The Impact of Hukou policy and E-commerce on land attractiveness in China

Gengze Li

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Institute for Transport Studies

March, 2024

Intellectual Property

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Section 2.1, Section 4.1 and Chapter 5 of the thesis has appeared in the publication as follows:

Gengze Li, Shu Wang, Jiyuan Li, The transport impedance disparity indicator: A case study of Xi'an, China, Environmental and Sustainability Indicators (JCR Q1), Volume 18, 2023, 100257, ISSN 2665-9727, https://doi.org/10.1016/j.indic.2023.100257. The candidate was responsible for the development of the research concept, data collection, data processing, coding computer programmes, development of the methodology, modelling work, analysis and interpretation of the results, paper writing and correction. Other co-authors contribute to the work of data production.

The work in Section 4.2 and Chapter 6 of the thesis has been submitted to Cities (JCR Q1) and received an invitation to major revision:

Gengze Li, Ian Philips, David Milne, The impact of E-commerce on land attraction and population distribution: a case study of emerging Taobao villages in China. The candidate was responsible for the development of the research concept, data collection, data processing, coding computer programmes, development of the methodology, modelling work, analysis and interpretation of the results, paper writing and correction. Other co-authors contribute to the work of development of the research concept, suggesting modelling work, analysis and interpretation of the results, review & editing.

Acknowledgements

First, I must express my sincere thanks to my supervisors Dr Ian Philips and Dr David Milne, for their patient guidance and constant support. They are very positive about every idea I had on my research, which enables me to work on my interests. They provide guidance for me. Every comment they made on my drafts and every suggestion they made to me during our meetings had made me closer to my destination. They are strict with my studies but always give me great encouragement when I make progress. They care so much about my work and responds to my questions and queries so promptly. The topic of this thesis is derived from the discussion with them.

I would like to extend my great appreciation to Professor Ronghui Liu. I used to work with her as her teaching assistant. Working with them enriches my knowledge and experience outside my field of study.

I am very grateful to Professor David Watling and Professor Simon Shepherd. They gave me a lots of helpful comments and guidance. Especially Simon helped me to learn system dynamics.

I sincerely appreciate Dr Gillian Harrison. She generously shared her Vensim codes and model with me.

My thanks also goes to Mr. Daniel Johnson and Dr Yue Huang. They help me a lot in ITS.

Special thanks to the data providers, YongSheng Li and Dr Shu Wang for the POI data, Xi'an mapping and surveying Institute for Land use data.

I would like to pay my special regards to my family. My parents try their best to provide me with the best learning and living conditions and environment. Their endless love and unreserved support are indispensable to complete the four years of study.

Finally, to my wife Mengqi, who has been by my side throughout this PhD, and without whom, I would not have had the courage to embark on this journey in the first place.

Abstract

Since 1950, the number of large cities has increased very rapidly – close to 400 now exceed 1,000,000 in the world. From 1980 to 2020, China's Urbanization rate increased from 19.39 to 63.89%. The influx of large number of people into cities in a short period of time brings not only economic prosperity, but also caused urban economic advantages are accompanied by overcrowding, environmental pollution, increasing criminal rate and other negative effects.

To better understand the mechanism of urban development and provide insights that can help policymakers solve these problems, theories and models have been developed. However, the application of the urban model, which originated from Western countries, in China has not been smooth because the model does not take into account China's unique household registration system, also called the Hukou system, which impacts population flows and distribution. Meanwhile, with the development of science and technology, especially Information and Communication Technologies (ICT), e-commerce activities enable regions to break location limitations and join in national and even global industrial divisions. This triggers a new bottom-up rural urbanization process in China that gradually changes population distribution and influences transportation needs. The impact of the Hukou policy and e-commerce on population distribution and human travel needs and behaviours needs to be integrated into urban models.

In order to achieve this goal, this thesis has accomplished the following two objectives:

i) This research proposes a transport mobility disparity indicator to describe the gap between traffic demand and accessibility for different communities. The indicator identifies the accessibility by urban points of interest (POI) (e.g. distribution of working place, park, school, hospital, supermarket, bus, and metro stations, etc.), by community/hukou policy types (local hukou/city residents vs non-local hukou/migrant residents), along with their average transport cost to income ratios. The lower the ratio of the average transport costs to average incomes of a community and the more POI factors in the surrounding areas, the lower the transport impedance indicator of the community, and vice versa. Greater transport impedance disparity indicators between different communities in the same city indicate a larger gap in transport impedance. This can help urban planners identify how severe the unequal relationship is between different areas and provide decision-making support for future urban planning and administrative policies. In this study, Xi'an, capital city of Shaanxi province in China, was chosen as the research object to calculate the transport impedance disparity indicator between local and non-local hukou residents' communities. The result indicates that the transport impedance disparity indicator effectively reveals the different treatments between local and non-local hukou residents' communities in the same city. Compared to local hukou residents, non-local hukou residents experience clear disadvantages in terms of accessibility.

ii) To better understand and analyse the impact of e-commerce on population distribution, based on a concept framework of E-urbanism, which is defined to describe the impact of e-commerce and migration on urbanism, this thesis employs a term "land attraction" to represent the ability of an area to provide job opportunities and life services to meet the needs of residents (e.g., people that are required to be housed, educated, employed, shopping and etc.). Based on that,

the land attraction significantly impacts on population distribution, this research did a case study in China to explore the relationship between land attraction and emerging Taobao villages from 2019 to 2021. The result shows that land attraction can effectively predict population distribution.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Context

China has experienced rapid urbanisation (19.39% in 1978 to 63.89% in 2021) after its reform and opining-up. The population of many cities were increasing rapidly. The largest cities (Called first tier cities) have populations in excess of 10 million, especially Shanghai, which had a population nearly 25 million in 2021. These cities have high population densities in the core urban areas. Even Xi'an, which has just become a first-tier city in recent years, is greater than 30000 people per square kilometre in its core urban area. The influx of large number of people caused those cities grow dramatically in size.

There have been benefits of urbanisation, China's economic aggregate has grown to the second place in the world, and the total manufacturing industry ranks first in the world. However, this has also caused significant problems including traffic congestion, infrastructure overwhelming, environmental pollution, increasing crime rate, deterioration of public health and other urban social environmental and economic issues. In China, while population brings economic advantages to cities, it also causes "*the perennial urban curses of overcrowding*" (Yang, 2013).

However, there is still a gap between China's urbanization rate and the average urbanization rate of developed countries (e.g., the urbanization rate of UK is 83.65% in 2019). It is expected that the urbanisation level will increase until it reaches similar levels to the G7 countries.

According to the Chinese government's plan (*Outline of the Fourteenth five-year Plan for National Economic and Social Development of the people's Republic of China and long-term objectives for 2035*, 2021), from 2021 to 2026, China's urbanization level will increase by 0.8% per year. China is a country with a population of 1.4 billion, which means about 11 million rural people will enter the cities each year. A huge population brings significant challenges to urban management.

Meanwhile, some of urban models that are widely used to support decision-making in urban planning in developed countries, do not work so effectively in China because the context, governance and urban planning systems in China are different to other countries. For example, China's unique Hukou policy, a policy closely related to the social welfare and rights of all Chinese people, significantly impacts on personal incomes, socioeconomic status, travel behaviours and needs. Those models which did not take Hukou policy into account will struggle to establish an accurate population movement sub-model (Niu, 2017; Niu et al., 2014).

Moreover, with the development of emerging technologies, especially Information and Communication Technologies (ICT), certain economic activities that rely on the Internet to organize have been created, also called digital economy which includes digital banking, e-commerce, remote education, smartphone apps, collaboration platforms and etc. The impact of the digital economy on global development continues to deepen. The proportion of the digital economy in the global economy continues to rise (Figure 1.1). In 2021, the scale of the digital economy in China is 45.5 trillion yuan, accounting for 39.8% of Gross Domestic Product (GDP) (Figure 1.1). Based on ICT infrastructure and the prosperity of digital economy, e-commerce activities enable regions to break location limitations and join in national and even global industrial divisions. This triggers a new bottom-up rural urbanization process in China (Zhendong et al., 2017) that gradually changes population distribution and influences transportation needs. However, there is a lack of comprehensive theoretical frameworks for studying the impact of e-commerce on urbanization and population distribution in the world wide yet.

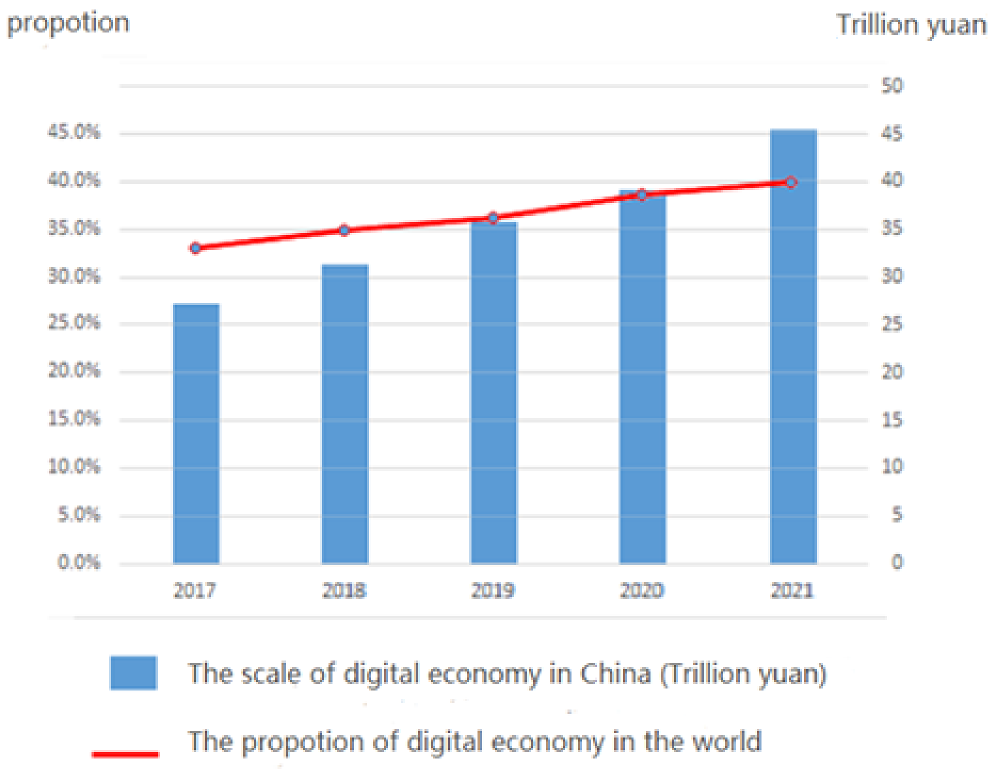

Figure 1.1 Digital economy data of China, 2017 to 2021 (Data source: CAICT)

In general, China needs powerful urban planning tools to face unprecedented challenges. China is experiencing rapid urbanization, and the urbanization rate will continually increase in the future. However, Western urban models don't perform well in China. Because the political contexts are different. Unique local contexts of China, such as Hukou policy, proposes challenges to urban models Meanwhile, emerging technologies, such as e-commerce, are changing Chinese people's life styles and trigger a new urbanization process in China which causes urban management and planning are facing unprecedented challenges. Therefore, powerful tools which adapt China's contexts, are needed to support urban planning decision-making.

1.2 Urban studies

A city is a complex system, it involves social, economic, demographic, transportation, and other aspects (Batty, 2007a) and constantly interacts with external systems by population, material, economy and information exchange (Kennedy et al., 2011, 2007; Wolman, 1965). Any urban area is a highly complex system, dynamic system of interrelated activities and components with many factors continually change over time. To better understand the mechanism of urban development, related models, which are simplified abstractions of real-world phenomena (Batty, 1976a), had been established to analyse and reproduce the process of urban development, evaluate and predict the consequence of urban planning.

Existing research on urban issues has provided many theories and methods to describe and predict urban development. Different disciplines approach this issue from different angles. For example, based on a mixture of natural environment factors (e.g., the locations of natural resources), human factors (e.g., the distribution of population and economic activities) and technology (e.g., transportation), different models have been established to analyse the influence of science, technology, economy, culture, site, policy, transportation, and social structure on urban sprawl, growth, form, function, structure, and development. This has led to many leading theories and meaningful achievements.

1.2.1 Urban modelling

The aims of urban modelling is to establish a framework to understand, analyse, evaluate and reproduce the development of cities, and predict the consequence of certain urban planning and policy. Studies on the mechanism of urban development gradually realized the influence of science, technology, economy, culture, site, policy and social structure on urban sprawl (Osborne, 2005). Land has the ability to attract technological and socioeconomic resources, and the term land attractiveness is defined to describe the ability (Servillo et al., 2012; Tanis, 2018). There is an inseparable relationship between urban development and land attractiveness. Moreover, some problems that did not exist before have emerged over time in cities. Therefore, urban modelling is facing more and more challenges. To accommodate an urbanizing world population, to provide insights which can help policy makers to solve that problem, a theoretical model that accurately analyses and predicts "*urban evolution* (Sjoberg, 1965)" is needed. The main challenge of urban modelling has always been around the unprecedented increasingly complex urban system (Batty, 2014). To understand the mechanism of urban development, elements of cities such as natural and socioeconomic resources and population are divided into many urban factors by character, for example educational, medical, financial, working and other factors. The geographical distribution of urban factors has played an important role in urban development (Batty, 2007a; Osborne, 2005).

Based on the comprehensive philosophical method of analysing urbanization, (Gu, 2019) poses five driving forces of urbanization, "*which are industrialization, modernization, globalization, marketization and administrative/institutional power*". Except the administrative/institutional power, other factors are related to socioeconomic. There is a viewpoint that socioeconomic is very important for urban development, and urban sprawl was significantly associated with urban population density, gross domestic product (GDP) per capita, and industrial structure (Li and Li, 2019). Further, when the spatial heterogeneity was considered, the driving forces of urban sprawl exhibited different magnitudes and directions (Li and Li, 2019). The results indicate that to formulate effective urban planning and land use policies, decision-makers should seriously consider the differences in urban sprawl depending on region, urban size, and hierarchy (Li and Li, 2019).

Although urban models are different in various aspects and details, the purpose of urban modelling is the same. The aim of urban modelling is trying to simulate and analyse the process of how factors formed the drivers and predict the consequence. They try to provide a set of practical tools for urban analyses which seeks firstly to understand and describe the mechanism of urban evolution, including govern urban land use and behaviour of urban system, and secondly to predict the consequences of future policy decisions and scenario of urban planning (Batty, 1976a);(Batty, 2009);(Foot, 2017).

1.2.2 Overview of Urban models

Urban models are designed to help urban planners and decision makers to understand and analyse urban development, and predict the consequence of certain urban planning.

The natural and socioeconomic condition of different areas are different, and causes people who live in the area have different needs. These needs determine the function of cities. Because within a certain area, any city has a certain or several functions to meet the needs of local residents and socioeconomic activities. For example, to meet the needs of transportation, cities, which main function is transportation, are formed (port city, railway hub, highway center and etc.). The needs usually are consequences of the interaction between local natural conditions and socioeconomic activities. For example, factories are often set up in areas where minerals are discovered and creating industrial-related needs. Therefore, urban models have to be sensitive to the context of the cities in order to better serve urban planning and decision making.

In China, emerging technologies and economic activities bring unprecedented challenges. For example, e-commercial activities enable regions to break the location limitation and join in the national or even global industry division, and that significantly impacts on economic organizational forms, people's travel needs and patterns. That requires new theories to understand and explain those phenomena. Urban modelling needs to root in those new theories in order to improve themselves and adapt those changes.

In summary, the main challenge of urban modelling has always been around the unprecedented increasingly complex urban system (Batty, 2014). The number of large cities in China is continually increasing. This in turn leads to cities that are more and more complex.

1.3 Research gaps

Although the previous studies and applications have made a great effort in urban modelling, some research gaps are remaining. The reserch gaps are as follows.

1. Firstly, urban models need to adapt China's political contexts. Many existing urban models are widely used to support decision-making in urban planning in developed countries, and practices have proved that they are powerful tools. However, the application of those models in developing countries are not smooth, because the contexts are so different. For example, the application of LUTI models, which regard

the free location choice as the main driving force for urban land use and transportation development (Cordera et al., 2017), in China is not smooth, because the model does not take into account China's unique Hukou policy (Niu, 2017; Niu et al., 2014), which significantly impacts on people's location choices and travel behaviours. Therefore, transferability is an indispensable part of urban modelling, and there are unsolved transferability issues affecting the application of urban models in China.

2. Secondly, with the development of sciences and technologies, some emerging socioeconomic activities and organization forms have altered people's life style. For example, the Internet, an ICT based communication network, is changing our world and the way people living in a city (Foth et al., 2020), because some socioeconomic and other activities which depend on the internet, they have been called Digital economy (Tapscott, 1996) or e-commercial activities, have dramatically influence on population distribution and gradually changed the city. That require new approaches to socioeconomic simulation. E-commercial activities loosen the location constraints that people may no longer move to cities to interact with urban economies. Therefore, urban modelling, which is rooted in those traditional theories, needs deal with the challenge.
3. Thirdly, to understand, analyse, evaluate and reproduce the urban evolution, and predict the consequence of certain urban planning, an urban model is needed to integrate the impact of local context and emerging phenomena on urban development. Urban development involves social, economic, demographic, transportation, and other aspects to form a complex system. The Hukou policy significantly impacts on people's location choices, and that might well amplify the location of E-commerce and the digital economy. Therefore, in China, to face the

challenges of urbanization goal, an urban development model which integrate the impact of Hukou policy and e-commerce on urban development is needed.

1.4 Research aims and objectives

1.4.1 Aims

The overall aims of this thesis are: (i) to understand the impact of Hukou policy and e-commerce on land attractiveness; (ii) establishing a new urban development model to integrate Hukou policy and e-commerce with related sub-model (e.g., population sub-model and economic activities sub-model) to achieve more accurate and effective urban development modelling. Land use data and urban Point of Interest (POI) data will be the main data source to the spatial analysis model. Those papers, which study the relationship between the distribution of urban factors, land value and house price, are used to analyse and assign different weights for different urban factors. Then, based on different people's needs, those urban factors are selected and used in the spatial analysis model to calculate and visualize the influence zone of urban factors. To accomplish this aim, the following research objectives have been defined.

1.4.2 Objectives

1. Review literature to explore existing approaches to modelling population movements in China. Quantitatively describe the impact of Hukou policy on population distribution. (Research gap 1) According to the gap between local and non-local Hukou residents, including socioeconomic status (e.g., incomes, social welfare etc.) needs (travel needs, living needs, working needs etc.),

2. Describe how existing models could be adapted to consider the impact of e-commerce on the distribution of population and economic activities, modify and improve existing researches to adapt those changes. (Research gap 2)
3. On the basis of existing models, integrate the new population distribution sub-model and economic sub-model to establish a conceptual framework of urban development. (Research gap 3)

1.5 original contribution

The original contributions are as follows.

There is a concept of "*land attraction*" which indicates the collective performance of land use, transportation systems, and the spatial distribution and organization of urban facilities, and determines how well that complex system meets the needs of a certain group of residents. Land attraction is an indicator. The equation for land attraction and the detailed calculation methods are provided in Chapter 4. It usually does not exist independently; instead, it focuses on studying a specific group of residents. For example, Section 4.1 and Chapter 5 used the term "land attraction for local Hukou residents" and "land attraction for non-local Hukou residents" to analyse how well Xi'an meets the needs of local and non-local Hukou residents, respectively. Land attraction is not simply a proxy for land value (as in bid rent theory). For instance, some areas may have higher land value than other areas, but they may not provide working places or something else. Although the land value is very high, but only a small proportion of residents living there. The land attraction is an indicator of the number of residents an area could house (not simply the house value) and the number of accessible jobs which those people are qualified for (so again not simply the best paying jobs) the quantity of services that a person can access that are appropriate to their needs (e.g. affordable food shops not necessarily the best food shops). Based on the assumption that the land attraction significantly impacts on population distribution, Chapter 5 and 6 discuss the spatial distribution of land attraction and population, and find a positive relationship between them.

To quantify the impact of the Hukou policy on land attraction, a disparity index has been created to describe the unequal distribution of transportation resources between local and non-local populations.

On the basis of traditional spatial organization, this research proposes a concept of e-commerce organization to improve general LUTI models. The model not only analyses accessibility but also analyses the socio-economic relationship among urban factors.

In 2019, Yanliu Lin posed a concept framework of "E-urbanism" to divided real world into three interwoven layers by attributes as ICT based economic activities organisations, socioeconomic relationships, and physical world, "namely of ICT infrastructure and production networks, social networks and power relations, and urban form and land use" (Lin, 2019). Based on the E-urbanism framework, this research proposes the E-commerce Four Interwoven Layers (Figure 1.4). This model divides urban areas into four layers, and from bottom to top, those layers are the Urban Form and Land Use Layer (the physical world), the Social Network and Relations Layer (socioeconomic relationship), and the last two layers are both economic activities Layer. According to different organization forms, socioeconomic activities have been divided into a Traditional Spatial Economic Organization Layer, which depends on transport networks, and an E-commerce Organization Form, which depends on ICT infrastructure. It should be noted that there is an intersection between e-commerce organization and traditional spatial economic organization. For example, online retail where information exchanges occur via ICT-based platforms and products are delivered via transport networks.

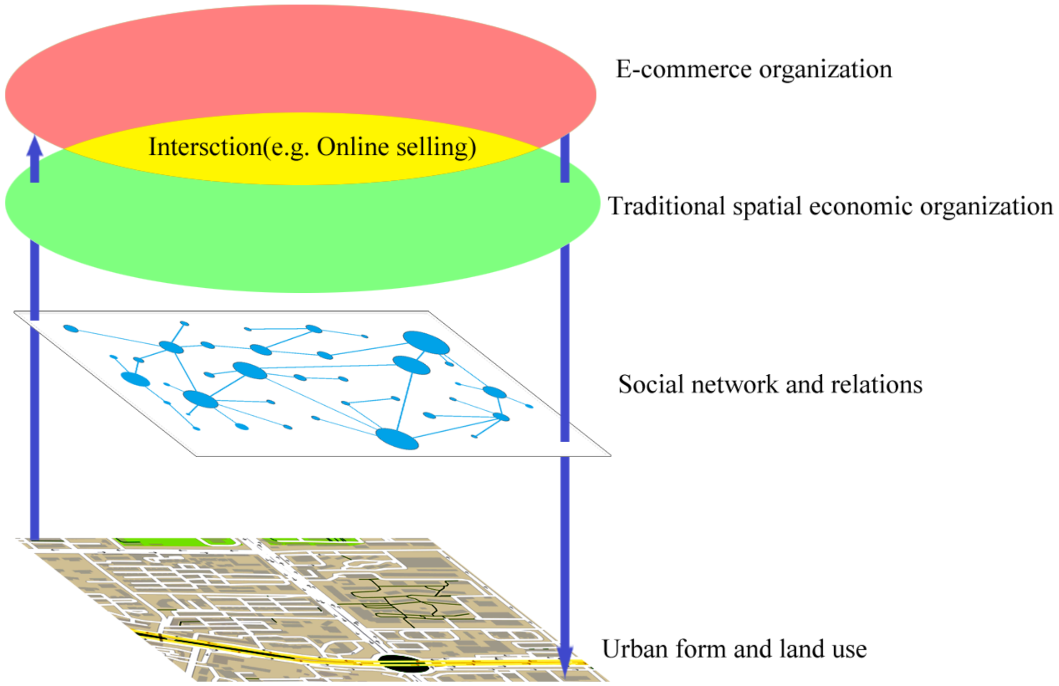

Figure 1.2 E-commerce four interwoven layers

1.6 Structure of the thesis

Chapter 1 introduces the background of the studying topic, the reason why this thesis needs to develop urban models. The most exiting urban models were developed for Western developed countries. Thus, the transferability of urban models is needed to deal with variations in other areas. Chapter 1 introduces in general terms the urban models and urban modelling, and their applications in urban planning. The chapter concludes with a summary of the research gaps, the objectives and the thesis outline.

Chapter 2 provides model details about the background, especially the unique context of China. The outlines of Chapter 2 are as follows.

- The difference between local and non-local Hukou, including social welfare and rights, travel pattern and etc.

- The basic situation of the research area, Xi'an, and why this research chooses it as the research object.
- The urban modelling challenges in China.

Chapter 3 is the literature review chapter which include previous multi-criteria assessment/evaluation approaches on land attractive assessment by Geographic Information System (GIS). The outlines of Chapter 3 are as follows.

- The review of Land use science and urban studies.
- The classification of urban models.
- The review of the impact of Hukou policy on population flows in China.
- The review of land/territorial/regional attractiveness.

Chapter 4 is the Methodology chapter which provides more detail on the steps taken to help other researcher to reproduce the approach. The outlines of Chapter 4 are as follows.

- Overview of methodology.
- The basic idea, methods and data sources used in case study of the impact of Hukou policy.
- The experimental framework used in case study of the impacts of Hukou and the ICT based activities.

Chapter 5 is a case study on Hukou policy. It analyses the different travel needs and travel behaviours between local and non-local Hukou residents. Then, this chapter proposed an indicator to describe the gap of traffic demand and accessibility between local and non-local Hukou residents.

Chapter 6 is another case study on e-commerce on population distribution. It analyses the impact of e-commerce on population distribution and human travel behaviours.

Chapter 7 is the conclusion chapter. It provides an overall summary of the thesis and give recommendations for future work.

Chapter 2
The context of China

To solve the problem of urban issues in China, the first step is to understand the context of China. The context of China is different from western countries. Unlike Western market-based population flow and land-use systems driven by economic factors, political factors have a greater impact on population flow systems and land-use change in China. Therefore, many urban models that originated in western countries usually will face unexpected challenges.

2.1 Hukou policy introduction

The function of Hukou (known as the household registration) policy is not only to control population mobility, but also affect people's lifestyles and fates (Chan and Zhang, 1999). It divides the people into urban (non-agriculture) and rural (agriculture) groups and assigns people to specific districts, shaping important elements of state-society relations in China (Cheng and Selden, 1994), maintaining a strong correlation between population and land, ensuring the stability of agricultural production, because China was an agricultural country in the past. In China, people l inherits the Hukou from their parents. In adulthood, people move their Hukou into another place if they meet the Hukou access conditions set by the local authority they wish to move to. Each Local Authority sets the conditions for moving local Hukous status to their jurisdiction. For example, the conditions for in-migrants to gain local Hukou status in Shanghai differ from those to gain local hukou status in Xi'an. These local conditions generally depend upon their income, education, property and etc. Thus, Hukou is related to the economic status of residents (Cheng and Selden, 1994) (Cui and Cho, 2020). Low economic status and low education are barriers for people to transfer their Hukou., which means that it is difficult for poor or low skilled people to transfer their Hukou to large cities, especially those first-tier cities in China.

At present, the Hukou policy has a significant influence on personal incomes, socioeconomic status, travel behaviours and needs, and it deeply influences the population distribution in China. Within the administrative framework of China, local governments divided into three levels as follows, province, city, “district / county”. District and county are equal-level governments. The difference among them is that usually the urbanization proportion of districts is very high, while the county has a high proportion of its population engaged in agriculture. The jurisdiction of a province includes multiple cities, and that of a city includes multiple districts or/and counties. Hukou policy is closely related to the administrative framework. The registration location of someone’s Hukou is specific to a certain district or county. Usually, different cities have different Hukou policy, including Hukou related social welfare (e.g., pension, social insurance, public services and etc.), rights (their children’s education rights, job opportunities and etc.) and etc. People, who live in the city consistent with their Hukou’s registration location, are regarded as local Hukou residents. Others are regarded as non-local Hukou residents.

The Hukou policy has created phenomena unique to China. Without an urban local Hukou, the rights and benefits of most migrants are limited to certain extent (Wei and Sheng, 2020). In most cases, people can only obtain Hukou related benefits and rights in the place where their Hukou are registered.

Firstly, it is necessary to purchase a car license plate for the car to legally drive on the road after the purchase of the car. And the right to buy car license plates is also related to Hukou policy. For example, without a Beijing Hukou, people are not allowed to buy a car license plate in Beijing. And without that license plate, cars are not allowed to drive into Beijing. Meanwhile, the average income of non-local Hukou residents is usually lower than that of local Hukou residents (Zhang et al., 2019). Therefore, in cities, non-local Hukou residents mainly depend on public transport or e-bikes for transportation needs.

Secondly, children of non-local Hukou residents cannot use local compulsory education resources, which means they have to leave their children back in the place where their children are Hukou registered for education creating a phenomena called “left behind children” (Hong and Fuller, 2019; Zhao et al., 2017). According to the public report of the Chinese government, in 2018, there were 6.97 million rural left behind children, who were not living with their parents and their parents usually working in urban areas.

In addition, disadvantages of non-local Hukou residents usually face a series of social discrimination such as high personal risks and difficulty in job hunting (Jian and Huang, 2007). For example, some job opportunities are only available to local Hukou residents, such as local civil servants and other public sector organizations. Usually, the average wage of public organizations is higher than that of private sector organizations. For example, according to income statistic report of Xi’an, 2020 (*Statistical Yearbook of Xi’an*, 2020), average wages in the public organizations are nearly twice as high as those in the non-public organizations. The wages of many non-local Hukou residents’ jobs do not even reach the starting line of personal income tax, and their employer pay social insurance for them. Thus, local Hukou residents get better pension, social insurance and job opportunities than non-local Hukou residents. Moreover, usually, the Hukou related social welfare and public services are mainly borne by local governments. Therefore, cities with better economic conditions not only proved more job opportunities, but also can often provide better welfare and public services than poor areas. Living without local Hukou in an economically developed areas can still be more economically advantageous to an individual than living in a relatively economically underdeveloped areas.

The change of Hukou policy usually triggers the change of population distribution in China. For example, before the reformation of Hukou system, demographic changes in various regions mainly depended on natural growth. From 1949 to 1980, before the reform and opening-up policy, the urbanization rate in China increased only from 10.64% to 19.39%, and there were very minor cross-regional population flows. However, since the reform and opening-up policy, the Hukou policy has been reformed, resulting in significant population flows in China, with the urbanization proportion increasing from 19.39% to 63.89%. According to the 7th census of China (Ning, 2021), 500 million Chinese people did not live in the district where they are Hukou registered. Therefore, the gap between local and non-local Hukou residents is an inevitable issue, especially in the first-tier Chinese cities. Table 2.1 shows that non-local Hukou resident's proportions of total population of first-tier Chinese cities are quite high. As the city with the highest urbanization rate in China, the proportion of non-local Hukou residents in Shenzhen is 70.84%. As the three largest and most important cities, which were the first to become first-tier cities, the proportion of non-local Hukou residents in Beijing, Shanghai and Guangzhou are 38.5%, 42.1% and 50.22% respectively. The proportion of non-local Hukou residents in Tianjin, Wuhan, Nanjing, Chongqing, Hangzhou, Chengdu and Xi'an are concentrate between 14.48% and 40.4%.

China's land-use process is top down. For example, the classical general LUTI model assumes that people are free to relocate and that interaction between economic participants occurs via transportation networks. However, political factors, such as Hukou system, are significantly impact on population flows in China. Thus, LUTI models usually focus on population distribution, not on land-use change. Moreover, Chinese land-use change is a top-down process that local governments develop detailed land-use plans, so they can predict consequence of urban planning decisions.

Table 2.1 The proportion of non-local Hukou residents is in first-tier cities in China, 2020. (Data source China seventh National census)

City	Permanent population (million)	Non-local Hukou population (million)	Non-local Hukou Proportion
Beijing	21.893	8.418	38.50%
Shanghai	24.871	10.479	42.10%
Tianjin	13.866	3.5348	25.49%
Wuhan	12.326	3.9454	32.01%
Nanjing	9.314	1.3484	14.48%
Guangzhou	18.676	9.3788	50.22%
Shenzhen	17.56	12.4387	70.84%
Chongqing	32.054	4.8114	15.01%
Hangzhou	11.936	3.7877	31.82%
Chengdu	20.937	8.4596	40.40%
Xi'an	12.952	3.7469	28.93%

The local government controls the proportion of local and non-local Hukou residents, and sets the total population according to its own needs. For example, in order to control the size of cities and avoid problems such as traffic congestion and infrastructure overloading, the Hukou access policies in Beijing and Shanghai set high conditions for those who wish to transfer their Hukou into those cities to avoid the rapid growth of the total local Hukou population. Meanwhile, local government decision about the Hukou system impacts on the total number of non-local Hukou population by affecting the total number of job opportunities available to non-local Hukou residents. A certain proportion of non-local Hukou residents not only contribute to local labour force, but also reduce the local government's social welfare expenditure. Because the social welfare expenditure provided by local government to local Hukou residents is higher than that of non-local Hukou residents.

The difference in Hukou related rights and the restrictions places on spatial location choices created a phenomenon in China's big cities, known as the village in city (also called Chengzhongcun) a crowded place that lacks public leisure areas and public services, and provides a large number of cheap rental houses. Almost all of the tenants living in the village in city are non-local Hukou residents. The detail of the village in the city will be discussed in the next section (Section 2.1.1).

The general gap between local and non-local Hukou residents significantly impacts their selection of settlement location, travel needs and behaviours in China.

2.1.1 Village in the city (Chengzhongcun)

The phenomenon of village in the city is the consequence created by China's land ownership system and rapid urban sprawl. Land ownership in China can be divided into two categories as follows, public ownership and collective ownership. The meaning of public ownership is that the land belongs to the state and is managed by government. The meaning of collective ownership is that the rural residential land belongs to all villagers who's Hukou located in the area and the area is self-managed by the villagers collectively to a certain extent. Except for rural residential land, all other land belongs to the state.

Because of rapid urban sprawl, local municipal governments bought land from villages, which are near the built-up area of cities, to develop new urban area. Some villages have only sold their farmland, and their residential land have been preserved. Although those residential lands are surrounded by the urban built-up areas and the Hukou of those villagers have been transformed from agriculture to non-agriculture group, the residential land belong to the villagers. The villagers built a large number of apartments for rent, and the rent fee is very cheap. Because unlike commercial community, villagers do not need to invest money to buy land from local government, so their cost is very low.

Since the income of non-local Hukou residents is relatively low, the cheap rent of the village in the city significantly contributes to solving the problem of house rent for non-local Hukou residents (Li, 2018). For example, according to "*Rental monitoring report for 40 key cities nationwide*"(*Report on the Development of Housing Rental Market in China*, 2023), the average rent per house in Xi'an's commercial residential area is 2,144 yuan per month which accounts for 61% of the per capital disposable income of residents (including local and non-local Hukou residents) in Xi'an. Non-local Hukou holders hardly can afford that because the average incomes of non-local Hukou holders are lower than that of local Hukou holders. In 2019, the average income of public sector organizations and private sector organizations is RMB 96867 and 50073 yuan per person per year, and the job opportunities of the public sector organizations are only available for local Hukou residents. The job opportunities of the private sector organizations are available for both local and non-local Hukou residents. Usually, the rent fee in village in the city is only 200 to 500 yuan per month. Non-local Hukou residents usually sent their money to their families and they will do everything possible to save money. In fact, in Xi'an, in 2009, almost all non-local Hukou residents, 2.4 million, were living in village in the city (Wang and Li, 2014). Thus, it is reasonable to assume that a large proportion of non-local Hukou residents live in village in the city at the present time. Therefore, this research regards all villages in the city as non-local Hukou residents' communities in Xi'an. And the rest communities are local Hukou residents' communities.

Figure 2.1 shows a typical village in the city. Because the land belongs to villagers, the local government is unable to make urban planning for this area. The villagers build as many houses as possible on their land for rent. Although the area is less than 0.2 square kilometres, about forty thousands of people live in this area. The average population density of built-up areas in Xi'an is about ten thousand people pre square kilometre in 2019. That means the population density of the village in the city is usually several times that of urban built-up areas in Xi'an.

Figure 2.1 A village in the city (inside the red lines), Xi'an

2.1.2 Hukou policy influence on population flows

China Hukou policy usually is regarded as a powerful tool for regulating the population size in urban areas (Chan et al., 1999) (Chan and Zhang, 1999) (Chan, 2019) (Zhang et al., 2019). For example, from 1949 to 1980, under the restriction of Hukou policy, there had been little cross-regional population flows in China.

People tend to migrate from areas with lower welfare levels to areas with higher welfare levels. For example, in 2019, the welfare expenditure in education, medical and pension of Xi'an is about RMB 1,400 yuan per person per year, and that of the Shaanxi province is about RMB 500 yuan per person per year. Therefore, those residents tend to migrate from Shangluo, Hanzhong, Weinan and other cities in the Shaanxi province to Xi'an. In the last 40 years, cities with higher welfare budget, such as Beijing and Shanghai have had the largest population growth. The inclination of policies (e.g., Hukou policy) has also driven the inland population, capital and foreign advanced technology resources to concentrate in these areas to form agglomeration effects in order to stimulate the rapid growth of Beijing, Tianjin and eastern seaboard cities (Houkai, 2014). This agglomeration effect will form a continuous and powerful attraction to the surrounding area, making resources and talents further gather themselves, thus further promoting the development of the city and formed a positive cycle. For example, from 1978 to 2018 in Shanghai, its permanent population increased from 11.04 million to 24.24million, and GDP increased from RMB 27.281 to RMB 3267.987 billion. The population and economic growth rate of Shanghai was much higher than the national average rate at the same period.

