the latest methods developed by network scientist to tackle the drawbacks of single-layer network analysis, are adopted on the urban flow models in this thesis. Moreover, diversity of spatial interaction (DSI) is a new metric defined and developed for the first time on quantifying diversity from flow data. We construct DSI by integrating multiple aspects of activity diversity being separately studied before. Its effectiveness is validated and further explored by intersecting with land characteristics data, offering powerful insights on revealing the positive and negative implications of flows for urban places. For urban vitality evaluation, latest metrics such as ridership variations, are integrated with the proposed flow diversity to develop a comprehensive framework of urban diversity, based on which the multiscale spatially varying relationship between diversity metrics and vitality is inferred and discussed. The use of spatial coefficients for profiling unique urban context is presented at the end.

As an interdisciplinary body of work conducted by a geography reseaecher, this thesis puts forward evidence on importance and effectiveness of multi-flows data analytics for fundamental questions in geography such as human-land relationship, and for latest topics in urban studies such as urban vitality and dynamic structure, substantiating the merits of numerical and multilayer network approach to urban flows in complex city.

Keywords: urban flows, complex city, spatial interaction, transport modes, urban diversity, urban vitality, dynamic structure

List of Papers

The following first-authored articles contribute to the research questions and extended discussions made for **Chapter 4-6** of this thesis, consisting of my sole efforts on research design, method development, and draft writing, but the achie vement would not be possible without the discussion and comments from the co-authors.

Paper I **Zhang, J.**, & Liu, X. (2022). Interaction diversity in geographical space: a novel index and its implications for urban development. *Annals of the American Association of Geographers* (Under review)

Paper II **Zhang, J.**, Liu, X., Tan, X., Jia, T., Senousi, A. M., Huang, J., Yin, L., & Zhang, F. (2021). Nighttime vitality and its relationship to urban diversity: An exploratory analysis in Shenzhen, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 309-322. https://doi.org/10.1109/JSTARS.2021.3130763

Paper III **Zhang, J.**, Liu, X., & Senousi, A. M. (2021). A multilayer mobility network approach to inferring urban structures using shared mobility and taxi data. *Transactions in GIS*, *25*(6), 2840-2865. https://doi.org/10.1111/tgis.12817

The following co-authored articles contribute to the methodology development and scope of this thesis. I contributed to the conceptualization, data processing, and draft revisions.

Paper IV Liu, X., Huang, J., Lai, J., **Zhang, J.**, Senousi, A. M., & Zhao, P. (2021). Analysis of urban agglomeration structure through spatial network and mobile phone data. *Transactions in GIS*, *25*(4). https://doi.org/10.1111/tgis.12755

Paper V Liu, X., Wu, J., Huang, J., **Zhang, J.**, Chen, B. Y., & Chen, A. (2021). Spatial-interaction network analysis of built environmental influence on daily public transport demand. *Journal of Transport Geography*, *92*, 102991. https://doi.org/10.1016/j.jtrangeo.2021.102991

Paper VI Senousi, A. M., Liu, X., **Zhang, J.**, Huang, J., & Shi, W. (2020). An empirical analysis of public transit networks using smart card data in Beijing, China. *Geocarto International*, 1-21. https://doi.org/10.1080/10106049.2020.1768594

The following co-authored articles are less relevant to the main topic of thesis but using similar data sources. I contributed to the data processing and conceptualization.

Paper VII Senousi, A. M., **Zhang, J.**, Shi, W., & Liu, X. (2021). A Proposed Framework for Identification of Indicators to Model High-Frequency Cities. *ISPRS International Journal of Geo-Information*, *10*(5), 317. https://doi.org/10.3390/ijgi10050317

Paper VIII Huang, J., Liu, X., Zhao, P., **Zhang, J.**, & Kwan, M. P. (2019). Interactions between bus, metro, and taxi use before and after the Chinese Spring Festival. *ISPRS International Journal of Geo-Information*, *8*(10), 445. https://doi.org/10.3390/ijgi8100445

“The result, therefore, of our present enquiry is, that we find no vestige of a beginning, - no prospect of an end.”

- James Hutton
Theory of the Earth (1788)

Acknowledgements

At the end of what has been a great chapter of my life, I am sincerely indebted to those who play a part in it. First and foremost, I would like to express the gratitude to my supervisor, Dr. Xintao Liu, who has been an excellent academic and life mentor to me - always offering patience and freedom to pursue curiosity and extend my limitation. I will carry his advice and spirit of scientific discovery with me for the new journeys. I am further grateful to my co-supervisor, Prof. Wenzhong Shi, for supporting me with his vision, expertise, and job opportunities in the early years that make my academic journey smoother. Thanks also to Prof. Zhilin Li, Prof. Bo Wu, Prof. Mengqian Lu, Prof. Shenghua Li, whom I may not have luck to work with at the moment but from whom I obtained valuable advice for applying Ph.D. program and career development. I have been fortunate to have comments from Prof. Michael Batty, Prof. Paul Longley, Prof. Chen Zhong, Prof. Ling Yin, Prof. Xiaolin Zhu, Prof. Tao Jia, and Prof Frank Yam, I'm humbled by the great work they do and for having short conversations with them to get inspired. Big thanks to my English teachers at PolyU, Joe Ching and Marshall Yin, for their passionate and fun classes on academic writing and presentation.

I would not have enjoyed the Ph.D. journey if without my lab members and friends, for having talks, coffees, dinners, and happy hiking hours together. I am lucky to have Jianwei Huang, Ahmad M. Senousi, Zijian Guo, Zidong Yu, Minglei Liao, Ningyezi Peng, Xinyu Gong, Yepeng Yao, and Xinyu Li. Special thanks to Wenzheng Fan and Xiaolin Zhou, for their weekly tutoring at the early stage to develop my professional skills and interests. Thanks to Christine Suen and Nora Ngai, for inviting me to join the innovation competition that indeed open my eyes to solve real-world problems by what I learned in academy, and for the inspiring talks and getting awarded at the end.

To my family who have always been there for me, my parents supporting me in their own different ways, my furry friends - Meme, Tara, and Tora offering me their small warm paw and cheek, and I'm blessed to have my wife – Echo Shang, not only for sharing the happiness, but for never giving me up in the darkest moments, for sharing the pain, sweat, and tears. She is the reason I started this journey and would be the light for the new horizons.

Junwei Zhang

Hong Kong, August 2022

Table of Contents

List of Abbreviations

AIC	Akaike Information Criterion
BCE	Before The Christian Era
BTH	Beijing-Tianjian-Hebei
CN	Local Condition Number
CPub/CTaxi	Consistency Index of Public Transport / Taxi Ridership
DOC	Denstity Of Catering
DSI	Diversity Of Spatial Interaction
FHV	For-Hired Vehicle
GAM	Generalized Additive Model
GPS	Global Positioning System
GWR	Geographically Weighted Regression
ICT	Information And Communication Technology
KDE	Kernel Density Estimation
LUM	Landuse Mixture
MAD	Median Absolute Deviation
MGWR	Multiscale Geographically Weighted Regression
ML	Machine Learning
NTL	Nighttime Light
NYC	New York City
OD	Orientation-Destination
OLS	Ordinary Least Squares
OSM	Openstreetmap
POI	Point-Of-Interest
RNR	Residential-Non-Residential Ratio
RSS	Residual Sum of Squares
SCD	Smart Card Data
TOD	Transit-Oriented Development
VDP	Variation Decomposition Proportions
VPub/VTaxi	Variability Index of Public Transport / Taxi Ridership

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Background and Motivation

City refers to a collection of settlement and way of living emerged around 5000 years ago, that is a human creation being ever-evolving but long-lasting. If looking humankind as a whole, city lives where people live. City adapts all the changes and survived when other human creation like cultures, empires, political regimes rise and fall. In 21st century, a lot of people including myself, a young boy born in China, have witnessed the wave of urbanization and globalization, during which small villages are transformed to cities, and connectivity of roads, railway, and airline make people be able to reach anywhere within few hours. It is magnificent to see how a single species (humankind) can make such transformation to the globe and its way of living, that motivates me to quest on what is the key to success of city and vibrant urban life? Alongside with urbanization, the time I grow up is also the time of Information Age (Castells, 1999). With advancement of Information and Communication Technology (ICT), I found that it doesn't even need hours to reach somewhere, but seconds and few clicks are enough to make me talk, see, or even feel someone/someplace. **Cities and people have been connected more than ever, that inspires my research interest and scope.**

Modern science was dominated by a mechanical view 'the worlds as a clockwork' for more than 200 hundred years. Human's great capability of creation make us have an imagination that everything is a product with fixed parts and rules, one typical example is Newton's gravity theory. Until the emergence of Einstein's relativity theory, chaos theory, and complexity science, scientists in many domains have shifted their 'view' of the studied entities and mechanisms to 'non-causality'. In machines rules are represented by 'ifs' and 'thens', but in reality many systems including city are consisting of uncountless number of 'ifs-thens'. Instead, city is viewed, studied, and operationalized as a complex system in contemporary urban and geography science. The way how a complex system present itself relies on the interactions of its elements. Simple rules and interactions at the bottom can lead to sophisticated function on the top, evidence of which can be found in many natural and human systems (Jiang & Yin, 2014; Bich, 2012). The complexity science reshapes my worldview and scientific thinking greatly, that urge me to comprehend a core feature of complex city, **flows**, on how it can be modelled to characterize urban places and urban life. On this quest, theories and methods are reviewed in **Chapter 2**, **setting the foundations** of this thesis on why flow is important for city, and how we can make sense of it using scientific methods.

Recent decades have witnessed urban big data to be widely available for research and application, such as smart card data (i.e., records of trip orientation and destination), mobile phone data (i.e., proxy of location and communication), and remote sensing images. Big data provides unprecedented resolution and volume on observing city and human activities, based on which research focus gradually shift from physical form of city (relatively static) to human-centric urban studies (more dynamic) (Liu et al., 2015). Driven by big data, a comprehensive set of new methods and tools have been improved or invented by computer and data scientist, known as The Fourth Paradigm (Tansley & Tolle, 2009). **Researchers from different domains are inevitably exploring and reflecting on how these big-data methodology can be used** in, such as Geography (Wu et al., 2016), Transportation (Tranos & Mack, 2019), and Urban Science (Batty, 2013a). Intersected with machine learning methods, geographers are not only able to extract objective from remote sensing images more efficiently (Maxwell et al., 2018), but also able to extract the semantics of places (Sui & Goodchild, 2011) or predicting travel flows (Tang et al., 2021). Such stream is summarized as GeoAI and Spatial Data Science (Janowicz et al., 2020). Intersected with Network Science, geographers are able to represent physical networks of city (Ter Wal & Boschma, 2009) as well as travel flows (Zhong et al., 2014). The network representations fill the methodological void of GIS to some extent, from which interactions and complexity can be captured, known as paradigm of Network Geography (Batty, 2005). **These new methodological developments have several implications for further studies on urban flow analytics in this thesis.** On one side, most machine learning (ML) models (except graph neural network) are not well suitable for representing flow data, because ML models emphasize more on the input (i.e., feature engineering) and output (e.g., classification and prediction), while complex interaction and structure captured by flow data are mostly aggregated in ML models. In addition, ML models are often criticized as 'black box' being lack of explainability. On another side, network analysis, although is suitable for urban flows, most existing work (including my early articles Paper IV, V, and VI) are relying on single-layer model that is insufficient for representing complex interaction of the real-world (Kivelä et al., 2014). Extensive research is required to bridge big data and inter-disciplinary methods on enhancing comprehension of urban flows in complex city (See more in **Chapter 2**, Section 2.2).

1.2. Research Questions and Objectives

The problem to be addressed by this thesis is the gap between the increasing awareness of complex city due to diverse flows and the limited comprehension and methods of urban flows and applications. More explanation importance of flows, complex cities, and limitation of current methods can refer to **Chapter 2**. The research questions emerged from the problem includes:

> **RQ1**: How to capture and model urban flows from big data of human activities?
>
> **RQ2**: What are human-land patterns in the new context of complex urban flows?
>
> **RQ3**: How the urban dynamics can be understood using urban flows patterns for guiding urban evaluation and policymaking?

The research objectives guided for the methodology design (**Chapter 3**) and case studies (**Chapter 4-6**) include:

> 1): To develop a framework of processing, integrating, and modelling multiple urban flow data using numerical and multilayer network models.
>
> 2): To develop metrics to quantify multi-facet patterns of flows in terms of versatility, variations, and diversity and their relationship to land characteristics.
>
> 3): To apply the quantified flow patterns to evaluate dynamics of urban vitality and urban structure.

With all objectives as a whole, this thesis contributes to contemporary mental image of complex urban flows for geographer (Martin, 2005) and urban planner (Batty, 2013b), addressing the inevitably intensified and diversified connectiveness at the moment and the future.

1.3. Synopsis of the Dissertation

- **Chapter 2:** In this chapter, we review historical and contemporary evidence on why flow is important for city. Complexity of flow is explained in terms of its high frequency and multiplicity. In geography and urban science, the shift from place-based to flow-based theory is discussed. On methods, the shift from classic gravity model to recent multilayer network is presented. At last, research challenges arose from the reviewed literatures are summarized.

- **Chapter 3:** In this chapter, the methods used in empirical studies of the thesis are organized by three domains: Spatial Data Science, Urban Science, and Network Science. Definitions, calculation processes, and rationales behind are elaborated in the methodological framework.

- **Chapter 4:** In this chapter, we present the study in New York city using multiplex network models and analysis. Six years taxi data and shared mobility data are integrated to investigate whether and how the emergence of shared mobility services influence on travel behaviour. Statistical and spatial analysis are performed on multiplex centralities and community detection.

- **Chapter 5:** In this chapter, we provide a new definition on interaction diversity related to geographical flows. We develop and test the metric using one-month taxi data in Shenzhen and intersect with build environment data for validation and interpretation. The positive and negative implications of flow diversity are discussed for several applications.

- **Chapter 6:** In this chapter, we advance the framework of urban vitality by including flow diversity, ridership diversity, built environment diversity, and night-time light remote sensing images. The spatially varying relationship between diversity metrics and vitality proxy are investigated using multiscale geographically weighted regression (MGWR). Application of model coefficients are further discussed on profiling distinct urban context.

In **Final Chapter** we draw the conclusion for the whole thesis, summarize the main contributions, and reflect on the promising future research directions.

Chapter 2. Literature review

2.1. *Flows of City, Flows for City*

2.1.1. Historical view on urban flow

The first great civilizations thrive next to river, as flow of water provides reliable resources for drinking, hygiene, and agricultural production. Dated back to 4000 BCE, Mesopotamia civilization emerged along Tigris and Euphrates rivers. Ancient Egypt grows around 3100 BCE along the Nile valley. In the east, great civilization appears around 1700 BCE along the Yellow River in China. All examples share the same indication that ‘flow’ is an intrinsic feature of city. While flow of water is neither the only factor that matters in human history, nor the central topic of this dissertation, but it is one first condition laying the foundation for other types of flows that shape society and city. The commonality inspires this study and motivates deeper exploration of what, why, and how flows are made of and for city.

Flows are not merely a feature of city, but vital for its liveness. Even tracing back to the ancient times, ‘flow’ can be in various forms to play critical role all along the history of city. How simple flow (such as water) generates diverse types of interaction (flows) can be evidently identified in Neolithic revolution, providing necessary pre-conditions for origin of city (Bairoch & Goertz, 1986). In this period, agriculture is conducted in scale, which makes food storage possible for the first time and leads to large population accumulation. Migration flows transform sparse communities into the larger territorial region when foragers became farmers. The magic happens when there is no need for everyone to search foods by themselves. Hands are free for making other products and trade began (i.e., flow of goods). Until present days, the abovementioned activities are still the most important functions of city. Dense population and agglomeration reduce transport costs for exchanging resources and ideas. Furthermore, this process (generation and benefit of flow) keeps being intensified in modern times (since 18th century), resulting in phenomenal urbanization in global scale. The evolution history of city and flow urges several features to be well considered in contemporary urban studies: complexity and multiplicity. This study is dedicated to this active and promising direction to develop new models and methods for urban flows.

2.1.2. Complexity of urban flows

Due to primitive state and limited productivity before industry resolution, percentage of people living in cities remains low for centuries (around 10% of total population). While with the onset of several waves of technological revolutions, unprecedented growth and rapid urbanization happens in recent two hundred years (more than 50% population move to cities). Machinery, electricity, and internet empower modern manufacturing, transportation, and communications. By its nature, new technologies are serving as boundary-breaking medium to make individuals, organizations, and places more connected than ever. Unlike classical geography representations, multiple types of flow are not simply different attributes overlaying together, but entangled each other (i.e., depend or influence), and happened across scales (e.g., intra-city and inter-city) and spaces (e.g., online and offline). Multi-flow in multi-space interact and evolve rapidly, diminishing classic understanding of distance, dynamics, and urban structure.

This paragraph elaborates more on complexity of urban flows, notably, the next paragraph highlights another aspect being less analysed in recent research, the multiplicity. Before 20th century, the mechanistic world view – 'the world as a clockwork' – dominates modern sciences on understanding social and natural universe. The machine metaphor is adopted on city (Portugali, 2021), in which the focus is on knowing all 'ifs' and 'thens' to control the mechanism. However, at the beginning of 20th century, the world view and scientific thinking are altered by Einsten's relativity, quantum theory, and more recently, complexity theory. All these theories have a common feature that differs from 'machine' is non-causality. To characterize flows in complex system by studying the 'connectiveness' is a particularly interested subject in complex science. A complex system is characterized by 'far from equilibrium', meaning that the changes never at rest. More importantly, high-level functions and patterns are organically grown from the bottom-up: order emerges from disorder through interaction. Fruitful evidence is reported that city suits such characterizations. If we look at urban morphologies (or costal line), they are messy but ordered and self-similar across spatial scales. Applying complexity 'thinking' on city, it is not hard to tell that origin of city is similar to emergence of life that simple interaction and structure repeats themselves (Courtat et al., 2011; Jiang & Yin, 2014). From classical point of view, such urban spaces with new technologies make more flows, while in turn, flows are changing the old way how we develop and make use of urban spaces. For example, more convenient transportation enables people to live far away from city centre and avoid high rents. Thereby a new paradigm has obtained growing attentions on complexity of city. The shift lies on that previous theorists regard city more as system with different components providing different functions, while recent researcher view that how city function as a whole is determined by the flows in the bottom. This debate and shift will be elaborated more in the literature review in next section.

To study complexity and diversity of modern city, we must make clear the multiplicity. Although we have witnessed dramatic urbanization in both 19th and 20th in North America, Europe, Japan and the 21st in China, Africa, and India, the characteristics are quite different in terms of what these cities rely on, how they work, and how they develop. Measuring multiplicity of city is thereby obtaining great attentions from urban researcher, for example, to study multilateral relations in politics (Ku, 1998), multi-level planning strategy (Ye & Björner, 2018), and social relations and identities (Verkuyten et al., 2019). Literally and generally, the term multiplicity depicts the state of being multiple. However, the meaning and impact of multiplicity can be beyond the simple overlapping of multi-factors. Growing evidence shows that city, as a complex system, its sub-systems are also interacting with each other, which means that multiplicity is not only 'multiple' but depicting 'interdependent' behaviour or relations (D'Agostino & Scala, 2014). For example, transportation system depends on communication infrastructure depending on electricity system and so on, each of which might already be well design to be robust (i.e., resilient for attack), while their interdependence may result in widespread function failures due to the failure of even one system (Buldyrev et al., 2010). This cascading failure is the typical example showing the facts that multiplicity is essential for survival of city.

In face of rapidly changing technologies and globalization, there are uncountable numbers of entities interacting and interdependent on each other. In urban and geography context, this study argues that the under-going influence of multiplicity on flows and spaces should be further investigated. The former represents the interaction by different infrastructures and relation in different scenarios, while the latter means such flows are no longer existing in physical urban space but across internet (Shaw & Yu, 2009) or social (Andris, 2016). Therefore, we can tell that studying multiplicity of urban flows are not simply overlaying data layers but considering fruitful connections across systems or spaces. Although fruitful works have been done on spatial or social flows using recently available individual-level big data, very limited research have proposed a holistic framework and model flow multiplicity across physical and virtual spaces. In some domains, data with multi-space information are extracted for same group of people to address this issue. For example, in social computing, computer and social scientist work together to understand how social interaction relates to spatial behaviour, resulting in applications such as recommendation system (Li et al., 2020). While the scope of multiplicity is not limited in social interaction, but is challenging classic notation of some geography principles. For example, how to extend classic spatial dependence into physical-virtual spaces? How to model interdependence of different spatial flows? This dissertation differentiates from previous work studying on flows in isolated manner, and to propose holistic framework and models to account for interdependence of people and places.

2.2. Theories And Methods of Multi-Flow Analytics

Understanding urban flow is paramount for a wide range of applications in city, including transport planning, social inclusion, and designing urban spaces for economic development. Theorizing and developing methods to analyse flow data have been a core subject for researchers. In this section, we review the scientific literatures on urban flows analytics, particularly focusing on theories and methods in different context and applications. Research gaps are discussed in face of the widely available urban sensing data and advanced methods in network science. This section contributes to construct framework and methods of this dissertation towards analysing multi-flow in multi-space for sustainable and vibrant city.

2.2.1. From place to flow theories

This section will not construct a new theory for urban flows but reflect on theoretical components in literatures to project new research directions towards evolving urban context. The selected literatures are particularly categorized as two streams: one is the location-based theories, and the other is the flow-based theories.

Flow is not an invention in modern age but exist ever since the onset of city. Put it simply, most human activities are flows by large in the form of interaction and relations. In Geography, classical view on human activity is generally dominated by Environmental Determinism (Peet, 1985), describing that physical environment plays a crucial role in determining social and cultural development (Wu et al., 2016). Following this principle, we could come up with the idea that space generates flow. A range of cases can be found in both ancient and modern cities. Locations near river are more attractive for migrant and trade flows due to the developed agriculture. In industrial cities, the separation and agglomeration of work and home facilities make it common to see prominent number of commuting flows every day (Yang et al., 2018). These typical examples are the supportive evidence for the location-based theory, setting the foundation for urban flows: Spatial heterogeneity of function, form, and demographics are the underlying forces to drive flows, in short, space makes flow.

Christaller's (1933) elaborates the 'space makes flow' thinking in the central place theory in economic geography. In its initiative, spatial patterns (e.g., size and number) of human settlements, towns, and cities are expected to be generalized by universal laws. Hereby 'central places', 'hinterlands', and their hierarchical relationship are introduced to describe that central place serve surrounding areas (i.e., hinterland) that are also serving their neighbourhoods. This theory explains spatial patterns well in various scenarios, for example, urban-rural structure in intra-city scale (Dutt, 1969), and city-town structure in regional scale (Taylor et al., 2010). However, central place theory is not adaptive for all, and major criticisms are on its simplified assumptions of how people behave and flow (e.g., homogenous spreading direction). Another problem lies on its key assumption on the dominant role of physical distance on the relationship between central place and its neighbours. In other words, nearer locations are most like to be the hinterlands of central place, because flow cost (e.g., travel time) is expected to be minimized. This is correct when the medium bearing the flow is largely constrained by physical distance, such as the limited mobility by foot or horse cargo. Spatial patterns induced by central place, therefore, often present a limit of influential range (scale), which might not be cases in modern cities where a location is not only related to its nearest neighbours (e.g., while people who live suburban can travel long distance to central area; A phone call can instantly travel to the other sphere of Earth). Fast transportation and internet communication significantly reduce or even remove the flow cost, in other words, space collapse (Marston et al., 2017; Tranos & Nijkamp, 2013).

These facts are hinting that a location theory like central place may not be sufficient to explain complexity of modern city where distance law and notation of space is being reshaped by widely used new technologies. Flow theories are introduced to present a new perspective: 'flow makes space'. One may feel confused on its grammatic structure as opposed to 'space makes flow' dominated by classical thinking of Environmental Determinism, while these two are indeed not contradictory but together describe a precise figure of contemporary city. This paragraph elaborates 'flow makes space' by the several important theories in geography and urban science.

Researcher in sociology and psychology might be some earliest pioneers accepting and studying the idea of 'flow makes space', such as social capital theory (Huggins et al., 2012). This is partly because that the studied 'space' in these domains is not obviously visible (like geographical space) but constructed by human relations and interactions. To theorize and analyse flow in geography space, it is necessary to make it clearer on the scope of this dissertation. Flow theory is not denying the objectivity of geographical space but to, in a highly connected society like nowadays, to delineate how location/place/city function and position differently due to the dynamical flow. With human interaction and flow, some places weigh more than others in the whole city system, underlying which flow-driven heterogeneity serve as the forces for newly emerged and ever evolved structures (e.g., social segregation and city hub). Evidence from different recent theories is reviewed in the following parts.

In economic geography, reflection on 'flow makes space' is triggered by the increasing need for understanding interurban processes alongside the fact that a city is no longer living by itself. The central flow theory proposed by Taylor et al. (2010) is the one typically towards such goal. Central flow distinguishes with the central place theory in terms of scale and role of flow: the former is depicting the town-ness process driving spatial organisation dependent on physical proximity from central place to neighbours (hinterland); central flow is linked to the city-ness process connecting locations/places/cities to each other (hinterworld) in where physical distance may not be primary constrain but how they interact (flow) matters. While in inter-city relations, distribution of different types of cities are less explainable through central place thinking, and even some places within city (e.g., wall street) are heavily weighted in global scale. Whereas from flow perspective, the intense, dynamic, and diverse flows (e.g., financial, transport, internet) seem to be more indicative on explaining regional and global agglomerations, and how those clusters places are serving rest of world. Based on such, researchers summarized the higher-level mechanism distinct between place and flow-based process: through central place (i.e., town-ness) process, city may never grow into a metropolitan, while through flows the mutual benefits among interacting cities work as strong facilitator for massive growth (Taylor & Derudder, 2015). This makes it clear on the importance of looking at the city from flow perspective. Furthermore, metropolitan cities like London, New York, and Hong Kong are more complex examples that these two processes could play simultaneously across sectors and spaces, considering the decent development of both physical (offline) and virtual (online) infrastructures.

Defining space has long been a core objective since the onset of the geography research. Therefore, similar grounds on 'flow makes space' are also found in recent geography research dedicated to the notation of 'space' in Information Age. In these studies, we see the natural exploration to extend geo-space to virtual space (Shaw & Yu, 2009; Davies, 2004; Castells, 2020); notably some unique characterizations of new spaces are also reported. First, notation of both spaces contains infrastructures serving as the foundation of human-related processes. Geographical (physical) space and its physical infrastructure such as roads, buildings, ports, etc. bear material flows and human movement. Cyber (virtual) space and its networking infrastructure such as routers and servers bear information flows and communication. However, flows in cyber space is less environment-constrained, which means the activity is less related to the distribution of the infrastructure itself. Comparing to physical space, flow in cyber space relatively has unlimited speed spinning over any spatial scales. If without flows, cyber space is nothing than dead servers, which is not the case in physical space that contains materials far richer than human-being. Second, relationship between human and environment constitutes to the major research content in both spaces. Traditional studies (geographical space) focus on human and land relationship. A wide of factors are constructed to study the coordination / mismatch between people (performance, well-being, and culture) and land (land-use, transport, urban forms) (Eagle et al., 2010; Hajrasoulih et al., 2018; Wu et al., 2016). While recent studies (virtual/hybrid space) take advantages of new observations (e.g., telecommunication, social media, online log) to study human-land relationship. Because it is undeniable that virtual space (and its infrastructure) cannot exist and perform without the physical space. Thereby many scholars prefer to assume the virtual's dependency on the physical, for example, mapping online perception to the urban space (Ghahramani et al., 2021). However, many other studies also reported the virtual structure (e.g., online communities) and dynamics (e.g., information transmission) that are independent from or even influence the physical space (Moss & Townsend, 2000; Dugundji et al., 2008; Althoff et al., 2017). Based on such evidence, although this dissertation agrees that virtual space depends on the existence of physical space, but more importantly, argue that virtual space is working by the unique ways that require further investigation. Based on flow thinking, it will be more interesting and challenging in future studies to not only address how flow, interaction, and relations form and function the virtual space, but also by turn, interdependent on the activity and environment in physical space.

2.2.2. From gravity model to multilayer network

Many human-induced flows are spatial data. Based on geographic coordinates of starting and ending locations, geographers developed metrics and models to quantify spatial interaction patterns (Mikkonen & Luoma, 1999; Griffith, 2007). While flows are not merely spatial data, but observations on topological relationship among entities on which network is often used as model representation (Ye & Liu, 2019). Among fruitful methods on flow analytics, this review is inevitably selective. Readers can further refer to Dickison et al. (2016) if interested on sociology, Crainic et al. (2022) on transportation, and Finn et al. (2019) on animal interaction behaviour. In this section, we particularly highlight the flow analytics in geography context, ranging from gravity-based interaction models, statistical models, to graph-based analysis. Furthermore, we introduce the advance developments in network science, 'multilayer network', as promising methodological components for multi-flows in multi-space.

Gravity model and its variation are typically popular to study flows across geographic space. Its basic form is largely determined by the Isaac Newton's theory of universal gravitation, that is $F_{ij} = km_i m_j (d_{ij})^{-2}$ where the gravitational force (F_{ij}) between entity i and entity j is positively related to the mass of two entities (m_i, m_j) and inversely proportional to the distance between the two entities (d_{ij}). Such form has natural analogues in urban context where number of human movements from place to place are inversely related to the travel distance and positively related to the population of each place (Ravenstein, 1885). In early 20th century, application of gravity model is further applied in regional science by Stewart (1941) and Huff (1963), investigating the gravitational parameter calibration and distance effect. Until 21st century, gravity model is still as feasible method to in delineating intensity and potential of flows in migration (Simini et al., 2012), trade (Silva & Nelson, 2012) and tourism (Morley et al., 2014). If the model is well fitted, the calibrated model can be used for either prediction of the flow volume, or depicting the relationship between flow and place characteristics. **However**, **several limitations of gravity model should be noted when applying on spatial interaction (flow) data**. **First**, gravity model performs better on aggregated data, both spatially and temporally. In other words, gravity model is more likely to be well fitted for zonal (or continental) and monthly (or yearly) flow patterns (Mikkonen & Luoma, 1999; Van Bergeijk & Brakman, 2010; Sen & Smith, 2012). On fine-grained spatial and temporal data, gravity model could be worse than simple statistical features to describe spatial interactions (Hilton et al., 2020; Hsu et al, 2021). **Second**, gravity model by nature is constrained in depicting the discrete choice of bilateral pairs. When constructing the interaction model, the 'discrete' means that there is no consideration on potential dependency and influence between flows, thereby each flow is treated as independent data entries in the gravity model. Another problem might lie on that the 'pairs' is always required in the model, such as pairs of cities. While in a general sense, spatial interaction patterns can be characterized in more diverse manners (will be discussed in the paragraphs) beyond pairs relationship. Third, gravity model is a deterministic physical model which is relatively static in dealing with changing scenarios in reality. In simple words, number of flows could be related to more than three factors, for example, it's common to see a decent list of environmental and behavioural factors influencing the traffic flows (Medina-Salgado et al., 2022). Flexibility of gravity model is not as good as linear regression model. But still, we conclude that gravity-form model is an important and useful alternative for flow analytics with awareness of its limitations.

