Complexity

Flow Space and the Complexity of Urban Spatial Networks

Lead Guest Editor: Zhipeng Tang
Guest Editors: Fangqu Niu, Haitao Ma, Jun Li, and MingXing Chen

Flow Space and the Complexity of Urban Spatial Networks

Flow Space and the Complexity of Urban Spatial Networks

Lead Guest Editor: Zhipeng Tang
Guest Editors: Fangqu Niu, Haitao Ma, Jun Li, and MingXing Chen

Contents

Hindawi
Complexity
Volume 2022, Article ID 2291508, 19 pages
https://doi.org/10.1155/2022/2291508

Review Article

State of the Art on Artificial Intelligence in Land Use Simulation

M. Luz Castro[1,2] **Penousal Machado**[3] **Iria Santos**[1,4] **Nereida Rodriguez-Fernandez**[1,2] **Alvaro Torrente-Patiño**[1,4] **and Adrian Carballal**[1,4]

[1] *CITIC-Research Center of Information and Communication Technologies, University of A Coruña, A Coruña, Spain*
[2] *Department of Computer Science and Information Technologies, Faculty of Communication Science, University of A Coruña, A Coruña, Spain*
[3] *CISUC, Department of Informatics Engineering University of Coimbra, Coimbra, Portugal*
[4] *Department of Computer Science and Information Technologies, Faculty of Computer Science, University of A Coruña, A Coruña, Spain*

Correspondence should be addressed to Iria Santos; iria.santos@udc.es

Received 1 November 2021; Revised 28 April 2022; Accepted 16 May 2022; Published 26 June 2022

Academic Editor: Haitao Ma

This review presents a state of the art in artificial intelligence applied to urban planning and particularly to land-use predictions. In this review, different articles after the year 2016 are analyzed mostly focusing on those that are not mentioned in earlier publications. Most of the articles analyzed used a combination of Markov chains and cellular automata to predict the growth of urban areas and metropolitan regions. We noticed that most of these simulations were applied in various areas of China. An analysis of the publication of articles in the area over time is included.

1. Introduction

As pointed out by Levy in his urban planning manual [1], planning is an activity that is highly political and inseparable from the legislation. Regional and urban planning decisions often require copious amounts of public and private money, which can bring great benefits to some and losses to others. The complexity of modern communities makes the simplest and most direct approaches inadequate. In recent decades, several modeling techniques have been developed to understand and predict urban growth. In this day and age, much of the planning efforts and modeling are focused on environmental issues to manage development by minimizing environmental damage.

As mentioned earlier, urban growth is a complex process [2], and different approaches have been used to model it. AI-based methods have the advantage of being able to capture the nonlinearities and heterogeneities that exist in urban development. However, a method's superiority depends, among other factors, on how the algorithm's configuration parameters are determined, the size of the teaching and check sets, the architecture of the classifiers, or the selection of the teaching and verification datasets.

Examples of approaches using Markov chain models [3–5], spatial logistics regression [6], cellular automaton (CA) [7–22], agent-based models [23, 24], and artificial intelligence (AI) and machine learning methods such as artificial neural networks (ANNs) [25–28], support-vector machines (SVMs) [29–31], genetic algorithms (GAs) [32–35], particle swarm optimization (PSO) rules [36–38], or data mining [39] can be found in the scientific literature.

The CA is one of the approaches that is most widely used. The CA's ability to replicate urban growth assumes that past urbanism will influence potential trends through interactions with local land use. Its simplicity, flexibility, and intuition, along with its ability to integrate the spatiotemporal dimensions of the mechanism, make CA models competent for sculpting complex dynamic systems like metropolitan systems.

This review article addresses the different applications of AI in urban planning from 2016 to 2020, and particularly, the latest AI techniques for land-use prediction are organized in detail and summarized. This state of the art follows a comprehensive search methodology. We use keywords such as artificial intelligence, urbanism, or urban growth, for example, to search through Google Scholar and ResearchGate. We review the articles found and all their citations, and we collect all relevant information in this study.

We begin by briefly summarizing the previous state of the art on the subject and then delving into the different subsequent articles, which were not included in any of the previous reviews. Before starting this in-depth study, we can find in Table 1 a summary of the different topics covered in the articles mentioned throughout our text. Finally, an objective analysis of the results is reached, which perfectly summarizes the most important aspects that can be observed during the reading, and other relevant aspects. The conclusion highlights the issues that remain unsolved thus far and explains the AI applications that need to be addressed in future research.

2. Literature Review

In 2010, 33 of these models applied to true-world situations were researched by Santé, García, Miranda, and Crecente [63], to show an ordered picture that would facilitate the selection of a certain procedure for a given implementation problem. The authors pointed out that the main flaw is the relative simplicity of CA-based models, while flexibility conditions the model's right to replace events in reality, leading to certain regulatory measures categorized as follows:

(i) Irregular cell space: such as Iovine et al. [64] with their hexagonal cells or Semboloni [65] whose three-dimensional matrices allow to represent the growth of urban areas around their height, and using irregular tessellations such as Voronoi polygons [66] or graphs [67], or cadastral plots instead of regular cells [68], which can provide a representation that is more adjusted to reality.

(ii) Nonuniform cell space: considering other attributes of the land, and examples include accessibility, slope, or elevation.

(iii) Extended neighborhood: districts can be described as adjacent units belonging to spaces made up of irregular units. For example, units that are at a certain distance or employing the Voronoi spatial model [66].

(iv) Nonstationary district: different definitions of slum space for each cell [69]. Models in which each cell is weighted within the neighborhood according to its state and location [70]. This allows the application of districts of various shapes and sizes by adding zero weights.

(v) Transition rules with greater complexity: it may include suitability for land use, accessibility, urban planning, or socioeconomic conditions, reflecting urban theories that are based on theories of microeconomic planning [71], centrality, and potential models [72].

(vi) Transition guidelines (nonstationary): the transition guidelines are adapted to the distinctive properties of the specific periods and areas. This spatiotemporal modification can be obtained via calibration [73, 74], SLEUTH self-modification rules [75], or changes in external parameters and configuration at each stage as suggested by Phipps and Langlois [76].

(vii) Growth constraints: urban demand is often defined by economic, environmental, or social limitations, like urban planning or demographic change, which restrict overall urban development.

(viii) Irregular steps in time: where cells can be aided by different durations [68] or variable steps in time [77] to mimic different durations of specific events.

Other challenges cited for the urban CA models are the need for data and the lack of easily configurable and useable software. Uses of urban CA models were mostly limited to academic exercises because of these shortcomings. The authors concluded that implementing CA models with different techniques, for example, transport models or multiagent systems, could lead to mixed models overcoming some CA difficulties [78].

In 2016, a new literature review on urban cellular automaton (CA) models was published by Aburas et al. [79]. The authors analyzed the data and the most important factors used in CA studies to simulate and predict urban growth patterns and concluded that the model limitations, such as its inability to include the driving forces of urban growth in the simulation process [80] and its implementation with other quantitative models, can be minimized; for example, through the analytical hierarchy process (AHP) [81, 82], the Markov chain models [83, 84], and frequency relationship [85], realistic simulation is achieved when socioeconomic and spatial and temporal factors are integrated into the simulation process.

2.1. Latest Articles (2016–2020). The CA is a discrete cell in which states characterize each cell. The state of each cell depends on its previous state and on neighboring cells according to a set of transition rules. For urban simulation, the different types of the possible cellular automata are as follows: binary values (urban or nonurban); qualitative values representing different land uses; quantitative values representing the degree of development, population density, value of buildings, elevation, or number of roads, for example; or a vector model with several attributes [63].

We will focus our study on the direction taken by land-use prediction research conducted in recent years, as summarized in Table 1, where 19 of the 24 studies have employed some CAs.

Tong published a model in 2016 [40] describing the pattern of distribution of building construction in green spaces that complied with Chinese regulations. The author presented the distribution index (DAI) based on the

Table 1: Summary table of research cited.

Author/s	Year	Objective	Method	Location	Results
Tong [40]	2016	Planning and management of building layout in green zones	GA	Yuhuatai and Qingliangshan Parks in Nanjing (China)	IOD is the only criterion used for assessing results
Naghibi et al. [41]	2016	Predicting urban growth from satellite images	CA + artificial bee colony	Urmia (Iran)	Overall accuracy: 90.1%
Feng et al. [31]	2016	Predicting urban growth from satellite images	CA + LS-SVM	Shanghai Qingpu-Songjiang (China)	Maximum accuracy of 81.2% in the 16th iteration
Perez-Molina et al. [42]	2017	Simulation of urban growth scenarios and their consequent flooding	CA	Kampala (Uganda)	Overall accuracy: 97% *y* 98%. Edge index differential of 0.10 (with a land-cover map index of 49.05)
Chen et al. [43]	2017	Simulation of urban land changes	LP-CA	Shenzhen (China)	Higher average accuracy: 73.08%. Coefficients of correlation of 0.902, 0.883, and 0.881 between the industrial, residential, and commercial land change areas observed and simulated
Jat et al. [44]	2017	Predicting urban growth from satellite images	CA (SLEUTH model)	Ajmer, Rajasthan (India)	Overall accuracy: 80% urban area, 83% urban borders, and 60% for urban clusters
Li et al. [45]	2017	Predicting urban growth from satellite images	Segmentation-Patch-CA	Guangzhou (China)	Overall accuracy: 96%
Liu et al. [46]	2017	Future land-use simulation (FLUS)	CA + ANN	China	Overall accuracy: 84.7%
Feng and Tong [47]	2018	Predicting urban growth from geometric maps and satellite images	DE-CA	Kunming (China)	Overall accuracy: 92.4%
Traore et al. [48]	2018	Predicting urban growth from satellite images	CA-Markov	Conakry (Guinea)	Overall accuracy: 92%
Pazos-pérez et al. [49]	2018	Prediction of urban vertical growth	GA + EC	Minato, Tokyo (Japan)	Overall accuracy number of buildings: 100%, with a 19.5% deviation in building height
Fu et al. [50]	2018	Land-use simulation	CA-Markov	Hamilton County, Ohio (USA)	Overall accuracy: 91,07%
Feng et al. [51]	2018	Land-use simulation	CA + GA CA + PSO CA + GSA CA + LR	Yangtze River Delta (China)	Overall accuracy: 88%
Lipinget al. [52]	2018	Land-use simulation	CA-Markov	Jiangle (China)	Overall accuracy: 92.33%
He et al. [53]	2018	Predicting urban growth and land-use simulation	CA + UMCNN	Pearl River Delta (China)	Overall accuracy: >93%
Yuliantoe et al. [54]	2019	Land-use simulation	CA-Markov	Citarum Watershed, West Java (Indonesia)	Overall accuracy in the most optimistic scenario: merit figure 72.5%, accuracy of producer 78.5%, and accuracy of user 79.6%
Lu and Wu [55]	2019	Land-use simulation	CA-Markov	Hefei (China)	Overall accuracy: 90,48%, 87,76%, 85,1%, and *y* 82,36%, for the 3-, 5-, 10-, and 15-year intervals, respectively
Devendran and Lakshmanan [56]	2019	Predicting urban growth from satellite images	CA-Markov + NNACA	Chennai (India)	Overall accuracy: 84%
Huang et al. [57]	2020	Land-use simulation	CA-Markov	Beijing (China)	Relative error on construction land <0.3%
Khawaldah et al. [58]	2020	Land-use simulation	CA-Markov	Irbid (Jordan)	Overall accuracy: 78.4%
Mohamed and Worku [59]	2020	Land-use simulation	CA-Markov	Addis Ababa (Ethiopia)	Overall accuracy: 87%
Nurwanda & Honjo [60]	2020	Prediction of urban growth and land surface temperature	ANN-Markov	Bogor City (Indonesia)	Overall accuracy >90%

Table 1: Continued.

Author/s	Year	Objective	Method	Location	Results
Anand & Oinam [61]	2020	Land-use simulation	ANN-Markov	Manipur River (India)	Overall accuracy: 88%–93%
Mansour et al. [62]	2020	Predicting urban growth from satellite images and land-use simulation	CA-Markov	Nizwa, Al Dakhiliyah, (Oman)	Overall accuracy >80%

geostatistical methods to explain the pattern of distribution of houses in natural areas, and a model based on genetic algorithms that generated the building plan in correlation with a specific DAI. The Nanjing Yuhuatai and Qingliangshan Parks were used as cases for verifying the IOD's effectiveness. The author claimed that the model provided outstanding flexibility in the location of the buildings and that there is organic uniqueness and wide variety in the result of the calculation. He proposed to use it as a guide and reference when planning green spaces during the construction of the building layout. By using just the IOD, the limitation required of the model is relatively restrictive, making the results random and difficult to apply. Besides, the building project will also be influenced by different factors in practical projects, for example, the surface, the shape of the natural zones, the entrances and paths, the terrain and the location of existing buildings, and the practical conditions were not considered under the model presented.

To overcome this disadvantage, the new integrated CA models were recently proposed that were optimized via methods based on the swarm intelligence algorithms.

In 2016, Naghibi et al. [41] proposed a new urban growth model based on CA, using an artificial bee colony (ABC) algorithm to extract optimal transition rules. The ABC is an advanced algorithm based on meta-heuristic swarm intelligence that performs well in solving optimization problems. The authors applied the ABC-CA model to project urban growth with Landsat images from 1997, 2006, and 2015 in Urmia (Iran). The year 1997 was chosen as the base year for simulating future urban growth, and 2006 and 2015 land uses were used, respectively, to evaluate and validate the results. Finally, results for urban growth in 2016 were obtained from the simulation.

The reproduction results were tested using various statistical methods, for instance, overall accuracy, total operating characteristics (TOCs), and total operation Ant colony optimization (ACO) calibrated the CA model to evaluate the productivity of the model raised against similar methods with swarm intelligence algorithms. The authors noted that the ABC-CA model's overall accuracy and merit figure are 90.1% and 51.7%; 2.9%; and 8.8% higher than the ACO-CA model, respectively. In relation, the disparity in the ABC-CA model's allocation of simulation results is 9.9%, which is 2.9% lower than the ACO-CA. To conclude, with fewer allocation errors the ABC-CA model exceeded the ACO-CA model (Figure 1).

In 2016, Feng, Liu, and Batty [31] presented an automatic CA learning model (called MachCA) with nonlinear transition rules based on the LS-SVM to simulate city growth. By launching the input data using the LS-SVM method in a high-dimensional space, a perfect hyperplane is built to move away from the more complex limits between the two terrains (urban and nonurban), allowing the recovery of transition rules from nonlinear CA. For each iteration of the model implementation, the transition rules are dynamically updated in the MachCA model. Applying MachCA to simulate metropolitan growth on China's Qingpu-Songjiang surface in Shanghai revealed that rural-urban trends can be translated into spatial structures. A comparison of the MachCA model with a traditional log-adjusted CA model (called LogCA) resulted in fewer failures and more hits with the MachCA model because of its capability to capture the spatial complexity of urban dynamics (Figure 2). This translated into improvements in the accuracy, however, with a deviation of less than 1% in overall errors produced between the MachCA and LogCA models. However, the way the MachCA model is used to retrieve the rules of transition provided a new way to project the active procedure of metropolitan growth.

Feng and Tong [47] later developed a new mixed model (called DE-CA) that integrated differential evolution (DE) in CA to decipher the objective function and rescue the perfect CA parameters. The DE-CA has been adjusted through spatial data from the past to mimic land use at Kunming in 2016 and to predict many environments for 2026. The quantitative accuracy evaluation showed that the NA-DEC gives an overall accuracy of 92.4%, where 6.8% is the suitably collected growth of the urban area. Besides, the model only reports 2.6% failures and 5% false alarms. The authors projected three scenarios for 2026 with the use of DE-CA to adequately collect reference district development, conservation of the environment, and metropolitan design to explain their new model's robust predictive capabilities (Figure 3).

A comparative study of four CA models incorporating logistic regression (LR) and three metaheuristics was published by the same researchers [51] to simulate the land-use change in the Yangtze River Delta from 2005 to 2015. The metaheuristic methods were driven by an objective function representing the transformation rules' root mean square error (RMSE) and were considerably diverse in terms of optimization iteration, algorithm structure, and computation time. The authors argued that in cases where the complexity of the algorithm and the computation time are not of immense importance, any of the three metaheuristics could be used to build CA models for land use, as equivalent results can be attained (Figure 4).

In 2017, Perez-Molina et al. [42] merged a cellular automaton model with the flood modeling tool openLISEM to

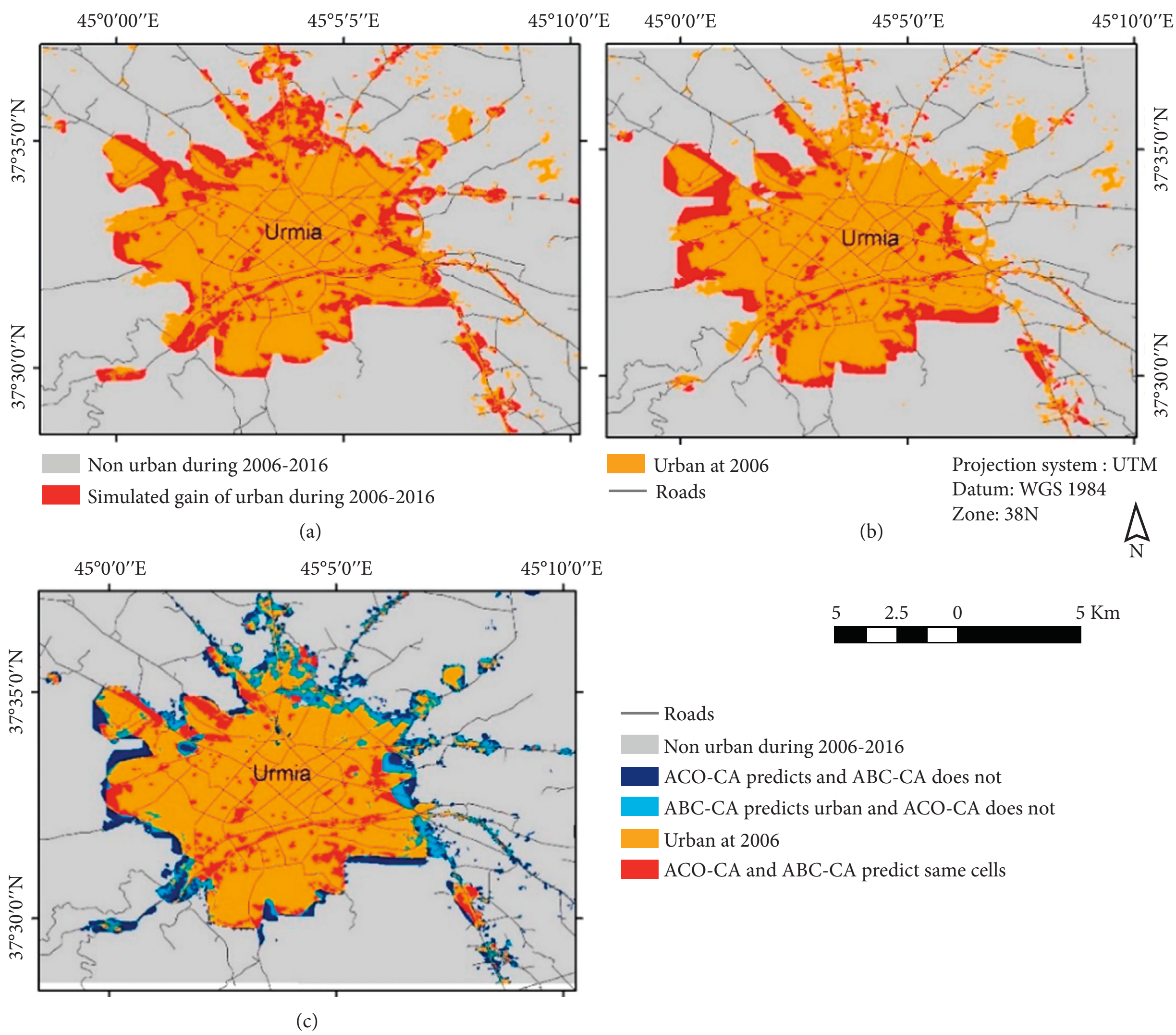


Figure 1: Results of Urmia town prediction for 2016. (a) ABC-CA model; (b) ACO-CA model; and (c) ACO-CA and ABC-CA models' comparison [41].

project district growth and subsequent flooding scenarios. This model was adjusted for the upper Lubigi (Kampala, Uganda), a sub-basin that had rapid metropolitan growth from 2004 through 2010. The model of the cellular automaton has been validated at Nalukolongo (Kampala, Uganda). Then, the authors employed the adjusted model set to project the upper Lubigi district growth contexts by 2020. There were simulations of two atmospheres, predilection requirements, and a strict policy for the protection of existing wetlands. The result of the upper Lubigi-projected scenario showed the ineffectiveness of a policy of exclusive wetland protection; likewise, a significant rise in the impacts of floods due to metropolitan growth is expected for 2020. The authors stated that the tool demonstrated its utility in creating significant land-cover change scenarios and in analyzing flood mitigation measures in a low-data number environment and that this strategy could also be applied to other spatially differentiated hazards that are seen affected by changes in the land cover. Liu et al. also studied simulation models for land-use and land-cover changes (LCCs) considering the background climate effects and proposed [46] a coming land-use simulation model (FLUS) that mimics the long-term spatial trajectories with multiple LCC's. During the projection period, the top-down system dynamics and bottom-up cellular automata interactively docked, and within the CA model, a self-adaptive competition and inertia

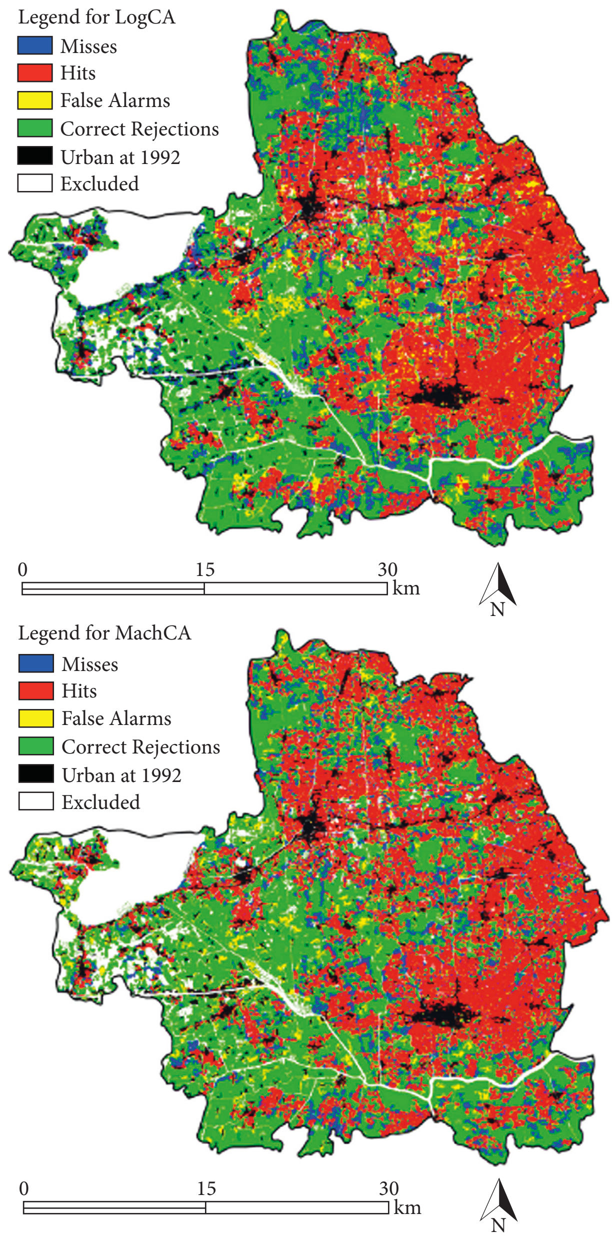


Figure 2: In 2008, metropolitan growth in the Qingpu-Songjiang area was projected using LogCA and MachCA models. Land use whose nomenclature is "excluded" shows water masses and wetlands. Hits (H) show that the increase observed was projected like growth. Misses (M) show that increase observed was projected like persistence. False alarms (FAs) show that the persistence observed was projected like growth. Correct rejections (CRs) show that persistence observed was projected like persistence (after Chen and Pontius 2010) [31].

mechanism was developed to conduct complicated tasks and interactions between several types of land uses. The model was introduced from 2000 to 2010 in China to a LUCC projection where the results indicated a promising agreement between the networks about the actual use of the land, and the precision was higher than other accepted systems like CLUE-S and CA. The model was even better applied to situations from 2010 to 2050 that encompassed various augmentation techniques that consider climatological, natural, and socioeconomic factors. The authors claimed that the simulation results indicated the correct use of the FLUS model in the hot spot regions and examined the causes and effects of potential active uses of land, which could assist academics and judgment-makers to develop correct policies to improve acclimation in the context of global warming to the rapidly changing natural setting.

Chen et al. [43] studied the reliability of fine-scale Earth simulations for raster cellular automaton (CA) models, because regular pixels/grids cannot accurately represent irregular geographic entities and their interactions. The authors proposed that vector CA models can overcome these deficiencies due to the vector data structure's ability to represent realistic urban entities. In their 2017 study, they presented a model of cellular plot automation (LP-CA) to

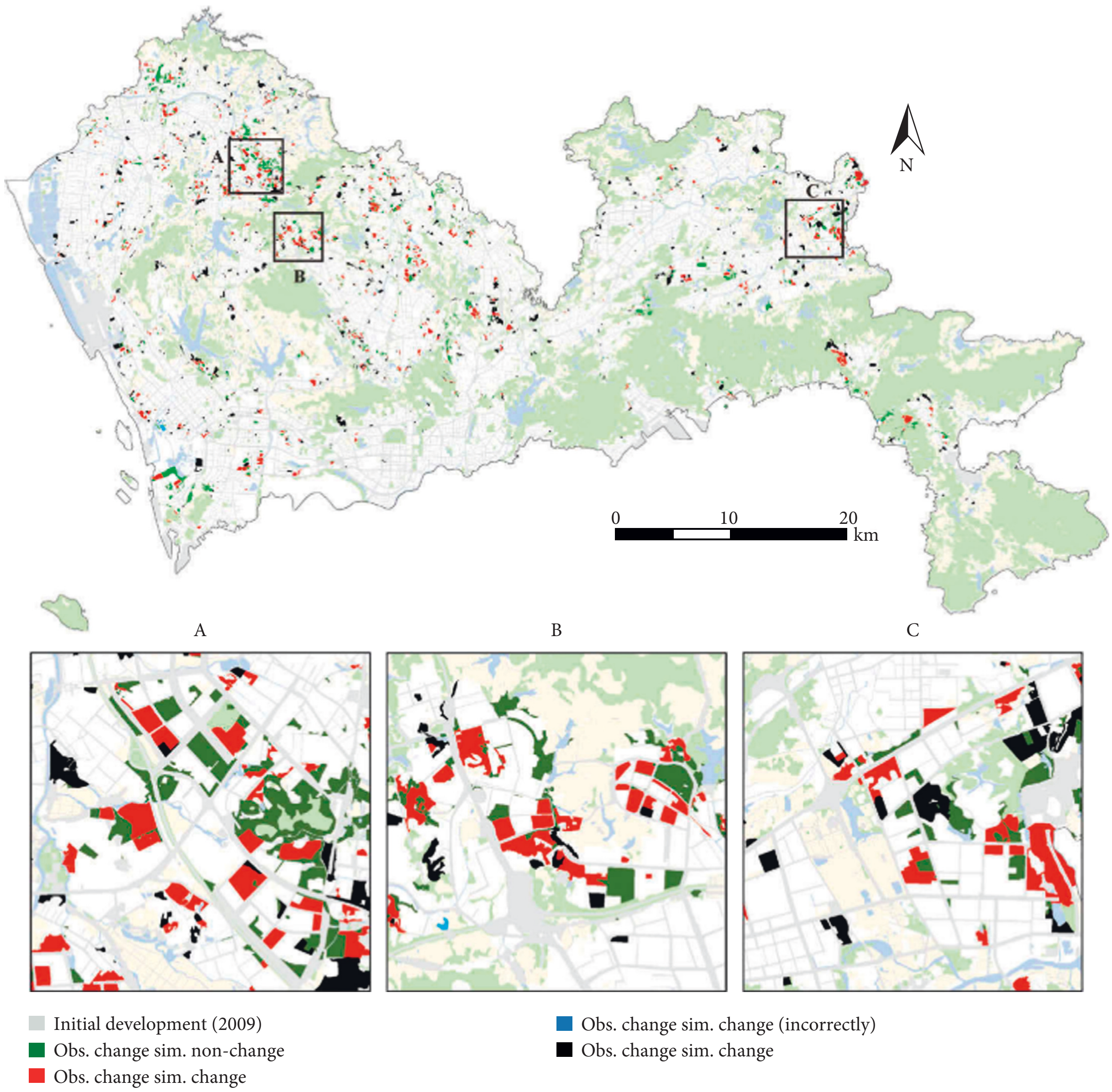


Figure 3: Three extended regions displaying the changes between projected and observed results. The disparities between the CA models are marked with circles [47].

simulate changes in urban land, where the model's innovation lies in the use of the automatic calibration joint-learning method. Three commonly used classifiers, Naïve Bayes, neural retro-diffusion networks, and decision trees were chosen to create a joint classifier as the base classifiers. The proposed model was applied in Shenzhen (China), and the experimental results showed the maximum calibration accuracy produced by bagging Naïve Bayes among a selected set of classifiers (Figure 5). District sensitivity assessment suggested that the LP-CA model with a neighboring radius $r = 2$ achieved the highest simulation accuracy. The calibrated LP-CA was used to challenge future changes in urban land use in Shenzhen, and the results were consistent with those specified in the official city plan according to the authors.