By raising or lowering the conditions for Hukou relocate policy, the probability of the population flows entering a certain area will increase or decrease. Super-lager cities (e.g., Beijing, Shanghai and Guangzhou) set up complex and difficult Hukou access policy conditions to control the size of local Hukou population. Those conditions for migrants, who wish to relocate their Hukou to one of these super-large cities, have weakened people's willingness to migrate. For example, people, who are not local Hukou residents of Shanghai and wish to relocate their Hukou to Shanghai, should meet certain conditions as follows. They should work, live, pay their tax (e.g., personal income tax) and social insurance (e.g., medical insurance and old-age insurance) in Shanghai for at least seven years, and they should have a master degree or professional and technical intermediate certificate. Meanwhile, conditions for non-local Hukou residents have become less attractive making it difficult for them to reside in the city for long enough to change hukou status. For example, children non-local Hukou residents cannot gain compulsory education services in the local city. Their parents will choose to return home to accompany their children after earning a certain amount of money. From 2016 to 2021, the number of non-local Hukou residents in Shanghai decreased, and the total permanent residents remained at 24.7 million. By contrast, some cities, such as third-tier and fourth-tier cities, have cancelled or reduced the conditions for Hukou access policy to increase the size of the local population.

2.1.3 History and role of Hukou policy in rapid urbanization in China

The process of urbanization in China is slow at first and then increased suddenly, especially in the 21st century. In the 20th century, urban and rural development was significantly unequal in China, and high-quality public resources were concentrated in cities. For example, at the beginning of 1980s, one-tenth of China's total urban population concentrated in the three metropolises of Beijing (Capital city of China), Shanghai (Economic centre of China), and Guangzhou (the primary port in the south of China), making up 1% of the total country population (Mann, 1984).

Before 1978, China created an urban-rural dual world under the Hukou system, with very limited population flow and huge differences between rural and urban areas (Wu and Treiman, 2004). The inequality between urban and rural areas is not only reflected in the price differential between urban industrial products and rural raw materials, but also in the polarization in the distribution of infrastructure and public resources. For example, for a long period in China, there has been a huge difference of social welfare and treatment between urban and rural Hukou in the same district. For example, the progress of infrastructure investment in rural areas is significantly slower than that of urban areas. The water system and streetlights are common in cities and did not cover rural areas until 2005.

After (1978), China has gradually reduced restrictions on migration (Shen, 2002). Due to those advantages that cities have over rural areas, a large number of rural migrants are attracted to work in cities (Chuankai et al., 2017). After a huge number of labours concentrate in cities, it strongly increases the development of cities. The size of the built-up areas of cities expanded rapidly during this period and the larger cities also created more jobs and enhance population's migration intentions of moving to cities. That formed a positive cycle. During this period, China's economy entered a long-term double-digit growth.

Part of peasants-workers, people with rural Hukou status whom home settlement in the countryside and working in urban areas, migration choose to use the money they earn in cities to support their families members who is living in their hometown in the countryside (Van der Ploeg et al., 2014). It has formed a trickle-down effect in which cities feed back to the countryside (Aghion and Bolton, 1997). However, cities attract a large number of rural people, which also deprived the human resources and potential markets for rural development, and further aggravated the polarization effect of the gap of rich and poor between urban and rural areas. This gap will accelerate the concentration of population from rural areas to cities (Chen et al., 2009).

Actually, from 1980 to 2020, while the urban population has grown rapidly and China's total population increased from 987 million to 1.4billion, the rural population has shrunk from about 800 million to 560 million. Economic development, population growth, urban expansion promote each other in the progress of urbanization. However, urban expansion is not always positive. On the one hand, those areas were covered by the influence of these cities will inevitably face the loss of population and resources which will increase the imbalance of regional development. For example, China's southeast coastal economic belt has concentrated more than 40% of China's population and 60% of GDP. The per capita GDP of Zhejiang Province, in the China's southeast coast, is several times that of Ningxia Province,in the northwest of China. In many places, entire village populations have migrated to cities or towns and left empty houses which called "*hollow villages* (Sun et al., 2011)".

Although there are "*home appliances to the countryside"* (the government provides subsidies to sell home appliances to the countryside at low prices)" and "*new rural construction* (the government invests heavily in the infrastructure, for example construction of rural tap water networks, roads, street lights, and communication networks)" policies which aim to reduce the gap between urban and rural areas, the gap between the two in employment, education, culture and entertainment, and medical care is still huge. Furthermore, it is still difficult to create enough jobs in rural areas to retain the population. Therefore, the decline of the countryside is almost inevitable.

It is noted that the concentration of population in cities has both positive and negative effects on urban development. For instance, between 1980 and 2012, China's Urbanization increased from 19.4 to 52.6%. Unfortunately, this growth "*accompanied by the perennial urban curses of overcrowding, air and water pollution, environmental degradation, contagious disease, and crime*" (Yang, 2013). As a consequence, serious infrastructure overloading, abnormally high land and housing prices, environmental degradation and pollution, and rising crime rates have hindered China's further urbanization and the economic benefits from it.

2.2 Background of urban issues in China, history and land policy

2.2.1 The history of China's urbanization, from 1950 to 2020

From 1950 to 1979, the internal mobility of the Chinese population was not high. Farming was almost the only means of livelihood for the agricultural population; Cities cannot provide new jobs rapidly. Thus, the population migration between urban and rural areas barely existed. Inside a city, urban residents had deep connections with their institutions, companies or other kind organizations where they worked. Almost every organization had its own living zone where can provide almost all living services such as living, dining, haircuts, etc. At that time, almost all organizations belong to the Chinese government and they provided permanent jobs. Therefore, the population migration inner urban areas hardly existed. Under these circumstances, the urban planning and management of China were concentred on state-owned organizations.

Since the reform and opening up in 1978, China has chosen to concentrate limited resources in specific areas to promote local development. The Chinese government relies on these developed areas to feedback other areas, and finally realize China's overall development (Chuang-Lin, 2009). First, in 1984, the 14 port cities of Dalian, Qinhuangdao, Tianjin, Yantai, Qingdao, Lianyungang, Nantong, Shanghai, Ningbo, Wenzhou, Fuzhou, Guangzhou, Zhanjiang, and Beihai were selected to be open. Later, the port cities were used as the fulcrum to establish the Bohai Rim Economic Circle, the Yangtze River Delta Economic Circle and the Pearl River Delta Economic Circle. China has successively established special economic zones such as Shanghai, Shenzhen, Zhuhai and Hainan. Therefore, China's opening-up is divided into regions, from the coastal areas to the inland areas.

In this period, China's reform and opening up is a sequence of policy reforms from the south to the north, from the coast to the inland. Therefore, the distribution of economic activities and population in China have a distinct characteristic which has been called one country, three worlds (Gao et al., 2014). This phenomenon creates three major sectors in China, developed sector in the eastern China, developing sector in the middle China and underdeveloped sector in the western China. A developed central city usually is the core of the region and drives the economic development of surrounding cities. The population and economic growth rate of Shanghai was much higher than the national average rate at the same period. Economic development, population growth and urban expansion promotes each other in the progress of urbanization. However, urban expansion is not always positive. On one hand, those areas covered by the influence of these cities will inevitably face the loss of population and resources, whichl increases the imbalance of regional development. For example, China's southeast coastal economic belt has concentrated more than 40% of China's population and 60% of its GDP. The per capita GDP of Zhejiang Province, which is in the China's southeast coast, is several times that of Ningxia Province, where in the northwest of China. Therefore, those first-tier cities which close to the eastern coastline of China, such as Beijing, Shanghai, Hangzhou, Shenzhen and Guangzhou, had formed a developed urban agglomeration along the eastern coastline. The distribution of the population is very similar to this phenomenon.

2.2.2 The land policy and land market in China

The land market in China is dominated by the government who defines the land use type (e.g., residential, medical, educational land use types and etc.) of different zones before the land enters the land market. After organizations or individuals have purchased the land usage right of a certain zone, the land use type of the zone is difficult to change. Therefore, the processes of urbanization and land policy making are top-down processes. Because of that, urban models are usually used to support land policy making by predicting the consequence of a certain urban planning scenario. For example, (Niu et al., 2019) used LUTI model to predict the consequence of Beijing's sub-centres planning object and suggested to consider the impact of migration on LUTI model to achieve a more accurate model. After the land policy is formulated, urban models usually will focus on population distribution and public service, rather than land use change. For example, By considering the specially residential layout systems in Changsha, (Wang et al., 2018) used LUTI model to design transport policy to increase the urban traffic efficiency. (Niu et al., 2018) considered the impact of housing rent distribution on LUTI model to forecast the change of population distritbution.

2.3 Introduction of research area, Xian

This study selected Xi'an as the research object which is representative of China's large cities in terms of geographical location, economic conditions, natural conditions, and political conditions. Xi'an, the capital city of Shaanxi province, is one of the most important and oldest cities in China. From the comprehensive consideration of location, geographical conditions, industrial distribution, economic situation, population and history, this research chooses Xi'an as the research object to do a case study. Xi'an is located in the geometric centre of China, on the geographical dividing line between North and South. Its natural conditions do not have particularly advantages, nor particularly disadvantages. It does not have a large amount of rich mineral resources, and it is located in plain area and does not lack water or other resources. In the China's economic rankings, Xi'an is in the middle rank of all big cities. Among the 35 most important big cities in China, Xi'an ranks 19^{th} in GDP in 2022 and ranks 15^{th} in total population. There are almost all kinds of industries in Xi'an, including health care, education, tourism, manufacturing, finance, agriculture and so on. In other words, Xi'an has the common characteristics of most Chinese big cities. Xi'an is representative of the new first tier cities, but is not in the eastern seaboard region. Therefore, the study of Xi'an is representative to a certain extent. More importantly, the data availability of Xi'an in this study is the best.

2.4 The impact of ICT based urban management and E-commerce on China

E-commerce activities, including but not limited to remote education, remote working, remote medical care, online sales and entertainment, greatly impacts on socioeconomic development and proposes challenges for urban planning. With the Information and Communication Technologies (ICT) boosting, ITC based E-commercial activities enable regions to break the location limitation and join in the national even global industry division, and that trigger a new bottom-up rural urbanization process in China (Zhendong et al., 2017). The 50th Statistical Report on China's Internet Development indicates which provided by China Internet network information centre (http://www.cnnic.net.cn/), China has the largest population of Internet users in the world. The coverage rate of ICT infrastructure in urban areas is close to 100% and 90% in rural areas. ICT based economic activities, has caused dramatic social and spatial reorganization of economic activities. It has greatly increased the demand for remote work that enforced and accelerated the impact of ICT on transportation. ICT has reorganized the way of economic activities that distribution of economy participants depends on the spatial relationship and social network relationship and ICT infrastructure. Certain economic activities, such as information exchange no longer be limited by spatial distance under the ICT support. It dramatically decreases related transportation needs. It affects the distribution, circulation and exchange of population, and further reforms urban planning, economic organization, and policymaking. Classical urban models are facing unprecedented challenges in that social and economic activities not only depend on the spatial neighbourhood and distance, but depend on ICT based on socioeconomic relational networks of urban factors.

The application of ICT technologies to urban development pose new challenges for urban planning. Those unprecedented technologies especially, information and communication technologies (ICT). have changed the lifestyle of people (Greenfield, 2004). Based on the use of ICT to integrate urban resources, it presents a new city pattern that urban resources can be allocated across multiple departments (Shen et al., 2018) which can provide urban planning and management decision-making support for decision makers to "*modify urban infrastructures, public and private services and governance activities* (Dameri and Ricciardi, 2017)".

Digital economy (Tapscott, 1996) based on ICT has gradually changed the city. Although a growing body of literatures has studied digital economy development, there are few of theoretical and empirical researches to understand the impact of digital economy on migration, being one of the key driving forces for urban development, on urbanism. In 2019, Yanliu Lin firstly posed a concept framework of "*E-urbanism*" to divide the real world into three interwoven layers by attributes as ICT based economic activities organisations, socioeconomic relationships, and physical world, "*namely of ICT infrastructure and production networks, social networks and power relations, and urban form and land use*" (Figure 2.2), and she believes that the impact of e-commerce and migration on urbanism is fundamentally revolutionizing social and spatial reorganization of the city (Lin, 2019). For example, the internet is changing our world and the way people living in a city (Foth et al., 2020), because some socioeconomic and other activities which depend on the internet, they have been called Digital economy (Tapscott, 1996) (Zhendong et al., 2017), have dramatically influence on population distribution.

To solve the population overloading in urban areas and alleviate the development gap between urban and rural areas, Chinese government has implemented a series of measures, including the development of e-commerce in rural areas. For example, many villagers in China are engaged in the sales industry on Taobao, the biggest online sales platform in China. In China, a huge number of villages, rural settlements, and villagers who belong to the agricultural Hukou group, choose to develop e-commerce on Taobao, a Chinese online trade market similar to Amazon or eBay, and thus form Taobao villages. Taobao Village refers to a village located within a specific administrative village area, where numerous small and micro enterprises and individual business owners have converged.

These enterprises and individuals primarily use Taobao as their trading platform for business transactions. Within this village, the number of active online stores exceeds 10% of the local household population, thereby establishing a sizeable e-commerce ecosystem. As a unique example, the rapidly developing Taobao village have greatly changed the economic structure of rural China, forming a new pattern of urban-rural division of labour, and through global penetration, achieving urban-rural integration and rural revitalization (Juan Lin et al., 2021). It's worth noting that administrative villages are not solely located in rural areas; some are also distributed within urban settings (Chengzhongcun, Section 2.1.1). According to related research, the spatial and dynamics of Taobao village' development in Zhejiang Province from 2014 to 2020 reveal a trend, which is that many villages in the city (Chengzhongcun) close to urban centres are transforming into Taobao villages (Juan Lin et al., 2021).

Based on the market requirement and infrastructure conditions, especially ICT related infrastructure, online commerce tradesmen select site to form "*Taobao villages* (Aggarwal and Morrison, 1998)" by integrating and organizing capital flows, human power, goods flow, technical flow, etc, and that will influence on the surrounding areas of Taobao villages (Wan et al., 2017). Wan and other researchers (Wan et al., 2017) argue that E-commerce villages influence people's spatial decision behaviours which impact on traditional land use change pattern, and they provide a concept "*flow factors* (Wan et al., 2017)", a characteristic indicator to describe population flows, cash flows, goods flows, technical flows and etc. They believe the flow factors can be used in digitalized governance system to guide planning and management for rural areas, and fulfil "*the gap in terms of the interaction between E-commerce village and space of flow*s (Wan et al., 2017)". The mechanism of Taobao villages' development shows that ICT based E-commerce activities have reformed population flows in China (Zhendong et al., 2017; Wan et al., 2017; Lin, 2019). Statistical data of Taobao in China shows there are 7023

Taobao villages (*List of Taobao villages*, 2021). Taobao villages have created more jobs in rural areas, absorbed part of the urban surplus population, but also narrowed the urban-rural development gap. Meanwhile, the mechanism of Taobao villages' development shows that E-commerce activities have reformed population flows in China (Zhendong et al., 2017; Wan et al., 2017; Lin, 2019).

Although ITC increases the capital and talents back to the rural areas and contributes to rural urbanization, it is still cannot change the overall trend of population reduction in the entire rural areas in the process of urbanization in China. Relate research has found that the distribution of ICT infrastructure plays a decisive role in the development of Taobao villages (Yehua Dennis Wei et al., 2019). Relevant study also has used regression analysis methods to demonstrate that the distribution of Taobao villages has a positive correlation with road networks, land prices, industry, labour distribution and etc (Juan Lin et al., 2021). Therefore, most Taobao villages are located in developed regions of China, where ICT and other infrastructure construction are more complete. Nationwide, Taobao villages exhibit a pronounced concentration in coastal regions, displaying a sharp gradient from the eastern coastal areas to the western inland (Liu et al., 2020; Zang et al., 2023), which mirrors a clear tiered pattern that aligns with the spatial distribution of developed and developing regions in China. The interaction relationship between urban factors is significant and spatial distribution. Thus Lin (Lin, 2019) posed a network to describe socioeconomic relations between urban factors, and by comprehensive considering infrastructure, urban form, land use, and transportation network, she proposed a conceptual framework of E-urbanism (Figure 2.2) to understand the impact of E-commerce and migration on urbanism. Lin (Lin, 2019) used the framework to study Taobao villages in Guangzhou city and concluded that E-commerce had changed the city's urban structure, which is the arrangement of land use and the spatial organization of urban factors in a certain urban area, by providing remote coordination and organization of social and economic activities. Lin argued that E-commerce activities bring an unprecedent phenomenon that people, who live in different areas, can be involved in the same economic activity at the same time by ICT infrastructure.

In summary, the concept of E-commerce, ICT based socioeconomic organization, consist of land use, population and economic activities sub-models. The basic concept comes from Lin's E-urbanism framework (Lin, 2019) (Figure 2.2). E-commerce is continually changing social and economic human activities organization form and population distribution. Online information and capital flowing of E-commerce replace certain traditional human travel needs. The fact that E-commerce work is not limited by geographical location makes it possible to provide job opportunities in places where the cost of living is lower, for example rural areas. According to related research, the spatial and dynamics of Taobao village' development in Zhejiang Province from 2014 to 2020 reveal a trend, which is that many villages in the city (Chengzhongcun) close to urban centres are transforming into Taobao villages (Juan Lin et al., 2021). It helps capital and talent remain in or return to the rural areas and contributes to rural urbanization. In the 21st century, an urban phenomenon is emerging that refers to as the networked polycentric mega-city region (Su et al., 2017). Developed around one or more cities of global status, it is characterized by a cluster of cities and towns, physically separated but intensively networked in a complex spatial division of labour. They described and analysed eight such regions in North West Europe (Hall and Pain, 2006). In many cases, the relationships between urban factors are nonlinear and hierarchic and their interactions produce the urban form as consequence. In other words, city should be regarded as an integral organization.

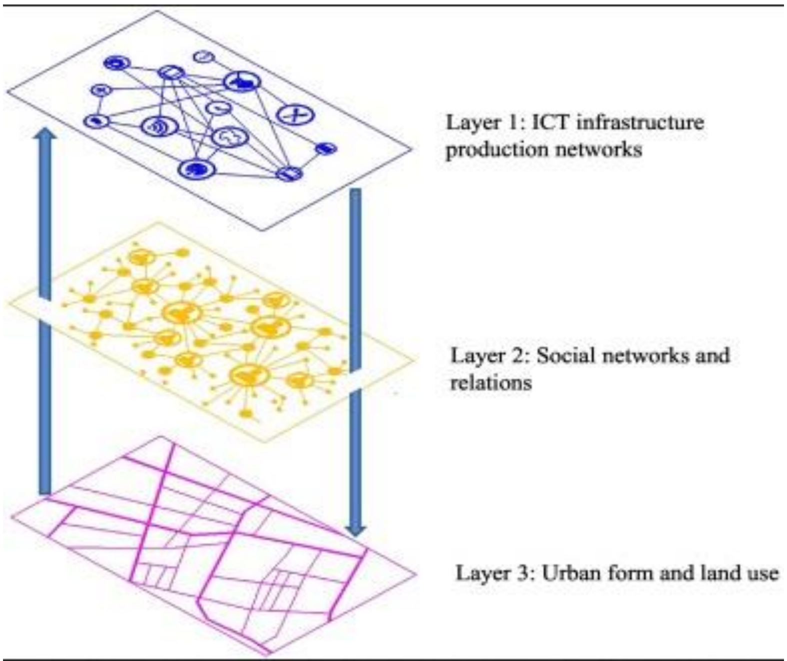

Figure 2.2 Three interwoven layers (Lin, 2019)

2.5 Urban modelling challenges in China

The main challenge of urban modelling in China is transferability. Because the conditions, governance and urban planning systems in China are different to any other country. It should be noted that, Chinese cities have different levels of affiliation and administration. Central government determines the overall plan for the whole country, and the provinces need to follow this plan. In a province, the provincial government has an overall plan for the whole province, and all local government need to follow the provincial plan. In the plan, some small cities develop around a big city, just as satellites orbit the earth. So those small cities are called "satellite city" (Mann, 1984). When designing a city development model in China, it needs to fully consider the local context, such as the imbalance among regional, urban and rural development, the inefficient distribution of population, the limitations posed by the household registration system, and so on. For example, the main attraction to floating population is from the eastern region, the most developed area in China (Hu et al., 2017), which concentrates the most of population and non-local Hukou residents. Meanwhile, Northeast China, the recession area, has experienced continuous population loss in recent years. Thus, perhaps an urban model designed for the developed sector of China to attract population is not suitable for a city within the underdeveloped sector. Therefore, transferability is an indispensable part of urban modelling in China.

Generally, the complex local context in China causes urban modelling to face two main challenges, the impact of Hukou policy and E-commerce on urban modelling.

On one hand, Hukou system is significantly impacts on population flows. Therefore, urban modelling in China is facing challenges of Hukou policy. On the other hand, the concept of E-commerce, ICT based socioeconomic organization, is consisted of land use, population and economic activities sub-models. For example, considering traditional spatial organization, formed by land use, transportation, population and economic activities sub-models, ICT based organization also influences the equilibrant status of the traditional general LUTI model.

The classical general LUTI model assumes that people are free to relocate and the interaction between economic participants via transportation networks. However, the situation has been changed. Therefore, Hukou and E-commerce need to be considered in LUTI model practice. To contribute the research gap, this research amendments of the impact of the interaction between Hukou policies and E-commerce on existing traditional classical LUTI models developed for Chinese cities had been integrate to promote LUTI models.

LUTI models in China need to incorporate the effects of Hukou and E-commerce on Transportation and Land-use. To contribute to the development of LUTI model in China, this research provides a conceptual approach for dealing with Hukou (as an example of China specific policy) and e-commerce platforms (as an example of the way ICT is influencing Land use and transportation).

Chapter 3

Literature review

To understand the mechanism of urban land use changes and represent possible future changes within urban systems, urban models are required to describe and analyse the spatial structure of urban areas and the interaction between various urban features by equations. Any urban area is a highly complex system. It involves social, economic, demographic, transportation, and other aspects (Batty, 2007a) and constantly interacts with external systems by population, material, economy and information exchange (Kennedy et al., 2011, 2007; Wolman, 1965). The urban model is a functional formula supported by economics, geography, sociology, and mathematics. It regards the city as a complex hierarchy system which contains multiple subsystems, and the subsystem (e.g., population distribution, transportation and economics) can be quantified by using abstract mathematical formulas. As a powerful urban research tool, it has the following features: multi-system collaborative analysis; multi-disciplinary theoretical support; and quantitative evaluation (Batty, 2007b);(Batty, 1976b);(Allen and Sanglier, 1978; Batty, 2008)). For simplicity of understanding and ease of manipulation, it can be divided into subsystems (Foot, 2017), for example, the transport subsystem and land use subsystem. Usually, urban models are not focused on a certain subsystem, instead, they emphasize the comprehensive influence of urban subsystems on urban spatial structure and individual behaviour, especially the transportation system and land use system (Cordera et al., 2017). Existing research on urban issues has provided many theories and methods to describe and predict urban development. Different disciplines approach this issue from different angles, such as the influence of science, technology, economy, culture, site, policy, transportation, and social structure on urban sprawl, growth, form, function, structure, and development. This has led to many leading theories and meaningful achievements.

3.1 The theoretical context of urban studies

3.1.1 Land Use Science

Land use science is a discipline, which focuses on the dynamic changes of land systems. The interaction between humans and the natural environment formed land systems as a consequence (Verburg et al., 2015). The survival and development of human society depend on the foundation and services such as food, energy, and resources provided by the land system (Christian, 1957) and human activities change the land system. It is the core element of sustainability science (Müller and Munroe, 2014) and interdisciplinary research form different disciplines provides great methodological diversity (Aspinall, 2006);(Verburg et al., 2015);(Verburg et al., 2019).

Land-use changes have a significant impact on regional environments and human well-being and play a vital role in global changes (Lambin and Geist, 2008);(Verburg et al., 2015). For example, the consequences, such as climate change, increasing demand for food and the ongoing biodiversity crisis, caused by accelerating urbanization and other human activity issues indicate that the connection between human society and the environment is inextricably linked (Rindfuss et al., 2008). Thus, an in-depth understanding of the dynamic changes in land systems can promote territorial spatial management and regional planning (戴尔阜 et al., 2019).

One of the most important tools for studying dynamic land use and land cover changes is model simulation (Verburg et al., 2004);(Brown et al., 2013);(Li et al., 2017). Models are simplified abstractions of real-world phenomena(Batty, 1976a). To understand the mechanism of how the drivers, status, trends and impacts of different land systems affect the functioning of the socio-ecological system, related models have been established to analyse land use changes, based on a mixture of natural environment factors (e.g., the locations of natural resources), human factors (e.g., the distribution of population and economic activities) and technology (e.g., transportation). (Johnston, 2013) argue that in most cities, economic factors directly relate to urban form. For example, cities dominated by transportation have developed into port-based cities and dominated by industry developed into industrial cities. (Clark and Kjelgren, 1990) believe water supplies determine the population and urban size. (Kiet, 2011) argues that culture affects urban form. Different religions, regimes or families have formed different forms of socioeconomic organization, which makes the different urban forms. UK geographer (Johnston, 2013) argues that social and productive development significantly impacts on urban development. For example, from the agricultural age to the industrial age, with the development of the industrialization process, productivity has been greatly developed, which drives rapid urbanization and reforms administrative systems to adapt the economic development.

Based on land use science, different disciplines have developed a series of alternative approaches for understanding the relationship between the natural environment, human factors and technology according to their emphases. For example, urban geography focuses on the organization form and mechanism of natural environment factors and human factors in urban areas (Johnston, 2013) and transportation geography focuses on studying the role of transport and mobility in the territorial combination of productivity and the territorial structure of transportation networks (Rimmer, 1988).

Nowadays, Land use science is facing three main challenges as follows: (i) globalization of land use; (ii) rapid urbanization; and (iii) rising competition for land (Müller and Munroe, 2014), especially in China. Under the influence of globalization, the urbanization rate (The ratio of urban population to total population) in China has grown by 44.56% over the past 40 years. Moreover, the urbanization rate of China in 2021 is 64.72%, which is lower than developed countries, for example, in 2021, the urbanization rate of the UK reached 90% and that of the US was 82%. The Chinese government remains committed to promoting urbanization.

3.1.2 Interdisciplinary theories and system dynamics

The field of urban studies is drawing on theoretical and conceptual propositions from a wide range of disciplines including sociology, economics, geography, informatics, geography, psychology, and complexity science. These theories provide different entry points for urban studies to approach urban issues from different perspectives. In the initial stage, the location was the root point of the urban study. A certain region was dominated by the highest population density gradient zone, which became the urban centre, and the urban centre was regarded as an entry point to study the relationship between urban centre location and the distribution of urban factors. A series of simple models were created to describe this pattern. For example, in the early 1920s, Ernest W. Burges proposed his famous concentric-zone model (Ernest, 1925; Meyer, 2000) and "*wedge-shaped sectors*" (Hoyt, 1939). However, large cities usually have more than one urban centre. Meanwhile, the urban system is an open system, which constantly interacts with external systems by population, material, energy and information exchange. Therefore, based on those simple models of urban morphology, multiple disciplines, such as transportation and economics, have been involved in urban modelling.

To understand the mechanism of urban development, elements of cities such as natural, socioeconomic resources and population are divided into many urban factors by character, for example educational, medical, financial, working and other factors. The geographical distribution of urban factors has played an important role in urban development (Batty, 2007a; Osborne, 2005). Concepts, such as urban structure and urban agglomeration, have been proposed to analyse and describe the distribution and organization of urban land use and urban factors. Those urban factors create driving forces of urbanization (Gu, 2019).

In summary, several theories can be applied either at an aggregate or disaggregate level of understanding decision-making behaviour (Acheampong and Silva, 2015) as Figure 3.1 shows.

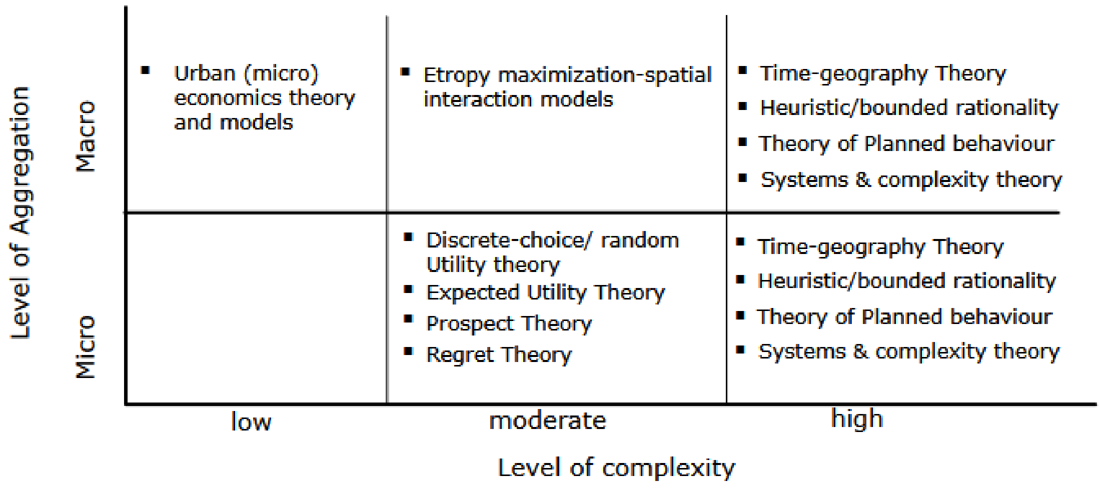

Figure 3.1 Level of aggregation and degree of complexity involved in operationalizing theories (Acheampong and Silva, 2015)

From a macro viewpoint, individual interaction behaviours between urban factors are influenced by the context of macro trends. Those individuals are regarded as several groups and the behaviour of each group will show homogeneity on a certain level. Based on that, top-down models, including the gravity model and entropy maximization, have been proposed.

From a micro viewpoint, those urban factors will interact with others subconsciously or unconsciously and spontaneously produce urban form as a consequence. This process is called "*self-organization*" (Allen and Sanglier, 1981; Ashby, 1991; Boonstra and Boelens, 2011). Based on that, bottom-up models, which focus on the micro-level, elemental or individualistic, to represent behaviours and interactions of those objects through space and time (Batty, 1976a), have been developed, such as CA and agent-based models.

As urban issues become more and more complex, it is difficult to accurately analyse and understand urban development using a single theory. Meanwhile, there are complex interactive relationships between urban factors, such that a change in one urban factor often triggers a chain reaction that causes multiple urban factors to change. To deal with the interactions and causal changes between multiple urban factors, system dynamics are used.

System dynamics is a discipline of crossover and integration which was initially developed by Forrester from MIT (Forrester, 1958). It is a cross-discipline, which uses information science as an entry point, and uses organisational theory, control theory, tactical decision-making, cybernetics, system theory and military games to analyse and explore the transmission, exchange, organization and control of information and the casual relationship between decision-making and consequence (Shepherd, 2014). System dynamic theory argues that every system has its structure and the organization of the structure decides the functions of the system. Thus, it prefers to focus on the interactions and relationships between the internal subsystems rather than external interference and stochastic events. The subsystems and elements inside are regarded as stakeholders of system dynamics. It uses a standard causal loop approach to describe the causal relationship and feedback between stakeholders. System dynamic approaches use qualitative Causal Loop Diagrams and quantitative stock-flow models to provide a powerful tool to explain and understand the basic structure and the nature of the problem within such systems (Shepherd, 2014). It provides an insight for modellers to investigate general dynamic tendencies (Shepherd, 2014). System dynamic approaches allow systems and policies to interact across space and time. Meanwhile, a feature of the Land-use/Transport Interaction Model is that transport and land-use systems operate on different time scales, because the response from transport users may relatively quickly to changes in transport policy or cost, but the response from land-use users may not (Shepherd, 2014). Therefore, system dynamics is ideal for investigating the interaction between land use and transport, because such systems "*contain feedbacks and delays which are often outside of the mental model of the decision maker or where feedbacks cross stakeholder boundaries* (Shepherd, 2014)."

System dynamics model can be used to predict the consequence of a certain urban planning pattern and they developed related methods to support decision making. There is a growing body of literature and models which explore and analyse the interaction and spatial organisation of economy, population, migration, infrastructure, land use and transportation (Shepherd, 2014). They developed urban-related methods, such as the dynamic Land Use and Transport Interaction (LUTI) model, and they argue that is part of a structured decision-making process (Pfaffenbichler et al., 2008).

3.2 Studies of Urban Models

Urban models are abstract conceptual frameworks that help urban researchers and decision-makers to better understand and analyse urban socioeconomic, environmental and other related issues. Urban models have two very important concepts, "factors" and "drivers", which will provide a better view to understand the process of urban development. Urban factors, such as population, economy, facilities and other urban resources also called development units, which are the engines of urban development (Makse et al., 1995). The organization and interaction of factors create drivers, which are certain socioeconomic activities which provide driving forces for land use changes (Liu et al., 2005). For example, in many cities, industrial economic activities have caused the conversion of agricultural land use into industrial land use in the last two centuries.

Usually, based on the mechanism of modelling, urban models are divided into data-driven and process-driven approaches (Li and Gong, 2016). The former is based on historical data to build a relationship between urban development and urban factors. Usually, the historical data is easy to observe and data-driven approaches can capture the main trajectory of urban development by statistical relationships (Li and Gong, 2016). Thus, the prediction is made with hard empirical experience. However, They "*focus on their correlations instead of their causal relationships* (Li and Gong, 2016)" which only provides weak evidence for urban development mechanisms. The latter "*is based on a clear mechanism and is generalized in a conceptual model* (Li and Gong, 2016)". There is a viewpoint that, theoretically, process-driven approaches can more realistically reflect socioeconomic activities (Li and Gong, 2016). However, the process of urban development is influenced by complex nonlinear and hierarchic interactions between various factors. Any change in any single one of the urban factors will affect other urban factors. Therefore, these complicated relationships are difficult to express based on structural models. Furthermore, "*available datasets in support of the construction of structural models are limited both in time and space, particularly for detailed socioeconomic datasets* (Li and Gong, 2016)." Because in many cases, the dataset tends to be generalized. Urban systems are very complex. Interactions between urban factors usually are complex and nonlinear. Therefore, in most cases, these two approaches coexist in urban models (Council and Committee, 2014).

In the initial stage, location was the root point of urban modelling. A certain region was dominated by the highest population density gradient zone, which became the urban centre, and the urban centre was regarded as an entry point to study the relationship between urban centre location and the distribution of urban factors. A series of simple models were created to describe this pattern. For example, in the early 1920s, Ernest W. Burges proposed his famous concentric-zone model (Ernest, 1925; Meyer, 2000) and "*wedge-shaped sectors*" (Hoyt, 1939). However, large cities usually have more than one urban centre. Therefore, based on those simple models of urban morphology, multiple disciplines, such as transportation and economics, have been involved in urban modelling.

Then, people were realizing that the distribution and organization of urban land use and urban factors create driving forces of urbanization (Gu, 2019), which is also called urban development driving force. From a micro viewpoint, those urban factors will interact with others subconsciously or unconsciously and spontaneously produce urban form as a consequence. This process is called "*self-organization*" (Allen and Sanglier, 1981; Ashby, 1991; Boonstra and Boelens, 2011). Based on that, bottom-up models, which focus on the micro-level, elemental or individualistic, to represent behaviours and interactions of those objects through space and time (Batty, 1976a), have been developed, such as CA and agent-based models. From a macro viewpoint, individual interaction behaviours between urban factors are influenced by the context of macro trends. Those individuals are regarded as several groups and the behaviour of each group will show homogeneity on a certain level. Based on that, top-down models, including the gravity model and entropy maximization, have been proposed.

It needs to be emphasized that, the urban development driving force is not always positively related to the size and number of urban factors. The research found that the concentration of urban factors, which will increase the city size, not only brings prosperity but also causes various problems. For example, high resource consumption with low utilization efficiency, and environmental pollution weaken the urban economic performance (Figure 3.2)(Guan et al., 2018). In the initial stage, increasing city size has a positive effect on economic performance, the negative effect will appear when it across the inflexion point (Borukhov, 1975). The negative effects are caused by many factors, such as traffic congestion, infrastructure overwhelming, environmental pollution, deterioration of public health and other urban social environmental and economic issues.

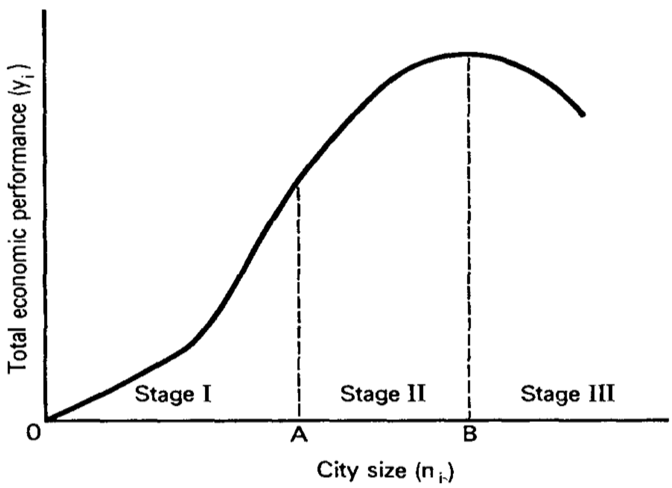

Figure 3.2 The relationship between city size and economic performance (Borukhov, 1975).