Abovementioned spatial interaction models focus more on collective level flow patterns, **while human mobility research in recent years greatly contribute to induvial level flow analytics.** Comparing to spatial interaction models, mobility methods enrich how we characterize movements, in other words, a more complete view on the physics of individual flows. For example, GPS data in hour/minute/seconds reveal that power-law of the jump length (i.e., travel distance in a given period) widely exists in different travel modes such as walks, public transport, and taxi (Han et al., 2011). Furthermore, some use physical equations of particle motion to explain human movements, such as Brownian motion (Jiang et al., 2009), Lévy flight (Rhee et al., 2011). Commonality in human movements are also quantified in terms of spatial coverage (using Radius of gyration), frequently visited locations, and returning patterns (i.e., motifs) (Xu et al., 2018; Schneider et al., 2013). The general pattern of human movements help build applications on traffic forecasting and epidemic spreading models (Medina-Salgado et al., 2022). More importantly, developing metrics and universal models not only expand our knowledge on human movements, but also applicable for analyzing other urban flows, interactions, and relations (Barbosa et al., 2018). **From several ways these developments are distinct from spatial interaction (gravity) model. First**, what's observable is no longer limited in the OD pairs, but contains more dynamic and detailed flow patterns. Dimension and meaning of flow are richer. **Second**, gravity model focusses more on 'space' while new metrics and physical models are more on 'human'. The way that gravity model depicts the spatial flows heavily depend on the law of distance decay, as distance is a key component in the equation. Thereby gravity model explains more how space influence the flows (i.e., human behaviour). While in abovementioned human mobility metrics, most are depicting the characteristics of the flow itself, that enables research questions related to human and society. For example, how urban flows are related to the urban environment and demographic (e.g., poverty and race). **Third**, flow metrics have more opportunities to intersect with other methods and analysis, while gravity model has less. Specifically, flow metrics can serve as location-based factors that are easily intersected with other socio-environmental factors using spatial operations. Flow factors can be either integrated with other factors for urban evaluation (e.g., measuring urban vitality) (Zhang et al., 2021b), or fed into regression framework to support planning policies (Liu et al., 2021). However, metrics and physical models are aggregative to some degree. While graph (or network) is regarded as a natural way to represent the whole system's interaction/flow, from which useful diagnosis can also be conducted.

Graphs have paramount role in flow analytics, contributing to methodological development in many domains such as biology, economics, climatology, and sociology (Lewis, 2011; Barabási, 2013). The key is that graph represent interaction effectively. Whether it is human body, ecosystem, or a society as a whole, one cannot function without interaction between two cells in the body, two animals in the ecosystem, or two people in the society. Graphs represent all flows as a whole network by vertices (nodes) and edges (links), upon which a comprehensive set of tools can be used. Put it simply, first, classic graph methods evaluate topological, statistical, and dynamic features of node, edge, and network blocks. For example, node versatility metrics can indicate traffic conditions of road segments (Senousi et al., 2022), and community detection is for revealing agglomeration and urban hubs (Zhong et al., 2014). Second, in the latest development, graph and its analysis have been extended to multi-layered context and graph-based machine learning models. The former is dedicated to more realistic representation such as inter-connected social relations (Kivelä et al., 2014), and the latter is improving prediction and classification performance on flow-related context such as traffic prediction considering multi-modal flows (Wang et al., 2020).

However, a fundamental question with regard to graph-based urban studies concern how to transform real-world complex flow into an abstract graph. Since even same set of graph analysis could result in rather different knowledge of city, it is crucial to distinguish, in urban flow context, what entities are nodes, edges, and how to deal with time. **Therefore, this part particularly summarizes the related literatures into two categories: space as nodes and human as nodes**. **First**, the 'space as nodes' frameworks treat location or area at graph nodes among which interactions and relations donate the edges (Barthélemy, 2011). Types edges between locations are diverse, such as mobile communication (Eagle et al., 2010), trade (De Benedictis & Tajoli, 2011), individual movement (Liu et al., 2014) and transport transit (Ding et al., 2019). There are clear advantages in such construction that real-world urban networks can be projected in the graph model realistically, on which diagnosis results can be mapped in the geographic space in the direct manner. It enables easy visualization and computation of flow-induced features, which are normally used for validating cohesive and friendly urban design for transport, planning, and well-being policies (Batty, 2013b). **Second**, the 'human as nodes' frameworks treat induvial or group as graph nodes among which edges are donated by people's spatial behaviour, social relations, or similarities defined by additional information. Related methods are largely benefited from increasingly feasible human-generated data that contains both geographic locations and fruitful semantic information (Watts & Strogatz, 1998; Grabowicz et al., 2014; Kwan et al., 2015). In this sense, graphs are used to represent social network, interests' similarity, and information transmission. Unlike sociology studies, geography researchers overlay diagnosis of human nodes with spatial context exposed (Wang et al., 2011; Andris, 2016; Barbosa et al., 2018; Ye & Liu, 2019; Dang et al., 2019), for understanding social segregation (Xu et al., 2019), work-life balance (Renne et al., 2016), and urban diversity (Zhang et al., 2021a).

Although both space- and human- graph have shown its effectiveness of modelling urban flows, most studies focus on single type of graph or treat multiple flows separately. **While increasing evidence have suggested that methods without considering multiple flows and their relationship may fail to explain complex urban interactions.** For example, dependency between spatial mobility and social interaction is arbitrary in different context (Baldassare, 1978; Stehlé et al., 2013; Ai et al., 2019). On one side, ICT-powered social interaction (e.g., phone call) technically can flow across the globe, but frequent contacts are geographically close to some what (Eagle et al., 2010). On another side, distance effect on social interaction could be weak for migration worker who outside the resident country and still contact family members lived very far. In this case critical nodes in social graph is dominating behaviour in spatial graph. A similar case exists in research collaborations that are more driven by social connections, furthermore, even determine where a researcher is travelling to give a seminar (Kivelä et al., 2014). Towards modelling different flow in a more integrative manner, hereby we review recent development of multilayer graphs to shed light on the urban flow analytics.

Early conceptualization of multilayer networks is led by sociology researchers, as complex types of links are widely existing in social network studies (Vasilyeva et al., 2021). For same set of people (i.e., nodes), one can interact with another as colleagues and friends at the same time. Similar complex interactions are also reported in animal migration (Albery et al., 2021) and disease transmission (Finn et al., 2019). In urban context, multi-flows analytics is familiar to transportation researchers who studies multi-modal travel behaviour (e.g., bus, vehicle, bike) across same set of places (Aleta et al., 2017). Although network analysis has long been adopted in these areas, methods on multiple flows are not mature in early studies. For example, multilayer problems are investigated by using modified representations based on classic network, such as attaching different strengths for edges (Barret et al., 2004), constructing bipartite networks (Breiger, 1974), and constructing networks across time (Holme & Saramäki, 2012). Only until recently (Kivelä et al., 2014; D'Agostino & Scala, 2014), multilayer network is formally defined to provide clear mathematical notions, diagnosis metrics, generative models, and analysis tools on dynamics and structure. Multilayer networks extend monoplex (i.e., single-layer) graph $G = (V, E)$ to $M = (V_M, E_M, V, L)$ which containing nodes and edges across different layers (L). Such extension enables flexible representation on multiple flows, and a growing number of extended analysis methods, such as multiplex node centralities (Kivelä et al., 2014; Domenico et al., 2015b), edge versatility (Solé-Ribalta et al., 2016), community extraction across layers (Jeub et al., 2017), and multilayer visualization (McGee et al., 2019). **Few studies have applied multilayer networks in urban context, and existing ones primarily focus on transport systems or social networks.** Buldyrev et al. (2010) for the first time empirically uncover the potential cascading failure using multi-modal transport network, on which the hidden interdependence is the main risk contributor. Comparting the single-layer transport network, further analysis showed that multi-modal transport network is more correlated to population density in term of degree distribution (Gu & Wang, 2022). Similarly, mobility behaviours quantified in multilayer networks are reported as significant features for simulating and predicting travel demands (Ma et al., 2016; Strano et al., 2015; Aleta et al., 2017). Considering time in layers, multilayer diagnosis can provide valuable insights on the dynamics of urban flows (Ducruet, 2017). Relying on community detection in multilayer network, one can investigate the specific structure of how grain disasters impact on provincial economy (Qu et al., 2022), and how shared mobility services impact on travellers choices regarding to traditional taxi (Zhang et al., 2021b). Based on a spatial-social network, Hristova et al. (2016) developed novel metrics measure how urban places can attract different types of people.

2.2.3. Contributions from GIS

Geographical Information Science (GIS) also plays key roles in understanding complex urban flows mainly from two aspects. **First**, the emergence of dynamic urban big data provides unpredented resolution for integrating space and time (Kwan & Neutens, 2014; Chen et al., 2014). Decades ago, spatial data such as travel surveys and land parcels was relatively sparse in the time dimension, limiting the temporal analysis in a screenshots manner. While time is indispensable component in GIS analysis to understand the underlying processes of city (Claramunt & Thériault, 1995). Several achievements have been made for representing human activities more closely, for example, the space-time prism (Miller 1992; Neutens et al., 2008), and space-time trajectories (Forghani et al., 2020). Another major achievement in the context of urban big data is addressing the importance of volunteered geographic information (Goodchild, 2007). Citizens with mobile devices are seamlessly collecting surrounding information, namely citizens as sensors. The perspective provides the legitimacy of aggregating human activities features to understand the whole city (the bottom-up approaches). This stream of studies affects the approaches throughout the whole thesis to drive urban knowledge from individuals.

Second, GIS researchers included ‘semantics’ as a key component in the analysis of space and human activities. In early years, semantics mainly focus on the spatial aspect, theorizing the spatial relations of geographical objects as well as the processes within (Claramunt & Theriault, 1996 June). Typical approaches for representing spatial relations include topological spatial analysis (Jiang & Claramunt, 2004) and space syntax (Hillier et al., 1976; Jiang et al. 2000). The graph-based representation captures all connections in the same model, enabling the analysis of semantics and spatial features to be conducted in an integrative way. The semantics in GIS analysis has become ever richer nowadays, including the human semantics derived from geo-located social networking data (Andris, 2016), and the urban semantics derived from point-of-interests and human activities data (Gao et al., 2017). These new semantics impose the opportunities to enrich the GIS analysis to understand, beyond the geographical surface, how people interact and perceive places, and how urban places are actually functioned.

Third, the importance of geographical meaning has been raised in recent analysis of human activities and spatial flows, as the graph representation itself focus more on the topological features. Geographical meanings of human activities and flows are indispensable, such as direction of movement, distance, and spatial heterogeneity of places. Two studies in GIS domain are identified to address this issue (Wang et al., 2021; Wang et al., 2023), and they are relevant to the case study conducted in the Chapter 5 of the thesis to derive locational indicator from urban flows. Locational indicators are useful in quantitative geographical analysis to enhance location-based policymaking. Many studies have adopted network analysis to induce locational indicators for urban and transport planning, such as number of flows (Cats et al., 2015; Sun et al., 2016), entropy-weighted flows (Xia et al., 2019), flow ratio (Xu et al., 2017), and centrality-based index (Senousi et al., 2020; Liu et al., 2021). However, network metrics are more representative for the characteristics in topological space, which may ignore or lose fruitful geographical meanings. New location-based metrics addressing this fundamental issue has appeared in recent literatures, for example I-index (Wang et al., 2021) and X-index (Wang et al., 2023). I-index depicts the irreplaceability of location (destination) as a function of travel distance and flow volume from the orientations. A place is irreplaceable when there are many OD flows with both large flow volume and long distance. The X-index measures centrality of location (destination) as a function of flow volume and flow directions. The value of X-index is determined when the xth angle (ranked by rose diagram) exceeds at least a number of in-flows. The flow analytics that the thesis is trying to explore is inspired by both abovementioned three major contributions from the GIS domain: namely the use of high frequency space-time individual data to understand the city from the bottom-up, the use of topological analysis to capture of spatial semantics, and the thorough consideration of geographical meanings embedded in human activities and urban flows.

2.3. Summary of Research Challenges

In this review we discussed how flow theory and analytics are obtaining more attention in urban studies, at the intersection of high-frequency urban sensing data and perspective of complex science. On theoretical side we cover the conceptual frameworks delineate the relationship between space and flow: place theory weighs more on spatial dependence and distance effect across geographical space, while flow theory argue that it is the flow itself shaping the space we live. One methodological side, we have witnessed significant contributions from geographer in early years, and from inter-disciplinary researchers in recent years such as physicist, sociologist, and complex scientist. In this context, a growing number of date representation, metrics, and models are developed to allow the comprehensive study of the topic.

Whilst increasing evidence have suggested that we live a highly connected city with abundant new infrastructure diminishing the interaction friction across geographical space. It is urgent to reflect the research challenges from geography and urban perspective, on how complex flows, relations, and interactions can contribute to better understanding of city. One can find more details in the bellow:

2.3.1. On GIS

- **Interaction and relations** are challenging representations in Geography Information System that are originally designed for geometrics, in other words, GIS is for location not for flows. The challenges lie in several aspects:

  1) interaction and relations are not a local property of location or the location per se, but a flow (or connection) between pairs of locations. How to represent flows and extract metrics to reflect characteristics of location requires

extensive work; 2) interaction and relations are intrinsically dynamic, while traditional GIS method is not adaptive to capture evolvement of flow in temporal dimension.

- **Dependence** in GIS is largely replying on proximity in Euclidean space. It is useful to explain many spatial patterns such as density of population or human settlements. While growing evidence are challenging traditional notion of spatial dependence due to intensifying flows generated by new technologies. For example, multi-modal transportation systems are making distant places more accessible, resulting in that 'time' is a more representative metric for spatial dependence rather than distance. Furthermore, communication systems have dramatically shift physical flows to virtual flows. In such manner, how people are connected is not limited by physical distance. For example, people in Covid-19 can easily work from home and join meetings with different groups via online tools. The multi-flow context will also challenge how we define the segregation of people (i.e., some are more physically active while others are virtually).

- **Multiplicity** in contemporary GIS should not limited in data layers. GIS has great initiatives and advantages in representing different aspects of entities of real world into digital manner. While it is not sufficient to represent and handle increasing complex activities and relations in city nowadays. As mentioned in previous section, this dissertation aims to include inter-dependence between different systems as well to enhance GIS embrace the multiplicity. Data of different systems (e.g., multi-modal transportation) will be integrated in the model construction phase, instead of analysing each and then aggregate or overlay the results.

2.3.2. On Urban Science

- **Co-evolution** describes places and flows are not separately produced, but as organizations that are intrinsically co-developed: In the bottom-up process, different levels of stakeholders interact with each other and lead to emergence of city. In the top-down process, places (e.g., function and configuration) bears the way and potential of interaction. The challenge lies on that these two processes are evolving simultaneously, modelling such process requires research framework adopting multi-flows perspective and data to integrate place and flows more closely.

- **Urban Structure** has rather rich meaning in urban studies when studying spatial arrangement in socio-economic context, such as economic agglomeration, transport-oriented development zone, and social segregation. In this sense, it is widely agreed that there is no definite and conclusive determination of boundaries in urban structure. Although different data can be involved for deriving boundaries of interest, it is challenging to compare or integrate different structures when multiple variables are involved. Thereby further research is required to find inter-relationship of different data and develop integrative data model at the original places, then the derived urban structure can represent multiple aspects of urban activities. The multi-flow study is the promising research direction toward such purpose and provide better solution than traditional methods that overlay aggregative demographic attributes to spatial entities.

- **Urban diversity, vitality, and sustainability** are the key initiatives in recent urban studies and practices. Developing quantitative measurement towards these concepts obtain great attentions in recent studies to sustain benefits and success of city. Although fruitful urban big data has been used to quantify the key element, diversity, there is no study measuring urban diversity from the interaction (i.e., flow) perspective. The benefit and challenge of flow-based diversity rely on that multi-flow provide observations on diversity spinning over scales and entities type, breaking the boundary where diversity of a place is often measured by its surrounding composition. On the basis of new diversity measurement based on flow, we need to further investigate how diversity is linking to vitality, which is key concept relating to prosperity and sustainability of city. These research challenges are not the first time proposed here, but this dissertation crafts the research methods from flow data, addressing the future trend of understanding places and individual in a more connected manner.

Chapter 3. Methodology

This chapter develops the methodology of this thesis for modelling space of flows, combining knowledge and methods from three domains: Urban Science (Section 3.1), Spatial Data Science (Section 3.2), and Network Science (Section 3.3). We explain why these domains are important for urban flow analytics at the beginning of each section (Figure 1). Specific methods and tasks related to the appended papers are demonstrated in Figure 2, from which we can see Urban Science mainly contributes to the domain knowledge on urban issues, Spatial Data Science contributes to Flow Pattern Analysis, and Network Science contributes to Data Models and Feature Extraction.

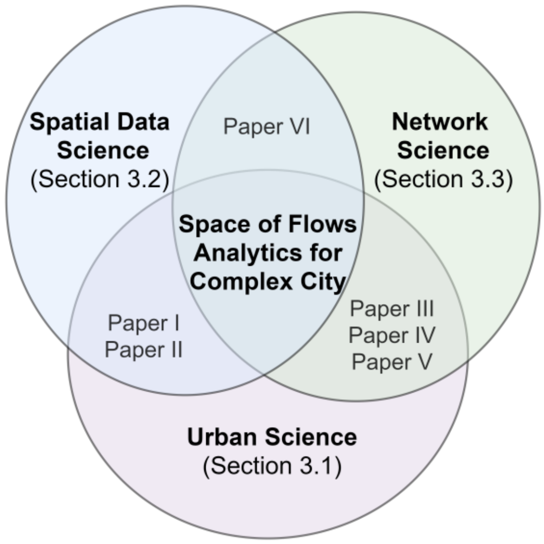


Figure 1. Domain knowledge of this thesis.

Same colors are used in the following figure with detailed methodological components.

The empirical studies of this thesis are surrounding a keyword, flow analytics, in the context of urban and transport geography. The theories and methods reviewed in **Chapter 2** explained why urban flows is important in the past and weight more in the current, and how inter-disciplinary research contributes to methodological development on flow analytics. The rest of empirical studies (**Chapter 4-6**) are the latest works done towards this direction, integrating classic data (e.g., public transport) with newly emerged data (e.g., shared mobility), and combing classic analysis methods (e.g., statistical and spatial regression) with advance network analytics (e.g., multilayer network). All empirical studies can fall into the framework shown in Figure 1, being conducted by three main stages. First, spatial-temporal mobility data need to be processed as OD matrix (a typical data format of pairwise flow, interaction, or relations), from which we need to define the unit of interest in particular study for further determining the conceptual linkages between flow data and network models. Second, based on the processed flow data models, various types of technics are used for extracting quantitative measurements to characterize flows, statistically, spatially, temporally, and topologically. Third, the extracted patterns (e.g., metrics, model coefficients, visualization) will be overlayed with other urban data for further analysis and discussion, such as socio-economic data and landuse. In this stage deeper insights can be obtained for urban-social applications.

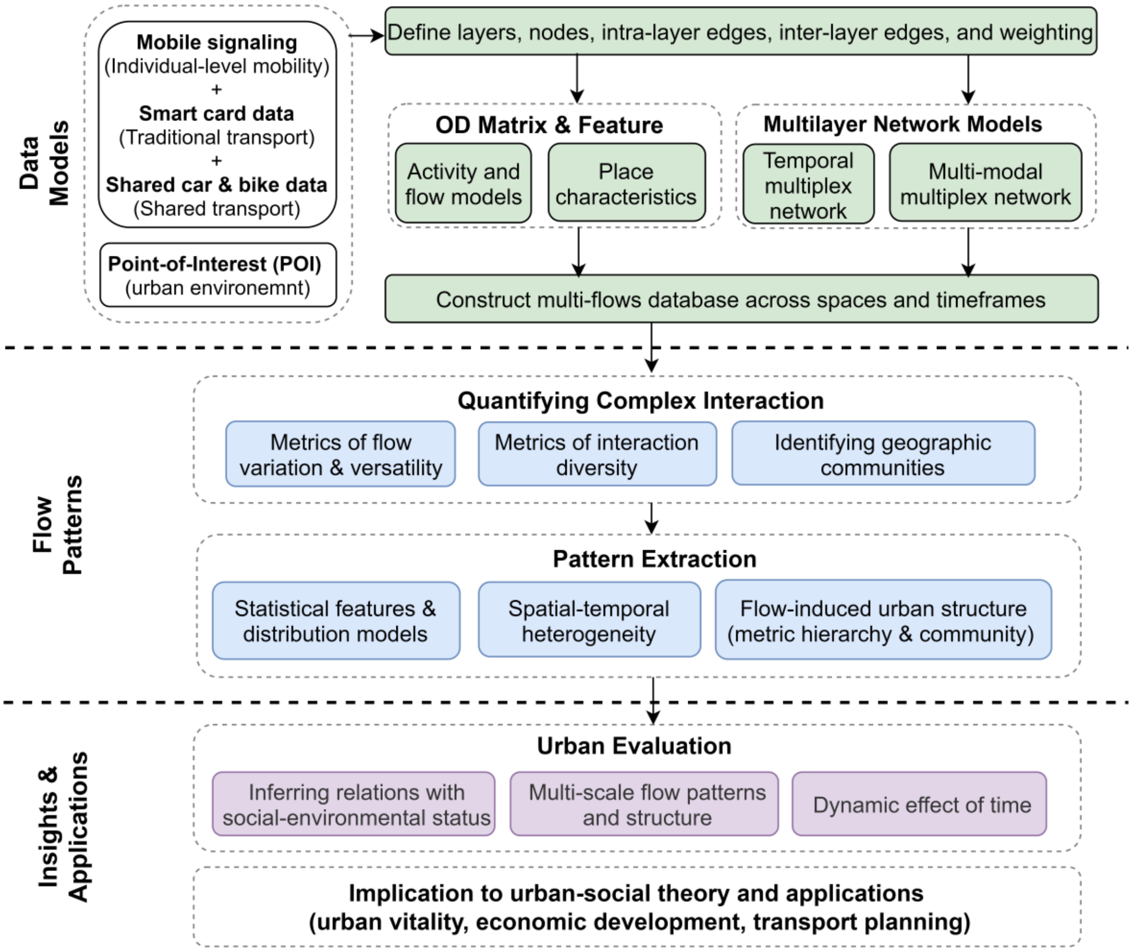


Figure 2. Methodology framework on space of flows analytics.

Although this thesis proposes that urban flows research is at the intersection of three domains, it should be noted that case studies presented in **Chapter 4-6** have their own preferences in the method design. **Chapter 4** focuses more network and urban science, constructing a multilayer mobility network to investigate the potential shift of travel patterns and flow-induced urban structure during the period of emergence of shared mobility services. **Chapter 5** focuses more spatial data and urban science, developing a new metric to measure interaction diversity from taxi flow data. **Chapter 6** extends research context to urban vitality, investigating how interaction diversity together with built environment diversity and ridership diversity are associated with the intensity of night-time vitality.

The computational complexity of the metrics introduced and especially the proposed flow diversity metric (DSI) has been tested using the hardware: CPU –i7 12700H, RAM – 16gb, DRIVE – SSD, ENV – python 3. Given a study area with 5000 grids, the calculation time of all urban science metrics are within 0.1 seconds and present a linear pattern between the data size and time. In the same area, given 7 days of hourly time series data, the calculation of ridership diversity is within 37 seconds and present a linear pattern as well. Given a network with 2 thousand nodes and 100 thousand edges, the calculation time of DSI is within 15 seconds, presenting a bilinear pattern between the number of nodes and time.

3.1. Urban Science

Research of urban science is often interdisciplinary, such as combining knowledge from sociology, transportation, environmental, and GIS to solve urban issues. Contemporary urban science is largly data- and computational-driven, exploiting potential of big earth data and massive human-generated contents to validate, delineate, and improve urban theories and practices (Batty, 2013a; Engin et al., 2020). Computational paradigm also motivates the methodological development of this thesis at large. In the following parts we introduce the methods used in case study chapters, with a focus on quantitative measurements for the topics of urban environment, dynamics, and urban structure.

3.1.1. Urban Environment

Recent developments on quantifying urban environment have largely benefited from fruitful types of urban sensing data (Liu et al., 2015; Kharrazi et al., 2016). In empirical studies of this thesis (**Paper I**, **Paper II**, **and Paper VI**), POI data is selected as major source to quantify urban environments for several reasons. First, spatial resolution and coverage of POI data is satisfactory in most big cities (e.g., as in Shenzhen city, China), which is continuously maintained by internet and map companies and widely used in other research (Jiang et al., 2015; Liu et al., 2016; Yue et al., 2017; Huang et al., 2019). Second, POI benefit from its ‘Point’ geometry type, thereby its attributes (e.g., usage type) can be easily aggregated into flexible spatial scales and units of interest (Niu et al., 2017). Third, increasing evidence showed that POI-derived metrics are related to human activity-related metrics (Maat et al., 2005; Sardari Sayyar & Marcus, 2011; Mouratidis, 2019), therefore is useful in the context of investigating how physical environment of urban places may influence urban flows (**Paper II**).

Diversity of the built environment has prominent role in urban vitality evaluation (Montgomery, 1998; Sung et al., 2013; Zeng et al., 2018). Places with diverse urban context attract travellers with a wider range of purposes (Yue et al., 2017). In the urban vitality study, this research includes three built environment metrics: entropy-based landuse mixture (LUM) (Frank et al., 2010), ratio of residential–non-residential (RNR), and density of catering (DOC). LUM metric is calculated as equation 1:

$$\mathrm{LUM} = -1\left(\frac{\Sigma_{i=1}^{n} pi * \ln(pi)}{\ln(n)}\right) \quad (1)$$

where pi is the proportion of the POI of category I in the spatial unit, and n refers to the total amount of all POI categories (totally 9 in our dataset). The higher the LUM, the higher degree of land use mixture it indicates.

RNR is also calculated from amount of POI categories but only limited to residential and others:

$$\mathrm{RNR} = 1 - \left|\frac{R_i - NonR_i}{R_i + NonR_i}\right| \quad (2)$$

where R_i is the proportion of residential category, and $NonR_i$ is the proportion of non-residential category. The integrated value, RNR, reflect relative amount between these two category: High RNR indicates residentials and others have similar amounts in the spatial unit.

Growing literatures have linked catering density to urban vitality (Long & Huang, 2017; Ye et al., 2018). The rationale behind is that catering are often built prior to other businesses as a basic need, thereby indicating there are people and economic activities contracted in the area. DOC metric is simply obtained as equation 3:

$$DOC_i = N(Catering\ POI)_i \quad (3)$$

where DOC in grid i is the number of catering POIs. $i \in \{1, 2, \dots, 9009\}$.

3.1.2. Urban Dynamics

City and citizens are changing so rapidly, thereby it is urgent to adopt new data, methods, and perspective to provide a holistic view of dynamics of city (Senousi et al., 2021). In past decades urban dynamics research focus on depicting landscape changes (Solon, 2009) and human mobility intensities (Yue et al., 2017). However, night-time activities and human mobility variations are less reported due to lack of feasible data and metrics. Furthermore, dynamics perspective has also been adapted in urban structures, definitions of which is no longer limited in the fixed boundaries led by administrative or natural powers but actual division of space is largely being influenced by human interaction. The methods introduced in this section link to these three directions.

To proxy night vitality, this study utilized gap-filled NTL radiance extracted from Daily Lunar BRDF-Adjusted Nighttime Lights dataset (Black Marble - VNP46A2). The 3 x 3–pixel grid was used to process the NTL radiance over days to reduce the value error caused by geographic mismatch of NTL imageries over days (Román et al., 2018). Although Black Marble products have been corrected using state-of-art strategy, errors are still reported due to lower accuracy in nighttime cloud detection (Wang et al., 2021). Therefore, this study further extract the maximum value of each week as vitality proxy for each pixel. The final vitality proxy based on the processed NTL value represents the best observed magnitude of night vitality across city locations.

To depict human mobility variations, two indices are developed with referring to (Zhong et al., 2016; Sulis et al., 2018). Mobility variations are useful metrics for further applications, such as constituting urban vitality indicator. In more specific context such as urban transport, places with non-regular ridership variations are more capable of attracting travellers with diverse travel purposes, which collectively make the place more vibrant. In this thesis, mobility variations are measured by Variability (V) and Consistency (C), where V is the day-to-day temporal variation of ridership volume and C is the within-day nonregularities.

For each spatial unit and time window, we calculate the variability of public transport and taxi transit as:

$$\mathrm{V} = \frac{\Sigma_i^n \Sigma_j^n (1 - Corr(TS_i, TS_j))}{(n * (n-1)/2)} \tag{4}$$

where variability index V is calculated based on time-series of travel flow volumes (TS_i,); numerator is the summation of dissimilarity (1 − Corr) in pairs of time series volumes (TS_i, TS_j), that is then standardized by the total number of pairs; n is the number of TS to be compared, in our weekly model, equal to 7 (days). Pearson correlation is used as our Corr() function. A higher V means more non-regular temporal changes from day to day.

Consistency is another metric of ridership diversity, being donated by extreme irregular within-day flow volume. High consistency value, practically, means more hourly peaks hinting the diverse use of lands. This metric is calculated by median absolute deviation (MAD) as equation 5:

$$\mathrm{C} = \sum_{j=1}^{N} \left(\frac{\sum_{i=1}^{n} \left| \frac{OutD_i}{MAD} \right|}{n} \right)_j \tag{5}$$

where C is the consistency metrics obtained by a N days time-series (7 days is set in this thesis). For day j, n is the number of outliers determined by MAD, and $OutD_i$ is the volume deviation from the ith outlier to the median volume. High C value at the end indicates big difference between the peak flows and the normal flows.

Urban structure, in traditions, is the spatial arrangement of units attributed to socio-economic or environmental features (Rodrigue, 2020). While recently, the view of urban structure has been extended by enormous observations on individuals' mobility. Spatial structure inferred from urban flows may mismatch or agree on the pre-designed urban boundaries, for example, using graph analysis on the mobility network (Sarkar et al., 2017; Zhang et al., 2018).