Li et al. [45] proposed a new model of metropolitan growth centered on patches with heuristic regulations that used a basin segmentation algorithm (Segmentation-Patch-CA) of the logistic CA model. Section objects derived from the properties of the metropolitan CA model were considered potential conversion patches, when determining a useful function that perceived the suitability and heterogeneity of the internal pixels. Two distinct kinds of metropolitan development were then recognized and individually recreated: organic growth and spontaneous growth, via the implementation of a neighborhood density analysis-based landscape expansion index (LEI). This proposal was applied to the city of Guangzhou, China (Figure 6). The results revealed an improvement in the landscape similarity index (LSI) that reached 20%–50% since the

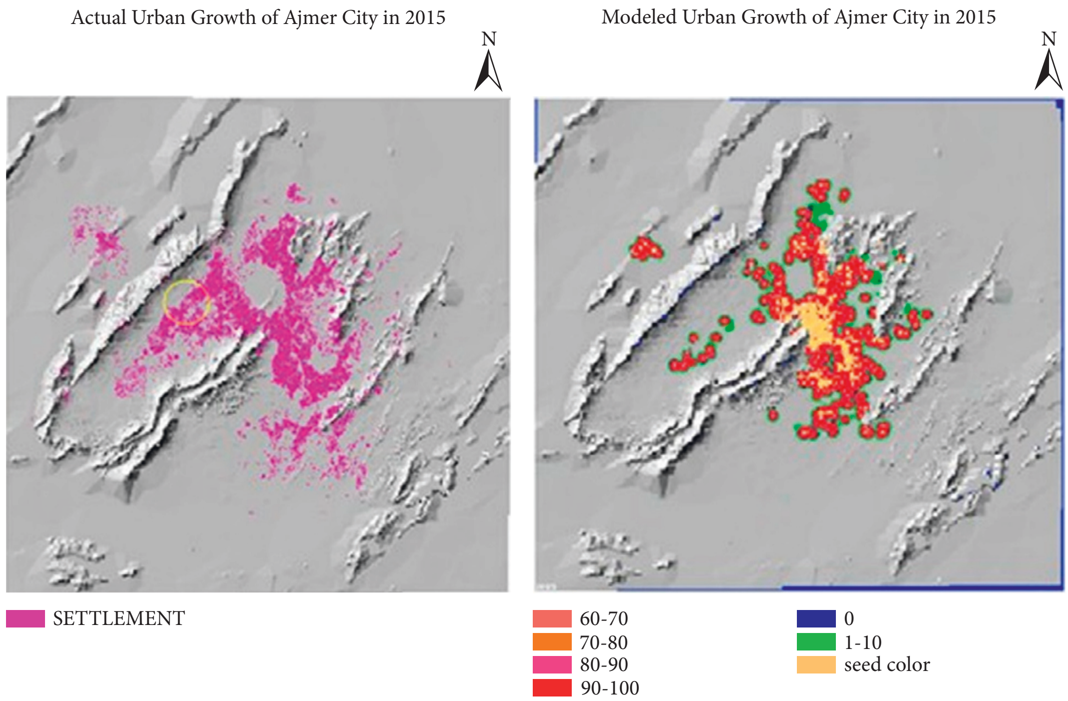


FIGURE 4: Simulation of 2015 LR-CA patterns and three metaheuristic CA models [51].

proposed system generated a more credible district landscape (96 and 97.38%) than the pixel-based one (45.14% and 74.82%) for two 2003–2008 and 2008–2012 modeling periods.

In 2017, Jat et al. [44] presented a study to evaluate the performance of the CA-based SLEUTH model to project the metropolitan rise of a complicated and slightly more heterogeneous area, the city of Ajmer in Rajasthan, India, which is different from the other areas tested by SLEUTH. Seven multispectral satellite images covering more than 21 years were processed and used for the parameterization of SLEUTH. The results of the urban growth predicted by SLEUTH were compared with other land-use/land-cover extraction methods. The authors stated that the study was successful in offering a meaningful insight into the issues benefiting to the risks in forecasting urban development in heterogeneous metropolitan areas. They have, however, identified some issues related to SLEUTH's sensitivity regarding the scale and spatial resolution of the input variables. Furthermore, the effects of the model indicated that SLEUTH is not capable of capturing the development of small size units, i.e., how unfragmented outdoor expansion is common in developing countries (Figure 7). The model underestimated the fragmented urban growth that can be compared to the approximate resolution implemented throughout the phases, and to the smallest average size below the resolution of the units created and errors in the input data due to classification errors in satellite images. This can be attributed to the heterogeneity of the procedure and the material used for construction.

With the use of the cellular automaton and Markov (CA-Markov) model incorporated with the geographic information system (GIS) and remote sensing (RS) [48], Traore, Mawenda, and Komba simulated land-cover change in Conakry, Guinea. Old information on land-cover modification was taken from Landsat data for the years 1986, 2000, and 2016.

The simulated result was compared for evaluation with the use of the relative operating characteristic (ROC) curve to the land-cover map of 2016. The ROC outcome demonstrated a substantial degree of agreement between both maps. Based on the result, using the CA-Markov model, they simulated the land-cover change map for 2025. The result showed that the metropolitan area ratio was 49% in 2016 and is anticipated to rise to 52% by 2025. On the contrary, the vegetation will fall from 35% in 2016 to 32% in 2025. The authors are optimistic about that study's results. They believe that the model will provide a basis for assessing the sustainability and urban area management and taking measures to mitigate urban environmental degradation.

Liping et al. [52] also used the CA-Markov model to predict potential land-use patterns through remote sensing and geographic information systems based on dynamic changes in usage patterns. They obtained a map predicting land use for 2014 in Jiangle County, China, based on the CA-Markov model, which was validated with actual 2014 results with a Kappa index of 0.8128. The authors also set the 2025 and 2036 land-use patterns.

In the upstream Citarum River Basin (West Java), Yulianto et al. [54] with the inclusion of remote sensing data and the CA-Markov model studied the dynamics of land-use reform and its estimation for the coming year.

Lu & Wu [55] used the method of spatial-temporal data fusion (STF) to obtain summer-scale images of Landsat over the past 30 years in Hefei, China. They used the CA-Markov model to simulate and predict future maps of land-use/land-

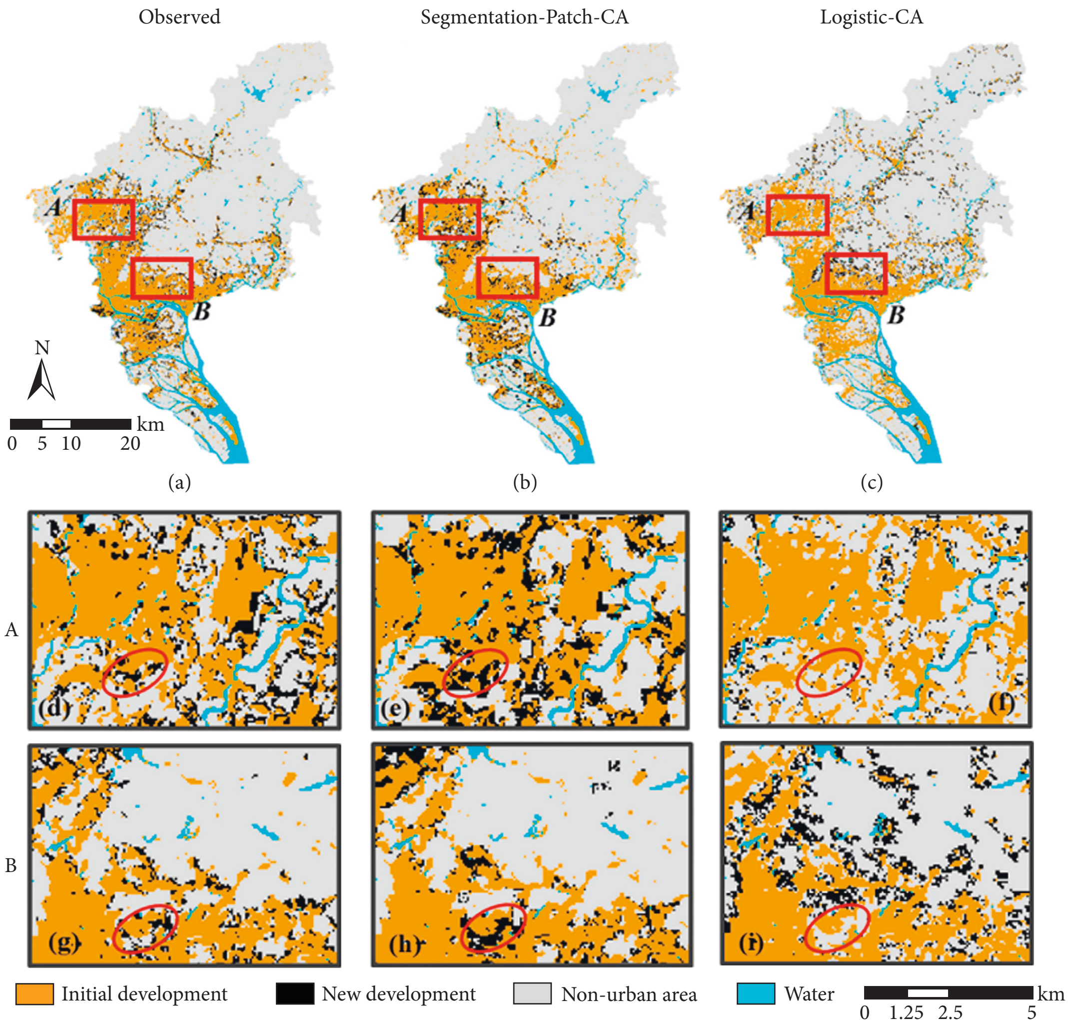


Figure 5: Chen's CA model applied in Shenzhen (China) 100 simulations overlapping ($r = 2$) for (a) industrial land; (b) residential land; and (c) commercial land [43].

cover change (LULC). The authors showed that merged data can improve the accuracy of LULC detection and prediction by shortening the year-to-year interval and obtain results for a specific year from LULC predictions. In their research, they observed that the areas of cultivated land, vegetation, and water declined by 33.14%, 2.03%, and 16.36%, respectively, and the area of land under building extended by 200.46% from 1987 to 2032.

Similar research was conducted by Huang et al. for the Beijing Territory [57]; Khawaldah et al. in Irbid, Jordan [58]; Mohamed and Worku in Addis Ababa and its environs [59]; and Mansour, Al-Belushi, and Al-Awadhi in the Nizwa mountain area of Oman [62].

Anand & Oinam [61] centered their research on the Manipur River Basin (India) wetlands and developed a future area LULC map using the Markov chain and an artificial neuron network. In Bogor City, Indonesia, Nurwanda & Honjo [60] applied the combination of multilayer perceptron and Markov chain (MLP-MC) to predict the future expansion of the city and land surface temperature (SST).

In the multicriteria evaluation (MCE), the Markov model of cellular automaton (CA) was used as a tool for decision-making about land use, analysis, and simulation [50]. Fu et al. explored the possibility of using historical data from a specific area for factor selection, scoring, and quantification. The authors created logistic regression models calibrated to choose and score every potential factor for historical land-use modifications and used the entropy technique for selecting the weights of the chosen variables. The SCM output is used as an input to the CA-Markov model to simulate changes in land use from 2001 through 2011. The simulation result was compared with the land use observed in 2011 to examine the method's performance. The result showed that the use of the SCM factors derived from old data creates a fitness of fit. The biggest advantage of this method is that it derives the selection of variables, ratings, and weights from local data that reflect the actual pattern. This numerical approach leaves for effective adjustment of the CA-Markov model and the growth of various land-use planning scenarios by adapting the rankings and weights of the various issues with the understanding of global change.

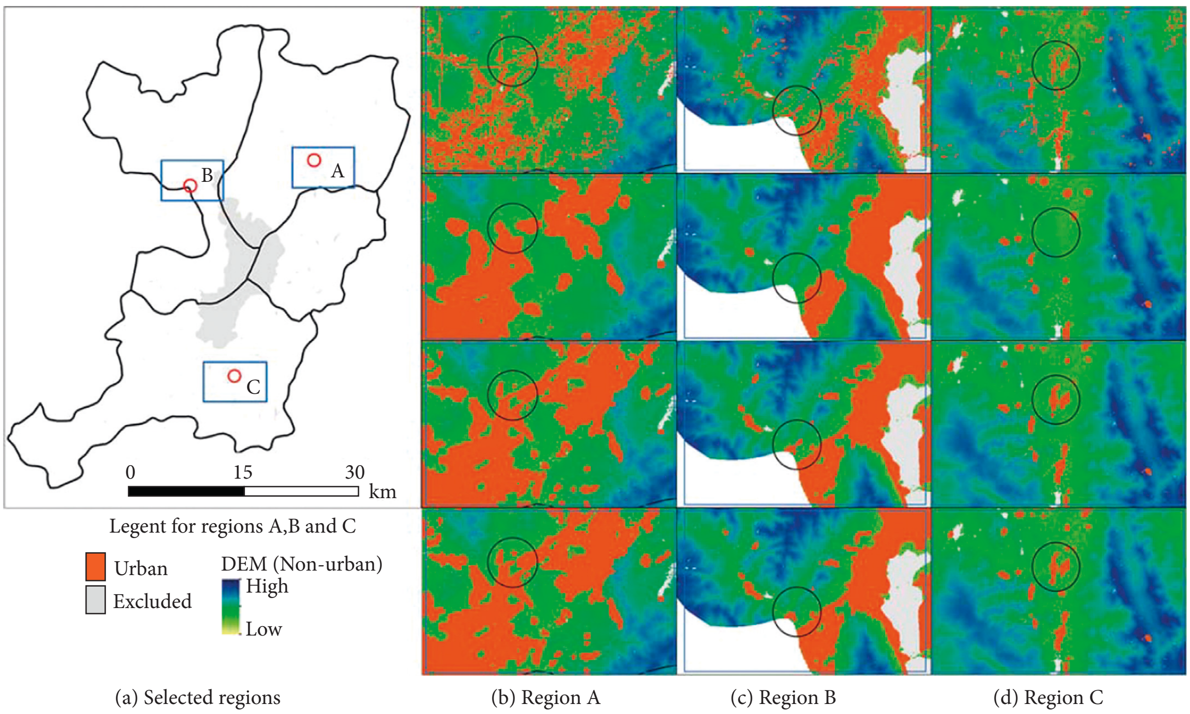


FIGURE 6: Comparison between two different approaches to growth with what was observed in Guangzhou (China), 2012. (a), (b), and (c) are overall results, and fusions of (d), (e), (f) and (g), (h), and (i) are zoomed-in viewpoints, respectively, of local areas A and B [45].

Devendran & Lakshmanan [56] used an urban growth model with cellular automata based on agent-coupled neural networks (NNACA) and historical datasets to forecast the growth of the Chennai metropolitan area (India). The prediction model used eight different urbanization agents, including transport, access points, and industries. Validation of the results showed that the most influential agent was the neighborhood. The authors measured urban expansion through Shannon entropy, and the values obtained suggested the need for more careful planning in the future development of the area.

He et al. [53] also focused on the effects of the neighborhood on the prediction of urban growth and used a CA coupled with a neural convolution network for united mining (UMCNN) and Markov chains to improve the performance of the simulation of urban expansion processes. They chose the Pearl River Delta (China) as a study area with the aim of verifying the effectiveness of deep learning in urban simulation, compared with three machine learning-based CA models (LR, ANN, and RFA). The proposed method achieves the highest simulation accuracy (>93%) and similarity to the landscape index (>89%). However, the authors warn that, although the accuracy of the model is greatly affected by the size of the training window, its lowest result is still higher than the traditional CA model.

Pazos-Pérez et al. investigated the use of evolutionary genetic algorithms to predict metropolitan vertical evolution scenarios in the big central districts [49]. Tokyo's Minato district was used for the case study, as it has rapidly grown over the past 20 years. A genetic algorithm that replicates vertical urbanization was used to make predictions based on initial set parameters, calculating not only the number of potential high-rise structures but also the specific locations most probable to support new high-rise advancements in the future. To assess the accuracy of the genetic algorithm in projecting future vertical district growth, the results of the evolutionary model were compared with continuous high-rise evolutions. After the test, the genetic algorithm's predictions for the period 2016–2019 (Figure 8), and sharing the results with the actual ongoing projects now, the researchers concluded that the algorithm's growth projections were accurate in terms of a complete number of properties and their likely location (±6.67 percent). On the other hand, the algorithm did not accurately determine the year of growth (±1 year) with exactitude and building height (19.5% deviation), indicating that more studies must be carried out in the areas. This experiment proved that the use of evolutionary genetic computation is a method to predict vertical metropolitan growth with a lot of possibilities concerning space and the number of potential edifications.

As seen in the previous section, there are different variables that affect the prediction. The studies mentioned above have used different input variables for their predictions. These data can be seen in Table 2. Although the variables used are different depending on the study, there are a number of factors that stand out because they are present in many of them: population, race, diffusion, elevation, slope, environment, density, accessibility, constraint factors, stochastic or random factors, probability of development, socioeconomic changes, DEM, and historical land use through neighborhood factors and distances (between urban or nonurban spaces), roads, rivers, railroads, etc. We can appreciate the predominance of qualitative and quantitative

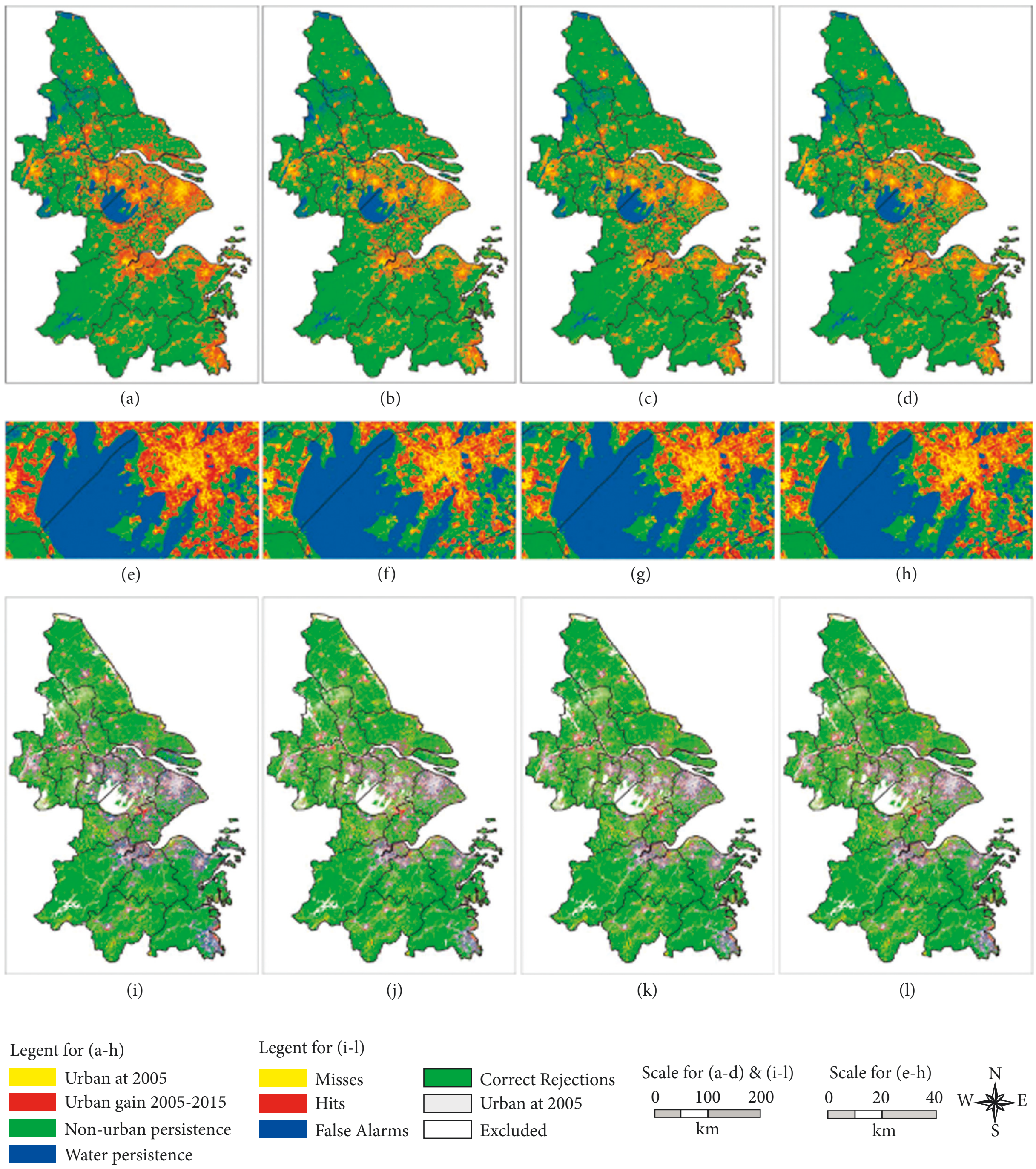


Figure 7: Modeled and actual urban growth in Ajmer (India) [44]. (a) Stimulated change by LR-CA. (b) Stimulated change by PSO-CA. (c) Stimulated change by GSA-CA. (d) Stimulated change by GA-CA. (e) Enlarged area of LR-CA. (f) Enlarged area of PSO-CA. (g) Enlarged area of GSA-CA. (h) Enlarged area of GA-CA. (i) Stimulation success and error by LR-CA. (j) Stimulation success and error by PSO-CA. (k) Stimulation success and error by GSA-CA. (l) Stimulation success and error by GA-CA.

CA models, to the detriment of the use of binary or vector CA models or other methods (ANN, GA, etc.).

2.2. More Relevant Data. The implementation of AI for urban planning is a flourishing field with an increasing number of researchers studying and publishing papers in this field. If we consider the different articles mentioned, we can see that since 2016 the number of research articles published in this regard has been increasing, with a decrease in 2018, which equals the number of publications with 2016 (3), but again the course of studies on the subject is redirected in 2020. We have not included the year 2021 within this state of the art, since it is the year of data collection and no results were found in this regard. The year with the highest number of publications was 2018 (7) followed by 2020 (6).

In addition, when it comes to the use of AI for urban planning, it is worth highlighting the countries used in the

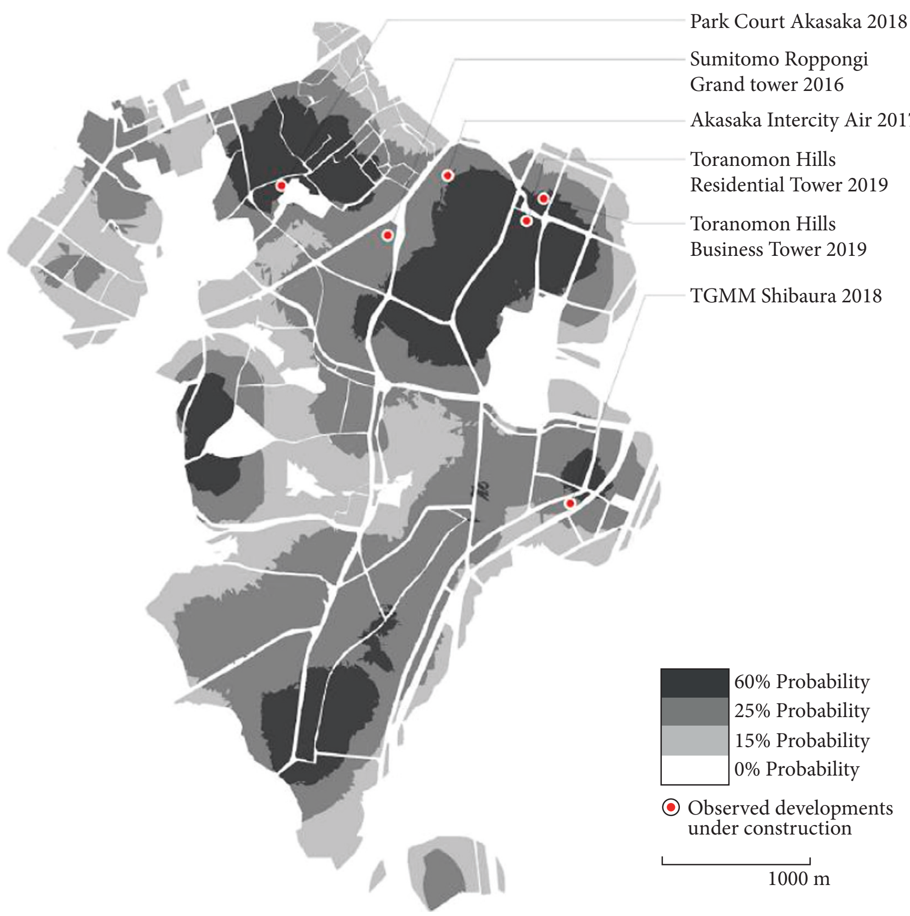


Figure 8: Genetic algorithm prognostications on a probabilistic grayscale in Tokyo's Minato district. Darker tones indicate a greater probability of new high-rise advancements (130 m approximately). Dots constitute discharge structures already planned for completion between 2016 and 2019 [49].

Table 2: Variables used for prediction in the articles of this review.

Author/s	Variables
Tong [40]	IOD is the only criterion used. The study mainly focuses on the geometric importance of the building, and the building types are ignored. The existing buildings in green spaces should also be taken into account in the calculation, but these constraints and practical conditions are not taken into account.
Naghibi et al. [41]	Remote sensing image (Landsat images) to land-use maps; distance to roads map; distance to business center map; distance to population center map; environmental-sensitive area map; slope map; and elevation map.
Feng et al. [31]	Conversion label; distance to urban center; distance to town center; distance to commercial housing area; distance to main roads; distance to agricultural land; and stochastic.
Perez-Molina et al. [42]	Spatial factors: calibrated model suitability, neighborhood factor, travel time to CBD factor, and wetland factor + 0,25 ∗ random.
Chen et al. [43]	Model control parameters: diffusion, breed, spread, slope, and roads.
Jat et al. [44]	SLEUTH parameterization with supervised classification (land use and urban) and the use of statistical measures: sng (cumulative number of urbanized pixels by spontaneous neighborhood growth), og (cumulative number of urbanized pixels by organic growth), rt (cumulative number of urbanized pixels by road influenced growth), area (total number of urban pixels), edges (number of urban to nonurban pixel edges), clusters (number of urban pixel clusters), rad (radius of cluster, which encloses the urban area), slope coefficient, spread coefficient, breed coefficient, road gravity coefficient, percent urban (percent of urbanized pixels divided by the number of pixels available for urbanization), urban growth rate, and number of growth pixels each year.
Li et al. [45]	Two different urban growth types: organic growth (based on segment) and spontaneous growth (based on pixels), which where identified and separately simulated introducing a landscape expanse index (LEI) that built on neighborhood density analysis. CA components: suitability surface, neighborhood, stochastic perturbation, and development probability.

Table 2: Continued.

Author/s	Variables
Liu et al. [46]	Socioeconomic and natural climatic factors: climate change, socioeconomic changes, historical land use, and interactions between variables to obtain land-use demand in each decade. Later, to ANN for land use in 2010: neighborhood influence, weight factor, self-adaptively land inertia, and converting cost. These combined probabilities with probability-of-occurrence surfaces and with roulette wheel selection detect the land use in time.
Feng & Tong [47]	Constrained relations among factors were applied in DE to generate different sets of CA parameters for the prediction of future scenarios. Variables that affect land-use changes: distance to urban center; distance to district center; distance to main roads; distance to the roads along Dianchi Lake; distance to protected areas; and DEM.
Traore et al. [48]	Prior to classifying the images using a supervised classification algorithm, unsupervised classification and normalized difference vegetation index (NDVI) were calculated to help select suitable polygons as training sites and to improve the overall classification process. The classification scheme was established based on auxiliary information from the field survey, local knowledge of the study area, and visual interpretation of the images. Image classification was performed using the maximum likelihood classification (MLC) algorithm, which is a supervised classification, and one of the most widely applied parametric classification algorithms.
Pazos-Pérez et al. [49]	A series of grayscale probabilistic maps with different parameters were produced to be used as the basis for the evolutionary model. The parameters were captured in the following gradient maps: land ownership, regulatory master plans, vertical urban consolidation, accessibility, and allocation. Land ownership: public vs. private; land redevelopment master plans; vertical density; accessibility; allocation parameters; and economic and real estate parameters.
Fu et al. [50]	They used the entropy method to determine weights for the selected factors. Potential factors for multicriteria evaluation: population change; change in employment; population density; median housing income; and highway accessibility, transit accessibility, slope, distance to each of the existing land use types, administrative constraints, and natural constraints.
Feng et al. [51]	Criteria for comparing CA metaheuristic models: best objective function value; iteration or generation; computational time; initial urban area; hit; correct rejection; failure; false alarm; assignment; and quantity. Spatial input variables: distance to city center; distance to country center; distance to main road; distance to railroad; distance to coast; DEM; and restricted areas.
Liping et al. [52]	This study uses remote sensing and geographic information. From 1992 and 2003 Landsat 5 TM images, and 2014 Landsat 8 OLI images and DEM, a land-use classification map was obtained for each year. The cell automaton model is mainly composed of cell, cell space, neighbor, ruler, and time. The closer the distance between the nuclear cell and the neighbor, the higher the weight factor. The weight factor is combined with transition probabilities to predict the state of adjacent grid cells so that land-use change is not a completely random decision. The Markov chain model component controls the temporal dynamics between LULC classes based on the transition probabilities, while the spatial dynamics are controlled by local rules determined by the CA spatial filter or transition potential maps.
He et al. [53]	They use spatial variables in UMCNN. For RFA-CA, the factors used are neighborhood effects, constraint factors, development suitability, and stochastic factors.
Yulianto et al. [54]	Inputs for the training phase of the CA-Markov model: time-1 land-use map; time-2 land-use map; simulated n-time transition area matrix; and simulated n-time Markov conditional probability image. Inputs for the simulation phase of the CA-Markov model: simulated n-time transition area matrix and simulated n-time Markov conditional probability image.
Lu & Wu [55]	Preprocessing tasks, such as radiation calibration, FLAASH atmospheric corrections, image mosaicking, and image cropping, were applied before classifying the images with the ENVI tool.
Devendran & Lakshmanan [56]	Agents of urbanization: existing built-up; hot spots; commutation, high-preference roads; medium-preference roads; least-preference roads; railways; red category industries; orange category industries; green category industries; white category industries; high land prices; low land prices; medium land prices; places of public interests; public utility centers; and population.
Huang et al. [57]	Driving factors: DEM, slope, aspect, GDP, population, highway, rail, river, road, and roc index.
Khawaldah et al. [58]	Image preprocessing techniques include the following: layer stacking; mosaicking; and subsetting or clipping to study area boundaries. The LULC classification scheme comprised seven LULC classes, identified by codes, to prepare different LULCs to simulate future land use.
Mohamed & Worku [59]	The research described the continuing historical increase in built-up space through the consumption of other ecologically valuable LULC classes. Driving factors: elevation, slope, road distance, highway distance, rail distance, and urban centers.
Nurwanda & Honjo [60]	Model control parameters: slope, distance to roads, distance to toll road, and elevation were also used as variables that influenced land-use change.
Anand & Oinam [61]	The ANN was trained with the driver variable, i.e., distance to roads, distance to settlement, elevation, and slope.
Mansour et al. [62]	The analysis was based on three equal interval LULC maps derived from satellite images: Landsat TM for 1998, 2008, and 2018, together with topographic spatial layers (elevation aspects and terrain slopes) derived from the ASTER digital elevation model. Other spatial parameters (population density, proximity to urban centers, and proximity to major roads) were also incorporated into the simulation process.

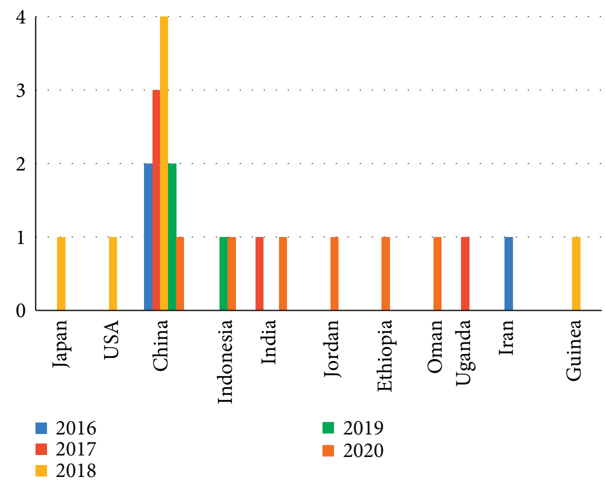


Figure 9: Country of urban development/year.

different works. In the different approaches mentioned in this review, AI techniques are applied to urban planning in large metropolitan areas in various parts of the world, although most papers used in this review focus on urban development in China. For this, Figure 9 is included, which shows the number of articles related to each country per year. In the graph of Figure 9, we can see that most of the works of research urban development in China as mentioned. It is the only country in which investigations have been carried out in all the years included (2016–2020), with 4 of them in 2018. On the other hand, it should be noted the countries used in the scientific research in 2020 (Oman, Ethiopia, Jordan, India, Indonesia, and China).