To better understand the mechanism of the negative effect and solve the problem, (Arnott, 1979; Capello and Camagni, 2000a; Fisch, 1975) studied the relationship between city size, agglomeration economies, urban evolution and decision-making. They argue that there is a relationship between city size and agglomeration economies. However, in 2000, Roberta Capello and Roberto Camagni argued that the question of optimal city size tended to be expressed misleadingly. The real issue is not 'optimal city size' but 'efficient size', which depends on the functional characteristics of the city and the spatial organisation within the urban system and has a positive effect on urban development (Capello and Camagni, 2000b). In other words, optimizing the spatial distribution and organisation of urban resources will solve urban negative issues. Therefore, the main task and challenge of urban modelling is to better understand the mechanism of the interaction, distribution and organization of urban factors.

3.3 Classification of Urban Models

Batty defined three main classes of urban models by their modelling styles and applications as follows (Batty, 2009):

- The first class is built around the aggregate static models of economic and spatial interaction. The theoretical roots of this class mainly are socioeconomics, regional economics, urban economics, and location theory, for example, land use and Transportation Interaction (LUTI) models. These models represent the spatial equivalents of classical macroeconomics and microeconomics theory. They are mainly used to understand and predict how urban land use and transportation policies affect the spatial distribution of urban socioeconomics and transportation activities (Niu et al., 2014). In short, this class of models is the most operational (Batty, 2009).
- The second class is urban dynamics models. In 1969, JW. Forrester's publication Urban Dynamics (1969) introduced an insight into "*forming a bridge between engineering and the social sciences* (Sanders and Sanders, 2004)". These models use

system dynamics methodology to analyse urban socioeconomic processes (Sanders and Sanders, 2004);(Batty, 2009). BY system dynamics methodology, socioeconomic processes are divided into elements and there are causal relationships between those elements. The change of elements casts their influence via certain causal relationships and triggers chain reactions. System dynamics methodology is usually used to deal with nonlinear growth issues.

- The last class is the focus of current urban research which involves models built around representing the action and behaviour of individual agents linked to spatial location, for example, cellular automata and agent-based models. Those models are bottom-up models, which focus on micro-level interactions, for example, the neighbourhood interactions between cell agents through spatial location analysis (Batty, 2009);(Gilbert, 2019);(Li et al., 2019). However, the actions and behaviour of individual agents are different to quantify and sometimes justify.

Foot provided another classification, which according to their theoretical basis, can be categorised into the three main types as follows: based on gravity, linear and optimizing mathematics models (Foot, 2017).

- The gravity-based model, for example, space economic and space equilibrium models, provide an opportunity for urban models to simulate individual behaviours in which attractiveness and distance decay usually are the most important parameters (Batty and Mackie, 1972). The attractiveness of a factor decays with distance. Therefore, the onset strength of attractiveness and rate of decay dominate the influence area of a factor.

- The linear-based model usually has been used to deal with the socioeconomic and land use activities in a certain urban area by deriving a set of linear equations (Foot, 2017). This model regards socioeconomic and land use activities as the independent variables. The information of a variable is collected for each zone in the region to calculate the dependent variable, such as population and employment, in the region.
- The optimizing models try to provide an optimal solution for urban development by performing a similar function. For example, simple optimal urban size theories, which are based on mainstream economics and focus on cost-benefit analyses, provide the same optimal size for cities when they have the same production function (Yang, 2020).

According to different modelling methods, urban models can be classified into two categories: top-down and bottom-up.

- The top-down models are based on spatial interaction models, including the gravity model, entropy maximization etc., which argue that individual behaviours are influenced by the context of macro trends and those individuals are regarded as several groups and the behaviour of each group will show homogeneity on a certain level (Alonso, 1964);(Anas and Hiramatsu, 2012);(万励 and 金鹰, 2014). By setting global behavioural assumptions, the model avoids calibration and calculation difficulties and reduces the requirement for individual behavioural data. It is helpful for related researchers to reveal the inherent mechanism of selection behaviour from a macro perspective.

- With the development of microeconomics and stochastic utility theory (Anas and Hiramatsu, 2011), bottom-up methods have become an important research direction, and they provide a better choice for describing the complex system (戴尔阜 et al., 2019). The model, which operates from the bottom up, mainly focuses on the micro-level, elemental or individualistic, to represent behaviours and interactions of those objects through space and time (Batty, 1976a). The model starts from the definition of the rules of subjective behaviours to build the framework to avoid the understanding and analyses of the entire system, and makes up for the difficulty of quantitative description of complex systems by traditional mathematical and physical models (Parker et al., 2003);(Le Page et al., 2017).

Figure 3.3 shows that urban models can be classified into the above catalogues. For example, the LUTI model can be classified into top-down and macro-perspective models, and CA and ABMs can be classified into bottom-up and micro-perspective models. Depending on different usage scenarios, the focus of the model is different.

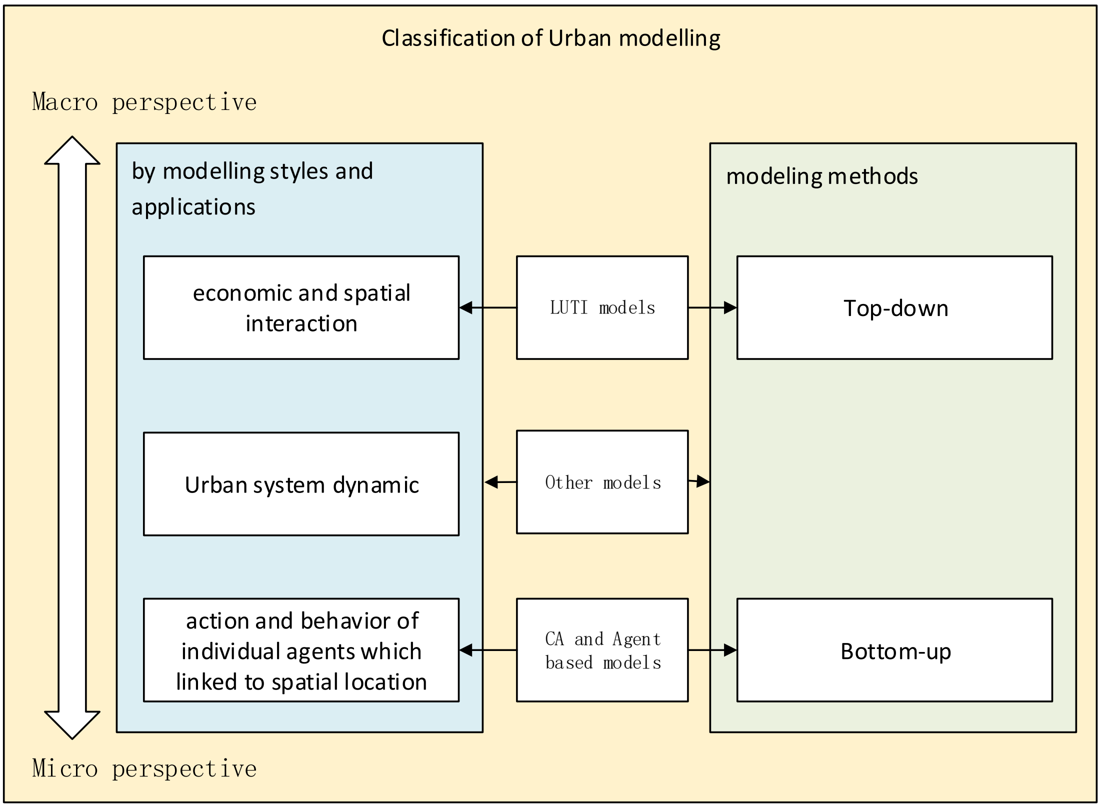

Figure 3.3 Classification of Urban Modelling

CA model is a kind of model that simulates spatial and temporal discrete complexity by simple local calculation. This local calculation is based on five elements as follows: Cell, Status, Neighbourhood, Transfer function and Temporal. The basic algorithm of the CA model is that in a moment, the status of a cell depends on its previous status and the status of its contiguous cells (Smith III, 1971). Such models usually are described as two-dimensional rectangle grids of cells, where each cell has a fixed location and attribute information, for example, land cover, land use or terrain (Green and Sadedin, 2005), to represent regional land use and "*embodying processes of change in the cellular status are determined in the local neighbourhood of any and every cell* (Batty, 1976a). "

ABMs are one of the most common approaches to "*analyse and simulate Land-use/cover change (LUCC) as the result of individual decisions* (Valbuena et al., 2010)". The model has been used to test urban policies in urban growth (Brown and Robinson, 2006), spatial economy (Farmer and Foley, 2009) and transportation (Batty, 1976a);(Davidsson et al., 2005). Under the guidance of its experience and knowledge, it will consider varying related factors and make a decision. As a consequence of the decision, the entity produces a certain behavioural pattern. Generally, the agent has four main features: autorhythmicity, spontaneity, sociality and reactiveness (Brown et al., 2005). Moreover, urban researchers believe that different modelling problems can lead to very different solutions. They suggest to use of multilevel models, which include macro, intermediate and micro levels, to build a conceptually or causally linked system of grouping objects or processes along an analytical scale to help the modeller identify the difficulties peculiar to each of the approaches (Gil-Quijano et al., 2010).

CA and ABMs both focus on spatial neighbourhood relations and decision-making on an individual level of the subject, emphasize the autonomy and dynamics of individual behaviour, and believe that individual behaviours depend on their social, economic, political and other conditions (Anas and Liu, 2007). In addition, those behaviours will interact with each other and the context to produce urban form as a consequence subconsciously or unconsciously, and related researchers named this process "*self-organization*" (Allen and Sanglier, 1981; Ashby, 1991; Boonstra and Boelens, 2011).

It should be noted that urban models can combine with others to better support urban planning and decision-making. CA and ABMs usually combine with other spatial interaction models to develop new models for specific conditions, for example, Land use-transportation interaction models.

The LUTI model is used to describe the relationship between urban land use and the accessibility of urban features. Because almost all human activities need to interact through transportation lines, in a certain context, transportation networks determine the spatial separations of human activities and land use, such as the distribution of population, commercial, industrial, etc. (Hansen, 1959a). This idea is used as the basis of LUTI models and laid their theoretical framework. The combination of space economy and the LUTI model creates a new type of urban model which is called spatial equilibrium model (Wieand, 1987), and the new model has become the most widely used type of urban model in the field of simulating the process of socioeconomic evolution, planning policy evaluation and decision-making support in developed countries ((Niu and Li, 2019);(Wegener, 2004);(Batty, 2007b);(Batty, 2012).

LUTI models have two key components which are land use and transport components (Acheampong and Silva, 2015). The land use component focuses on the distribution of all urban factors—residential, employment, and ancillary activities such as shopping, medical treatments, schools, entertainment etc. The spatial intercations between them will impact on people's location choice behaviours and create "*Residential–job location choice interdependencies*" (Acheampong and Silva, 2015). The transport component can be divided into two sub-components, which are transport demand and activity-base modelling approach (Acheampong and Silva, 2015). The former sub-component focuses on "*four distinct steps of trip generation, trip distribution, modal split, and route assignment* (Acheampong and Silva, 2015)." It approaches urban modelling from a macro viewpoint which is used to estimate and evaluate how well the transport system meets people's travel needs in a large-scale area. The later sub-component usually focuses on analysing and simulating individual transportation behaviours, such as location and route choice.

3.4 The impact of Hukou policy on human mobility

Unlike Western market-based population flow systems that are driven by economic factors, the Hukou system, a political factor, significantly impacts urban residential distribution (Chan and Zhang, 1999) and inhibits population flows (Xiaojun, 2017). After 1978, in the reform and opening-up, the Hukou policy was also been reformed to release population flows and many people were concentrated in urban areas. After decades of economic reform, especially on February 23rd, 2012, the Status Council of China determined the reform of China's Hukou system (Hu, 2012). According to statistical yearbooks of China, there is a direct relationship between Hukou policy and population movements. Although the restrictions of the Hukou system on population migration have been weakened, the effect of the Hukou system effect on urban China's labour markets (Zhang and Wu, 2017). Because of the existence of the Hukou system in China, the population immigration is not decided by socioeconomic activities. Many Chinese cities have their own population-related policy which is called "*population access*". Chinese people must meet certain conditions in order to relocate their household registration to a city if their household registrations are not in this city yet. The household registration system is very important because it is directly related to our social welfare, including but not limited to Medical, education, housing and other right. Guangdong Li, Siao Sun, Chuanglin Fang and other related Chinese researchers (Li et al., 2018; Xiaojun, 2017) argue that Chinese local governments can adjust policy to adjust 'population access' policy, particularly in large cities, in order to potentially control and plan urban expansion. Cities may face the problem of population overload and the Hukou system can help cities move closer to efficient size by controlling and limiting population size.

On one hand, because of different economic conditions, welfare which depends on Hukou policy is different in different cities. People tend to migrate from areas with lower welfare levels to areas with higher welfare levels. In the last 40 years, cities with higher welfare budgets, such as Beijing and Shanghai have the largest population growth. Table 3.1 shows the relationship between welfare spending statistical data of local government and population in Shanghai, Wuhan, Xi'an and Yinchuan, 2019. The inclination of policies has also driven the inland population, capital and foreign advanced technology resources to concentrate in these areas to form agglomeration effect in order to stimulate the rapid growth of local cities (Houkai, 2014). This kind of agglomeration effect will form a continuous and powerful attraction to the surrounding area, making resources and talents further gather themselves, thus further promoting the development of the city. For example, from 1978 to 2018 in Shanghai, its permanent population increased from 11.04 million to 24.24 million, and GDP increased from RMB 27.281 billion to RMB 3267.987 billion. The population and economic growth rate of Shanghai was much higher than the national average rate at the same period.

Table 3.1 Welfare spending statistical data of local government in Shanghai, Wuhan, Xi'an and Yinchuan, 2019

City	Social security	Medical	Education	Local Hukou Residents	Total Residents
Shanghai	42.01 billion Yuan	20.57 billion Yuan	28.34 billion Yuan	14.504 million	24.281 million
Wuhan	31.72 billion Yuan	14.37 billion Yuan	28.71 billion Yuan	9.064 million	11.212 million
Xi'an	4.666 billion Yuan	3.31 billion Yuan	5.836 billion Yuan	9.567 million	10.203 million
Yinchuan	3.33412 billion Yuan	2.23161 billion Yuan	4.083 billion Yuan	1.9957 million	2.2931 million

On the other hand, the Hukou system also significantly impacts on the transportation needs. It divides the population into urban and rural groups and related people into specific districts which define the city-countryside relationship and shape important elements of status-society relations in China (Cheng and Selden, 1994). Without an urban local Hukou, the rights and benefits of most migrants have been limited in certain contexts (Wei and Sheng, 2020). Because of the restriction of Hukou policy and income differences, residences in central urban areas are only available to local Hukou holders. Non-local Hukou holders live in more peripheral regions known as "*village in the city*" and local and non-local residents differ in commuting behaviour (Li et al., 2021). 500 million Chinese people (a reference to the China census) did not live in the district where they are Hukou registered. With such a large population, the urban transportation system must consider the transportation needs differences between local and non-local Hukou residents. Economic development, population growth, and urban expansion promote each other in the progress of urbanization. However, urban expansion is not always positive. Those areas were covered by the influence of these cities will inevitably face the loss of population and resources which will increase the imbalance of regional development. For example, China's southeast coastal economic belt has concentrated more than 40% of China's population and 60% of GDP. The per capita GDP of Zhejiang Province, where in China's southeast coast, is several times that of Ningxia Province, where in the northwest of China.

In general, the Hukou system dramatically affects population flows in China. Thus, it is necessary to establish a population flow mode to simulate population migration in China (Niu et al., 2019).

3.5 Land/Territorial/Regional Attractiveness

Land has the ability of an area to attract financial, technological and human resources, and the term "*Territorial/Regional attractiveness* (Servillo et al., 2012; Tanis, 2018)", also called land attractiveness, is defined to describe the ability. The land attractiveness depends on natural resources and human activities. On one hand, natural resources, such as topography, land cover, climate, hydrology, soil and other natural conditions are a prerequisite for urban development (Anttiroiko, 2012) and effect on land attractiveness. For example, because Aberdeen is the closest port city to the British North Sea oil fields in the UK, the local land attractiveness makes it easier for the city to attract oil-related and shipping industries. On the other hand, usually, human activities are changing over time. For example, during working hours, human activities are more likely to concentrate in the workplace. After working hours, human activities are more likely to concentrate in other places, such as home, restaurants, bars etc. In summary, because the contexts are different in different regions and human activities usually change over time, the attractiveness of land is also different and changing over time. Different regions have different land attractiveness, resulting in regional development and land values.

Moreover, land attractiveness also impacts on population distribution. For example, regional attractiveness significantly impacts on location choices of immigrants in Germany (Tanis, 2018). Residents have needs for work and daily life. Therefore, they need to interact with urban characteristics, and that forms human travel behaviours, one of the most important activities in cities. The ability of a region to effectively meet the needs of its residents significantly impacts on the local land attractiveness. There is a concept of "*accessibility*" which indicates the collective performance of land use and transportation systems and determines how well that complex system serves its residents (El-Geneidy and Levinson, 2006). Classical accessibility has two main elements as follows: attractiveness and impedance function (Mack et al., 2021). The attractiveness indicates the possibility and opportunities for residents to interact with facilities, which include workplaces, parks, schools, hospitals, supermarkets, buses, metro stations, etc., in a certain zone. The impedance function indicates that as distance or other costs increase, the possibility for residents to interact with facilities is decreased (Cui and Levinson, 2018; Geurs et al., 2012). Accessibility shaped urban land use (Hansen, 1959b) and transportation systems have deep influences on urban form, development and population migration (Cordera et al., 2017). Therefore, it is necessary to study the relationship between community traffic demand and accessibility. The gravity-based model, for example, space economic and space equilibrium models, provide an opportunity for urban models to simulate individual behaviours in which attractiveness and distance decay usually are the most important parameters (Batty and Mackie, 1972). The attractiveness of a factor decays with distance. Therefore, the onset strength of attractiveness and rate of decay dominate the influence area of a factor.

Land value is the value of a piece of property including both the value of the land itself as well as any improvements that have been made to it. It relates to local policy, facilities (e.g., schools, hospitals, bus stops, metro stations etc.), land use, economic activities, etc. The calculation of land values can use weighted calculation. For example, Zhou Liping's housing price model (Zhou, 2008) provides a method for setting the weight of urban factors, such as hospitals, supermarkets, green leisure spaces etc., to calculate and estimate rent and housing fees. The relationship between urban characteristics (workplace, park, school, hospital, supermarket, bus, metro stations, etc.) and land value indicated that there was a positive relationship between them in New York City (Ma et al., 2020). GIS plays a significant role in the calculation of land value. It can achieve timely and accurate tracking and understanding of land use status through the collection, processing, analysis, and output of spatial (e.g., the distribution of urban characteristics etc.) and non-spatial data (e.g., residential income levels, ages etc.), as well as the analysis of spatial databases for remote sensing images, maps, and statistical data sources. For example, to calculate the land value of a community, GIS can use buffer analysis to calculate and analyse the number and types of urban characteristics within a certain area, thereby providing a basis for the calculation of land values.

Nowadays, the impact of e-commerce on global development continues to deepen which enables regions to break location limitations and join in national and even global industrial divisions. More and more information and capital flow through the Internet, replacing traditional human travel needs and gradually changing people's travel behaviours. From a social point of view, e-commerce can alleviate traffic congestion and e-commerce has the ability to re-direct certain people from physical transport to online service (Shao et al., 2016), such as online retail, remote meetings and information exchanges. Thus, ICT infrastructure partially replaces the role of transportation networks and exerts an influence on land attractiveness.

3.6 Summary

To understand the mechanism of urban development, elements of cities such as natural and socioeconomic resources and population are divided into many urban factors by character, educational, medical, financial, working and other factors. The geographical distribution of urban factors has played an important role in urban development (Batty, 2007a; Osborne, 2005). Concepts, such as urban structure and urban agglomeration, have been proposed to analyse and describe the distribution and organization of urban land use and urban factors that create driving forces of urbanization (Gu, 2019).

Studies on the mechanism of urban development gradually realized the influence of science, technology, economy, culture, site, policy and social structure on urban sprawl (Osborne, 2005). Although urban models differ in various aspects and details, the purpose of urban modelling is the same. They try to provide practical tools for urban analyses seeking firstly to understand and describe the mechanism of urban evolution, including governing urban structures and behaviour of the urban system. Secondly, aiming to predict the consequences of future policy decisions and scenarios of urban planning (Batty, 1976a);(Batty, 2009);(Foot, 2017).

Urban studies have achieved a great deal of progress and developed numerous theories and models over the past decades. As urban scales expand, the complexity of cities grows geometrically, posing increasingly severe challenges for urban modelling. One of the main challenges is the impact of ICT-based economic organizational forms on urban models. Currently, research in this area is relatively lacking, and existing studies are not very well-equipped to address the current challenges. Therefore, the integrated planning of urban land use, transportation, and information space is worthy of further research and exploration.

Chapter 4
Methodology

In this thesis, the methodology chapter holds a pivotal position. It thoroughly describes how we systematically collected, organized, and analysed data, and explains the reasons behind our selection of specific research methods. This study employs a mixed-methods research design, integrating both quantitative and qualitative approaches, aiming to comprehensively explore the research question and arrive at more reliable and insightful conclusions.

However, it must be noted that any research method has its inherent limitations. This study has endeavoured to mitigate the potential shortcomings of a single method through the mixed-methods design and maintained a cautious and objective attitude throughout the data collection and analysis process to ensure the accuracy and reliability of the research results.

Overall, the methodology chapter of this thesis comprehensively elaborates on the research design, data collection and analysis methods, as well as the limitations of these methods. Through this chapter, readers can clearly understand the methods adopted in this study and the logic and considerations behind them, thereby facilitating a better understanding and evaluation of the research outcomes.

4.1 Methods used in examining Hukou

The hypothesis is that the degree of traffic difficulty experienced by different communities in the same city is related to the distribution of urban characteristics, average incomes, and average transport costs of those communities.

There is a hypothesis that some urban facilities, such as schools, medical facilities, working places, supermarkets, malls, shops, entertainment facilities, parks and urban green belt. are more favourably associated with ease of travel. Based on the types of services provided by these urban facilities, this research divides urban facilities into seven aspects based on their functions: workplace, medical facilities, road systems, business district, public transport system, compulsory education resources, and green leisure spaces (Table 4.1). According to those urban facilities' attributes, they are divided and stored in different subset. For example, all hospital data is stored in one data subset, while supermarket data is stored in another data subset of the database.

Table 4.1 Classification of urban factors

Main aspects	**Urban facilities (Detailed classification)**
Workplace distance	Factory Company Business office district
Medical resource conditions	Pharmacy Clinic Health centre Hospital
Road system	High way Main road Branch road
Business district	Shopping facilities (e.g. supermarket, mall) Entertainment facilities (e.g. cinema, karaoke) Living facilities (e.g. bank, post office)
Public transport system	Bus system Metro system
Green leisure space	Park Urban green belt
Children's Education conditions	Kindergarten

	Primary school Middle school

Related to the first two hypothesis there is a hypothesis that there is a disparity in the ease of travel (Currie, 2010) between areas inhabited by Local and non-local Hukou residents.

Key definitions of indicators calculated

- The transport impedance indicator reflects the level of difficulty for a group of residents to meet their traffic demand.
- The transport impedance disparity indicator represents the difference in the transport impedance indicator among different residential groups.

The purpose of the transport impedance indicator is to measure the disparity between traffic needs and accessibility for different communities. Then compare the impedance indicator in Local and non-local hukou areas in order to understand whether there is a disparity and if so its extent.

Regarding the relationship between average incomes and the transport impedance indicator of a community, there is an inversely proportional relationship, while there is a positive proportional relationship between the average transport costs and transport impedance indicator.

4.1.1 Basic idea

The hypothesis is that the variation in difficulty in accessing activities and destinations (transport impedance (Li et al., 2023)) experienced by different communities in the same city is related to the distribution of urban characteristics, average incomes, and average transport costs experienced in those communities. Based on above discussion (Chapter 2), non-local Hukou residents are disadvantaged in terms of income compared to local

Hukou residents. Meanwhile, based on historical studies (Li, 2018; Wang and Li, 2014; Xue et al., 2017), non-local Hukou residents in Xi'an are generally concentrated in the villages in the city (Chengzhongcun). Therefore, this study assumes that all non-local Hukou residents live in the villages in the city (Chengzhongcun), and all local Hukou residents live in commercial residential communities.

Figure 4.1 shows the analysis framework of transport impedance indicator. In the first row, L community (the diamond on the left) represents local Hukou residents' community, and N community represents non-local Hukou residents' community. The data collection box represents the case study collected relevant data (including electronic maps, road network data, land use data, POI (Points of Interest) data, population data, and economic data) using various methods and from multiple sources, and the Table 4.2 shows data types and sources. In the second row, Data process box represents different types of data were processed, such as converting textual data and graphical data into point, polyline, polygon or raster data, and based on the attributes of those data, the data was stored in different datasets within the database (in the third row, Data set). In the third row, the Tavel needs and behaviours analyses boxes represent that based on the Hukou status, this research analyses the requirements of local and non-local Hukou residents (e.g., people need to work and live, such as commute, communication, shopping, leisure and entertainment) and spatial location choice behaviours. Then, in the fourth row, the Urban POI data selection boxes represent that urban facilities will be selected as travel destinations by local and non-local Hukou residents based on their needs, and use urban POI to collect information on these urban facilities. The urban POI data provide by open database of commercial mapping and navigation companies have spatial information on almost all urban facilities. These data can represent various types of information, such as buildings, transportation hubs, attractions, etc. POI data usually contain attributes such as name, address, location, category and other information, and can be abstracted into points in GIS for management, spatial analysis and calculation. According to the certain residents' requirements, related subset has been selected from the spatial database This study collects urban POI data, land use data, economic data, and other relevant data to establish a comprehensive dataset. According to the certain residents' requirements, related subset has been selected from the spatial database. For example, because only local Hukou residents' children can gain local compulsory education service, the POI data of middle and primary schools are only selected in the calculation of impedance indicator for local Hukou residents' communities. In the fourth row, the Weight set boxes represent that based on the distinctive travel needs of different communities, the corresponding data is selected and weighted accordingly. The weight of an urban facility depends on the certain group of residents' needs. When choosing a place to live, people consider a variety of factors, including transportation, education, work, environment and etc. How important of a factor in people's location decision-making process is regarded as the weight of that factor. Those factors are provided by urban facilities. Meanwhile, the distribution of urban facilities significantly impact on housing prices (Wang, 2011). For example, the distribution of supermarkets, schools, and other urban facilities influences people's spatial decision behaviours, related to housing prices and rent fees (Glaeser et al., 2005; Liang et al., 2018). In the study of urban housing prices, the impact of these urban facilities on housing prices is quantified. For example, Zhou (Zhou, 2008) quantified and scored the impact of urban facilities on housing prices in Xi'an in 2008. In this study, the weight of these urban facilities that affect housing prices is used as an important reference. For example, Zhou's paper (Zhou, 2008) was studied on the mechanism of the house price in Xi'an and argued that the bus stops, schools (including kindergarten, primary, junior and high schools) and living facilities (e.g. supermarket, bank, post office, and etc.) within a community or within 1 km of the community have a significant impact on the housing prices of this community. ("Commuter Monitoring report of Major cities in China 2022,"

2022) indicates that almost 90% of long-tern rental apartments are located within 2 km of a rail station in Beijing. Meanwhile, the report indicates that the average commuting distance of the city is greater than the average distance of residents to other destinations. For example, the average commuting distance in Beijing, Shanghai, Guangzhou, Hangzhou, Xi'an is over 8 km, with Beijing exceeding 11 km. Other destinations for residents to travel are usually within a radius of 5 km. Based on the above studies, this thesis proposes a method for assigning weight to different urban facilities. The detailed weight calculation method and formula will be explained in the section 4.1.3. In the fifth row, the spatial analyses and calculation boxes represent that spatial analysis and calculation tools were used to calculate transport impedance indicator and transport impedance disparity indicator, and the boxes on the left and right sides of the sixth row represent the results for local and non-local Hukou residents. Then, the results were used to calculate the transport impedance disparity indicator for local and non-local Hukou residents (the box in the middle of the sixth row). The detailed calculation method and formula will be explained in the next section.

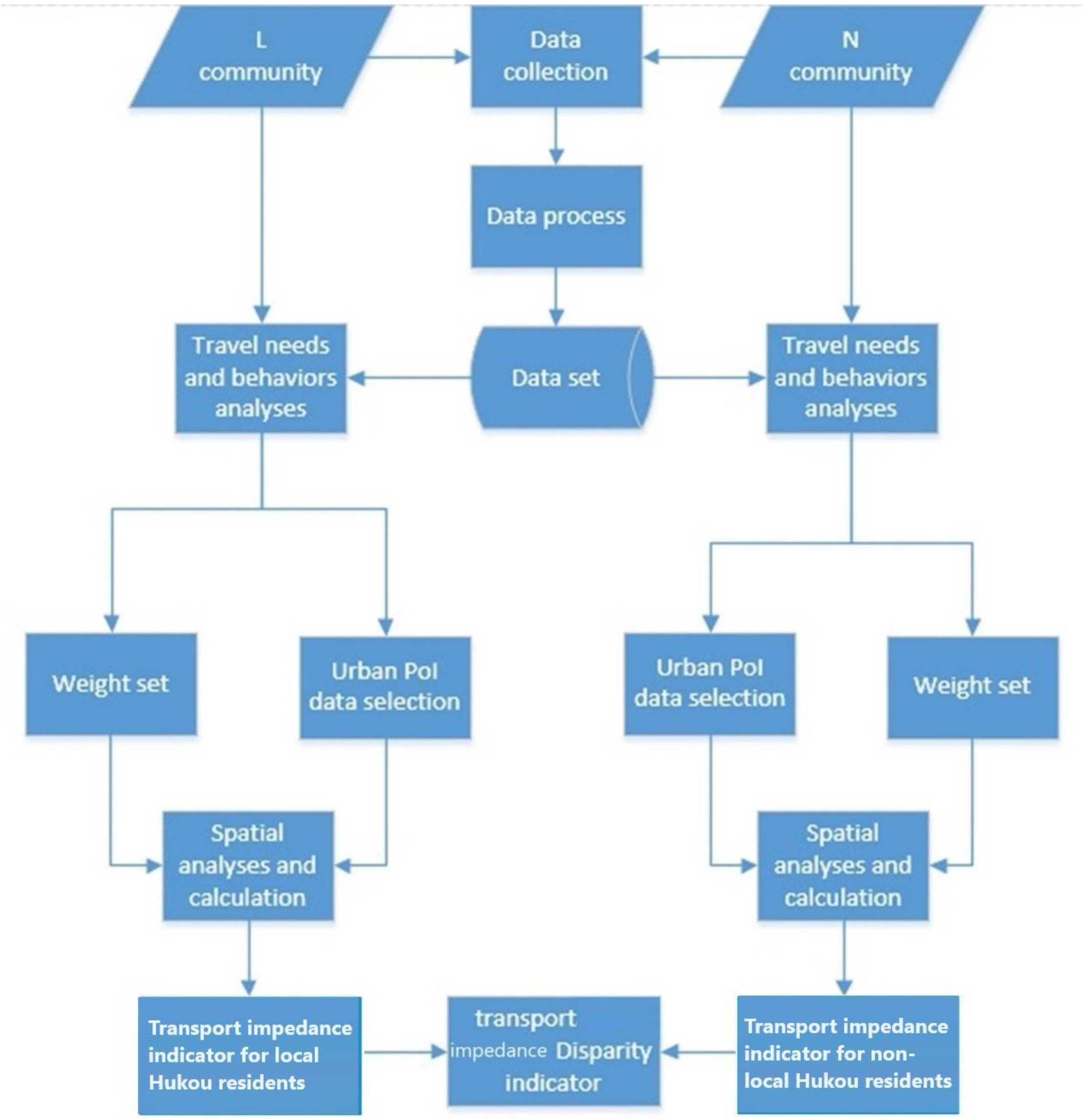

Figure 4.1 Calculation framework of transportation difficulty indicator and transportation disparity difficulty indicator

4.1.2 Hukou indicator

This research proposes the use of the transport impedance indicator to measure the level of difficulty for a community in accessing urban characteristics by transportation system, as well as a transport impedance disparity indicator to describe the disparities in transport impedance among different communities.

The transport impedance indicator is related to the distribution of urban facilities such as the distance and number of working places, parks, schools, hospitals, supermarkets, bus and metro stations, and so on, while also taking into account the average incomes and transport costs of a given community.

The definition of the transportation impedance indicator is as follows:

$$\boldsymbol{D_i = \frac{C_i}{I_i S_i}} \qquad [1]$$

D_i……………………Transport impedance indicator of region i (an indicator).

C_i……………………Average transportation cost (yuan per person) of region i.

I_i……………………Average income (yuan per person) of the region i.

S_i……………………Local residents' satisfaction of the region i (an indicator).

Urban characteristics are subdivided into three categories: transport resources, living resources and working resources.

If people lack access to employment, education, healthcare and other basic amenities, they will suffer the social exclusion (Lucas, 2012). Meanwhile, there is significant correlation between social exclusion and transport poverty (Lucas, 2012). Therefore, the living and working resources is very important to residents. This research uses parameter S_i to describe how well the living and working resources meet local residents' need in region i.

4.1.3 Spatial analytic model and Weighting the accessibility to each type of destination

To calculate parameter S_i (Section 4.1.2, equation [1]) a number of techniques are employed. Broadly these include weighting of layers, spatial overlay and density analysis.

The procedures employed are explained in detail in the following subsections.

4.1.3.1 Accessing POI Data

Point of Interest (POI) data are urban facilities closely related to people's lives (vector point spatial data layered by theme such as Health facility, retail outlet, school and etc.) as were explained in section 4.1.1 above.

POI data may be used as a proxy for understanding the level of satisfaction with an area.

The advantages of POI data are mainly reflected in the following aspects:

a) Data richness: POI data covers a large amount of geographical entity information, such as facilities and services in various categories including catering, shopping, entertainment, transportation, etc. These rich data provide a wide range of basic information for various applications, helping to more comprehensively understand and analyse geographical spatial phenomena.

b) High positional accuracy: POI data usually has high positional accuracy, which can accurately reflect the spatial position of geographical entities. This is crucial for applications that require precise positioning (such as navigation, location search, etc.).

c) Rich attribute information: In addition to location information, POI data also contains rich attribute information such as name, category, business hours, ratings, etc. This attribute information helps to better understand the characteristics and laws of geographical entities, providing more valuable information for decision-making.

A high number of high-quality services within or close to an area indicate that the population is likely to be satisfied with the services. The thesis on the formation mechanism of housing prices in Xi'an (Zhou, 2008) divided the relevant POI data into several main aspect, including education (including kindergarten, primary school, junior school and high school), medical conditions (including clinic, hospital and etc.), shopping, leisure space and so on. The study on the spatial location choice behaviours of online selling workers in Guangzhou (Lin, 2019) divided relevant facilities into road, public transport, sanitation, children education, public spaces and other online selling support facilities. The study on the population distribution in Beijing (Niu and Li, 2018) indicated that commuting distances and workplaces significantly impact on people's location choice behaviours. The study on housing prices in Ningbo, China (Liang et al., 2018) divided relevant facilities into Transportation infrastructure, Educational facilities, Medical facilities, Commercial facilities and landscape. This study classified facilities into different main aspects based on their attributes. For example, the main aspect of education facilities includes kindergartens, secondary schools and universities. Considering the classification methods of above studies, this thesis employed the five main aspects from Xi'an (Zhou, 2008), Guangzhou (Lin, 2019) and Ningbo (Liang et al., 2018) and the two main aspects form Beijing (Niu and Li, 2018). Based on this, some adjustments were made. Those adjustments were based on the needs of local and non-local Hukou residents. For example, because of the phenomena of left behind children (Section 2.1) amongst non-local Hukou groups, the distribution of compulsory education resources will not impact on the spatial location choices of non-local Hukou residents. Meanwhile, because the allocation of educational resources in high school and earlier stages is related to Hukou policy (Section 2.1), while there are no restrictions on Hukou policy for education at universities and beyond, in studying the residents' demand for children's education, this thesis only used data from kindergartens, primary schools, junior high schools, and high schools. Moreover, Due to the differences in travel behaviours between non-local and local Hukou residents, non-local Hukou residents rely more on public transportation resources, while local Hukou residents who own private cars will consider the traffic capacity of the road system more. Based on above discussions, the final result show in Table 4-1 (Section 4.1.1).