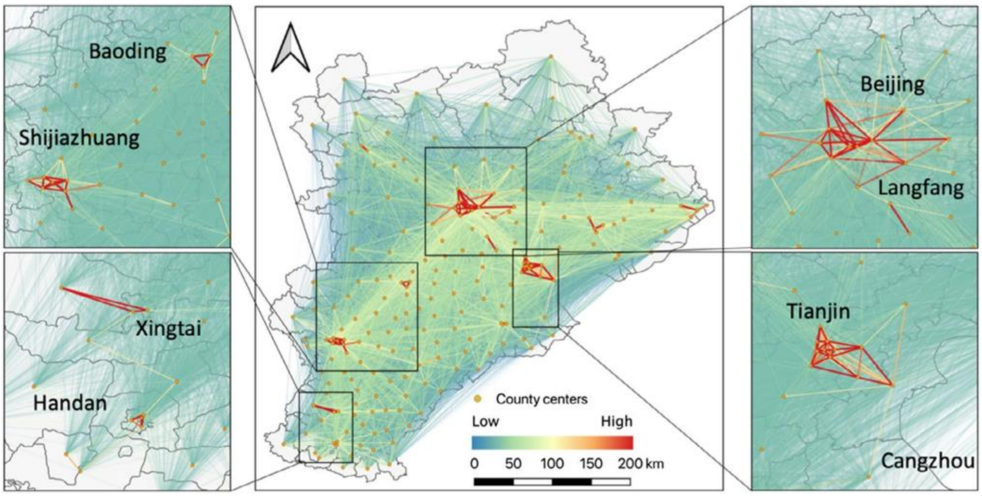


Figure 3. Spatial flow patterns in the Beijing–Tianjin–Hebei region

(Source: **Paper IV**).

These types of new data and methods bring new insights, because the inferred structure are produced by a bottom-up process, in other words, human flows-induced spatial structure are self-organized. In reality, the choices of frequency and direction of travel flows, are linked to the heterogeneity of many aspects of urban configuration. For example, resident and work buildings are concentrated in certain areas (static aspect of configuration), at the same time, nature of work (e.g., factory workers, IT staffs) also influence the distance of travel (dynamic aspect of configuration). The rationale behind the methodology inferring urban structure from urban flows is that such complexity of spatial-social configuration can be reflected by just looking deeper on the ways of flows are generated. One can see the power of flow-based method for urban structure study from just simply visualizing the travel flows. Different clusters and flow intensity emerge on Figure 3. Beyond visualization, other methods are elaborated in following sections.

3.2. Spatial Data Science

Big data imposes new challenges on urban flows analytics, specifically the complexity comes from three aspects. First, the large scale and huge volume of spatial big data require methods of summarizing main trend of data, that referring to statistical models. Second, the second-order nature of big flow data (e.g., OD matrix) should be well considered in the method design, rather than pre-processing in an aggregation manner (e.g., merge to time series). This part refers to our new development of diversity metric for urban flow data. Third, the ‘spatial’ nature of big flow data is also an indispensable aspect, as the patterns extracted from flows are intrinsically related to the spatial location (and its configuration). This part refers to methods for investigating spatially varying relationship. With all three parts as a whole, methods introduced below aims to contribute to the domain knowledge of spatial data mining, especially for urban flows.

3.2.1. Statistical models

In order to understand flow patterns, some studies have focused on the statistical characterization of flow-induced networks. For example, in human mobility network, statistical models have revealed important findings such as power-law (Gonzalez et al., 2008), and small-world properties (Watts & Strogatz, 1998). This thesis includes statistical models with two considerations. **First**, in high level, statistical models work as an indispensable lens to summarize collective behavior no matter from small data or big data (Chen et al., 2016). A rich body of scientific literatures across domains provide solid background for interpreting statistical results. **Second**, although the power-law paradigm has been a popular way to explain and quantify some urban flows, many studies have argued such law (or statistical distribution) is not always true and could still depend on transport mode (Han et al., 2011) and geographical scale (Alessandretti et al., 2020).

Table 1. Common statistical distribution for fitting human mobility patterns.

Statistical Model	**Equation**	**Parameters**
Weibull	$\frac{k}{\lambda}\left(\frac{k}{\lambda}\right)^{k-1} e^{-(d/\lambda)^k}$	k and λ
Gamma	$\frac{\beta^{\alpha} d^{\alpha-1} e^{-\beta d}}{\Gamma(\alpha)}$	α and β
Exponential	$\lambda e^{-\lambda d}$	λ
Power Law	$(d + d_0)^{-\beta}$	d_0 and β
Lognormal	$\frac{1}{d} \cdot \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(lnd - \mu)^2}{2\sigma^2}\right)$	μ and σ

For example, researchers found other statistical models such as using Weibull, Gamma, and Lognormal functions (Table 1), also have decent goodness of fit for various types of movements (Plötz et al., 2017; Kou & Cai, 2019). Experiments in Paper III of this thesis extend this open initiative to multi-flows context, inferring the best fitted distribution for the centrality diagnosis in multiplex mobility network. The results contribute to knowledge on urban flow patterns and the commuting behavior in the era of mixed shared mobility and traditional transportation (**Chapter 4**).

3.2.2. Composite indicator

In mathematics, composite indicator refers to the form of combining (or aggregating) a set of indicators, often being used to summarize complex and multi-dimensional problems. Constructing composite indicator is a popular methodology in urban and geography research, such as living planet index in environmental study (Loh et al., 2005), wellbeing index (Cummins et al., 2003) in social study, and accessibility index based on urban morphology data (Sevtsuk et al., 2016). First, popularity of composite indicator benefits from its explainability at large. Factors of composite indicator often have clear physical meaning (e.g., building height, block density, income), and selection of which often is guided by a theoretical and policy-driven framework such as transit-oriented development (TOD). Second, from GIS perspective, overlaying spatial data according to location or neighborhood is an essential initiative in the origin of GIS theories and tools, thereby many spatial data available nowadays make the process of constructing composite indicator easier.

This thesis is dedicated to urban flow analysis for advancing understanding of urban dynamics. One large contribution of this work is developing a metric to project the dynamic nature of urban flows in understanding an essential concept of urban evaluation: diversity. Awareness of importance of urban diversity has ever growing in contemporary research and society, as its association with innovation (Vormann, 2015), inclusion (Thompson, 2020), resilience (Loh et al., 2005), and vitality (Jacobs, 1961; Kang et al., 2021) of city. However, existing urban diversity metrics are mostly relying on static data, such as POI-induced landuse mixture. An essential feature of complex city, urban flows, is rarely linked to urban diversity.

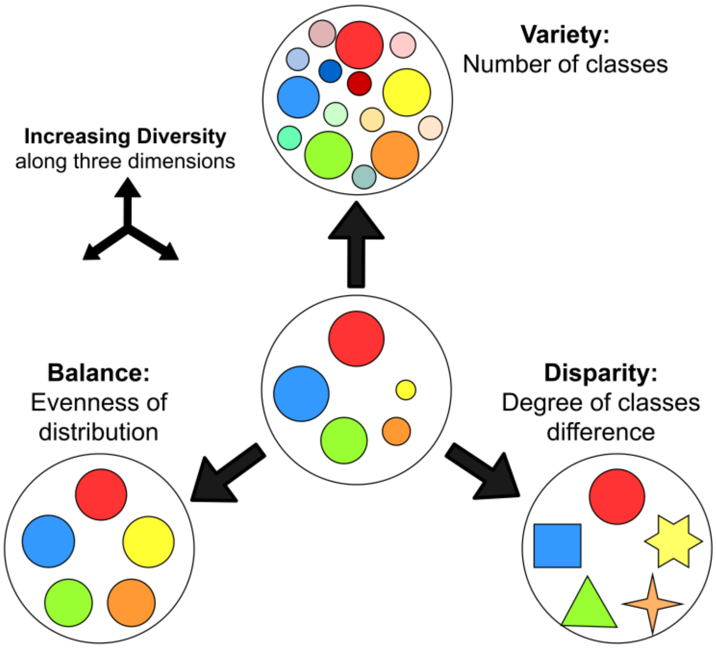


Figure 4. Semantic Framework of Diversity
(Adapted from: Leydesdorff et al., 2019).

More explicitly, for the first time we develop a metric to quantify interaction diversity as indicating how a location can attract flows in a diverse manner (See more results in **Chapter 5**). This diversity of spatial interaction (DSI) is inspired by insightful discussion on diversity in scientometrics and informatics fields studying on evaluation of scientific literatures. A latest framework (Figure 4) on measuring journal citation diversity is modified in the spatial context to combine three essential aspects of diversity (Rao, 1982; Rafols & Meyer, 2010; Leydesdorff et al., 2019). Variety means the richness of different classes in the population, Balance means whether there is equity in the number of each classes, and Disparity means how classes differ from each other considering their intrinsic attribute. Three aspects complement to each other to reduce the blindness of one-side diversity evaluation. In geographical context, the detailed metric development, evaluation, and empirical analysis are demonstrated in **Chapter 5**.

$$DIV = Varity * Balance * Disparity \quad (6)$$

3.2.3. Multiscale geographically weighted regression

Geographically Weighted Regression (GWR) is a typical method for revealing spatially varying relationship. This thesis adopted an improved GWR, multiscale geographically weighted regression (MGWR) in the case study, as MGWR outperforms global regression model by inferring a local relationship that varies from location to location for more implications. MGWR is similar to GWR in terms of both capabilities to infer spatially varying relationship. While the new feature as opposed to GWR is the flexibility allowing different scales to be used for different explanatory variables to fit significant coefficients, in other words, not only the relationship vary from location to location, but for each location and variable, the spatial range of borrowing data into model fitting and explanation vary as well (Fotheringham et al., 2017). One can refers to **Chapter 6** for more discussions on the results. According to several research, such flexible scales enable MGWR to handle the collinearity issues better, from which the overall model performance can be improved (Murakami et al., 2018; Fotheringham et al., 2019), outperforming the fixed scale used in GWR. In the results section, the comparison between MGWR and other models is provided, showing its effectiveness for the purpose of urban vitality evaluation as well.

The general form of MGWR model is presented as equation 7:

$$y_i = \sum_{j=0}^{m} \beta_{bwj}(u_i v_i) x_{ij} + \varepsilon_i \tag{7}$$

where x_{ij} is the jth explanatory variable of observation i at location $(u_i v_i)$, $\beta_{bwj}(u_i v_i)$ is the coefficient of jth variable inferred by using the bwj bandwidth, ε_i is the error term, and y_i is the response variable.

In GWR-series models, the bandwidth (or scale) is the size of the moving kernel for including data points for estimating model parameters. To infer flexible scales in MGWR, the whole model is regarded as a generalized additive model (GAM) (Rigby & Stasinopoulos, 2005) for the calibration process, upon which the back-fitting algorithm is adopted.

Therefore, the $\beta_{bwj} x_i$ of MGWR is the jth additive term f_i of a GAM :

$$y = \sum_{j=0}^{m} f_i + \varepsilon \tag{8}$$

At the beginning of calibration process, initial values of model parameters are provided by fitting a traditional GWR model as approximate estimation. Then an iteration process is required to refine the model parameters under different bandwidth.

This process is conducted on the form of GAM, where f_i, is treated like the varying variable and other terms are treated like constants. In the core, the coefficient β_j and the optimal bandwidth bwj are indeed estimated by GWR model. The iteration process will end when all term f_i have been inferred for their coefficients and bandwidths. It should be noted that in higher level there is another iteration process resulting in multiple GAM models to produce what we see multiple bandwidths at the end. In order to stop the whole iteration, a smooth function (SOC-f) is used to compare the multiple GWR models generated during the process, and the SOC-f is suggested to be smaller than 10^{-5}.

3.3. Network Science

Network science is a field of mathematician and physicists, studying on mathematical notion and diagnosis of graph (i.e., an abstraction of relations by nodes and edges). This field has provided a wide range of useful methods to urban and geography studies. In decades ago, graphs have shown it effectiveness in modelling static city networks (Batty, 2005). In contemporary research, graphs are constructed in a more dynamic context, being donated by human interaction and urban flow across spaces and time (**Paper III, Paper V**). Network science is associated with the view of complexity, under which a system's function is beyond the simple addition of parts. In this thesis, graphs are regarded as promising methodological components responding to the challenges raised in complex cities (Batty & Cheshire, 2011; Aleta et al., 2017; Portugali, 2021). This part introduces the graph models, metrics, and analysis used for the discussion on urban dynamics and structure.

3.3.1. Constructing multilayer metwork

Prior to understanding multilayer network models, it is reasonable to explore a simple graph. The general form of a monoplex (single-layer) network is $G = (V, E)$, where V is the set of nodes and $E \subseteq V \times V$ is the set of edges connecting each pair of nodes. In a preliminary study (**Paper V**), we extract spatial trajectory from mobile phone data to construct the flow matrix between orientation and destination zones in the BTH region (Figure 5). The spatial interaction network is then donated by the flow matrix, in which zones are represented as nodes (N), and travel volumes adjust the weights (w) of edges (E), resulting in $G = (N, E, w)$. It is disgustable from static network that mobility network is dynamic in terms of its nodes (e.g., some places may opt out traffic flows in certain period) and edges (e.g., flow volume and associated weights could change in every minute). Thereby we may assume the network structure and its characteristics may change through time, which is essentially useful to monitor how city works (Zhong et al., 2014; Sarkar et al., 2017; Zhang et al., 2018). Single-layer mobility network can be easily developed from other data with coordinates.

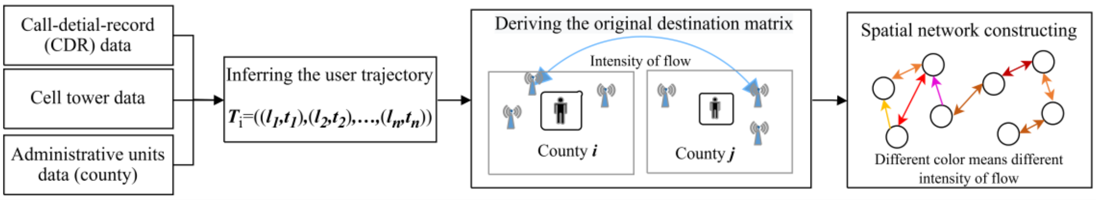
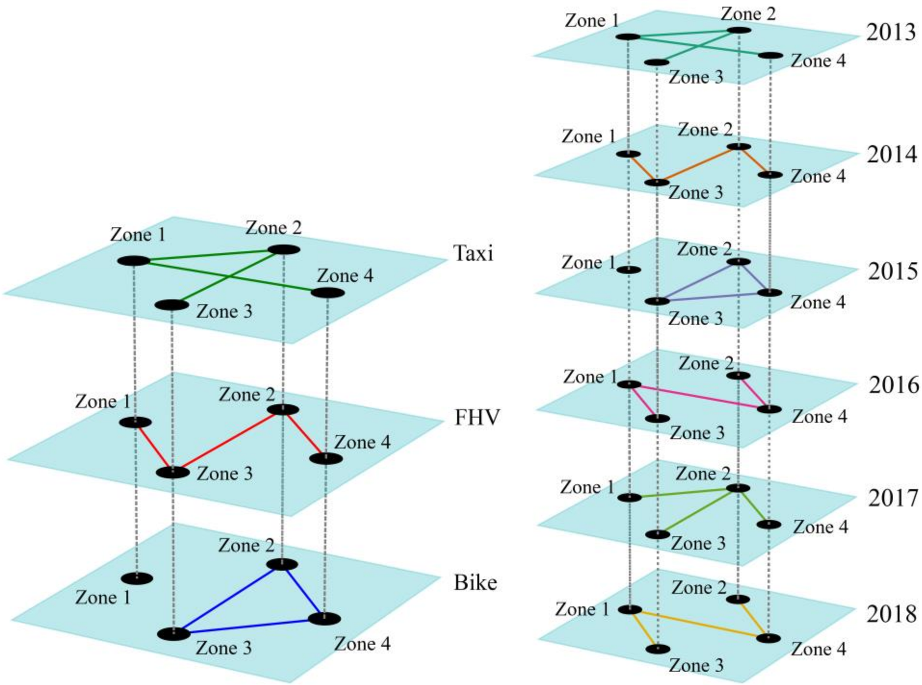


Figure 5. Constructing single-layer mobility network from mobile phone data (Source: **Paper IV**).

For a multilayer travel network, it is a natural extension of above notation. By referring to Kivelä et al. (2014), the general form of a multilayer network can be represented as $M = (V_M, E_M, V, L)$ by adding layers to the previous definition $G = (N, E)$. The node links on a single layer are the intralayer edges, and node links flying over different layers are interlayer edges. There can be multiple aspects of layers in the set L, where L = {L_1, L_2, … L_d}. An aspect of layers, for example, can represent transport modes or different time frames. In **Chapter 4** and **Paper III**, we construct the multiplex networks (a particular model) to integrate shared mobility flows and traditional taxi flows in the unified model. To achieve this, aspects in the general multilayer notion is set to 1 (Nicosia et al., 2013; Cardillo et al., 2013; Yildirimoglu & Kim, 2018), resulting in L = {L1}. Consequently, in a set of mobility networks $\{(V_\alpha, E_\alpha)\}_{\alpha=1}^{\beta}$, different layers can share the same set of nodes $V_\alpha = V_\beta$ for all α *and* β. Inter-layer links are the edges from node to its counterparts of other layers. This representation is especially in the context of multi-flows of city, in which different flows (represented by layers) exist in the same set of geographical locations (Figure 6).



(a) (b)

Figure 6. Construction of multiplex network models:

(a) Layers defined by modes; (b) Layers defined by time. Note that every single layer is a directed network donated by travel flows, and intralayer edges are weighted by flow volume; We use a common way to determine weight of interlayer edges in transport networks, namely all equal to 1 and nodes connected to its counterpart on other layers. The multiplex network here is mainly used for representing the multiplicity among same set of locations.

3.3.2. Characterizing multilayer network

This part focuses on introducing the centrality metrics in traffic flow context (**Paper III** and **Paper VI**). For a comprehensive review on multilayer network diagnosis, one can refer to Kivelä et al. (2014) and Interdonato et al. (2020). Centrality is one of the essential features reflecting how each node connect to the whole system (Ding et al., 2009; Agryzkov et al., 2016; Jia et al., 2019). In mobility network, high centrality nodes could indicate a transportation hub bearing high travel volumes from decent number of directions, that could both face attacking risk and prosperity. Among centrality metrics, degree and PageRank is selected as indicators to analyse urban dynamics.

For node i, multilayer degree can be simply obtained by the summation of degree k_i^α of multiple layers. Such method is effective but is mostly used in model without interlayer links. Instead, the case studies of the thesis use multiplex network as general form, considering multiplicity nature of place better and including interlayer to nodes counterparts. In this sense, degree summation is not suitable while a new variate of multilayer degree is found in De Domenico et al. (2015a):

$$k_i = M^{i\alpha}_{j\beta} U^{\beta}_{\alpha} u^j \qquad (9)$$

where k_i is the multilayer degree of node *i*, $M^{i\alpha}_{j\beta}$ is the adjacency matrix depicting the pairwise flow between node *i* on layer α and node *j* on layer β, u^j is a first-order tensor with entries equal to 1, and $U^{\beta}_{\alpha} = u_{\alpha} u^{\beta}$ is a second-order tensor with all entries equal to 1.

PageRank reveal node importance with more consideration on flows of influential neighbours. This definition, to some extent, suit the context of travel flow analysis very well (Wen, 2015; Xu et al., 2017), in which transfer behaviours are normal and relying on influential nodes (e.g., a station accessible to multiple lines). This metric is originally proposed by the founder of Google (Page et al., 1999), and then applied in many domains such as biology (Yu et al., 2017) and urban flows (Zhou & Qiu, 2018). While extension of PageRank in multilayer context is still at its infancy (Battiston et al., 2014). A viable solution can be found in Halu et al. (2013) to develop PageRank beyond single-layer network. However, the metric is initially designed for a two-layer network, based on which De Domenico et al. (2015b) further develop PageRank in a more general context, known as multiplex PageRank. This thesis relies on this solution to infer the structures in mobility network.

Multiplex PageRank utilized an essential concept of original version, that is the random walk algorithm to sample nodes for centrality evaluation. What determines the walk behaviour is the transition matrix as show in equation 11. Then the Multiplex PageRank centrality of node *i* can be calculated as equation 10:

$$\omega_i = \Omega_{i\alpha} u^{\alpha} = \sum_{\alpha=1}^{L} \Omega_{i\alpha} \qquad (10)$$

where ω_i is the aggregated PageRank centrality of node i, and $\Omega_{i\alpha}$ is the eigenvector of tensor $R^{i\alpha}_{j\beta}$.

$$R^{i\alpha}_{j\beta} = \tau T^{i\alpha}_{j\beta} + \frac{(1-\tau)}{NL} u^{i\alpha}_{j\beta} \qquad (11)$$

where τ is the walking rate being set as fixed value (e.g., 0.85), $T_{j\beta}^{i\alpha}$ is the transition tensor containing the jumping probabilities enabling the walk can be proceeded across layers, N is the size of nodes set, L is the number of layers, and $u_{j\beta}^{i\alpha}$ is a 4th-order tensor with all entries equal to 1.

Multilayer centralities then enable the investigation of versatile locations with considering multiple travel flows, which can be checked layer by layer or in an aggregative way. We illustrate and discuss more in **Chapter 4**.

3.3.3. Multiplex community detection

Community in network refers to that nodes can be seem as belonging to different groups according to connection patterns. This is like 'Clusters' in the general machine learning context, meaning that nodes have strong links within its community but weak links to other communities. In single-layer network, community detection method has been well developed and tested using empirical data, for example, revealing different travel preferences (Zhong et al., 2014; Liu et al., 2015).

The process of community detection is to divide network nodes into groups, which generally aims for a balanced resulting of maximizing between-group distances and minimizing within-group node distances (Grünwald & Grunwald, 2007). One widely used metric for evaluation is modularity (Newman & Girvan, 2004). Through iterations of trying different ways of nodes grouping, a large modularity is expected to be found. Although community detection has been used in mobility networks (Zhong et al., 2014), the representation is still relying on single-layer network, which suffers from difficulty of comparing community results of different network in a direct manner. Instead, this thesis adopted a recent developed algorithm, multiplex-Infomap algorithm (De Domenico et al., 2015a) enabling direct comparison that being used to investigate urban dynamics from the varying community structure perspectives. The multiplex modularity is calculated as equation 12:

$$Q_{multilayer} = \frac{1}{2\mu}\sum_{ijsr}\left[\left(A_{ijs} - \gamma_s \frac{k_{is}k_{js}}{2m_s}\right)\delta_{sr} + \delta_{ij}\omega\right]\delta(g_{is}, g_{js}) \tag{12}$$

where A_{ijs} is the edge weight between node i and j on layer s; k is calculated by adding up all edge weights of a node; k*is* represents the total weighted flow volumes of node i on layer s; kjs is similar to above; δ is the Kronecker delta function; $k_{is} = \sum_j A_{ijs}$; $\mu = \frac{1}{2}\sum_{jr} k_{jr}$; $m_s = \frac{1}{2}\sum_{ij} A_{ijs}$; gis will be the community label given to node i on layer s; γ_s is a resolution parameter being suggested to 1 by default; and ω is the weight for interlayer coupling that range from 0 to 1, we set as 1 for the multiplex networks.

A community detected in the end is equivalent to a cluster of nodes with similar interaction patterns in the whole network, practically, could means travel flow patterns inward and outward a location. The community detection mainly results in a community label for each node (location). More importantly, a set of labels can be generated for a node in multiplex network, associating with which community a node belongs to on different layers. In this sense, the dynamics of interaction/structure now can be easily investigated across layers that are specially constructed for comparison of interests, such as spinning over different time. A benefit of constructing multiplex network from urban flows data is that we can then project the community labels from network space to geographical space based on the spatial attributes of nodes. We discussed more how these are useful for understanding urban dynamics in **Chapter 4**.

Chapter 4. A Multiplex Network Approach for Profiling Urban Dynamics

4.1. Motivation

Urban structures attributed to different spatial-socio factors has been an important topic for geographers and urban planners. While in recent decades, feasibility of fine-grained data on urban flows has dramatically reshaping the way we look at urban structure, that is now more dynamic and activity-based, rather than fixed or static urban boundaries (Zhong et al., 2014; Sarkar et al., 2017; Zhang et al., 2018; Yildirimoglu & Kim, 2018). However, there are several limitations in existing literatures. **First**, different mobility data could tell different stories on urban structure, as the travel behaviour linked to transport modes originally differs. Thereby single source data is insufficient and biased to some extent. **Second,** although community detection of mobility network has been applied in inferring urban structure, the comparison of community across different network is not efficient. This is similar to compare result of clustering analysis: the cluster labels themselves have no meaning, but only for distinguishing on cluster from another in current experiment. The cluster label (e.g., cluster no.2) may not indicate the same meaning for another minorly adjusted experiment (e.g., also has cluster no.2 but the indication may be totally different). This community comparison problem, particularly, can be solved by technics in multilayer network. **Third**, although characteristics of human mobility are widely reported and accepted, such as scale-free properties, the characteristics of multilayer mobility network is less reported. Particularly, the influence of shared mobility has not been studied using multilayer network.

4.2. Research Questions and Methods

This study attempts to investigate the following three questions:

- What are the travel patterns in the multilayer mobility network consisting of shared mobility for traditional taxis?
- How flow-induced communities are spatially arranged and varying in the different layers (i.e., transport modes and years) of a multilayer network?
- Whether and how emergence of shared mobility influences the travel behaviour from place to place?

To answer those questions, we collect 6 years (2013 to 2018) trip records of traditional taxi, for-hired vehicle (FHV), and shared bike in New York City (NYC) (Figure 7). During this period the shared mobility services emerge and gradually take decent market shares. To investigate multi-flows behaviour, we construct two types of multilayer network (Figure 6). First, a temporal multiplex network is constructed using traditional taxi data, in which each year serve as a layer. $M_{Time} = (V_t, E_t, L_t)$, where $L_t = \{2013, 2014, \ldots 2018\}$, and $E_t \subseteq V_t \times V_t$. Second, a multimodal network is constructed to combine shared mobility and traditional taxi data in the last year (2018) when share mobility services already take an important part in daily transit (according to the number of ridership). $M_{Mode} = (V_m, E_m, L_m)$, where $L_m = \{Taxi, FHV, Bike\}$ and $E_m \subseteq V_m \times V_m$. The multilayer network centralities and community detection are conducted on above two models, and whose layer architecture enable us to profile and compare the characteristics across years or transport modes. These results are overlayed with the inferred statistical models and other spatial data for further discussions.

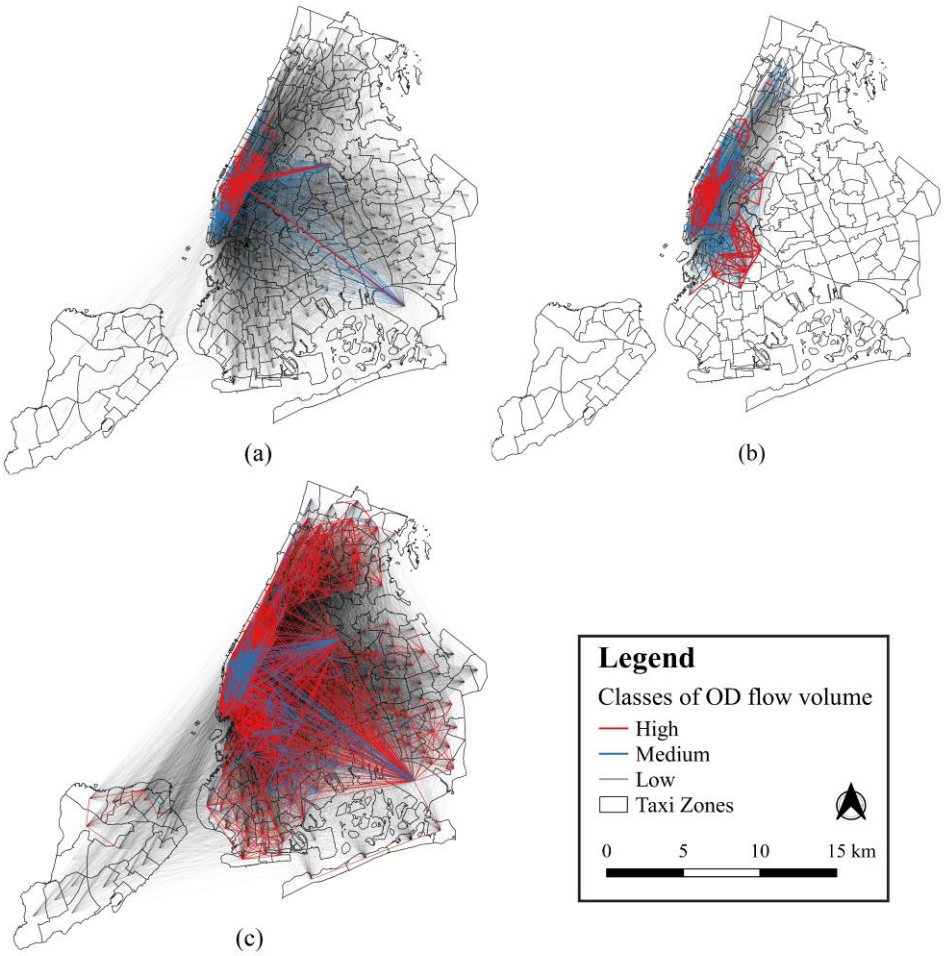


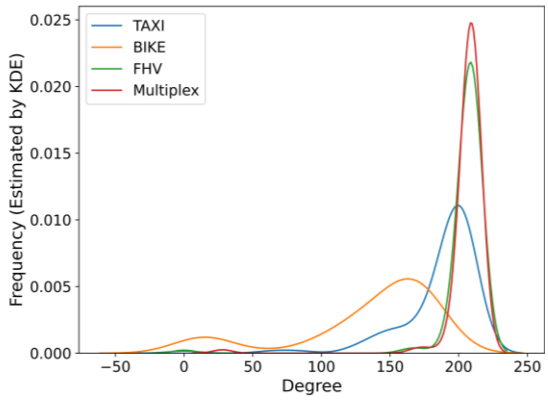
Figure 7. Spatial Distribution of Multimodal Travel flows.

(a) Traditional Taxi; (b) Shared Bike; (c) FHV.

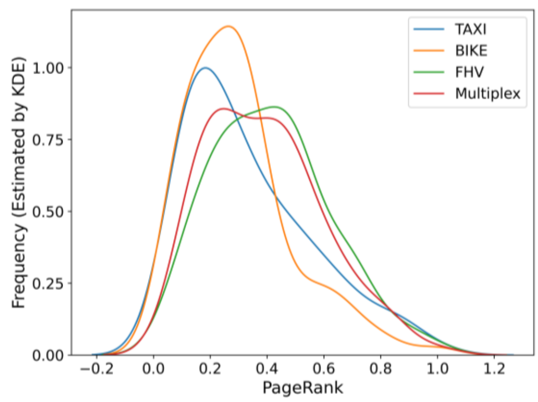
4.3. Results and Implications

4.3.1. Travel patterns characterized by multilayer mobility network

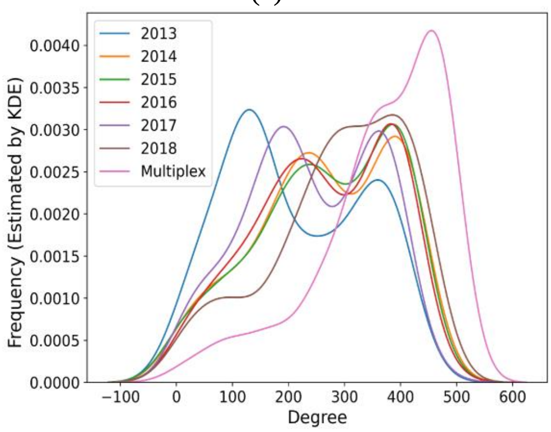
For both the multimodal and temporal networks, we compute both multiplex degree and PageRank. Then, the statistical distribution is inferred using distfit, a python library fitting more than 80 models. Performance of all statistical models will be ranked based on a widely used metric, the residual sum of squares (RSS). We found that both two types of centralities does not adhere to a power-law or exponential distribution in both multi-modal and temporal networks. On the basis of the least RSS, the beta distribution is found as the optimal model for depicting centralities structure in multimodal network. That means majority of zones (nodes) in multimodal network have strong degree and moderate PageRank (Figure 8a). The degree distribution possesses a substantial left tail, while the PageRank distribution possesses a tiny right tail. In temporal network, however, high degree but low PageRank are found for the majority of nodes (Figures 8c & 8d). In terms of degree centrality, there are substantial differences between the layers (i.e., 2013 and 2018) Chronologically, more zones with greater degree appear. Similar PageRank distributions with slightly longer right tails exist across layers.