Perhaps one of the most important pieces of data that we have been able to collect concerns the techniques used in the different experiments. In the papers used in this review, most of the systems combine Markov chains and cellular automata. A less number of papers use other techniques, like artificial neural networks. This contrasts with the substantial number of papers related to architectural design using genetic algorithms [86–93]. You can read more about it in a state of the art that we have previously made [94].

In particular, the AI techniques applied to urban planning used by the authors of the different works that we have found are as follows: the combination between CA and Markov stands out, with three jobs in 2018, two in 2019, and one in 2020. The rest of the jobs in 2020 also use Markov, but in this case, combined with ANN. For its part, the other remaining work from 2019 also uses CA and Markov, but with the peculiarity that it uses neural networks as well.

We can see in Figure 10 that for the rest of the experiments all the works used CA, except [40, 49], which in 2016 used simple GA and combined with EC in 2018, successively. On the other hand, the research that uses CA only has two examples in 2017 of its individual use [42, 44] and the second with a specific model (SLEUTH). In the rest of the studies, they use it in combination with other techniques or the case of [45] use Segmentation-Path-CA.

The best results obtained could be considered by those of [49] with the use of GA + EC, who in 2018 obtained a global

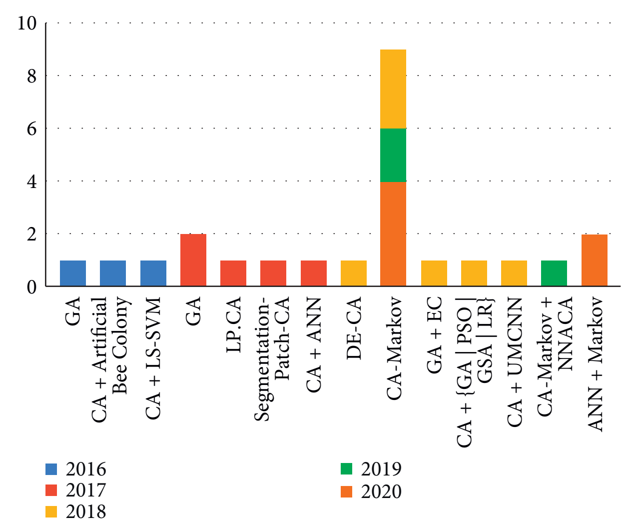


Figure 10: Techniques/year.

precision of the number of buildings of 100%, with a deviation of the height of the building of 19.5%. The results of [42] were also very good with the use of CA achieving an overall precision of 97% and 98%, and an edge index dispersion of 0.10 (with a land cover map index of 49.05), and [45] through Segmentation-Patch-CA obtained an overall precision of 96%. Both are prior to [49], so the work after 2018 has not achieved results that improve the union of AC and EC.

3. Conclusions

For years now, CAs have been successfully used in the exploration of a wide variety of urban phenomena [95], from regional-scale urbanization and urban development to traffic simulation [96–98]. The CA models have been developed for topics as varied as sprawl [99, 100] or gentrification [101–103], and simulations of city shape, growth, and location [73, 104–107]. This is because CAs have many advantages for modeling urban phenomena, such as their flexibility, their decentralized and dynamic approach, the relative ease with which model results can be visualized, and also their affinities with geographic information systems and remote sensing data [108]. In our opinion, its simplicity may be its most significant quality. By mimicking the way in which large-scale urban structures can emerge from the myriad interactions of simple elements, CAs provide a framework for the exploration of complex adaptive systems.

We have seen that in recent years these techniques are being used to simulate the urban planning approach and changes in land use, considering population migration or the effects of global warming (such as flooding or the disappearance of water and vegetation, for example), with the aim of planning and achieving sustainable urban development.

Although the best results dating from 2018, it is considered that the other works also obtain satisfactory results, since they are very diverse uses, and as mentioned, different countries and techniques are used.

Among the different deficiencies shown are the following:

(i) The only technique with published examples from a sufficient number of countries to compare the results is the use of CA + Markov.

(ii) This technique has not exceeded the results of the first experiment carried out in 2018 [48] with data from Conakry (Guinea). The results were of an average precision of 92%.

(iii) The only country with which enough techniques have been used to make a comparison between them is China.

(iv) Despite being the same country, the location is different, and therefore, the comparative results will be less consistent.

(v) The use of GA + EC has not been replicated, although it is the option with the best results so far.

(vi) For the most part, the predictive data have not been compared to actual subsequent results.

With the above in mind, we recommend that comparisons be made with actual results when possible. It also makes sense to us that the use of CA has been investigated, as the 2017 results showed its potential. However, following the contained explanation of the literature review we can understand that it does not solve the previous problems. That is why it would also be of interest to show the results of combining CA and EC since it is a combination that has not yet been worked on until now.

We also consider it interesting for future research to carry out more experiments under similar conditions, either of localization and varying the technique used, or of techniques modifying the location. Finally, we recommend promoting the use of GA + EC for its satisfactory results.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This project was supported by the General Directorate of Culture, Education and University Management of Xunta de Galicia (ref. ED431G/01 and ED431D 2017/16), the Spanish Ministry of Economy and Competitiveness via funding of the unique installation BIOCAI (UNLC08-1E-002 and UNLC13-13-3503), and the European Regional Development Funds (FEDER). CITIC, as Research Center accredited by Galician University System, is funded by "Consellería de Cultura, Educación e Universidade from Xunta de Galicia," supported in an 80% through ERDF Funds, ERDF Operational Programme Galicia 2014–2020, and the remaining 20% by "Secretaria Xeral de Universidades" (grant no. ED431G 2019/01).

References

[1] J. M. Levy, *Contemporary Urban Planning*, Prentice-Hall, Hoboken, 7th Edition, 2006.

[2] X. Li, Y. Chen, X. Liu, D. Li, and J. He, "Concepts, methodologies, and tools of an integrated geographical simulation and optimization system," *International Journal of Geographical Information Science*, vol. 25, no. 4, pp. 633–655, 2011.

[3] Q. Wu, H.-q. Li, R.-s. Wang et al., "Monitoring and predicting land use change in beijing using remote sensing and GIS," *Landscape and Urban Planning*, vol. 78, no. 4, pp. 322–333, 2006.

[4] D. Guan, H. Li, T. Inohae, W. Su, T. Nagaie, and K. Hokao, "Modeling urban land use change by the integration of cellular automaton and Markov model," *Ecological Modelling*, vol. 222, no. 20-22, pp. 3761–3772, 2011.

[5] H. Shafizadeh Moghadam and M. Helbich, "Spatiotemporal urbanization processes in the megacity of Mumbai, India: a Markov chains-cellular automata urban growth model," *Applied Geography*, vol. 40, pp. 140–149, 2013.

[6] R. Store and J. Kangas, "Integrating spatial multi-criteria evaluation and expert knowledge for GIS-based habitat suitability modelling," *Landscape and Urban Planning*, vol. 55, no. 2, pp. 79–93, 2001.

[7] G. Engelen, R. White, and I. Uljee, "Integrating constrained cellular automata models, GIS and decision support tools for urban planning and policy making," *Decision Support Systems in Urban Planning*, pp. 125–155, E&FN Spon, London, 1997, chapter 8.

[8] K. C. Clarke and L. J. Gaydos, "Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore," *International Journal of Geographical Information Science*, vol. 12, no. 7, pp. 699–714, 1998.

[9] F. Wu and C. J. Webster, "Simulation of land development through the integration of cellular automata and multicriteria evaluation," *Environment and Planning B: Planning and Design*, vol. 25, no. 1, pp. 103–126, 1998.

[10] M. Batty, Y. Xie, and Z. Sun, "Modeling urban dynamics through GIS-based cellular automata," *Computers, Environment and Urban Systems*, vol. 23, no. 3, pp. 205–233, 1999.

[11] X. Li and A. G.-O. Yeh, "Modelling sustainable urban development by the integration of constrained cellular automata and GIS," *International Journal of Geographical Information Science*, vol. 14, no. 2, pp. 131–152, 2000.

[12] P. Rocha, "Calibration of stochastic cellular automata: the application to rural-urban land conversions," *International Journal of Geographical Information Science*, vol. 16, no. 8, pp. 795–818, 2002.

[13] S. Leao, I. Bishop, and D. Evans, "Simulating urban growth in a developing nation's region using a cellular automata-based model," *Journal of Urban Planning and Development*, vol. 130, no. 3, pp. 145–158, 2004.

[14] F. Aguilera, L. M. Valenzuela, and A. Botequilha-Leitão, "Landscape metrics in the analysis of urban land use patterns: a case study in a Spanish metropolitan area," *Landscape and Urban Planning*, vol. 99, no. 3–4, pp. 226–238, 2011.

[15] J. Jokar Arsanjani, M. Helbich, W. Kainz, and A. Darvishi Boloorani, "Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion," *International Journal of Applied Earth Observation and Geoinformation*, vol. 21, pp. 265–275, 2013.

[16] Y. Feng and Y. Liu, "A heuristic cellular automata approach for modelling urban land-use change based on simulated annealing," *International Journal of Geographical Information Science*, vol. 27, no. 3, pp. 449–466, 2013.

[17] M. Aljoufie, M. Zuidgeest, M. Brussel, J. van Vliet, and M. van Maarseveen, "A cellular automata-based land use and transport interaction model applied to jeddah, Saudi arabia," *Landscape and Urban Planning*, vol. 112, no. 1, pp. 89–99, 2013.

[18] H. Wang, S. He, X. Liu et al., "Simulating urban expansion using a cloud-based cellular automata model: a case study of jiangxia, wuhan, China," *Landscape and Urban Planning*, vol. 110, no. 1, pp. 99–112, 2013.

[19] J. A. Verstegen, D. Karssenberg, F. van der Hilst, and A. P. C. Faaij, "Identifying a land use change cellular automaton by bayesian data assimilation," *Environmental Modelling & Software*, vol. 53, pp. 121–136, 2014.

[20] Y. Sakieh, B. J. Amiri, A. Danekar, J. Feghhi, and S. Dezhkam, "Simulating urban expansion and scenario prediction using a cellular automata urban growth model, SLEUTH, through a case study of karaj city, Iran," *Journal of Housing and the Built Environment*, vol. 30, no. 4, pp. 591–611, 2015.

[21] A. Mustafa, I. Saadi, M. Cools, and J. Teller, "Measuring the effect of stochastic perturbation component in cellular automata urban growth model," *Procedia Environmental Sciences*, vol. 22, pp. 156–168, 2014.

[22] J. Liao, L. Tang, G. Shao, X. Su, D. Chen, and T. Xu, "Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations," *Environmental Modelling & Software*, vol. 75, pp. 163–175, 2016.

[23] P. H. Verburg, P. P. Schot, M. J. Dijst, and A. Veldkamp, "Land use change modelling: current practice and research priorities," *Geojournal*, vol. 61, no. 4, pp. 309–324, 2004.

[24] I. Benenson and P. M. Torrens, "Geosimulation: object-based modeling of urban phenomena," *Computers, Environment and Urban Systems*, vol. 28, no. 1-2, pp. 1–8, 2004.

[25] X. Li and A. G.-O. Yeh, "Neural-network-based cellular automata for simulating multiple land use changes using GIS," *International Journal of Geographical Information Science*, vol. 16, no. 4, pp. 323–343, 2002.

[26] B. C. Pijanowski, D. G. Brown, B. A. Shellito, and G. A. Manikd, "Using neural networks and GIS to forecast land use changes: a land transformation model," *Computers, Environment and Urban Systems*, vol. 26, no. 6, pp. 553–575, 2002.

[27] A. Tayyebi, B. C. Pijanowski, and A. H. Tayyebi, "An urban growth boundary model using neural networks, GIS and radial parameterization: an application to tehran, Iran," *Landscape and Urban Planning*, vol. 100, no. 1-2, pp. 35–44, 2011.

[28] Y. Qiang and N. S. N. Lam, "Modeling land use and land cover changes in a vulnerable coastal region using artificial neural networks and cellular automata," *Environmental Monitoring and Assessment*, vol. 187, no. 3, Article ID 57, 2015.

[29] B. Huang, C. Xie, R. Tay, and B. Wu, "Land-use-change modeling using unbalanced support-vector machines," *Environment and Planning B: Planning and Design*, vol. 36, no. 3, pp. 398–416, 2009.

[30] M. Samardžić-Petrović, S. Dragićević, M. Kovačević, and B. Bajat, "Modeling urban land use changes using support vector machines," *Transactions in GIS*, vol. 20, no. 5, pp. 718–734, 2015.

[31] Y. Feng, Y. Liu, and M. Batty, "Modeling urban growth with GIS based cellular automata and least squares SVM rules: a case study in Qingpu-Songjiang area of Shanghai, China," *Stochastic Environmental Research and Risk Assessment*, vol. 30, no. 5, pp. 1387–1400, 2016.

[32] X. Li, Q. Yang, and X. Liu, "Genetic algorithms for determining the parameters of cellular automata in urban simulation," *Science in China - Series D: Earth Sciences*, vol. 50, no. 12, pp. 1857–1866, 2007.

[33] J. Tang, L. Wang, and Z. Yao, "Spatio-temporal urban landscape change analysis using the Markov chain model and a modified genetic algorithm," *International Journal of Remote Sensing*, vol. 28, no. 15, pp. 3255–3271, 2007.

[34] P. Taylor, X. Li, J. Lin, Y. Chen, X. Liu, and B. Ai, "Calibrating cellular automata based on landscape metrics by using genetic algorithms," *International Journal of Geographical Information Science*, vol. 27, no. 3, pp. 594–613, 2013.

[35] Y. Liu, Y. Feng, and R. Pontius, "Spatially-explicit simulation of urban growth through self-adaptive genetic algorithm and cellular automata modelling," *Land*, vol. 3, no. 3, pp. 719–738, 2014.

[36] Y. Feng, Y. Liu, X. Tong, M. Liu, and S. Deng, "Modeling dynamic urban growth using cellular automata and particle swarm optimization rules," *Landscape and Urban Planning*, vol. 102, no. 3, pp. 188–196, 2011.

[37] M. H. Khalilnia, T. Ghaemirad, and R. A. Abbaspour, "Modeling of urban growth using cellular automata (CA) optimized by Particle Swarm Optimization (PSO)," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W3, pp. 231–234, 2013.

[38] J. Liao, L. Tang, G. Shao et al., "A neighbor decay cellular automata approach for simulating urban expansion based on particle swarm intelligence," *International Journal of Geographical Information Science*, vol. 28, no. 4, pp. 720–738, 2014.

[39] A. Tayyebi and B. C. Pijanowski, "Modeling multiple land use changes using ANN, CART and MARS: comparing tradeoffs in goodness of fit and explanatory power of data mining tools," *International Journal of Applied Earth Observation and Geoinformation*, vol. 28, pp. 102–116, 2014.

[40] Z. Tong, "A genetic algorithm approach to optimizing the distribution of buildings in urban green space," *Automation in Construction*, vol. 72, pp. 46–51, 2016.

[41] F. Naghibi, M. R. Delavar, and B. Pijanowski, "Urban growth modeling using cellular automata with multi-temporal

remote sensing images calibrated by the artificial bee colony optimization algorithm," *Sensors*, vol. 16, no. 12, Article ID 2122, 2016.

[42] E. Pérez-Molina, R. Sliuzas, J. Flacke, and V. Jetten, "Developing a cellular automata model of urban growth to inform spatial policy for flood mitigation: a case study in kampala, Uganda," *Computers, Environment and Urban Systems*, vol. 65, pp. 53–65, 2017.

[43] Y. Chen, X. Liu, and X. Li, "Calibrating a land parcel cellular automaton (LP-CA) for urban growth simulation based on ensemble learning," *International Journal of Geographical Information Science*, vol. 31, no. 12, pp. 2480–2504, 2017.

[44] M. K. Jat, M. Choudhary, and A. Saxena, "Application of geospatial techniques and cellular automata for modelling urban growth of a heterogeneous urban fringe," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 20, no. 2, pp. 223–241, 2017.

[45] X. Li, P. Gong, L. Yu, and T. Hu, "A segment derived patch-based logistic cellular automata for urban growth modeling with heuristic rules," *Computers, Environment and Urban Systems*, vol. 65, pp. 140–149, 2017.

[46] X. Liu, X. Liang, X. Li et al., "A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects," *Landscape and Urban Planning*, vol. 168, pp. 94–116, 2017.

[47] Y. Feng and X. Tong, "Calibration of cellular automata models using differential evolution to simulate present and future land use," *Transactions in GIS*, vol. 22, no. 2, pp. 582–601, 2018.

[48] A. Traore, J. Mawenda, and A. Komba, "Land-cover change analysis and simulation in Conakry (Guinea), using hybrid cellular-automata and Markov model," *Urban Science*, vol. 2, no. 2, Article ID 39, 2018.

[49] R. I. Pazos Perez, A. Carballal, J. R. Rabuñal, O. A. Mures, and M. D. García-Vidaurrázaga, "Predicting vertical urban growth using genetic evolutionary algorithms in tokyo's Minato ward," *Journal of Urban Planning and Development*, vol. 144, no. 1, Article ID 04017024, 2018.

[50] X. Fu, X. Wang, and Y. J. Yang, "Deriving suitability factors for CA-markov land use simulation model based on local historical data," *Journal of Environmental Management*, vol. 206, pp. 10–19, 2018.

[51] Y. Feng, Y. Liu, and X. Tong, "Comparison of metaheuristic cellular automata models: a case study of dynamic land use simulation in the Yangtze River delta," *Computers, Environment and Urban Systems*, vol. 70, pp. 138–150, 2018.

[52] C. Liping, S. Yujun, and S. Saeed, "Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques — a case study of a hilly area, Jiangle, China," *PLoS One*, vol. 13, no. 7, Article ID e0200493, 2018.

[53] J. He, X. Li, Y. Yao, Y. Hong, and Z. Jinbao, "Mining transition rules of cellular automata for simulating urban expansion by using the deep learning techniques," *International Journal of Geographical Information Science*, vol. 32, no. 10, pp. 2076–2097, 2018.

[54] F. Yulianto, T. Maulana, and M. R. Khomarudin, "Analysis of the dynamics of land use change and its prediction based on the integration of remotely sensed data and CA-markov model, in the upstream Citarum watershed, West Java, Indonesia," *International Journal of Digital Earth*, vol. 12, no. 10, pp. 1151–1176, 2019.

[55] Y. Lu and P. Wu, "Detection and prediction of land use/land cover change using spatiotemporal data fusion and the cellular automata – Markov model," *Environmental Monitoring and Assessment*, vol. 191, Article ID 68, 2019.

[56] A. A. Devendran and G. Lakshmanan, "Analysis and prediction of urban growth using neural-network-coupled agent-based cellular automata model for Chennai metropolitan area, Tamil nadu, India," *Journal of the Indian Society of Remote Sensing*, vol. 47, no. 9, pp. 1515–1526, 2019.

[57] Y. Huang, B. Yang, M. Wang, B. Liu, and X. Yang, "Analysis of the future land cover change in beijing using CA – Markov chain model," *Environmental Earth Sciences*, vol. 79, Article ID 60, 2020.

[58] H. A. Khawaldah, I. Farhan, and N. M. Alzboun, "Simulation and prediction of land use and land cover change using GIS, remote sensing and CA-markov model," *Global Journal of Environmental Science and Management*, vol. 6, no. 2, pp. 215–232, 2020.

[59] A. Mohamed and H. Worku, "Urban climate simulating urban land use and cover dynamics using cellular automata and Markov chain approach in Addis Ababa and the surrounding," *Urban Climate*, vol. 31, Article ID 100545, 2020.

[60] A. Nurwanda and T. Honjo, "The prediction of city expansion and land surface temperature in bogor city, indonesia," *Sustainable Cities and Society*, vol. 52, Article ID 101772, 2020.

[61] V. Anand and B. Oinam, "Future land use land cover prediction with special emphasis on urbanization and wetlands," *Remote Sensing Letters*, vol. 11, no. 3, pp. 225–234, 2020.

[62] S. Mansour, M. Al-belushi, and T. Al-awadhi, "Land use policy monitoring land use and land cover changes in the mountainous cities of Oman using GIS and CA-markov modelling techniques," *Land Use Policy*, vol. 91, Article ID 104414, 2020.

[63] I. Santé, A. M. García, D. Miranda, and R. Crecente, "Cellular automata models for the simulation of real-world urban processes: a review and analysis," *Landscape and Urban Planning*, vol. 96, no. 2, pp. 108–122, 2010.

[64] G. Iovine, D. D'Ambrosio, and S. Di Gregorio, "Applying genetic algorithms for calibrating a hexagonal cellular automata model for the simulation of debris flows characterised by strong inertial effects," *Geomorphology*, vol. 66, no. 1-4, pp. 287–303, 2005.

[65] F. Semboloni, "The dynamic of an urban cellular automata in A 3-D spatial pattern," in *Proceedings of the XXI National Conference, Aisre: Regional and Urban Growth in a Global Market*, Palermo, 2000.

[66] W. Shi and M. Y. C. Pang, "Development of Voronoi-based cellular automata -an integrated dynamic model for Geographical Information Systems," *International Journal of Geographical Information Science*, vol. 14, no. 5, pp. 455–474, 2000.

[67] D. O'Sullivan, "Graph-cellular automata: a generalized discrete urban and regional model," *Environment and Planning B: Planning and Design*, vol. 28, pp. 687–705, 2001.

[68] D. Stevens and S. Dragićević, "A GIS-based irregular cellular automata model of land-use change," *Environment and Planning B: Planning and Design*, vol. 34, no. 4, pp. 708–724, 2007.

[69] H. Couclelis, "Cellular worlds: a framework for modeling micro-macro dynamics," *Environment & Planning A*, vol. 17, no. 5, pp. 585–596, 1985.

[70] R. White and G. Engelen, "Cellular automata and fractal urban form: a cellular modelling approach to the evolution of

urban land-use patterns," *Environment & Planning A*, vol. 25, no. 8, pp. 1175–1199, 1993.

[71] C. Webster and F. Wu, "Coase, spatial pricing and self -organising cities," *Urban Studies*, vol. 38, no. 11, pp. 2037–2054, 2001.

[72] M. C. Polidori and R. Krafta, "Simulando crescimento urbano com integração de fatores naturais, urbanos e institucionais," *GeoFocus*, vol. 5, pp. 156–179, 2005.

[73] S. Geertman, M. Hagoort, and H. Ottens, "Spatial-temporal specific neighbourhood rules for cellular automata land-use modelling," *International Journal of Geographical Information Science*, vol. 21, no. 5, pp. 547–568, 2007.

[74] X. Li, Q. Yang, and X. Liu, "Discovering and evaluating urban signatures for simulating compact development using cellular automata," *Landscape and Urban Planning*, vol. 86, no. 2, pp. 177–186, 2008.

[75] K. C. Clarke, S. Hoppen, and L. Gaydos, "A self-modifying cellular automaton model of historical urbanization in the san francisco bay area," *Environment and Planning B: Planning and Design*, vol. 24, no. 2, pp. 247–261, 1997.

[76] M. Phipps and A. Langlois, "Spatial dynamics, cellular automata, and parallel processing computers," *Environment and Planning B: Planning and Design*, vol. 24, no. 2, pp. 193–204, 1997.

[77] H. Couclelis, "From cellular automata to urban models: new principles for model development and implementation," *Environment and Planning B: Planning and Design*, vol. 24, no. 2, pp. 165–174, 1997.

[78] P. M. Torrens and I. Benenson, "Geographic automata systems," *International Journal of Geographical Information Science*, vol. 19, no. 4, pp. 385–412, 2005.

[79] M. M. Aburas, Y. M. Ho, M. F. Ramli, and Z. H. Ash'aari, "The simulation and prediction of spatio-temporal urban growth trends using cellular automata models: a review," *International Journal of Applied Earth Observation and Geoinformation*, vol. 52, pp. 380–389, 2016.

[80] N. Bihamta, A. Soffianian, S. Fakheran, and M. Gholamalifard, "Using the SLEUTH urban growth model to simulate future urban expansion of the isfahan metropolitan area, Iran," *Journal of the Indian Society of Remote Sensing*, vol. 43, no. 2, pp. 407–414, 2014.

[81] M. Mohammad, A. Sahebgharani, and E. Malekipour, "Urban growth simulation through cellular automata (CA), analytic Hierarchy process (AHP) and GIS, case study of 8th and 12th municipal districts of isfahan," *Geographia Technica*, vol. 18, no. 2, pp. 57–70, 2013.

[82] S. Park, S. Jeon, S. Kim, and C. Choi, "Prediction and comparison of urban growth by land suitability index mapping using GIS and RS in South Korea," *Landscape and Urban Planning*, vol. 99, no. 2, pp. 104–114, 2011.

[83] S. Q. Wang, X. Q. Zheng, and X. B. Zang, "Accuracy assessments of land use change simulation based on markov-cellular automata model," *Procedia Environmental Sciences*, vol. 13, pp. 1238–1245, 2012.

[84] B. H. Aithal and S. V. T. Ramachandra, "Landscape dynamics modeling through integrated Markov, fuzzy-AHP and cellular automata," in *Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium*, pp. 3160–3163, Quebec, Canada, October 2014.

[85] A. A. A. Al-sharif and B. Pradhan, "Spatio-temporal prediction of urban expansion using bivariate statistical models: assessment of the efficacy of evidential belief functions and frequency ratio models," *Applied Spatial Analysis and Policy*, vol. 9, no. 2, pp. 213–231, 2016.

[86] N. Behboudi, F. Butt, and A. Abhari, "Automatic building design with genetic algorithms and artificial neural networks (WIP)," in *Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design*, vol. 7, Society for Computer Simulation International, Orlando, FL, USA, March 2012.

[87] A. Gaspar-Cunha, D. Loyens, and F. Van Hattum, "Aesthetic design using multi-objective evolutionary algorithms," in *Proceedings of the International Conference on Evolutionary Multi-Criterion Optimization*, pp. 374–388, Springer, Ouro Preto, Brazil, April 2011.

[88] C. T. Mueller and J. A. Ochsendorf, "Combining structural performance and designer preferences in evolutionary design space exploration," *Automation in Construction*, vol. 52, pp. 70–82, 2015.

[89] C. Ugurlu, I. Chatzikonstantinou, S. Sariyildiz, and M. F. Tasgetiren, "Evolutionary computation for architectural design of restaurant layouts," in *Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC)*, pp. 2279–2286, IEEE, Sendai, Japan, May 2015.

[90] I. G. Dino, "An evolutionary approach for 3D architectural space layout design exploration," *Automation in Construction*, vol. 69, pp. 131–150, 2016.

[91] Z. Guo and B. Li, "Evolutionary approach for spatial architecture layout design enhanced by an agent-based topology finding system," *Frontiers of Architectural Research*, vol. 6, no. 1, pp. 53–62, 2017.

[92] A. De Almeida, B. Taborda, F. Santos, K. Kwiecinski, and S. Eloy, "A genetic algorithm application for automatic layout design of modular residential homes," in *Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 002774–002778, IEEE, Budapest, Hungary, October 2016.

[93] B. Taborda, A. De Almeida, F. Santos, S. Eloy, and K. Kwiecinski, "Shaper-G. A.: Automatic shape generation for modular house design," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 937–942, ACM, Kyoto Japan, July 2018.

[94] M. L. Castro, A. Carballal, N. Rodriguez-Fernandez, I. Santos, and J. Romero, "Artificial intelligence applied to conceptual design. A review of its use in architecture," *Automation in Construction*, vol. 134, Article ID 103550, 2021.

[95] M. Batty, H. Couclelis, and M. Eichen, "Urban systems as cellular automata," *Environment and Planning B: Planning and Design*, vol. 24, no. 2, pp. 159–164, 1997.

[96] M. Rickert, K. Nagel, M. Schreckenberg, and A. Latour, "Two lane traffic simulations using cellular automata," *Physica A: Statistical Mechanics and Its Applications*, vol. 231, no. 4, pp. 534–550, 1996.

[97] D. E. Wolf, "Cellular automata for traffic simulations," *Physica A: Statistical Mechanics and Its Applications*, vol. 263, no. 1–4, pp. 438–451, 1999.

[98] S. F. Hafstein, R. Chrobok, A. Pottmeier, M. Schreckenberg, and F. C. Mazur, "A high-resolution cellular automata traffic simulation model with application in a freeway traffic information system," *Computer-Aided Civil and Infrastructure Engineering*, vol. 19, no. 5, pp. 338–350, 2004.

[99] B. Wang and F. Qian, "Three dimensional gas dispersion modeling using cellular automata and artificial neural network in urban environment," *Process Safety and Environmental Protection*, vol. 120, pp. 286–301, 2018, ISSN 0957-5820.

[100] F. Wu and C. J. Webster, “Simulation of land development through the integration of cellular automata and multicriteria evaluation,” *Environment and Planning B: Planning and Design*, vol. 25, no. 1, pp. 103–126, 2018.

[101] D. O’Sullivan, “Toward micro-scale spatial modeling of gentrification,” *Journal of Geographical Systems*, vol. 4, no. 3, pp. 251–274, 2002.

[102] A. Nara and P. M. Torrens, “Inner-city gentrification simulation using hybrid models of cellular automata and multiagent systems,” in *Proceedings of the Geocomputation 2005 Conference*, University of Michigan, Michigan, August 2005.

[103] P. M. Torrens and A. Nara, “Modeling gentrification dynamics: a hybrid approach,” *Computers, Environment and Urban Systems*, vol. 31, no. 3, pp. 337–361, 2007.

[104] M. Batty, “Cellular automata and urban form: a primer,” *Journal of the American Planning Association*, vol. 63, no. 2, pp. 266–274, 1997.

[105] F. Wu, “Calibration of stochastic cellular automata: the application to rural-urban land conversions,” *International Journal of Geographical Information Science*, vol. 16, no. 8, pp. 795–818, 2002.

[106] E. A. Silva and K. C. Clarke, “Complexity, emergence and cellular urban models: lessons learned from applying SLEUTH to two Portuguese metropolitan areas,” *European Planning Studies*, vol. 13, no. 1, pp. 93–115, 2005.

[107] K. R. Dahal and T. E. Chow, “Characterization of neighborhood sensitivity of an irregular cellular automata model of urban growth,” *International Journal of Geographical Information Science*, vol. 29, no. 3, pp. 475–497, 2015.

[108] P. M. Torrens, *How Cellular Models of Urban Systems Work”, WP-28*, Centre for Advanced Spatial Analysis (CASA), University College, London, ISSN: 1467-1298, 2000.