POI data was acquired for Xi'an https://lbs.amap.com/?ref=http%3A%2F%2Flbs.amap.com%2Fdev%2Fkey%2Fapp, https://wiki.open.qq.com/wiki/lbs/get_poi see Table 4.3. Then, the POI data were divided into different datasets (Figure 4.2, line 1 to 3). For example, one of the seven main aspects was Health data. the following specific types of point were extracted. This was carried out using a query in ARCGIS to subset the health data into four layers pharmacy.shp, clinic.shp, healthcentre.shp and hospital.shp (Figure 4.2, line 3). The following table shows the correlations between the data and figures.

Table 4.2 Correlations between Data types, source and figures

<table>
<tr><th>Related data</th><th>Description</th><th>Data types</th><th>Data source</th><th>Illustration</th></tr>
<tr><td>Workplace distance</td><td>Workplace (e.g., factory, company, Finical facilities and commercial facilities.)</td><td>POI</td><td rowspan="6">ArcMaps Company (It has basically the same functions as Google Maps and is a local map navigation service company in China.) https://lbs.amap.com/?ref=http%3A%2F%2Flbs.amap.com%2Fdev%2Fkey%2Fapp, Tencent Map Navigation Service (a company that competes with Amap in map navigation services)</td><td>Figure 4.5 (a)</td></tr>
<tr><td>Medical resource conditions</td><td>Pharmacy
Clinic
Health centre
Hospital</td><td>POI</td><td>Figure 4.5 (d)</td></tr>
<tr><td>Road system</td><td>High way
Main road
Branch road</td><td>Polyline</td><td>Figure 4.3 (b)</td></tr>
<tr><td>Business district</td><td>Shopping facilities
Entertainment facilities
Finical facilities</td><td>POI</td><td>Figure 4.5 (b)</td></tr>
<tr><td>Public transport system</td><td>Bus system
Metro system</td><td>POI and polyline</td><td>Figure 4.3 (a)</td></tr>
<tr><td>Green leisure space</td><td>Park
Urban green belt</td><td>POI</td><td>Figure 4.5 (a)</td></tr>
</table>

Children's Education conditions	Kindergarten Primary school Middle school	POI	https://wiki.open.qq.com/wiki/lbs/get_poi	Figure 4.5 (c)
Land use data	Residential area Commercial area Public service Industrial land leisure land, including urban greening land Natural geographical data	Polygon	Xi'an Surveying and Mapping Institute (a subordinate agency of the Xi'an Municipal Government responsible for urban surveying, mapping, and cartography)	Figure 4.4

Figure 4.5d Table layout showing the columns in the health POI data.

4.1.3.2 Spatial analysis model

The next step as shown in Figure 4.2 was spatial analysis which was performed as follows. The first step is to convert point data into Kernel density raster data. By using Kernel density analysis tool of ArcMap, those points data were used for density analysis and the result are converted into raster data. The Kernel density analysis tool's search radius of a type of urban facility depends on the maximum distance residents can tolerate when select this type of facility as a traveling destination. The research radius for different urban facilities is different. Some urban facilities have a constant search radius, for example, based on the discussion of section 4.1.1, the research radius of bus stop is 1 km and metro station is 2 km. While others have different values in different cities. That depends on the average commuting distance of the city where this type of urban facility is located, as well as the weight of this type of urban facility in its main aspect. The calculation formula for weight distribution (Equation [2]) is as follows:

$$Weight\ distribution\ = \frac{N_1 * L_1}{N * L_1 + N_2 * L_2 + \cdots + N_x * L_x} \quad [2]$$

N........................The number of resident this type of the facility can serve in a certain time-span (person-time)

L........................The level of service quality (index).

x........................The total number of the facility types in a main aspect

For example, the layer pharmacy.shp (Figure 4.2, line 3) was used with the Kernel density analysis tool of ArcMap. The numerical values involved in the following calculations are all hypothetical data, which are only used for illustration and do not represent the actual numerical values in the calculation. Assuming that pharmacies receive an average of 100 people per hour, clinics receive 200 people per hour, community health centres receive 300 people per hour, and hospitals receive 500 people per hour, and their service quality indexes are 1, 3, 5, and 10, respectively. Taking the index of service quality and Substituting Equation [2], it can calculate that the weight of pharmacy in the main aspect of medical condition resources is (100 * 1) / (100 * 1 + 200 * 3 + 300 * 5 + 500 * 10) and the result is approximately equal to 0.0139. The search radius the average commuting distance of the city multiplied by 0.0139. For example, if the average commuting distance of the city is 6 km, then the research radius of pharmacy is 6 * 0. 0139 and the result is 0.0834 km. This calculated the distribution range and intensity of the accessibility to pharmacies. The resulting layer was named Pharmacy raster layer (Figure 4.2, line 5). For the convenience of calculation, these raster data had been resampled, with the original values scaled proportionally to the range of 0 to 100. The reclassify tool in the Reclass box of Spatial Analyst tools was utilized to reclassify the value of every raster pixel into classes ranging from 0 to 100, which keep relative distances between values, using the equal interval method (Figure 4.2, line 6). The resulting layer was named Pharmacy-R raster layer (Figure 4.2, line 7). According to the calculation results of the Equation [2], different weights are assigned to different layers (Figure 4.2, line 8). Then, the weighted spatial overlay tools were used to calculate the results of those main aspects (Figure 4.2, line 9). For example, to calculate the result of medical resource conditions, weighted spatial overlay tools were used to accumulate the raster values of the four layers according to the spatial overlap relationship. For the convenience of calculation, the result raster data had been resampled, with the original values scaled proportionally to the range of 0 to 100 (Figure 4.2, line 10), and named Medical condtions-R raster layer (Figure 4.2, line 11).

The weight of a main aspect depends on a certain group of residents' needs (Li et al., 2023). To analyse the impact of Hukou policy, this research has established two community groups: local and non-local Hukou residents' communities. Chapter 2 discussed the different of socioeconomic status, travel needs and behaviours between local and non-local Hukou residents. Therefore, all those seven main aspects need to be assigned different weight for local and non-local Hukou residents respectively (Figure 4.2, line 12). The next step is to use the weighted spatial overlay tools to calculate the Land attraction of those seven main aspects (Figure 4.2, line 13 and 14).

Land attraction is one of the core concepts in this chapter, and its calculation formula is as follows:

$$LM_G = T_1 * Weight_{T1} + T_2 * Weight_{T2} + \cdots + T_n * Weight_{Tn} \quad [3]$$

LA_G.................The land attraction of a particular main aspect for a certain resident's gropu.

T.................The density index of a certain type of urban facility.

$Weight_T$.......... The percentage of the demand score for that type of urban facility by a specific resident group. The sum of all the weight coefficients equals 100%.

The parameter T represents the calculation result of the 11th row in Figure 4.2.

The percentage of parameter $Weight_T$ is used to measure the degree of influence of the type T of urban facility on the location choice behaviours of a certain group of residents. The greater the impact, the higher the weight, and vice versa.

The calculation of weights employs quantitative analysis, a method that involves collecting and analysing specific data to assess the degree of influence of different factors on a particular matter. Firstly, data is collected from published papers, such as those studying the impact of urban facilities on residents' relocation and settlement choices. Then, statistical analysis and other methods are used to process and analyse the data in depth. Finally, weights are assigned to different factors based on the analysis results. Finally, weights are assigned to different factors based on the analysis results.

This study either directly extracted weight parameters from other research or calculated them based on the processing of findings from that research. For example, in studying the extent to which various urban facilities influence the choice of spatial locations for both local and non-local residents, researchers assigned different weights to different urban facilities based on the needs of different Hukou groups (Li et al., 2023). The weight parameters from this thesis can be directly used. In studying the extent to which various factors influenced Taobao villagers when choosing their spatial location, Lin assigned scores to different factors (Lin, 2019). The scores ranged from 0 to 100, with higher scores indicating a greater influence and lower scores indicating a lesser influence. The weight of a certain factor is calculated by dividing its score by the total score of all factors. For example, if we calculate the land attraction for Taobao Village residents by taking sanitation, children's education, road and public transportation systems, as well as government support as all relevant factors, and these factors score 75, 50, 95, 65, and 50 respectively in Lin's paper (Lin, 2019). Then, the weight of sanitation would be 75 / (75+50+95+65+50), which is equal to 22.39%.

Subsequently, based on the spatial distribution of the residential areas of the specific Hukou group, the Clip tool in the ArcGIS toolbox was used to extract the land attraction value of the residential areas (Figure 4.2, line 15). This value represents the satisfaction parameter S of the resident group (Figure 4.2, line 16).

The formula [4] is used to calculate the transportation impedance disparity indicator. The transport impedance disparity indicator is obtained by dividing the transport impedance indicator of non-local Hukou residents by that of local Hukou residents within the same region. The transportation impedance disparity indicator is as follows:

$$T_i = \frac{D|N_i}{D|L_i} \quad [4]$$

$D|N_i$……………… Transport impedance for non-local hukou community within region i.

$D|L_i$………………… Transport impedance for local hukou community within region i.

T_i…………………… Transport impedance disparity indicator of region i.

The larger the transport impedance disparity indicator value is, the more difficult it is for non-local Hukou residents to access relevant urban facilities through the transportation network compared to local Hukou residents.

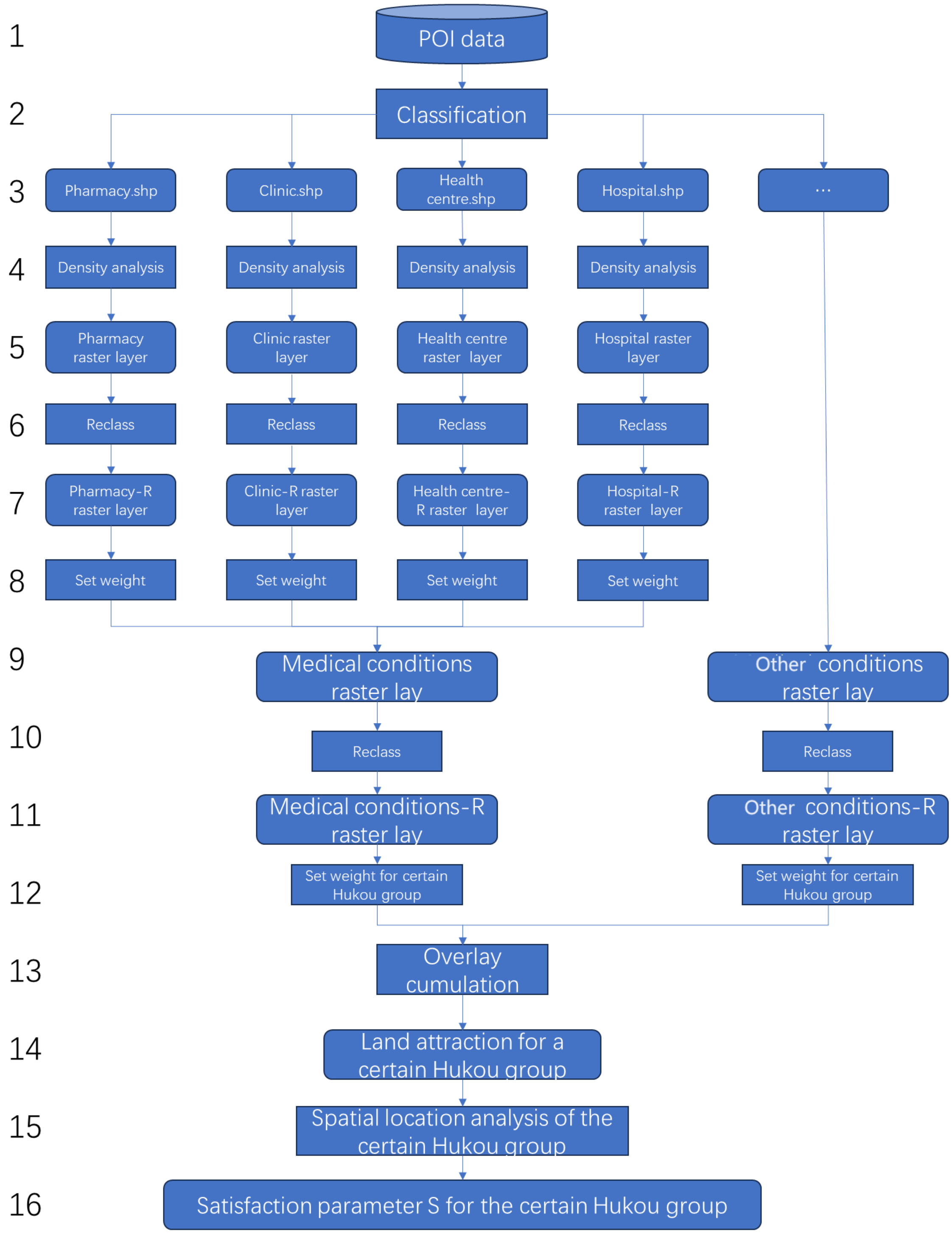

Figure 4.2 Spatial analytic model overview showing the processing sequence of processing for health facilities, one of 7 POI data types included

4.1.4 Data sources

Xi'an, the capital city of Shaanxi province in China, is the research subject to calculate the Transport impedance Indicator for both local and non-local Hukou residents' communities. The Transport impedance Disparity Indicator is also calculated to reveal any differences in treatment between these two groups within the same city.

To reach this aim, transportation system, land-use data, POI data, the distribution of different types of residential area such as areas reserved for local Hukou holders and "village in the city" neighbourhoods - which are dense urban neighbourhoods on the site of former villages that have been subsumed into rapidly growing cities, and other urban data (Table 4.3) of Xi'an were collected by fieldwork, online Python (Python is a programming tool for bulk data acquisition from websites. As shown in Table 4.3, using Python tools to collect POI data from map service interface, and collecting house prices and rental data from website) tools and other measures in the case study.

I conducted field surveys on Shajing Village and Zhongtian Community. Shajing Village is a typical and one of the most populous Chengzhongcun located in the core built-up area of Xi'an. With an area of approximately 0.21 square kilometres, approximately 100,000 tenants lived there in 2019. Zhongtian Community is a typical commercial residential community that abuts Shajing Village. as shown in Figure 4.6, there are locations and photos of Shajing Village and Zhongtian Community. Then, this study randomly selected several malls and supermarkets, randomly chose a location near the centre of their business areas, and then used a rangefinder to measure a 10-meter by 10-meter square area. This study recorded the total number of people in this area. In this experiment, this study chose to start recording the total number of people in this area at 8 pm for each sampling, and then recorded the total number of people every half hour for a total of three times. The average value was taken as the final traffic volume (Unit person-time) result. For example, this study investigated the Sunshine World Shopping Mall and Vanguard (Wanke Store) supermarket in Chang'an District, Xi'an City. The average values of them are 130 person-time and 36 person-time respectively.

Figure 4.3a shows the subway stations (red dot icons), subway lines (red lines), bus stops (cyan dot icons), and bus routes (cyan lines) in Xi'an. Figure 4.3b illustrates the main road network in Xi'an, including expressways (red bold lines), main roads (black lines), and branch roads (purple thin lines). Figure 4.4 is a land use planning map of the main urban area of Xi'an. The yellow areas represent commercial and residential land, the purple areas are rural collective land (Chengzhongcun), the red-brown areas are industrial land, the red areas are commercial land, the dark green areas are parks, and the light green areas are urban greenbelts. Figures 4.5a to 4.5d show the POI data of Xi'an City. Among them, Figure 4.5a illustrates the distribution of companies (including factories), parks, and scenic and historical sites. Figure 4.5b displays the distribution of financial facilities, catering facilities, entertainment facilities, hotels, and large supermarkets. According to (*Statistical Yearbook of Xi'an*, 2020), in 2019, Xi'an had 1.6454 million employees in the secondary industry and 3.7995 million employees in the tertiary industry. The POI data of Figure 4.5a (companies (including factories)) and Figure 4.5b (financial facilities, catering facilities, entertainment facilities, hotels) is used to analyse the distribution of working place. Figure 4.5c shows the distribution of kindergartens, primary schools, and secondary schools. The POI data of Figure 4.5c is used to analyse the distribution of Children's education conditions. Figure 4.5d presents the distribution of pharmacies, clinics, community health centres, and hospitals. The POI data of Figure 4.5d is used to analyse the distribution of medical resource conditions. Using Points of Interest (POI) data, this research analyses and calculates the density of each urban POI data, including schools, malls, supermarkets, universities, metro stations, bus lanes, hospitals and etc.

Meanwhile, field work investigation has provided qualitative details, such as the disparities in living conditions indifferent communities.

Figure 4.3 Bus and metro route network (a) and road system (b) data of Xi'an, 2019 (Data source: https://lbs.amap.com/?ref=http%3A%2F%2Flbs.amap.com%2Fdev%2Fkey%2Fapp https://wiki.open.qq.com/wiki/lbs/get_poi)

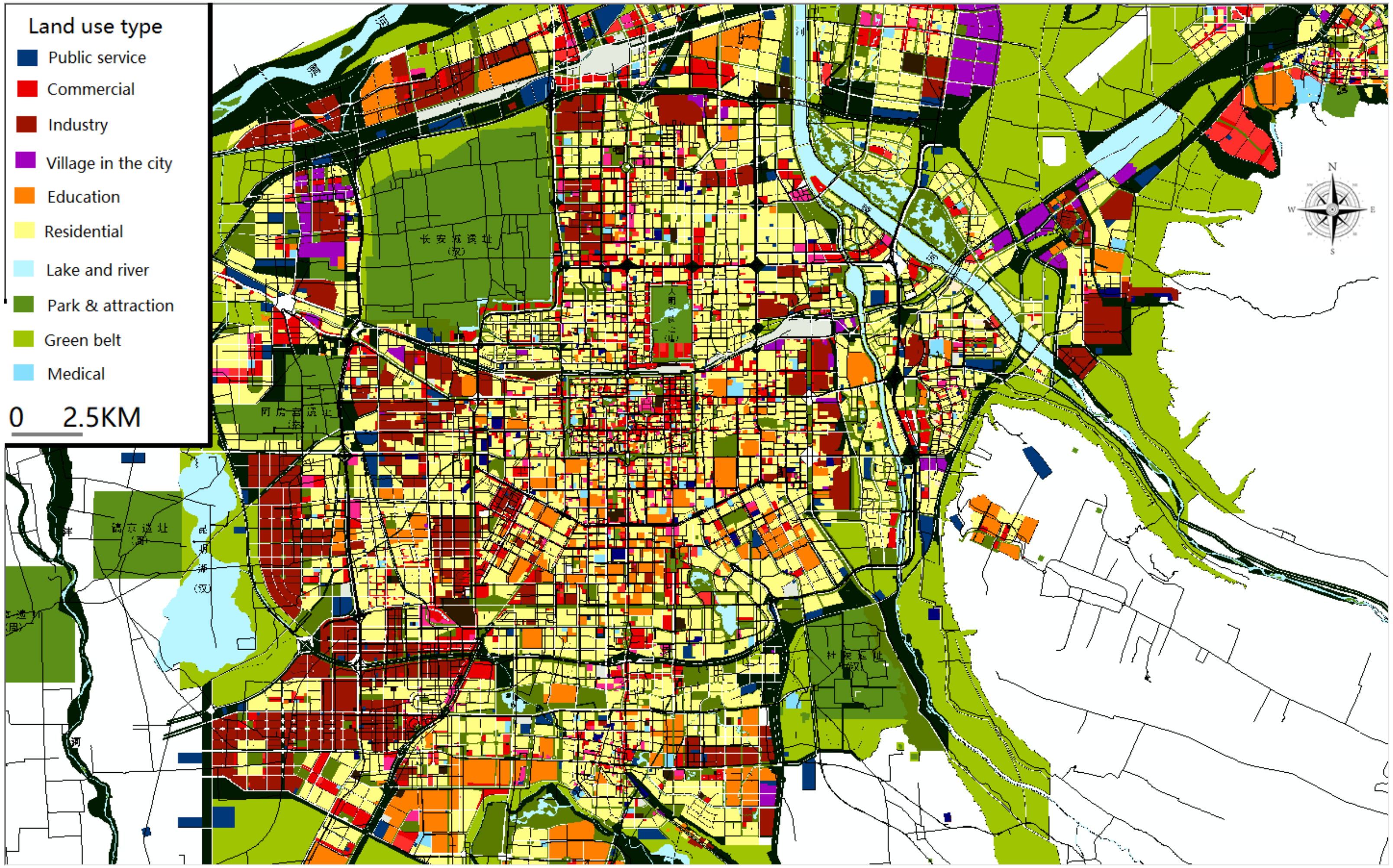

Figure 4.4 Urban land use data of Xi’an, 2019

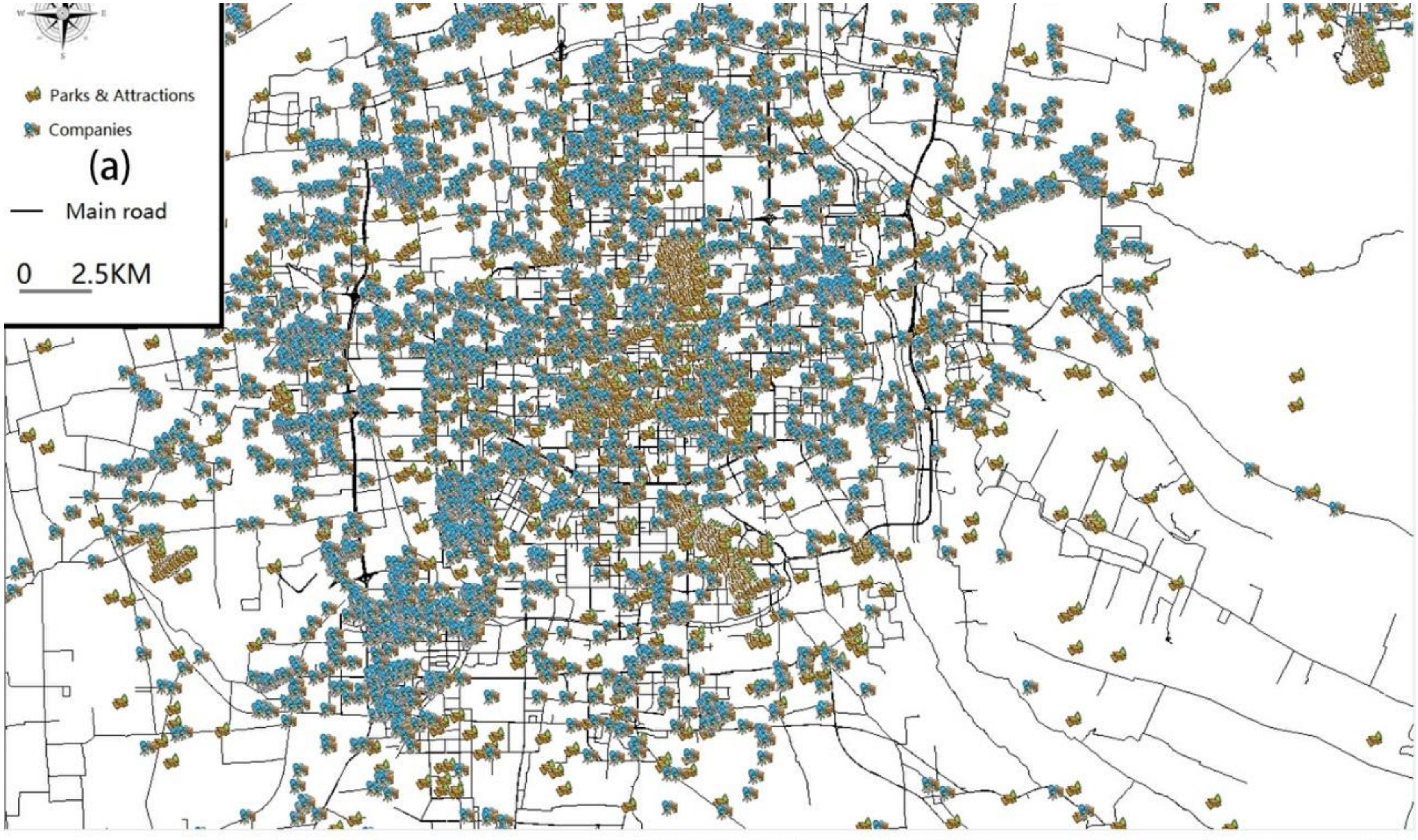

Figure 4.5a Parks, attractions and companies' (including factories) POI data of Xi'an, 2019

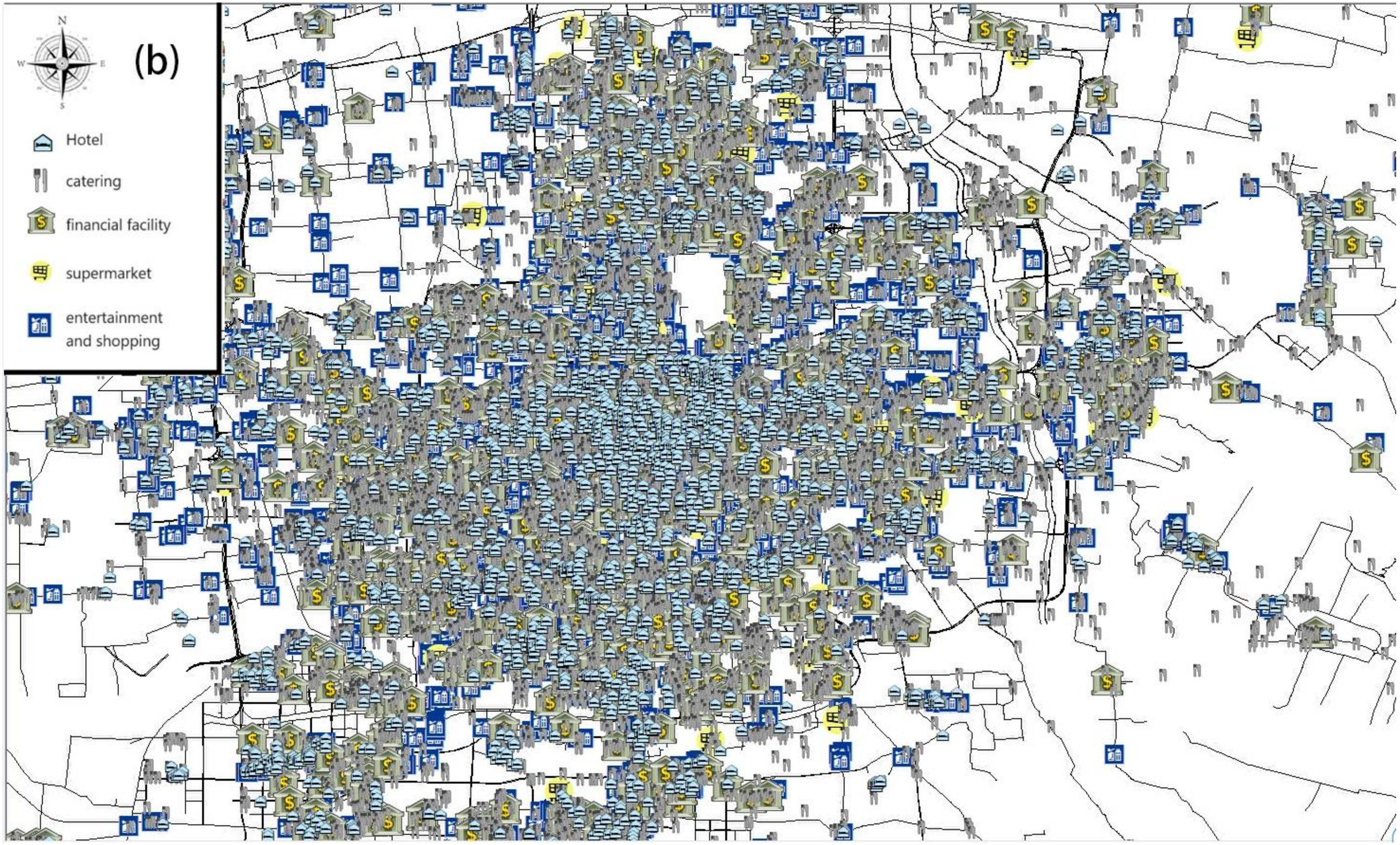

Figure 4.6b Tertiary industry facilities' POI data of Xi'an, 2019

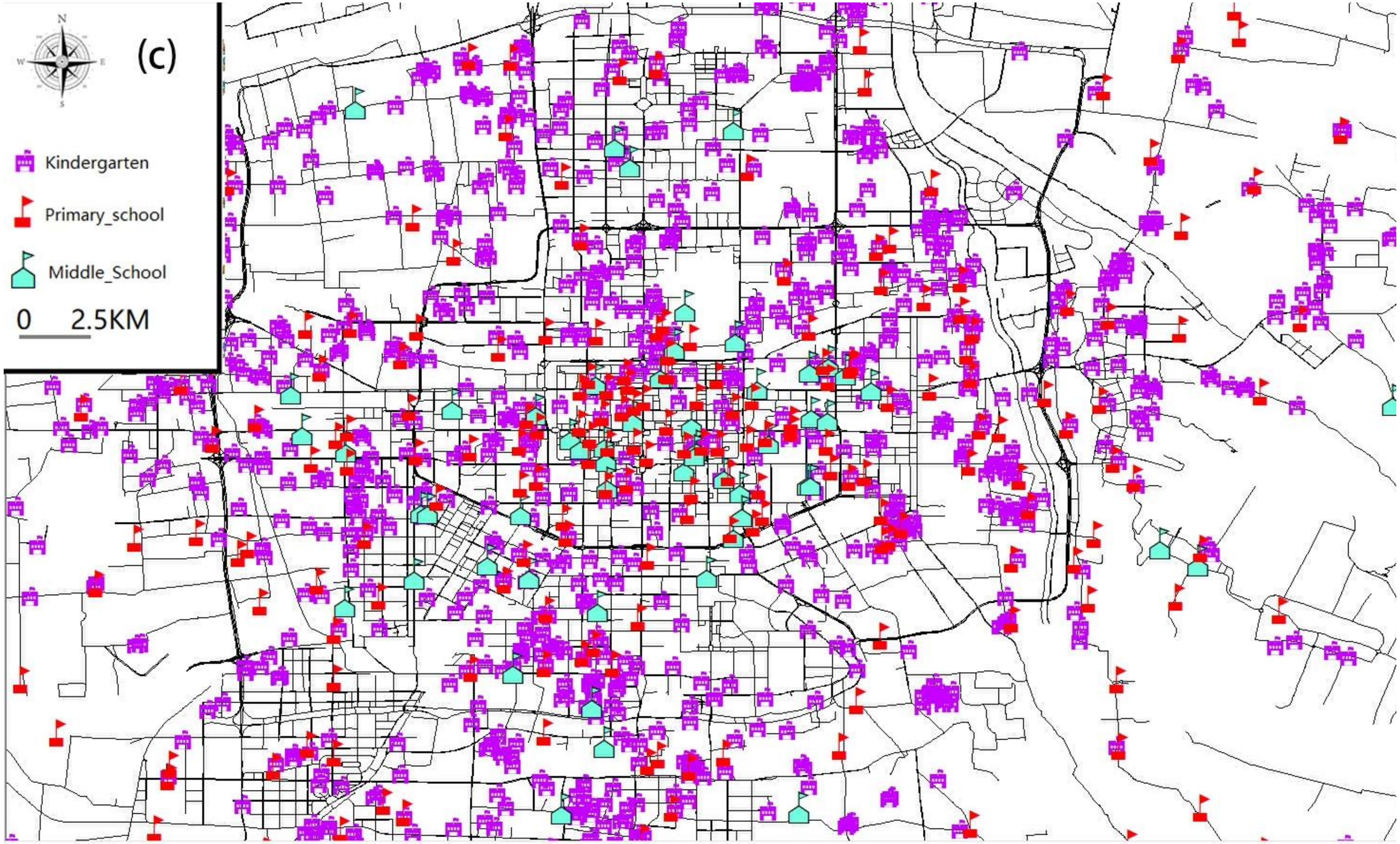

Figure 4.7c Kindergartens, primary schools, and middle schools' POI data of Xi'an, 2019

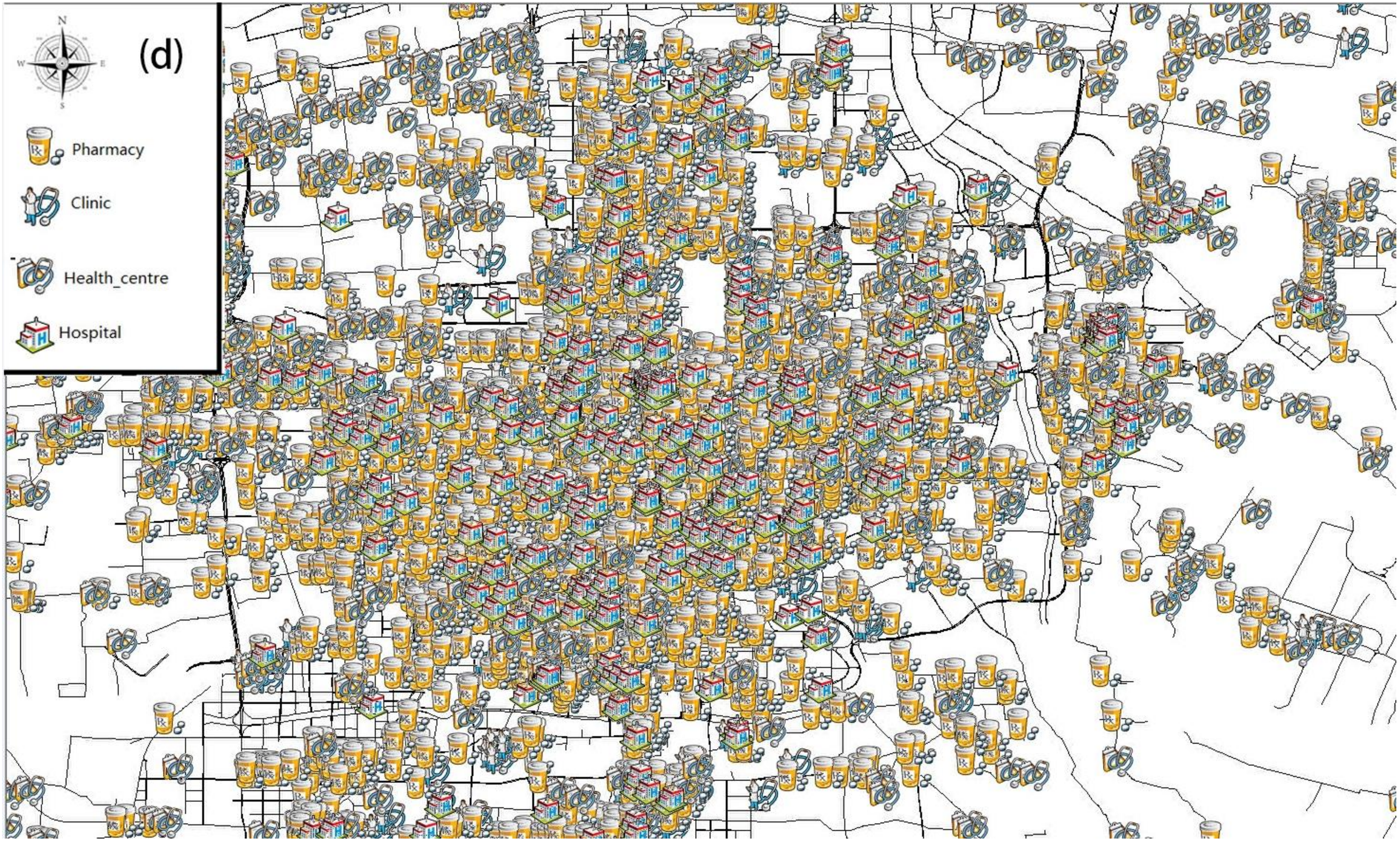

Figure 4.8d Medical facilities' POI data of Xi'an, 2019

Figure 4.9 The location of Shajing village and Zhongtian community

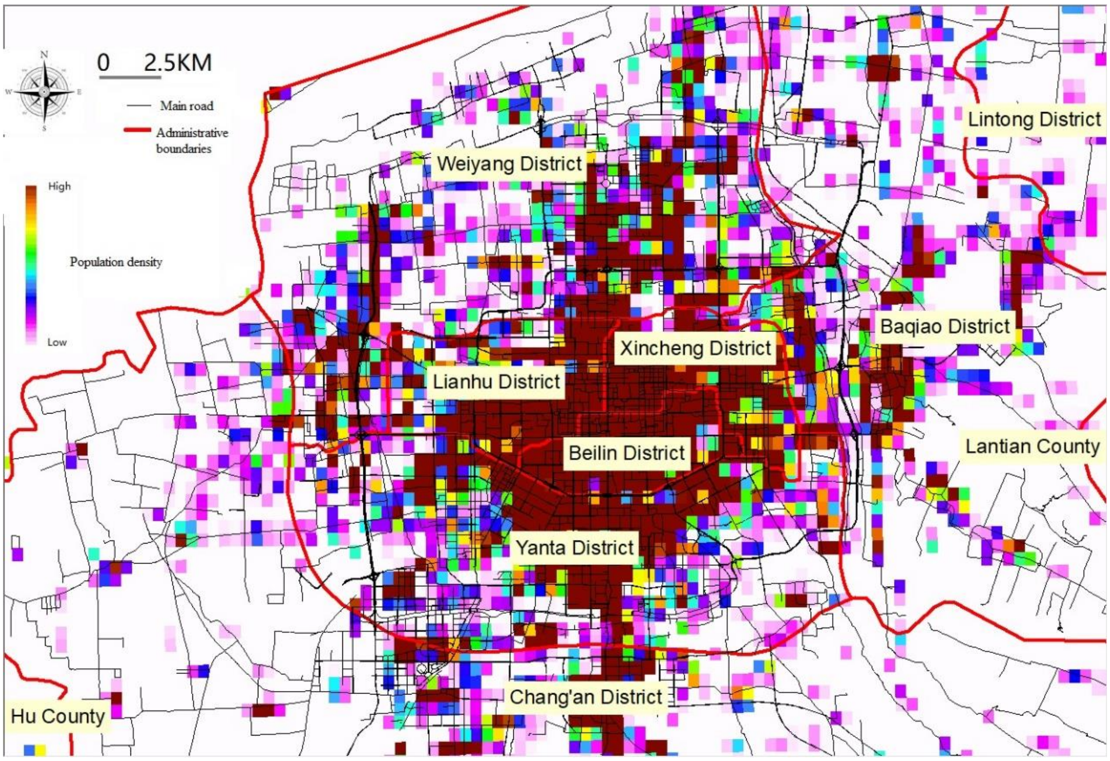

Figure 4.10 Population distribution of Xi'an, 2019

Table 4.3 **Data source**

Data source	fields	Description

Statistical Yearbook provided by the Xi'an Municipal Government. (http://tjj.xa.gov.cn/tjnj/2020/zk/indexch.htm)	Statistic data of Xi'an in 2019	Including population distribution in all districts and counties, average income by job type, location, Hukou type, average cost and other data
AMaps Company (It has basically the same functions as Google Maps and is a local map navigation service company in China.) https://lbs.amap.com/?ref=http%3A%2F%2Flbs.amap.com%2Fdev%2Fkey%2Fapp, Tencent Map Navigation Service (a company that competes with Amap in map navigation services) https://wiki.open.qq.com/wiki/lbs/get_poi	Open database of Point of Interest (POI) data	Factories, hospitals, hotels, companies, malls, supermarket, parks, bus stops, metro stations, financial facilities, public service facilities.
Land use planning and mapping of Xi'an (Porvided by Xi'an Surveying and Mapping Institute (a subordinate agency of the Xi'an Municipal Government responsible for urban surveying, mapping, and cartography))	Specific urban Land use raster and vector data.	Provided by Xi'an Institute of Surveying and Mapping. Polygons with specific land use as residential, commercial, industrial, educational and others land use data

Data provided by the official website of Anjuke, a real estate agency company (https://xa.fang.anjuke.com/fangjia)	House price and rent market data	Using python to collect on the Internet

4.2 Integrate the influence of e-commerce and Hukou policy

4.2.1 Basic idea

Land attraction measures the collective impacts of people's spatial decisions that higher the land attraction, more people will work and live in the zone. It can be quantified by first calculating the accessibility to urban factors as separate layers and then weighting the layers to consider the needs of different groups (e.g. hukou and non-local hukou). Accessibility depends on various factors, such as the distribution and capacity of road systems, public transport facilities, and other relevant factors when calculating the accessibility of urban facilities like schools or hospitals (Li et al., 2023). This research employs the term "*weight*" to represent how important the various factors are to Taobao villagers' spatial decision behaviours.