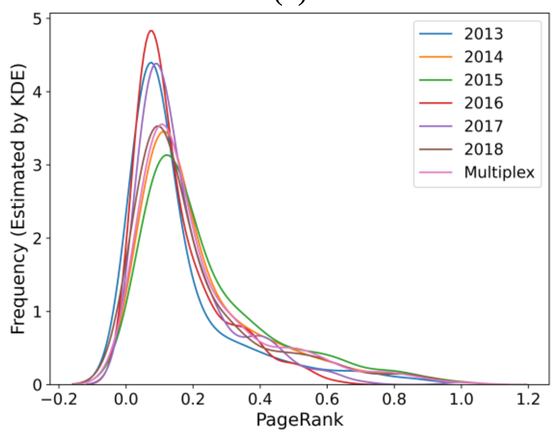
(a)



(b)



(c)

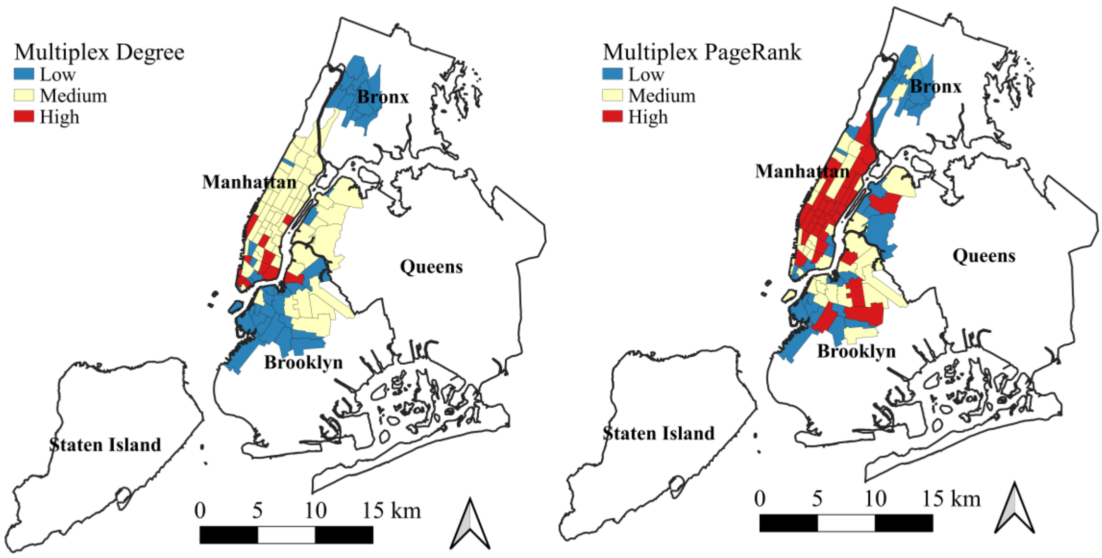


(d)

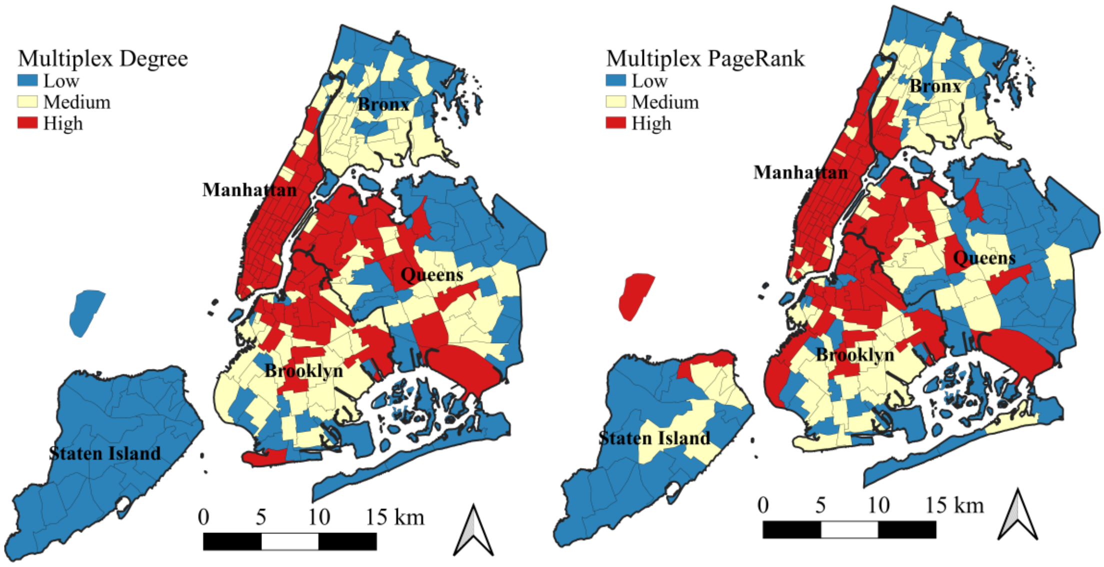
Figure 8. Distribution of node centralities

a) multilayer degree of modes network; (b) multilayer PageRank of modes network; (c) multilayer degree of temporal network; and (d) multilayer PageRank of temporal network.

Centralities reflect connectiveness of a location in flow network, which may hint underlying urban structure that generates such flows (Jia et al., 2019). As shown on Figure 9, spatial distribution of zone centralities is an effective means to reveal polycentric spatial structure of NYC. While we also notice the differences between multilayer degree and PageRank (Figures 9a & 9b). Extremely high degree are most prevalent in the Manhattan downtown, whereas high PageRank are identified for nearly the entire borough. Possible explanation for the degree's extreme lean to the left is that the Manhattan zones are highly interconnected when multiple modes of transportation are considered together (i.e., traditional taxis, shared bikes, and FHVs). On such a scale, the degree may not be the most appropriate metric for distinguishing a node's importance, whereas PageRank reflect more variation in this densely connected network. Degree and PageRank calculated from temporal network exhibit more comparable patterns (Figures 9c & 9d). In terms of node centralities, Manhattan is the most "important" borough; however, certain zones in Brooklyn and Queens also exhibit high values. Multiplex centralities in New York City provide evidence for a polycentric urban structure that reflects travel demand.



(a) (b)



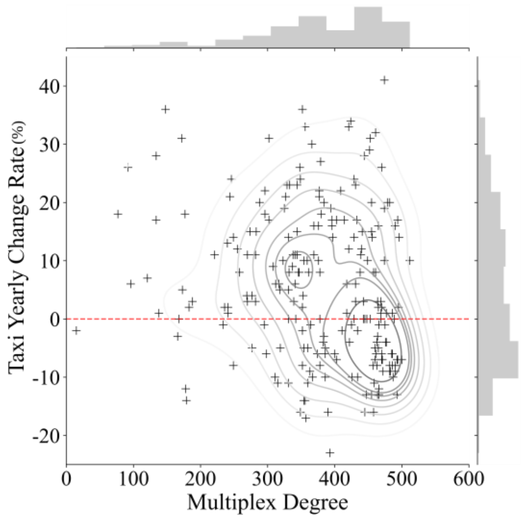
(c) (d)

Figure 9. Spatial distribution of node centralities.

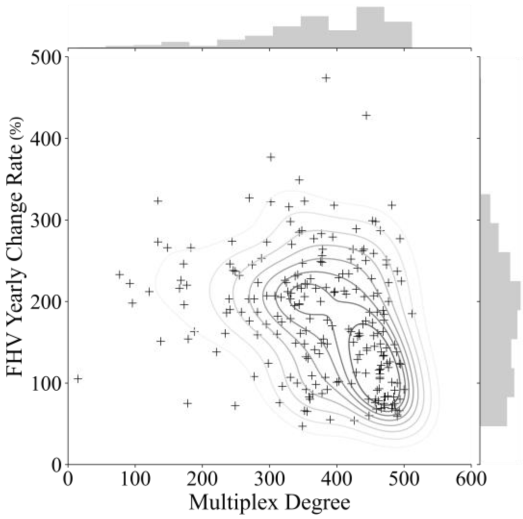
(a) multilayer degree of modes network; (b) multilayer PageRank of modes network; (c) multilayer degree of temporal network; and (d) multilayer PageRank of temporal network.

Year change rate is an essential feature to depict how people's travel choices shift among different transport modes. Herby we further investigate whether such behaviour shift is particularly related to certain areas revealed by multilayer zone centralities. In specific, we choose centralities calculated from temporal network to discuss relationship for two reasons. First, the spatial coverage of traditional taxi cover the whole city, enabling multiplex centralities in a wider range and diversity context for correlation analysis. Second, temporal network considers time variations in the network and is more aligned with the year change rate that is also calculated from time series flow data.

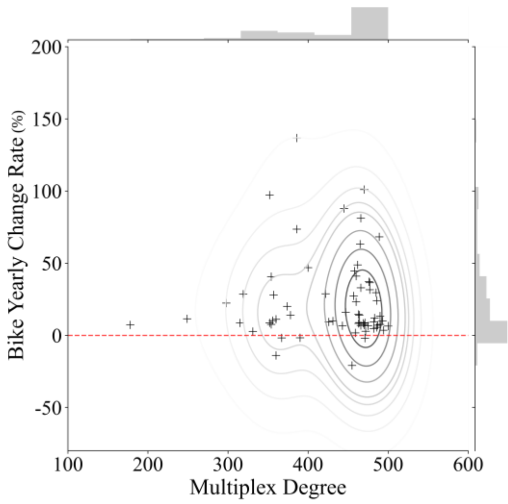
In each spatial unit (zone), the yearly change rate of three transport modes are all calculated. In Figure 10, X axis depict the multiplex centrality of each zone, and Y axis shows the yearly change rate of the specified transport mode. Some visual analytics are added to enhance the comprehension. First, the KDE function adds contour lines to highlight density of data points. Second, we add the reference line at the position where yearly change rate equal to 0, distinguishing data points (zones) with positive or negative flow volume changes. In general, both negative and positive change rates are recorded for traditional taxis, whereas both two shared mobility services are noticed with dramatic increase of use in many zones. In terms of overall relationship between yearly change rate and zone centrality, similar distribution is found in the two vehicle-based mobility services, FHV and traditional taxi. In Figures 7a and 7b, zones generally fall into a relatively low centrality group (350) and the high centrality group (450). While the interesting finding is that the yearly change rate of both these two vehicle services are low in high centrality group of zones. In some zones with high centrality, the annual change rate for traditional taxis is even negative. In other words, people of zones with versatile flows are losing interests on vehicle mobility. Another side of the same story is revealed by conducting the same analysis by plotting shared bike data (Figure 10c). First, the data points are obviously less than other two modes, mainly available in Manhattan and its surrounding areas. Second, we found that 450 centrality of shared bike network is what most zones present. While number of shared bike flows are significantly increasing in these zones, in contrast to the vehicle-based services.



(a)



(b)



(c)

Figure 10. Correlations between multilayer degree and yearly change rate.

(a) Traditional Taxi; (b) FHV; (C) Shared Bikes.

4.3.2. Investigating variation of multi-flow-induced urban structure

Multiplex community detection is another way to reveal urban structure, more importantly, dynamics of flow-induced urban structure can be better investigated than the centralities or single-layer community detection. In this section, we adopt the multiplex-Infomap algorithm for both two network models proposed in the study: multimodal mobility network and temporal taxi network. As we discussed in **Chapter 3**, multiplex community detection will group zones with similar flow patterns, and such identification is layer-wise. In other words, the layer-wise labels of each zone can evaluate the possible changes of flow patterns (travel behavior) over transport modes or time. All labels together, on a higher level, can be used to understand the zonal characteristics and its associated the urban structure based on if and how the labels vary.

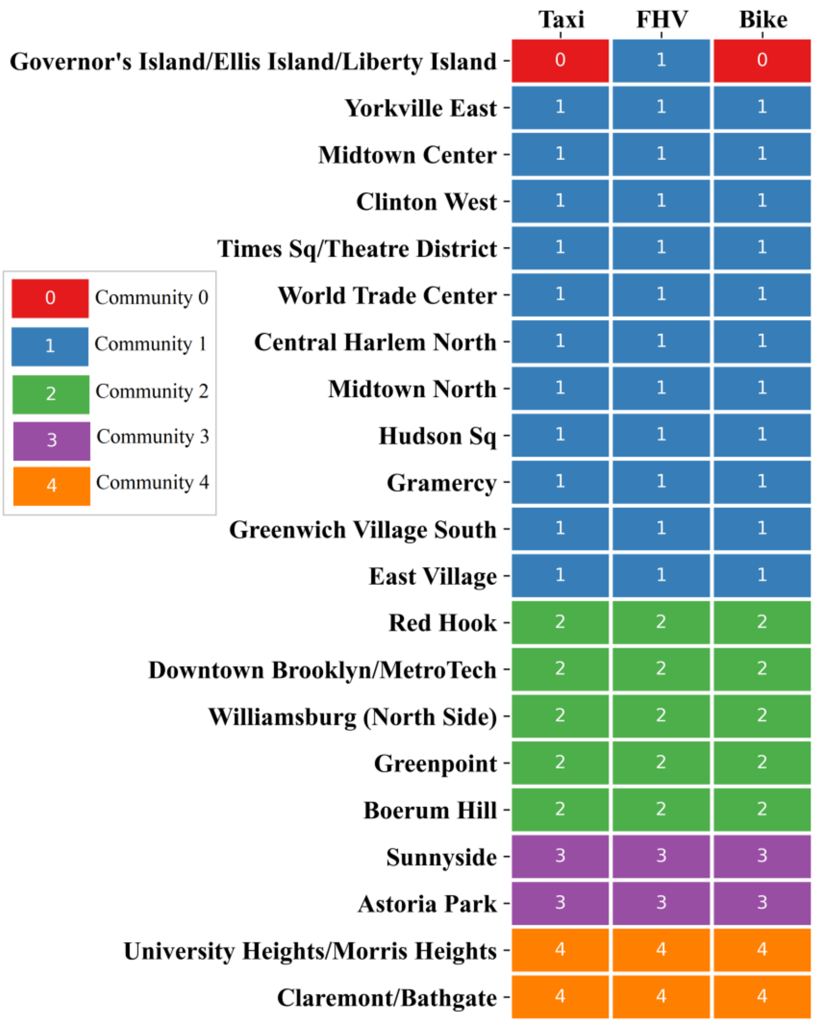


Figure 11. Zonal labels of network community across transport modes (layers).

Note that only partial zones are demonstrated in the figure due to limited space. There are total 1 zone with varying labels across layers, 57 zones with all labels as 1; 27 zones with all labels as 2; 8 zones with all labels as 3, and 13 zones with all labels as 4.

For multimodal network {Taxi2018, FHV2018, and Bike2018}, the methods mentioned above specifically investigates, in 2018 the time when shared mobility obtain decent proportion of ridership, whether the flow patterns captured by network community are varying among different transport modes? Does the dramatic rise of shared mobility ridership mean the varying travel bahviour as well? In total, five types of community were identified, for which we assign 0 to 4 as community labels to nodes across all three layers. We plot the label distributions using a matrix plot (Figure 11), on which Y axis is different zones (i.e., nodes) and X axis is the three transport modes constructing as layers. Surprisingly, we found that most zones have identical labels among different transport modes. Below the Figure 11 we provide the extract number of zones belonging to different labels.

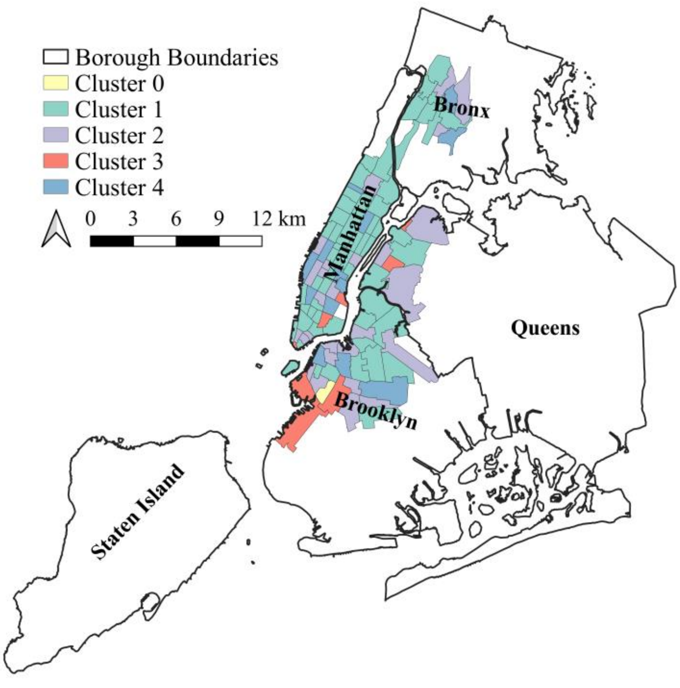


Figure 12. Spatial distribution of zonal labels of modes network.

On above we present the overall pattens of multiplex communities, while hereby the more detailed trend is discussed. Due to limited space, only partial zones are represented on Figure 11. Whereas most zones have same labels, a classification scheme is reasonable to explore the flow patterns in the aggregative manner. In particular, each zone can be generalized as a set of three labels {taxi_community, FHV_community, bike_community}. When all labels are same, we use the label value to name the class of the zone. For example, Yorkville East with all community labels as 1 is given the name of 'Cluster 1'. These zones with consistent labels across transport modes reflect that the travel flow patterns are similar no matter for traditional taxi or shared mobility services (shared bike and FHV). The community labels of the first zone, Governor's Island, however, are 0, 1, 0. That means that the flow patterns by taking traditional taxis and shared bike are similar, while shared vehicle present in different manner. Figure 12 projects the multimodal network communities on NYC map, from which we find that zones with consistent modes flow patterns are strongly spatially clustered, for example, label 1 in the centre regions, and label 2 and 3 in more distant regions.

The results obtained from temporal networks share some similarity (Figure 13 & 14). Using time as layers, thereby this analysis examines whether the flow patterns of traditional taxis vary from 2013 to 2018, during when shared mobility services rapidly expand their market share. Except, Green-Wood Cemetery, Gothenburg, Riverdale, and Country Club, most zones have consistent community labels. This result indicates that the flow patterns of the majority of zones for traditional taxi remain consistent. The result is comparable to the findings of the multimodal network experiment. In other words, despite the large variance in market share between regular taxis and shared transportation, flow patterns did not vary considerably. Further discussions are given in the next section.

It is important to note that ground truth data for community detection is often difficult to obtain, because spatial structure of a city is complex and dynamic, and the detected communities exatracted from dynamic data often not fully match the fixed boundaries. However, one exploratory way for the evaluation is to compare the detected communities to the geographic features of the city. Using our case as example, the selection of geographic feature could consider the population density. Because what we observed in two multiplex mobility networks is the stability of variation of community structure across layers. That hints although a new transport mode (shared mobility) was introduced, the travel behavior between locations presents a robust pattern (Figure 13). In other words, travel demands remain although passengers are free to choose other transport modes (Figure 10). While the analysis of reasons for travel demand is not the objective of this study, and the correlation analysis can be done in future studies. The community structures hereby are used as indication of the travel behaviour.

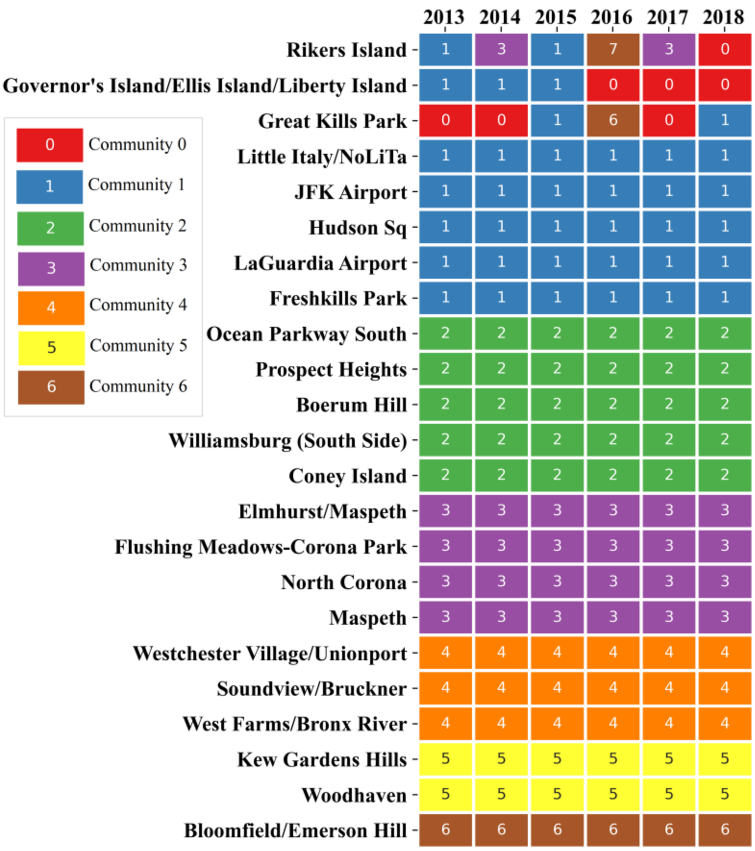


Figure 13. Zonal labels of network community across years (layers).

Note that only partial zones are demonstrated in the figure due to limited space. There are total 3 zones with varying labels, 65 zones with all labels as 1, 62 zones with all labels as 2, 29 zones with all labels as 3, 46 zones with all labels as 4, 37 zones with all labels as 5, and 18 zones with all labels as 6.

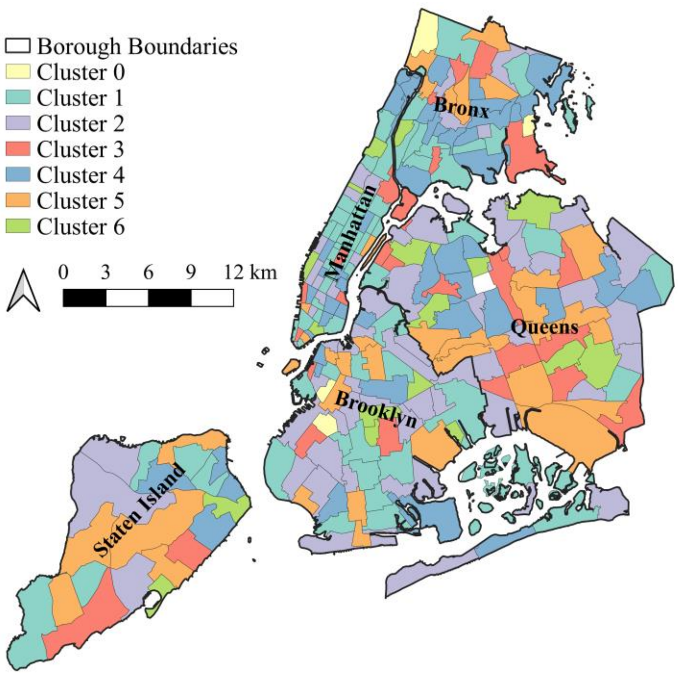


Figure 14. Spatial distribution of zonal labels of temporal network.

4.4. Conclusion

In recent years, shared mobility services have expanded considerably. In light of recent advancements in multilayer network analysis, this research develops empirical multiplex network models to investigate how the spatial structure of city is shaped and varies due to dynamic urban flows. The logic complexity of multiplex models is O(n) where n is the number of types of layers. For each layer's type, the users need to determine what donates to the inter-layer edges. The computational complexity of multilayer methods highly depends on the specific algorithms or tools used, thereby was not examined in this study while one can further refer to the Boccaletti et al. (2014), Domenico et al. (2013), and Kivelä et al. (2014).

A set of diagnosis methods in multilayer network is used to depict the flow patterns. First, multiplex degree and PageRank are calculated to reflect the hierarchical structure of NYC zones. In a multimodal network, we discovered major distinctions between Manhattan's uptown and downtown, being similar to results in Zhou et al. (2019). The spatial distribution of node degree in temporal network is more intriguing, showing that Manhattan and its surrounding areas have similar high contact flows. The disparity may be attributed to the different layer architecture between modes network and temporal network. The results indicate that the chosen layer adequately captures the variation attributed to the selected context, namely the relative magnitude of the flow between modes and the relatively minor variation over years. These results are valuable as lenses of spatial structures using multiple flows, while presenting two distinct perspectives. The statistical distribution of network centrality is contrary to findings from earlier research (Gonzalez et al., 2008; Zhao et al., 2015). Rather than power-law, degree is found to be left-skewed and PageRank is found to be a small right-skewed. The non-power-law is especially pronounced in the multimodal network, indicating that multimodal transit alternatives increase connectivity between zones. That leads to a more equal chance of travel. Using multilayer degree, the found empirical evidence support the argument made by Kong et al. (2020) on that supplement of shared mobility for traditional transportation is significant for distant areas.

Multilayer community detection enables direct comparison of the identified role of node under different circumstances. Despite the fact that the ridership amount of traditional taxis has been largely supplanted by FHVs and shared bikes, the flow pattern depicted by community detection indicates that the majority of zones are consistent no matter in the context of transport modes or multiple years. The indication is interesting because the shared mobility is often seemed as totally new way of travel comparting to traditional mode, while the destination where people travel to may not vary significantly. The collective travel behavior spinning over zones is not influenced by new technology and travel modes, but instead, may be associated with the demographic attributes of the travellers.For instance, regardless of the traveler's means of transportation, the relationship between a home zone and a working zone remains unchanged. The constancy of neighborhood patterns across transit modes in New York City is consistent with another agent-simulation-based study (Lokhandwala & Cai, 2018). In simple words, shared mobility do competes with traditional modes severely, which is however may not influence the travel demand and supply collectively for the travellers.

Consistent community labels are also found in temporal taxi network, which is more surprisingly than the findings of multimodal network. Because we do see the dramatic decline of taxi ridership amount from 2013 to 2018. Our findings imply that the human mobility in long period of time might be generalizable to other cities. In this sense, taxi data is an appropriate candidate as observing long-term human mobility (Riascos and Mateos, 2020). A further implication is that environmental factors may have a greater impact on flow pattern and travel behaviour than the advent of shared mobility services (Zhang et al., 2020).

The association with land use was not explicitly investigated, while the clustering of zone degree also provides some hints on how ‘location’ influence on flow patterns. Figure 10 reveals similar suggestions that urban context characteristics rather than the emergence of shared transportation may influence the change in mode selection preferences. We have witnessed the increase of flow volume of vehicle-based mobility in 350-degree zones (suburban), but decline in 450-degree zones (city centrals). This phenomenon is attributed to the convenience of shared bike comparting to car-driving in the dense area. While multiple transport modes do provide more choices for the distant areas, thereby transforming more travel demands into real flows. The results indicate that travel demand is on the rise and that regular taxis and shared vehicles do not have to be "competitors," but offer their own advantages for travellers in different locations. The multilayer network methods here are, therefore, useful tools for location-based policies on which modes should be encouraged or suppressed for specific regions. Such policies should also well consider the built environment and local demographics, as what community detection presents is the consistent spatial interaction patterns among zones.

This work possesses certain limitations. This study does not incorporate public transport flows when examining travel behavior and urban form. We concern that public transportation is more constrained by the pre-planned infrastructure routes, in which flow patterns may not be comparable to shared mobility service being operated by mixed-mode mechanism (face-to-face and online matching). In addition, further studies are deserving on investigation how contextual information influence the flow patterns, providing more practical implications. With all limitations being aware, our approach contributes to a new perspective to comprehend urban dynamics by integrating multiple urban flows into multilayer network models.

Chapter 5. Diversity of Spatial Interaction: A Novel Metric for Geographical Flows

5.1. Motivation

Diversity has been a vital concept in urban life and development. In urban context, many big-city problems nowadays are the consequences of lacking diversity, such as traffic overcrowding, income inequality, and social discrimination (Jacobs, 1961; Dincer & Hotard, 2011; Thompson, 2020). It is urgent to develop quantitative measurement on various facets of urban diversity to guide policymaking. However, an essential feature of complex city, urban flows, has still been neglected in diversity evaluation. Besides the methodological void, quantifying diversity embedded in flows are necessary for several reasons. **First**, evaluating diversity because it is widely regarded a beneficial component for city, however, growing evidence from many have suggested that flow is the fundamental force for the benefit, such as for ecological resilience, social inclusion, economic prosperity, and innovations (Vreeker et al., 2004; Pardikes et al., 2018; Smith et al., 2018; Vormann, 2015). In other words, diversity benefits system through enhancing flow and interaction. **Second,** although quantitative measurement on urban diversity has long been a core topic in geography (Zimmerer, 1994; Low et al., 2009; Nash, 2012), **most existing methods focuses on first-order attributes of each location** (e.g., either land characteristics or time-series human activities of each unit). Increasing attention nowadays has been attracted to extract spatial knowledge from second-order data (i.e., flow). In new urban sciences, flow is also regarded as a fundamental lens to observe complex city. Thereby developing a metric for flow diversity contributes to the latest initiatives of both geography and urban science.

5.2. Research Questions and Methods

This study attempts to investigate the following three questions:

- What are the basic components of interaction diversity and how to form the integrative metric?
- What are the spatial-temporal characteristics of interaction diversity?
- How interaction diversity can be used in understanding urban issues?

To answer those questions, we collect taxi trip data in Shenzhen of a whole month as the observation on spatial flows (Figure 15), and test our newly proposed metric, diversity of spatial interaction (DSI). We refer to the philosophy of diversity metric in scientometrics field (Figure 4) to define the basic components of flow diversity in geographical context (Figure 16). The calculation of DSI is conducted in weekly manner, covering the week before, within, and after the Spring Festival of China. Spatial statistics methods are used to quantify spatial patterns of DSI, and temporal changes are discussed. Relationship between DSI and landuse mixture are investigated to validate the usefulness of DSI. In the final parts, we further examine the intersection between DSI with specific built environment types (using POI), to address practical implications on the positive and negative sides of interaction diversity.