Hindawi
Complexity
Volume 2022, Article ID 5912696, 11 pages
https://doi.org/10.1155/2022/5912696

Research Article

Spatial Pattern and Evolution of Global Innovation Network from 2000 to 2019: Global Patent Dataset Perspective

Yuna Di,[1] Yi Zhou,[2] Lu Zhang ,[1] Galuh Syahbana Indraprahasta ,[3] and Jinjin Cao[2]

[1]*School of International Economics and Management, Beijing Technology and Business University, Beijing 100048, China*
[2]*School of Economics, Beijing Technology and Business University, Beijing 100048, China*
[3]*Research Center for Population, National Research and Innovation Agency Indonesia (BRIN), Jakarta, Indonesia*

Correspondence should be addressed to Lu Zhang; zhanglu113@mails.ucas.ac.cn

Received 9 February 2022; Revised 3 May 2022; Accepted 18 May 2022; Published 10 June 2022

Academic Editor: Haitao Ma

In the era of the knowledge economy, the improvement of national innovation systems is playing a significant role in the global entrepreneurship ecosystem. Entrepreneurs are accelerating international intellectual property applications to be competitive. What remains to be explored is the evolution of international intellectual property network in the globe. With the application of social network analysis and intellectual property application database, the global innovation network structure from 2000 to 2019 is explored. Results showed that (1) in the period 2000–2019, the global innovation network has been expanding rapidly from a sparse network to a dense and complex one. (2) Patent application is unevenly distributed in the globe. Countries such as the US, China, and Canada have been the top countries flowing in, while Japan, Korea, EU, and Switzerland have been the main countries flowing out. (3) Global innovation network shows an obvious "core-periphery" pattern. The distribution pattern presents a quadrilateral structure with the four core regions of "US, Japan, EU, and China" as the apex. This analysis contributes to the visualization of the global layout of intellectual property and the evolution trend by analyzing intellectual property application networks. This can provide important experience reference for enterprises to study the global entrepreneurship ecosystem.

1. Introduction

Innovation is becoming the key driver for the survival and growth of firms in emerging and developed countries. With the continuous development of technology and the deep division of the global value chain, the development strategy of enterprises is entering an era in which growth is driven by knowledge innovation. For nations, innovation is far more important than land, physical capital, or labor in a knowledge-based economy, and is the dominating factor affecting different economic growth and development. As a result, the effective management of knowledge and innovation has gained increasing interest in the public debate, calling upon the contribution of scholars and practitioners [1]. One of the important views is that innovations never occur in isolation but on the contrary generated by networks of interacting actors (e.g., organisations, multinational enterprises, and individuals) [2]. For instance, multinational enterprises engage in overseas innovation for adapting products to meet the unique requirements in foreign markets [3]. Thus, inbound and outbound flows of knowledge activated by the stakeholders' interactions within a country make today's innovation ecosystem is highly globally interlinked [4].

In the era of the knowledge economy, intellectual property has become the core innovation factor, infiltrating into all fields of international trade. The identification of the main actors involved in regions has been revised across the years by, most recently, including the emerging markets [5, 6]. The concept of an innovation network was first proposed by Ernst [7] after he studied the internal relationship between regional production and global production. Since then, increasing studies have started related research, such as topological properties [8–10], spatial pattern [11], evolution process [12, 13], and mechanisms. The previous studies generally revealed the characteristics of innovation network, such as scale-free [8, 9], hierarchical [8, 9], and spatial agglomeration [12, 13]. It is also widely verified that cognitive proximity, social proximity, organizational proximity, institutional proximity, and geographic proximity are important factors affecting the evolution of innovation networks [14–18]. However, in view of the research scale, these studies have been conducted on either country or region scale because of the data availability [19, 20]. As the importance of interactions between different national innovation systems, the question arises that how nations perform in the global innovation network. Understanding knowledge network from a global view and monitoring the network evolution change in a global scale is necessary for enhancing innovation capability.

The intellectual property is an important driving force to encourage entrepreneurship, and an important means to maintain entrepreneurial achievements and promote the sustainable development of enterprises [21]. The intellectual property network has gradually become the main way to study regional knowledge spillover, innovation and technology diffusion, regional development path, innovation cluster, and other practical problems. The improvement of international intellectual property system, which is led by World Intellectual Property Organization (WIPO), is playing a significant role in the construction of global entrepreneurship ecosystem. WIPO provides us with a novel approach to obtaining innovation flows in global scale. Thus, the aim of this study is to explore how the structure of globalization of technology via intellectual property networks has changed longitudinally. The identification of the main countries in the global innovation network can not only describe the innovation performance and development strategy of each country but also can help to figure out the trend of global economy.

The rest content is divided into four sections. The first section briefly describes the data collection and method applied in this analysis. The results of properties, "core-periphery" pattern of global innovation network during 2000–2019 will be presented in the third section. The fourth section discusses the main findings and potential limitations outlined in this section. Following this section, an overview of potential avenues for further research is presented. This analysis will contribute to current studies by visualizing the global layout of intellectual property, which provides references for enterprises to study the global entrepreneurship ecosystem.

2. Methodology

2.1. Data Source and Network Generation. In this study, we use a comprehensive dataset from WIPO to investigate the structure and dynamics of global innovation network from 2000 to 2019. Innovation has been the primary driving force for development in the past decades. In the twenty-first century, especially, globalization, information and intelligent development, which are inseparable from technological innovation, have provided large support for the world economy. In order to explore the evolution of global innovation since the 21th century, the year 2000 is chosen as the starting year while the year 2019 is chosen as the ending year due to the emergence of COVID-19 in 2020.

WIPO, which manages the international application system and the global intellectual property system, is an intellectual property management agency affiliated with the United Nations. Compared with other intellectual property databases, it has a long history (since 1967) with a large volume of data (more than 55 million patent records) and a large number of member countries (193 member countries). It provides free intellectual property information as well as patents from the African Regional Intellectual Property Organization (ARIPO), the Eurasian Patent Organization (EAPO), and the European Patent Office (EPO) and regional patent office's records. Furthermore, the WIPO dataset can provide the information on filing office and application's origin which is basic for constructing networks in our analysis. Therefore, the dataset of the WIPO Intellectual Property Statistics Data Center (https://www3.wipo.int/ipstats) is applied in this analysis. Based on the availability and reliability of data, this analysis selects 203 countries and regions as the research objects. The gross domestic product (GDP) and patent applications of these countries and regions account for more than 90% of the global value, which can represent the basic situation of the global innovation cooperation network.

In the process of data collection, this analysis selects the "total number of patent applications (direct application and entry into the PCT national phase)" instead of granted applications. Due to the complicated process of application and differences between countries, the approval process of one patent might take years. For this reason, the number of granted applications cannot reflect the latest technological development and changes in the field of innovation. The report type selection is counted by the filing office and the origin of the applicant. In order to analyze the evolution of the global innovation cooperation network, data from 2000 to 2019 are collected and applied in this analysis. Besides, there are some missing data in individual countries in some years. To minimize the impact of missing data, the processing method of this analysis is (1) to supplement the missing data with interpolation; (2) to average all the data every five years. In addition, this analysis merges the countries that joined the European Union (EU, hereafter) that year into the EU. Finally, this analysis constructs a patent data matrix for four periods: 2000–2004, 2005–2009, 2010–2014, and 2015–2019. According to the selected countries identified above, there were 185 nodes in the global innovation network in 2000–2004, 175 nodes in 2005–2009, 173 nodes in 2010–2014, and 172 nodes in 2015–2019. The change in the number of nodes is due to the accession of some countries to the EU during the study period. Thus, directed adjacent matrices of $185 * 185$, $175 * 175$, $173 * 173$, and $172 * 172$ were constructed based on the strength of patent edges (Figure 1). The values in the matrix represent the number of patents between the two countries, and the matrix is directed.

2.2. Indicators of Network Analysis. By constructing a weighted network, we explore the direction and strength of the connection between countries. This analysis adopts a

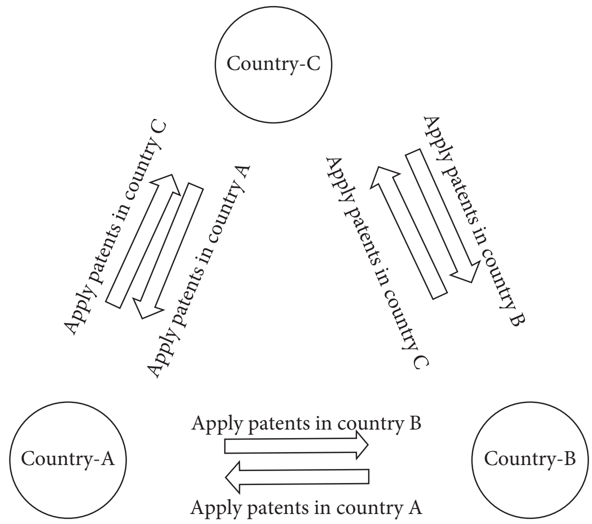


Figure 1: Construction of innovation network with patent.

directed weighted network, which can reasonably evaluate the spatial structure characteristics of innovation network.

2.2.1. Key Indicators. This analysis aims to analyze the network structure and evolution of the global intellectual property network (GIPN) from network scale, small-world property, scale-free property, centrality properties' views. In general, $G = (V, E)$ stands for a directed network, V is the set of nodes while E is the set of edges. Density and network diameter are used to indicate the scale of the GIPN. The value of density refers to the ratio of the number of actual network relationships to the maximum number of possible relationships. The network diameter is the longest path of any two nodes in the entire network. The specific calculation formulas are

$$D = \frac{2M}{N(N-1)},$$
$$R = \max_{\forall v_i, v_j \epsilon V}\left(r\left(v_i, v_j\right)\right), \tag{1}$$

where D refers to the network density, R refers to the network diameter, M refers to the number of relationships existing in the actual network, and N represents the number of nodes in the network. v_i, $v_j \in V$, $r(v_i, v_j)$ represents the longest path between nodes i and j.

The average clustering coefficient and average path length are used to analyze small-world properties of the GIPN. The average cluster coefficient is the mean of the clustering coefficient of all nodes. The cluster coefficient of a node is the density of its open neighborhood [22]. The neighborhood (N_i) of node v_i is the set of its adjacent nodes. N_i is defined as

$$N_i = \left\{v_j\colon e_{ij} \epsilon E \vee e_{ji} \epsilon E\right\}. \tag{2}$$

Cluster coefficient is calculated by means of the following formula:

$$C_i = \frac{\left|\left\{e_{jk}\colon v_j, v_k \epsilon N_i, e_{jk} \epsilon E\right\}\right|}{N(N-1)}, \tag{3}$$

where C_i refers to the cluster coefficient of v_i. Average clustering coefficient (C) is calculated as

$$C = \frac{1}{N}\sum_{i=1}^{n} C_i. \tag{4}$$

The average path length is the average number of edges along the shortest path for all possible pairs of nodes.

$$L = \frac{1}{N(N-1)}\sum_{i,j} d\left(v_i, v_j\right), \tag{5}$$

where L refers to the average path length and $d(v_i, v_j)$ represents the shortest path between nodes i and j.

Centrality measurements are used to identify nodes that are most critical and central in networks. Among which, degree, betweenness, and closeness centralities are the most widely used indexes. Degree centrality is calculated by means of the following formula:

$$DC_i = \sum_{j=1}^{N} w_{ij}. \tag{6}$$

The degree centrality can be calculated according to weighted network data or binary network data. In this analysis, weighted data are used to calculate the degree centrality of each node. w_{ij} is the number of patents between country i and j.

Closeness centrality refers to the mean geodesic path between nodes in networks. Here, the geodesic path is defined as the number of edges traversed from node i to j. Thus, a node with a high closeness centrality indicates a short communication edge to other nodes in networks. In general, geodesic paths are not unique, as there can be several paths between two given nodes with the same shortest length. However, at least one geodesic path always exists between any two nodes in the same connected component of a network. The mean geodesic distance between i and all other nodes in the network is given by

$$g_i = \frac{1}{N}\sum_{j=1}^{n} \gamma_{ij}. \tag{7}$$

Here, γ_{ij} refers to the number of edges traversed from node i to j. Then, the closeness centrality CC_i of node i is defined as follows:

$$CC_i = \frac{1}{g_i} = \frac{N}{\sum_{j=1}^{n} \gamma_{ij}}. \tag{8}$$

Another notion of centrality is betweenness centrality, which measures the number of short paths between nodes while they pass through a given node. Thus, we define

$$v_i(s,t) = \begin{pmatrix} 1, & \text{if } i \text{ lies on the geodesic path from } s \text{ to } t \\ 0, & \text{otherwise,} \end{pmatrix}. \tag{9}$$

Then, the betweenness centrality BC_i of node i is

$$BC_i = \sum_{s,t \in N} v_i(s,t). \quad (10)$$

DC_i and CC_i analyze the properties of nodes. Graph centrality is used when the focus is on the whole network. This analysis uses graph degree centrality and graph betweenness centrality to analyze the centrality trend of the whole network [23]. The specific calculation formulas are

$$DC = \frac{\sum_{i=1}^{N}(DC_{\max} - DC_i)}{N^2 - 3N + 2},$$
$$BC = \frac{\sum_{i=1}^{N}(BC_{\max} - BC_i)}{N^3 - 4N^2 + 5N - 2}, \quad (11)$$

where DC and BC refer to the graph degree centrality and graph betweenness centrality, and $DC_{\max}$ and $BC_{\max}$ represent the maximum value of DC_i and BC_i.2.2.2. In-Out Flow

The in-out flow is calculated based on out-flow and in-flow. The out-flow is defined as

$$O_i = \sum_{i \neq j}^{n} S_{ij}, \quad (12)$$

where S_{ij} refers to the number of patents flowing from country i to country j. O_i refers to the number of patents flowing out of country i. The in-flow is defined as

$$I_i = \sum_{i \neq j}^{n} S_{ij}', \quad (13)$$

where S_{ij}' refers to the number of patents filed by country i in country j. I_i refers to the total number of patents filed by country i. Then, we compare the out-flow and in-flow, which is referred as in-out flow.

$$\text{In} - \text{out flow}_i = I_i - O_i. \quad (14)$$

There are two possible results: (i) $\text{In} - \text{out flow}_i > 0$, which presents country i has more in-flow than out-flow, indicating the number of patents flowing into country i is higher than the number of patents flowing out. (ii) $\text{In} - \text{out flow}_i < 0$, which presents country i has more in-flow than out-flow, indicating that the number of patents flowing out from country i is higher than the number of patents flowing in.

2.3. Coreness. There is often a core edge structure in the network, so the coreness is introduced to quantitatively study the status of each node in the network, and have a quantitative understanding of where the node is (core, semicore, and periphery). When calculating the coreness, each node needs to be given a coreness c_i. The coreness is calculated by the following steps:

$$\delta_{ij} = c.c^T. \quad (15)$$

δ_{ij} is the pattern matrix of the network data matrix W_{ij} to be analyzed. c is the eigenvector, and c^T is its transpose vector. The core-periphery analysis method finds the eigenvector c, which can make the correlation coefficient maximum between the actual matrix W_{ij} and the pattern matrix δ_{ij}. The element c_i in eigenvector c is the coreness of each node in the network. This indicator is achieved by the Ucinet platform.

In order to construct a continuous core-periphery model, nodes in the GIPN are divided into the following four layers: nodes with a coreness greater than 0.2 belong to core layer; nodes with a core degree between 0.01 and 0.2 are semicore countries, nodes with a coreness between 0 and 0.01 are semicore countries, and countries with zero coreness are considered as peripheral nodes.

3. Results

3.1. Key Indicators. Table 1 presents the summary statistics of complexity of the GIPN from 2000 to 2019. The following observations can be gained.

(1) The density and the number of edges increase, and the network becomes denser. From 2000 to 2014, the scale of the GIPN expanded rapidly, and the number of nodes in the network declined. This is due to the increasing number of EU countries in the sample. At the same time, the number of edges expanded rapidly, from 1772 in 2000–2004 to 3070 in 2015–2019. The network density increased rapidly from 0.0521 to 0.1044, indicating that the global urban innovation network has gradually grown from a sparse network to a dense and complex network.

(2) In view of the centrality indicators, the weighted degree centrality increased from 8.14 to 15.08. The average intermediary centrality of the GIPN has gradually increased, indicating that some countries have gradually increased their ability to control and deliver innovation. Besides, closeness centrality increased from 2.87 to 7.23, indicating that the relationship between countries and the core countries of the innovation network is getting closer.

(3) Both degree centrality and betweenness centrality have large Gini coefficients and coefficients of variation, and the nodes in the network are very polarized. The Gini coefficient of degree centrality is relatively large, remaining above 0.88, indicating that the GIPN exhibits a strong discrete trend and extremely unbalanced characteristics.

(4) In view of the evolution over the years, the coefficient of variation of degree and betweenness centrality shows a downward trend, indicating that the discrete trend of the GIPN has declined.

3.2. Global Distribution. In this section, the topological relationship is converted to spatial connections with the application of the ArcGIS platform. The difference between in-flow and out-flow in 2000–2019 is visualized in Figure 2, and three observations were gained.

Table 1: Complexity statistics of global intellectual property network from 2000 to 2019.

Statistical characterstics	Index	2000–2004	2005–2009	2010–2014	2015–2019
Network scale	Number of nodes	185	175	173	172
	Number of edges	1772	1791	2650	3070
	Density	0.0521	0.0588	0.0891	0.1044
	Diameter	4	3	3	4
Small-word property	Average clustering coefficient	0.76	0.785	0.752	0.757
	Average path length	1.918	1.828	1.823	1.804
	Power law fitting of degree centrality	$y = 1618.2X^\wedge - 1.313$	$y = 1415.3x^\wedge - 1.271$	$y = 1965.8x^\wedge - 1.246$	$y = 1972.7x^\wedge - 1.216$
Scale-free property	$R2$	0.7984	0.7871	0.7405	0.7262
	Exponential fitting of degree centrality	$y = 88.082e^\wedge - 0.031x$	$y = 90.871e^\wedge - 0.032x$	$y = 125.85e^\wedge - 0.029x$	$y = 138.6e^\wedge - 0.029x$
	$R2$	0.9895	0.9847	0.983	0.9807
Degree centrality	Average degree centrality	8.143	9.051	13.503	15.028
	Graph degree centrality	0.539	0.577	0.61	0.641
	Coefficient of variation	1.487	1.443	1.19	1.135
	Gini coefficient	0.961	0.96	0.955	0.954
Betweenness centrality	Average betweenness centrality	0.369	0.379	0.473	0.458
	Graph betweenness centrality	0.143	0.144	0.138	0.157
	Coefficient of variation	4.615	4.403	3.677	3.852
	Gini coefficient	0.887	0.941	0.916	0.91
Closeness centrality	Average closeness centrality	2.872	2.049	7.213	7.23
	Coefficient of variation	0.0063	0.0066	0.0177	0.0177
	Gini coefficient	0.003	0.004	0.01	0.01

First, in view of the overall distribution pattern of the GIPN, countries that were most deeply embedded in GIPN include the US, China, Korea, Japan, EU, Switzerland, and Canada, as shown in Figure 2. Specifically, the US, China, and Canada have been the top countries with large in-flow patent and small out-flow patents during 2000–2004. Since then, China and the US have jointly occupied important positions in the GIPN. Besides, Japan, Korea, EU, and Switzerland have been the top countries with large out-flow patents and small in-flow patents. Second, the transferring direction has been concentrated among these top cities. For instance, the in-flow and out-flow patents among the USA, Japan, China, and Korea occupied more than 60% of the total. Besides, the transferring direction of patented technology is mainly east-west in the northern hemisphere. Third, increasing in-flows in Asian countries can be observed. These Asian countries include Viet Nam, Thailand, Malaysia, Singapore, and Indonesia.

3.3. In-Out Flow of the Global Intellectual Property Network. For our interests in the positions of a country, the countries' position is overlaid by differencing in-flow and out-flow. The analysis of ranking countries in the GIPN was conducted in the last 20 years periods from 2000 to 2019. In view of the in-flow, our results point to the evolution of countries' positions in the GIPN (Figure 3). The first point to make here is that the primary three countries, including the US, Japan, and China (Figure 3, dark lines), have remained the top three positions in the past years. Specifically, the total in-flow in the three countries share more than 52% of the total volume. Moreover, the concentration of in-flow in the three countries has been increasing. Specifically, the in-flow volume accounts for 52.2% during 2000–2014 while the percentage has increased to 63.6% during 2015–2019. This indicates that the three countries possess high position relative to their economic and innovation potential in the GIPN. The second point is that some countries have raised their position largely. These countries include Korea, India, and Indonesia reflecting the trend from the low rank to the high rank in the GIPN. The third point is that some countries, including Canada, EU, and Brazil, rank from higher to lower positions.

In view of the out-flow, our results point to the evolution of countries' positions in the GIPN (Figure 4). The first point to make here is that the primary three countries, including EU, Japan, and the US, have remained in the top three positions in the past years. Specifically, the total in-flow in the three countries shares more than 49% of the total volume. The concentration of the out-flow in the three countries has been decreasing. Specifically, the out-flow volume account for 51.7% during 2000–2014 while the percentage has decreased to 49.6% during 2015–2019. This

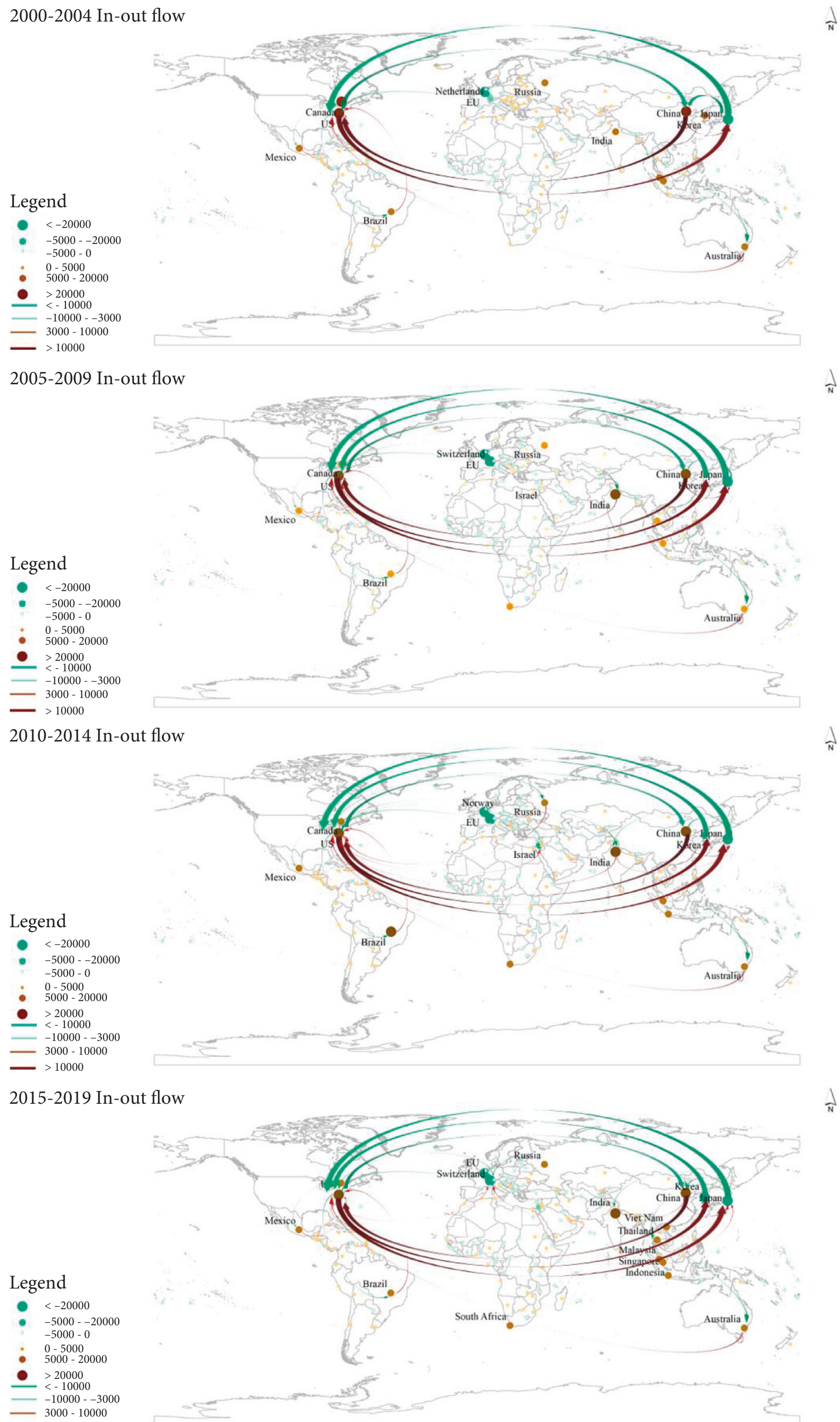


Figure 2: In-out flows of countries in 2000–2019.

indicates that the three countries possess high positions relative to their economic and innovation potential in the GIPN. The second point is that some countries have raised their position largely. These countries include China, India, Singapore, and Kazakhstan, reflecting the trend from the low rank to the high rank in the GIPN. The third point is that some countries, including Switzerland, Canada, and New Zealand, rank from higher to lower positions.

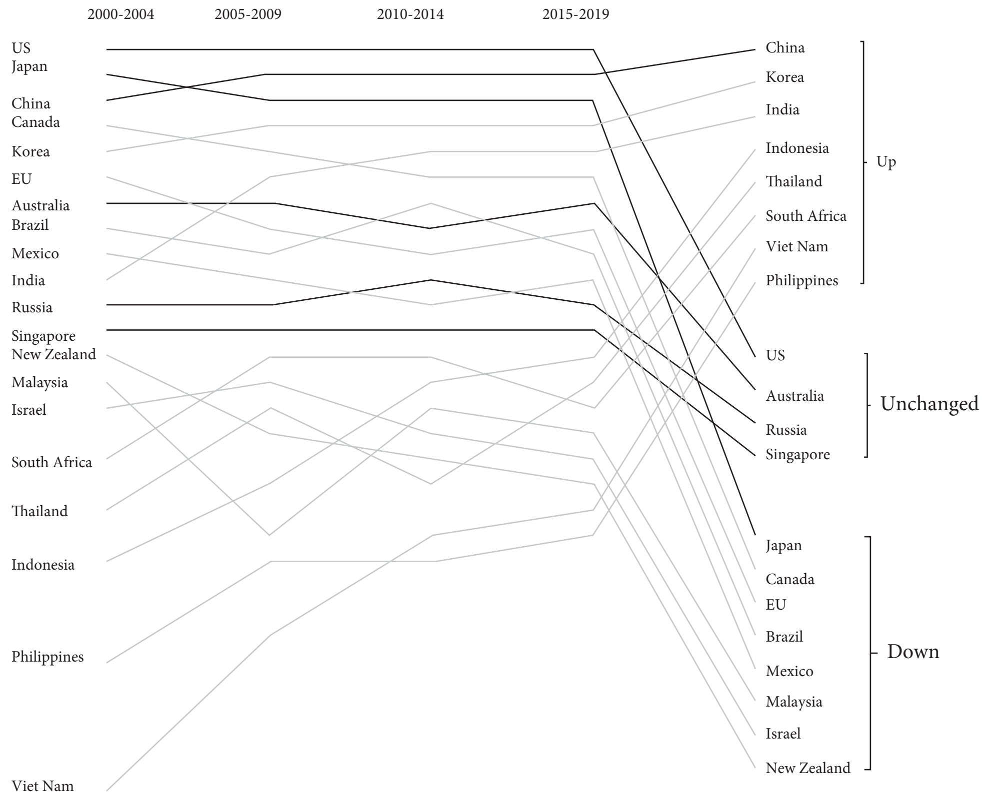


Figure 3: Positions of countries in in-flow.

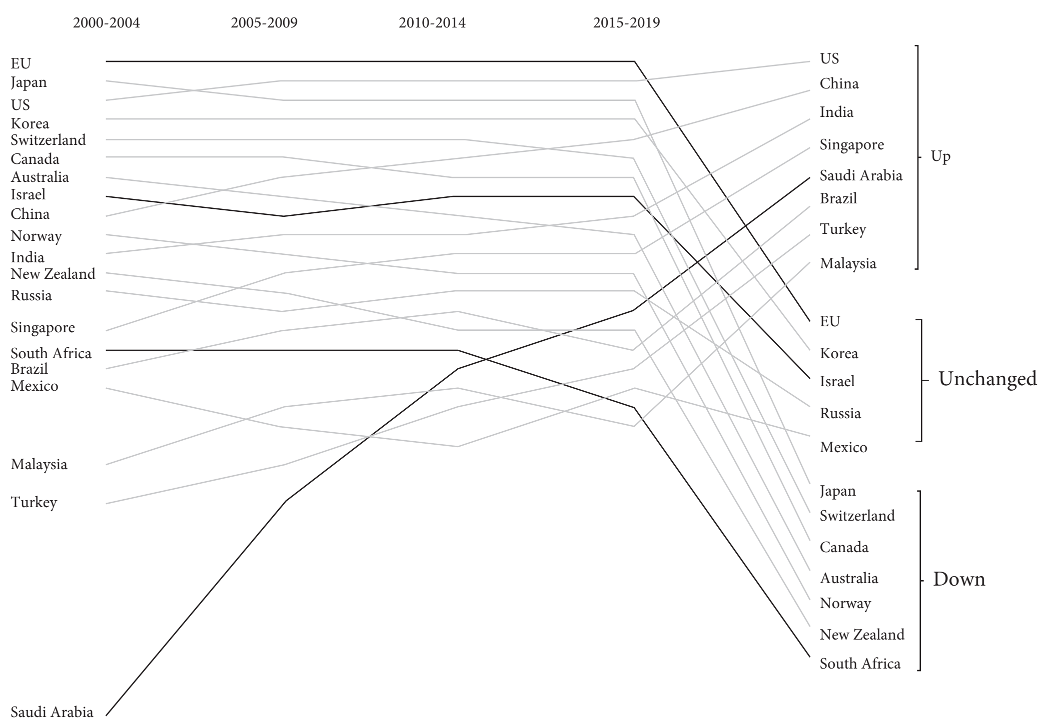


Figure 4: Positions of countries in-out flow.

3.4. Core and Periphery of the Global Intellectual Property Network. The centrality is an important indicator to measure the degree of centralization of the entire network. From 2000 to 2019, the degree centralization of the GIPN was basically maintained at about 0.6, and the closeness centralization was basically maintained at about 0.65, indicating that the entire network has a relatively obvious direction, indicating a significant core-periphery structure with hierarchies in the GIPN (Figure 5). In sum, the following three observations can be gained.

(1) In the core layer, the GIPN gradually developed from a triangular structure to a quadrilateral pattern. Specifically, the coreness of the four periods from 2006 to 2010 has always been higher than 0.8, becoming the absolute core of the network. EU and Japan have far lower patent transfers than the US. As a result, the core countries from 2000 to 2009 are the US, Japan, and the EU. China has performed well, and its core degree has always increased, from 0.105 in 2000–2004 to 0.262 in 2015–2019, further narrowing the gap with the EU. As a result, China ranks among the core countries of the innovation network.

(2) In the semicore (second) layer, it has maintained a relatively stable state. Specifically, Australia, Brazil, Canada, China, India, Israel, Mexico, Korea, Russia, Singapore, and Switzerland have stayed in the semicore layer.

(3) The innovation network presents an obvious phenomenon of cooperation aggregation-technical cooperation within core countries, between core and semicore countries, is intensive, while the participation of peripheral countries in the GIPN is limited.

In view of the patent transfer among layers, three observations can be observed (Table 2). First, in sum, patent within the core layer and between the core and semicore layer occupies more than 87% of the total patent. Second, the transfer of patents at the core layer is more concentrated. In the two phases of 2000–2004 and 2005–2009, the proportion of patent transfers at the core layer composed of the US, EU, and Japan was 39.45% and 34.34%, respectively, which decreased somewhat. From 2010 to 2019, due to China's entry into the core layer, the proportion has increased sharply to 49.76% and 50.03%, respectively. Third, the technology spillover effect from core layer to semicore is obvious, manifesting in the increment in patent transfer from 10.07% to 15%. The technology transfer from semicore to core layer has dropped significantly, and the number of patent transfers has increased from 39.56% to 22.4%.