4.2.2 Methods

To comprehensively understand the spatial distribution pattern of Taobao Villages and the evolution of their spatial structure, related research (Juan Lin et al., 2021) adopts a regional perspective, taking the number of Taobao Villages in each county of Zhejiang Province as the dependent variable. It points out that six categories of independent variables, including socio-economic and locational environmental factors such as land cost, transportation cost, industrial foundation, labour supply, agglomeration effect, and control variables, have influenced the formation of Taobao Villages (Juan Lin et al., 2021). Subsequently, the study employs panel regression models to analyse the correlation between the aforementioned dependent and independent variables, emphasizing that land and transportation costs significantly impact the development pattern of Taobao Villages (Juan Lin et al., 2021). This research method can determine the extent to which different factors influence the spatial distribution of Taobao Villages. It is a top-down approach that assumes macro-level policies, such as land policies, have a consistent impact on the choices of numerous individuals, and thus holds certain advantages when assisting relevant managers in formulating policies.

Meanwhile, this study argues that investigating the spatial distribution of newly emerging Taobao Villages necessitates a bottom-up approach that delves into the needs and decision-making mechanisms of the villagers residing in these villages. This approach emphasizes understanding the individual preferences and motivations that drive villagers to select specific locations for their Taobao businesses.

By focusing on the villagers' demands and spatial decision-making behaviors, the study can gain valuable insights into the micro-level factors that shape the emergence and distribution of new Taobao Villages. This bottom-up perspective complements the macro-level analysis by offering a nuanced understanding of how individual actions and choices interact with broader policy and market forces.

The spatial decision behaviours of Taobao villagers related to multiple types of facilities and services (Figure 4.8)(Lin, 2019). By both considering the needs of Taobao villagers and the distribution of regional characteristics, this research calculated the land attraction for rural e-commerce migrant who will work and live in Taobao villages.

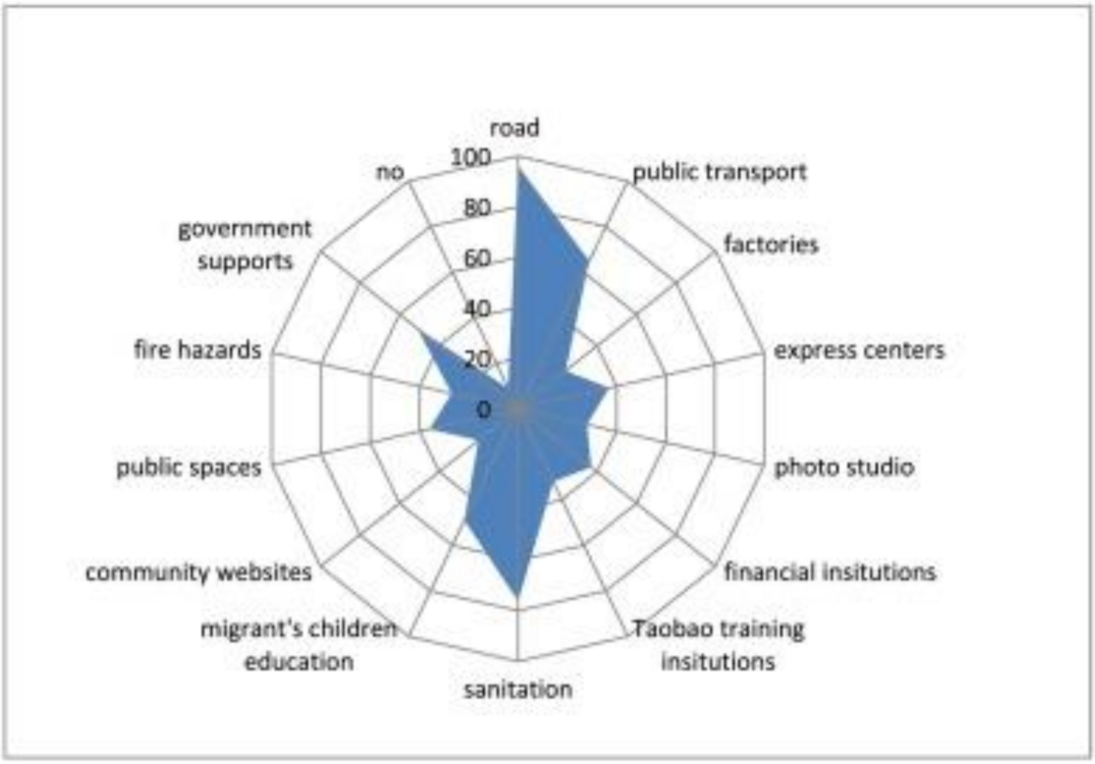

Figure 4.11 Factors and scores considered by Taobao villagers when Choosing spatial location (Lin, 2019).

Unlike other aspects can be analysed and quantitative described by spatial analysis (e.g., road, public transport), the evaluation of government support hardly can be quantitative described. The government support for e-commerce is aimed to improve working and living conditions of e-commerce workers. Therefore, Gross Domestic Product (GDP) of cities can approximately replace the government support in spatial analysis in China. Because the better the local economic development, the more resources local government can use for government support.

The weight of an aspect depends on its influence. Figure 4.8 uses different points to describe those influences. Based on that, this study uses the calculation formula for weight distribution (Equation [2]).

This study establishes a spatial analytic model for accessibility calculation by employing weight calculation, spatial overlay, and density analysis tools.

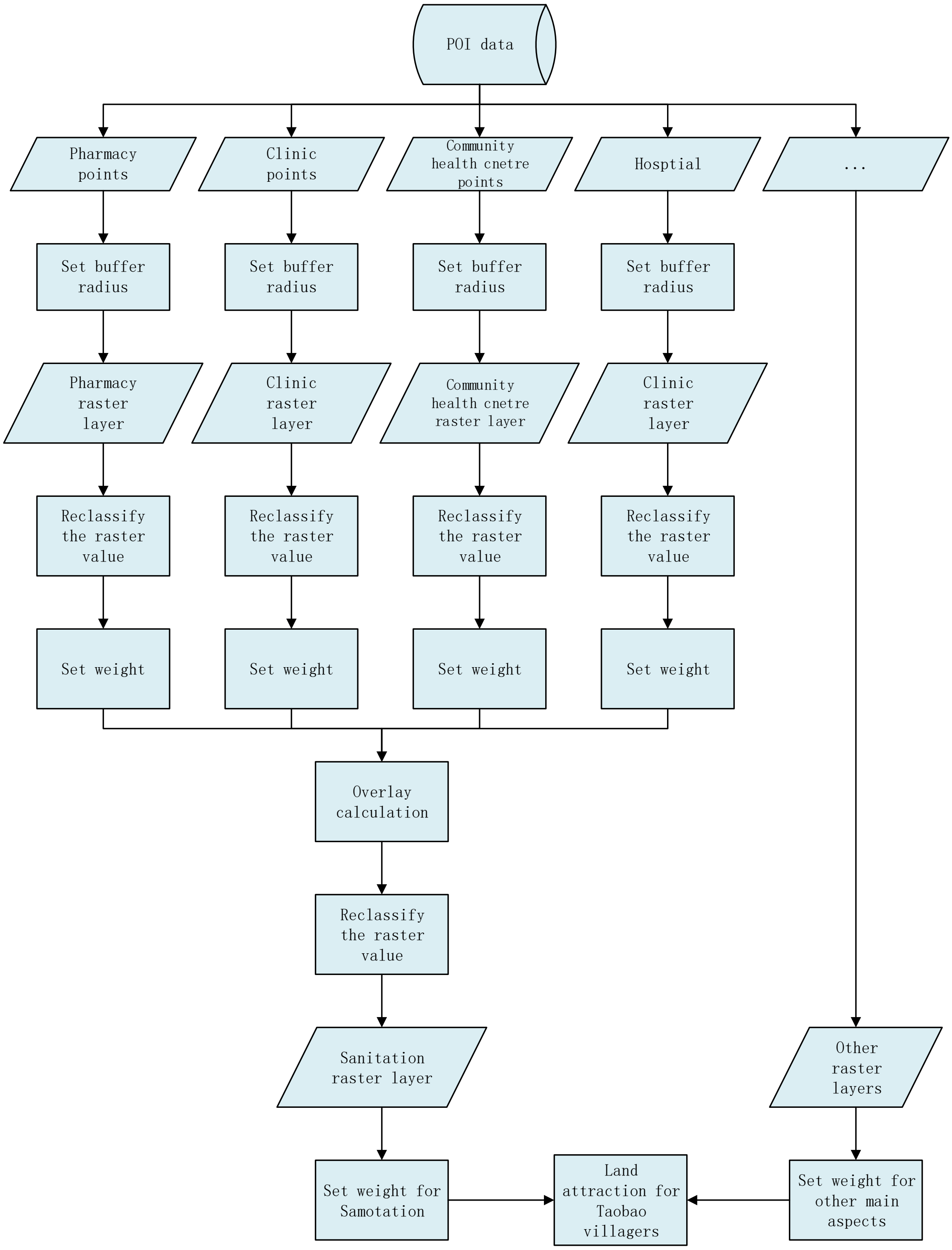

Figure 4.12 Spatial analytic model

Figure 4.9 shows the structure of spatial analytic model. Firstly, this study selected top five main categories from Figure 4.8 as follows: Sanitation, Road, public transport system, Children's Education conditions and Government support to calculate the land attraction.

Secondly, urban factors (supermarket, malls, schools, hospitals and others,) are converted to Point of Interest (POI) data. Those POI data are subdivided into those main categories. For instance, The Sanitation includes the distribution of hospital, health centre, clinic and pharmacy. The Road includes the distribution of high way, main road and branch road. The public Transport system includes the distribution of bus stops and metro stations. The Children's Education conditions include the distribution kindergarten, primary school, junior school and high school. Afterward, the buffer zone radius for each point is set based on the data type to convert point data to raster data. The value of spatial units (e.g., 0.1 square km) in the raster layer is reclassified on a scale of 0 to 100, and all raster layers are grouped into different classes. Within the same main category, different weights are assigned for each raster layer, and all layers are overlaid to calculate the value of influence accumulation. The weight of an urban factor is determined by the capacity of the urban factor, i.e., how many people it can serve within a certain time span. The value of a pixel represents the value of influence of a point at that location, which is affected by distance decay. Closer to the point, the higher the influence value, and vice versa. The influence value of a point's buffer zone decreases linearly from the centre to the surrounding areas, and outside the buffer zone, the influence value of points becomes zero. Then, based on their respective influence, radius is assigned for each type of POI data, and point data is converted to raster layers. Afterward, different weights are set for each layer, and the resulting raster layer of medical facilities is calculated. The value of pixels in the result raster layer is reclassified on a scale of 0 to 100. Finally, various weights are assigned to different layers, and spatial overlay tools are used to calculate the distribution of the land attraction for e-commerce migrant.

Table 4.4 Classification of urban factors and Weight assign for Taobao villagers (Based on Figure 4.8 and (Li et al., 2023)

Main aspects	Urban factors	Weight (inside the aspect)	Weight of main aspect
Sanitation	Pharmacy	0.01	22%
	Clinic	0.03	
	Health centre	0.08	
	Hospital	0.88	
Road	High way	0.51	29%
	Main road	0.36	
	Branch road	0.13	
Public transport system	Bus system	0.46	19%
	Metro system	0.54	
Children's Education conditions	Kindergarten	0.25	15%
	Primary school	0.25	
	Junior school	0.25	
	High school	0.25	
Government support	GDP		15%

This research used equation [2] to calculate the weight distribution in different main categories. It should be noted that inside the same category, different urban factors have different weights. The weight of an urban factor is determined by two main aspects: the impact of the urban factor on the residential location decision-making of urban residents and the capacity of the urban factor, i.e., how many people it can serve within a certain time span (Li et al., 2023). This research used the staff numbers of medical facilities to reflect the level of service quality, average speeds of bus and metro systems to reflect the level of service quality for public transport systems, traffic volume (Unit person-time) of business districts to reflect the level of service quality, and maximum speed limits and number of lanes for road systems to reflect the level of service quality. Primary, junior, and high schools are equally weighted. In China, 2019, the staff number of hospital is usually over one thousand, health centre is five or six, clinic is around five staff, and pharmacies is between one to three. The average operation speed of bus system is 20 km per hour and that of metro system 35 km. The traffic volume (Unit person-time) of a mall usually three or four times larger than that of a big supermarket, being dozens time larger than a small community shop. The maximum speed limit and number of lanes of highway is 120 km per hour and four or six lanes, that of main road is 70 km per hour and four or eight lanes, and that of branch road is 50 km per hour and two or four lands. This research used their average lanes number. Five for highway, six for main road and three for branch road. Based on that, this research set different weights for different factors (Table 4.4).

4.2.3 Experimental Data

According to 2021 statistical yearbooks of cities, this research collected GDP data (Figure 4.10) of cities in China mainland.

This study takes China as the research area to analyse the relationship between land attraction and e-commerce migration, and the impact of land attraction on the distribution of emerging Taobao villages. To reach this aim, highway, roads, railways, population distribution and other data of China had been collected. Depends on POI data, this work analyses and calculates the density of transport networks (Figure 4.11) and medical POI data, including highways, roads, railways, hospitals, community health centres, clinics, pharmacies, and etc. (Figure 4.12). Table 4.5 shows the data sources of this study.

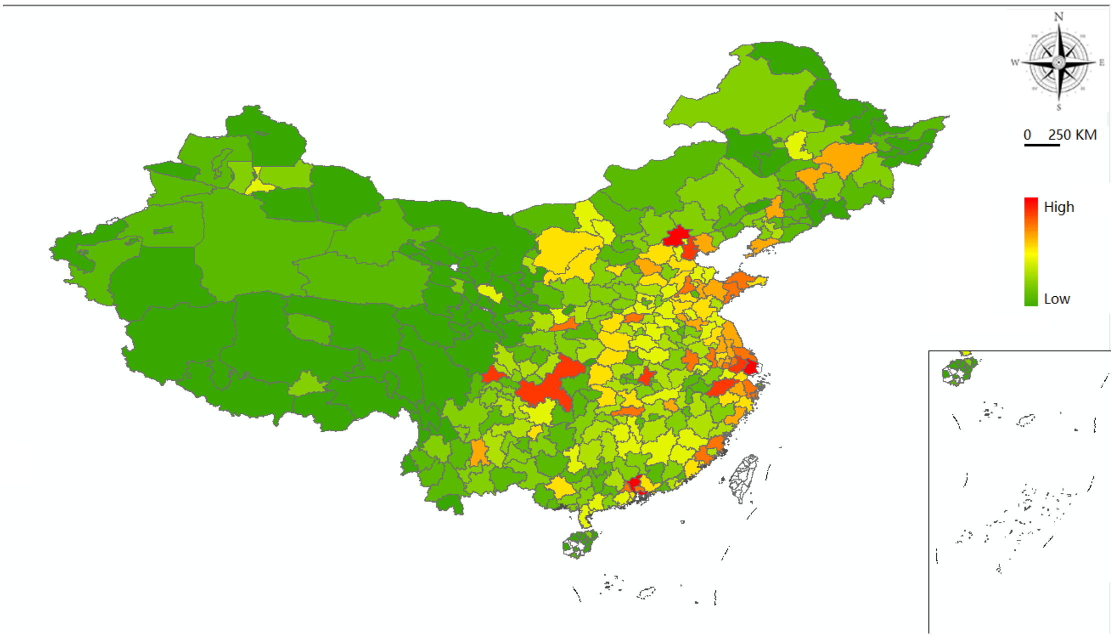

Figure 4.13 The distribution of cities' GDP in China mainland(City level), 2019

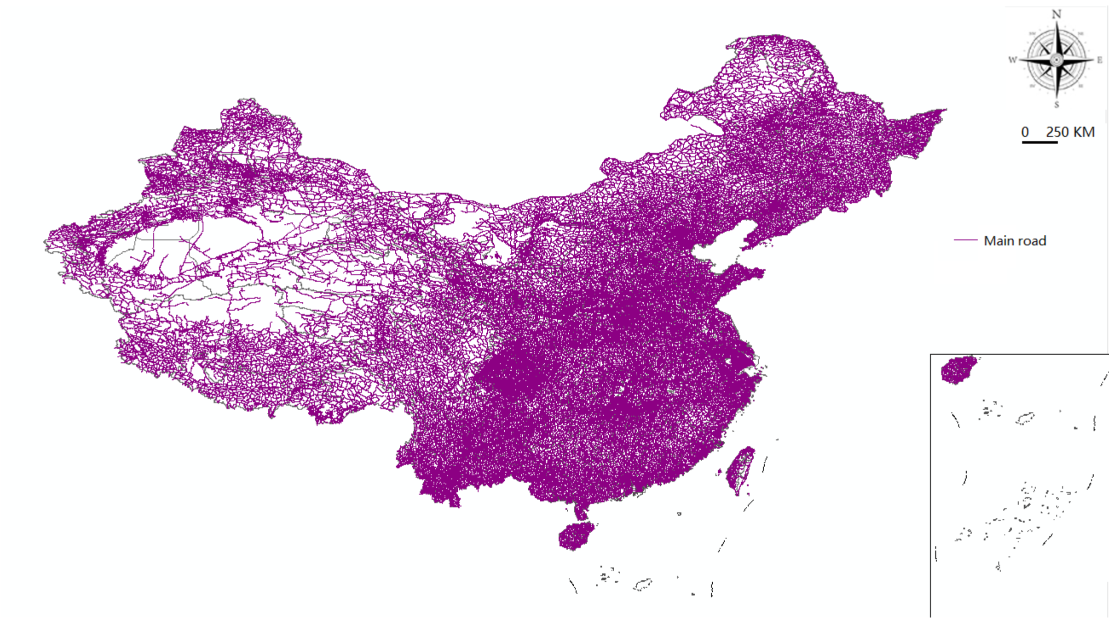

Figure 4.14 Transport networks in China, 2019

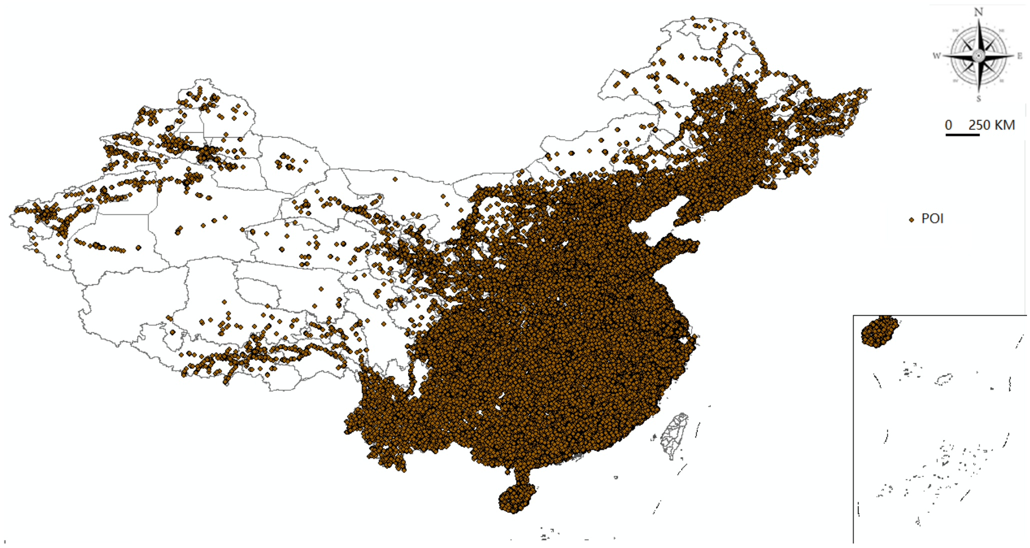

Figure 4.15 POI data of China mainland, 2019

Table 4.5 Data sources of this case study

Data source	fields	Description
CNKI (China National Knowledge Infrastructure), the largest academic paper database in China. http://www.cnnic.net.cn/	Internet related data	China Internet network information center
AMaps Company (It has basically the same functions as Google Maps and is a local map navigation service company in China.) https://lbs.amap.com/?ref=http%3A%2F%2Flbs.amap.com%2Fdev%2Fkey%2Fapp, Tencent Map Navigation Service (a company that competes with Amap in map navigation services) https://wiki.open.qq.com/wiki/lbs/get_poi	Open database of Point of Interest (POI) data	Factories, hospitals, hotels, companies, malls, supermarket, parks, bus stops, metro, etc.
AMaps Company (It has basically the same functions as Google Maps and is a local map navigation service company in China.) https://lbs.amap.com/?ref=http%3A%2F%2Flbs.amap.com%2Fdev%2Fkey%2Fapp,	Open database of Point of road data	Including highways, ways and railways network in China

Tencent Map Navigation Service (a company that competes with Amap in map navigation services) https://wiki.open.qq.com/wiki/lbs/get_poi		
Alibaba Research Institute (a subordinate agency of the parent company of Taobao) http://www.aliresearch.com/	Taobao villages	List of the name and location of taobao villages
Statistic year books of cities in China (Porvided by Chinese local governments)	Economic, population, governments' budget and other data	Local government public costs for local hukou residents

4.3 Summary of methodology

This chapter presents the methods proposed to understand and analyse the impacts of Hukou policy and e-commerce on land attraction for a certain resident group (e.g., local and non-local Hukou residents).

Conceptual models, and indicators have been created to describe the changes in distribution of economic activities, human spatial location choices, travel needs and behaviours for a certain resident group.

The main subsections of the methodology are

- Definition of the transport impedance indicator to calculate the unequal distribution of transportation resources between local and non-local Hukou residents.
- A spatial analysis to calculate the land attraction for local and non-local Hukou residents, and used that to predict population flows of Local and non-local Hukou to particular areas.
- A method to construct a conceptual model consisting of land use, population and economic activities sub-models to deal with the challenge of the digital economy on urbanism.

Chapter 5

The impact of Hukou policy: a case study in Xi'an

5.1 Introduction

Hukou policy significantly impacts on people's life styles, and local and non-local residents usually reside in different communities (Chapter 2). In this case study, Xi'an, capital city of Shaanxi province in China, was chosen as the research object to calculate the transport impedance disparity indicator between local and non-local hukou residents' communities. The result indicates that the transport impedance disparity indicator effectively reveals the different treatments between local and non-local hukou residents' communities in the same city. Compared to local hukou residents, non-local hukou residents experience clear disadvantages in terms of accessibility.

To understand and analyse the impact of Hukou policy on land attraction, this chapter proposes the Transport impedance indicator to analyse and describe the level of difficulty in a community to access urban characteristics through the transportation system. Then, this chapter proposes the Transport impedance indicator to reflect the difference of Transport impedance indicator in local and non-local Hukou communities.

This chapter presents visualisation of and spatial analysis using GIS to develop and present a transport impedance indicator for the different administrative zones in Xian city region.

5.1.1 General context and descriptive maps of the case study area

Xi'an is the capital city of Shaanxi province (Figure 5.1 (a)). Xian is the largest city in Shaanxi Province in terms of both population and GDP. In 2019, Xi'an has a total of 10.2035 million residents, including 7.6128 million urban residents and 2.5907 million rural residents. 6.6857 million urban residents have local Hukou, the remaining 0.9271 people are non-local Hukou residents.

The population distribution is as follows. The Density in the core area is between 765 and 29421 person per square km, the density in village in the city areas is between about 10,000 and over than 100,000, the density in County areas is between 201 and 267.

The administrative zones shown in the map are Figure 5.1a and Figure 5.1b. Figure 5.1a illustrates the geographical location of Shaanxi Province and Xi'an City in China. The yellow area represents the administrative region of Shaanxi Province, while the light red area represents the administrative region of Xi'an.

Figure 5.1b demonstrates the administrative divisions of districts and counties in Xi'an. The light-yellow area represents the administrative districts of Xi'an City, while the light green area represents the administrative counties of Xi'an (rural areas are mainly concentrated within the administrative area of the county).

Figure 5.1c shows the population density across the city region.

In Xi'an, nearly 80% of the population live in those districts, and the majority of them are urban Hukou residents. The population with rural Hukou mainly lives in those county areas. Administratively, those county areas and districts are part of the city. The administrative region and political status of a city in China are basically the same as those of a shire in the United Kingdom. For example, the administrative area of Xi'an is roughly equal to that of Yorkshire.

Figure 5.1a Location of Shaanxi province and Xi’an in China

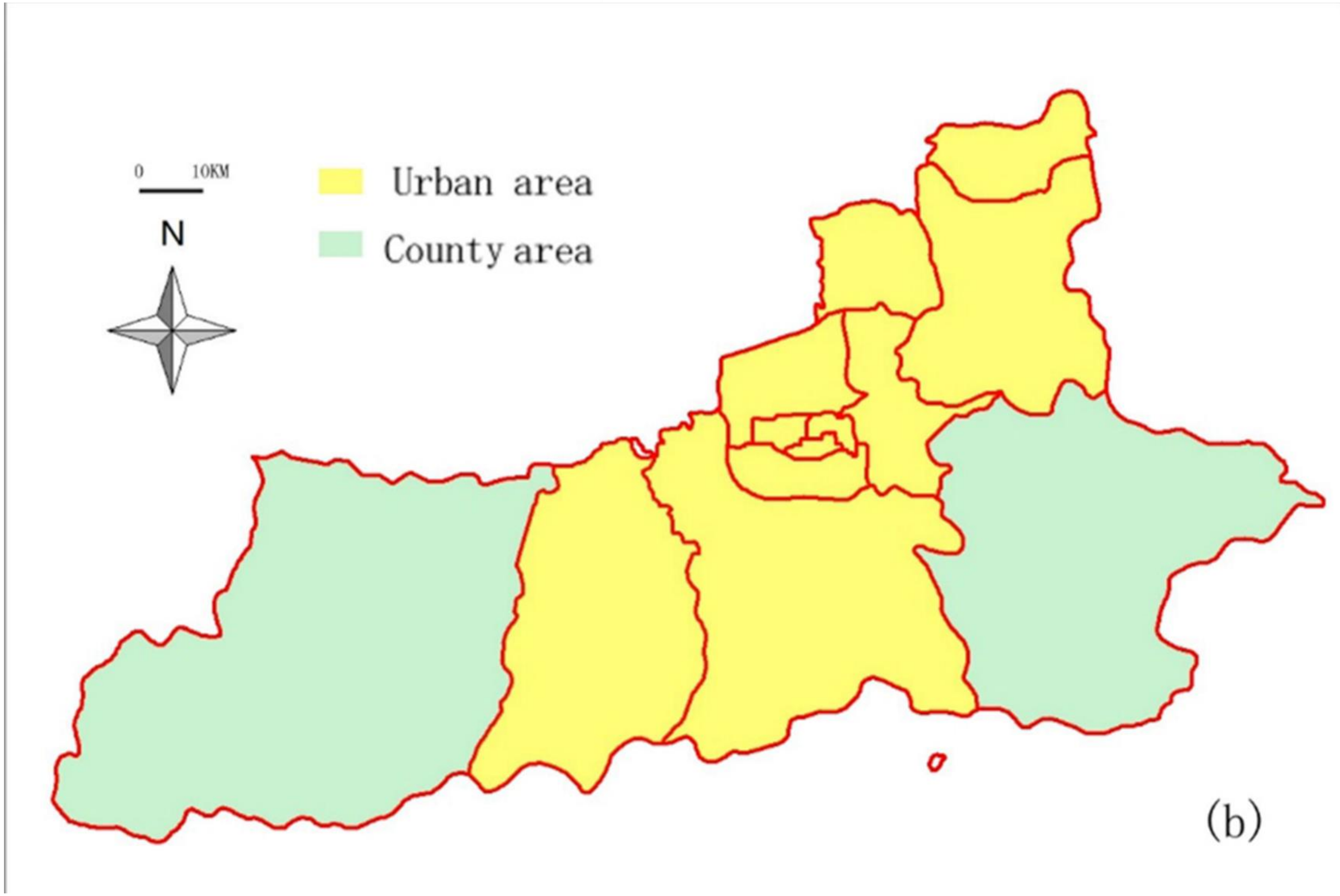

Figure 5.2b Urban and County areas of Xi’an (b)

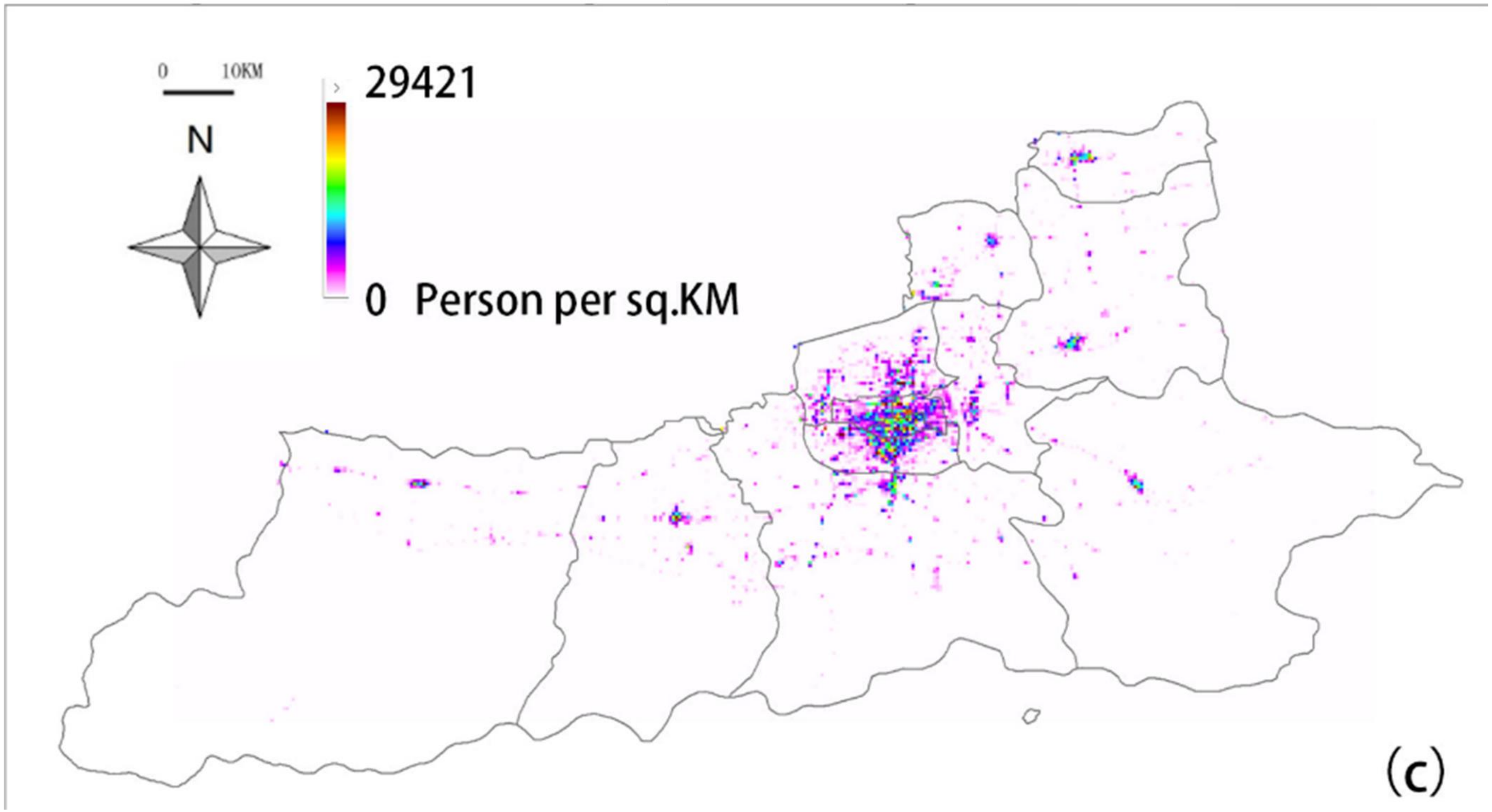

Figure 5.3c population density of Xi'an (c)

Figure 5.2a uses the brown colour to indicate the spatial location of the built-up area in Xi'an. Figure 5.2b is a land use distribution map of the built-up area, where dark blue represents public service land (including government agencies, police stations, fire stations), red represents commercial areas, brown represents industrial areas, purple represents urban villages, orange represents educational land (including universities, middle schools, primary schools, and kindergartens), yellow represents residential areas, cyan represents rivers and lakes, dark green represents parks and places of interest, light green represents urban green belts, and light blue represents land for medical facilities.

According to (*Rental monitoring report for 40 key cities nationwide*, 2021), the average monthly rent per house in Xi'an's commercial residential areas was 2,144 yuan, accounting for 61% of the per capita disposable income per person. Non-local Hukou holders may struggle to afford this due to their lower average incomes compared with local Hukou holders. Therefore, this study considers all villages in the city as non-local Hukou residents' communities in Xi'an, while the remaining communities are regarded as local Hukou residents' communities. According to history data, in 2009, all non-local hukou residents, 2.4 million, living in village in the city (Wang and Li, 2014). The GDP in Xi'an is 932.1 billion yuan. The land area of Xi'an is 10096.81 square kilometres, of which 51% are urban areas and 49% are County areas (Figure 5.1 (b)). The built-up area of Xi'an is 729.14 square kilometres (Figure 5.1 (c)), and there are about 186 villages in the city located in the built-up area.