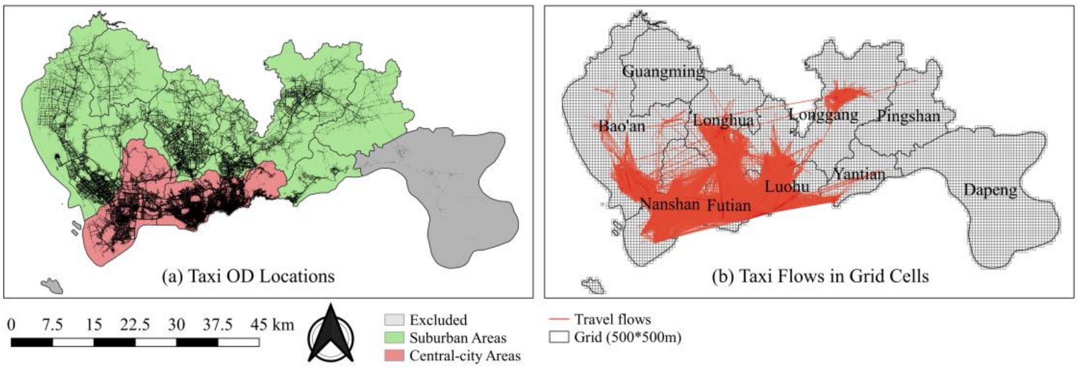


Figure 15. Study Area and Taxi OD Flows.

Note that central-city areas include Nanshan, Futian, Luohu districts (Zhou et al., 2022).

5.3. *Results and Implications*

5.3.1. Constructing DSI from geographical flows

The analogy between the two fields is natural because the format of citations among articles and travel flows are fairly similar. In geographic space, we describe a diversified spatial interaction as having a large number of origins (Variety), a balanced interaction volume (Balance), and origins of differing types (Disparity). The DSI index is comprised of the aforementioned three components, and its reasoning and use are described below.

For a location j, DSI index has three components as shown in Equation 13:

$$DSI_j = Var_j \times Bal_j \times Dis_j \tag{13}$$

where Bal_j values originally range from 0 to 1, as it is calculated based on 1-Gini coefficient. Var_j and Dis_j are normalized so that all DSI components are bounded in the same range.

Eech DSI components have clear physical meaning in geographical space:

Variety means the number of classes that have flows to the destination. We regard each origin location as a class, addressing the uniqueness of city places (Cao et al., 2018). Physical meaning of Variety is similar to the richness of places (Kang et al., 2021) or node degree in a spatial network (Ni & Weng, 2009). The more places a destination can attract flows, the higher the interaction diversity will be. Using travel as example, high Var value represents a destination can serve travel purposes from many locations.

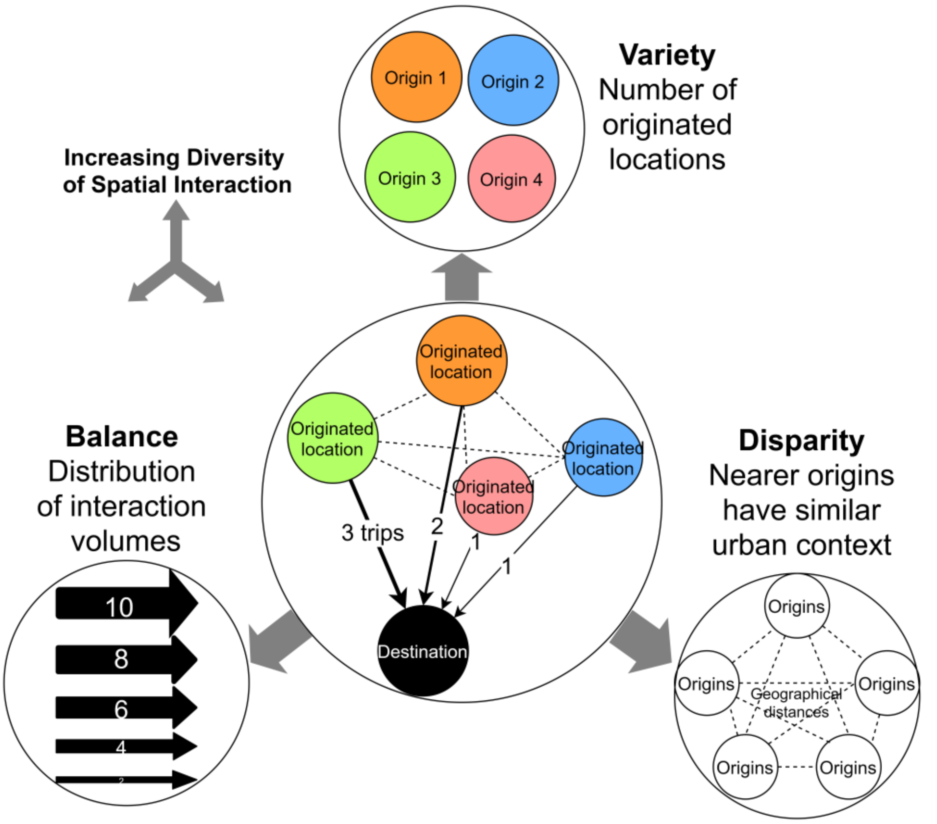


Figure 16. Components of Diversity of Spatial Interaction.

Balance means the evenness of flows spinning over originated locations. In other words, this component measures the inequality of interaction flows. For instance, a set of flow volumes (7,1,1,1) is more 'unbalanced' than the set (4,3,2,1). The Balance has two practical meanings in geographical space. First, distribution of interaction volume serves as 'weights' of Variety (i.e., number of originated locations) to have a more realistic evaluation of interaction diversity. If a location attracts flows from many origins but very few origins dominate the flows volume (i.e., overweigh than others), the destination is indeed not diversely interacting. Second, we are not pursuing totally homogenous interaction in city, but addressing the importance to reduce the gap between flows-dominant locations and flows-shortage locations. The gap of interaction volume is found in large cities, also known as the ’80-20’ rule (Jiang et al., 2009; Liang et al., 2013), while the huge gap also raises concern on equality and safety (Yin et al., 2018; Choudhury et al., 2020). A more balanced flows distribution matches the pursuit of ‘polycentric’ in modern urban planning (Kloosterman & Musterd, 2001), offering relatively equal opportunities for different communities to access important areas.

Disparity means the overall difference of originated locations. In other words, if a destination attracts flows from locations with the different urban context, the overall interaction diversity is decent. We acknowledge that there are sophisticated methods to determine the similarity of urban context (Cai et al., 2019; Chen et al., 2021), while we choose distance as Disparity metric in this study. First, it can keep Disparity and DSI index in simple form, which enable DSI to be easily implemented in other cities. Second, distance-based metric is widely used in other Geography studies to proxy different urban contexts (Solon, 2009; Chang & Liao, 2011; La Rosa et al., 2017). Based on the First law of Geography, we assume that near origins have similar urban context, while more distant origins have more diverse context (Tobler, 2004). Using travel as an example, a destination has higher interaction diversity when attracting flows from all over the city. Interaction diversity is relatively low when a destination attracts flows from origins only located in a certain area.

5.3.2. Effectiveness of DSI

Monotonicity is required for measuring diversity with multiple dimensions (Rousseau, 2018). It means that when two dimensions (components) of diversity remain the same (i.e., as constant), the increase of the third component will increase the overall diversity index. This concept is met by the DSI index. DSI integrates three components to depict the diversity of spatial interaction. Compared to a single diversity index alone, DSI captures more heterogeneity and dynamics.

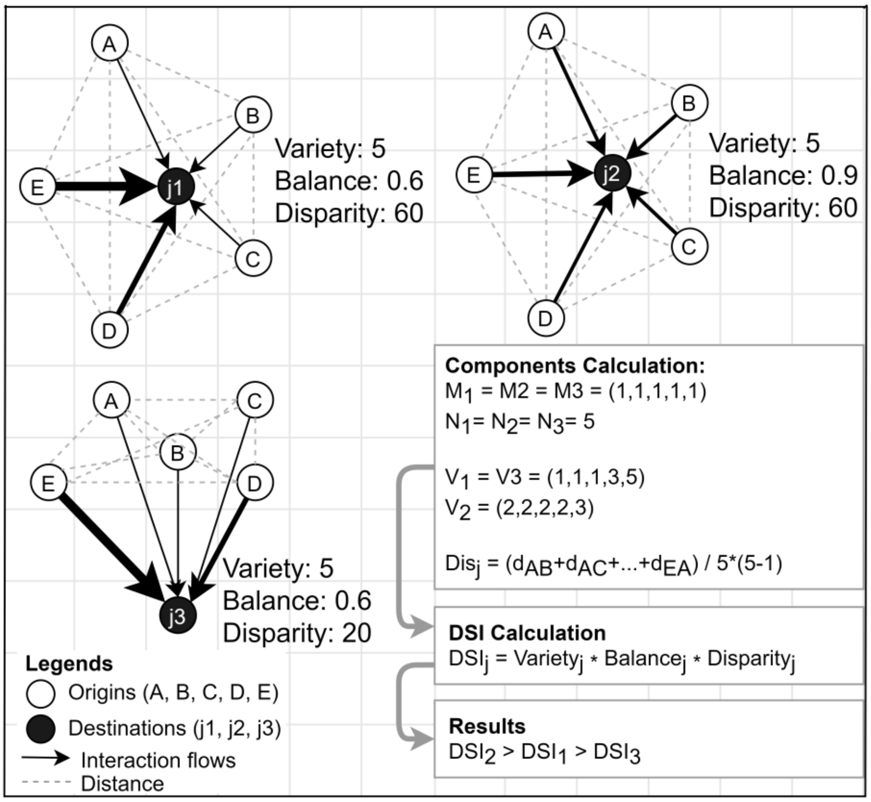


Figure 17. Examples of calculation of DSI for destination j1, j2 and j3.

Note that all destinations (j1, j2, j3) attract flows from 5 originated locations (A, B, C, D, E), interaction diversity cannot be differentiated using richness of origins. Regarding to the origins, Balance supplements information of flows volume distribution, and Disparity supplements different spatial context.

The effectiveness of DSI is demonstrated using example cases of spatial interaction. As shown in Figure 17, the example calculates the DSI of three destinations (location j1, location j2, and location j3). They separately interact with the same 5 origins (A, B, C, D, E) in different ways. It is obvious that if interaction diversity is only determined by the richness of places, all destinations (j1, j2, j3) have identical diversity. While the visual patterns obviously tell the different story that spatial interaction patterns are distinct. If considering two components, interaction diversity cannot be differentiated very well. It would be wrong to conclude that diversity of j1 equal to j2 because they have the same spatial layout of origins locations (i.e., same Disparity). While interaction volumes indicate overall diversity j2 is larger than j1. It would be wrong to conclude that diversity of j1 equal to j3 because they have the same distribution of interaction volume (V_j) (same Balance). While disparity of spatial context indicate overall diversity j1 is larger than j3. Therefore, diversity of spatial interaction can be effectively reflected when combining all three components, specifically found that $DSI_{j2} > DSI_{j1} > DSI_{j3}$.

5.3.3. Spatial-temporal Characteristics of DSI

This section reports how DSI, and its disaggregated components are spatially distributed in Shenzhen, and to what extent spatial heterogeneity of DSI is significant. In Figure 18, we use week 1 to discuss spatial pattern, and other four weeks will be discussed in the temporal heterogeneity. First, the DSI and components values are classified into 4 grades by user-defined percentiles. We are interested in locations with top and bottom interaction diversity, which are more indicative for specific policies. Therefore, we define the following symmetric categories: Grade 1 means 'very high' value (rank above 90th percentile), Grade 2 means 'high' value (between 50th and 90th), Grade 3 means 'low' value (between 10th and 50th), and Grade 4 mean 'very low' value (below 10th).

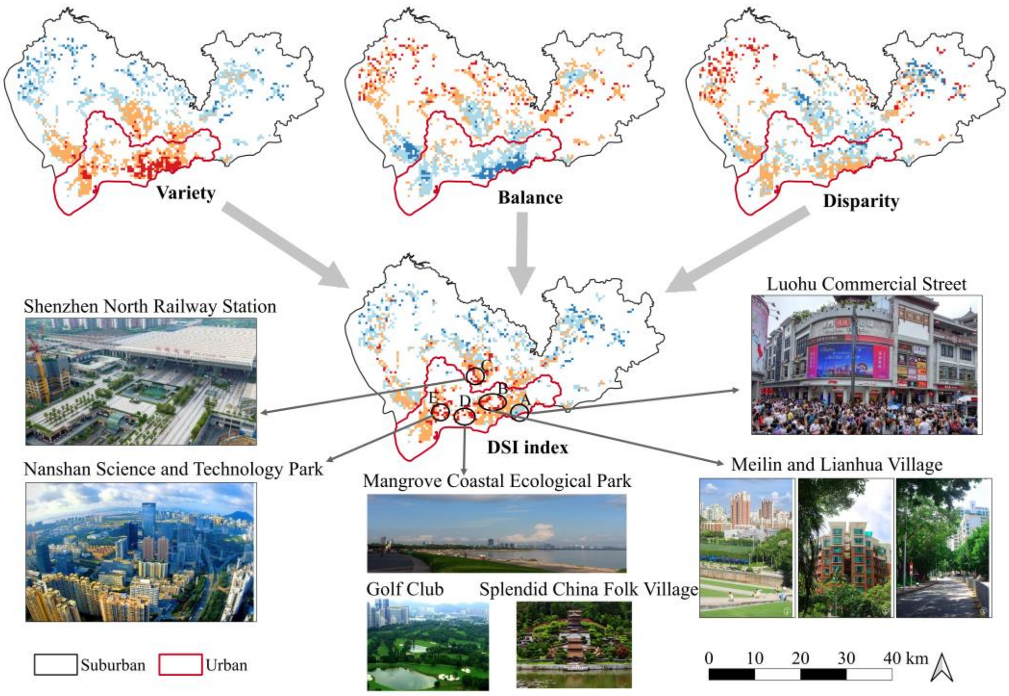


Figure 18. Spatial Distribution of DSI in Shenzhen.

We find that urban areas generally have higher interaction diversity than suburban areas. For instance, very high DSI values are noticed in the well-built districts such as Futian and Luohu. But a more interesting finding is that very high DSI places are indeed not in the most densely built areas near the south city boundary, such as location A. But very high DSI seems to be related to several urban complex (i.e., metropolitan planning area with integrated services), such as location C (Shenzhen North Railway Station), Location E (Nanshan Science and Technology where encompass high tech companies such as Tencent), Location D (Several large city parks and scenery), and Location B (Meilin and Lianhua Village where are large and cozy residential area). Overcrowding and compact built environment are not necessarily related to decent diversity of interaction. For example, we found that Location A (Luohu Commercial Street) has a high total trip amount while it is only from very few locations, resulting in low Balance and DSI value. This old city market is rebuilt and famous for catering and life-related business, while cannot attract diverse travel flows as other urban complex do. Beyond measuring urban diversity, the results show that DSI might be a better metric to identify vibrant locations of city, compared to population-density-based vitality proxy (Yue et al., 2017; Wu & Niu, 2019).

We further quantify the DSI heterogeneity using spatial autocorrelation statistics. Moran's I is used to test the assumption that the DSI of a location is influenced by its neighbours. To obtain spatial weights matrix, K-nearest neighbour method is used with considering that spatial unit (grid cells) in this study doesn't always have continuity-based neighbours. As shown in Figure 19, we compare the results of monte-carlo simulation with increasing k (1 to 16). The moran's I of DSI are all significant ($p=0.001$) and located on the right tail of distribution. It means that we can reject the null hypothesis that DSI is randomly distributed in urban space but conclude that similar DSI values are concentrated in specific areas. We choose $k=8$ for the following analysis.

The moran scatter plot classify locations with different DSI effect (Figure 20). The four quadrants correspond to four types of spatial autocorrelation: low-low (lower left, clustering of low DSI locations); high-high (upper right, clustering of high DSI locations); low-high (upper left, cold spot surrounding by high DSI neighbours); and high-low (lower right, hot spot surrounding by low DSI neighbours). The dual urban structure of DSI is obvious in Shenzhen, with low values in suburban areas (North) and high values in urban area (South).

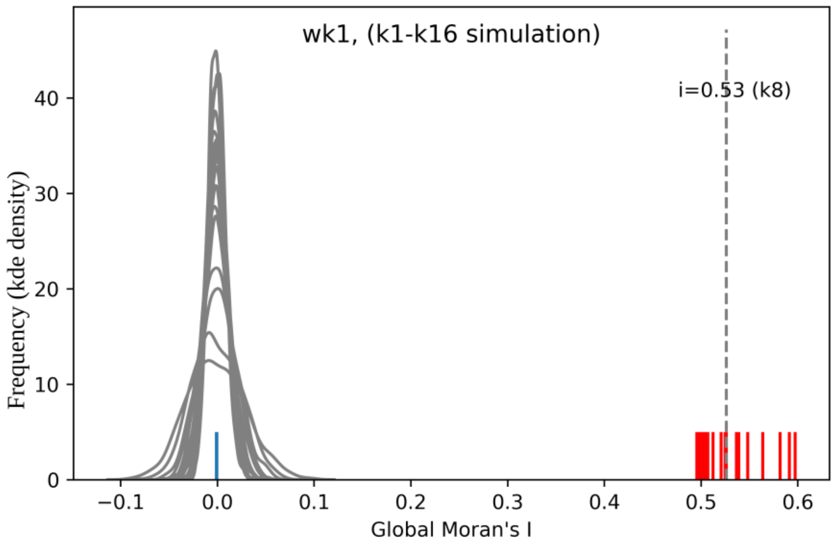
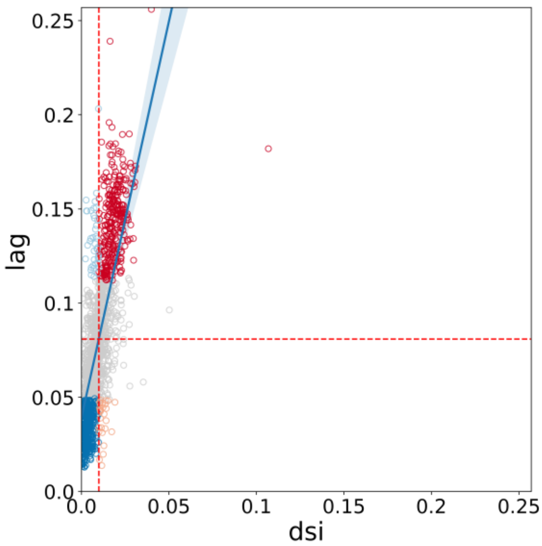
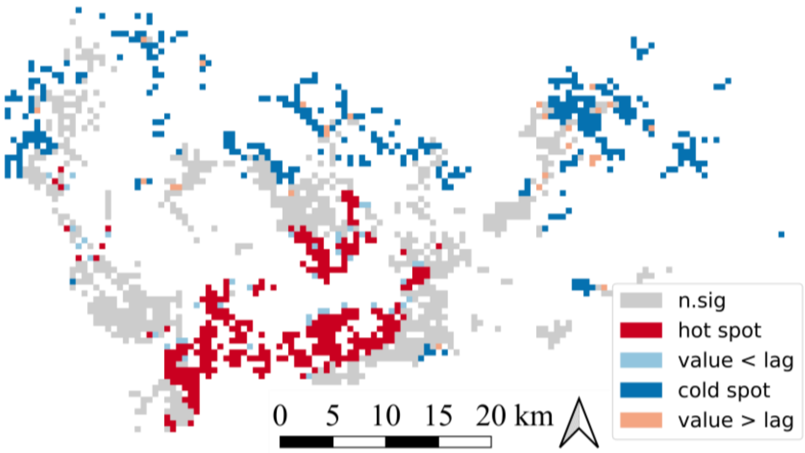


Figure 19. Monte-Carlo Simulation of DSI Moran’s I.



(a)



(b)

Figure 20. Local moran’s I plot.

(a) Relationship between DSI and its spatial lag; (b) Map of local morans’ I. Note that significance level is set as 0.05 for masking non-significant locations as grey.

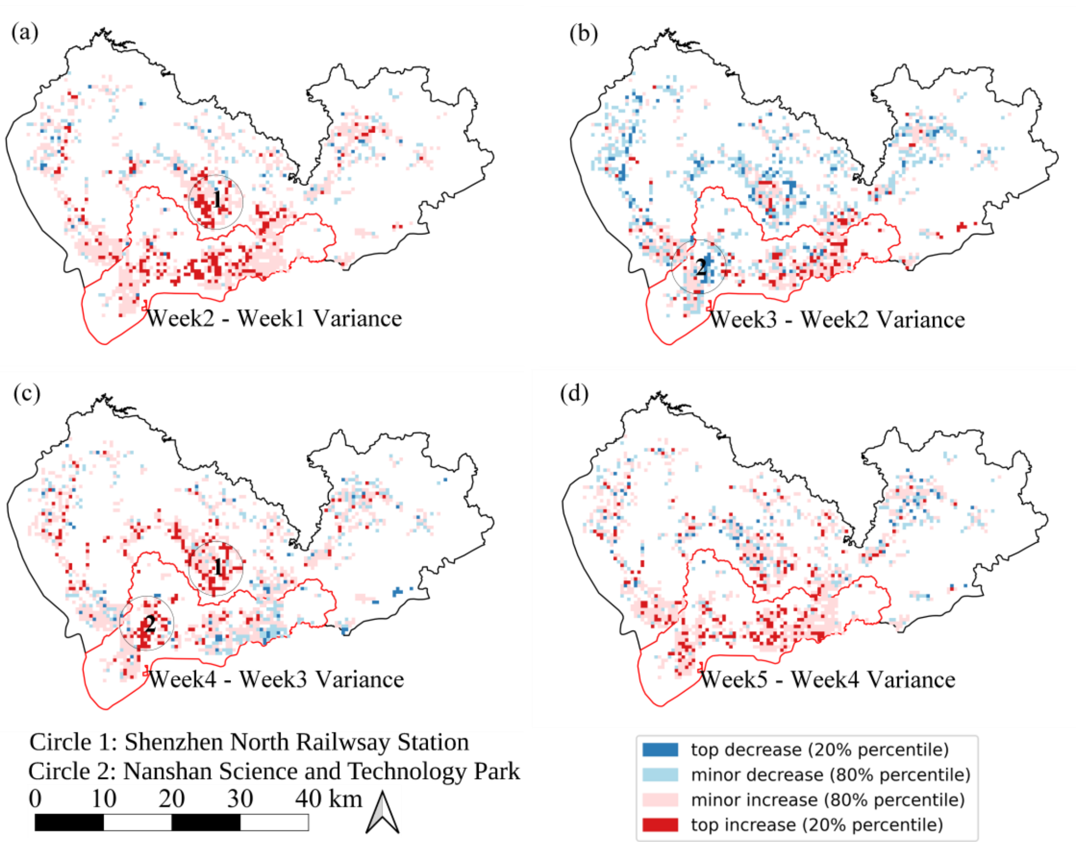


Figure 21. Temporal Changes of DSI from Week 1 to Week 5.

Note 'Variance' values are calculated by subtracting locational DSI values in consecutive weeks.

We further investigate whether and how DSI varies over time. Particularly, we conduct comparative experiments during the Spring festival travel season. It should be noted that the travel season is normally categorized into 5 weeks: The first (1st) and the last (5th) week are the normal working weeks; the 2nd week is the pre-holiday week; the 4th is the post-holiday week; the middle week (3rd) is the whole week of holiday.

In Figure 21, locational DSI values are subtracted in consecutive weeks. We found that seasonal backgrounds do have impacts on DSI. From a normal week to pre-holiday week (i.e., week 2), we find increasing DSI in widespread areas, particularly location 1 (Railway station). Because many workers will leave the city during the 2nd week, and other citizens will travel with more purposes such as purchasing goods and gathering. To the 3rd week (holiday), the results show that a wide range of suburban areas faces a dramatic decrease of DSI (Figure 21b). In the urban area, Futian and Luohu districts are very active while science and technology park in Nanshan district lack interaction diversity severely. In the post-holiday week (Figure 21c), the periodical patterns are observed for location 1 and location 2, where are active in week 1, cooldown in week 2, and being interactive again in week 4. It reflects the back-to-work behaviour in the festival season. In week 5 (Figure 21d), the uprise of DSI is found in a more dispersed manner across the whole city, showing the returning citizen are re-vitalize more urban locations.

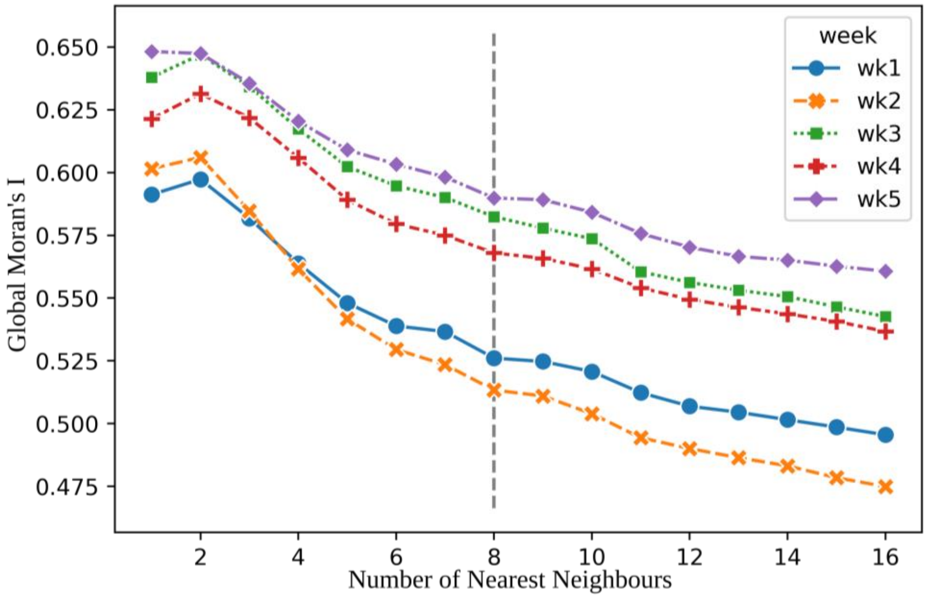


Figure 22. Monte-Carlo Simulation in All Weeks.

Note that y axis represents moran’s I value of DSI.

In the above results, DSI variance depicts the temporal effect in microscale (i.e., grid). We further investigate the temporal effect on the DSI macro structure (i.e city). With increasing k-nearest neighbours, we repeat the monte-carlo simulations in all weeks to quantify the DSI spatial autocorrelation (Figure 22). The results show that moran’s I of all 5 weeks are significantly large positive value (0.47-0.65), indicating locations with similar DSI are clustered together. It is reasonable to see the overall global moran’s I decrease with larger k, because more k neighbours involved for calculating spatial lag (i.e., average attribute value of neighbours), the more uncertainty (or spatial heterogeneity) involve. While the more useful finding is that decrease trend of moran’s I is similar in all weeks, indicating indeed the spatial structure of DSI is not influenced by temporal (seasonal) context. This pattern is revealed more clearly in maps of DSI spatial lags (Figure 23), where we use DSI ranks (i.e., percentiles) instead of absolute values in colour scheme.

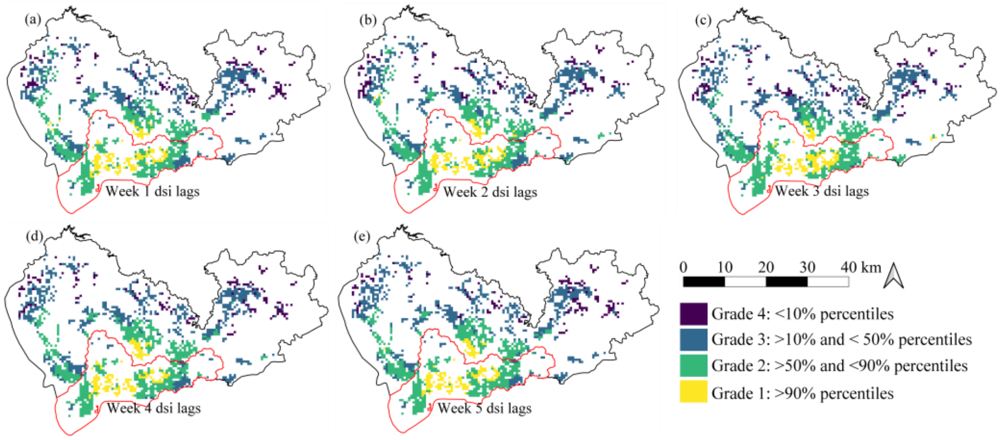


Figure 23. Spatial distribution of DSI lags.

Overall, the results show that temporal context does have an impact on the magnitude of interaction diversity at grid level, while it does not lead to changes of city level structure. Based on these results, we conclude that interaction diversity might be intrinsic characteristics of place, as a capability to attract diverse travel flows. To delineate what factor may affect such capability, we further explore the relationship between urban functions and DSI in the next section.

5.3.4. Implications of DSI

First, human behaviour is highly influenced by land characteristics, it is therefore natural to question whether interaction diversity is associated with land diversity? Based on urban vitality theory and knowledge spillover effect, a diverse urban environment will generate fruitful types of human activity and interaction (Jacobs, 1961; Talen, 2005; Firestone, 2010). To proxy land diversity, we calculate the entropy-based diversity index of POI types (Cervero, 1989; Zhang & Zhao, 2017).

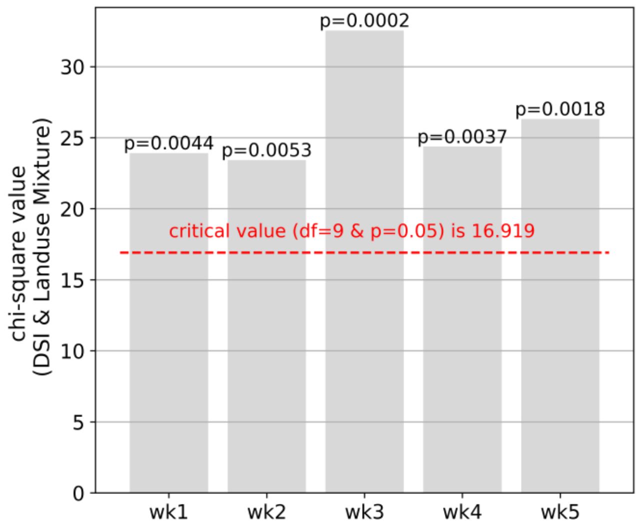


Figure 24. Chi-test between landuse mixture and dsi intensity grade.

Rather than comparing absolute magnitude, we categorize DSI and entropy (POI) values into classes for comparison, under the same scheme (i.e. [10%, 50%, 90%]). This scheme produces 4 classes based on value ranks. This step results in two types of diversity classes for each location, i.e. (grid_id, entropy_class, dsi_class). Using the chi-test, we find that the distribution of these two classes is significantly associated in all weeks. As shown in Figure 24, chi-square values are larger than the critical value 16.919 (with the degree of freedom 9=(4-1)*(4-1), and significance level 0.05). In another word, level of interaction diversity is associated with the level of landuse mixture in Shenzhen. The strongest association is observed in the week 3 (holiday week).