4. Discussion

The 21st century has seen rapid development and wide application of various emerging technologies such as the Internet, big data, cloud computing, artificial intelligence, and blockchain. The production, search, dissemination, and application of knowledge have effectively broken through the limitations of geographical distance. Innovation activities have expanded from within the organization to cross-organizational and cross-regional networks, and gradually evolved into a global network. As a result, the scale of the GIPN expanded rapidly from a sparse network to a dense and complex network. As the number of global patent cooperation increase continually, the patent cooperation network shows obvious small-world characteristic, and the integration of countries is high, which helps to obtain new information and new resources, and strengthen patent cooperation and innovation among countries.

Our results also suggest some important implications. First, the patents dataset is useful for exploring the global trends of technological diffusion. The strength of network analysis is that it describes the relationships among countries. In this study, network analysis presented not only which countries have higher technological capabilities but also how countries are mutually connected for technological collaboration or transfer. Since 2000, the GIPN has gradually lost its scale-free feature, and the small-world feature has been continuously strengthened [24].

Second, intellectual property is an important indicator to forecast global investment flow and entrepreneurial environment. As innovation shows the property of clustering, which provides an important driving force and encouragement for entrepreneurship [25]. In a long term, the pattern of the GIPN has presented a triangle structure with the three core regions of "US-EU-Japan." The triangle structure accumulated a large number of patents. Since 2005, the pattern of the GIPN has presented a diamond structure with the four core regions of the US, EU, Japan, and China, as the apex. Although the US, EU, and Japan have been the core of the GIPN, Asian countries have gradually improved their status in the cooperative innovation network and gradually entered the core layer [26, 27]. Patent plays a key role in transferring innovations and changing the social, economic, and political system on a global level. Through innovative production and technological cooperation, countries gained chances to surpass their original innovative production network system and achieve a certain degree of leapfrog development. Existing studies have shown that emerging marketing countries' positions in the GIPN have been raised gradually [27–29].

Third, the GIPN has been structured as a core-peripheral structure. During the period of 2000–2009, the US has been the absolute core in the GIPN. Our analysis also presents an interesting observation that the countries' positions in the GIPN have been relatively stable. Specifically, the US, EU, Japan, China, Australia, Brazil, Canada, China, India, Israel, Mexico, Korea, Russia, Singapore, and Switzerland have remained in the top positions in the GIPN. Enterprises, especially multinational companies, are the mainstay of innovation. Multinational companies transfer innovation to various countries through R&D alliances, cooperation agreements, subsidiaries, and affiliates. Therefore, the GIPN is an extension of the global production network [30]. In the GIPN, information technology and industry are combined to form a useful supplement to the internal innovation activities of enterprises and their core competitiveness. However, even so, in the industrial value chain, the innovation function is the most difficult to transfer. In comparison, patent greatly influences global technology collaboration among well-developed or major economically powerful countries while less developed countries have less potential to participate in the process of global technology transfer. In other words, the division of labor in the global

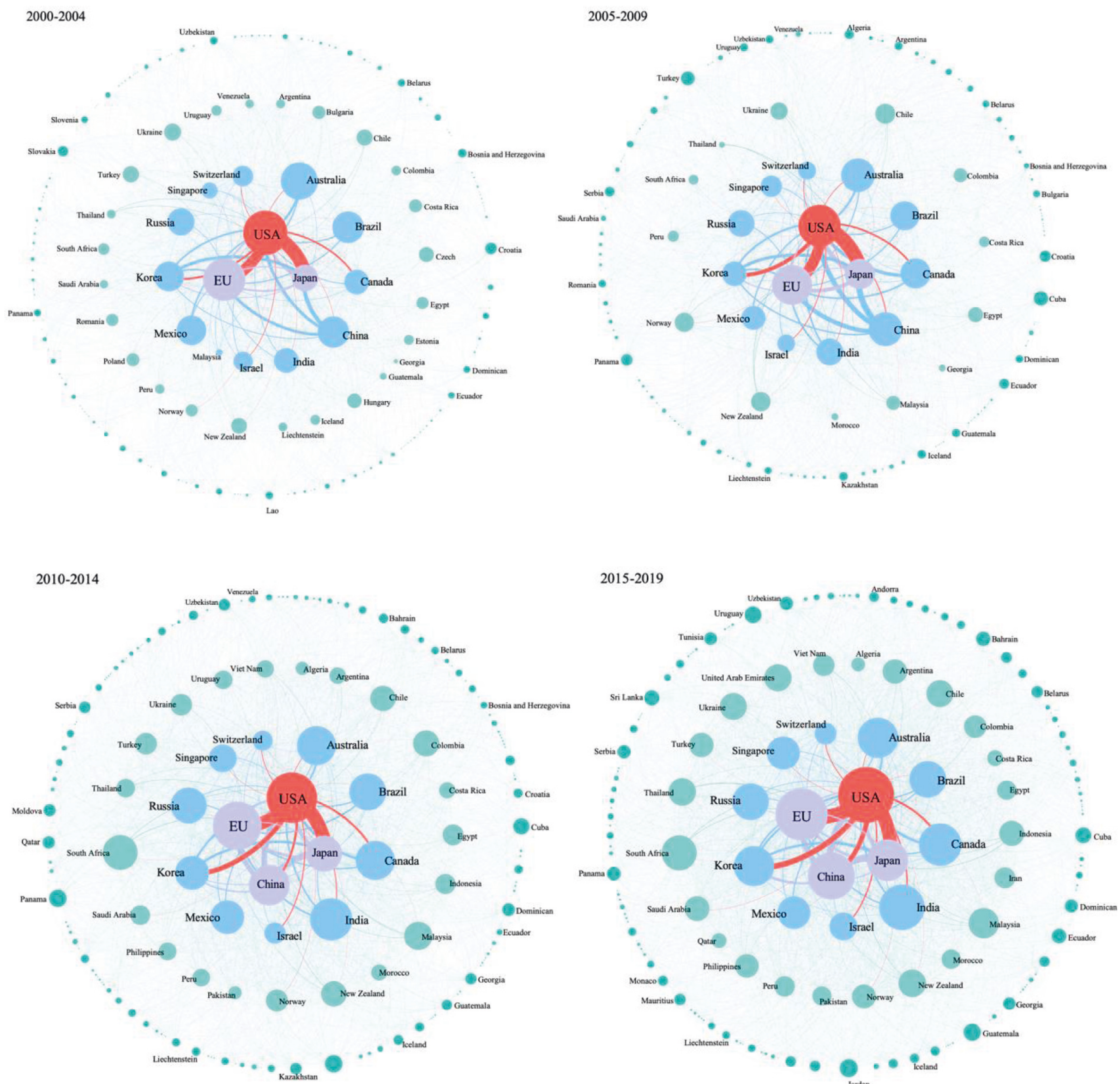


Figure 5: Core-periphery structure in the GIPN.

Table 2: The percentage of patents among layers in GIPN.

Layer	Percentage (%)			
	2000–2004	2005–2009	2010–2014	2015–2019
Core-core	39.45	34.34	49.76	50.03
Core-semicore	10.07	11.82	14.17	15.00
Core-semiperiphery	0.70	0.48	0.83	1.04
Core-periphery	0.07	0.26	0.26	0.22
Semicore-core	39.56	41.91	24.59	22.40
Semicore-semicore	4.40	5.94	2.87	2.64
Semicore-semiperiphery	0.56	0.47	0.42	0.40
Semicore-periphery	0.13	0.27	0.14	0.11
Semiperiphery-core	3.90	3.42	5.26	6.19
Semiperiphery-semicore	0.52	0.53	1.07	1.27
Semiperiphery-semiperiphery	0.08	0.04	0.12	0.15
Semiperiphery-periphery	0.03	0.03	0.05	0.05
Periphery-core	0.43	0.39	0.35	0.37
Periphery-semicore	0.07	0.08	0.09	0.10
Periphery-semiperiphery	0.02	0.01	0.01	0.02
Periphery-periphery	0.00	0.01	0.01	0.01
Sum	100	100	100	100

industrial value chain is still difficult to change. Therefore, similar to the global production network, only several countries determine the global innovation division of labor and the direction of technological development [31]. In addition, another point of view explaining this phenomenon is that the multilateral intellectual property system under the WTO is rooted in neoliberalism, which is the main economic ideology of western industrial capitalism. Although neoliberalism supports the use of multilateral rules and policies (such as proprietary intellectual property) as tools to promote technology transfer and commodities, empirical results show that the impact of such support is very limited [32].

5. Conclusion

This analysis explored the structure and evolution characteristics from both static and dynamic aspects. Static analysis reveals the overall distribution characteristics of the GIPN, while dynamic evolution analysis effectively identifies the evolution characteristics and development trends of the GIPN.

This analysis has aimed to explore the structural characteristics and evolution of the GIPN with the application of the patent dataset. Results gained from our analysis show that (1) in the period 2000–2019, the scale of the GIPN is expanding rapidly, gradually growing from a sparse network to a dense and complex network. (2) Patent application is unevenly distributed in the globe. The US, China, and Canada have been the top countries flowing in while Japan, Korea, EU, and Switzerland has been the main countries flowing out. (3) Some Asian countries have raised their position largely in the GIPN. These Asian countries include Viet Nam, Thailand, Malaysia, Singapore, and Indonesia. (4) The GIPN shows an obvious “core-periphery” pattern. The distribution pattern is unevenly distributed in space and presents a quadrilateral structure with the four core regions of “US, Japan, EU, and China” as the apex.

This analysis contributed to providing new insights both methodologically and theoretically. From the methodological view, based on the dynamic and static analysis of the GIPN, this analysis enriches network characteristics at the global scale and deepens the understanding of the global intellectual property transfer mechanism. From a theoretical view, given that the core of the GIPN is still distributed in a small number of core countries, countries are maintaining cooperation with core countries. The country at the key connection point can control and promote the technical exchange of nodes in the network. By linking different technologies together, technical barriers can be overcome, and the integration and innovation of different technologies can be accelerated. In addition, a deepening cooperation with countries in the core layer is also proposed in the future.

6. Limitations and Future Research

The analysis suffers from limitations in both methodological and dataset perspectives. First, the analysis suffers from limitations in a dataset perspective since we fail to distinguish the types of innovative areas, which may loss the shifting knowledge of intellectual property among the various fields. Second, we use the data of application rather than completed and approved. As in recent years, countries have become more and more strict about the protection of high-tech products and the term from application to approval is almost two years or more. The long application cycle may cause delays in changes in the GIPN. Third, this analysis did not focus on countries which have low rankings in the GIPN, paying little attention to the development of marginal countries in the GIPN. The overall goal of this analysis was to estimate the evolution structure of the GIPN. From the information collected from WIPO, we found that all the main country which occupy the main positions in global economy have been included. Thus, we believe the above effect to be of minor relevance. Future studies may consider the GIPN’s changes in the turbulent year of 2020. At the same time, the GIPN will focus on various industries [33], consider the deep reasons for the evolution of innovation networks, and explore the deep relationship between the world innovation pattern, transnational knowledge capital flow and international economic background.

Data Availability

The data collected during the study are freely available from the World Intellectual Property Organization (https://www.wipo.int/portal/en/index.html).

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This research was supported by the Beijing Philosophy and Social Science Foundation (No. 21JJC023), Key Project supported by Beijing Municipal Education Commission (No. SZ202110011006), and National Natural Science Foundation of China (No. 42101210).

References

[1] C.-Y. Tseng, “Technological innovation and knowledge network in asia: evidence from comparison of information and communication technologies among six countries,” *Technological Forecasting and Social Change*, vol. 76, no. 5, pp. 654–663, 2009.

[2] R. M. Grant and C. Baden-Fuller, “A knowledge accessing theory of strategic alliances,” *Journal of Management Studies*, vol. 41, no. 1, 2004.

[3] J. Cantwell and R. Mudambi, “Mnes competence creating subsidiaries mandates,” *Strategic Management Journal*, vol. 12, pp. 155–172, 2005.

[4] N. M. Coe and T. G. Bunnell, “Spatializing’ knowledge communities: towards a conceptualization of transnational innovation networks,” *Global Networks*, vol. 3, no. 4, pp. 437–456, 2003.

[5] X. Bai and Y. Liu, “Exploring the Asian innovation networks (AINs) characteristics,” *Information Development*, vol. 35, no. 1, pp. 21–35, 2019.

[6] C. Zhang and X. Wang, “The influence of ict-driven innovation: a comparative study on national innovation efficiency between developed and emerging countries,” *Behaviour & Information Technology*, vol. 38, no. 1, pp. 1–11, 2019.

[7] D. Ernst, “A new geography of knowledge in the electronics industry? Asia’s role in global innovation networks.” *Policy Studies*, vol. 54, 2009.

[8] D. Li, Y. D. Wei, and T. Wang, “Spatial and temporal evolution of urban innovation network in China,” *Habitat International*, vol. 49, pp. 484–496, 2015.

[9] F. Peng, Q. Zhang, Z. Han, Y. Ding, and N. Fu, “Evolution characteristics of government-industry-university cooperative innovation network of electronic information industry in liaoning province, China,” *Chinese Geographical Science*, vol. 29, no. 3, pp. 528–540, 2019.

[10] L. Zhang, Y. Zhao, D. Chen, and X. Zhang, “Analysis of network robustness in weighted and unweighted approaches: a case study of the air transport network in the belt and road region,” *Journal of Advanced Transportation*, vol. 2021, pp. 1–13, Article ID 8810254, 2021.

[11] J. Li, Y. Wang, and Y. Cheng, “Study of the collaborative innovation and operation mechanism in regional innovation network,” *Science & Technology Progress and Policy*, vol. 29, no. 13, pp. 32–36, 2012, (In Chinese).

[12] T. Buchmann and A. Pyka, “The evolution of innovation networks: the case of a publicly funded German automotive network,” *Economics of Innovation and New Technology*, vol. 24, no. 1-2, pp. 114–139, 2015.

[13] S. Valverde, R. V. Solé, M. A. Bedau, and N. Packard, “Topology and evolution of technology innovation networks,” *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, vol. 2, Article ID 056118, 2006.

[14] H. Chen and F. Xie, “How technological proximity affect collaborative innovation? An empirical study of China’s beijing–tianjin–hebei region,” *Journal of Management Analytics*, vol. 5, pp. 1–22, 2018.

[15] L. Lazzeretti and F. Capone, “How proximity matters in innovation networks dynamics along the cluster evolution. A study of the high technology applied to cultural goods,” *Journal of Business Research*, vol. 69, no. 12, pp. 5855–5865, 2016, Journal of Business Research.

[16] E. Marrocu, R. Paci, and S. Usai, “Proximity, networks and knowledge production in europe,” *Working Paper Crenos*, vol. 80, 2011.

[17] M. Steinmo and E. Rasmussen, “How firms collaborate with public research organizations: the evolution of proximity dimensions in successful innovation projects,” *Journal of Business Research*, vol. 69, no. 3, pp. 1250–1259, 2016.

[18] L. Zhang, Y. Zhao, B. Dessein, and P. De Maeyer, “Determinants of intercity air-passenger flows in the “belt and road” region,” *Complexity*, vol. 2021, pp. 1–14, Article ID 5514135, 2021.

[19] Y. Sun, “The structure and dynamics of intra- and inter-regional research collaborative networks: the case of China (1985-2008),” *Technological Forecasting and Social Change*, vol. 108, pp. 70–82, 2016.

[20] L. Yao, J. Li, and J. Li, “Urban innovation and intercity patent collaboration: a network analysis of China’s national innovation system,” *Technological Forecasting and Social Change*, vol. 160, Article ID 120185, 2020.

[21] M. Dost, Y. F. Badir, Z. Ali, and A. Tariq, “The impact of intellectual capital on innovation generation and adoption,” *Journal of Intellectual Capital*, vol. 17, no. 4, pp. 675–695, 2016.

[22] D. J. Watts and S. H. Strogatz, “Collective dynamics of ’small-world’ networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[23] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[24] G. Morrison, E. Giovanis, F. Pammolli, and M. Riccaboni, “Border sensitive centrality in global patent citation networks,” *Journal of Complex Networks*, vol. 2, no. 4, pp. 518–536, 2014.

[25] D. Reis, F. Moura, and I. Aragão, “Entrepreneurship, intellectual property and innovation ecosystems,” *International Journal for Innovation Education and Research*, vol. 9, no. 2, pp. 108–134, 2021.

[26] X. Pan, M. L. Song, J. Zhang, and G. Zhou, “Innovation network, technological learning and innovation performance of high-tech cluster enterprises,” *Journal of Knowledge Management*, vol. 23, no. 9, pp. 1729–1746, 2019.

[27] Z. Org, “Innovation offshoring:asia’s emerging role in global innovation networks,” *Economics Study Area Special Reports*, vol. 10, 2006.

[28] J. Wu, C. Wang, J. Hong, P. Piperopoulos, S. Zhuo, and E. Shuaihe Zhuo, “Internationalization and innovation performance of emerging market enterprises: the role of host-country institutional development,” *Journal of World Business*, vol. 51, no. 2, pp. 251–263, 2016.

[29] Y. Qiu and Y. Fan, “Rethinking global innovation strategy: emerging market perspectives,” *Business and Management Research*, vol. 2, no. 3, 2013.

[30] L. Ju, “Exploring the interplay, differences, and commonalities between global production networks and global innovation networks of two multinational companies,” *Papers in Innovation Studies*, 2014.

[31] L. Zhang, H. Du, Y. Zhao, P. De Maeyer, B. Dessein, and X. Zhang, “Drawing topological properties from a multi-layered network: the case of an air transport network in “the Belt and Road” region,” *Habitat International*, vol. 93, Article ID 102044, 2019.

[32] C. May, *The Global Political Economy of Intellectual Property Rights: The New Enclosures?*Taylor & Francis, 2013.

[33] T. Van der Valk, M. M. H. Chappin, and G. W. Gijsbers, “Evaluating innovation networks in emerging technologies,” *Technological Forecasting and Social Change*, vol. 78, no. 1, pp. 25–39, 2011.

Hindawi
Complexity
Volume 2022, Article ID 1816778, 11 pages
https://doi.org/10.1155/2022/1816778

Research Article

A Novel Model for Accessibility of Bus Station and Its Application to Bottlenecks Identification: A Case Study in Harbin

Chao Sun [1,2] and Wei Quan [1]

[1]*School of Traffic Science and Engineering, Harbin Institute of Technology, Harbin 150090, China*
[2]*School of Transportation, Southeast University, Nanjing 210000, China*

Correspondence should be addressed to Wei Quan; weiquan@hit.edu.cn

Received 20 December 2021; Revised 18 February 2022; Accepted 21 March 2022; Published 9 April 2022

Academic Editor: Zhipeng Tang

Accessibility plays a crucial role in evaluating and optimizing the service quality of public transport systems. Traditionally, bus accessibility research studies mainly focus on the evaluation from the perspective of the aggregate model and analyze the service radius of the site considering the front end of travel. This paper aims to model microlevel bus accessibility calculation and accurate bottleneck station identification according to the opportunity method with AVL (Automatic Vehicle Location) data. The research featured the following control elements: (1) redefining bus accessibility with bus station as the research object; (2) running time between interstation sections is selected as the control index; (3) A new measure index of accessibility. The farthest direct reachable distance is proposed. Moreover, by introducing the quality control algorithm, the study establishes a model of upper and lower limits of accessibility based on the probability index of bilateral norms, which can be employed to identify the bottleneck stations according to the significance level. Harbin City is selected as the case study for methodology verification and shows that the accessibility method is effective and reliable. Additionally, by taking the travel demand in the region as the adjustment parameter of the quality control algorithm, the study demonstrates the ideal accessibility interval of stations and accurately locates all bottlenecks. The study presented here adds to our understanding of accessibility in the construction and management of buses and provides an essential reference for the following optimization.

1. Introduction

With the development of the city, the urban spatial structure may change from a stable state to a constantly updated state continuously, and the urban design may be multicentered [1, 2]. Nevertheless, expansion into a marginal area will result in an evident increase in employment and housing imbalances and bring about greater travel obstacles, including the larger spatial-time block [3]. Accessibility is one of the most important abilities to decompress time and space. Its limitations would affect the balance of development between the central areas and others [4] and may even induce serious social problems [5]. Under normal circumstances, ensuring the balance of public transport accessibility plays a significant role in solving social equity [6]. Hansen first proposed the concept of accessibility in 1959, defined as the degree and difficulty of the interaction between nodes in the network [7]. Similarly, bus accessibility can be defined as the difficulty for travelers to get from one part of the city to another employing a bus system [8]. The evaluation of the bus accessibility index can comprehensively consider the interactive relationship between land use and bus to realize the guiding role of the bus on urban spatial structure and functional layout. Additionally, bus accessibility evaluation can guide urban construction by public transit, especially high-capacity rail transit systems. Urban construction can determine the type of land and the degree of intensification based on the distribution of bus accessibility, thus balancing the travel demand and reducing the pressure on the transportation system.

The evaluation of bus accessibility is varied and complex, involving many elements. Bus network is one of the modeling bases for measuring bus accessibility. The main method to build the topology of the bus network is to regard bus stations as nodes in the network and route sections or sections of two stations as edges in the network. Based on related theories, many scholars have studied the performance of bus networks, which provides a good breakthrough for accessibility evaluation. For example, two parameters, line function loss and connectivity, are often employed to measure the traffic function and connectivity of subway lines [9]. Some studies split a complex network with multiple weights into a complex network with single weights, investigated the global synchronization problem of the new multiweight network model, and gave simulation results of network equilibrium [10]. Wei et al. constructed a new type of bus route network and studied the spatial characteristics of urban bus routes using community detection [11]. Additionally, with the progress of graph theory and network theory, many novel networks such as weighted space-L networks [12], directed space-L networks [13], a hybrid of random network and scale-free network [14], and so on have been employed to model for the bus network.

Although the indexes and targets needed to evaluate the accessibility models in the existing literature are diversified and complicated, generally speaking, these can be divided into four types: Space Separation Model, Gravity Model, Cumulative Opportunities Model, and Utility Measure Mode. Among them, based on the general model framework, the concrete evaluation methods of public transport accessibility mainly include TTSAT (time-based traffic service area tools) [15], LUPTAL (land use and public transport accessibility index) [16], GTFS (general transport feed specifications) [17], PTAL (public transport accessibility level) [18] and so on. For example, Ding et al. employed principal component analysis to measure accessibility based on the gravity model [19]. Hyun et al. calculated the accessibility of three different stages based on time distance and proved its spatial transfer effect [20]. Mavoa et al. provided a measure of bus services through the bus and walking accessibility index and transit frequency [21]. Furthermore, accessibility-based location selection of urban POIs (Points of Interest) also has become a research hotspot in recent years, such as medical facilities [22], educational facilities [23], and leisure facilities [24]. For instance, Oviedo et al. deployed multisource data to study the impact of accessibility on employment, especially the contribution to the relatively poor groups [25]. As researches continue, the spatial-temporal dynamics of travel have been of increasing interest to scholars. By using an integrated commuter mode, Owen et al. focused on accessibility in continuous time and estimated a binomial logistic model of bus commuting [26]. Guida et al. studied the phenomenon of accessibility and equity by analyzing the travel behavior of the ageing population [27]. Jrv et al. presented a conceptual framework for location-based dynamic accessibility modeling that captures the temporal correlation of people, traffic, and social activity locations [28]. Significant data sources offer new possibilities for urban mobility and accessibility studies, especially mobile cellular data from cell phones simplify group travel patterns and traffic geography studies. Garcia et al. used the Google API to construct OD matrixes, ultimately analyzing the independent impact of dynamic accessibility and its associated components [29]. However, little attention has been paid to systematically analyzing the effect of temporal resolution on the results. Based on this, Marcin et al. addressed the loss of accuracy due to progressively lower temporal resolution, aiming to guide selecting the appropriate temporal resolution in accessibility studies [30].

Most of the studies are focused on the connotation of accessibility, the establishment of quantitative models, determination of evaluation indexes, and the optimization of accessibility. The theory of accessibility has been widely used in the actual process of urban planning and construction. However, most of the existing studies analyze bus accessibility from the macroscopic point of view, and few theories are developed at the microscopic level (such as stations). For the station, most studies only consider the front end of the trip to analyze the service radius of the station, that is, evaluating the difficulty level to reach the station. For passengers, besides the accessibility from the departure point (home, workplace, school, etc.) to the bus station, the spatial distance range that the bus can reach from the station is the key index to measure the service capacity of the station. Additionally, the models mainly limited evaluating the accessibility of public transport without identification of weak points and bottlenecks in the public transport network. The travel bottleneck is a relatively new and important concept applied to traffic congestion, but few theories on bottlenecks of public transport are researched. Bottleneck identification models for public transport can help us understand the operation mode and optimize service efficiency to the greatest extent possible. This paper aims to respond to the above-mentioned challenges by applying a chance model based on AVL and spatial data to address the following research questions:

(1) How to consider the time factor and measure the space range that can be reached from a bus station within the bus network?

(2) How to accurately locate bottlenecks in public transportation networks based on accessibility?

The paper is structured as follows. The methodology to model bus accessibility and identify the bottleneck stations is discussed in Section 2. The opportunity method with AVL data is employed to establish time and space limits. We use POI (point of interest) data to describe the land use that has been considered to calibrate the relevant parameters of the subsequent model. Additionally, we expound on the framework of using quality control to analyze specification range and probability indexes for bottlenecks identification. Section 3 presents the case study to demonstrate the results of accessibility calculation and bottlenecks identification. The paper ends with the discussion and conclusions in Section 4.

2. Methodology

2.1. Methodology Framework. The paper investigates the bus accessibility and bottleneck stations with the time-space variation. The methodology framework is mainly constructed based on data processing, accessibility calculation, bottleneck stations identification, and case verification, respectively, as shown in Figure 1.

2.2. Accessibility Model

2.2.1. Theoretical Basis. Based on the opportunity method, this paper employs a geographic information system to measure the accessibility of public transport from both micro and macroperspectives [31, 32]. The public transportation system includes three levels: station, line, and area network. Obviously, the smaller the research scale, the higher the accuracy of the research results. In addition, from the perspective of travelers' needs, time and distance are the most critical factors affecting travel decisions. Therefore, the single station is selected as the research object in this study. By considering the measurement of time, the accessibility of public transport is defined as the maximum sum of the directly available distance at a certain point within a reliable time. The sum of the accessibility of all stations constitutes the accessibility level of the whole network. The prominent advantage of this definition lies in ① taking a single station as the research foundation, indirectly reflecting the service level of the station by measuring the travel opportunities available of the station; ② providing a basis for finely identifying the bottleneck points of the public transport network and giving some reference for travel choices of users. According to the definition, three factors affect the spatial accessibility of bus stations: the number of stops, the longest distance, and the travel time of each shift. Therefore, each station's spatial accessibility calculation model is calculated by the following equations:

$$M_{\text{SDAD}} = \sum_{j=1}^{n} L_j(\max) f(t), \tag{1}$$

$$f(t) = \begin{cases} 0, & t_j > t_{\text{limit}}, \\ 1, & t_j \le t_{\text{limit}}, \end{cases} \tag{2}$$

$$t_j = \sum_{i=i}^{m_j} \overline{T}_i^{(j)}, \tag{3}$$

$$t_{\text{limit}} = m \times \overline{t}, \tag{4}$$

where M_{SDAD} is the spatial accessibility level of a station; $L_j(\max)$ is the maximum distance that the bus for the j-th line that stops at the station can reach from the station to continue driving; $f(t)$ is a discriminant function, which is used to determine whether the running time of public transport vehicles exceeds the limit; t_{limit} is the time limit, that is, the reliable travel time of the j-th line bus that stops at the station and continues to travel from the station to the destination, which belongs to the ideal time and takes the average level of the whole city; m_j is the number of inter-station sections passed by the j-th line bus that stops at the station and continues to drive from the station to the destination; t_j is the actual travel time required by the j-th line bus that stops at the station to continue driving from the station to the destination; $\overline{T}_i^{(j)}$ is the actual interval running time required for the bus on the j-th line stopping at the station to continue driving and pass through the i-th inter-station section; and $\overline{t}$ is average running time of all inter-station sections in the city.

The larger the value of MSDAD, the greater the travel opportunity obtained from the station, which means the higher the possibility of meeting the travel demand of passengers, and thus the higher the accessibility level. At the same time, for lines or networks, if the covered stations own great accessibility, their accessibility is also at a relatively good level. Therefore, this paper employs the accessibility of the stations to get the accessibility calculation model of the line, as shown in the following equation:

$$A_{\text{bl}} = \sum M_{\text{SDAD}}. \tag{5}$$

A_{bl} is the accessibility of bus line. Numerically, it is equal to the sum of all the accessibility levels of stations of this line. Similarly, if the accessibility level of a region needs to be calculated, the accessibility values of all stations in the region are summed up.

2.2.2. Calculation Procedure. Taking the running time between stations as the control index and dividing a day into three periods (morning peak, 6:00–9:30; Evening peak, 16:00–19:30), we calculate the bus accessibility of the whole city within a week. The running time between stations is a crucial index to determine the reliability of bus running time. In this study, the running time between stations refers to the time from entering the last station to the next station, including the stay time at the last station and the driving time before reaching the next station. The total calculation steps are as follows.

Step 1: Match the station number with the actual station according to the line name.

Step 2: Establish two lists for each interstation section, and the list of running time of interstation section starting from i station is $i_t = []$, and $i_a = []$ is used to store the accessibility of the i-th station. Traverse the inbound and outbound data of all routes and shifts and match the "O_TIME" index with stations to obtain the inbound time of all routes and shifts at each station. For a single shift of a single line, the arrival time of i-th station is T_i, and the arrival time of $(i+1)$-th station is $T_i + 1$, so $T_i + 1 - T_i$ is added to list i_t, as shown in Figure 2.

Step 3: According to the list i_t obtained from each inter-station section, calculate the average running time of i-th interstation section $(\overline{T}_i)$. Further, calculate the running time of all inter-station sections in the whole city, which is the $\overline{t}$ in equation (4).

Step 4: According to equation (3), calculate the theoretical time t_j required for the bus of the j-th line reaching the terminal from i-th station. If $t_j < t_{\text{limit}}$, go to step 5; otherwise, another line is calculated.

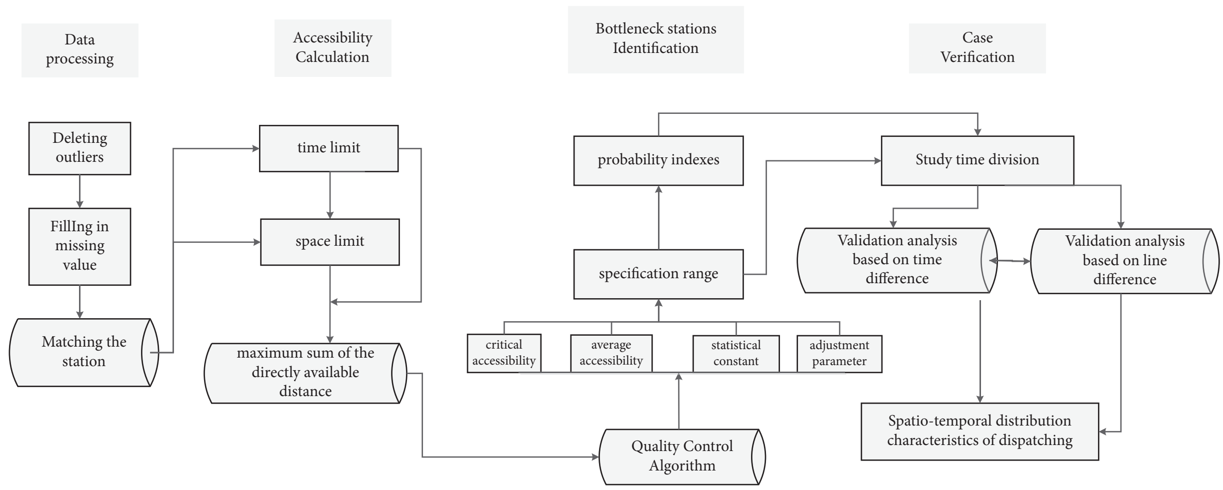


Figure 1: Methodology framework.