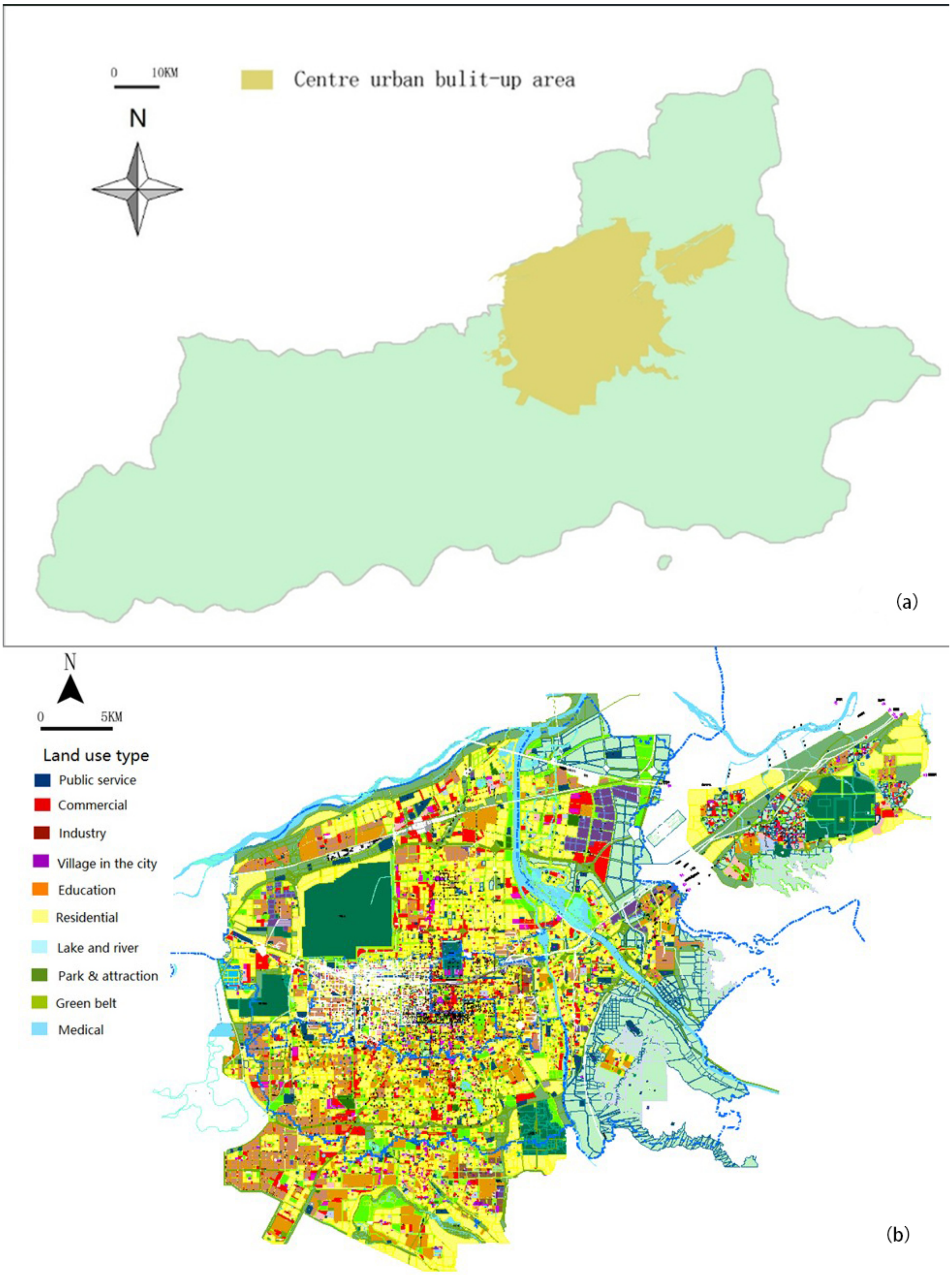

Figure 5.4 Land use of central urban built-up area in Xi'an

5.2 Calculation of transport disparity indicator for urban built-up areas in Xi'an

As discussed in Section 2.1, 2.2 and 3.4, the literature shows that local and non-local Hukou status residents are likely have different travel needs to certain extent. The transport impedance indicator for two Hukou groups, local and non-local Hukou residents, based on different travel needs and behaviour has been established in the present research. Then, the transport impedance disparity indicator was established to describe the transport impedance gap between those two groups.

5.2.1 Data used

This case study utilized data from Xi'an City in 2019, encompassing electronic maps, road network data, land use data, POI (Points of Interest) data, population data, and economic data. Detailed descriptions and listings of these data types are provided in Section 4.1.4 (Table 4.2 and 4.3).

5.2.2 Application of methodology

First, this study collected data on land use, economy, population, electronic maps and POI of Xi'an from multiple data sources, as shown in table 4.1. Then, utilizing database processing techniques in ArcCatalog and Arcmap, the collected data was converted into point, polyline, polygon, and raster data. These data were then stored in different subsets of the spatial database based on their respective attributes (Table 4.1). Third, according to the results obtained from the previous discussion (Section 2.1, 2.2 and 3.4), this study selected the POI data related to local and non-local Hukou residents' different travel needs and behaviours. Fourth, taking into account the needs of residents (Section 2.1 and

Section 4.1.1) and the inherent attributes of urban facilities, this study applied Equation [2] (Section 4.1.3.2) to calculate the weights of the main aspects and each sub-item in Table 4.2. The results are presented in Table 5.1. Fifth, this study utilized spatial analysis tools in GIS, including the kernel density tool, buffer analysis tool, reclassification tool, and spatial overlay weighted calculation tool, to calculate the travel impedance indicator for local and non-local Hukou residents based on their respective characteristics. Finally, based on the results from the fifth step, the transport impedance disparity indicator between local and non-local Hukou residents was calculated. This research used equation [2] (Section 4.2.3) to calculate the weight distribution in different main aspects (Table 4.1). According to the ("Statistical Bulletin of the Development of Health Undertakings in Xi'an City, 2020," 2020), issued by the Xi'an Municipal Government, the total number of employees in all hospitals in Xi'an was 112,323 in 2019, and the total number of employees in community health centres was 10,919, and the total number of employees in clinics is estimated to be approximately 6,893, with a total of 4,709 pharmacists and pharmaceutical technicians. Meanwhile, considering that pharmacies can only provide pharmaceutical services and lack diagnostic, inpatient, and rehabilitation services compared to other medical institutions, this thesis hypothesizes in this study that the weightage of pharmacies should be further reduced by three-quarters. According to the public news reports of Xi'an Public Transport Corporation and Xi'an Metro Corporation, in 2019, the total passenger volume of bus system is 1.405 billion and that of metro system is 0.946 billion. The average operation speed of bus system is 20 km per hour and that of metro system 35 km per hour (passenger km). Based on the data of field surveys, the number of mall visitors is usually about 3 to 4 times that of larger supermarkets. The maximum speed limit and number of lanes of highway is 120 km per hour and four or six lanes, that of main road is 70 km per hour and four or eight lanes, and that of branch road is 50 km per hour and two or four lands. This research used their average lanes number. Five for highway, six for main road and three for branch road.

Kernel Density tool of Density in Spatial analyst tools of Arc toolbox is used as research instrument. The tool was used to input each bus stop, metro station, hospital, and other urban characteristic data and convert density data into raster data layers respectively. The search radius was set at different distances for different urban characteristics based on the type of urban resources. The buffer radius of an urban factor determined by its weight and the average commute distance of the city. According to the "2020 National Commuting Monitoring Report on Major Cities" issued by the China Academy of Urban Planning and Design, the average commute radius, the average straight-line distance between the place where people work and live, Xi’an is 8.3 km in 2019. The buffer zone radius of different facilities was calculated by multiplying the average commuting distance by the weight. For example, the buffer zone radius of hospitals is equal to 8.3 km multiplied by the weight of hospitals, which is 0.88, resulting in approximately 7.3 km. In addition, some urban characteristics of the same type are closely adjacent to each other, such as restaurants and bus stops. Thus, the distance of an average adult walking in ten to fifteen minutes was used to set the radius for them as 1 km. In Xi'an, almost all bus lines have the same service frequency; thus, this research assumes that the density of bus stops can represent bus service ability in a certain way. Distance decay was used to analyse the metro station buffer zone. The subway stations are used as the central points of the buffer zones. The value at the centre of the buffer zone is the highest, and then it decreases linearly from the centre to the edge as the distance increases. Higher values indicate more bus stops in the vicinity.

Table 5.1 Classification of urban factors and weight distribution (Li et al., 2023)

Main aspects	Urban factors	Weight (inside the aspect)	Weight of main aspect for local Hukou residents	Weight of main aspect for non-local Hukou residents
Workplace distance	Workplace (e.g., factory, company and etc.)	1	10%	11%
Medical conditions	Pharmacy	0.01	16%	16%
	Clinic	0.03		
	Health centre	0.08		
	Hospital	0.88		
Road system	High way	0.51	10%	6%
	Main road	0.36		
	Branch road	0.13		
Business district	Mall	0.76	12%	14%
	Supermarket	0.21		
	Small shop	0.03		
Public transport system	Bus system	0.46	20%	51%
	Metro system	0.54		
Green leisure space	Park Urban green belt		2%	2%
Children's Education conditions	Kindergarten	0.25	30%	0%
	Primary school	0.25		
	Junior school	0.25		
	High school	0.25		

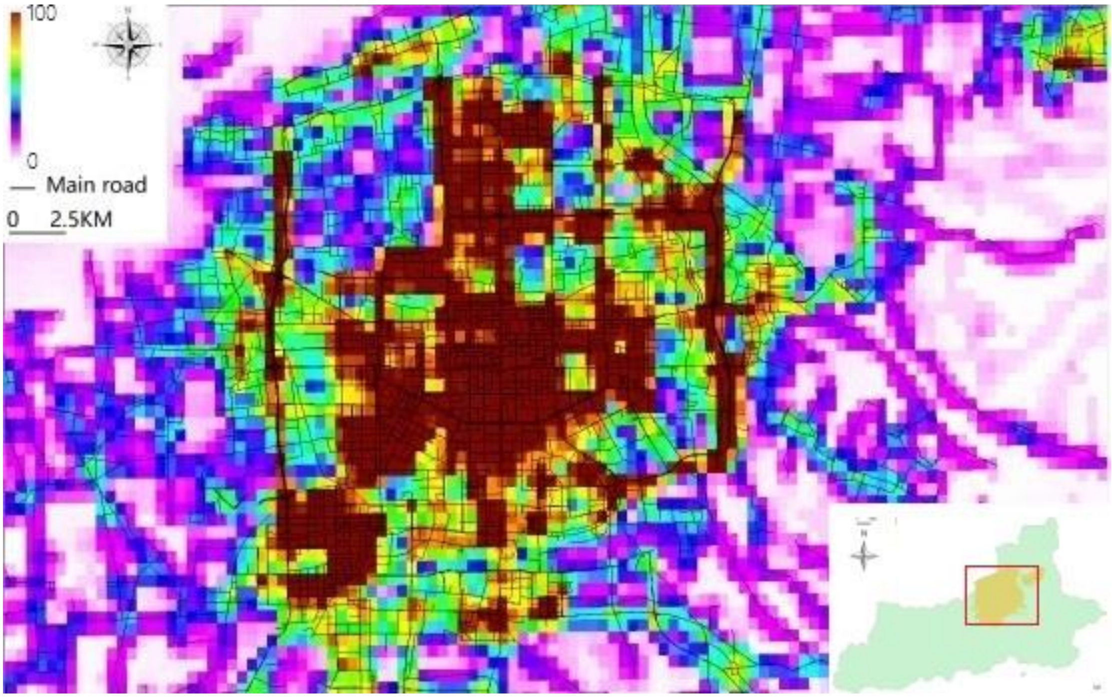

Figure5.5 Land attraction for local Hukou residents in the core built-up area of Xi'an, 2019

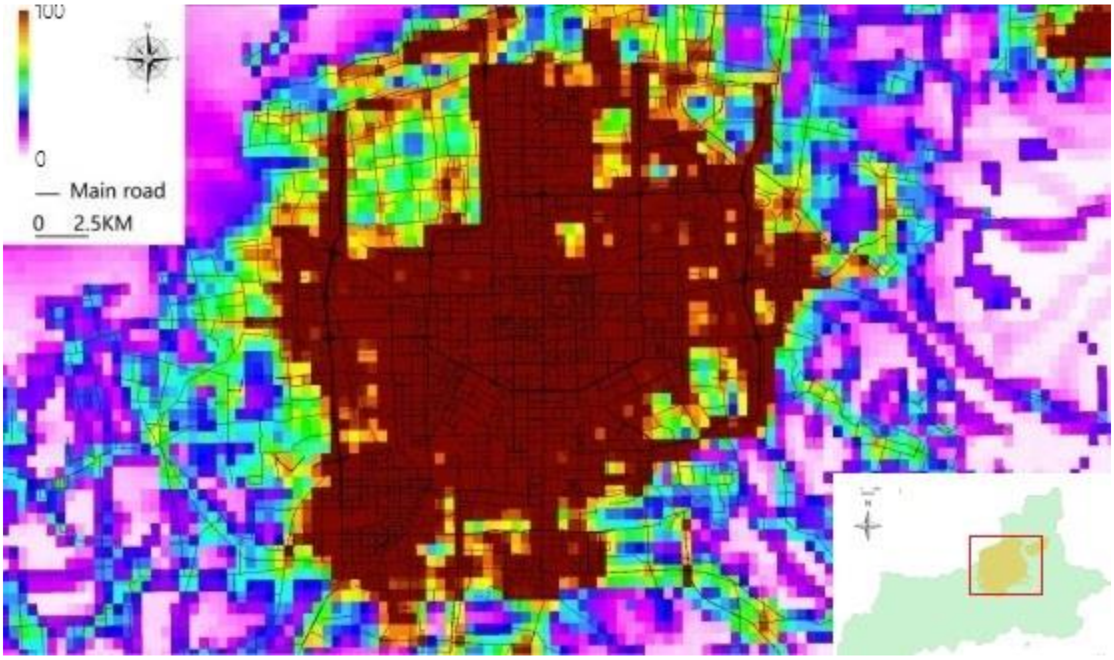

Figure 5.6 Land attraction for non-local Hukou residents in the core built-up area of Xi'an, 2019

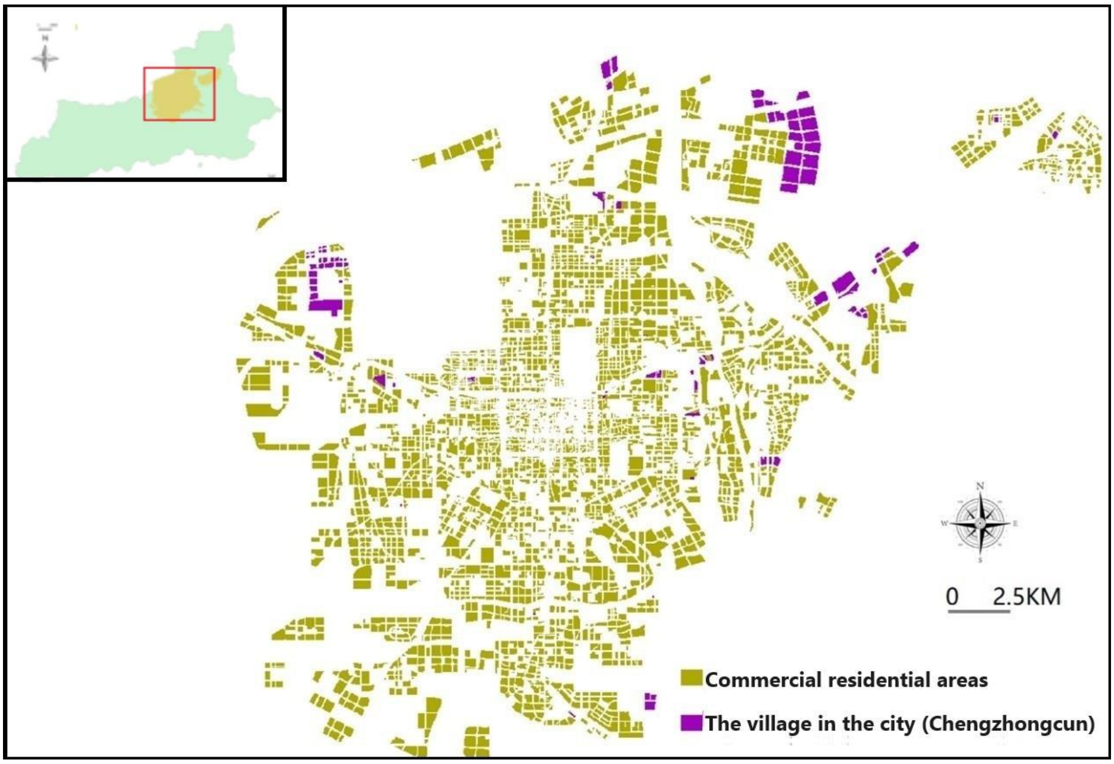

Figure 5.7 Commercial Residential areas and Chengzhongcun distribution of Xi'an in core built-up area

The final step in this research was to assign different weights to different rate data of different classes based on the different travel needs and behaviours of local and non-local Hukou residents' communities. This was done to calculate Land attraction for different communities. Figure 5.3 shows the Land attraction for local Hukou residents' communities, while Figure 5.4 shows the same for non-local Hukou residents' communities.

As Figure 5.6 shows, to extract value of raster image (Figure 5.3 and Figure 5.4) by residential areas (Figure 5.5), this research used the Extract by Mask tool of the Extraction toolbox in Spatial Analyst tools. By using statistical tools to calculate the extract data in certain region, it can indicate the land attraction value of the region. The value can be substitute into the parameter S in the Equation [2]. Then, by substituting the average income of residents in the region into the parameter I (Equation [2]) and the average cost on transportation into parameter C, the transport impedance indicator for the region can be calculated.

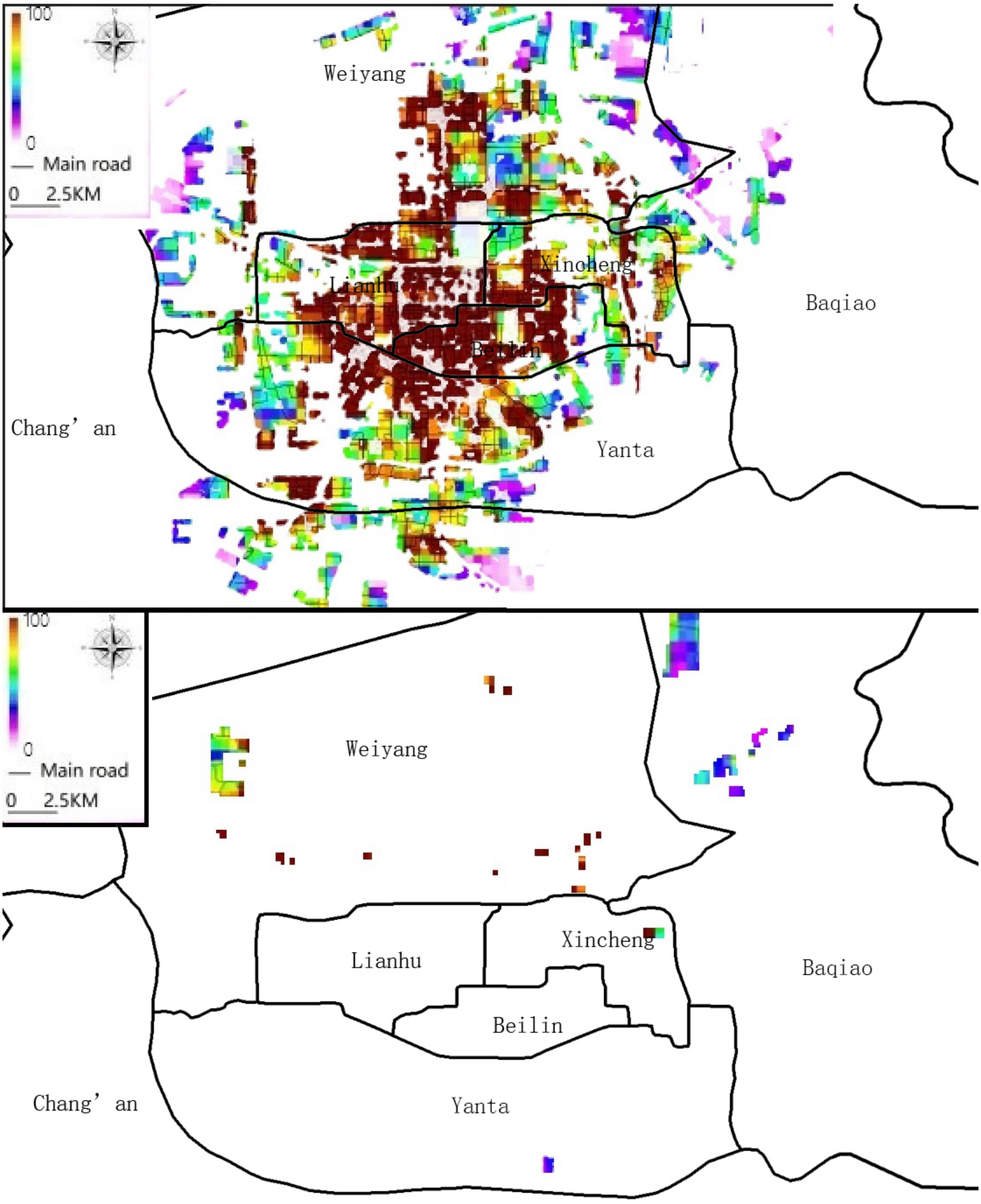

Figure 5.8 The satisfaction parameter S (Section 4.1.2, equation [1]) for commercial communities (a) and Chengzhongcun (b) of Xi'an, 2019

In 2019, Xi'an had 10.2 million residents distributed in thousands of communities and villages. The economic conditions of residents are different. For example, the housing prices in some communities can reach up to 100,000 yuan per square meter (approximately 11,800 pounds per square meter), while the housing prices in other communities are only around 8,000 yuan per square meter (approximately 940 pounds per square meter). It is difficult for me to independently complete the statistics of relevant information such as the average income of residents in all communities. Due to that, the calculation of the transport impedance indicator for each community was not possible in this study.

Therefore, in this case study, it used the official average income and average transportation cost of residents, statistically grouped by administrative districts (counties), for the calculation of transport impedance indicators.

According to (*Statistical Yearbook of Xi'an*, 2020), in 2019, the transportation expenditure of residents in Xi'an City was 2019.7 yuan and substitute this value into the parameter C in the Equation [1]. According to (*Statistical Yearbook of Xi'an*, 2020), in 2019, the annual per capita income of each administrative district and county in Xi'an is shown in Table 5.2, and substitute this value into the parameter I in the Equation [1].Calculate the transport impedance indicator of local Hukou residents for each administrative district and county using the Equation [1]. The result shows in the Figure 5.7a.

In this study, it is assumed that the educational background of non-local Hukou residents is below the bachelor's degree (Chapter 4, Section 4.1.1). Based on the recruitment information posted by companies in Xi'an City on job websites, and through my statistics, this study found that the annual income of workers with an educational background below the bachelor's degree is approximately 42,000 yuan in 2019, and substitute this value into the parameter I in the Equation [1]. Calculate the transport impedance indicator of non-local Hukou residents for each administrative district and county using the Equation [1]. The result shows in the Figure 5.7b. The result shows in the Table 5.3.

Table 5.2 annual per capita income of each administrative district and county in Xi'an

Name	Population2019(Person)	Average Income（yuan)
Beilin District	687700	76117.4
Xincheng District	647200	76348.4
Lianhu District	783300	76348.4
Yanta District	1385300	76825.4
Chang'an District	1064200	70191.4
Weiyang District	827300	72768.4
Yanliang District	304500	76825.4
Gaoling District	367100	66934.4
Baqiao District	703900	68378.4
Lintong District	700400	58950.4
Hu County	564600	52928.4
Lantian County	536100	52033.4
Zhouzhi County	592900	50547.4

Finally, the Transport impedance Disparity Indicator formula (Equation [2]) were used to calculate the result (Figure 5.8).

Table 5. 3 transport impedance indicator of each administrative district and county in Xi'an

Name	Local Hukou residents	Non-Local Hukou residents
Beilin District	0.055	0.089
Xincheng District	0.059	0.104
Lianhu District	0.063	0.121
Yanta District	0.062	0.123
Chang'an District	0.076	0.143
Weiyang District	0.077	0.149
Yanliang District	0.148	0.154
Gaoling District	0.155	0.164
Baqiao District	0.119	0.213
Lintong District	0.142	0.244
Hu County	0.242	0.259
Lantian County	0.251	0.262
Zhouzhi County	0.268	0.274

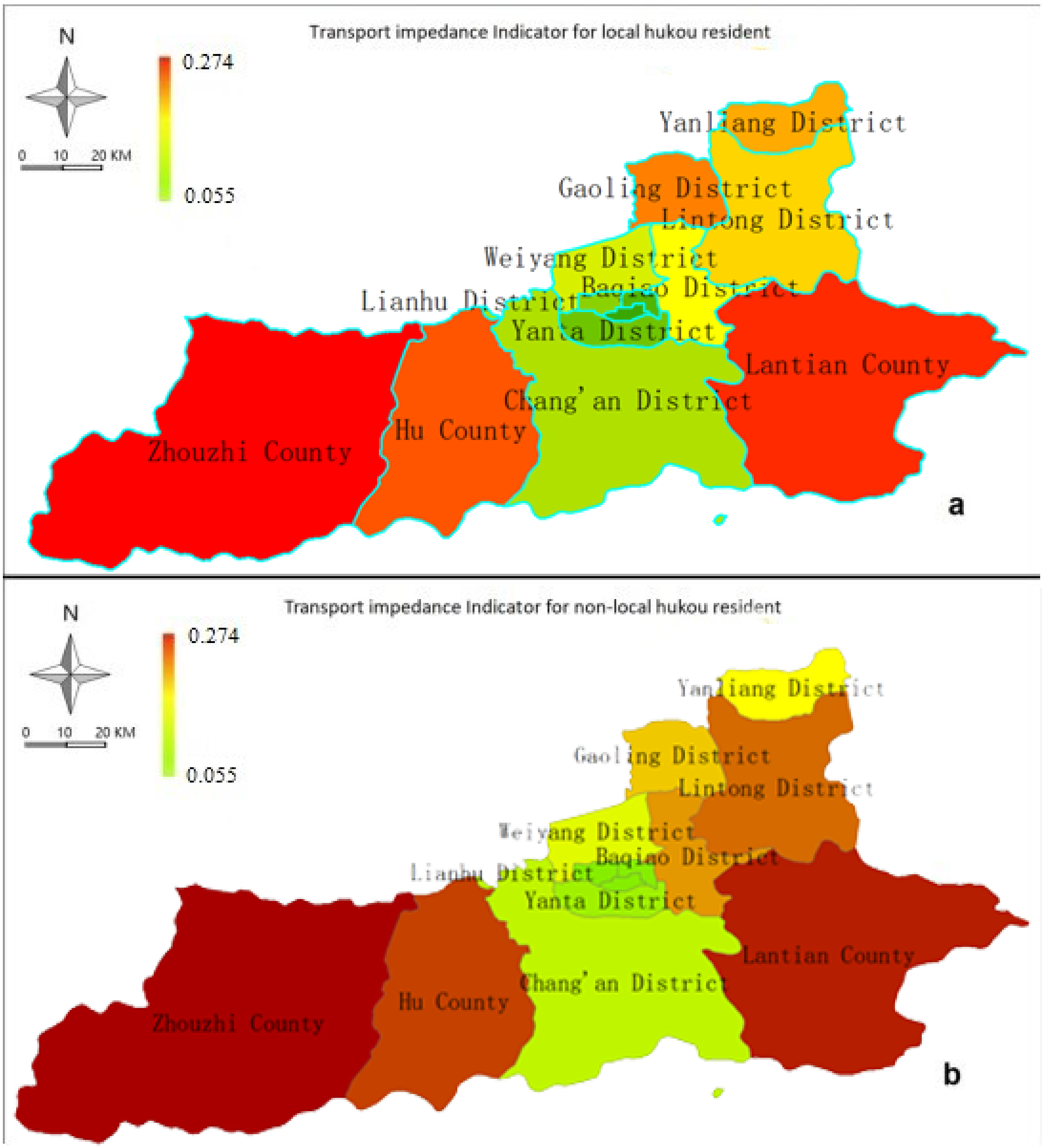

Figure 5.9 Transport impedance indicator of local (a) and non-local (b) Hukou residents in different districts of Xi'an, 2019

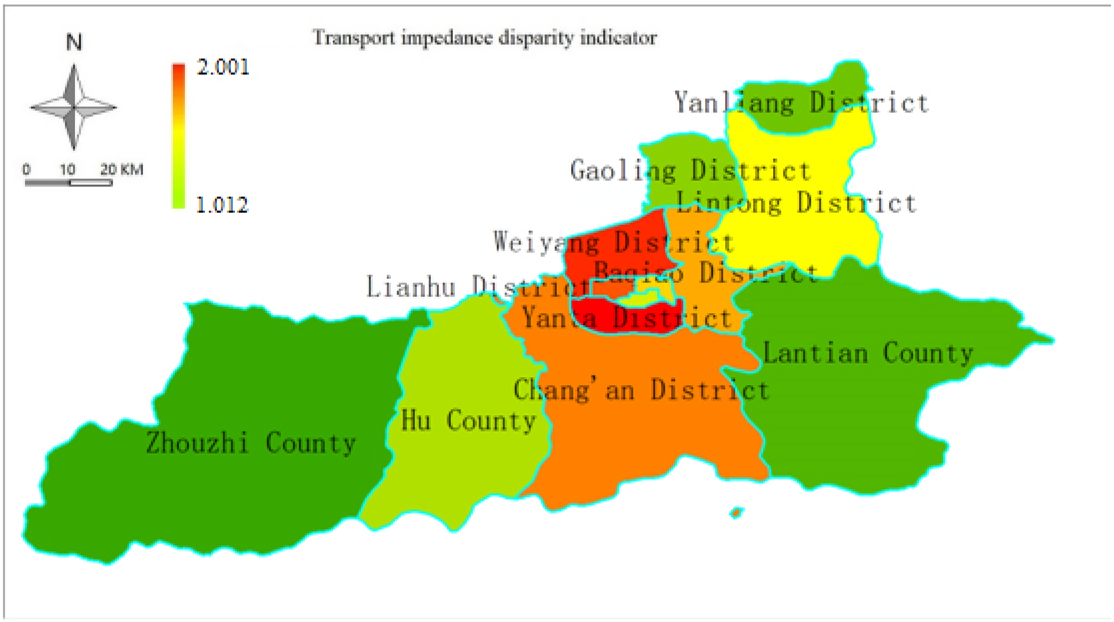

Figure 5.10 Transport impedance disparity indicator of different districts (counties) in Xi'an, 2019

5.3 Findings

By comparing commercial residential communities with Chengzhongcun (Section 4.1.1, assuming that local Hukou residents all reside in commercial residential communities and non-local Hukou residents all reside in Chengzhongcun), it is evident that non-local Hukou residents face disadvantages in Land attraction. Transport impedance is higher in non-local Hukou residents than in local Hukou residents, and it is higher in counties and suburbs than in core areas in Xi'an, indicating a disparity. For example, the urbanization rate of the Beilin district, the core built-up area in Xi'an, is 100%, which is about five times that of Zhouzhi County, where is the furthest place from the core areas. And the transport impedance indicator of local Hukou resident in Beilin is about 0.0556, is about one-fifth of that in Zhouzhi County.

The disadvantages faced by non-local Hukou residents are mainly caused by two factors: first, their income is generally lower than that of local Hukou residents; second, they reside in areas that are mostly on the periphery of the urban built-up area, where there is a relatively small distribution of public resources.

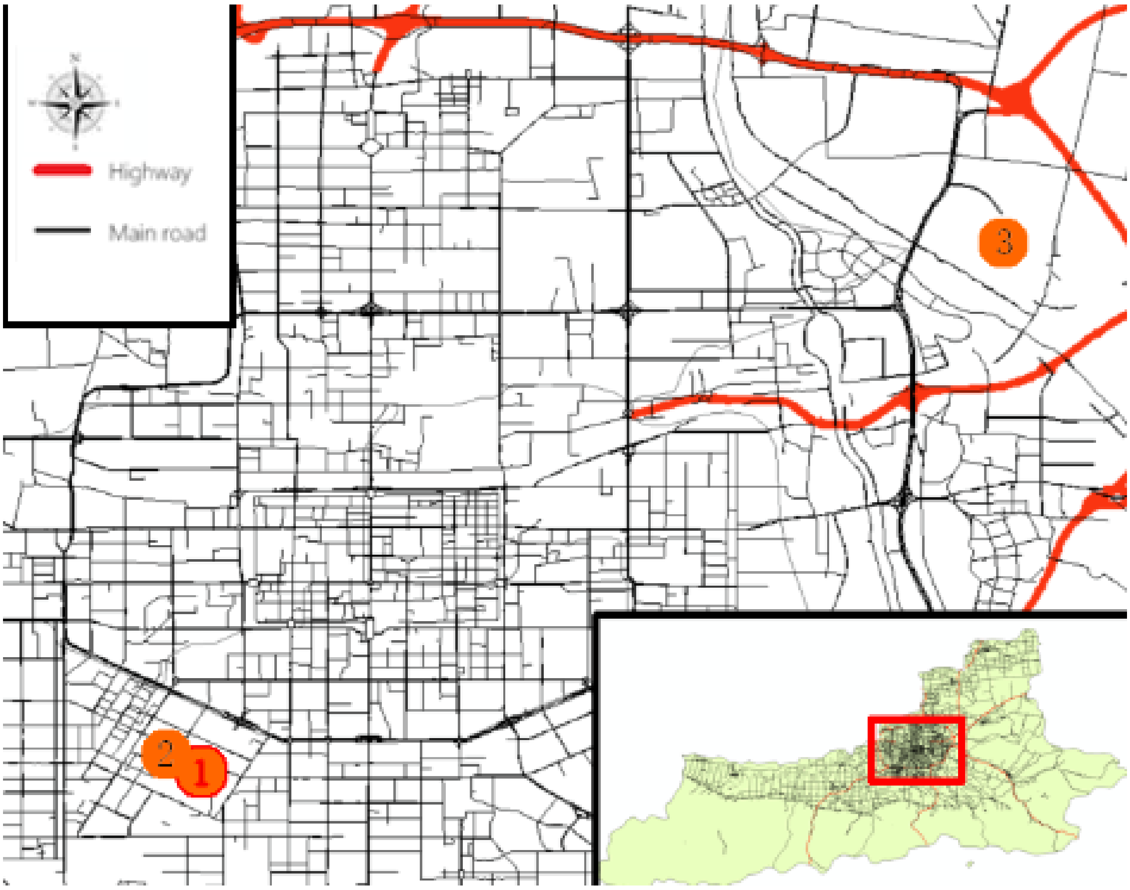

Figure 5.11 The location of Shajing village (Chengzhongcun,①), Zhongtian community (Commercial residential area, ②) and Dushu village (Chengzhongcun, ③)

This disparity is mainly caused by three aspects as following:

- Firstly, urban resources such as public service and economic activities are highly concentrated in the central region. This results in relatively low transport impedance indicators for both local and non-local Hukou residents in the central region (Figure 5.7). For example, as Figure 5.9 shows, Shajing Village (Figure 4.6, Section 4.1.4, Figure 5.9, ①) and Zhongtian Community (Figure 4.6, Section 4.1.4, Figure 5.9, ②) are located in the core area of the city, and both have relatively low transport impedance indicators. The transport impedance indicator of Shajing village is 0.105

and that of Zhongtian community is 0.059. The difference between them is mainly caused by the difference in income.

- Secondly, non-local Hukou residents live in the villages in the city (Chengzhongcun) which are mainly located in urban-rural junctions and decayed urban zones far away from the city's core (Figure 5.5), resulting in decreased Land attraction parameters and increased Transport impedance Indicators. As Figure 5.9 shows, Dushu village (Figure 5.9, ③) is located on the outskirts of the built-up area, and the transport impedance indicator of the village is 0.251.
- Finally, according to Xi'an local government's statistical yearbook 2020, the average income of non-local Hukou residents is lower than that of local Hukou residents, which further increases the Transport impedance disparity Indicator.

5.4 Discussion and Conclusion

5.4.1 Limitation of Transport impedance Indicator and Transport impedance Disparity Indicator

The accuracy of the Transport impedance Indicator and Transport impedance Disparity Indicator largely depends on the accuracy of the input data. For instance, to calculate Land attraction, data such as the number of jobs and salary levels of all working places are required, which may be difficult to collect for an entire city with over 10 million inhabitants. Therefore, in the formula for calculating the traffic impedance indicator (equation [1], Section 4.1.2), for parameter I, this study used the average income of residents as reported by the local government based on administrative divisions. For parameter C, this study utilized the average transportation expenditure of residents also provided by the local government according to administrative divisions. Although this study calculated parameter S for different communities, to ensure consistency in the final computation, this study calculated the average value of parameter S for each administrative division. This affects the accuracy of the Transport impedance Indicator and Transport impedance Disparity Indicator, limiting their application.