Second, although we report the association between landuse diversity and interaction diversity, it is still largely unknown what specific landuse types play important role in driving high/low DSI. Landuse types is proxied by the types of POI points intersected with DSI grid cells. This step results in a spectrum of POI types for each DSI classes (Figure 25). Technically, this spectrum is a barplot made based on the ranks of POI type proportion in each grade DSI locations. A higher rank of a specific POI type means that the type appears more in the selected DSI locations.

The results show that the combination of POI types is divergent in different DSI locations:

- In very high DSI locations (>90 percentiles), the top POI types are Utility, Transportation, Financial Service, Hotel, and Office Building. We name this group as Business-driven destination.
- In high DSI locations, Hotel, Administration, Cultural-Education, and Residential are dominated. We name this group as Entertainment-driven destination.
- The low and very low DSI locations present similar POI profiles, where Mechanical Service, Medical, Shopping, Catering and Tourist Sites in high ranks. We name these two DSI groups as Living-driven destination. Minor differences are that residential POI appear more in DSI low locations, and Office building appear more in DSI very low locations.

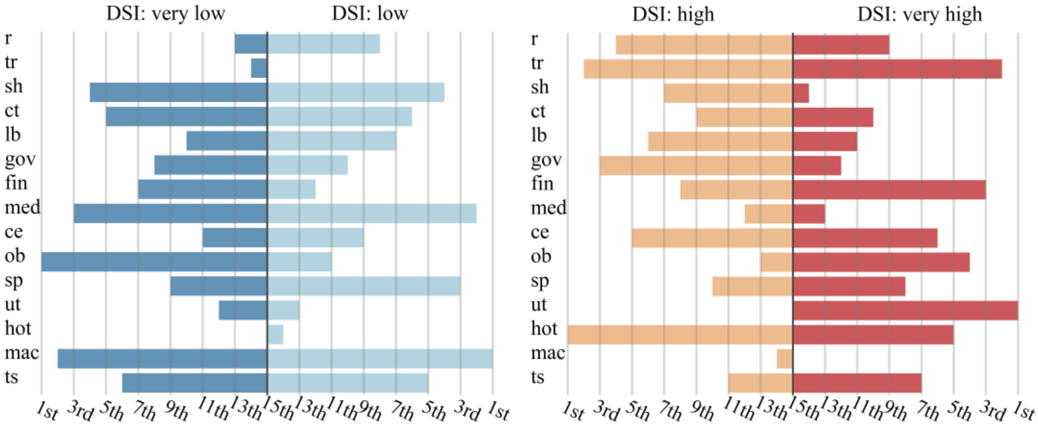


Figure 25. POI ranking in DSI areas.

Note that POI types are ranked by f(poi)ijk / Nij, where f(poi)ijk is the frequency of poi type i in week j and spatially intersected with the locations of DSI grade k, Nij is the total number of poi type i in week j. Full names of POI types: r – Residential, sh – Shopping, lb – Life service business, ct – Catering, tr – Transportation, ut – Utilities, ob – Office building, ce – Cultural & Education, med – Medical, sp – Sport Facilities, gov – Government & Administration, fin – Financial Service, mac – Mechanical Service & Factory, hot – Hotel, ts – Tourist Sites.

Practical implications can be also drawn from locational DSI value. Although high diversity is generally regarded as a good condition of city (Jacobs, 1961; Vreeker et al., 2004; Kang et al., 2021). We argue that interaction diversity might be the two sides of the same coin, influencing both life and death of city. Examples linked with prosperity (point 1 and 2 below) and danger (point 3 and 4 below) are futher discussed below. Specifically, we intersect DSI values with the density of four important POI types to showcase how DSI can be considered in urban development (Figure 26 & 27).

- Locations with high DSI have high potential benefit for economic success because attracting diverse travel flows will bring more opportunity for human interaction that relates to knowledge spillover (Talen, 2005; Firestone, 2010; Vormann, 2015). In other words, diverse interaction brings money, resources, and information exchange. As shown on Figure 26a, there are several clusters in urban area with both high office density and high DSI, where can be considered as the most competitive locations for choosing offices.
- DSI can improve evaluation of urban vitality. In previous studies, the catering business was used to proxy urban vitality as eating is an essential need even if other business has not been built up (Ye et al., 2018; Xia et al., 2020). Single view based on catering density lack observations on real human activity, while

DSI enriches the meaning of catering density for measuring urban vitality (Figure 26b). For locations with both high DSI and catering (urban area), it indicates vibrant places that connect/serve a wider range of city (high interaction diversity). Locations with low DSI and high catering also indicate good urban vitality, but the place may only serve locally (low interaction diversity).

- DSI reveals the inequality issue between urban and suburban areas. Due to historical reasons, unbalanced development exists between urban and suburban areas in many aspects, such as employment and commuting costs (Yang et al., 2018; Zhou et al., 2018). Figure 27a identify the same gap from interaction perspective. High DSI locations should be accessible for citizens. However, the majority of urban area in Shenzhen have both high transport density and high DSI, while the status in suburban area is less decent. Urban planners could prioritize transportation improvement in suburban locations with high DSI. Shenzhen government has long interests in improving connection between central and suburb districts. Urban expansion or industry re-vitalization in suburban can be conducted in the surrounding area of DSI hotpot, benefiting from materials and human resources brought by diverse spatial interaction.
- High DSI warns us of the potential risks of disease transmission, because diverse human mobility and interaction can lead to high possibility of epidemic transmission (Ni & Weng, 2009; Xiong et al., 2020). Considering the fact that the recent Covid-19 is spreading severely and globally in large cities, DSI provides a powerful tool access interaction diversity and enable location-based policy (Figure 27b). For instance, high DSI residential area might be given more strict controls while low DSI locations could be relatively relieved, reducing the overall impact on the economy and citizen’s life.

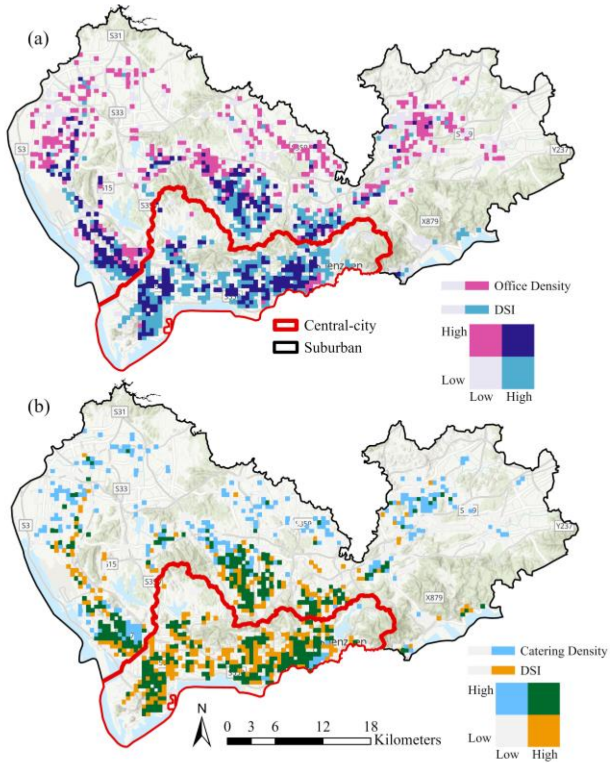


Figure 26. Positive Implications of DSI.

(a) Intersection with Industry Density suggests for business development; (b) Intersection with Catering Density enrich meaning of urban vitality

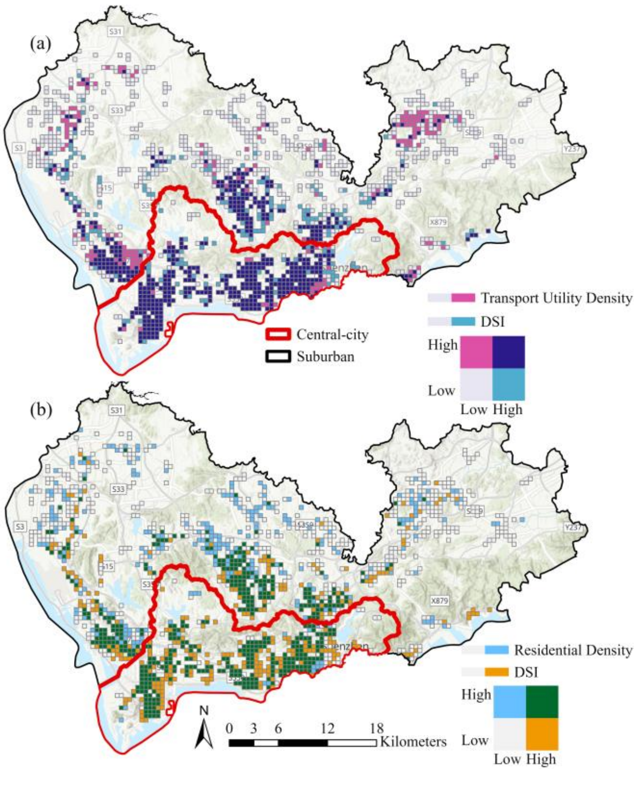


Figure 27. Negative Implications of DSI.

(a) Intersection with Transportation Utility Density reveal inequality issue; (b) Intersection with Residential Density warns on pandemic risk

5.4. Conclusion

This work defines the diversity of spatial interaction and develops a simple yet effective DSI index as a quantitative measurement. DSI index can capture diversity donated by both land and human activity through the lens of spatial interaction. DSI defines that interaction diversity of a location is determined by how it attracts diverse urban flows with consideration of three dimensions: Variety of originated locations, Balance of interaction volume among originated locations, and Disparity of originated locations. Inspired by informatics research, DSI index is the geography version of interdisciplinary citation index.

Some studies have extracted locational indicator from flow data based on network analysis (Cats et al., 2015; Xia et al., 2019; Liu et al., 2021). However, network metrics are more representative for the characteristics in topological space, which may ignore or loss fruitful geographical meanings. Similar to DSI, some recent flow metrics have considered more aspects of geographical meanings, for example the I-index (Wang et al., 2021) and X-index (Wang et al., 2023). I-index depicts the irreplaceability of location (destination) as a function of travel distance and flow volume from the orientations. X-index measures centrality of location (destination) as a function of flow volume and flow directions. However, DSI provides several unique contributions comparing to these two metrics. First, DSI captures and integrates three aspects of interaction diversity as locational diversity metric while I-index and X-index are mainly designed for measuring location importance (e.g., irreplaceability and centrality) based on two aspects. Although spatial diversity was considered as a component of X-index, it is only measured by the flow direction. Whereas the interaction diversity can have fruitful geographical meanings regarding to the not only flow but the orientation locations. Second, there are several parameters need to be determined before calculating I-index and X-index. Although the authors suggested rule to pre-define the parameters, the applicability of the rule may not perform consistently across different scales and cities. While the DSI is better in generality due to its simple form of composite index. The components of DSI are easy to calculate from any scales of geographical flows and there are no parameters to be pre-defined. In this regard, DSI can be used as a basic feature reflecting the locational diversity due to flows.

DSI index is not intended to replace an early glance at the spatial and temporal characteristics of interaction diversity: significant spatial autocorrelation, temporally varied intensity at grid scale, and stable spatial structure at city scale. On a theoretical aspect, quantifying interaction diversity reveals that interaction diversity could be an intrinsic capability of urban places affected by physical form, meanwhile, its magnitude also varies at local scale due to the dynamic nature of human behaviour. On the practical side, DSI is quite important because interaction fundamentally shapes many aspects of city and people. For instance, interaction enhance economy via the knowledge spillover effect, but interaction also increases possibility of disease transmission.

To best of our knowledge, this article is the first one to quantify diversity of spatial interaction. We are not tending to conduct exhaustive analysis on the characteristics of DSI in this article. But more importantly, this work attempts to address the importance of interaction diversity in geographical space and inspire more future work. The use of our empirical results should be aware of study area (Shenzhen) and interaction data used (taxi OD). But still, DSI index proposed in this study is easy to implement and interpret, thus a wide range of big data (e.g., communications, social network) available nowadays can be used to calculate it and provide more insights on interaction diversity. We validated the strong relationship between landuse diversity and DSI, while future work can be also conducted to investigate other influential factors on interaction diversity. With awareness of its contributions and limitations, the interaction diversity and DSI index can benefit for urban planning and policymaking towards a diverse, balance, and vibrant city.

Chapter 6. Urban Vitality: Insights from Flow, Ridership, and Built Environment Diversity

6.1. Motivation

Urban vitality has become an essential concept in assessing the quality of urban development (March et al., 2012; Sung et al., 2013; Sung & Lee, 2015), lack of which could lead to serious planning failures such as ghost cities—large and well-built residential areas within which few people live (Woodworth & Wallace, 2017; Williams et al., 2019). Suggested by abundant qualitative research (Jacobs, 1961; Montgomery et al., 1998; Gehl, 2011), a key component of generating vitality is diversity Yet, current diversity framework for vitality either relies on static feature (e.g., built environment) or first-order mobility feature (e.g., travel intensity), more importantly, lacks flow-induced diversity. Another gap lies on that the relationship between diversity metrics and urban vitality reported in existing literatures are either general pattern (using global regression) or varying patterns but in fixed scale (using GWR). The multi-scale varying patterns need to be further investigated in this research topic. The last minor gap is the necessity to advance current knowledge on application of night-time light images (NTL) in intra-city studies. Overall, for urban vitality analysis, this study aims to contributes to new diversity framework, and application of multiscale regression and new NTL data.

6.2. Research Questions and Methods

This study attempts to investigate the following questions:

- How to advance framework of urban diversity for capturing different facets of city?
- What is multi-scale spatially varying relationship between urban diversity and vitality?
- How model coefficients of diversity metrics can be used to depict spatial-temporal context?

To answer those questions, we develop a framework to analyse the relationship between Vitality Indices and Vitality Proxy (Figure 28). Multiple data source are collected during the 3 weeks of Spring Festival Month of China in 2017 (Figure 29): NTL remote sensing images for extracting vitality proxy; Trip records of metro, bus, and taxi for calculating ridership diversity; Taxi OD flows for calculating interaction diversity, POI data for calculating several metrics of built environment diversity.

Vitality proxy is the dependent variable and other metrics are independent variables, upon which MGWR models are calibrated to infer the multiscale spatial coefficients.

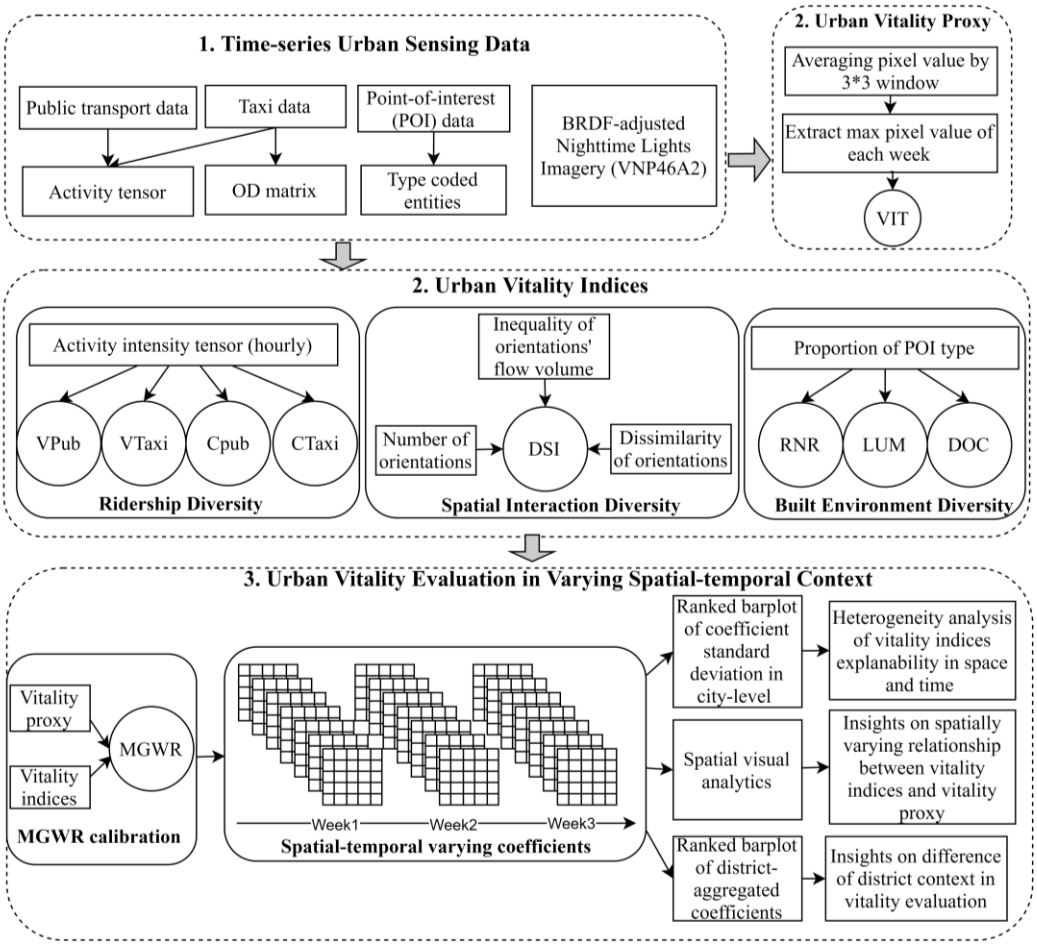


Figure 28. Methodological framework of urban vitality evaluation.

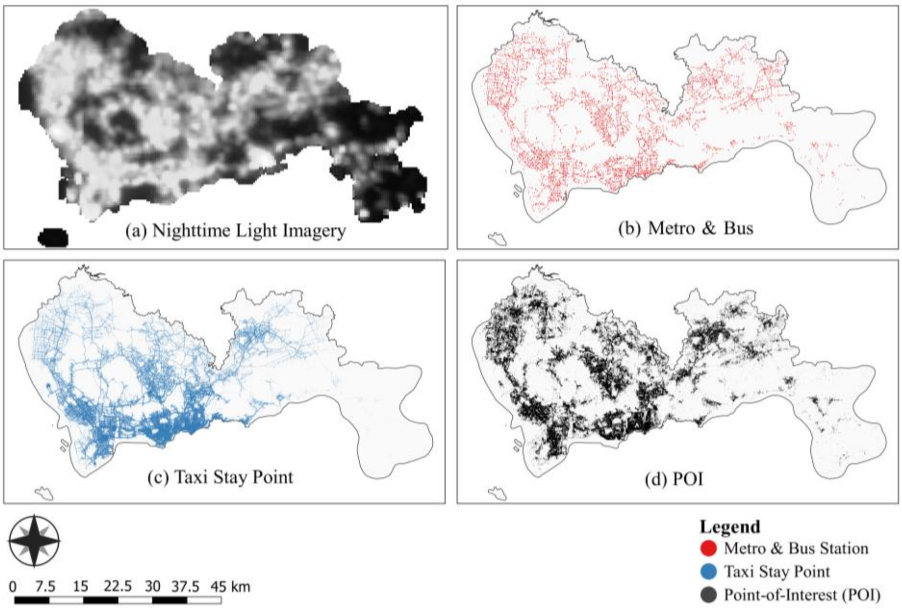


Figure 29. Multi-source Urban Big Data in Shenzhen City.

(a) NTL; (b) Smart Card Data; (c) Taxi Trips; and (d) POI.

6.3. *Results and Implications*

6.3.1. Spatial-temporal patterns of vitality proxy

The processed pixel values based on NTL data effectively proxy the spatial-temporal characteristics of urban vitality. It should be noted again the selected time period has its unique background: the first week as normal week, the second as pre-festical week, and the third as festival week. In Figure 5, intensity of vitality proxy is visualized. It shows that locations with decent vitality are mainly clustered in districts like Nanshan, Luohu, and Futian, and some other distant suburban centres. The value of the vitality proxy tends to diminish with increasing distance from these centres. Although the vitality variation in week 2 and 3 present differences, we still identify from the maps that overall spatial structure are nearly identical.

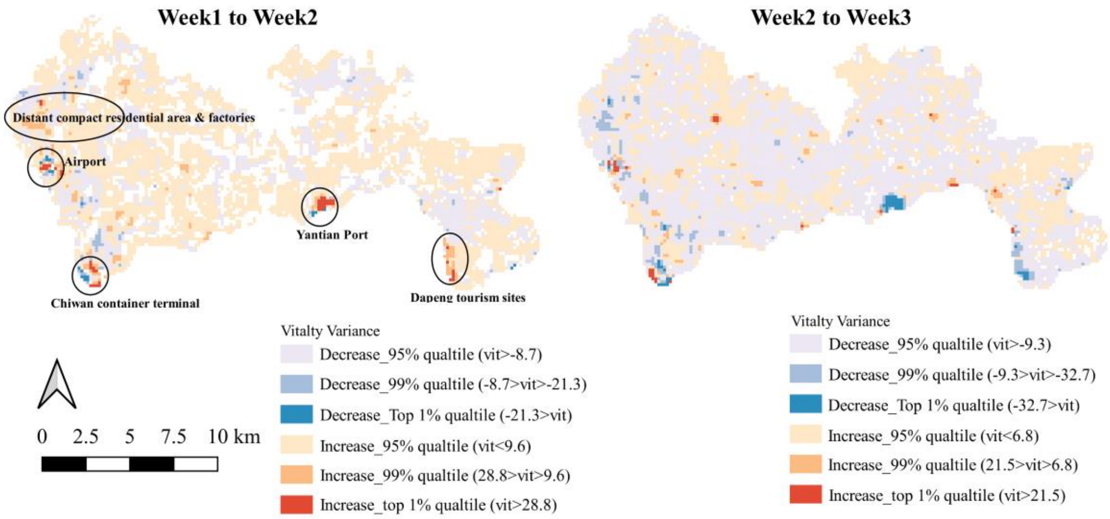


Figure 30. Vitality Proxy Distribution across Space and Time

Temporal fluctuation of the vitality proxy across multiple grids illustrates urban dynamics attributed to festival season. The subtraction is conducted for the vitality proxy of each pixel from week to week (Figure 30). Results show that vitality proxy increases in a variety of locations from the normal week to pre-festival week, as cultural-driven activities would be enhanced such as reunions and tourism (Huang et al., 2019; Liu et al., 2020). During transition to festival week, vitality proxy is found to be shifted to lower levels in wide range of location, while some hotspots in city still face dramatic rise of vitality in this period, such as Northern area of Shenzhen. This phenomenon may be largely attributed to the economic developing of Shenzhen, which has planned distant areas in the North as manufacturing clusters. During festival week, residents are not necessary to travel to urban area in the south for work, instead, sustain and even increase the vitality in those distant regions. These findings demonstrate that NTL-based vitality proxy can effectively capture spatial structure of city and dynamics of human behaviour in temporal dimension.

6.3.2. MGWR model performance and statistics

For each week, MGWR models are derived to explore the local relationship between the vitality proxy and vitality metrics. Assessing model performance is essential for obtaining reliable interpretation of the relationship. The evaluation here relies on goodness of fit of model and significance of the bandwidth (Table 2). The r-squared values for weeks 1 (R2 = 0.931), 2 (R2 = 0.908), and 3 (R2 = 0.918) indicate that the overall performance of all weekly models is satisfactory. Bandwidths of MGWR explicitly mean the size of spatial range to include data points for fitting local models. The results show that all variables have significant inference on bandwidth, judged by the alpha value smaller than 0.05. Notably, the bandwidth significance of flow diversity variable (DSI) is smaller than 0.001, showing the strong relationship between flow diversity and vitality.

Flexible bandwidth is the novel feature that makes MGWR superior to traditional GWR, in terms of additional information obtained on spatial relationship. In this research, obvious differences of bandwidths are noticed on vitality variables. In normal week, VPub, CPub, CTaxi, RNR, and DOC exhibit small bandwidths smaller than 100 m, whereas VTaxi and DSI exhibit around 300 m, and LUM exhibits bandwidth larger than 1000 m. The bandwidth sizes here can be understood as indication of resolution that a variable can significantly depict vitality proxy. According to Table 2, no vitality index explains vitality on a set scale. Most bandwidths differ between normal and festival week. LUM and DSI have consistent bandwidths against time. This is reason that LUM is measuring land diversity that has long-term and stable impact on urban vitality (Yue et al., 2017), scale of which is found to be large in this thesis (1000m). DSI, although as a metric extracted from dynamic human mobility data, exhibit a stable and small bandwidth over all weeks. It shows that diversity of spatial interaction is indeed very related to urban vitality across space and time, indicating its high potential as solid element that should be considered in the future studies.

Table 2. Diagnosis Metrics of MGWR.

	Bandwidth Diagnosis				Overall Diagnosis			
	Band-width	ENP_j	Adj t-val(95%)	Adj alpha	R2	AICc	Residual sum of squares	Model
Constant	10	307.19	3.8	0.000***				
VPub	96	26.89	3.1	0.002***				
CPub	48	41.23	3.2	0.001***				
VTaxi	294	4.04	2.5	0.012*				
CTaxi	82	30.12	3.2	0.002***	0.93	1939.41	88.28	Week1
LUM	1272	1.26	2.1	0.04*				
RNR	50	48.18	3.3	0.001***				
DOC	59	37.44	3.2	0.001***				
DSI	110	16.54	3.0	0.003***				
Constant	10	314.84	3.8	0.000***				
VPub	300	8.30	2.8	0.006***				
CPub	338	7.54	2.7	0.007***				
VTaxi	189	6.62	2.7	0.008***				
CTaxi	110	20.89	3.0	0.002***	0.91	1905.24	114.47	Week2
LUM	1054	1.78	2.2	0.028*				
RNR	179	12.79	2.9	0.004***				
DOC	51	42.96	3.3	0.001***				
DSI	97	19.23	3.0	0.003***				
Constant	10	285.07	3.8	0.000***				
VPub	210	11.23	2.9	0.004***				
CPub	666	4.08	2.5	0.012*				
VTaxi	290	3.45	2.4	0.015*				
CTaxi	34	63.27	3.4	0.000***	0.92	1768.03	94.91	Week3
LUM	1158	1.29	2.1	0.039*				
RNR	52	43.18	3.3	0.001***				
DOC	902	1.90	2.2	0.026*				
DSI	94	17.79	3.0	0.003***				

*** represents significance level of 1%.

* represents significance level of 5%.

In this study, the collinearity issue is also evaluated, although MGWR has been improved to better handle such problem (Fotheringham et al., 2017). The local condition number \and local variation decomposition proportions are used as diagnosis as suggested by Fotheringham et al. (2017). The former is calculated at variable level, while the latter is evaluated at model level (Table 4). All weeks obtained satisfactory results that CN is suggested to be under 30, and the VDP of should be under 0.5. These results argue that the MGWR collinearity is acceptable in this study. In terms of model performance, we compare MGWR with OLS and classic GWR on the same dataset (Table 3). In terms of goodness of fit, MGWR is superior to OLS and GWR, as evidenced by the high r-squared and low AIC. Together with the additional information provided on bandwidth, we argue that MGWR is more suitable for inferring urban vitality model as well as exploring the spatially varying relationship with diversity indices at local scale.

Table 3. Comparison among OLS, GWR, and MGWR Performance

	Week1			Week2			Week3		
	ols	**gwr**	**mgwr**	**ols**	**gwr**	**mgwr**	**ols**	**gwr**	**mgwr**
R2	0.216	0.729	0.931	0.216	0.697	0.908	0.176	0.680	0.918
Adj-R2	0.211	0.639	0.884	0.211	0.603	0.859	0.170	0.590	0.870
aic	3323	2591	1242	3248	2640	1434	3085	2478	1253
aicc		2807	1939		2827	1905		2621	1768

Table 4. Multicollinearity Evaluation of MGWR Models

		Week 1			Week 2			Week 3		
	metrics	**mean**	**std**	**max**	**mean**	**std**	**max**	**mean**	**std**	**max**
Local variation decomposition proportions (VDP)	VPub	0.109	0.150	0.809	0.053	0.083	0.568	0.063	0.091	0.612
	CPub	0.228	0.261	0.975	0.033	0.050	0.392	0.014	0.021	0.155
	VTaxi	0.215	0.245	0.848	0.494	0.299	0.945	0.321	0.264	0.867
	CTaxi	0.208	0.253	0.943	0.244	0.258	0.954	0.408	0.320	0.991
	DSI	0.450	0.342	0.977	0.574	0.292	0.986	0.553	0.333	0.981
	DOC	0.262	0.249	0.959	0.223	0.265	0.939	0.023	0.046	0.465
	LUM	0.006	0.008	0.075	0.013	0.019	0.165	0.009	0.018	0.184
	RNR	0.259	0.267	0.963	0.079	0.115	0.748	0.186	0.240	0.944
Local condition number (CN)		3.702	0.984	7.983	3.684	1.271	9.945	3.643	1.407	11.802

6.3.3. MGWR coefficient analysis

The inferred MGWR models are reliable to depict the relationship between vitality proxy and diversity indices, as evidenced by their model performance illustrated in the above section. In this section, we further extract patterns and insights from the coefficients at multiple scales (city and districts).

Coefficients give two types of information. Positive and negative numbers show whether an increase in the vitality indices predicts an increase or decrease in the vitality proxy. Alternatively, the value of the coefficient reflects the extent to which the vitality indices may explain the vitality proxy. In Table 4, statistical characteristics of model coefficients is presented. We found that vitality proxy is adversely related to public transport ridership diveristy (VPub and CPub) and built environment diversity. Positive relationship is generally found with taxi ridership diversity (CTaxi), DOC, and DSI.

Notably the relationship with ridership diversity differs from that reported by Sulis et al. (2018). In a prior study, variability of travels is positively associated with vitality, whereas consistency is inversely associated. Possible explanation for the contrast is the disparities in the city's general backdrop. Our empirical findings indicate that less within-day variations is indicative of more urban vitality in Shenzhen; in other words, the stability of various forms of transportation is more significant. The taxi industry's good within-day ridership variability (i.e., sudden peaks in daily trip numbers) indicates a more accurate vitality proxy. This result is fair since the abrupt spike in taxi demand represents a variety of travel goals at a given place, which corresponds to a high level of socioeconomic activity. A greater magnitude of CTaxi coefficients are discovered in week 2, indicating that taxis is a more favourable choice in Fesvital week than public transportation. There is comparable evidence that VPub is heavier in the first and the third week, showing that public transportation weighs more in non-festival period.