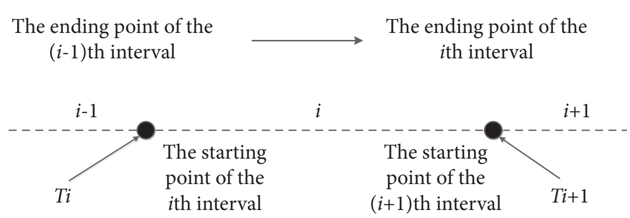


Figure 2: Interstation section running time calculation.

Step 5: Calculate the distance between i-th station and the terminal station of the j-th line according to the latitude and longitude of the station [33] and add the calculation results to i_a.

Step 6: Repeat all routes and all stations to calculate the accessibility of all stations.

2.3. Identification for Bottleneck Stations. The ultimate goal of accessibility research is to identify unreliable units (stations or lines) and provide a basis for the next step of optimizing the layout of stations or lines. Because of the large amount of bus station data, this paper employs QCA (Quality Control Algorithm) to judge the degree to which the test data meet the requirements of quality standards (such as specification range) through the probability indexes of bilateral specifications, to identify the bottleneck stations with low accessibility [34, 35]. The QCA compares the average accessibility of all stations on the same line and establishes the upper and lower limits for evaluating the accessibility of public transportation according to the significance level. When the accessibility of the evaluated station is less than the limit value, the station is the reachable bottleneck station. The specific calculation model is as follows:

$$\begin{cases} A_C^+ = \overline{A_C} + k\sqrt{\dfrac{\overline{A_C}}{M}} + \dfrac{1}{2M}, \\ A_C^- = \overline{A_C} - k\sqrt{\dfrac{\overline{A_C}}{M}} - \dfrac{1}{2M}, \end{cases} \quad (6)$$

where AC is the critical accessibility value, “+” represents the upper bound and “−” represents the lower bound; $\overline{A_C}$ is the average accessibility of the station; k is a statistical constant, corresponding to different confidence levels; and M is an adjustment parameter, which is the travel demand within the service radius of the station.

3. Case Study and Result Analysis

3.1. Study Data and Area. Part of the research data comes from the AVL data provided by Harbin Bus Company from May 1st to May 7th, 2017. The file size of single-day data is about 30M, containing about 300,000 inbound and outbound data. The relevant data attribute records are shown in Table 1. The other part comes from vector map data of Harbin provided by open street map, and crawler tools capture the latitude and longitude data of all stations in the city.

The Harbin City is employed for verification (as shown in Figure 3). As the capital of Heilongjiang province, it is located in the extreme northeast of China. As of February 2020, there are 256 bus lines in Harbin (the main urban area); among them, the length of the bus network reached 997.4 kilometers, the density of the line network reaches 2.47 km/km^2, and the average daily passenger volume of buses and trams is 3,676,200 (https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202102&filename=1021030779.nh&uniplatform=NZKPT&v=lc4N3PNCTp3TIbtOIgb1Doj9ikTLNXOVYLgIFoDQc2wms1oyRiuep2ZRWJBeY4QU). From the traffic analysis report released by the Amap, however, it can be seen that Harbin has been listed as one of the top ten traffic jam cities in China for many years (https://report.amap.com/share.do?id=a187527876d07ac50177142eba987ce0; https://baijiahao.baidu.com/s?id=1622808967853154807; https://wenku.baidu.com/view/95cddad46bdc5022aaea998fcc22bcd126ff42ca.html). At the same time, Harbin City is known as one of the most unreasonable cities in the country in terms of road planning. The unreasonable and unscientific road planning has a direct impact on the development of the city. Traffic

Table 1: The attribute of AVL data.

The attribute name	Property description
O_LINENO	Record line number
O_BUSNAME	Record the bus name
O_STATIONNO	Record station number
O_TIME	Record the stop time
O_STATIONSTATUS	Record the running status. 0 indicates the forward running and 1 indicates the reverse running

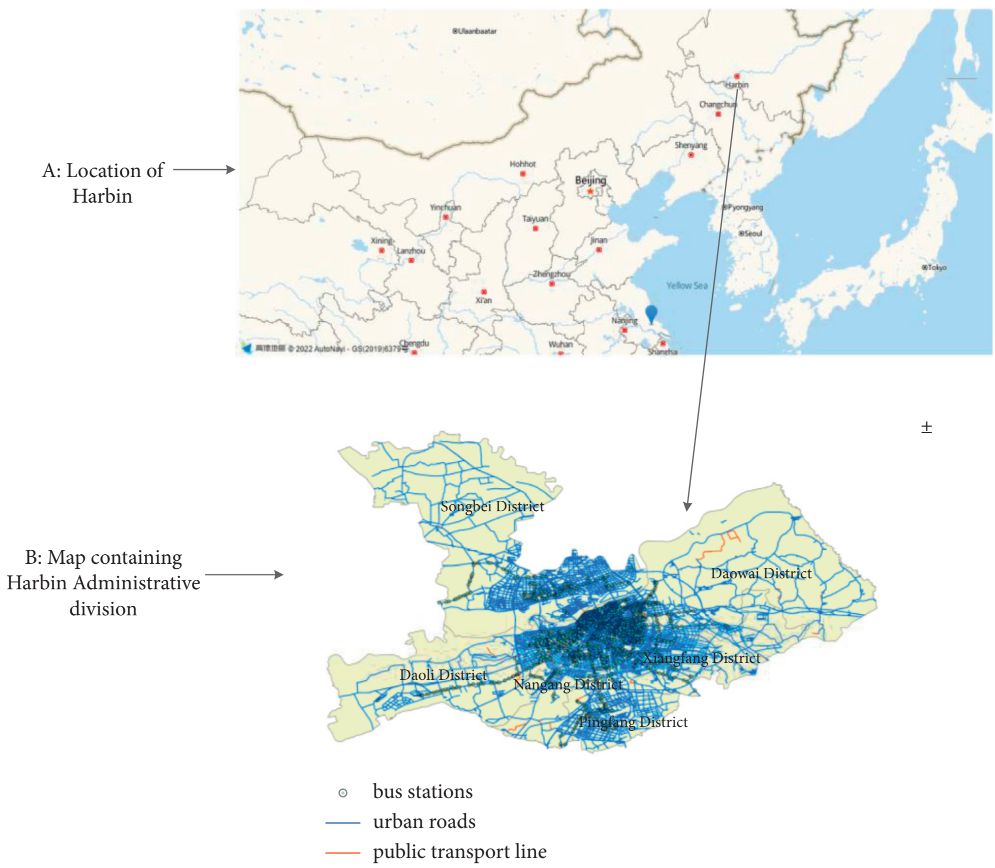


Figure 3: Demonstration of the research area.

congestion is increasing and induces a negative impact on the economic-social development and the function of the city. The most direct, essential, and obvious harm is the inconvenience of urban residents and the significant increase in travel costs. Therefore, compared with other cities, the application of the above-mentioned model to Harbin shows more urgent demand and practical value.

According to Figure 3, six districts are selected as research areas including Nangang district, Pingfang district, Daoli district, Xiangfang district, Daowai district, and Songbei district. The government has invested more in public transport in these areas, making them more representative as they have better infrastructure. However, due to the low level of economic growth, the traffic department has never recorded AVL data for other regions, which are relatively backward in bus development.

3.2. Analysis of Validity Based on Spatial-Temporal Difference. According to the calculation results of the proposed accessibility model, it can be seen that the performance of bus accessibility varies significantly in different time periods (flat peak and peak) and different areas (downtown and surrounding areas). Based on existing calculation results, the validity of the model is further analyzed. Referring to Wei et al.'s [36] modeling ideas in the accessibility calculation of rail transit stations, this paper verifies the model by comparing the horizontal distribution and vertical changes of accessibility values.

3.2.1. Validation Analysis Based on Time Difference of Accessibility. According to the calculated results, the accessibility level is classified by using Jenks natural breaks method: Grade 5: 0–1 500 m; Grade 4: 1500–3000 m; Grade

3: 3000–6000 m; Grade 2: 6000–12,000 m; Garde 1: greater than 12,000 m. Figures 4–9 show the accessibility performance in different time periods of a certain day. There is a significant difference in the accessibility of public transportation between morning and evening peak hours and off-peak hours. Only the proportion of stations with Grade 5 accessibility increases significantly during the morning and evening peak hours, while others do not. As mentioned above in the traffic analysis report, traffic congestion in Harbin is very serious, especially during peak hours. During peak hours, after all, the disruption to bus running will also reach the highest due to the surge in traffic volume. As a result, the maximum distance that buses can reach on most routes within a reliable time will drop steeply to a minimum, resulting in an increase in the number of stations with Grade 5 accessibility. By contrast, the number of stations with other Grade accessibility will decrease.

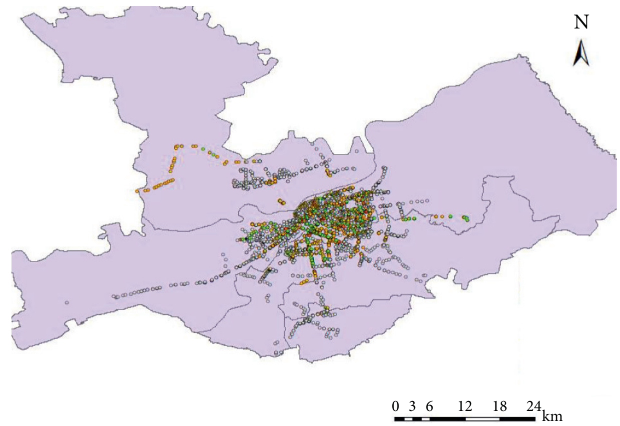


FIGURE 4: Spatial distribution of accessibility during off-peak hours.

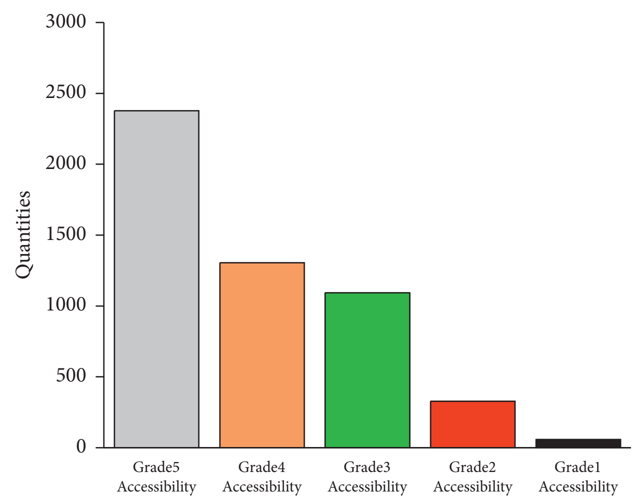


FIGURE 5: Off-peak accessibility performance.

Combined with Figures 4–9, according to the calculation results of average accessibility in a week (Figure 10), there is little difference between morning and evening peak hours. The average accessibility in off-peak hours is far greater than that in peak hours. The average accessibility during off-peak hours is about 5.76 times that during peak hours, which means that the average distance that can be reached by bus from a station during peak hours may be less than one-fifth of that during off-peak hours in the ideal time range. According to the above analysis, the time difference of bus accessibility is basically consistent with the performance of the Harbin urban traffic congestion index in different periods. In addition, the average off-peak accessibility shows the trend as "falling down first and rising up later in one week," especially the lowest point on Wednesday. It is an episodic condition that can be linked to a day of the week when something happens, such as bad weather or a big event, but the exact cause is unknown due to data limitations, which will be discussed in future research. The main purpose of this graph is to show the difference in accessibility between peak and off-peak hours, not the fluctuation of accessibility within a week. Furthermore, the one-week average of accessibility is used for model validation and bottleneck site identification below, which weakens the unanticipated event's effect.

3.2.2. Validity Analysis Based on Differences in Accessibility of Line. In this paper, equation (5) is employed to calculate the accessibility of all bus lines in urban areas, which helps evaluate the service level of all lines. The model's validity is also analyzed according to the accessibility difference between different lines and different stations on the same line. Bus lines No. 58 and No. 31 in Harbin are selected as examples, both of which run at Harbin's relatively prosperous downtown area, with large travel demand, large time, space span, and high research value. According to the calculation results of the model, the accessibility of the No. 58 bus line is better than that of the No. 31 bus line. For example, from Nanzhi Road Station to Haxi Wanda plaza Station, the accessibility values of the two lines are 162,839.2 and 135,167.5, respectively. This result is basically consistent with the comparison result of the time impedance provided by the Gaode map. The time impedance of the former is 46 minutes, which is obviously superior to the latter's 63 minutes. For bus line No. 31, its accessibility distribution is bounded by Harbin Railway Bureau located near the center of Harbin with a dense population and heavy traffic, which has a more prominent influence on public transportation, which is consistent with the traffic congestion in Harbin.

3.3. Bottleneck Stations Identification Based on Accessibility. The ultimate goal of this study is to employ the quality control algorithm to find the bottleneck stations in the line according to the accessibility of the same line station and provide the focus for the subsequent optimization work. As a public welfare infrastructure, public transportation serves the largest users. Therefore, the evaluation of public transportation services must be based on the size of travel

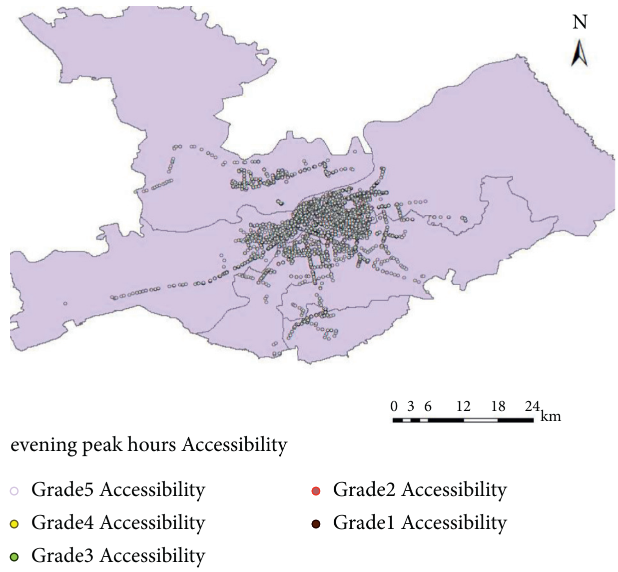


Figure 6: Spatial distribution of accessibility performance in evening peak hours.

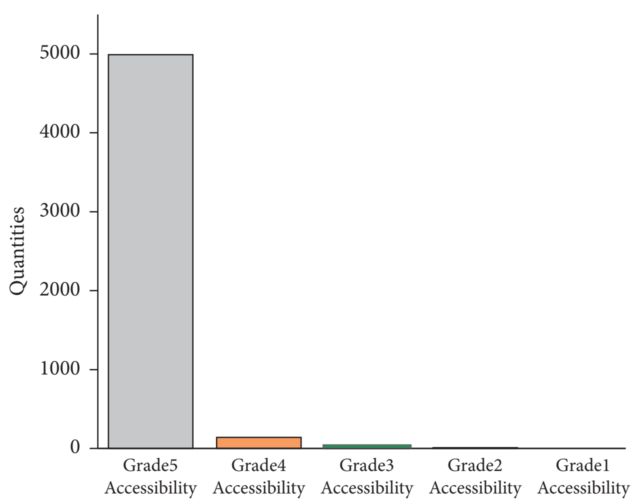


Figure 7: Accessibility performance during evening peak hours.

demand. Because the number of POI facilities reflects the level of travel demand in this area to a certain extent, the station is used as a buffer with a radius of 500 m, and the number of POI facilities is used as an adjustment parameter the quality control algorithm. According to equation (5), this paper takes Harbin No. 31 bus as an example for bottlenecks analysis. Among them, the statistical constant is 1.96, corresponding to a 95% confidence level.

As shown in Table 2, all bottleneck stations of the No. 31 bus are identified by judging the degree to which test data meet the requirements of quality standards. For example, in terms of Haxi Wanda plaza, its accessibility is 3366.1, significantly lower than the lower limit of 4465.98, so it is defined as the bottleneck station. Among them, the normalized travel demand of Heilongjiang University Station is 0, which cannot be calculated through the above model. Because it corresponds to the lowest travel demand at this

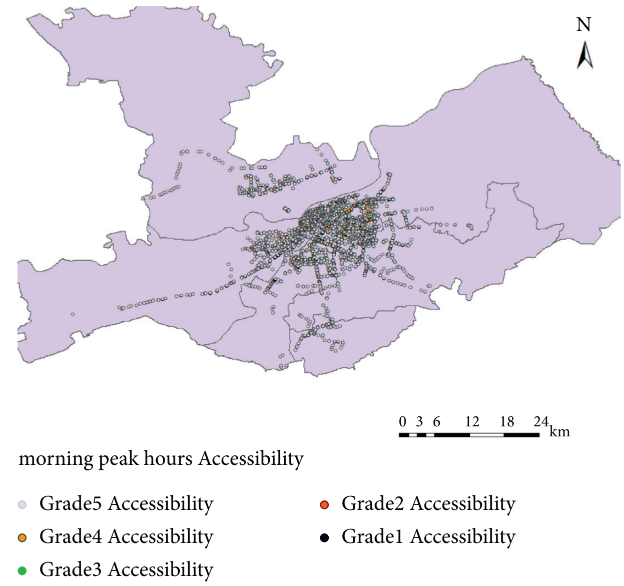


Figure 8: Spatial distribution of accessibility performance in morning peak hours.

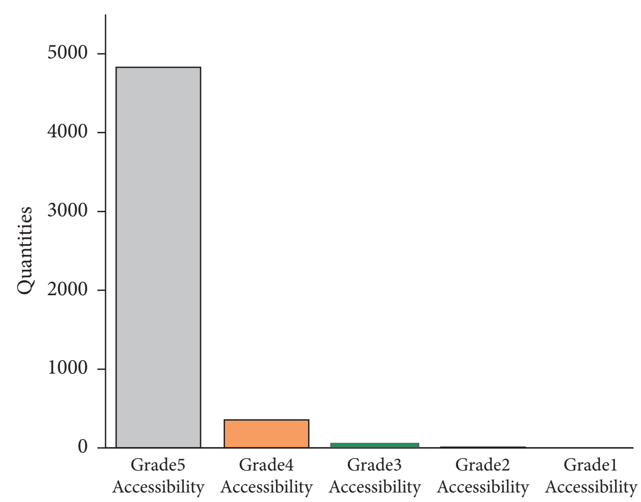


Figure 9: Accessibility performance during morning peak hours.

time, the upper and lower limits are the minimum values of other stations. At the confidence level of 95%, 17 bottleneck stations are located, which exceeds 50% of the total stations and explains the operation unreliability of the No. 31 bus from the side. Similarly, according to the above method, whether the stations inside all bus lines in Harbin are abnormal or not are identified. Starting from the bottleneck stations, optimizing the bus shifts and other related indicators will help to improve the accessibility of the whole public transportation system more accurately.

4. Discussion and Prospect

4.1. Discussion

(1) This paper explored a novel model for evaluating the urban bus accessibility from the microperspective by

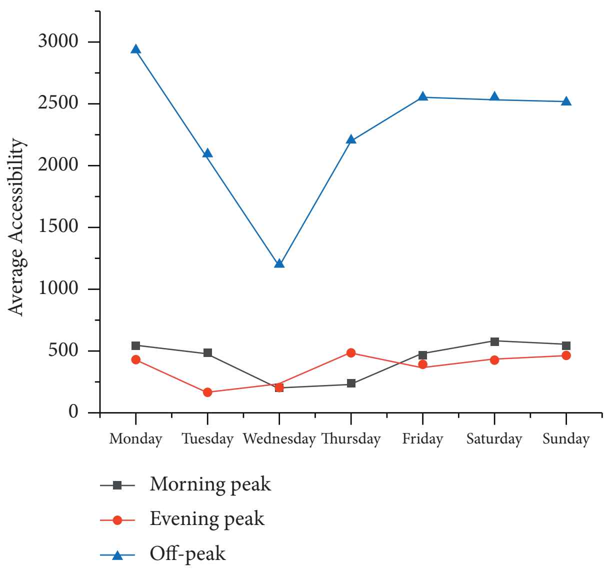


Figure 10: Time differences in accessibility.

using GIS software based on the AVL data, which realized the calculation of bus accessibility on different scales and accurate identification for the bottleneck stations with the worst balance between demand and supply in order to effectively optimize the level and fairness of urban public transport service. The paper employed the chance method and bus trajectory data to describe the buses' ability to meet travel demand, dynamically monitoring and tracking the variation of service level temporally and spatially. Furthermore, this paper took bus stations as the research object, took the interstation section time as the control index, and put forward a new measure index of station accessibility, called the farthest direct reachable distance. This index indirectly reflected the service level of the station by measuring the travel opportunities available, provided a basis for finely identifying the bottleneck points of the public transport network, and gave some reference for travel choices of travelers. The Tanimoto coefficient was finally used to discuss the coupling relationship between public transport accessibility and external development factors and describe the unbalancedness of the layout of bus facilities and bus service.

(2) Quality control theory was selected to establish a multiparameter evaluation probability model for identifying bottlenecks because the model systematically considered all stations of one line, which could solve the problem of confidence level and dynamically guarantee the implementation of the identification system. Furthermore, the paper improved the evaluation method of travel demand to reduce the difficulty of data collection and calculation, based on the point of interest instead of bus ridership data.

(3) Taking Harbin as an example, based on the AVL data within one week, we calculated the accessibility of stations and routes in the whole city during the morning and evening peak periods and the flat peak periods. ① Based on the spatial-temporal difference of accessibility and the coupling with congestion state, we verified the effectiveness of the refined bus accessibility model based on AVL data at the station scale. In engineering practice, the accessibility of different stations and lines could be deployed to optimize the public transport network, which provided some reference for actual traffic planning. ② Through comprehensive analysis of bus accessibility in different dimensions, there was little difference in bus accessibility between morning and evening rush hours, while the average accessibility in peak hours was about five times that in peak hours. There were significant differences in the accessibility of public transport in different areas of Harbin, with the maximum difference exceeding 30 times. The development of public transport performed obvious regional imbalance, but this imbalance was related to social and economic development. ③ For the direct quality control of bus stations, the statistical probability was 1.96, and then it could be known that there were 17 bottleneck stations in the No. 31 bus route of Harbin at a 95% confidence level. After joining the quality control of public transport accessibility, the bottleneck stations could be identified under different confidence levels, and the error range could be given.

(4) In general, this method is superior compared to other accessibility methods and its importance in urban traffic network or bus stations planning. ① Finer scale: Previous studies mainly evaluated the accessibility of public transport from the regional perspective. In this paper, the accessibility of a single station is modeled, forming a complete evaluation system from point to line to surface. In this way, the planning and management department can clearly grasp the rationality of the infrastructure layout and provide a reference for the next optimization strategy. ② More practical: There are few models about bus bottlenecks in existing research, and most of them stop at accessibility evaluation. With the help of a quality control algorithm, this paper realizes the location of weak points in public transport networks from the perspective of accessibility. Researchers can make more accurate transportation policies and management plans based on this. In particular, the probability value of each bottleneck site is given, which makes the identification model more reliable and flexible. More importantly, users can plan their travel strategies more reasonably according to the accessibility of each stop and whether it is a bottleneck or not and even change the selected departure bus stop.③ Smaller amounts of data: We only need to obtain the AVL data within the research area

Table 2: The abnormal station determines the result.

Station name	Accessibility	Travel demand (standardizing)	Upper bound of quality control	Quality control lower boundary	Is it a bottleneck?
Heping Community	4291.5	0.407	4943.102	4518.238	No
South Road	997.7	0.037	5444.656	4016.684	Yes
Hongqi Community	2654.2	0.370	4953.533	4507.807	Yes
Huaihe Road	2674.7	0.259	4997.357	4463.983	Yes
Liaohe Community	102.3	0.259	4997.357	4463.983	Yes
Songshan Community	159.6	0.407	4943.102	4518.238	Yes
Xianfeng Community	2119.5	0.444	4934.008	4527.332	Yes
Xianfeng Road Intersection	4854.9	0.296	4980.017	4481.323	Yes
Dacheng Street	1200.1	0.222	5018.892	4442.448	Yes
Children's Park	1880.3	0.074	5232.738	4228.602	Yes
First Medical Hospital	3421.0	0.444	4934.008	4527.332	Yes
Lin Qiu Company	3331.5	0.593	4906.635	4554.705	Yes
Provincial Expo Center	2569.6	0.630	4901.357	4559.983	Yes
Harbin Railway Bureau	27955.7	0.852	4877.318	4584.022	No
Harbin Institute of Technology	5063.1	0.815	4880.628	4580.712	No
West Bridge	3037.1	0.704	4892.083	4569.257	Yes
Orthopedic Hospital	3096.9	0.333	4965.665	4495.675	Yes
Tongda Street	740.9	0.667	4896.526	4564.814	Yes
Helu Road	2937.1	0.185	5046.637	4414.703	No
Hesandao Street	11286.2	0.667	4896.526	4564.814	No
Xingshiyi Avenue Street	3762.1	0.630	4901.357	4559.983	No
Harbin Normal University Affiliated Middle School	14400.2	0.741	4887.978	4573.362	No
Harbin University of Science and Technology	7200.1	0.667	4896.526	4564.814	No
Heilongjiang University	3286.8	0	4865.979	4016.684	Yes
Garment Mall	7933.5	0.778	4884.171	4577.169	No
Harbin Institute	4917.1	0.963	4868.566	4592.774	No
Political and Legal Cadre College	5743.8	0.407	4943.102	4518.238	No
Zhongxing Boulevard	4475.4	1	4865.979	4595.361	Yes
Haxi Wanda Plaza	3366.1	1	4865.979	4595.361	Yes
Haxi Railway Station East Square Station	2461.3	0.074	5232.738	4228.602	No

without assistance from other multisource spatial data, which can be conveniently complicated through ArcGIS. In addition, the model in this article does not need to forecast traffic congestion and other socioeconomic factors, which makes the application process more straightforward and saves time. ④ No complicated algorithm: We just need to calculate the departure and arrival times for each bus station, which are the simplest indexes in evaluating buses. Compared to other intricate algorithms, including nested logit "logsum" [8], comparative assessment [17], 3SFCA [22], online route planning algorithm [29], and so on, the model proposed in this manuscript does not require any special design algorithm and has a low threshold, which is suitable for different cities and has better application prospects.

4.2. Prospect. With the help of urban computing and big spatial data, it is easy to get all the data sources of bus accessibility evaluation. Every certain period, the repeated accessibility calculation process can dynamically evaluate and track the improvement degree of bus system services in time and space. By promoting the development of urban public transport, this method can realize the optimal allocation of urban public transport resources, thus improving the attractiveness of public transport to the greatest extent and increasing the sharing rate.

There are still some problems to be further discussed: (1) The relevant conclusions of this paper can only be verified by the temporal and spatial differences of urban public transport accessibility and the general characteristics of coupling. Compared with the traditional accessibility calculation model based on bus trajectory data, the accuracy and effectiveness of this method need to be further discussed. (2) Our primary concern is the accessibility of traditional public transportation while ignoring the influence of other shared transportation, such as bike-sharing and the subway under construction. (3) Although the peak and off-peak hours are distinguished, the influence of other objective factors (nonworking days, weather, etc.) on accessibility and bottleneck stations is not considered. Future research should consider the internal relationship between different vehicles in detail and attach importance to connection accessibility to find out the causes of bottlenecks. In addition, future research should be analyzed in different demand scenarios to establish a more stable and reliable bottleneck identification model.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 42171451.

References

[1] S. S. Rosenthal, W. C. Strange, and J. A. Urrego, "Are city centers losing their appeal? commercial real estate, urban spatial structure, and COVID-19," *Journal of Urban Economics*, vol. 127, 2021.

[2] L. A. Yuan and B. Ceka, "Topic modeling to discover the thematic structure and spatial-temporal patterns of building renovation and adaptive reuse in cities," *Computers, Environment and Urban Systems*, vol. 78, Article ID 101383, 2019.

[3] C. Sun and W. Quan, "Evaluation of bus accessibility based on hotspot detection and matter-element analysis," *IEEE Access*, vol. 99, no. 1, 2020.

[4] R. Benevenuto and B. Caulfield, "Measuring access to urban centres in rural Northeast Brazil: a spatial accessibility poverty index," *Journal of Transport Geography*, vol. 82, Article ID 102553, 2020.

[5] G. Porter, "Transport services and their impact on poverty and growth in rural sub-saharan africa: a review of recent research and future research needs," *Transport Reviews*, vol. 34, no. 1, pp. 25–45, 2014.

[6] Z. Song, M. Cao, T. Han, and R. Hickman, "Public transport accessibility and housing value uplift: evidence from the Docklands light railway in London," *Case Studies on Transport Policy*, vol. 7, no. 3, pp. 607–616, 2019.

[7] W. G. Hansen, "How accessibility shapes land use," *Journal of the American Institute of Planners*, vol. 25, no. 2, pp. 73–76, 1959.

[8] N. Nassir, M. Hickman, A. Malekzadeh, and E. Irannezhad, "A utility-based travel impedance measure for public transit network accessibility," *Transportation Research Part A Policy Practice*, vol. 88, pp. 26–39, 2016.

[9] J. Zhang, X. Xu, H. Liu, S. Wang, and F. Qi, "Networked analysis of the Shanghai subway network, in China," *Physica A: Statistical Mechanics and Its Applications*, vol. 390, no. 23–24, pp. 4562–4570, 2011.

[10] X.-l. An, L. Zhang, and J.-g. Zhang, "Research on urban public traffic network with multi-weights based on single bus transfer junction," *Physica A: Statistical Mechanics and its Applications*, vol. 436, pp. 748–755, 2015.

[11] S. Wei, W. Zheng, and L. Wang, "Understanding the configuration of bus networks in urban China from the perspective of network types and administrative division effect," *Transport Policy*, vol. 104, pp. 1–17, 2021.

[12] S. Yi, F. J. Shao, R. C. Sun, and S. J. Li, "Space evolution model and empirical analysis of an urban public transport network," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 14, pp. 3708–3717, 2012.

[13] N. Cao and H. Cao, "Exploring the robustness of urban bus network: a case from Southern China," *Chinese Journal of Physics*, vol. 65, pp. 389–397, 2020.

[14] S. D. Dimitrov and A. Ceder, "A method of examining the structure and topological properties of public-transport networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 451, pp. 373–387, 2016.

[15] C.-L. Cheng and A. Agrawal, "TTSAT: a new approach to mapping transit accessibility," *Journal of Public Transportation*, vol. 13, no. 1, pp. 55–72, 2010.