5.4.2 Prospect

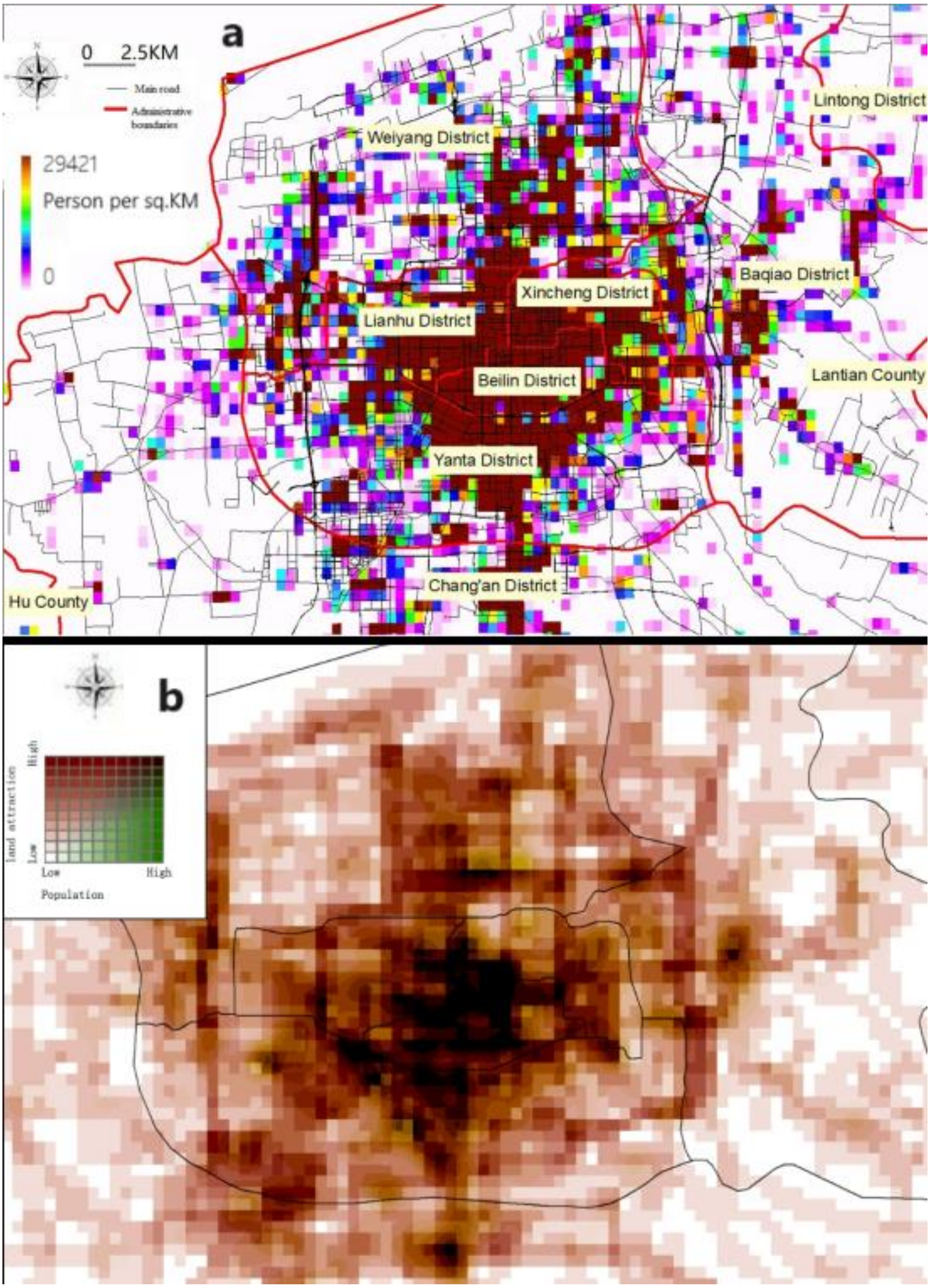

Figure 5. 12 Population distribution (a), and land attraction for local Hukou resident and population distribution (b, bivariate mapping) of Xi'an, 2019

The Transport impedance Indicator formula has good expansibility. For example, the parameter *S* of the spatial analytic model (Section 4.1.3.2) can be regarded as land attractions, which can be adjusted by parameters and weights of the sub-model to calculate attractiveness as a residential location or an activity destination. For example, in this case study, according to the different needs of local and non-local Hukou residents, this study compiled a corresponding weight table (Table 5.1) to calculate the land attraction for local and non-local Hukou residents respectively. If this study conducts statistical analysis on the travel needs of the elderly and develop corresponding weight tables, it can then calculate the land attraction for elders. From the perspective of urban planners, local government can formulate corresponding policies to influence land attraction, thereby achieving their certain goals. For example, if certain jobs (such as civil servants) are only available to residents with local Hukou, this will limit the accessibility of some working places to non-local Hukou residents. Consequently, the distribution of land attraction calculated based on accessibility will undergo corresponding changes. On the other hand, by spatially overlaying land attraction, the distribution of commercial residential areas and Chengzhongcun, corresponding rent fees and housing prices, and combining economic status data for a certain resident group, it is possible to analyse where the land attraction is highest for that resident group within the area where they can afford the housing prices or rent fees.

In this research, the urban resource supply of Transport impedance Indicator was used to calculated attractiveness as a residential location. By comparing the distribution of urban resource supply and population, the potential destinations of population movement can be predicted. For example, Figure 5.10 (a) shows the population distribution of Xi'an in 2019, and the Figure 5.10 (b, bivariate mapping) indicating that in the southwest of the Yanta district, the population is significantly lower than the Land attraction. There are similar situations in Baqiao, Lianhu, Weiyang and Chang'an District. According to the

(*XI'AN STATISTICAL YEARBOOK*, 2020) and (*XI'AN STATISTICAL YEARBOOK*, 2023), from 2019 to 2022, the population of Yanta, Baqiao, Chang'an, Lianhu and Weiyang District increased from 1.2775 million to 2.1093 million, 0.6551 million to 1.0523 million, 1.2618 million to 1.6359 million, 0.7613 million to 1.0341 million and 0.978 million to 1.6078 million respectively. During the same period, the total population of Xi'an increased by 2.3885 million (except the Xixian New Area, a new administrative district that originally belonged to Xianyang, a city nearby Xi'an). The increasing population mainly concentrated in areas where the population is significantly lower than the Land attraction (Figure 5.11). So, if a region has high land attraction but a relatively low population, there is a high probability that its population will grow rapidly in the future. This result indicates the calculation of Land attraction effectively and accurately predicting population movements in Xi'an.

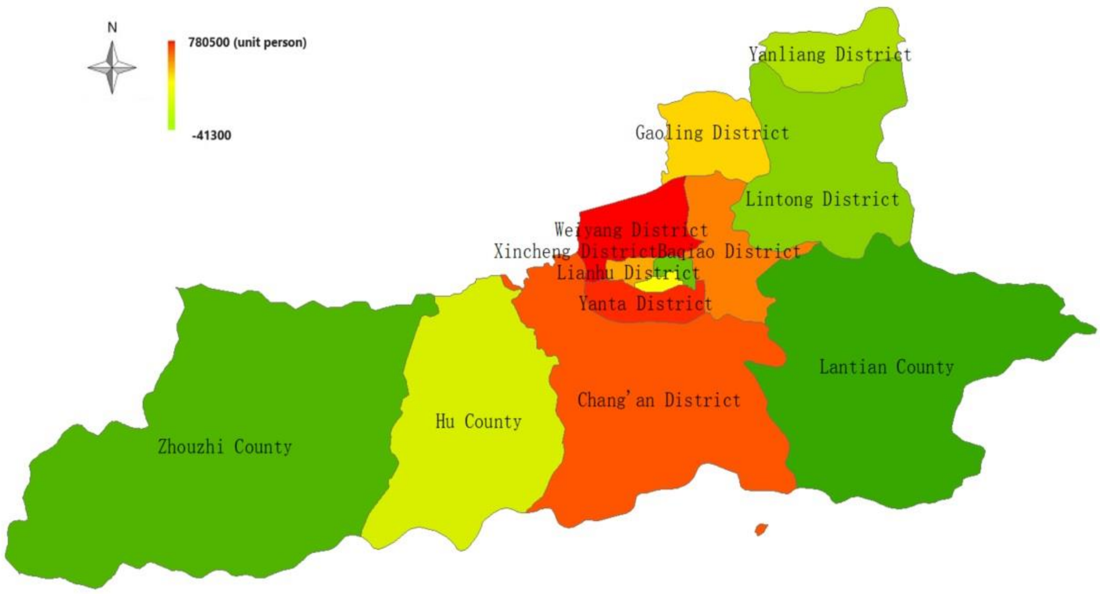

Figure 5. 13 Transport impedance disparity indicator of different districts in Xi'an, 2019

5.4.3 Conclusion

Transport impedance indicator reflects the level of difficulty for a community to meet its traffic demand. Transport impedance disparity indicator is the difference in the transport impedance indicator in different communities. The transport impedance indicator is actually used to evaluate the spatial distribution difference between the distribution of residents and the distribution of land attraction. Residents may not reside in the places where the land attraction value is the highest for them due to certain reasons. According to equation [1] (Section 4.1.2), assuming that both residents' income and transportation costs remain unchanged, the higher the number of residents residing in areas with high land attraction, the larger the value of parameter S, and the smaller the overall transport impedance indicator, and vice versa. Experimental results demonstrated that the Transport impedance indicator and Transport impedance disparity indicator effectively measure the difficulty and differences in different communities.

In summary, the Transport impedance Indicator and Transport impedance Disparity Indicator effectively reveal the difficulty of accessing urban characteristics for different communities in a city and the disparities between them. They can help urban planners identify areas with unequal relationships s and provide decision-making support for the future urban planning.

For example, this case study in Xi'an showed that non-local Hukou residents' communities face significant disadvantages in some areas, particularly on the border of urban built-up expansion. Although the transport impedance indicators for both local and non-local Hukou residents in Yanta, Weiyang, and Baqiao Districts are relatively low, the transport disparity indicators in those districts were over 1.94, indicating a need for local authorities to invest more resources to reduce the transport disparity indicator to reduce the gap of urban characteristics distribution between different Hukou groups. The transport disparity indicators in Zhouzhi, Lantian and Hu Counties, Yanliang and

Gaoling Districts approximately below to 1.06. This because the transport impedance indicator for both local and non-local Hukou residents are all quite high. That indicates a need for local authorities to invest more resources to reduce the transport indicator to reduce the gap of urban characteristics distribution between different districts and counties.

5.5 Summary

This chapter proposes the Transport impedance indicator, which reflects the level of difficulty for a community to meet its traffic demand. The lower the ratio of the average transport costs to average incomes of a community and the more POI factors in the surrounding areas, the lower the transport impedance indicator of the community, and vice versa.

Based on that, according to spatial distribution and travel needs of different resident group, this chapter studied the transport impedance disparity indicators for both local and non-local Hukou residents' communities. Greater transport impedance disparity indicators between different communities in the same city indicate a larger gap in transport impedance.

The indicators identify the accessibility by urban POI, by community/Hukou policy types (local hukou/city residents vs non-local hukou/migrant residents), along with their average transport cost to income ratios.

This can help urban planners identify how severe the unequal relationship is between different areas and provide decision-making support for future urban planning and administrative policies.

Chapter 6

The impact of land attraction on emerging Taobao villages: a case study in China

6.1 Introduction

In the experiment in chapter 5, this thesis found that land attraction has an impact on people's location choice behaviours. To further verify this influence, in this chapter, this study designed experiments to further verify the role of the land attraction in regional development. However, from 2019 to 2021, the number of Taobao villages in Xi'an increased from 0 to 11. This thesis believes that this number is too small to be used for analysing the role of land attraction. Therefore, study decided to analyse the distribution of Taobao villages throughout China from a national perspective.

6.1.1 Introduction of E-commerce

With the development of Information and Communication Technologies (ICT), the digitalization of the economy, also called e-commerce, is rapidly increasing worldwide. The proportion of the digital economy in the global economy continues to rise. For example, according to public news reports by the China Academy of Information and Communications Technology (CAICT) (http://www.caict.ac.cn/english/), from 2019 to 2022, the proportion of the digital economy in China's overall economy was increased from 36.2% to 41.5%. Meanwhile, driven by increased internet retail platforms, between April 2020 and April 2021, the total online retail volume in the U.S. increased by over 30% compared to the previous year (Banerjee et al., 2023).

The impact of E-commerce on global development continues to deepen, which enables regions and related employment opportunities to break location limitations and join national and even global industrial divisions. More and more information and capital flow through the Internet, replacing traditional human travel needs and gradually changing people's travel behaviours. This online shopping model has shifted the traditional face-to-face model of capital flow and information exchange online. It poses significant challenges for related decision makers to plan transport systems more wisely in order to support and adapt to e-commerce activities.

6.1.2 Introduction of the case study

I selected China's Taobao villages (see Section 2.4 for the description of Taobao village) as the experimental subjects to verify whether there is a positive correlation between the distribution of emerging Taobao villages in China from 2019 to 2021 and land attraction for Taobao villagers. In this chapter, this study have calculated the land attraction for Taobao villagers throughout China (main land). Then, this study conducted a regression analysis to investigate the relationship between the number of newly emerged Taobao villages in various regions and the local land attraction values. The result shows that from 2019 to 2021, there is a positive relationship between the land attraction for Taobao villagers and the emerging Taobao villages in China.

6.2 Experimental program

Section 4.2 provides detailed explanations of the methods and data used in this experiment. This section provides explanations of other data used, including the methods (Section 4.2.2) and the data sources (Section 4.2.3).

This case used the same spatial analysis model (Figure 4.2, Section 4.1.3, and Figure 4.9, Section 4.2.2) as Chapter 5, but made adjustments to some key data. On one hand, although people engaged in e-commerce industry in Taobao villages do not need to consider commuting issues, to calculate the land attraction for Taobao villagers in China, it is necessary to calculate the buffer zone radius of different urban facilities. The buffer radius of an urban facility is equal to the average commuting distance of the city multiplied by the weight of such urban facilities within their main aspects (Section 4.1). This case study used average commuting distance data for major cities in China. According to the "*2020 National Commuting Monitoring Report on Major Cities*" issued by the China Academy of Urban Planning and Design, the average commute radius, which is the average straight-line distance between the place where people work and live, in most cities in China was 6 km in 2019 ("Commuter Monitoring report of Major cities in China 2022," 2022). Therefore, the buffer radius of a hospital is the average commuting distance is multiplied by a weight. According to Table 4.4, the weight of the hospital is 0.88. Then, the buffer radius of the hospital is 6 km multiplied by 0.88, resulting in 5.28 km. On the other hand, Based on Lin's research findings (Figure 4.8, Section 4.2.2, (Lin, 2019)), this case study selects the five most significant aspects that influence the spatial location choice behaviour of Taobao villagers. As table 4.4 shown, there are Sanitation, Road, Public transport system, Children's Education conditions and Government support. The weights of these five aspects are determined by their scores in Figure 4.8. Specifically, the weight of a particular aspect is calculated by dividing the sum of the total scores of all five aspects by the score of that particular aspect. The sources for Sanitation, Road, Public transport system (although both the workplace and the living space are at home, those Taobao villagers still need to rely on public transport system to get to other places like hospitals, schools and parks), Children's Education conditions and Government support approximately are 75, 95, 65, 50 and 50. Therefore, the weight of the sanitation is equal

75 divided by the sum of all five scores: 70 / (75+95+65+50+50). The result is approximately 0.22. Thus, the weight of the sanitation is 22%. For the weights of the subcategories within the main aspects, this study has used the same values as those presented in Table 5.1 (Chapter 5), as the discussion in subsections 2.3 and 5.2.2 demonstrates that Xi'an is representative of China. As Figure 4.9 shown (Section 4.2.2), the spatial analysis conducted in this experiment employed the same spatial analysis model (Figure 4.2, Section 4.1.3.2) as in chapter 5 to calculate the land attraction for Taobao village in China, but with relevant parameters modified to be related to Taobao village residents instead of local and non-local household Hukou residents. The results are shown in Figure 6.1.

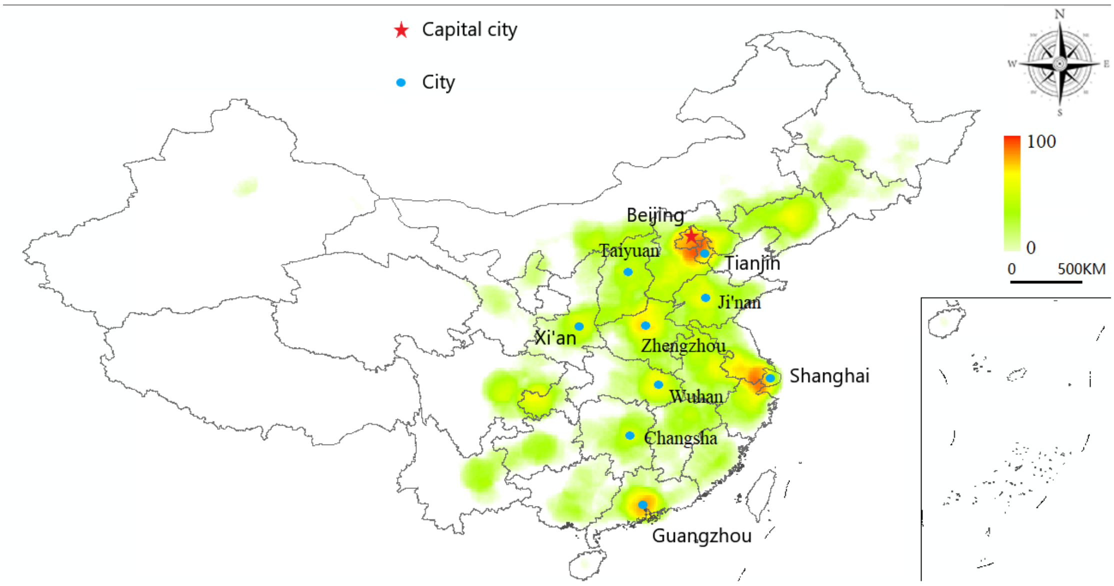

Figure 6.1 Land attraction for Taobao villagers in China, 2019

Figure 6.2 Density of Taobao village in China mainland, 2021(Data source: http://www.aliresearch.com/)

According to Figure 6.1, at the national level, the value of the land attraction for Taobao villagers of Beijing, Shanghai, and Guangzhou are the highest.

6.2.1 Regression analysis

To further verify the correlation between the distribution of emerging Taobao villages from 2019 to 2021 and the land attraction for Taobao villagers, this research selected Beijing, Shanghai, Tianjin, Xi'an, Zhengzhou, Taiyuan, Changsha, Wuhan and Ji'nan as experimental subjects (Table 6.1). These ten cities cover the core region of China from east to west and from south to north. The administrative areas of these cities are relatively similar in size, and the population of each city is about ten million people or more. In terms of political status, they are basically the same, except that Beijing and Shanghai are municipalities directly under the central government, while the others are provincial capital cities. Therefore, these cities are able to represent the overall research subjects.

Regression analysis is a statistical method used in statistics to study the relationship between one or more independent variables and a dependent variable. This study selects the number of emerging Taobao villages from 2019 to 2021 (Y) as the dependent variable and the value of the land attraction for Taobao villagers in 2019 (X) as the independent variable. Through descriptive statistical analysis of the data, it is found that there is a certain correlation between the variables. Based on these variables, a linear regression model is established (formula [5]):

$$Y = \beta 0 + \beta 1X + \varepsilon \quad [5]$$

β0 represents the intercept term, β1 represents the regression coefficient, and ε represents the residual term. SPSS software is used to conduct a regression analysis on the data, and the results are shown in Table 6.2.

Table 6.1 Increase in number of emerging Taobao village from 2019 to 2021

City	2019	2021	The number of emerging Taobao villages	Value of Land attraction for Taobao villagers in China (2019)
Beijing	11	127	116	65.38548
Shanghai	0	78	78	62.11815
Guangzhou	124	273	149	92
Tianjin	14	52	38	32.30908
Xi'an	0	11	11	25.17119
Zhengzhou	19	30	11	21.3856
Taiyuan	0	3	3	10.68374
Changsha	3	10	7	19.70666
Wuhan	0	5	5	18.15338
Ji'nan	1	16	15	26.84386

Table 6.2 The regression analysis results of the correlation between the number of emerging Taobao villages from 2019 to 2021 and the land attraction for Taobao villagers in 2019

Regression Statistics	
Multiple R	0.983703459
R Square	0.967672495
Adjusted R Square	0.963631557
Standard Error	10.04824776
Observations	10

	df	SS	MS	F	Significance F
Regression analysis.	1	24178.36174	24178.36174	239.4672907	3.02581E-07
Residual error.	8	807.7382647	100.9672831		
Total	9	24986.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-29.82811153	5.694599713	-5.237964569	0.000785231	-42.95988201	-16.6963
X Variable 1	1.956567613	0.126436295	15.47473072	3.02581E-07	1.665004995	2.24813

6.2.2 Parameter description and analysis

This subsection will provide a description and explanation of the parameters in Table 6.2, as well as an analysis of the meanings represented by the parameter values.

Regression Statistics:

1. Multiple R: This is the square root of the correlation coefficient R, used to measure the degree of linear correlation between independent variables and the dependent variable. In this case, the Multiple R value for Guangzhou is very high (0.9837), indicating a very strong linear relationship between the independent variable (land attractiveness value) and the dependent variable (possibly the number of Taobao villages).

2. R Square (Coefficient of Determination): It represents the proportion of total variation explained by the model. A high R^2 value (such as 0.9677 for Tianjin) indicates that the model can explain the variation in the dependent variable very well. However, it should be noted that the R^2 value does not directly indicate the accuracy of model predictions, but only explains the extent to which the model explains the variability of the data.
3. Adjusted R Square (Adjusted Coefficient of Determination): This is an adjustment to R^2 that considers the number of independent variables in the model. It is used to assess the actual goodness-of-fit of the model given the number of independent variables. The adjusted R^2 value is still very high (such as 0.9636 for Xi'an), further confirming the goodness-of-fit of the model.
4. Standard Error: This is the average difference between the predicted values of the model and the actual observed values. A smaller standard error (such as 10.05 for Zhengzhou) indicates that the model's predictions are relatively accurate.
5. Observations: This is the number of data points involved in regression analysis. A sufficient number of observations can improve the stability and reliability of the model.
6. Analysis of Variance (ANOVA):
   - df (degrees of freedom): Used to calculate the degrees of freedom for statistical measures.
   - SS (sum of squares): The total variation of the dependent variable.
   - MS (mean square): The average value obtained by dividing SS by the corresponding df.
   - F-value: The ratio of the mean square of the regression to the mean square of the residuals, used to test the overall significance of the model.

- Significance F: The significance level of the F-value, used to determine whether the independent variables have a significant impact on the dependent variable. In this example, the significance level of the F-value is very low (such as 3.0258E-07 for Guangzhou), indicating that the independent variables have a significant impact on the dependent variable.

Coefficient Section

1. Intercept: The predicted value of the dependent variable when the independent variable is 0. In this example, the intercept is -29.83, which may indicate that without land attractiveness value, the predicted number of Taobao villages or related economic indicators is negative. This may not have practical significance in reality but reflects the model's prediction level when the independent variable is 0.
2. X Variable 1 (Independent Variable 1, i.e., land attractiveness value): Its coefficient is 1.96, indicating that for every one-unit increase in land attractiveness value, the predicted number of Taobao villages increases by 1.96 on average. This coefficient is highly significant (with a very small P-value), indicating a positive and significant relationship between land attractiveness value and the number of Taobao villages.

6.2.3 Result analysis

1. **Highly Significant Linear Relationship**: According to the statistical results, there is a very strong linear relationship between the independent variable (Value of Land attraction for Taobao villagers in China (2019)) and the dependent variable (the number of emerging Taobao villages). This may be because areas with high land attractiveness value are more prone to attracting e-commerce entrepreneurs to gather, thereby forming more Taobao villages.

2. **Goodness of Fit of the Model**: Both the determination coefficient (R^2) and the adjusted determination coefficient of the model are high, indicating that the model can well explain the relationship between the independent variable and the dependent variable. However, a high R^2 value does not necessarily mean that the model can accurately predict in all cases, especially for extreme or outlier values.
3. **Limitations of the Data**: Although the model performs well overall, its predictions may not be accurate for cities or regions with low value of Land attraction for Taobao villagers. This may be due to data heterogeneity, limitations in data collection, or limitations in model assumptions. For example, land value may also be a factor affecting the formation of Taobao villages, but since land value data for the whole of China could not be collected, land value was not included in the calculation of Land attraction for Taobao villagers in this experiment. Land attraction for Taobao villagers is a multi-factor system, and in cities or regions with high value of Land attraction for Taobao villagers, the dominance of other factors (such as transportation, healthcare, and children's education) may overshadow the impact of land value on the results. This could be the reason why the model performs well in areas with high value of Land attraction for Taobao villagers. In other situations, when the dominant factors weaken, the role of land value may become more prominent. This could be the reason why the fit may not be as good for lower values of Land attraction for Taobao villagers.

In summary, the regression analysis shows a significant linear relationship between land attractiveness value and the number of Taobao villages, and the model has a high goodness of fit. However, attention should still be paid to the limitations of the data and the assumptions of the model in practical applications.

6.3 Finding

From 2019 to 2021, according to the data from Ali Research, the total number of Taobao villages increased from 4310 to 7023. Among them, Beijing added 116 new Taobao villages, totalling 127 Taobao villages by 2021. Shanghai added 78 new Taobao villages, also totalling 78 by 2021. Guangzhou added 149 new Taobao villages, reaching a total of 273 by 2021. The number of emerging Taobao villages and the total number of Taobao villages in these three cities are the highest in China, as Figure 6.2 shown. This situation is consistent with the characteristics of the distribution of the land attraction for Taobao villagers in Figure 6.1. Meanwhile, the distribution characteristics of Taobao villages' density in 2021 are also very similar to the distribution characteristics of the land attraction for Taobao villagers in 2019. The development of e-commerce and logistics industries in Taobao villages has stimulated the growth of local tertiary industries (e.g., catering, entertainment, education, and finance). This results in the number of villagers in Taobao villages often reaching several thousand to tens of thousands, with the majority of them not engaged in agricultural production. In terms of scale and function, a Taobao village is equivalent to a town in the UK, with local facilities such as bank branches and paid entertainment venues like KTVs and theatres. That creates more jobs in the tertiary industry and triggered a new bottom-up rural urbanization process.

The land attraction for Taobao villagers effectively predicts the distribution of emerging Taobao villages.

According to the above study, the distribution of emerging Taobao villages is concentrated in developed regions extending approximately 200 kilometres inland from the southeastern coastal line of China, and there is a correlation with the land attraction for Taobao villagers. This research suggests that there are three reasons for this phenomenon, as follows:

1. The density of highways, railways and roads is higher in developed areas than that in developing and underdeveloped areas in China, which is more conducive to package delivery. It is beneficial for online sales.
2. Developed areas have a greater number and higher quality of resources, such as education and medical resources. Therefore, more Taobao villages have emerged in developed areas.
3. The economic condition level in developed areas is higher than that in developing and underdeveloped areas, which causes the quality of government support to be better in developed areas than that in underdeveloped areas for encouraging e-commerce.
4. Taobao villages are not all concentrated inside or near first-tier cities. For example, Zhejiang Province had 2,202 Taobao villages in 2021, the highest number among all provincial administrative regions in China. However, Hangzhou (Hangzhou is adjacent to Shanghai and is located approximately 160 kilometres south of Shanghai.), the capital city of Zhejiang Province and a first-tier city, only had 224 Taobao villages in 2021. More Taobao villages are distributed in small and medium-sized cities, county towns, and townships.

6.4 Discussion and Conclusion

6.4.1 Discussion

According to the Chinese government's public news report, in 2021, national online sales were 13088.4 billion yuan, and the parcel delivery industry delivered 108.3 billion packages in China. By encouraging e-commerce activities, such as online retail, job opportunities are increased.

E-commerce changes the distribution of job opportunities. ICT has reorganized the way of economic activities so that the distribution of E-economic participants no longer depends solely on spatial relationships, social network relationships, and ICT infrastructure. In particular, it effectively reduces the shopping and income gap between different areas and creates certain e-commerce job opportunities that are no longer limited by spatial distribution. In Chapter 5, the distribution of workplaces has a significant impact on the residential location decision making of Chinese urban residents, including both local and non-local Hukou residents, and is one of the important parameters for calculating land attraction. Therefore, the E-commerce can distribute job opportunities to areas that would not have existed in traditional economic models, thereby affecting the distribution of the land attraction.

In the last 40 years, a large number of people have entered cities in China. This is because factors that influence land attraction, such as education, public transport, and medical resources, are highly concentrated in urban areas. While the large population brings advantages, such as a labour force and a consumer market, to the city, it also causes problems, such as environmental pollution, overwhelming infrastructure and maintenance costs, overburdened public resources like health care and education, and rising crime rates.

If population resources cannot be used and controlled reasonably, the negative effects on economic development are likely to outweigh the positive effects. The integrated development of population flows, control of the total urban population, and reasonable planning of urban size are issues that city managers must consider in a balanced manner.

Today, the Chinese government is aware of this problem, but at this stage, although the Chinese government has begun to limit the size of super-large and large cities and the area of new urban land, because high-quality public resources are concentrated in developed cities, immigrants still tend to seek to live in developed cities. Therefore, at this stage, restrictions on urban size will lead to the scarcity of land in developed cities, which in turn will push up housing prices, fuel the economic bubble, and fail to solve the problem of over-concentration of population.

In order to solve the problem that the size of the city is too large and affects economic performance, the Chinese government has been trying to guide the flow of the population to make the distribution of the population as reasonable as possible. On one hand, the central government of China has begun to strictly limit the expansion of urban areas in major cities. For example, Xi'an has been required not to increase the area of land used for construction until 2030. On the other hand, local governments have begun to steer a portion of the population from areas with high population density to those with lower population density. For example, Beijing has constructed the Xiong'an New Area in the hope of alleviating the infrastructure overloading issue in the old urban areas by transferring a portion of the population from the old urban areas to the newly established urban areas. In this process, e-commerce and Taobao villages have played significant roles. They alleviate the infrastructure overload in areas with high population density by transferring some employment opportunities and population to areas with lower population density. For example, by supporting the development of e-commerce, especially online sales, in surrounding areas, Shanghai has transferred some retail job opportunities from its core districts to other cities, thereby effectively controlling the population of working-age individuals. With the cooperation of related policies such as household registration policies and land-use policies, Shanghai has successfully controlled its population size over the past decade as Figure 6.3 shown.

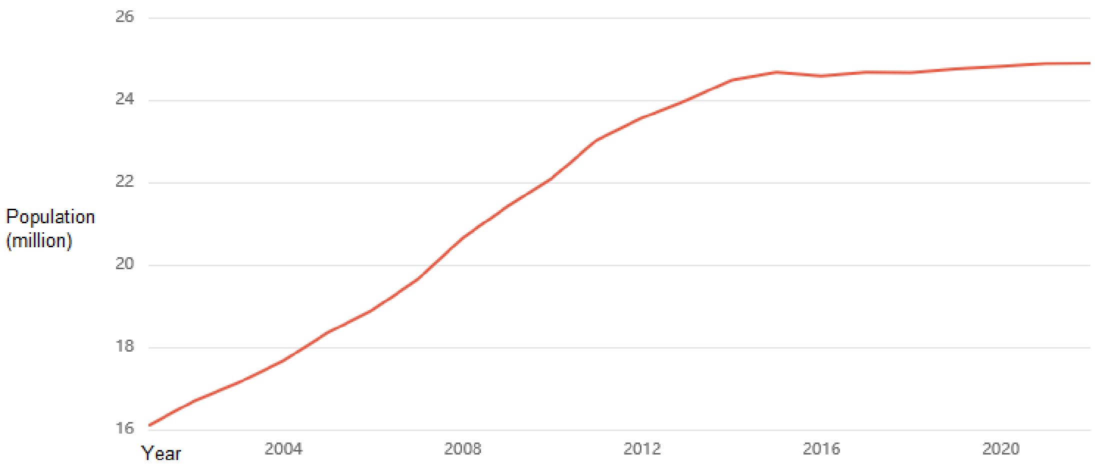

Figure 6. 3 Population size of Shanghai (2000 - 2021)(Unit million)

6.4.2 Conclusion

Based on the above discussion, this research proposes the Four Interwoven Layers (Figure 6.6). The first layer represents the physical world, and the last two layers represent socioeconomic activities. Based on that, this research uses the physical world, which is the same as the first layer of the Three Interwoven Layers (Lin, 2019). Then, the second layer, the Social Network and Relations Layer, only represents logical relationships between socioeconomic participants. According to different organization forms, socioeconomic activities have been divided into a Traditional Spatial Economic Organization Layer, which depends on transport networks, and an E-commerce Organization Form, which depends on ICT infrastructure. It should be noted that there is an intersection between e-commerce organizations and traditional spatial economic organizations. For example, online retail, where information exchanges occur via ICT-based platforms and products are delivered via transport networks.

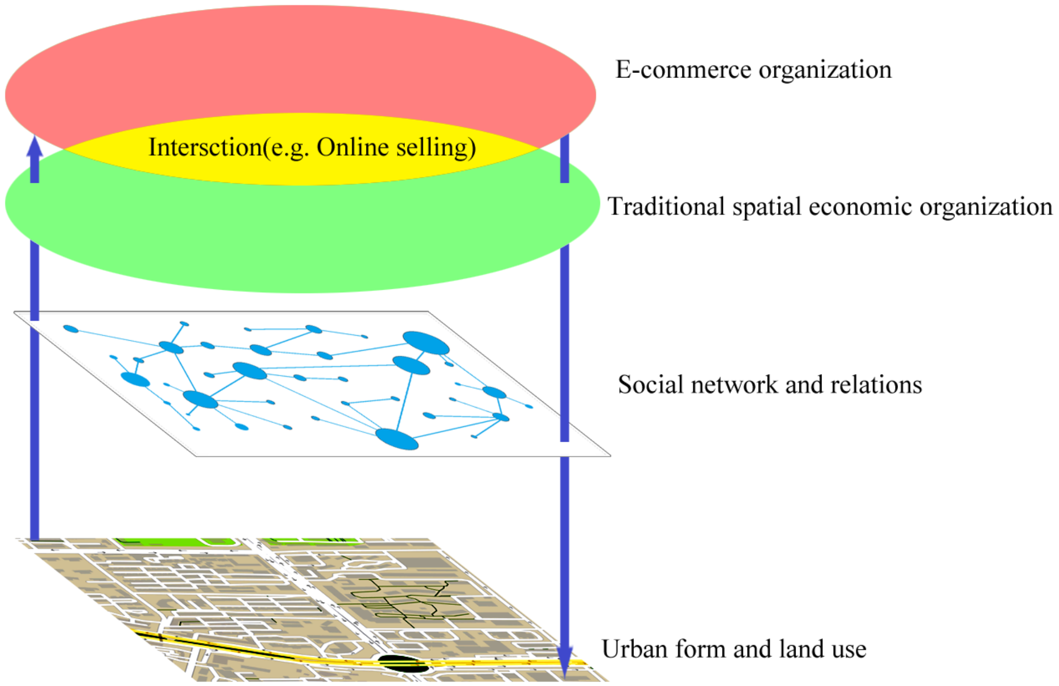

Figure 6. 4 Four interwoven layers

Although e-commerce activities have significantly overcome geographical limitations, allowing individuals to participate regardless of where they choose to settle, the distribution of facilities related to e-commerce activities still influences the behaviour of e-commerce participants in making residential location decisions. For example, Taobao villagers require a developed road network and public transportation system, as well as good medical and educational conditions for their children and themselves. Since developed regions in China possess better and denser road networks, as well as more hospitals and schools, Taobao villages tend to be more concentrated in these developed regions from a national level (Figure 6.2).

Land attraction effectively reveals the relationship between population distribution and urban characteristics. It can help urban planners to provide decision-making support for future planning. For example, by adjusting the distribution of urban characteristics, urban planners can change land attraction to guide the urban population to a certain area. This can avoid or relieve negative effects on urban development caused by the conflict between the distribution of the urban population, urban infrastructures, and urban socioeconomic activities, which cause a series of problems, such as traffic congestion, overwhelmed infrastructure, environmental pollution, deterioration of public health, and other urban, social, environmental, and economic issues.

6.5 Summary

E-commerce has gradually changed the city and triggered a new bottom-up rural urbanization process, especially in China. Therefore, the relationship between people's spatial decision-making behaviours and e-commerce is worth further investigation. To better understand and analyse the impact of e-commerce on population distribution, this thesis, based on a concept framework of E-urbanism, which is defined to describe the impact of e-commerce and migration on urbanism, employs a term, “land attraction”, to represent the ability of an area to provide job opportunities and life services to meet the needs of residents (e.g., people who need to be housed, educated, employed, able to shop, etc.). Based on that, land attraction significantly impacts population distribution. This research did a case study in China to explore the relationship between land attraction and emerging Taobao villages from 2019 to 2021. The results show that land attraction can effectively predict population distribution.

Chapter 7

Conclusion

7.1 Achievement

The aim of this thesis was to propose an urban development theory suitable for China's local contexts. To achieve this aim, this study has chosen the Hukou policy unique to China to study its impact on urban development. Then, this study choses e-commerce, which has flourished in China over the past two decades, to investigate the relationship between economic activities and regional development.