Table 5. Statistics of MGWR Coefficients

	Min	Median	Max	Mean	SD	Model
Constant	-1.974	0.130	2.815	0.168	0.787	
VPub	-0.165	-0.014	0.629	-0.005	0.108	
CPub	-0.812	0.022	1.237	0.094	0.308	
VTaxi	-0.339	-0.106	0.012	-0.129	0.118	
CTaxi	-0.200	0.013	0.206	0.013	0.075	Week1
LUM	-0.039	-0.037	-0.036	-0.038	0.001	
RNR	-0.462	-0.026	0.365	-0.032	0.108	
DOC	-0.161	0.087	0.451	0.093	0.100	
DSI	-0.096	0.098	0.473	0.117	0.119	
Constant	-1.951	0.222	3.045	0.233	0.797	
VPub	-0.071	-0.024	0.118	-0.020	0.030	
CPub	-0.175	-0.010	0.028	-0.028	0.048	
VTaxi	-0.496	-0.086	0.190	-0.102	0.212	
CTaxi	-0.140	0.014	0.260	0.021	0.087	Week2
LUM	-0.048	-0.017	-0.003	-0.022	0.014	
RNR	-0.134	-0.023	0.159	-0.026	0.064	
DOC	-0.116	0.088	0.681	0.111	0.115	
DSI	-0.163	0.099	0.397	0.115	0.105	
Constant	-2.256	0.218	3.045	0.255	0.851	
VPub	-0.049	-0.003	0.273	0.020	0.059	
CPub	-0.098	-0.048	-0.005	-0.045	0.023	
VTaxi	-0.309	0.020	0.122	-0.033	0.135	
CTaxi	-0.654	0.015	1.430	0.046	0.270	Week3
LUM	-0.035	-0.033	-0.026	-0.032	0.002	
RNR	-0.315	-0.041	0.275	-0.037	0.094	
DOC	0.065	0.096	0.109	0.090	0.015	
DSI	-0.113	0.199	0.572	0.215	0.155	

The LUM and RNR coefficients contradict the commonly held belief that variety of land use is associated with a more accurate vitality proxy. In the majority of instances, the coefficients have negative values according to our findings. This does not imply that built environment is detrimental to vitality; rather, we suggest that other indices, such as ridership diversity, are more positively associated. Due to the fact that all indices jointly describe the vitality proxy together, the substantial positive relationship of some variables may make coefficients of other variables to be negative. In our case, indices extracted from travel flows show more prominent roles, namely ridership variations and DSI. This is also supported by the smaller bandwidth as described in the preceding section. On the basis of our findings, we contend that flow-based metrics explain vitality proxy better. What’s consistent to previous theory is the role of DOC

(Long & Huang, 2017). Although it is not a flow-based indicator, is found to be a robust positive variable for vitality in all weeks.

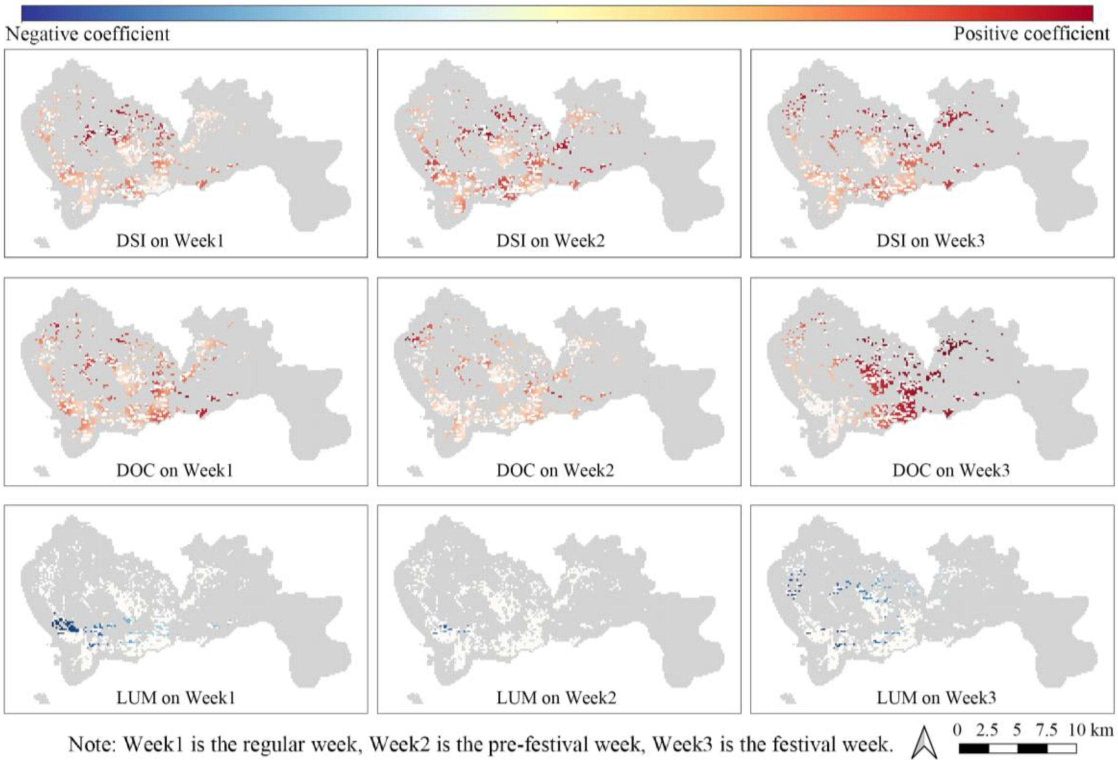
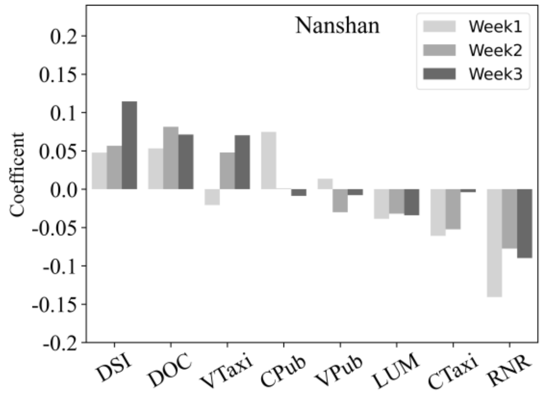
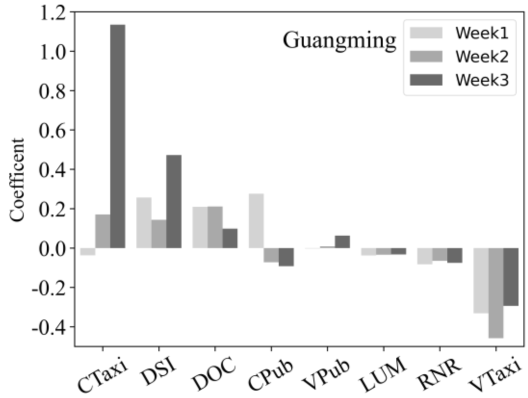


Figure 31. Spatial Distribution of MGWR Coefficients.

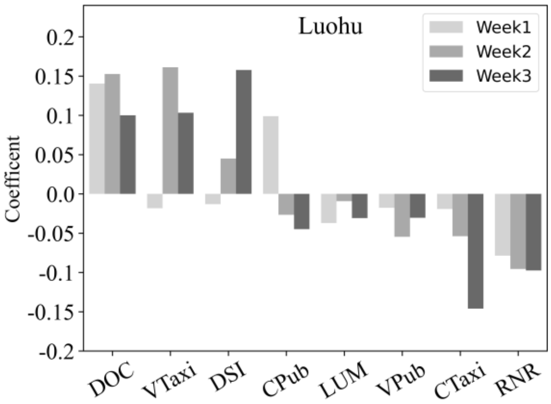
Second, coefficients are projected on maps to enhance comprehension of urban vitality (Figure 31). Different weeks are presented horizontally, whilst diversity indices are presented along the vertical axis. Positive coefficients are depicted in red, and blue is used for negative values. White colour is used to mask the locations with non-significant coefficients determined by t-value. Previous sections only depict the over patterns of coefficients (Table 5), while the spatially varying effects are explained in details in Figure 11. An interesting finding appears on the differences between DSI and DOC, both are stronger positive variable for vitality. We found that DSI is positively correlated in a wider range of places, whereas DOC only in a few hot regions. The maps do show the relationship in urban vitality evaluation could vary from place to place. Consequently, earlier research relying on global linkages to overlook the spatial heterogeneity by large (Delclòs -Alió & Miralles-Guasch, 2018), may result in biased estimation or explanation in some areas.

Standard deviation permits observation of temporal changes in coefficients (Table 5 and Figure 32). This statistic is calculated across multiple places, therefore larger standard deviations suggest greater variation of relationship in space. We ranked all diversity indices by standard deviation value. VTaxi, with the largest standard deviation coefficient, and LUM has smallest deviation for describing city vitality. The coefficients of the public transportation and taxi trip indices vary both spatially and temporally. DSI and DOC vary across space while are stable over time. From a higher level, this analysis offers important guide for future study that may consider how the variation may introduce biases to their model.





(a) (b)



(c)

Figure 32. Coefficient Values across Districts and Weeks.

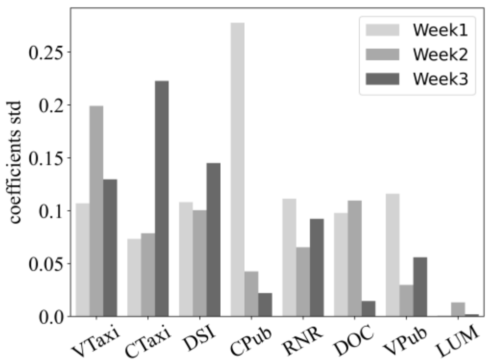


Figure 33. Standard Deviation of Coefficient Values across Weeks.

Third, we undertake additional analysis by plotting district-level coefficients. Due to the fact that administrative boundaries are obstacles that influence numerous crucial urban dynamics and developments, characterizing coefficients for district may yield differing vitality conclusions. According to Lai et al. (2021) and Zhou et al. (2022), Shenzhen's districts have developed unevenly. Using the combination of prominent vitality indices, one can provide an alternate perspective for profiling district vitality. Figure 32 depicts three districts for illustration purposes. Nanshan is a neighborhood recognized for its concentration of technological industries, that is reflected by the strong relationship with DSI. High rank of DSI in Nanshan may relate to diverse business-related motives. In contrast, high rank of CTaxi in Guangming may attribute to the context that it is through a distant district for travellers while being famous as regional hub of manufacturing. Thereby, taxi peaks of daily ridership is stronger variable to depict overall vitality. Luohu’s vitality can be effectively reflected by DOC, which is reasonably related to abundant entertainment and catering streets in this district. Overall, the ranking analysis of MGWR coefficients can deliver fruitful information on the district context, and furthermore on the polices being interested in travel flow or built environment diversity.

6.4. Conclusion

This research developed a comprehensive framework to integrate multisource urban data to examine the spatially varying relationship between urban vitality and urban diversity. By conducting comparative experiments on both normal and festival weeks, the following highlights importance implications:

In multiple ways, the MGWR framework outperforms OLS- and GWR- models. In all weeks, better performance of MGWR models are obtained. The flexible scales of the new models provide additional information beyond coefficients. Throughout the investigation, we found that NTL effectively capture the spatial-temporal structure of city, suggesting its viability for understanding intracity dynamics. Substantial correlation between the vitality proxy and diversity indices contribute to the empirical evidence on the detailed mechanism. Overall, both flow-based diversity and land-based diversity are essential for vitality.

It is important to highlight that the practical implications reported should be limited in Shenzhen. While using the framework developed in this research, further comparison research can be undertaken in other cities. It should be noted that urban vitality is a broad notion encompassing numerous facets of the city. Although we adopt a standard rule (i.e., economic activity strength) to proxy vitality from NTL, it is important to recognize the representativeness of such data. NTL represents more on socioeconomic intensity from the reflective light attributed to infrastructures and human activities. Notwithstanding, integrating NTL in urban vitality framework is still useful because this data is widely available, free, and with decent spatial coverage.

Chapter 7. Conclusion

7.1. Contributions of Each Chapter

First, the thesis starts with a review of urban flows from geography, economic, and complex science perspectives. The **Chapter 2** is a comprehensive review dedicated to urban flows, including not only the examples and historical views on the impacts of flows, but also the shifting theories and methods of flows. The chapter then highlights the research significance and challenge of multi-flow in multi-space that has potential to open up a new research direction to extend the foundation of geographical notion on space and human activity. Multi-flow analytics contributes to realistic representation of complex interaction of city, which benefit many applications such as transportation and travel policies.

Second, after reviewing on historical and contemporary research on urban flows, we develop a methodological framework in **Chapter 3** to address several important aspects and applications of flows in transport sector. This framework clearly sketches out the key methods and related domains. The framework addresses the urban dynamics by including ridership variation metrics, a new metric on flow diversity, and mobility network, and associated urban dynamics applications. Then the construction and analysis of multiplex network is presented in detail, contributing to literatures of multi-modal travel behaviour studies using advance network method.

Third, the case study in New York City (**Chapter 4**) is one of first academic papers to use multiplex network to study shared mobility. The multi-modal network and temporal network developed in this study address our first research question on how to properly integrate multiple flows for quantifying interaction patterns. The multilayer network is widely reported in transportation studies, while is relatively new for geography and urban science researcher, thereby our study contributes an early practice to inspire more future work. Some empirical findings are interesting and new. First, statistical distribution of centralities in multiplex network is not the power-law reported in many human mobility studies. It may indicate the across layers (i.e., modes) connectivity make flow among places smoother than the case of single transport mode. Second, community structure in multiplex network is directly comparable, in which we found that even shared mobility dramatically arise in term of total ridership, while flow-induced urban structures remain similar across layers, indicating the stability of total travel demand (i.e., volume and travel direction).

Fourth, the **Chapter 5** contributes to geography and quantitative urban research by defining and implementing a new metric on diversity of spatial interaction. This work is the first attempt towards such issue, and address the first and the second research question of this thesis. This diversity metric can be calculated from geographical flow data, and the notion of 'diversity' is a key component for vibrant and sustainable city. The metric itself is easy to implement (generalizable) with three highly explainable metric components addressing on location numbers, travel volumes, and location differences. At the end of this study, we demonstrate how interaction diversity can be used as an effective tool for urban evaluation and policy implications in various aspects. Overall, this study contributes to both method and application.

Fifth, the urban vitality study in **Chapter 6** is an updated application of flow analytics towards a better life of city. For urban science, we advance the framework of measuring urban diversity, the core component of urban vitality, by including flow diversity, ridership diversity, and built environment diversity. Analysis of model coefficients contributes to few quantitative research on the relationship between diversity and vitality. From methods perspective, this study provides new application using new data (NTL images) and new method (MGWR) in vitality evaluation. Flow diversity is found to be significantly related to vitality proxy, which is evident for its effectiveness argued in **Chapter 5**.

Overall, this thesis set out to contribute to the modelling and analytics of urban flows by integrating multi-source data and implementing explainable metrics, to make applications in the emerging topics of urban science (diversity, vitality, and dynamic structure). The works made in above chapters have addressed the research question proposed, and we believe that there are still plenty of room for advancing urban flow analytics in order to make a comprehensive understanding of city and its dynamics for a better future.

7.2. Wider Impacts of the Thesis

The framework proposed in this thesis has a significant potential impact on urban flow analysis beyond the current use of mobility datasets. By modifying the definition of flows and layers, the framework can be adapted to represent a wide range of urban flows analysis using various types of locational data, both big and small. For example, geo-located social media and population migration extracted from national surveys can be directly adopted using the framework. The spatial behaviour can be modelled using the same layer definition as did in NYC study, while the social interaction can be treated as other layers where the intra-layer edges are donated by intensity of social communications. In addition, the framework can also represent implicit interactions between locations, such as co-location of people in nearby locations and the correlation of groups of people visiting the same locations. This is achieved through the representation of both explicit and implicit scenarios, allowing for a comprehensive understanding of the urban flows. The representation of implicit interaction in the flow models require more aggressive modification on the layer and edges definition. For example, the co-location / co-interests matrix need be inferred by intersecting spatial-temporal information of entities (e.g., both locations or peoples), similarity of online contents, similarity of friends lists, and so on. Pairs of entities do not have direct interactions / flows with each other, while they are correlated based on other features, so that we call the extracted data model as the correlational matrices. The correlational matrices in nature have the same format as a OD matrix has, thereby we can treat correlational matrices as the OD matrix to construct the multilayer network models, or use the metrics proposed in the thesis to quantify the implicit flow patterns.

The canvas that the thesis attempts to lay out is the intimate relationship between flows and cities. *Places and urban structure influence urban flows*: The physical design of urban spaces, built environment, and even cultural and historical legacies can influence the way people move through and interact with the city. For example, the distribution of public transportation, the design of streets and sidewalks, and the location of amenities have been largely reported to be influential factors in spatial morphology studies. *Urban flows shape the function of places and urban structure*: The movement of people, goods, and information within the city can shape the function and character of urban spaces. For example, areas with high traffic flow can become centers of economic activity and cultural exchange even though the areas might not have dominant spatial characteristics, while low-traffic areas may gradually become neglected or underutilized as the local business can not be sustained. It is no wonder that methodologies implemented is not merely for the case studies for travel behaviour in NYC, urban diversity revealed by DSI, and the economic vitality in Shenzhen, but aiming for longer impacts on future perspectives of flows in the fields of Urban Planning, Economic Studies, and Transportation Studies.

For urban planning, it is urgent to face the challenges raised by highly dynamic urban flows nowadays. Population migration, in conventional perspective, might be determined by the change of the residing cities in decades time window, while a more frequent migration behaviour is witnessed in both China and western cities. Massive moving into cities are triggered by the rapid urbanization process, while in some larger cities in China people are moving out in most recent years due to unbearably high housing price, air pollution, and narrowed margin of leisure time. The regional planning and strategic policies are thereby in much need to model and quantify the spatial flows over cities, to obtain an updated view of regional structure and role of cities in the whole systems, for example, using the network centralities metrics. Similar problems also exist in intra-city level to require quantitative tool to pace up the urban design with urban flows that are densified and interconnected. To promote 'polycentric' design, one could use the proposed framework to unveil the underlying urban structures and explore their relationship to the built environment. Certain environmental elements with high association with urban flows can be built more the over the designed centres. Overall, considering urban flow analytics as a key component in the urban planning practices would provide a more connected view to inform policy-making.

For economic studies, the framework could aid in the development of economic models by integrating spatial-, social-, and economic flows. By doing such, one can better understand how human behaviour is associated with economic status of their own and with the city's vitality. The thesis presented in the Chapter 5 and Chapter 6 that quantifying flow is not an end of the research direction, but a beginning to identify stronger links to urban diversity and urban vitality. In macro level, the flow-based analytics raise new consideration of the dominants factors in driving economic activities. Similar to the impact in urban planning, the identified key amenities and services should be sustained and enhanced to main the role of key spatial nodes (financial centres and hubs). In micro level, site selection is a classic problem in economics. From conventional geography perspectives, the common practices in determining sites wit business potentials depend on the neighbouring features: whether population density is sufficient and the number of competitive businesses. While due to rapid transportation even online shopping, the analysis for site selection limited in local places is obviously not enough. The new definition and implementation of the flow diversity metric, is an effective indicator for quantifying attractiveness and potential in second-order form that consider the activities without constraints of certain geographical scales. Overall, the economic studies involve a great amount of observations on human behaviour, thereby should be ignore the increasing complexified and connected context.

The analysis of multi-dimensional flows is early than the thesis, namely the multi-modal transportation analysis. Constructing multi-modal models match the real scenarios better and therefore provide more accurate simulation on the traffic volumes. However, the urban flows framework of the thesis not only draw on the travel behaviour, but also urban environment factors, and incorporate geographical meaning in the flow metrics and analysis, implying for future transportation studies in several ways. First, the influential factors for traffic volumes not only rely on the transportation system itself, but the geographical features attracting the travel behaviour. For example, the job-housing distributions, the places with high flow diversity value, and the place with high catering densities are the typical factors driving and changing travel behaviour. Second, to further understanding the relationship in the first point, a multi-scale geographically regression approach has indications for transportation studies when evaluating accessibility. The spatial design the transportation facilities normally follow some rule of thumb value of accessibility, for example, a ’30mie life circle’. While the relationship and accessibility could function in varying scales as reported in Chapter 6. Overall, the traffic volume estimation, the infrastructure design, and travel behaviour analysis in transportation studies could embrace a wider and integrated scope of geography, urban, and transportation, by adopting or extend the urban flow framework of the thesis.

7.3. Future Directions and Outlook

There are two main limitations of this doctoral research. First, the proposed framework on urban flow analytics is mainly tested by mobility data (geographical movements flowing among locations), while it should be highlighted that all the analysis of this research can be adopted on other flow data with minor modifications such as the definitions of network nodes, edges, and layers. Although the case studies in the thesis emphasize much on mobility data, while the author totally agrees with the view that human activities not only work for the representations but are intrinsic components to generate more other types of urban flows and behaviours. Future works in testing this framework in other cities will provide a general view on characteristics of urban flows and its relationship with places. Second, one city for testing all the mentioned methods would generate deeper and more comprehensive insights, while the reasons for selecting two cities (New York and Shenzhen) in the case studies of the thesis mainly are the different data privacy standards. The open data in NYC is more comprehensive in terms of spatial-temporal resolution and multiplicity, while at the moment of conducting this thesis, multi-modal travel data in Shenzhen is not accessible for the general researcher. While it should be noted that all the proposed analysis can be directly to one city (e.g., Shenzhen) with increasingly available open data. With awareness of these limitations, this body of doctoral research has several wider and longer impacts as discussed in Section 7.2. Beyond the methodological framework, the conceptual contributions and the perspectives of this thesis may lay the ground for future investigation on the following topics:

1) **Reassessment of basic laws of geography**. Thousands of goods and millions of people are interacting every day through online and offline technologies. The multiple flows studied in this thesis are in travelling context, while the more complex flows beyond geographical space are diversifying our way of life (at the bottom) that eventually may lead to dramatic transformation of the physical form of city. The classic laws in geography largely depend on spatial dependence to explain spatial configuration of city. However, in the multi-flow context, whether spatial dependence need to be extended to, for example, spatial-social dependence is still an open question (See our illustration in Figure 34). The following related question would be the notion of distance. If new notion of dependence is investigated, the notion of distance across the multi-spaces where multi-flows are in will be a key question to explore. During the limited time of my PhD study, I'm interested in involving other flow data such as social media or online logs to study on this direction.

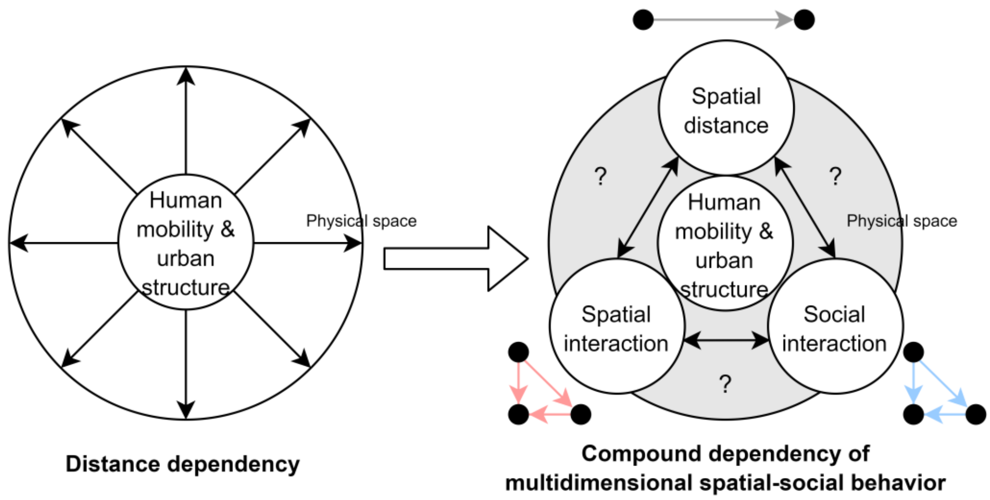


Figure 34. Thematic illustration of spatial-social dependency.

2) **Predicting socio-economic status using multi-flows.** People are more influential than they think, to others and to themselves. How people behave and interact with others are highly influencing socio-economic background of themselves and the object they are interacting with. Reversely, it seems feasible and interesting to evaluate ones or groups' socio-economic status according to behaviour data that is more dynamic and frequently updated. In contrast, traditional way relying on survey is less efficient and costly. The integrated spatial-social flow model proposed should incorporates both implicit (e.g., similar interests) and explicit (e.g. re-post) connections between people for realistic evaluation. The multilayer network introduced in the **Chapter 3** and **Chapter 4** are the promising tool for such purpose. The results have a wide range of applications for studying social segregation and policies for individual development.

3) **Co-evolution of physical form and urban flows.** A research challenge identified in **Chapter 2** has not been studied due to time limitation, that is how urban physical network co-evolve with urban flows. Because settlements and infrastructure are not developing for no reasons, but to serve (potential) travel demand, whilst the demand may change dramatically and make places are reinvented, which will influence how flows can be generated backwards. This is a circle between physical form and human activity, but in the context of flow, this topic has not been well studied. This thesis addresses more on dynamic

network that is constructed from human activity data, but that is not to dismiss the value of static network of city. In future works, we can integrate both static and dynamic in the same network model to better understand the co-evolution issue. This research direction will provide comprehensive tools and results for urban planning.

With more than half the world's population living in cities, I feel strong interest and urge to study on city, and work for its bright future. In face of challenges, city shows magnificent prosperities and resilience. Back to two years ago when Covid-19 started to take over the world, we are concerned about city and regular life. But what we have seen is the strong capability of city and people to adapt: working from home via online software, house price drop at city centres where the spaces are spared for public use. Although technologies do enhance the connectiveness, but resilience and prosperity of city may not hold forever. It is our duties to deepen the understanding of how city work and transform, specifically, urban flows is a promising lens.

References

Agryzkov, T., Tortosa, L., & Vicent, J. F. (2016). New highlights and a new centrality measure based on the Adapted PageRank Algorithm for urban networks. *Applied Mathematics and Computation*, *291*, 14-29.

Ai, P., Liu, Y., & Zhao, X. (2019). Big Five personality traits predict daily spatial behavior: Evidence from smartphone data. *Personality and Individual Difference*s, 147, 285-291.

Albery, G. F., Kirkpatrick, L., Firth, J. A., & Bansal, S. (2021). Unifying spatial and social network analysis in disease ecology. *Journal of Animal Ecology*, *90*(1), 45-61.

Alessandretti, L., Aslak, U., & Lehmann, S. (2020). The scales of human mobility. *Nature*, *587*(7834), 402-407.

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

Althoff, T., Jindal, P., & Leskovec, J. (2017, February). Online actions with offline impact: How online social networks influence online and offline user behavior. In *Proceedings of the tenth ACM international conference on web search and data mining* (pp. 537-546).

Andris, C. (2016). Integrating social network data into GISystems. *International Journal of Geographical Information Science*, *30*(10), 2009-2031.

Bairoch, P., & Goertz, G. (1986). Factors of urbanisation in the nineteenth century developed countries: a descriptive and econometric analysis. *Urban Studies*, *23*(4), 285-305.

Baldassare, M. (1978). Human spatial behavior. *Annual Review of Sociology*, 29-56.

Barabási, A. L. (2013). Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1987), 20120375.

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco J., Simini F. & Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports, 734*, 1-74.

Barrett, C., Eubank, S., Kumar, V. A., & Marathe, M. V. (2004). Understanding large scale social and infrastructure networks: a simulation based approach. *SIAM news*, 37(4), 1-5.

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

Battiston, F., Nicosia, V., & Latora, V. (2017). The new challenges of multiplex networks: Measures and models. *The European Physical Journal Special Topics*, *226*(3), 401-416.

Batty, M. (2005). Network geography: Relations, interactions, scaling and spatial processes in GIS. *Re-presenting GIS*, 149-170.

Batty, M. (2013a). Big data, smart cities and city planning. *Dialogues in human geography*, *3*(3), 274-279.

Batty, M. (2013b). *The new science of cities*. MIT press.

Batty, M., & Cheshire, J. (2011). Cities as flows, cities of flows. *Environment and Planning B: Planning and Design*, *38*(2), 195-196.

Beck, U. (2018). *What is globalization?*. John Wiley & Sons.

Bich, L. (2012). Complex emergence and the living organization: an epistemological framework for biology. *Synthese*, *185*(2), 215-232.

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., & Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics reports, 544*(1), 1-122.

Breiger, R. L. (1974). The duality of persons and groups. *Social forces*, *53*(2), 181-190.

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., & Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, *464*(7291), 1025.

Cai, L., Xu, J., Liu, J., Ma, T., Pei, T., & Zhou, C. (2019). Sensing multiple semantics of urban space from crowdsourcing positioning data. *Cities*, *93*, 31-42.

Cao, H., Feng, J., Li, Y., & Kostakos, V. (2018). Uniqueness in the city: Urban morphology and location privacy. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(2), 1-20.

Cardillo, A., Zanin, M., Gómez-Gardenes, J., Romance, M., del Amo, A. J. G., & Boccaletti, S. (2013). Modeling the multi-layer nature of the European Air Transport Network: Resilience and passengers re-scheduling under random failures. *The European Physical Journal Special Topics*, *215*(1), 23-33.

Castells, M. (1999). Grassrooting the space of flows. *Urban geography*, *20*(4), 294-302.

Castells, M. (2020). Space of flows, space of places: Materials for a theory of urbanism in the information age. In *The city reader* (pp. 240-251). Routledge.

Cats, O., Wang, Q., & Zhao, Y. (2015). Identification and classification of public transport activity centres in Stockholm using passenger flows data. *Journal of Transport Geography, 48,* 10-22.

Cervero, R. (1989). Jobs-housing balancing and regional mobility. *Journal of the american planning association, 55*(2), 136-150.

Chang, H. S., & Liao, C. H. (2011). Exploring an integrated method for measuring the relative spatial equity in public facilities in the context of urban parks. *Cities*, *28*(5), 361-371.

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation research part C: emerging technologies*, *68*, 285-299.

Chen, S., Claramunt, C., & Ray, C. (2014). A spatio-temporal modelling approach for the study of the connectivity and accessibility of the Guangzhou metropolitan network. *Journal of Transport Geography*, *36*, 12-23.

Choudhury, P., Koo, W. W., Li, X., Kishore, N., Balsari, S., & Khanna, T. (2020). Food security and human mobility during the covid-19 lockdown. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (20-113), 20-113.

Christaller, W. (1933). Die zentralen Orte in Suddeutschland, Jena: Gustav Fischer. *English translation:(1966) Baskin CW, Central Places in Southern Germany, Engelwood Cliffs, NJ*.

Claramunt, C., & Thériault, M. (1995). Managing time in GIS an event-oriented approach. In Recent Advances in Temporal Databases: Proceedings of the International Workshop on *Temporal Databases, Zurich, Switzerland, 17–18 September 1995* (pp. 23-42). Springer London.

Claramunt, C., & Theriault, M. (1996, June). Toward semantics for modelling spatio-temporal processes within GIS. In *Advances in GIS II.*

Courtat, T., Gloaguen, C., & Douady, S. (2011). Mathematics and morphogenesis of cities: A geometrical approach. *Physical Review E, 83*(3), 036106.