[16] M. Pitot, T. Yigitcanlar, N. Sipe, and R. Evans, "Land use and public transport accessibility index (LUPTAI) tool - the development and pilot application of LUPTAI for the gold coast," *29th Australasian Transport Research Forum, ATRF 06*, Australia, 2006.

[17] D. Luo, O. Cats, H. van Lint, and G. Currie, "Integrating network science and public transport accessibility analysis for comparative assessment," *Journal of Transport Geography*, vol. 80, Article ID 102505, 2019.

[18] B. Adhvaryu, A. Chopde, and L. Dashora, "Mapping public transport accessibility levels (PTAL) in India and its applications: a case study of Surat," *Case Studies on Transport Policy*, vol. 7, no. 2, pp. 293–300, 2019.

[19] J. Ding, Y. Zhang, and L. Li, "Accessibility measure of bus transit networks," *IET Intelligent Transport Systems*, vol. 12, 2018.

[20] K. Hyun, L. Keumsook, P. J. Soo, and Y. Song, "Transit network expansion and accessibility implications: a case study of Gwangju metropolitan area, South Korea," *Research in Transportation Economics*, vol. 69, pp. 544–553, 2018.

[21] S. Mavoa, K. Witten, T. Mccreanor, and D. O'Sullivan, "GIS based destination accessibility via public transit and walking in Auckland, New Zealand," *Journal of Transport Geography*, vol. 20, no. 1, pp. 15–22, 2012.

[22] L. Ma, N. Luo, T. Wan, C. Hu, and M. Peng, "An improved healthcare accessibility measure considering the temporal dimension and population demand of different ages," *International Journal of Environmental Research and Public Health*, vol. 15, no. 11, Article ID 2421, 2018.

[23] H. C. Chin and K. W. Foong, "Influence of school accessibility on housing values," *Journal of Urban Planning and Development*, vol. 132, no. 3, pp. 120–129, 2006.

[24] Y.-s. Nam, "Evaluation of locational accessibility to leisure facilities using GIS: focusing on Jeju city Dong area," *Korean Journal of Leisure, Recreation & Park*, vol. 43, no. 3, pp. 69–80, 2019.

[25] D. Oviedo, L. Scholl, M. Innao, and L. Pedraza, "Do rapid transit bus systems improve accessibility to job opportunities for the poor? The case of Lima, Peru," *Sustainability*, vol. 11, 2019.

[26] A. Owen and D. M. Levinson, "Modeling the commute mode share of transit using continuous accessibility to jobs," *Transportation Research Part A: Policy and Practice*, vol. 74, pp. 110–122, 2015.

[27] C. Guida and G. Carpentieri, "Quality of life in the urban environment and primary health services for the elderly during the Covid-19 pandemic: an application to the city of Milan (Italy)," *Cities*, vol. 110, 2020.

[28] O. Järv, H. Tenkanen, M. Salonen, R. Ahas, and T. Toivonen, "Dynamic cities: location-based accessibility modelling as a function of time," *Applied Geography*, vol. 95, pp. 101–110, 2018.

[29] P. Garcia-Albertos, M. Picornell, M. Henar Salas-Olmedo, and J. Gutierrez, "Exploring the potential of mobile phone

records and online route planners for dynamic accessibility analysis," *Transportation Research Part A: Policy and Practice*, vol. 125, pp. 294–307, 2018.

[30] M. Stpniak, J. P. Pritchard, K. T. Geurs, and S. Goliszek, "The impact of temporal resolution on public transport accessibility measurement: review and case study in Poland," *Journal of Transport Geography*, vol. 75, pp. 8–24, 2019.

[31] T. L. Lei, Y. Chen, and K. G. Goulias, "Opportunity-based dynamic transit accessibility in southern California," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2276, no. 1, pp. 26–37, 2012.

[32] K. S. Sahitya and C. Prasad, "Evaluation of opportunity based urban road network accessibility using GIS," *Spatial Information Research*, vol. 28, no. 1, 2020.

[33] G. Tarroni, O. Oktay, W. Bai, A. Schuh, H. Suzuki, and J. Passerat-Palmbach, "Learning-based quality control for cardiac MR images," in *IEEE Transactions on Medical Imaging*vol. 38, no. 5, p. 1127, 2019.

[34] G. Tarroni, O. Oktay, W. Bai, A. Schuh, and D. Rueckert, "Learning-based quality control for cardiac MR images," *IEEE Transactions on Medical Imaging*, vol. 99, 2018.

[35] W. Ji and S. M. AbouRizk, "Simulation-based analytics for quality control decision support: pipe welding case study," *Journal of Computing in Civil Engineering*, vol. 32, no. 3, Article ID 5018002, 2018.

[36] P. Y. Wei, J. L. Huang, and Y. Y. Chen, "Urban rail transit accessibility calculation method," *Journal of Chongqing Jianzhu University*, vol. 38, no. 10, pp. 1–6, 2019.

Hindawi
Complexity
Volume 2022, Article ID 5327023, 13 pages
https://doi.org/10.1155/2022/5327023

Research Article

Research on Urban Expansion and Population Density Change of an Urban Agglomeration in the Central-Southern Region of Liaoning Province, China

Min Guo[1] **and Shijun Wang**[2]

[1]*College of Earth Sciences, Jilin University, Changchun 130061, China*
[2]*College of Geographical Sciences, Northeast Normal University, Changchun 130024, China*

Correspondence should be addressed to Shijun Wang; wangsj@nenu.edu.cn

Received 26 January 2022; Revised 3 March 2022; Accepted 9 March 2022; Published 30 March 2022

Academic Editor: Haitao Ma

The relationship between urban expansion and population density changes is complex and plays a fundamental role in urban sustainable development research. This relationship has been studied in multiple large cities. However, there is no report of the relationship of the two factors mentioned above in urban agglomeration in a particular region of China. Ten cities located in the central-southern region of Liaoning province are selected as research samples in this study. The spatial growth rate and urban compactness index of the sample cities were calculated using the land use and population data of these cities in three time phases: 1995, 2005, and 2015. Then, the geographical growth features, the population density changes, and the link between these two in the sample cities were investigated at these three periods. Our results revealed that the studied cities of central-southern Liaoning province expanded at the three time phases under the driver of positive population density growth, but the effects of urban population change on urbanization growth modes across time were uneven across different periods. Our study showed a trend that active transition of different urbanization growth modes was associated with decreased population density.

1. Introduction

Urban expansion is a popular area of human geography study. Evolution characteristics of urban construction land's temporal and spatial patterns are evaluated by assessing gravity center shifts and intensity expansion through time and space [1, 2]. To date, most of the research on urban expansion focuses on analyzing its temporal and spatial evolution characteristics, such as the magnitude, elements, structures, and manners of urban spatial growth [3]. Primary studies have also discussed the mode, mechanism, and measurement method of urban spatial growth [4–7] and analyzed the urban growth process through constructing indicators, such as the degree of freedom, compactness, and sprawl of urban expansion [8, 9].

Urban population density is an important feature that is used to characterize the development of a city [10, 11]. With respect to urban population density, previous research has mainly focused on the following two aspects: (1) the spatial distribution patterns of urban population density, including the spatial distribution patterns and model research of urban population density at the national and provincial levels [12–14]; (2) temporal and spatial changes in urban population density, mainly from a global perspective and the perspective of urban districts to study population density trends and influencing factors [15–17]. Beyond these aspects, population density changes can affect and determine environment and urban development [18–20].

Over the past four decades, China has experienced the world's fastest urbanization process [21]. With a continuous increase in urban population across the whole country and an ongoing outward expansion of urban land, urban area expansion has developed at a quicker pace than urban population growth, resulting in a drop in urban population density over time [22, 23]. The feature of quick urban expansion is closely related to many urban factors and can be used to measure the development of a city [24, 25]. When there is an imbalance between land urbanization and population density, the reduction in urban population density due to quick urban expansion will result in low-density and fragmented city growth [18]. Eventually, this imbalance will lead to the destruction of city growth. In the context of rapid urbanization in China, it is crucial to comprehensively study the characteristics and mechanisms of urban population density fluctuations on urban expansion because this type of assessment can help guide the coordinated growth of urban land and population and thus establish a healthy form of urbanization development [19].

The central-southern region of Liaoning province is one of the earliest formed urban agglomerations in China [26, 27]. Some researchers believe that this area is the fourth-largest urban agglomeration after the Yangtze River Delta, the Pearl River Delta, and the Beijing Tianjin Hebei region [28]. The cities in this region provide the most significant sources of growth for Liaoning province and northeastern China [29, 30]. In 2010, China issued the "National Major Function Zone Plan," the first national land and spatial development plan [31]. The urban agglomeration in the central-southern area of Liaoning province was designated as an optimized development zone at the national level in this plan [31]. During the last decades, the development of the cities in this particular urban agglomeration was relatively slower than that of the other cities in China but still in process [32, 33]. In line with the urbanization of these cities, more and more countryside people have emerged in cities and settled down as citizens in these cities [34, 35]. To meet the growing need of these newcomers for living and production, urban land in urban agglomerations shows a trend toward continuous expansion [36, 37].

Although scholars have conducted a significant amount of research on urban expansion and population density [10, 14, 38, 39], the relationship between urban expansion rates and changes in population density within the urban agglomeration in the central-southern region of Liaoning has not yet been studied. The following questions guided our research on the relationship that has emerged between urban expansion rates and population density changes in this particular region: (1) "Will the increased population in this region promote city expansion of the urban agglomeration in the central-southern part of Liaoning province?" (2) "Can the theories of preexisting studies about the relationship between urban expansion rates and population density changes be applied to this urban agglomeration?"

In this study, the urban agglomeration in central-southern Liaoning province was selected as the study region. In total, ten cities were taken as research samples. We studied urban land density based on the cities' land use data from 1995, 2005, and 2015. The urban spatial growth rate at each period was calculated based on fitted parameters using the inverse sigmoid function (Section 2.3.2) to fit the relationship between urban land density and population density change. Based on the fitted parameters, the relationship between urban expansion and population density was assessed, and a compactness index that characterizes urban expansion was created. The study holds the potential of serving as the basis of sustainable urban development research and a valuable reference for future urban planning decisions and land policy formulations.

2. Data and Methods

2.1. Overview of the Research Area. Ten cities in the urban agglomeration located within the central-southern region of Liaoning province were taken as research samples (Figure 1). The urban agglomeration is located in the hinterland of Liao He Plain, one of the three major plains in northeast China. The city cluster is centered on two sub-provincial cities, which are Shenyang and Dalian. Besides the two centers, the urban agglomeration contains other eight prefecture-level cities (Anshan, Fushun, Benxi, Dandong, Liaoyang, Yingkou, Panjin, and Tie Ting) and 12 county-level cities, as well as 18 counties. At the end of 2015, the total population of the central-southern Liaoning urban agglomeration was 31.266 million, accounting for 73.69% of Liaoning province's total population. The total land area of this region is 96,690 km^2 with a built-up area of 1,666 km^2. The total urban land use development intensity reached 85.3% in 2015.

2.2. Data Sources. The research data consisted of regional land use data and population data that were collected in 1995, 2005, and 2015. The land use data comes from Landsat remote sensing images with a spatial resolution of 30 m. The land use types in this region were divided into six categories (construction land, waterbody, cultivated land, forest land, grassland, and unused land) based on the interpretation of remote sensing data. Urban population data comes from the census data of each city as is presented on the China Demographics Bureau's website (https://www.stats.gov.cn/).

2.3. Research Methods

2.3.1. Circle Analysis Method. Circle analysis refers to the establishment of a series of equidistant buffer zones with the city center as the center of the circle. Each circle was used as the basic unit for three separate purposes: (1) to describe the spatial differentiation of urban expansion, (2) to calculate relevant spatial indicators, and (3) to analyze the urbanization process to discover the spatial characteristics of different periods [39, 40].

We then analyzed the buffer zone of the urban land expansion process of each city. A buffer zone with a width of 1 km was established with the geometric center of the main urban area of each city in 1995 as the center and gradually expanding the outermost layer contained the main urban area of each city. Outer ring roads of the studied cities and contiguous built-up regions were used to determine the scope of the urban area (the scope determination of Shenyang city was selected as an example and illustrated in Figure 2). Based on the formed circle layer structure, the construction land density was calculated in each circle layer (i.e., the ratio of the construction land area in each circle layer to the total land area minus the nonconstructible land

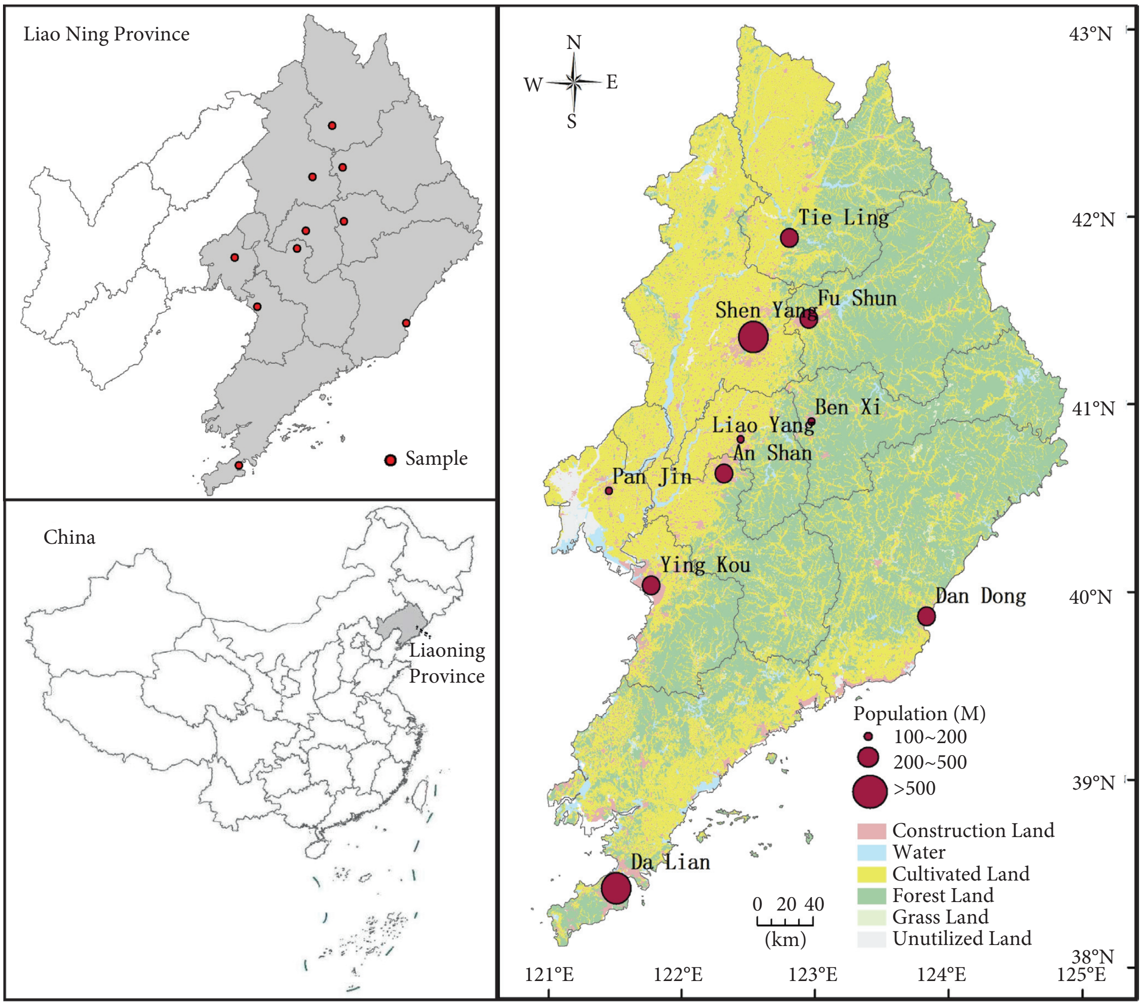


Figure 1: Schematic diagram of the study area location.

[water body]) [41]. The calculation formula for construction land density is as follows:

$$\text{Density} = \frac{S_{\text{construction}}}{S_{\text{land}} - s_{\text{water}}}. \quad (1)$$

In the formula, $S_{\text{construction}}$ is the construction land area of each circle in the buffer zone, S_{land} is the total land area of each circle in the buffer zone, and S_{water} is the water area of each circle in the buffer zone.

2.3.2. Inverse Sigmoid Method of Characterizing Urban Expansion. Jiao analyzed the decreasing law of impervious surface density with distance from the city center using three time phases of remote sensing photos from 28 major Chinese cities as the research sample and developed an inverse sigmoid function to fit the distribution of impervious surface density [42]. Using the inverse sigmoid function is helpful in modeling city expansion. Here, we took the function as a reference and used it to perform a fitting analysis on the land circle density distribution of ten sample cities, and the urban morphology of each city was determined based on the fitting results. The inverse sigmoid function is as follows:

$$f(\text{r}) = \frac{1-c}{1+e^{\alpha((2r/D)-1)}} + c. \quad (2)$$

In the formula, $f(r)$ indicates the density of the city; r represents the distance to the city center; e is Euler number; and α, c, and D are parameters for this investigation. The α value is a parameter that controls the slope of the inverse sigmoid function curve, the c value represents the background value of the urban hinterland construction land density, and the D value represents the fitted estimated value of the radius of the main urban area.

The growth rate of city radius D value is used to characterize urban spatial growth rate (V) in the following formula:

$$V = \sqrt[n]{\left(\frac{D_t}{D_o}\right)} - 1. \quad (3)$$

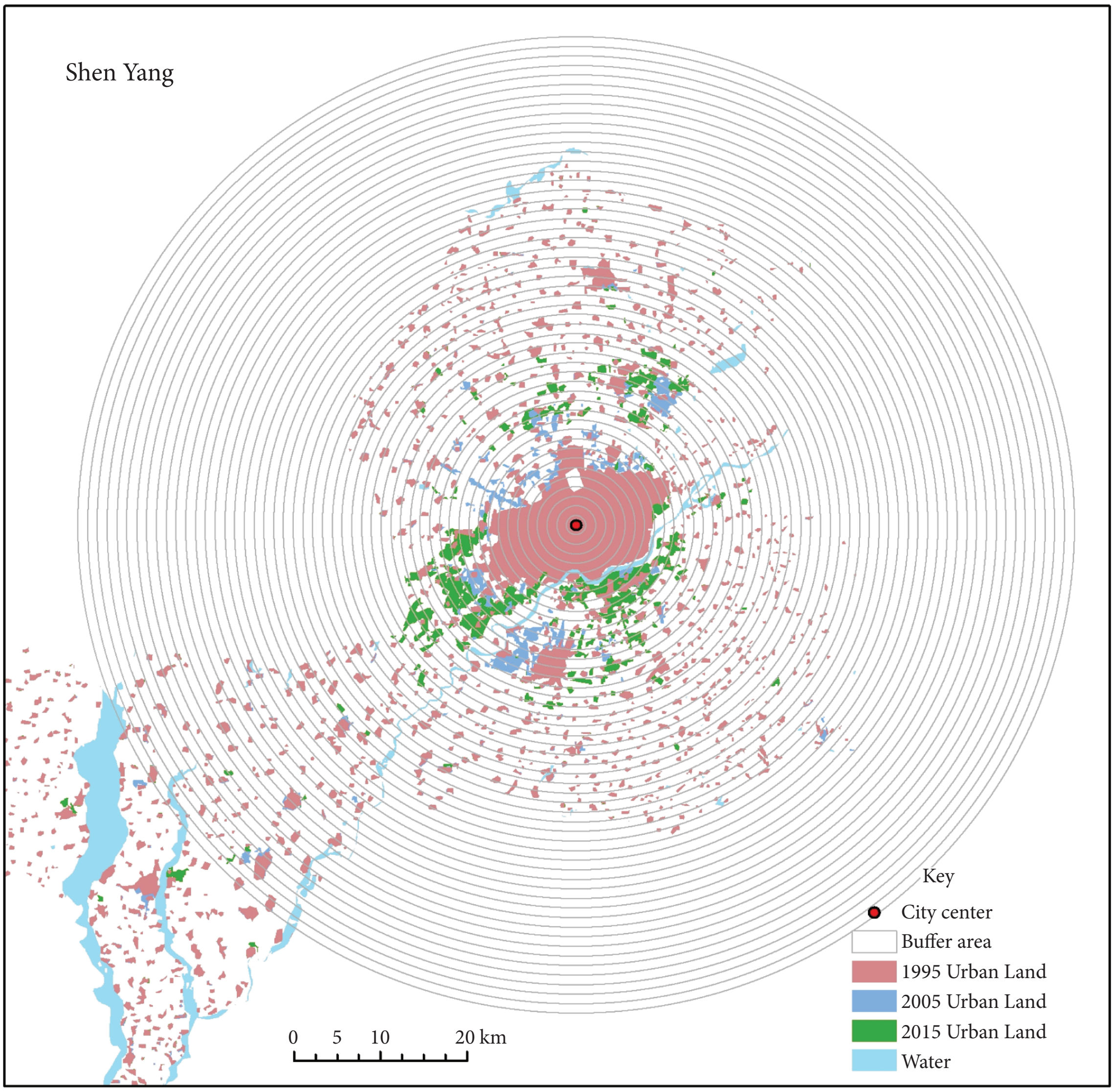


Figure 2: The buffer area and urban land area at three time points in Shenyang.

In the formula, D_t is the city radius at the end of the study, D_0 is the city radius at the beginning of the study, and n is the study period.

The compactness of each research point is calculated according to the parameters of the inverse sigmoid function fitting formula (2). The compactness value is the ratio of the inner urban area to the urban radius given by the function fitting. The compactness K_p value and the parameter α are inversely proportional to each other. If the K_p value at a later point in time decreases relative to the previous point in time, it indicates that the development of city urbanization followed a compact direction during the study period. The calculation formula for compactness (K_p) is as follows [42]:

$$K_p = \frac{r_2 - r_1}{D} = \frac{1.316957}{\alpha}. \quad (4)$$

In the formula, $r_2 - r_1$ represents the inner-city area, D is taken as the city radius, and α is the parameter that is used to control the slope of the inverse sigmoid function curve. The parameter D obtained by inverse sigmoid function fitting is the estimated value of the main urban area radius. By investigating the parameter D value change, the growth rates of each city space can be compared.

2.3.3. Method of Calculating the Relationship between City Expansion and Urban Population Density. The urban population density for each corresponding city is calculated by counting the population within the defined urban land area. Based on population density, population density changes are calculated during different research periods (1995–2005, 2005–2015, and 1995–2015). City expansion rates are obtained from the output result of (3). The correlation value of these two values with p value is calculated based on R statistics programming.

The relationship between city compactness and population density is studied separately with three different city growth modes, "compact," "maintenance," and "spread," characterized based on city compactness value K_p from (4). Generally, when the K_p value of a city increased from the previous point in time, the urban spatial growth model was defined as a "spread" mode. When the K_p value decreased, the urban spatial growth mode was defined as "compact." If the difference in K_p values between two periods was less than 0.01, the type of city growth mode is regarded as "maintenance."

3. Results

3.1. Population Density Change. In this study, we investigated urbanization by studying the urban expansion effects of urban population density. The urban population density was based on the census data from the government annual release of each city and is calculated as the ratio of urban permanent population to the built-up area (Supplementary Table 1). Figure 3 shows the change of overall population density between the three research time points (i.e., 1995, 2005, and 2015) in each city of the studied area. Comparing the population densities of different periods showed that the population density of the studied area in 2015 was lower than that in 1995 and 2005 (Figure 3(a)). The average population density of the studied area in 2015 was 12.42 people/hm^2, while the average population density in 1995 and 2005 was 13.901 people/hm^2 and 13.813 people/hm^2, respectively.

A comparison of the population density of each sample city as is depicted in Figures 3(b) and 3(c) reveals that three city types can be distinguished according to population density changes in the three different research periods (1995–2005, 2005–2015, and 1995–2015). The first type was characterized by a yearly increase in population. The second type consisted of cities whose population has increased and then declined. The third type included cities with annual population decline. The population densities of the studied cities Dalian, Shenyang, Panjin, and Yingkou increased during the studied periods. The population density increase rate of these four cities was 0.67%, 0.65%, 1.21%, and 1.72%. The population density of Benxi, Anshan, and Liaoyang increased between 1995 and 2005, while it decreased between 2005 and 2015, with a decline rate of 1.81%, 2.02%, and 4.23%, respectively. The population density of the other cities (Fushun, Dandong, and Tieling) showed a downward annual trend. Among these three cities, the population density of Fushun declined the most from 2005 to 2015, with a decline rate of 6.52%.

3.2. Urban Sprawl. Nonlinear least square method was used in the R software [43] to fit the land density of the ten cities by calculating the land density value of the corresponding circle layer of each city according to (2) in Section 2.3.2. Fitting parameter results of the inverse sigmoid function of the urban land density distribution are shown in Table 1, and the raw land density of the corresponding circle layer of each city is provided in Supplementary Table 2.

As depicted in Table 1, the fitting function effect is excellent with less than 0.01 of p values, indicating that the parameters from the fitted functions can accurately reflect city expansion. Generally, the parameters were summarized as follows. The α value reflects the compactness of the city form (the larger the α value, the more compact the city). The α values were taken at three time points in the sample cities with a value range of 1.319–7.198. The value of c is the land density value of the urban fringe hinterland, and the value range of c in the sample city was from 0.037 to 0.181. For most of the cities included in this study, the c value was less than 0.1. Some cities had a relatively large value, such as Panjin, which reached 0.181, suggesting that cities with this characteristic are almost always connected with the surrounding cities. The D value represents the city radius, and the city radius of each city changes with time. These fitting values increased with time, which suggests that the radiuses of these cities gradually increased during the periods analyzed in this study, but the growth rates were different. In 1995, the cities with the smallest and largest urban radiuses were Tieling and Dalian. In 2015, the cities with the smallest and largest urban radiuses were Tieling and Shenyang.

The spatial attenuation mode and fitting curve of the land density function of the ten cities are plotted and presented in Figure 4. The curve shape is used to reflect city expansion based on the following definition. (1) The curve of a compact city is relatively flat near the city center. (2) The curve of a maintenance city has a rapid decline from a steep shape on the middle part. (3) The curve of a rapidly expanding city shifts back to the right by a large margin. Overall, most of the sample cities were relatively compact in 1995, especially Dandong, Benxi, Fushun, and Shenyang. A few cities were relatively loose, such as Dalian and Anshan. The curves of these cities shifted backward to the right, which indicates that these cities were in a looser state. After 2005, some sample cities began to expand rapidly, and the extent of the expansion in the second period was greater than it was in the first one. For example, Shenyang and Yingkou experienced more significant expansions during the two studied periods. Compared with the cities that were in a status of significant expansions, the compactness degrees in Dandong, Fushun, Benxi, and Liaoyang has not changed significantly since 1995.

3.3. The Relationship between Urban Expansion and Population Density

3.3.1. The Relationship between Urban Expansion Rates and Population Density Change. The estimated value of the city radius D was obtained based on the inverse sigmoid function fitting, and city radius change was used to characterize the change in the urban spatial growth rate. The urban spatial growth rate of each city in three periods was calculated using (4) (Supplementary Table 3 and Figure 5(a)), and a correlation analysis was performed in accordance with the speed of population density change that was recorded during the same period (Figure 5(b)).

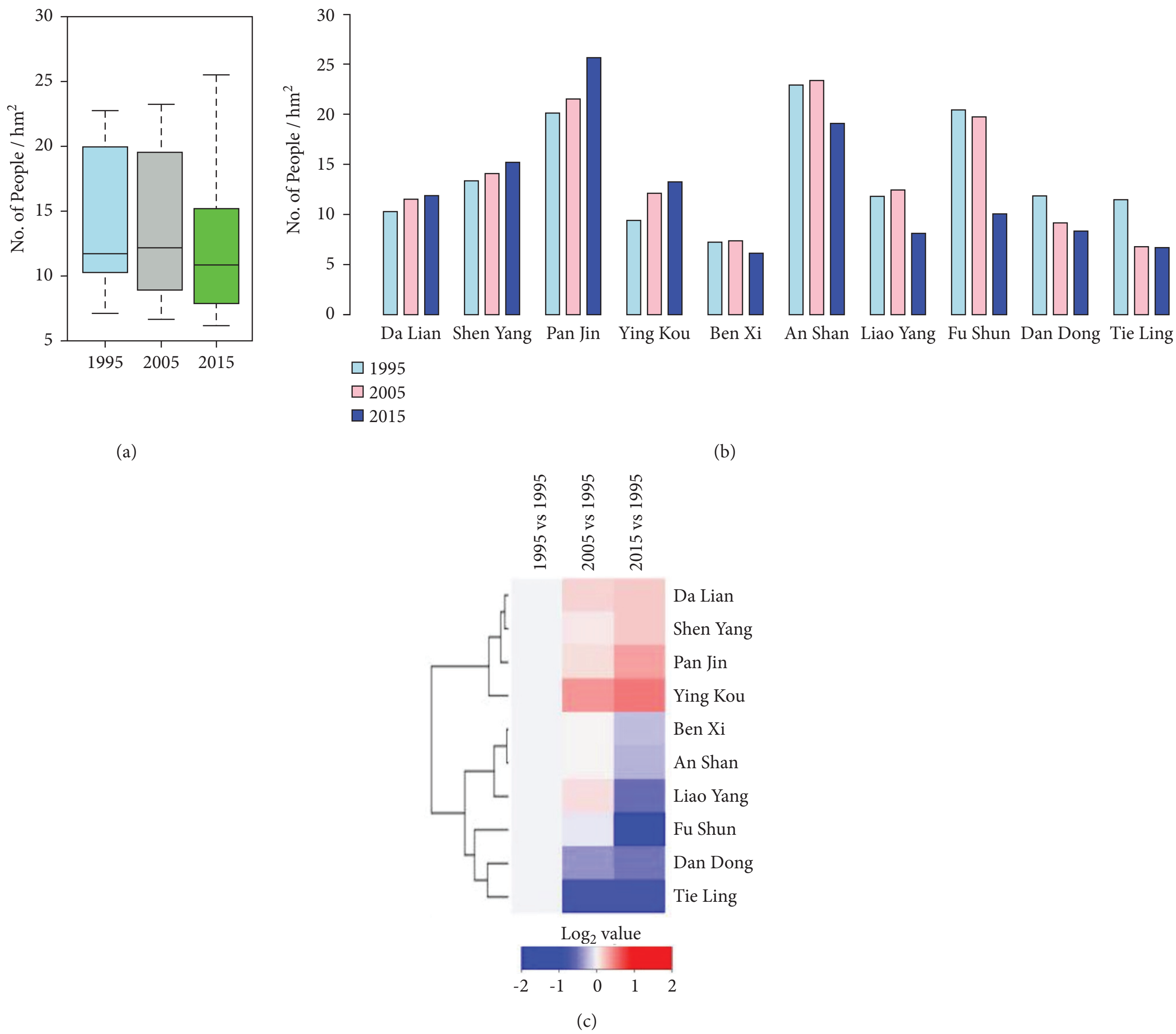


Figure 3: The population density of the sampled cities. The box plot (a) shows the average population of the studied cities during the three studied periods. The bar plot (b) represents the population density of the cities in the studied area in 1995, 2005, and 2015. The heatmap plot (c) depicts the population changes of each city between the three studied periods.

Table 1: Parameters of fitted inverse S-shaped functions in ten sample cities.