To understand and analyse the impact of Hukou policy on population distribution and human travel behaviours (Section 1.4.1 Research Gap 1 and Section 1.4.2 Objectives 1), this research proposes a transport mobility disparity indicator to describe the gap between travel demand and accessibility for different communities. The indicator identifies the accessibility of urban points of interest (POI) (e.g., distribution of workplace, park, school, hospital, supermarket, bus, and metro stations, etc.) by community/hukou (household registration) policy types (local hukou/city residents vs non-local hukou/migrant residents), along with their average transport cost to income ratios. The lower the ratio of average transport costs to average incomes in a community, and the greater the number of urban facilities in the surrounding area that meet the needs of local residents, the lower the transport impedance indicator of the community will be. Conversely, the higher the ratio and the fewer facilities available, the higher the transport impedance indicator will be. Greater transport impedance disparity indicators between different communities in the same city indicate a larger gap in transport impedance. This can help urban planners identify how severe the unequal relationship is between different areas and provide decision-making support for future urban planning and administrative policies. In this study, Xi'an, capital city of Shaanxi province in China, was chosen as the research object to calculate the transport impedance disparity indicator between local and non-local hukou residents' communities. The result indicates that the transport impedance disparity indicator effectively reveals the different treatments between local and non-local hukou residents' communities in the same city. Compared to local hukou residents, non-local hukou residents experience clear disadvantages in terms of accessibility. Non-local Hukou residents often face restrictions in accessing urban public services. They may not be able to enjoy the same welfare resources as local Hukou residents, such as healthcare, pension, children' education, and housing. This increases their living costs and lowers their quality of life. Based on the experimental results presented in Chapter, the transport impedance indicator of non-local Hukou residents is higher than that of local Hukou residents.

To better understand and analyse the impact of e-commerce on population distribution, describe how existing models could be adapted to consider the impact of e-commerce on the distribution of population and economic activities (Section 1.4.1 Research Gap 1 and Research Gap 2), and integrate the new population distribution sub-model and economic sub-model to establish a conceptual framework of urban development (Section 1.4.2 Objectives 2 and Objective 3), this study takes Taobao villages as the research object and analyses the distribution mechanism of newly established Taobao villages from 2019 to 2021. Based on a concept framework of E-urbanism (Figure 1.4), which is defined to describe the impact of e-commerce and migration on urbanism, this research proposes the Four Interwoven Layers (Figure 6.6). Based on that, the land attraction significantly impacts population distribution. This research did a case study in China to explore the relationship between land attraction and emerging Taobao villages from 2019 to 2021. The result shows that land attraction can effectively predict population distribution.

7.2 Advancements

Land attraction is a brand-new concept that analyses the appeal of a particular region to a specific group of people based on their needs, thus exhibiting strong scalability. In this thesis, Chapter 5 calculates the land attraction for local and non-local Hukou residents, while Chapter 6 computes it for Taobao villagers. In the future, we can also calculate the land attraction for skilled workers, high-level intellectuals, or other specific groups. Another concept proposed in this thesis, the transport impedance indicator, effectively reflects the difference between the spatial distribution of a specific group of people and land attraction for them. This provides information support and decision-making assistance for the fine-grained development of cities.

This thesis innovatively proposes an urban development model, the Four Interwoven Layers (Figure 6.6). It explores and analyses the interaction and spatial organization of economy, population, migration, infrastructure, land-use and transportation, and sets up an empirical model to understand the impact of e-commerce on urban models. It is a tool to adapt to the world's current situation and solve problems effectively.

7.3 Contributions

On one hand, this thesis employs the term "land attraction" to indicate the collective performance of land use, transportation systems, and the spatial distribution and organization of urban facilities, and determines how well that complex system meets the needs of a certain group of residents. These needs include, but are not limited to, housing, education, employment, and shopping opportunities. The distribution of policies and economic activities can affect the land attraction, and the study of the interaction mechanism between them will help people better understand the development patterns of Chinese cities, providing information support for relevant researchers and decision-makers.

On the other hand, E-commerce activities significantly enable regions to break location limitations and join national and even global industrial divisions. This forms a new pattern of urbanization that gradually changes population distribution. The Four Interwoven Layers this thesis proposed provide a viewpoint to study the impact of e-commerce on urban development, which can provide scientific evidence and reference for urban decision-makers. Decision-makers can formulate more precise urban development plans and policies based on the trends and characteristics of e-commerce, in order to promote the sustainable development.

7.4 Limitations

Despite some progress made in this study, there are also some limitations. This research heavily relies on data, and thus the limitations of the data also become the limitations of the study. For instance, with ten million residents in Xi'an, it is difficult for me to obtain accurate information on their addresses, incomes, jobs, expenses, and other details. This leaves me with the option of using statistical averages in my calculations. Although averages can provide valuable information in many cases, they also have some limitations. For example, averages can be easily influenced by extreme values, and they fail to capture the individual characteristics of each data point. Therefore, the accuracy of the calculation results in this study is quite limited.

There are also some limitations in the methodology. There are also some limitations in the methodology. For example, the gravity model adopted in calculating the influence of urban facilities simplifies the assumption that the influence decays linearly, which may affect the accuracy of the results. For example, the gravity model adopted in calculating the influence of urban facilities simplifies the assumption that the influence decays linearly, which may affect the accuracy of the results.

7.5 Further work

Based on this study, the scientific question that this study intend to address in my future research is how to achieve fine-grained urban development in the era of digital economy. Traditional transportation accessibility has, to a significant extent, shaped the urban morphology. However, in the digital economy era, the electronic economy has broken the constraints of geographical regions, thereby exerting a tremendous influence on the distribution of economic activities. This influence, whether direct or indirect, shapes the urban morphology by affecting population distribution and transportation demand, and the underlying mechanisms need to be further explored. To address this issue, building on the new urban development model proposed in this study, it introduces the concept of electronic information accessibility, which will be combined with traditional transportation accessibility to investigate the urban development mechanisms under the new economic paradigm. Subsequently, this thesis propose optimizing the distribution of facilities in large and megacities through digital economy means to address urban issues such as congestion. The interdisciplinary research methods required for future studies, encompassing geography, transportation science, information science, economics, management, and surveying and mapping, are expected to bring about theoretical and practical breakthroughs.

List of References

Acheampong, A.R., Silva, A.E., 2015. Land use–transport interaction modeling: A review of the literature and future research directions. J. Transp. Land Use Vol. 8, 11–38.

Aggarwal, V.K., Morrison, C.E., 1998. Asia-Pacific crossroads: Regime creation and the future of APEC. St. Martin's Press New York.

Aghion, P., Bolton, P., 1997. A theory of trickle-down growth and development. Rev. Econ. Stud. 64, 151–172.

Allen, P.M., Sanglier, M., 1981. Urban evolution, self-organization, and decisionmaking. Environ. Plan. A 13, 167–183.

Allen, P.M., Sanglier, M., 1978. Dynamic models of urban growth. J. Soc. Biol. Struct. 1, 265–280.

Alonso, W., 1964. Location and land use. Toward a general theory of land rent. Locat. Land Use Gen. Theory Land Rent.

Anas, A., Hiramatsu, T., 2012. The effect of the price of gasoline on the urban economy: From route choice to general equilibrium. Transp. Res. Part Policy Pract. 46, 855–873.

Anas, A., Hiramatsu, T., 2011. RELU-TRAN: Applications and challenges. 国際公共政策研究 16, 153–162.

Anas, A., Liu, Y., 2007. A regional economy, land use, and transportation model (relu - tran©): formulation, algorithm design, and testing. J. Reg. Sci. 47, 415–455.

Anttiroiko, A.-V., 2012. Urban Planning 2.0. Int. J. E-Plan. Res. IJEPR 1, 16–30.

Arnott, R., 1979. Optimal city size in a spatial economy. J. Urban Econ. 6, 65–89.

Ashby, W.R., 1991. Principles of the self-organizing system, in: Facets of Systems Science. Springer, pp. 521–536.

Aspinall, R., 2006. Editorial: Land use science. J. Land Use Sci. 1, 1–4.

Batty, M., 2014. Can it happen again? Planning support, Lee's Requiem and the rise of the smart cities movement. Environ. Plan. B Plan. Des. 41, 388–391.

Batty, M., 2012. A generic framework for computational spatial modelling, in: Agent-Based Models of Geographical Systems. Springer, pp. 19–50.

Batty, M., 2009. Urban modeling. International encyclopedia of human geography. Elsevier, Oxford.

Batty, M., 2008. Fifty years of urban modeling: Macro-statics to micro-dynamics, in: The Dynamics of Complex Urban Systems. Springer, pp. 1–20.

Batty, M., 2007a. Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals. The MIT press.

Batty, M., 2007b. Complexity in city systems: Understanding, evolution, and design. Planner's Encount. Complex. Roo G Silva EA Eds 99–122.

Batty, M., 1976a. Urban modelling. Cambridge University Press Cambridge.

Batty, M., 1976b. Urban modelling. Cambridge University Press Cambridge.

Batty, M., Mackie, S., 1972. The calibration of gravity, entropy, and related models of spatial interaction. Environ. Plan. A 4, 205–233.

Boonstra, B., Boelens, L., 2011. Self-organization in urban development: towards a new perspective on spatial planning. Urban Res. Pract. 4, 99–122.

Borukhov, E., 1975. Optimality in city size and systems of cities: a comment. Urban Stud. 12, 325–328.

Brown, D.G., Riolo, R., Robinson, D.T., North, M., Rand, W., 2005. Spatial process and data models: Toward integration of agent-based models and GIS. J. Geogr. Syst. 7, 25–47.

Brown, D.G., Robinson, D.T., 2006. Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. Ecol. Soc. 11.

Brown, D.G., Verburg, P.H., Pontius Jr, R.G., Lange, M.D., 2013. Opportunities to improve impact, integration, and evaluation of land change models. Curr. Opin. Environ. Sustain. 5, 452–457.

Capello, R., Camagni, R., 2000a. Beyond optimal city size: an evaluation of alternative urban growth patterns. Urban Stud. 37, 1479–1496.

Capello, R., Camagni, R., 2000b. Beyond optimal city size: an evaluation of alternative urban growth patterns. Urban Stud. 37, 1479–1496.

Chan, K.W., 2019. China's hukou system at 60: continuity and reform, in: Handbook on Urban Development in China. Edward Elgar Publishing.

Chan, K.W., Liu, T., Yang, Y., 1999. Hukou and non - hukou migrations in China: comparisons and contrasts. Int. J. Popul. Geogr. 5, 425–448.

Chan, K.W., Zhang, L., 1999. The hukou system and rural-urban migration in China: Processes and changes. China Q. 160, 818–855.

Chen, M.X., Lu, D.D., Zhang, H., 2009. Comprehensive evaluation and the driving factors of China's urbanization. Acta Geogr. Sin. 64, 387–398.

Cheng, T., Selden, M., 1994. The origins and social consequences of China's hukou system. China

Q. 644–668.

Christian, C.S., 1957. The concept of land units and land systems.

Chuang-Lin, F., 2009. The urbanization and urban development in China after the reform and opening-up. Econ. Geogr. 29, 19–25.

Chuankai, Y., Ye, L., Wei, X., 2017. The determinants for peasants' migration intentions of moving to cities in China: An analysis based on the CGSS 2010. Geogr. Res. 36, 2369–2382.

Clark, J.R., Kjelgren, R., 1990. Water as a limiting factor in the development of urban trees. J. Arboric. 16, 203–208.

Commuter Monitoring report of Major cities in China 2022, 2022.

Cordera, R., Ibeas, Á., dell'Olio, L., Alonso, B., 2017. Land Use–Transport Interaction Models. CRC press.

Council, N.R., Committee, G.S., 2014. Advancing land change modeling: opportunities and research requirements. National Academies Press.

Cui, H., Cho, J., 2020. Does the Revised Hukou System Facilitate or Restrain the Short-Term Labor Inflows into Chinese Cities? Sustainability 12, 1295.

Cui, M., Levinson, D., 2018. Full cost accessibility. J. Transp. Land Use 11, 661–679.

Currie, G., 2010. Quantifying spatial gaps in public transport supply based on social needs. J. Transp. Geogr. 18, 31–41.

Dameri, R.P., Ricciardi, F., 2017. Leveraging smart city projects for benefitting citizens: the role of ICTs, in: Smart City Networks. Springer, pp. 111–128.

Davidsson, P., Henesey, L., Ramstedt, L., Törnquist, J., Wernstedt, F., 2005. An analysis of agent-based approaches to transport logistics. Transp. Res. Part C Emerg. Technol. 13, 255–271.

El-Geneidy, A.M., Levinson, D.M., 2006. Access to destinations: Development of accessibility measures.

Ernest, B., 1925. Concentric Zone" model of urban structure and land use. Landmark Publ. 125.

Farmer, J.D., Foley, D., 2009. The economy needs agent-based modelling. Nature 460, 685.

Fisch, O., 1975. Optimal city size, the economic theory of clubs and exclusionary zoning. Public Choice 24, 59–70.

Foot, D., 2017. Operational urban models: an introduction. Routledge.

Forrester, J.W., 1958. Industrial Dynamics. A major breakthrough for decision makers. Harv. Bus. Rev. 36, 37–66.

Forrester, J.W., Forrester, J.W., 1969. Urban dynamics (Vol. 114). mIt press Cambridge.

Foth, M., Mitchell, P., Estrada-Grajales, C., 2020. Today's Internet for Tomorrow's Cities: On Algorithmic Culture and Urban Imaginaries. Second Int. Handb. Internet Res. 725–746.

Gao, Q., Yang, S., Zhang, Y., Li, S., 2014. Three Worlds of the Chinese Welfare State: Do Health and Education Change the Picture, in: IARIW 33rd General Conference. Rotterdam. Citeseer, pp. 24–30.

Geurs, K.T., Krizek, K.J., Reggiani, A., 2012. Accessibility analysis and transport planning: challenges for Europe and North America. Edward Elgar Publishing.

Gilbert, N., 2019. Agent-based models. Sage Publications, Incorporated.

Gil-Quijano, J., Louail, T., Hutzler, G., 2010. From biological to urban cells: lessons from three multilevel agent-based models, in: International Conference on Principles and Practice of Multi-Agent Systems. Springer, pp. 620–635.

Glaeser, E.L., Gyourko, J., Saks, R.E., 2005. Why have housing prices gone up? Am. Econ. Rev. 95, 329–333.

Green, D.G., Sadedin, S., 2005. Interactions matter—complexity in landscapes and ecosystems. Ecol. Complex. 2, 117–130.

Greenfield, S., 2004. Tomorrow's people: how 21st-century technology is changing the way we think and feel. Penguin UK.

Gu, C., 2019. Urbanization: Processes and driving forces. Sci. China Earth Sci. 62, 1351–1360.

Guan, X., Wei, H., Lu, S., Dai, Q., Su, H., 2018. Assessment on the urbanization strategy in China: Achievements, challenges and reflections. Habitat Int. 71, 97–109.

Hall, P.G., Pain, K., 2006. The polycentric metropolis: learning from mega-city regions in Europe. Routledge.

Hansen, W.G., 1959a. How accessibility shapes land use. J. Am. Inst. Plann. 25, 73–76.

Hansen, W.G., 1959b. How accessibility shapes land use. J. Am. Inst. Plann. 25, 73–76.

Hong, Y., Fuller, C., 2019. Alone and "left behind": a case study of "left-behind children" in rural China. Cogent Educ. 6, 1654236. https://doi.org/10.1080/2331186X.2019.1654236

Houkai, W., 2014. Regional economic development in China: agglomeration and relocation, in: The Micro-Analysis of Regional Economy in China: A Perspective of Firm Relocation. pp. 1–27.

Hoyt, H., 1939. The structure and growth of residential neighborhoods in American cities. US Government Printing Office.

Hu, S., 2012. Complexities of hukou reform must not deter China from pressing ahead. South China Morning Post May 8th.

Hu, X., Li, H., Bao, X., 2017. Urban population mobility patterns in Spring Festival Transportation:

Insights from Weibo data, in: 2017 International Conference on Service Systems and Service Management. IEEE, pp. 1–6.

Johnston, R.J., 2013. City and society: An outline for urban geography. Routledge.

Juan Lin, Han Li, Mingshui Lin, Chuhai Li, 2021. Rural e-commerce in China: Spatial dynamics of Taobao Villages development in Zhejiang Province. Growth Change 53, 1082–1101. https://doi.org/10.1111/grow.12560

Kennedy, C., Cuddihy, J., Engel - Yan, J., 2007. The changing metabolism of cities. J. Ind. Ecol. 11, 43–59.

Kennedy, C., Pincetl, S., Bunje, P., 2011. The study of urban metabolism and its applications to urban planning and design. Environ. Pollut. 159, 1965–1973.

Kiet, A., 2011. Arab culture and urban form. focus 8, 10.

Lambin, E.F., Geist, H.J., 2008. Land-use and land-cover change: local processes and global impacts. Springer Science & Business Media.

Le Page, C., Bazile, D., Becu, N., Bommel, P., Bousquet, F., Etienne, M., Mathevet, R., Souchere, V., Trébuil, G., Weber, J., 2017. Agent-based modelling and simulation applied to environmental management, in: Simulating Social Complexity. Springer, pp. 569–613.

Li, G., Li, F., 2019. Urban sprawl in China: Differences and socioeconomic drivers. Sci. Total Environ. 673, 367–377.

Li, G., Sun, S., Fang, C., 2018. The varying driving forces of urban expansion in China: Insights from a spatial-temporal analysis. Landsc. Urban Plan. 174, 63–77.

Li, G., Wang, S., Li, J., 2023. The transport impedance disparity indicator: A case study of Xi'an, China. Environ. Sustain. Indic. 100257. https://doi.org/10.1016/j.indic.2023.100257

Li, S.Y., Liu, X.P., Li, X., 2017. Simulation model of land use dynamics and application: Progress and prospects. J. Remote Sens. 21, 329–340.

Li, X., Gong, P., 2016. Urban growth models: progress and perspective. Sci. Bull. 61, 1637–1650.

Li, Y., Chen, M., Dou, Z., Zheng, X., Cheng, Y., Mebarki, A., 2019. A review of cellular automata models for crowd evacuation. Phys. Stat. Mech. Its Appl.

Li, Y., Geertman, S., Hooimeijer, P., Lin, Y., Yang, H., 2021. Do migrants and locals differ in commuting behavior? A case study of Xiamen, China. Transp. Policy.

Liang, X., Liu, Y., Qiu, T., Jing, Y., Fang, F., 2018. The effects of locational factors on the housing prices of residential communities: The case of Ningbo, China. Habitat Int. 81, 1–11.

Lin, Y., 2019. E-urbanism: E-commerce, migration, and the transformation of Taobao villages in urban China. Cities 91, 202–212.

List of Taobao villages, 2021. . aliresearch.

Liu, J., Zhan, J., Deng, X., 2005. Spatio-temporal patterns and driving forces of urban land expansion in China during the economic reform era. AMBIO J. Hum. Environ. 34, 450–456.

Liu, M., Zhang, Q., Gao, S., Huang, J., 2020. The spatial aggregation of rural e-commerce in China: An empirical investigation into Taobao Villages. J. Rural Stud. 80, 403–417. https://doi.org/10.1016/j.jrurstud.2020.10.016

Lucas, K., 2012. Transport and social exclusion: Where are we now? Transp. Policy 20, 105–113.

Ma, J., Cheng, J.C., Jiang, F., Chen, W., Zhang, J., 2020. Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques. Land Use Policy 94, 104537.

Mack, K., McDonnell, E., Jain, D., Lu Wang, L., E. Froehlich, J., Findlater, L., 2021. What do we mean by "accessibility research"? A literature survey of accessibility papers in CHI and ASSETS from 1994 to 2019, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. pp. 1–18.

Makse, H.A., Havlin, S., Stanley, H.E., 1995. Modelling urban growth patterns. Nature 377, 608.

Mann, S., 1984. Urbanization and historical change in China. Mod. China 10, 79–113.

Meyer, W.B., 2000. The other Burgess model. Urban Geogr. 21, 261–270.

Müller, D., Munroe, D.K., 2014. Current and future challenges in land-use science. Taylor & Francis.

Niu, F., Li, J., 2019. An activity-based integrated land-use transport model for urban spatial distribution simulation. Environ. Plan. B Urban Anal. City Sci. 46, 165–178.

Niu, F., Li, J., 2018. Modeling the population and industry distribution impacts of urban land use policies in Beijing. Land Use Policy 70, 347–359.

Niu, F., Wang, F., Chen, M., 2019. Modelling urban spatial impacts of land-use/transport policies. J. Geogr. Sci. 29, 197–212.

Niu, F.-Q., Liu, W.-D., Chen, M.-X., 2018. An Integrated Land Use and Transport Model to Examine Polycentric Policies of Beijing. Int. Rev. Spat. Plan. Sustain. Dev. 6, 1–17.

Osborne, R., 2005. Urban sprawl: what is urbanization and why does it matter?, in: PROCEEDINGS-BRITISH ACADEMY. OXFORD UNIVERSITY PRESS INC., p. 1.

Outline of the Fourteenth five-year Plan for National Economic and Social Development of the people's Republic of China and long-term objectives for 2035, 2021.

Parker, D.C., Manson, S.M., Janssen, M.A., Hoffmann, M.J., Deadman, P., 2003. Multi-agent systems for the simulation of land-use and land-cover change: a review. Ann. Assoc. Am. Geogr. 93, 314–337.

Pfaffenbichler, P., Emberger, G., Shepherd, S., 2008. The integrated dynamic land use and transport model MARS. Netw. Spat. Econ. 8, 183–200.

Rental monitoring report for 40 key cities nationwide, 2021.

Report on the Development of Housing Rental Market in China, 2023.

Rimmer, P.J., 1988. Transport geography. Prog. Hum. Geogr. 12, 270–281.

Rindfuss, R.R., Entwisle, B., Walsh, S.J., An, L., Badenoch, N., Brown, D.G., Deadman, P., Evans, T.P., Fox, J., Geoghegan, J., 2008. Land use change: complexity and comparisons. J. Land Use Sci. 3, 1–10.

Sanders, P., Sanders, F., 2004. Spatial urban dynamics, in: Proceedings of the 2004 International System Dynamics Conference, Oxford, UK.

Servillo, L., Atkinson, R., Russo, A.P., 2012. Territorial attractiveness in EU urban and spatial policy: a critical review and future research agenda. Eur. Urban Reg. Stud. 19(4), 349–365.

Shao, J., Yang, H., Xing, X., Yang, L., 2016. E-commerce and traffic congestion: An economic and policy analysis. Transp. Res. Part B Methodol. 83, 91–103. https://doi.org/10.1016/j.trb.2015.11.003

Shen, J., 2002. A study of the temporary population in Chinese cities. Habitat Int. 26, 363–377.

Shen, L., Huang, Z., Wong, S.W., Liao, S., Lou, Y., 2018. A holistic evaluation of smart city performance in the context of China. J. Clean. Prod. 200, 667–679.

Shepherd, S.P., 2014. A review of system dynamics models applied in transportation. Transp. B Transp. Dyn. 2, 83–105.

Sjoberg, G., 1965. The origin and evolution of cities. Sci. Am. 213, 54–62.

Smith III, A.R., 1971. Cellular automata complexity trade-offs. Inf. Control 18, 466–482.

Statistical Bulletin of the Development of Health Undertakings in Xi'an City, 2020, 2020. . Xian Munic. Health Comm.

Statistical Yearbook of Xi'an, 2020.

Su, S., Liu, Z., Xu, Y., Li, J., Pi, J., Weng, M., 2017. China's megaregion policy: Performance evaluation framework, empirical findings and implications for spatial polycentric governance. Land Use Policy 63, 1–19. https://doi.org/10.1016/j.landusepol.2017.01.014

Sun, H., Liu, Y., Xu, K., 2011. Hollow villages and rural restructuring in major rural regions of China: A case study of Yucheng City, Shandong Province. Chin. Geogr. Sci. 21, 354–363.

Tanis, K., 2018. Regional distribution and location choices of immigrants in Germany. Reg. Stud. Volume 54, 483–494.

Tapscott, D., 1996. The digital economy: Promise and peril in the age of networked intelligence. McGraw-Hill New York.

Valbuena, D., Verburg, P.H., Bregt, A.K., Ligtenberg, A., 2010. An agent-based approach to model land-use change at a regional scale. Landsc. Ecol. 25, 185–199.

Van der Ploeg, J.D., Ye, J., Pan, L., 2014. Peasants, time and the land: The social organization of farming in China. J. Rural Stud. 36, 172–181.

Verburg, P.H., Alexander, P., Evans, T., Magliocca, N.R., Malek, Z., Rounsevell, M.D., van Vliet, J., 2019. Beyond land cover change: towards a new generation of land use models. Curr. Opin. Environ. Sustain. 38, 77–85.

Verburg, P.H., Crossman, N., Ellis, E.C., Heinimann, A., Hostert, P., Mertz, O., Nagendra, H., Sikor, T., Erb, K.-H., Golubiewski, N., 2015. Land system science and sustainable development of the earth system: A global land project perspective. Anthropocene 12, 29–

41.

Verburg, P.H., Schot, P.P., Dijst, M.J., Veldkamp, A., 2004. Land use change modelling: current practice and research priorities. GeoJournal 61, 309–324.

Wan, L., Yun, Y., Ni, J., 2017. The geo-space of Taobao village in China: theoretical framework based on space of flows. City Plan. Rev. 41, 27–34.

Wang, S.-Y., 2011. State misallocation and housing prices: theory and evidence from China. Am. Econ. Rev. 101, 2081–2107.

Wang, Y., Peng, Z., Chen, Q., 2018. The choice of residential layout in urban China: A comparison of transportation and land use in Changsha (China) and Leeds (UK). Habitat Int. 75, 50–58.

Wegener, M., 2004. Overview of land-use transport models. Handb. Transp. Geogr. Spat. Syst. 5, 127–146.

Wei, H., Sheng, G., 2020. Rural-Urban Integration and Household Registration System Reform, in: Comprehensive Deepening of Reforms in China. Springer, pp. 107–130.

Wieand, K.F., 1987. An extension of the monocentric urban spatial equilibrium model to a multicenter setting: The case of the two-center city. J. Urban Econ. 21, 259–271.

Wolman, A., 1965. The metabolism of cities. Sci. Am. 213, 178–193.

Wu, X., Treiman, D.J., 2004. The household registration system and social stratification in China: 1955–1996. Demography 41, 363–384.

XI'AN STATISTICAL YEARBOOK, 2023.

XI'AN STATISTICAL YEARBOOK, 2020.

Xiaojun, Y., 2017. The impact of household registration system reform on population immigration of big cities in China: An empirical study based on urban panel data from 2000 to 2014. Popul.

Res. 41, 98–112.

Yang, X.J., 2013. China's rapid urbanization. Science 342, 310–310.

Yang, Z., 2020. Development of Optimal City Size Theory: A Critical View. J. Resour. Ecol. 11, 100–110.

Yehua Dennis Wei, Juan Lin, Ling Zhang, 2019. E-Commerce, Taobao Villages and Regional Development in China. Geogr. Rev. https://doi.org/10.1111/gere.12367

Zang, Y., Hu, S., Zhou, B., Lv, L., Sui, X., 2023. Entrepreneurship and the formation mechanism of Taobao Villages: Implications for sustainable development in rural areas. J. Rural Stud. 100, 103030. https://doi.org/10.1016/j.jrurstud.2023.103030

Zhang, J., Wang, R., Lu, C., 2019. A quantitative analysis of Hukou reform in Chinese cities: 2000–2016. Growth Change 50, 201–221.

Zhang, Z., Wu, X., 2017. Occupational segregation and earnings inequality: Rural migrants and local workers in urban China. Soc. Sci. Res. 61, 57–74.

Zhao, C., Zhou, X., Wang, F., Jiang, M., Hesketh, T., 2017. Care for left-behind children in rural China: A realist evaluation of a community-based intervention. Child. Youth Serv. Rev. 82, 239–245. https://doi.org/10.1016/j.childyouth.2017.09.034

Zhendong, L., Heming, H., Caige, L., 2017. A New Bottom-Up Process: The Rural Urbanization Driven by E-Commerce in China. China City Plan. Rev. 26.

万励 (Wan Li), 金鹰 (Jin Ying), 2014. 国外应用城市模型发展回顾与新型空间政策模型综述 (Review on Applied Urban Modeling and New Trends of Urban Spatial Plocy Models). 城市规划学刊 (Urban Planning Forum) 1, 81–91.

周丽萍(Zhou Liping), 2008. 商品住宅特征价格模型与指数的应用研究(Application Reasearch of Merchandise Housing HedonicPrice Model and IndexSpecialty). 财经理论 与实践.

宁吉喆, 2021. 第七次全国人口普查主要数据情况. 中国统计 (Main Data of the Seventh National Population Census. China Statistics) 36.

戴尔阜 (Dai Erbu), 马良 (Ma Liang), 杨微石 (Yang Weishi), 王亚慧 (Wang Yahui), 尹乐 (Yin Le), 童苗 (Tong Miao), 2019. 土地系统多主体模型的理论与应用 (Agent based model of land system: Theory, application and modelling framework). 地理学报 (Acta Geographica Sinica) 74, 2260–2272.

李冠元(Li Guangyuan), 2018. 西安市城中村用地去留问题探析(Analysis on the problem of land use in village in the city of Xi'an). 小城镇建设 11.

汪丽(Wang Li), 李九全(Li Jiuquan), 2014. 西安城中村改造中流动人口的空间剥夺(The Spatial Deprivation of the Floating Population in the Transformation of Urban Villages in Xi'an City: Analysis Based on Web Text). 地域研究与开发 33.

牛方曲 (Niu Fangqu), 2017. LUTI 模型的概念结构, 实现方法及发展趋势 (Overview of Urban Land-use/Transport Interaction Model: Origin, Techniques and Future). 地理科学 (Scientia Geographica Sinica) 37, 46–54.

牛方曲 (Niu Fangqu), 刘卫东 (Liu Weidong), 宋涛 (Song Tao), 2014. LUTI 模型原理, 实现及应用综述 (THEORY IMPLEMENTATION AND APPLICATIOINS OF LUTI MODEL). 人文地理 (Human Geography) 29, 31-35,118.

简新华 (Jian Xinhua), 黄锟 (Huang Kun), 2007. 中国农民工最新情况调查报告. 中國人口· 資源與環境 (Latest Survey Report on Chinese Migrant Workers. Chinese Population, Resources and Environment) 17, 1–6.

薛东前 (Xue Qiandong), 贾金慧 (Jia Jinhui), 罗正文 (Luo Zhengwen), 马蓓蓓 (Ma Peipei),

2017. 西安城市贫困阶层的空间聚居研究. 干旱区资源与环境 (A Study on the Spatial Aggregation of Urban Poor Classes in Xi'an. Journal of Arid Land Resources and Environment) 31, 45–50.

Appendix A

```
#coding:utf-8
import re
import urllib2
import chardet
from BeautifulSoup import BeautifulSoup

def remove_js_css (content):
    """ remove the the javascript and the stylesheet and the comment content (<script>....</script>
and <style>....</style> <!-- xxx -->) """
    r = re.compile(r'''<script.*?</script>''',re.I|re.M|re.S)
    s = r.sub ('',content)
    r = re.compile(r'''<style.*?</style>''',re.I|re.M|re.S)
    s = r.sub ('', s)
    r = re.compile(r'''<!--.*?-->''', re.I|re.M|re.S)
    s = r.sub('',s)
    r = re.compile(r'''<meta.*?>''', re.I|re.M|re.S)
    s = r.sub('',s)
    r = re.compile(r'''<ins.*?</ins>''', re.I|re.M|re.S)
    s = r.sub('',s)
    return s

def remove_empty_line (content):
    """remove multi space """
    r = re.compile(r'''^\s+$''', re.M|re.S)
    s = r.sub ('', content)
    r = re.compile(r'''\n+''',re.M|re.S)
    s = r.sub('\n',s)
    return s

def remove_any_tag (s):
    s = re.sub(r'''<[^>]+>''','',s)
    return s.strip()

def remove_any_tag_but_a (s):
    text = re.findall (r'''<a[^r][^>]*>(.*?)</a>''',s,re.I|re.S|re.S)
    text_b = remove_any_tag (s)
    return len(''.join(text)),len(text_b)

def remove_image (s,n=50):
    image = 'a' * n
```

```
    r = re.compile (r'''<img.*?>''',re.I|re.M|re.S)
    s = r.sub(image,s)
    return s

def remove_video (s,n=1000):
    video = 'a' * n
    r = re.compile (r'''<embed.*?>''',re.I|re.M|re.S)
    s = r.sub(video,s)
    return s

def sum_max (values):
    cur_max = values[0]
    glo_max = -999999
    left,right = 0,0
    for index,value in enumerate (values):
        cur_max += value
        if (cur_max > glo_max) :
            glo_max = cur_max
            right = index
        elif (cur_max < 0):
            cur_max = 0

    for i in range(right, -1, -1):
        glo_max -= values[i]
        if abs(glo_max < 0.00001):
            left = i
            break
    return left,right+1

def method_1 (content, k=1):
    if not content:
        return None,None,None,None
    tmp = content.split('\n')
    group_value = []
    for i in range(0,len(tmp),k):
        group = '\n'.join(tmp[i:i+k])
        group = remove_image (group)
        group = remove_video (group)
        text_a,text_b= remove_any_tag_but_a (group)
        temp = (text_b - text_a) - 8
        group_value.append (temp)
    left,right = sum_max (group_value)
    return left,right, len('\n'.join(tmp[:left])), len ('\n'.join(tmp[:right]))
```

```
def extract (content):
    content = remove_empty_line(remove_js_css(content))
    left,right,x,y = method_1 (content)
    return '\n'.join(content.split('\n')[left:right])

def extract_news_content(web_url,file_name):
    request = urllib2.Request(web_url)


    request.add_header('User-Agent','Mozilla/5.0 (Windows; U; Windows NT 6.1; en-US;
rv:1.9.1.6) Gecko/20091201 Firefox/3.5.6')
    opener = urllib2.build_opener()
    html= opener.open(request).read()
    infoencode = chardet.detect(html)['encoding']#
    if html!=None and infoencode!=None:
        html = html.decode(infoencode,'ignore')
        soup=BeautifulSoup(html)
        content=soup.renderContents()
        content_text=extract(content)
        content_text= re.sub(" "," ",content_text)
        content_text= re.sub(">","",content_text)
        content_text= re.sub(""",'"',content_text)
        content_text= re.sub("<[^>]+>","",content_text)
        content_text=re.sub("\n","",content_text)
        file = open(file_name,'a')#append
        file.write(content_text)
        file.close()


key_word =input(“Please, input the keyword”)

def search(key_word):

search_url='http://news.baidu.com/ns?word=key_word&tn=news&from=news&cl=2&rn=20&ct=
1'
    Myrequest=urllib2.urlopen(search_url.replace('key_word',key_word))
    real_visited=0
    for count in range(30):#Extract information from the first 30 pages of the search engine
        html     =   Myrequest.read()
        soup     =   BeautifulSoup(html)
        content  =   soup.findAll("li", {"class": "result"}) #resultset object
        num      =   len(content)
        for i in range(num):
            p_str= content[i].find('a') #if no result then nontype object
```

```
            contenttitle=p_str.renderContents()
            contenttitle=contenttitle.decode('utf-8', 'ignore')#need it
            contenttitle= re.sub("<[^>]+>","",contenttitle)
            contentlink=str(p_str.get("href"))
            visited_url=open(r'E:\Python27\ NewsBaidu.txt','r')#Is it already accessed?
            visited_url_list=visited_url.readlines()
            visited_url.close()
            exist=0
            for item in visited_url_list:
                if contentlink==item:
                    exist=1
            if exist!=1:
                p_str2= content[i].find('p').renderContents()
                contentauthor=p_str2[:p_str2.find("    ")]
                contentauthor=contentauthor.decode('utf-8', 'ignore')
                contenttime=p_str2[p_str2.find("    ")+len("    ")+ 1:]
                contenttime=contenttime.decode('utf-8', 'ignore')
                real_visited+=1
                file_name=r"D:\Python27\newscn\%d.txt"%(real_visited)
                file = open(file_name,'w')
                file.write(contenttitle.encode('utf-8'))
                file.write(u'\n')
                file.write(contentauthor.encode('utf-8'))
                file.write(u'\n')
                file.write(contenttime.encode('utf-8'))
                file.write(u'\n'+contentlink+u'\n')
                file.close()
                extract_news_content(contentlink,file_name)
                visited_url_list.append(contentlink)
                visited_url=open(r'E:\Python27\NewsBaidu.txt','a')
                visited_url.write(contentlink+u'\n')
                visited_url.close()
            if len(visited_url_list)>=120:
                break
        if len(visited_url_list)>=120:
            break
        if count==0:
            next_num=0
        else:
            next_num=1
        next_page='http://news.baidu.com'+soup('a',{'href':True,'class':'n'})[next_num]['href'] #
search for the next page
        print next_page
        Myrequest=urllib2.urlopen(next_page)
```

```
if __name__=='__main__':
    key_word=raw_input('input key word:')
    search(key_word)
```