Crainic, T. G., Gendron, B., & Kazemzadeh, M. R. A. (2022). A Taxonomy of Multilayer Network Design and a Survey of Transportation and Telecommunication Applications. *European Journal of Operational Research*.

Cummins, R. A., Eckersley, R., Pallant, J., Van Vugt, J., & Misajon, R. (2003). Developing a national index of subjective wellbeing: The Australian Unity Wellbeing Index. *Social indicators research*, *64*(2), 159-190.

D'Agostino, G., & Scala, A. (Eds.). (2014). *Networks of networks: the last frontier of complexity* (Vol. 340). Berlin: Springer.

Dang, L., Chen, Z., Lee, J., Tsou, M. H., & Ye, X. (2019). Simulating the spatial diffusion of memes on social media networks. *International Journal of Geographical Information Science*, 1-24.

Davies, C. (2004). Virtual space. *Space: In science, art and society*, 69-104.

De Benedictis, L., & Tajoli, L. (2011). The world trade network. *The World Economy*, *34*(8), 1417-1454.

De Domenico, M., Porter, M. A., & Arenas, A. (2015a). MuxViz: a tool for multilayer analysis and visualization of networks. *Journal of Complex Networks*, *3*(2), 159-176.

De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., Gómez, S., & Arenas, A. (2013). Mathematical formulation of multilayer networks. *Physical Review X, 3*(4), 041022.

De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S., & Arenas, A. (2015b). Ranking in interconnected multilayer networks reveals versatile nodes. *Nature communications*, *6*(1), 1-6.

Delclòs-Alió, X., & Miralles-Guasch, C. (2018). Looking at Barcelona through Jane Jacobs's eyes: Mapping the basic conditions for urban vitality in a Mediterranean conurbation. *Land Use Policy*, *75*, 505-517.

Dickison, M. E., Magnani, M., & Rossi, L. (2016). *Multilayer social networks*. Cambridge University Press.

Dincer, O. C., & Hotard, M. J. (2011). Ethnic and religious diversity and income inequality. *Eastern Economic Journal*, *37*(3), 417-430.

Ding, R., Ujang, N., Hamid, H. B., Manan, M. S. A., Li, R., Albadareen, S. S. M., ... & Wu, J. (2019). Application of complex networks theory in urban traffic network researches. *Networks and Spatial Economics*, 1-37.

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, *60*(11), 2229-2243.

Ducruet, C. (2017). Multilayer dynamics of complex spatial networks: The case of global maritime flows (1977–2008). *Journal of Transport Geography*, *60*, 47-58.

Dugundji, E., Páez, A., & Arentze, T. (2008). Social networks, choices, mobility, and travel. *Environment and Planning B: Planning and Design*, *35*(6), 956-960.

Dutt, A. K. (1969). Intra-city hierarchy of central places: Calcutta as a case study. *The Professional Geographer*, *21*(1), 18-22.

Eagle, N., Macy, M., & Claxton, R. (2010). Network diversity and economic development. *Science*, *328*(5981), 1029-1031.

Engin, Z., van Dijk, J., Lan, T., Longley, P. A., Treleaven, P., Batty, M., & Penn, A. (2020). Data-driven urban management: Mapping the landscape. *Journal of Urban Management*, *9*(2), 140-150.

Finn, K. R., Silk, M. J., Porter, M. A., & Pinter-Wollman, N. (2019). The use of multilayer network analysis in animal behaviour. *Animal behaviour*, *149*, 7-22.

Firestone, S. (2010). Diverse cities and knowledge spillovers. *The Annals of Regional Science*, *44*(1), 1-20.

Forghani, M., Karimipour, F., & Claramunt, C. (2020). From cellular positioning data to trajectories: Steps towards a more accurate mobility exploration. *Transportation Research Part C: Emerging Technologies*, *117*, 102666.

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247-1265.

Fotheringham, A. S., Yue, H., & Li, Z. (2019). Examining the influences of air quality in China's cities using multi‐scale geographically weighted regression. *Transactions in GIS, 23*(6), 1444-1464.

Frank, L. D., Sallis, J. F., Saelens, B. E., Leary, L., Cain, K., Conway, T. L., & Hess, P. M. (2010). The development of a walkability index: application to the Neighborhood Quality of Life Study. *British journal of sports medicine, 44*(13), 924-933.

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.

Gehl, J. (1987). Life between buildings (Vol. 23). New York: Van Nostrand Reinhold.

Ghahramani, M., Galle, N. J., Ratti, C., & Pilla, F. (2021). Tales of a city: Sentiment analysis of urban green space in Dublin. *Cities*, *119*, 103395.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *nature*, *453*(7196), 779-782.

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

Grabowicz, P. A., Ramasco, J. J., Gonçalves, B., & Eguíluz, V. M. (2014). Entangling mobility and interactions in social media. *PloS one*, *9*(3), e92196

Griffith, D. A. (2007). Spatial structure and spatial interaction: 25 years later. *Review of Regional Studies*, *37*(1), 28-38.

Grünwald, P. D., & Grunwald, A. (2007). *The minimum description length principle*. MIT press.

Gu, Y., & Wang, Y. (2022). Using weighted multilayer networks to uncover scaling of public transport system. *Environment and Planning B: Urban Analytics and City Science*, 23998083211062905.

Hajrasoulih, A., del Rio, V., Francis, J., & Edmondson, J. (2018). Urban form and mental wellbeing: Scoping a theoretical framework for action. *Journal of Urban Design and Mental Health*, *5*(10).

Halu, A., Mondragón, R. J., Panzarasa, P., & Bianconi, G. (2013). Multiplex pagerank. *PloS one*, *8*(10).

Han, X. P., Hao, Q., Wang, B. H., & Zhou, T. (2011). Origin of the scaling law in human mobility: Hierarchy of traffic systems. *Physical Review E*, *83*(3), 036117.

Hillier, B., Leaman, A., Stansall, P., & Bedford, M. (1976). Space syntax. *Environment and Planning B: Planning and design*, *3*(2), 147-185.

Hilton, B., Sood, A. P., & Evans, T. S. (2020). Predictive limitations of spatial interaction models: a non-Gaussian analysis. *Scientific reports*, *10*(1), 1-10.

Hochmair, H. H. (2016). Spatiotemporal pattern analysis of taxi trips in New York City. *Transportation research record*, *2542*(1), 45-56.

Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics reports*, *519*(3), 97-125.

Hristova, D., Noulas, A., Brown, C., Musolesi, M., & Mascolo, C. (2016). A multilayer approach to multiplexity and link prediction in online geo-social networks. *EPJ Data Science*, *5*(1), 24.

Hsu, C., Fan, C., & Mostafavi, A. (2021). Limitations of gravity models in predicting fine-scale spatial-temporal urban mobility networks. *arXiv preprint arXiv:2109.03873*.

Huang, J., Liu, X., Zhao, P., Zhang, J., & Kwan, M. P. (2019). Interactions between bus, metro, and taxi use before and after the Chinese Spring Festival. *ISPRS International Journal of Geo-Information*, *8*(10), 445.

Huff, D. L. (1963). A probabilistic analysis of shopping center trade areas. *Land economics*, *39*(1), 81-90.

Huggins, R., Johnston, A., & Thompson, P. (2012). Network capital, social capital and knowledge flow: how the nature of inter-organizational networks impacts on innovation. *Industry and Innovation*, *19*(3), 203-232.

Interdonato, R., Magnani, M., Perna, D., Tagarelli, A., & Vega, D. (2020). Multilayer network simplification: approaches, models and methods. *Computer Science Review*, *36*, 100246.

Jacobs, J. (1961). The life and death of great American cities. New York, NY: Random House.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, *34*(4), 625-636.

Jeub, L. G., Mahoney, M. W., Mucha, P. J., & Porter, M. A. (2017). A local perspective on community structure in multilayer networks. *Network Science*, *5*(2), 144-163.

Jia, C., Du, Y., Wang, S., Bai, T., & Fei, T. (2019). Measuring the vibrancy of urban neighborhoods using mobile phone data with an improved PageRank algorithm. *Transactions in GIS*, *23*(2), 241-258.

Jiang, B., & Claramunt, C. (2004). Topological analysis of urban street networks. *Environment and Planning B: Planning and design*, *31*(1), 151-162.

Jiang, B., & Yin, J. (2014). Ht-index for quantifying the fractal or scaling structure of geographic features. *Annals of the Association of American Geographers*, *104*(3), 530-540.

Jiang, B., Claramunt, C., & Klarqvist, B. (2000). Integration of space syntax into GIS for modelling urban spaces. *International Journal of Applied Earth Observation and Geoinformation*, *2*(3-4), 161-171.

Jiang, B., Yin, J., & Zhao, S. (2009). Characterizing the human mobility pattern in a large street network. *Physical Review E, 80*(2), 021136.

Jiang, S., Alves, A., Rodrigues, F., Ferreira Jr, J., & Pereira, F. C. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, *53*, 36-46.

Kang, C., Fan, D., & Jiao, H. (2021). Validating activity, time, and space diversity as essential components of urban vitality. *Environment and Planning B: Urban Analytics and City Science, 48*(5), 1180-1197.

Kharrazi, A., Qin, H., & Zhang, Y. (2016). Urban big data and sustainable development goals: Challenges and opportunities. *Sustainability*, *8*(12), 1293.

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

Kloosterman, R. C., & Musterd, S. (2001). The polycentric urban region: towards a research agenda. *Urban studies*, *38*(4), 623-633.

Kong, H., Zhang, X., & Zhao, J. (2020). How does ridesourcing substitute for public transit? A geospatial perspective in Chengdu, China. *Journal of Transport Geography, 86,* 102769.

Kou, Z., & Cai, H. (2019). Understanding bike sharing travel patterns: An analysis of trip data from eight cities. *Physica A: Statistical Mechanics and its Applications*, *515*, 785-797.

Ku, S. C. (1998). Southeast Asia and cross-strait relations: The policy of separation of politics and economics. *Journal of Contemporary China*, *7*(19), 421-442.

Kwan, M. P., & Neutens, T. (2014). Space-time research in GIScience. International *Journal of Geographical Information Science, 28*(5), 851-854.

Kwan, M. P., Richardson, D., Wang, D., & Zhou, C. (2015). *Space-time integration in geography and GIScience*. Netherlands: Springer.

La Rosa, D., Privitera, R., Barbarossa, L., & La Greca, P. (2017). Assessing spatial benefits of urban regeneration programs in a highly vulnerable urban context: A case study in Catania, Italy. *Landscape and Urban planning*, *157*, 180-192.

Lai, Y., Tang, B., Chen, X., & Zheng, X. (2021). Spatial determinants of land redevelopment in the urban renewal processes in Shenzhen, China. Land use policy, 103, 105330.

Lewis, T. G. (2011). *Network science: Theory and applications*. John Wiley & Sons.

Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics, 13*(1), 255-269.

Li, W., Liu, X., Yan, C., Ding, G., Sun, Y., & Zhang, J. (2020). STS: spatial–temporal–semantic personalized location recommendation. *ISPRS International Journal of Geo-Information*, *9*(9), 538.

Liang, X., Zhao, J., Dong, L., & Xu, K. (2013). Unraveling the origin of exponential law in intra-urban human mobility. *Scientific reports*, *3*(1), 1-7.

Liang, X., Zheng, X., Lv, W., Zhu, T., & Xu, K. (2012). The scaling of human mobility by taxis is exponential. *Physica A: Statistical Mechanics and its Applications*, *391*(5), 2135-2144.

Liu, K., Yin, L., Lu, F., & Mou, N. (2020). Visualizing and exploring POI configurations of urban regions on POI-type semantic space. *Cities*, *99*, 102610.

Liu, X., Kang, C., Gong, L., & Liu, Y. (2016). Incorporating spatial interaction patterns in classifying and understanding urban land use. *International Journal of Geographical Information Science*, *30*(2), 334-350.

Liu, X., Wu, J., Huang, J., Zhang, J., Chen, B. Y., & Chen, A. (2021). Spatial-interaction network analysis of built environmental influence on daily public transport demand. *Journal of Transport Geography*, *92*, 102991.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., ... & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, *105*(3), 512-530.

Liu, Y., Sui, Z., Kang, C., & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PloS one*, 9(1), e86026.

Loh, J., Green, R. E., Ricketts, T., Lamoreux, J., Jenkins, M., Kapos, V., & Randers, J. (2005). The Living Planet Index: using species population time series to track trends in biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1454), 289-295.

Lokhandwala, M., & Cai, H. (2018). Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of NYC. *Transportation Research Part C: Emerging Technologies*, *97*, 45-60.

Long, Y., & Huang, C. C. (2019). Does block size matter? The impact of urban design on economic vitality for Chinese cities. *Environment and Planning B: Urban Analytics and City Science, 46*(3), 406-422.

Low, S. M., Taplin, D., & Scheld, S. (2009). *Rethinking urban parks: Public space and cultural diversity*. University of Texas Press.

Ma, J., Han, W., Guo, Q., Zhang, S., Wang, J., & Wang, Z. (2016). Improved efficient routing strategy on two-layer complex networks. *International Journal of Modern Physics C*, *27*(04), 1650044.

Maat, K., Van Wee, B., & Stead, D. (2005). Land use and travel behaviour: expected effects from the perspective of utility theory and activity-based theories. *Environment and Planning B: Planning and Design*, *32*(1), 33-46.

March, A., Rijal, Y., Wilkinson, S., & Firidin Özgür, E. (2012). Measuring building adaptability and street vitality. *Planning Practice & Research*, *27*(5), 531-552.

Marston, S. A., Jones, J. P., & Woodward, K. (2017). Human geography without scale. In *Theory and Methods* (pp. 337-354). Routledge.

Martin, G. J. (2005). All possible worlds: A history of geographical ideas. *OUP Catalogue*.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784-2817.

McGee, F., Ghoniem, M., Melançon, G., Otjacques, B., & Pinaud, B. (2019, September). The state of the art in multilayer network visualization. *In Computer Graphics Forum* (Vol. 38, No. 6, pp. 125-149).

Medina-Salgado, B., Sánchez-DelaCruz, E., Pozos-Parra, P., & Sierra, J. E. (2022). Urban traffic flow prediction techniques: A review. *Sustainable Computing: Informatics and Systems*, 100739.

Mikkonen, K., & Luoma, M. (1999). The parameters of the gravity model are changing–how and why?. *Journal of Transport Geography*, *7*(4), 277-283.

Miller, H. J. (1991). Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information System*, *5*(3), 287-301.

Montgomery, J. (1998). Making a city: Urbanity, vitality and urban design. *Journal of urban design*, *3*(1), 93-116.

Morley, C., Rosselló, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: theory and use. *Annals of Tourism Research*, 48, 1-10.

Moss, M. L., & Townsend, A. M. (2000). How telecommunications systems are transforming urban spaces. *Cities in the Telecommunications Age: the fracturing of geographies, Routledge, London*.

Mouratidis, K. (2019). Built environment and leisure satisfaction: The role of commute time, social interaction, and active travel. *Journal of transport geography*, *80*, 102491.

Murakami, D., Lu, B., Harris, P., Brunsdon, C., Charlton, M., Nakaya, T., & Griffith, D. A. (2019). The importance of scale in spatially varying coefficient modeling. *Annals of the American Association of Geographers*, *109*(1), 50-70.

Nash, C. (2012). Genetics, race, and relatedness: Human mobility and human diversity in the genographic project. *Annals of the Association of American Geographers*, *102*(3), 667-684.

Neutens, T., Van de Weghe, N., Witlox, F., & De Maeyer, P. (2008). A three-dimensional network-based space–time prism. *Journal of Geographical Systems*, *10*, 89-107.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, *69*(2), 026113.

Ni, S., & Weng, W. (2009). Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks. *Physical review E*, *79*(1), 016111.

Nicosia, V., Bianconi, G., Latora, V., & Barthelemy, M. (2013). Growing multiplex networks. *Physical review letters*, *111*(5), 058701.

Niu, N., Liu, X., Jin, H., Ye, X., Liu, Y., Li, X., ... & Li, S. (2017). Integrating multi-source big data to infer building functions. *International Journal of Geographical Information Science*, *31*(9), 1871-1890.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab.

Pardikes, N. A., Lumpkin, W., Hurtado, P. J., & Dyer, L. A. (2018). Simulated tri-trophic networks reveal complex relationships between species diversity and interaction diversity. *PloS one*, *13*(3), e0193822.

Peet, R. (1985). The social origins of environmental determinism. *Annals of the Association of American Geographers*, *75*(3), 309-333.

Plötz, P., Jakobsson, N., & Sprei, F. (2017). On the distribution of individual daily driving distances. *Transportation research part B: methodological*, *101*, 213-227.

Portugali, J. (2021). Cities, complexity and beyond. In *Handbook on Cities and Complexity* (pp. 13-27). Edward Elgar Publishing.

Qu, H., Li, J., Wang, W., Xin, W., Zhou, C., & He, Z. (2022). New Insight into the Coupled Grain–Disaster–Economy System Based on a Multilayer Network: An Empirical Study in China. *ISPRS International Journal of Geo-Information*, *11*(1), 59.

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, *82*(2), 263-287.

Rao, C. R. (1982). Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 1-22.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the statistical society of London*, *48*(2), 167-235.

Renne, J. L., Hamidi, S., & Ewing, R. (2016). Transit commuting, the network accessibility effect, and the built environment in station areas across the United States. *Research in Transportation Economics*, *60*, 35-43.

Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., & Chong, S. (2011). On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking*, *19*(3), 630-643.

Riascos, A. P., & Mateos, J. L. (2020). Networks and long-range mobility in cities: A study of more than one billion taxi trips in New York City. *Scientific reports*, *10*(1), 1-14.

Rodrigue, J. P. (2020). The geography of transport systems. *Routledge*.

Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., ... & Masuoka, E. J. (2018). NASA's Black Marble nighttime lights product suite. *Remote Sensing of Environment, 210*, 113-143.

Rousseau, R. (2018). The repeat rate: from Hirschman to Stirling. *Scientometrics*, *116*(1), 645-653.

Sardari Sayyar, S., & Marcus, L. (2011). Urban diversity and how to measure it: An operational definition of classes and scales. In *18th International seminar on urban form, Montréal, Canada, 26–29 August 2011.*

Sarkar, S., Chawla, S., Ahmad, S., Srivastava, J., Hammady, H., Filali, F., Znaidi, W. & Borge-Holthoefer, J. (2017). Effective urban structure inference from traffic flow dynamics. *IEEE Transactions on Big Data, 3*(2), 181-193.

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, *10*(84), 20130246.

Sen, A., & Smith, T. E. (2012). *Gravity models of spatial interaction behavior*. Springer Science & Business Media.

Senousi, A. M., Liu, X., Zhang, J., Huang, J., & Shi, W. (2022). An empirical analysis of public transit networks using smart card data in Beijing, China. *Geocarto International*, *37*(4), 1203-1223.

Senousi, A. M., Zhang, J., Shi, W., & Liu, X. (2021). A proposed framework for identification of indicators to model high-frequency cities. *ISPRS International Journal of Geo-Information*, *10*(5), 317.

Sevtsuk, A., Kalvo, R., & Ekmekci, O. (2016). Pedestrian accessibility in grid layouts: the role of block, plot and street dimensions. *Urban Morphology*, *20*(2), 89-106.

Shaw, S. L., & Yu, H. (2009). A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography*, *17*(2), 141-149.

Silva, S. J., & Nelson, D. (2012). Does aid cause trade? Evidence from an asymmetric gravity model. *The World Economy*, 35(5), 545-577.

Simini, F., González, M. C., Maritan, A., & Barabási, A. L. (2012). A universal model for mobility and migration patterns. *Nature*, *484*(7392), 96.

Smith, G., Archer, R., Nandwani, D., & Li, J. (2018). Impacts of urbanization: Diversity and the symbiotic relationships of rural, urban, and spaces in-between. *International Journal of Sustainable Development & World Ecology*, *25*(3), 276-289.

Solé-Ribalta, A., De Domenico, M., Gómez, S., & Arenas, A. (2016). Random walk centrality in interconnected multilayer networks. *Physica D: Nonlinear Phenomena*, *323*, 73-79.

Solon, J. (2009). Spatial context of urbanization: Landscape pattern and changes between 1950 and 1990 in the Warsaw metropolitan area, Poland. *Landscape and Urban Planning*, *93*(3-4), 250-261.

Stehlé, J., Charbonnier, F., Picard, T., Cattuto, C., & Barrat, A. (2013). Gender homophily from spatial behavior in a primary school: A sociometric study. *Social Networks*, *35*(4), 604-613.

Stewart JQ (1941), An inverse distance variation for certain social distances. *Science* 93: 89–90

Strano, E., Shai, S., Dobson, S., & Barthelemy, M. (2015). Multiplex networks in metropolitan areas: generic features and local effects. *Journal of The Royal Society Interface*, *12*(111), 20150651.

Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: challenges for GIScience. *International journal of geographical information science*, *25*(11), 1737-1748.

Sulis, P., Manley, E., Zhong, C., & Batty, M. (2018). Using mobility data as proxy for measuring urban vitality. *Journal of Spatial Information Science*, *16*, 137-162.

Sun, Y., Fan, H., Li, M., & Zipf, A. (2016). Identifying the city center using human travel flows generated from location-based social networking data. *Environment and Planning B: Planning and Design*, *43*(3), 480-498.

Sung, H. G., Go, D. H., & Choi, C. G. (2013). Evidence of Jacobs's street life in the great Seoul city: Identifying the association of physical environment with walking activity on streets. *Cities*, *35*, 164-173.

Sung, H., & Lee, S. (2015). Residential built environment and walking activity: Empirical evidence of Jane Jacobs' urban vitality. *Transportation Research Part D: Transport and Environment*, *41*, 318-329.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(3), 507-554.

Talen, E. (2005). Land use zoning and human diversity: Exploring the connection. *Journal of Urban Planning and Development*, *131*(4), 214-232.

Tang, J., Liang, J., Liu, F., Hao, J., & Wang, Y. (2021). Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network. *Transportation Research Part C: Emerging Technologies*, *124*, 102951.

Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). A. J. Hey (Ed.). Redmond, WA: Microsoft research.

Taylor, P. J., Hoyler, M., & Verbruggen, R. (2010). External urban relational process: introducing central flow theory to complement central place theory. *Urban studies*, *47*(13), 2803-2818.

Taylor, P., & Derudder, B. (2015). *World city network: a global urban analysis*. Routledge.

Ter Wal, A. L., & Boschma, R. A. (2009). Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science*, *43*(3), 739-756.

Thompson, N. (2020). *Anti-discriminatory practice: Equality, diversity and social justice*. Bloomsbury Publishing.

Tobler, W. (2004). On the first law of geography: A reply. *Annals of the Association of American Geographers*, *94*(2), 304-310.

Tranos, E., & Mack, E. (2019). Big data: A new opportunity for transport geography?. *Journal of Transport Geography*, *76*.

Tranos, E., & Nijkamp, P. (2013). The death of distance revisited: Cyber-place, physical and relational proximities. *Journal of Regional Science, 53*(5), 855-873.

Van Bergeijk, P. A., & Brakman, S. (Eds.). (2010). The gravity model in international trade: Advances and applications.

Vasilyeva, E., Kozlov, A., Alfaro-Bittner, K., Musatov, D., Raigorodskii, A. M., Perc, M., & Boccaletti, S. (2021). Multilayer representation of collaboration networks with higher-order interactions. *Scientific reports*, *11*(1), 1-11.

Verkuyten, M., Wiley, S., Deaux, K., & Fleischmann, F. (2019). To be both (and more): Immigration and identity multiplicity. *Journal of Social Issues*, *75*(2), 390-413.

Vormann, B. (2015). Urban diversity: Disentangling the cultural from the economic case. *New Diversities, 17*(2), 119-129.

Vreeker, R., De Groot, H. L., & Verhoef, E. T. (2004). Urban multifunctional land use: Theoretical and empirical insights on economies of scale, scope and diversity. *Built environment*, *30*(4), 289-307.

Wang, D., Pedreschi, D., Song, C., Giannotti, F., & Barabasi, A. L. (2011, August). Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1100-1108). Acm.

Wang, X., Chen, J., Pei, T., Song, C., Liu, Y., Shu, H., Guo, S., & Chen, X. (2021). I-index for quantifying an urban location's irreplaceability. *Computers, Environment and Urban Systems, 90*, 101711.

Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., ... & Yu, J. (2020, April). Traffic flow prediction via spatial temporal graph neural network. In *Proceedings of The Web Conference 2020* (pp. 1082-1092).

Wang, X., Pei, T., Song, C., Chen, J., Liu, Y., Guo, S., Chen, X., & Shu, H. (2023). X-index: A novel flow-based locational measure for quantifying centrality. *International Journal of Applied Earth Observation and Geoinformation, 117*, 103187.

Wang, Z., Román, M. O., Kalb, V. L., Miller, S. D., Zhang, J., & Shrestha, R. M. (2021). Quantifying uncertainties in nighttime light retrievals from Suomi-NPP and NOAA-20 VIIRS day/night band data. *Remote Sensing of Environment, 263*, 112557.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440.

Wen, T. H. (2015). Geographically modified PageRank algorithms: Identifying the spatial concentration of human movement in a geospatial network. *PloS one*, *10*(10), e0139509.

Williams, S., Xu, W., Tan, S. B., Foster, M. J., & Chen, C. (2019). Ghost cities of China: Identifying urban vacancy through social media data. *Cities, 94*, 275-285.

Woodworth, M. D., & Wallace, J. L. (2017). Seeing ghosts: Parsing China's "ghost city" controversy. *Urban Geography, 38*(8), 1270-1281.

Wu, W., & Niu, X. (2019). Influence of Built Environment on Urban Vitality: Case Study of Shanghai Using Mobile Phone Location Data. Journal of Urban Planning and Development, 145(3), 04019007.

Wu, W., Wang, J., & Dai, T. (2016). The geography of cultural ties and human mobility: Big data in urban contexts. *Annals of the American Association of Geographers*, *106*(3), 612-630.

Xia, C., Yeh, A. G. O., & Zhang, A. (2020). Analyzing spatial relationships between urban land use intensity and urban vitality at street block level: A case study of five Chinese megacities. *Landscape and Urban Planning, 193*, 103669.

Xia, C., Zhang, A., Wang, H., Zhang, B., & Zhang, Y. (2019). Bidirectional urban flows in rapidly urbanizing metropolitan areas and their macro and micro impacts on urban growth: A case study of the Yangtze River middle reaches megalopolis, China. *Land use policy, 82*, 158-168.

Xiong, C., Hu, S., Yang, M., Luo, W., & Zhang, L. (2020). Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proceedings of the National Academy of Sciences*, *117*(44), 27087-27089.

Xu, J., Li, A., Li, D., Liu, Y., Du, Y., Pei, T., ... & Zhou, C. (2017). Difference of urban development in China from the perspective of passenger transport around Spring Festival. *Applied Geography*, *87*, 85-96.

Xu, Y., Belyi, A., Santi, P., & Ratti, C. (2019). Quantifying segregation in an integrated urban physical-social space. *Journal of the Royal Society Interface*, *16*(160), 20190536.

Yang, X., Fang, Z., Yin, L., Li, J., Zhou, Y., & Lu, S. (2018). Understanding the spatial structure of urban commuting using mobile phone location data: a case study of Shenzhen, China. *Sustainability*, *10*(5), 1435.

Ye, L., & Björner, E. (2018). Linking city branding to multi-level urban governance in Chinese mega-cities: A case study of Guangzhou. *Cities*, *80*, 29-37.

Ye, X., & Liu, X. (Eds.). (2019). *Cities as Spatial and Social Networks*. Springer.

Ye, Y., Li, D., & Liu, X. (2018). How block density and typology affect urban vitality: An exploratory analysis in Shenzhen, China. *Urban Geography*, *39*(4), 631-652.

Yildirimoglu, M., & Kim, J. (2018). Identification of communities in urban mobility networks using multi-layer graphs of network traffic. *Transportation Research Part C: Emerging Technologies*, *89*, 254-267.

Yin, C., He, Q., Liu, Y., Chen, W., & Gao, Y. (2018). Inequality of public health and its role in spatial accessibility to medical facilities in China. *Applied Geography*, *92*, 50-62.

Yu, X., Lilburn, T. G., Cai, H., Gu, J., Korkmaz, T., & Wang, Y. (2017). PageRank influence analysis of protein-protein association networks in the malaria parasite Plasmodium falciparum. *International Journal of Computational Biology and Drug Design*, *10*(2), 137-156.

Yue, Y., Zhuang, Y., Yeh, A. G., Xie, J. Y., Ma, C. L., & Li, Q. Q. (2017). Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science, 31*(4), 658-675.

Zeng, C., Song, Y., He, Q., & Shen, F. (2018). Spatially explicit assessment on urban vitality: Case studies in Chicago and *Wuhan. Sustainable cities and society, 40*, 296-306.

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

Zhang, J., Liu, X., Tan, X., Jia, T., Senousi, A. M., Huang, J., Yin, L., & Zhang, F. (2021b). Nighttime Vitality and Its Relationship to Urban Diversity: An Exploratory Analysis in Shenzhen, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 309-322.

Zhang, M., & Zhao, P. (2017). The impact of land-use mix on residents' travel energy consumption: New evidence from Beijing. *Transportation Research Part D: Transport and Environment, 57*, 224-236.

Zhang, X., Huang, B., & Zhu, S. (2020). Spatiotemporal Varying Effects of Built Environment on Taxi and Ride-Hailing Ridership in New York City. *ISPRS International Journal of Geo-Information, 9*(8), 475.

Zhang, X., Xu, Y., Tu, W., & Ratti, C. (2018). Do different datasets tell the same story about urban mobility—A comparative study of public transit and taxi usage. *Journal of Transport Geography*, *70*, 78-90.

Zhao, K., Musolesi, M., Hui, P., Rao, W., & Tarkoma, S. (2015). Explaining the power-law distribution of human mobility through transportationmodality decomposition. *Scientific reports*, *5*(1), 1-7.

Zhong, C., Arisona, S. M., Huang, X., Batty, M., & Schmitt, G. (2014). Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, *28*(11), 2178-2199.

Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., & Schmitt, G. (2016). Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. *PloS one, 11*(2), e0149222.

Zhou, J., & Qiu, G. (2018). China's high-speed rail network construction and planning over time: a network analysis. *Journal of Transport Geography*, *70*, 40-54.

Zhou, X., Yeh, A. G., Yue, Y., & Li, W. (2022). Residential-employment mixed use and jobs-housing balance: A case study of Shenzhen, China. Land Use Policy, 119, 106201.

Zhou, Y., Chen, H., Wu, Y., Wu, J., Li, J., & Chen, L. (2019, February). A Method for Analyzing Pick-Up/Drop-Off Distribution of Taxi Passengers' in Urban Areas Based on Dynamical Network View. In *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 1-8). IEEE.

Zimmerer, K. S. (1994). Human geography and the "new ecology": The prospect and promise of integration. *Annals of the Association of American Geographers*, *84*(1), 108-125.