Cities	1995				2005				2015			
	a	c	D	p	a	c	D	p	a	c	D	p
Dandong	2.314	0.047	7.932	0.004	2.364	0.048	7.977	0.004	2.302	0.050	8.079	0.004
Dalian	1.319	0.039	17.885	0.043	1.666	0.037	22.271	0.020	1.798	0.044	22.581	0.012
Fushun	2.902	0.139	8.957	<0.001	3.021	0.149	9.087	<0.001	3.053	0.159	9.062	<0.001
Benxi	3.381	0.040	7.386	<0.001	3.389	0.041	7.374	<0.001	3.446	0.054	7.285	<0.001
Shenyang	4.604	0.095	17.779	<0.001	4.149	0.100	19.383	<0.001	3.486	0.098	22.829	<0.001
Panjin	3.066	0.178	6.310	<0.001	3.777	0.181	6.967	<0.001	3.292	0.177	8.318	<0.001
Yingkou	2.638	0.091	16.209	<0.001	2.543	0.082	17.958	<0.001	2.802	0.137	19.235	<0.001
Liaoyang	7.198	0.163	7.804	0.002	7.046	0.164	7.832	0.002	5.685	0.171	8.022	0.003
Tieling	5.897	0.049	6.306	0.001	5.615	0.065	6.354	<0.001	5.942	0.105	6.292	<0.001
Anshan	1.770	0.050	11.811	0.009	2.008	0.051	12.805	0.010	1.920	0.044	13.969	0.056

The p value represents the significance of function fitting.

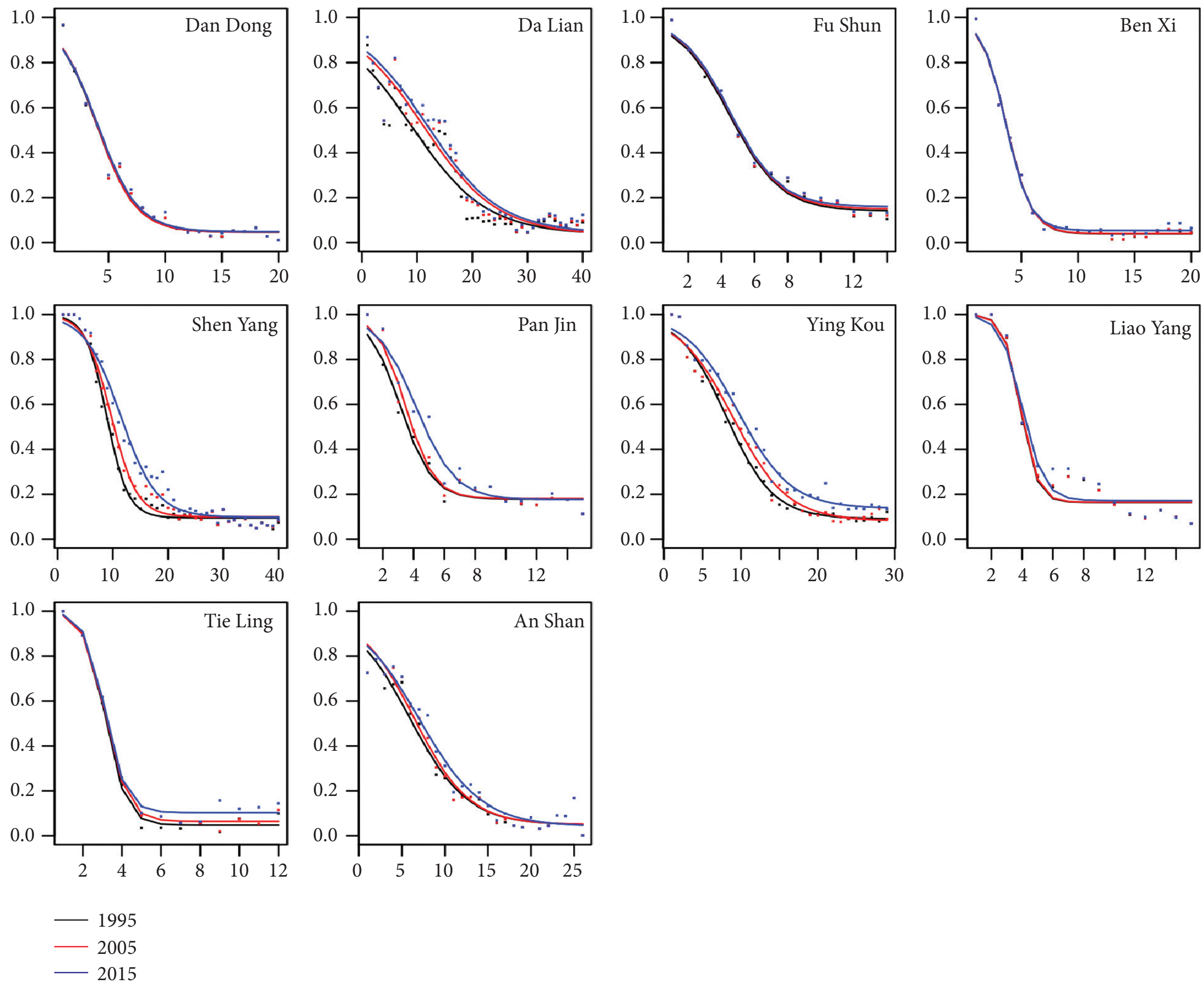


FIGURE 4: Spatial variation of urban land density in different circles with the distance to the city center and the curve fitted by inverse sigmoid function.

From 1995 to 2015, the spatial growth rates of the sample cities were between 0.00 and 0.015. Spatial growth rates of cities such as Dalian, Panjin, and Shenyang were higher than 0.01, while lower values (less than 0.01) were observed for the cities of Yingkou, Anshan, Liaoyang, and Dandong. The urban spatial growth rate decreased in five cities from 2005 to 2015, including Dalian, Fushun, Benxi, and Yingkou. However, the urban spatial growth rate of Dandong, Shenyang, Panjin, Liaoyang, Tieling, and Anshan increased during this same period. This means that the sampled cities are all expanding at different growth rates and that the changes in spatial growth rates are inconsistent. The correlation coefficient was calculated to study the correlation between the urban spatial growth rate and the population density change rate. The overall correlation coefficient was 0.613 with a p value of less than 0.01. A positive correlation between these two values indicated that the population growth might be one of the driving forces of urban expansion in the sampled cities.

3.3.2. The Relationship between City Compactness and Population Density Change. City expansion rate is one important attribute characterizing city expansion while city compactness determines urban sustainable development and growth [44]. City compactness is highly related to rising urban populations [45, 46]. To characterize the relationship between city expansion compactness and population density changes in the areas of this study, the value K_p was used to characterize the urban form at a certain time across an entire city radius and calculated as (4). The smaller the K_p value, the more compact the urban city forms and vice versa. The K_p values at two time points can be used to describe urban compactness change that occurs during two studied periods. In this article, if the K_p value of a city increased from the previous point in time, the urban spatial growth model was defined as a spread mode. If the K_p value decreased, the urban spatial growth mode was defined as compact. If the K_p values remained unchanged between the two periods (i.e., the difference in K_p values at each time point was less than 0.01), the urban spatial growth model was regarded as a maintenance type. Xu et al. also used this standard to classify city explanation in different periods [47] and showed robust results in studying population effects on city explanation. Based on this method, we calculated the compactness values K_p of the ten

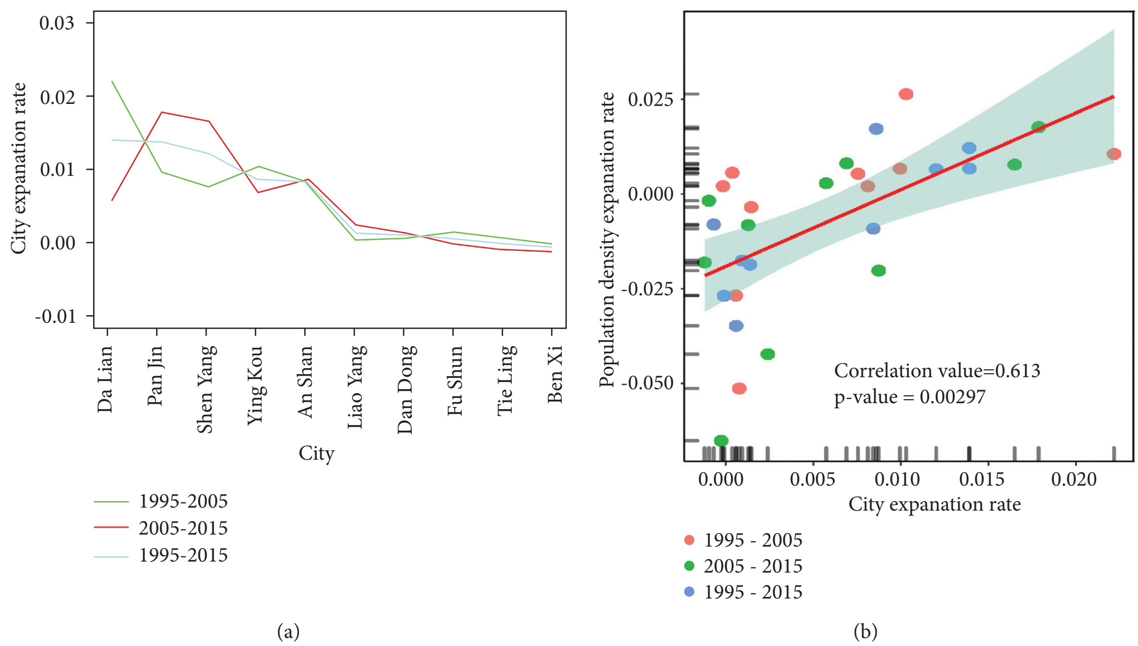


Figure 5: An overall observation of spatial growth rates and population density changes in the sample cities. The line plot (a) shows the spatial growth rate of sample cities in different periods. The point plot (b) shows the relationship between the urban expansion rate and the population density change rate.

sample cities at three time points (Figures 6(a)–6(c)) and then characterized urban spatial growth mode changes between the periods of 1990–2005 and 2005–2015 according to the difference in the compactness K_p values of the two time points (Supplementary Table 4 and Figure 6(d)).

The geometric growth rates of urban population density were calculated in relation to three different urban spatial growth modes of varying periods (Figure 7). Based on the analysis mentioned above, the relationship between the two was studied. Almost all the sample cities grew in a compact manner during the first period (Figure 6(d)), and the population density change rates of the cities in a compact mode had no significant changes (Figure 7). On the contrary, the cities in a spread mode (i.e., Shenyang, Yingkou, and Tieling) showed population density increase (Figures 6(d) and 7). In the second period, most of the sampled cities spread outward as they grew (Figure 6(d)), and their population density decline rates increased (Figure 7). This observation is consistent with Chinese development. During the period from 2005 to 2015, China was experiencing rapid urbanization, and cities expanded in a sprawling spatial growth mode, which led to an increase in the rate of population density decline. These findings suggest that a comparative analysis of sample cities at different development stages can serve as a valuable reference point for individuals who are making urban development decisions in rapidly urbanizing areas.

Our analysis also indicated that population density's influence on urban expansion was complex and manifested itself in different ways throughout the study periods. In the first period (1995–2005), the population density rate of sustaining and sprawling cities showed an upward trend. The increase in population promoted the growth of urban space to a certain extent, which suggests that population growth is the driving force of urban expansion. This result fits with the previous research [48]. However, in the second period (2005–2015), the population density of compact cities increased significantly compared with sprawling cities. At this time, it seems that an increase in population density was no longer the main factor of urban expansion, and economic development may have driven urban expansion [49]. In the longer period (1995–2015), both contraction and expansion cities had high growth rates, indicating the superposition of the different effects of urban population density change on urban expansion in the first two periods. In contrast, the maintenance-oriented cities show a downward trend, suggesting that the long-term reduction in urban population change will slow the rate of urban expansion.

Next, the urban spatial growth modes were analyzed in combination with the rate of population density change in each studied city in detail (Figure 8). In the two periods of 1995–2005 and 2005–2015, different types of transitions between cities were more likely to be manifested as a decline in urban population density, which suggests that the decline in population density promoted a shift in city type from spread to maintenance and then to compact, or vice versa. For example, in Fushun, the decrease in population density has caused the city to transform from the original spread type to the maintenance type. Population density decrease in Dandong and Anshan encouraged them to shift from being spread-type cities to becoming compact-type cities. Moreover, population density decrease brought on a transition change between urban types. The population density decrease that took place in Tieling between 2005 and 2015 was smaller than it was between 1995 and 2005,

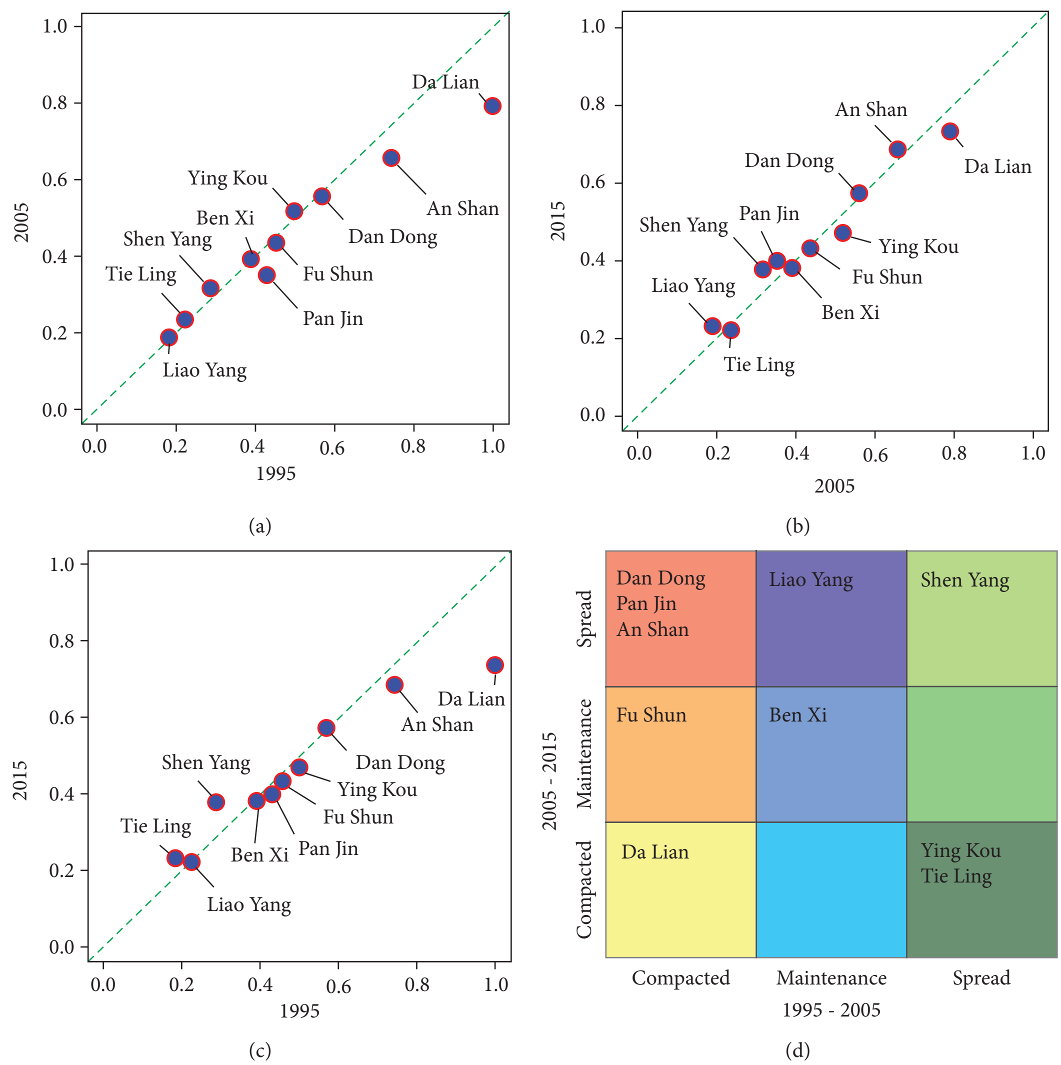


FIGURE 6: Land growth mode of each city. The point plots ((a)–(c)) show the comparison of the K_p values of cities between the periods of 1995–2005, 2005–2015, and 1995–2015. The schematic plot (d) represents the land growth modes of the sampled cities during each period.

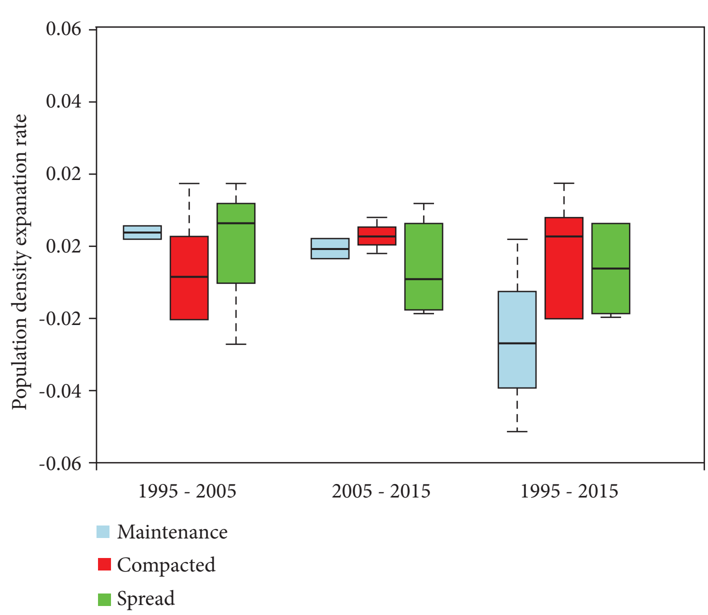


FIGURE 7: The population density variation rate of sample cities with different growth modes in each research period is depicted in a box plot.

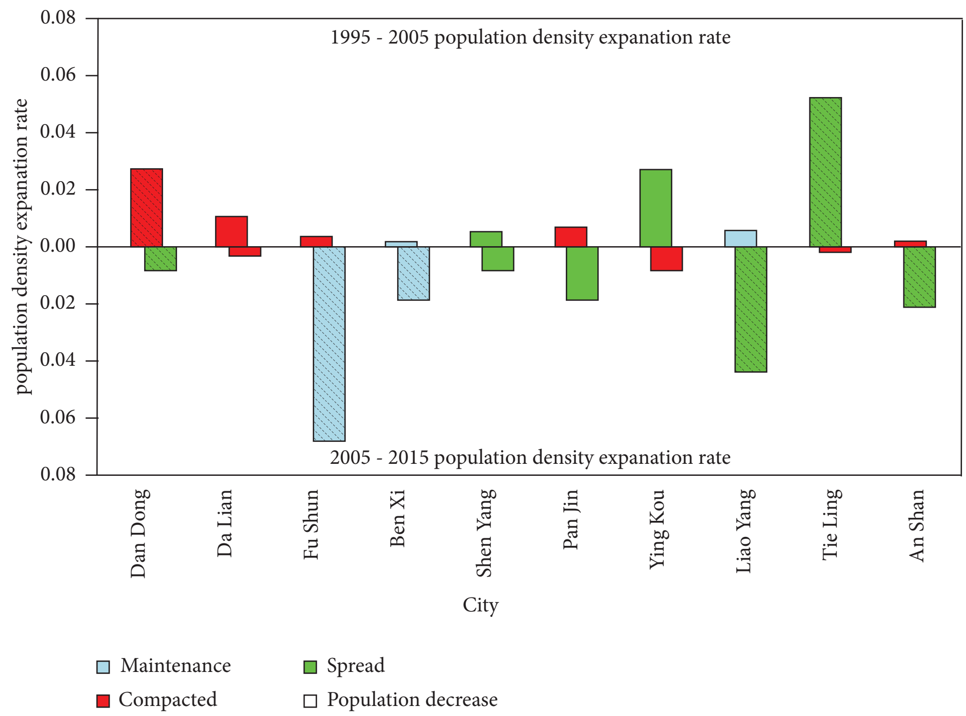


Figure 8: The relationship between the expansion type of the sampled cities and the rate of population density change is depicted in a bar plot.

which suggests that population density decrease promoted the transformation of urban types from compact to spread types. In summary, the types of urban expansion in the selected area of this study have changed significantly, except for Shenyang, Benxi, and Dalian, whose expansion types have not changed during the two periods. These three cities remained spread, maintenance, and compacted cities, respectively.

4. Conclusion

Ten cities located in the urban agglomeration of the central-southern region of Liaoning province were selected as research objects in this study. The studied area took around 73.69% of Liaoning province's total population, suggesting that population density plays important roles in this studied area. Our result showed that populations of some cities (e.g., Dalian, Shenyang, Panjin, and Yingkou) in this studied area increased while the population of others decreased. The population study provided a foundation for this study in understanding the effects of population on urban land use. The urban land density was studied through circle analysis, which is a method of analyzing the expansion of urban land use based on calculating construction land use on each circle from the city center [50, 51]. Based on the urban land growth analysis on each circle through fitting to inverse sigmoid function, we found that all the studied cities expand in construction area across the three studied periods. The fitting to inverse sigmoid function can not only provide the size of urban area but also be used to calculate compactness, which tells the quality or density of city expansion, and to reflect city expansion modes [38, 39, 42]. The calculation of compactness showed that the city expansion modes changed during different periods of this study in the study area of the urban agglomeration in Liaoning province.

We integrated population density changes and urban land use information by systematically analyzing the relationships between these two factors. Integrating the urban expansion rate and urban population density data showed that the urban population density was positively correlated with the urban expansion rate, suggesting that urban population density growth was a driving force for urban expansion of the cities in the research area of this study. Although we observed that low population density and a decline in moderate population density were prevalent in the area of this study, the population increase was strongly associated with significant city expansion in relatively large cities (e.g., Shenyang and Dalian), strongly proving that population increase was a driver of city expansion.

Furthermore, the relationship between population density and city compactness was different in the two studied periods (i.e., 1995–2005 and 2005–2015). Our results showed that urban population density drives the cities to a "compact" development mode during the period from 1995 to 2005. On the contrary, the increase in urban population density from 2005 to 2015 is related to the "spread" development mode. These results clarified that the effects of urban population on urban expansion changes were inconsistent in different periods and highlighted the flexible development of urban growth in the study region. More interestingly, detailed analysis showed that the decrease in population density had a positive effect on the conversion of urban types, suggesting that population density is a determinant factor in controlling city expansion modes.

5. Discussion

This study took the urban agglomeration in the central-southern area of Liaoning province as a research region. The cities in this area are representative and comparable to other city clusters in northeastern China because the region is more early developed and is one of the economic centers of the northeastern region of China [26, 37]. Our results showed that the urban population density of some cities (such as Benxi, Anshan, Liaoyang, Fushun, Dandong, and Tieling) in the central-southern area of Liaoning province followed a downward trend while the population increased in some other cities (i.e., Dalian, Shenyang, Panjin, and Yingkou). Generally, the population in northeast China decreases dramatically [52]. However, this study showed that the population of the relatively larger cities (e.g., Shenyang and Dalian, which are the sub-provincial cities of the studied area) increased during the two studied periods while the population in the small cities showed a decreasing trend. This is possible due to population movement during urbanization from countryside to city and from small cities to large cities [53, 54]. The further gathering of the population data in countryside and small cities can help in figuring out population movement and understanding urban population density change.

We revealed that the urban expansion rate was positively correlated with population density changes. This finding is different from the negative correlation of the relationship in another study that focused on large cities in Europe and China [38]. One possibility is that the urban expansion rates of medium-sized cities in this study are different from those of large cities. The results of this study thus fill several gaps that exist in research on urban expansion and population density changes in medium-sized cities. This finding suggests further interesting research on the comparison of the relationship between population density changes and urban expansion in cities of different sizes.

On the other hand, the relationship between population density and city compactness was different in the two studied periods (1995–2005 and 2005–2015), which suggests that population density changes show significant temporal characteristics for urban expansion in different years. This is probably due to other determination factors in urbanization, such as population loss during the studied periods and the urbanization policy of China during recent years (2010–2020) [55]. We suggest further research should consider the combined effects of these factors with population density due to the complicated relationship between population density changes and urban expansion. Based on the fact that the severe population loss observed in the studied region in recent years (2010–2020) will gradually become significant [55], it is possible that the relationship between population density changes and urban expansion in the region will become more complicated in the future.

This study specifically focused on the relationship between urban expansion and changes in urban population density in ten cities of the central-southern area of Liaoning province. However, urban expansion is a complex process that is influenced by the joint action of multiple variables [56, 57]. Because of limitations in data availability, this paper only discussed the difference between urban spatial growth and population density changes that have taken place since 1995 in the urban agglomeration located in the central-southern region of Liaoning province. Horizontally, future research is needed to compare the various urban agglomerations in China, examine the differences and similarities between typical urban agglomerations in different countries, and explore the development differences between different urban agglomerations. A comprehensive analysis of the relationship between urban growth and population density changes in these different urban agglomerations can be used to compare and analyze the impact of regional characteristics on city types and urban expansion. Vertically, it is necessary to go deep into the interior of the city and explore the association between spatial growth and population density changes on a smaller scale, such as polycentric urban structures of large cities. We did not consider polycentric urban structures of the studied cities due to the lack of population density datasets at that scale even though one of the studied cities (Shenyang) has been proved with three subcenters [58]. Further study should be carried out to understand the relationship of population density and city expansion at polycentric urban structures. Moreover, the relationship at the county or town level is worth studying if population dataset in this level is available. In future, with the continuous emergence of massive amounts of data [59], combining multisource data to perceive the density changes of different population types will help broaden our understanding of the micro mechanisms of urban spatial growth and urban population density changes.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

G. M. contributed to methodology, software, and formal analysis. W. S. J. was responsible for supervision, project administration, and funding acquisition. Both authors participated in conceptualization, validation, investigation, resources, data curation, and visualization; prepared the original draft; revised and edited the manuscript; and read and agreed to the published version of the manuscript.

Acknowledgments

The authors thank Dr. Dong Liguo for revising the manuscript. This research was funded by the key project of the National Natural Science Foundation of China, "Research on Spatial Process and Comprehensive Effect of Northeast Revitalization," grant no. 41630749.

Supplementary Materials

Supplementary Table 1: population density and population density change rate. Supplementary Table 2: construction area in circle analysis of each city. Supplementary Table 3: city expansion rate. Supplementary Table 4: K_p values and urban expansion mode. (*Supplementary Materials*)

References

[1] H. Li, Y. D. Wei, F. H. Liao, and Z. Huang, "Administrative Hierarchy and Urban Land Expansion in Transitional China," *Applied Geography*, vol. 56, pp. 177–186, 2015.

[2] L. Muxuan, S. Yingchun, C. Yangfen, Y. Deqing, H. Qiongfeng, and W. Liangjian, "A study on spatial-temporal features of construction land expansion in Changsha urban area," *Geographical Research*, vol. 26, no. 2, pp. 265–274, 2007.

[3] Z. Chunshan and Y. Changdong, "Features and causes of urban spatial growth in Chinese metropolises," *Acta Geographica Sinica*, vol. 68, no. 6, pp. 728–738, 2013.

[4] J. Jinliang, Z. Liang, W. Wenjia, S. Dongqi, and X. Jiangang, "A compatative study on spatiotemporal characteristics of urban expansion for the central cities along the Yangtze River--A case study of NanJing,WuHan and ChongQing," *Resources and Environment in the Yangtze Basin*, vol. 24, no. 9, pp. 1528–1536, 2015.

[5] L. Xiaolong and S. Jianfa, "Urban growth between cities: the case of jiangyin economic development zone in jingjiang," *Acta Geographica Sinica*, vol. 061, no. 4, pp. 435–445, 2006.

[6] C. Jianglong, G. Jinlong, X. Mengyue, and C. Wen, "Characteristics and mechanism of construction land expansion in Nanjing metropolitan area," *Geographical Research*, vol. 34, no. 8, pp. 946–954, 2014.

[7] M. Ronghua, M. A. Ronghua, G. U. Chaolin, P. U. Yingxia, M. A. Xiaodong, and Z. H. U. Chuangeng, "Urban spatial sprawl pattern and metr ics in south of jiangsu province along the Yangtze River," *Acta Geographica Sinica*, vol. 62, no. 10, pp. 1011–1022, 2007.

[8] T. Luyi and H. Shougeng, "Characterizations of urban sprawl in major Chinese cities," *Resources Science*, vol. 38, no. 1, pp. 50–61, 2016.

[9] L. Jialin, X. Jiqin, L. Weifang, and L. Chuang, "Spatio-temporal characteristics of urbanization area growth in the Yangtze River Delta," *Acta Geographica Sinica*, vol. 62, no. 4, pp. 437–447, 2007.

[10] V. Volpati and M. Barthelemy, "The spatial organization of the population density in cities," 2018, https://arxiv.org/abs/1804.00855.

[11] P. Francesca and S. Christopher, "Analysis of Urban Land Cover and Population Density in the United States," *Photogrammetric Engineering & Remote Sensing*, vol. 71, no. 6, pp. 719–726, 2015.

[12] S. Zhuoran, H. Yaping, and Z. Xianchun, "Spatial distribution of population density of typical megacities in Central China case study of wuhan," *Economic Geography*, vol. 35, pp. 33–39, 2015.

[13] W. Anhua, Z. Bo, P. You, and Z. Meng, "Forecasting urban population density using a multi-variable grey system mode," *Statistics & Information Forum*, vol. 33, no. 8, pp. 60–67, 2018.

[14] F. Jian, "Modeling the spatial distribution of urban population density and its evolution in Hangzhou," *Geographical Research*, vol. 21, no. 5, pp. 635–646, 2002.

[15] L. Xiaoling, "Comparing spatial pattern of population density of sino-Russian large costal metropolitans: case study of St.Petersburg and dalian," *Economic Geography*, vol. 38, no. 9, pp. 78–86, 2018.

[16] H. Jie and Z. Yaxi, "China's urban population density and changes," *Urban Problems*, vol. 2014, no. 10, pp. 17–22, 2014.

[17] X. Gang, L. Jiao, M. Yuan, T. D. B. Zhang, and C. Du, "How does urban population density decline over time? An exponential model for Chinese cities with international comparisons," *Landscape and Urban Planning*, vol. 183, pp. 59–67, 2019.

[18] Z. Fengkai, Z. Fengrong, L. Can, J. Pengfei, and W. Jingxia, "Coordination and regional difference of urban land expansion and demographic urbanization in China during 1993-2008," *Progress in Geography*, vol. 33, no. 5, pp. 647–656, 2014.

[19] C. Tonne, L. Adair, D. Adlakha et al., "Defining pathways to healthy sustainable urban development," *Environment International*, vol. 146, Article ID 106236, 2021.

[20] M. L. Hunter, *Population and Environment: A Complex Relationship*, RAND Corporation, Santa Monica, CA,USA, 2000.

[21] J. Limin and L. Yaolin, *Sustainable Urbanization and Territorial Spatial Optimization*, pp. 1–13, Geomatics and Information Science of Wuhan University, Wuhan, China, 2020.

[22] J. K. Brueckner, "Urban sprawl: diagnosis and remedies," *International Regional Science Review*, vol. 23, no. 2, pp. 160–171, 2000.

[23] S. Angel, J. Parent, D. L. Civco, A. Blei, and D. Potere, "The dimensions of global urban expansion: estimates and projections for all countries," *Progress in Planning*, vol. 75, no. 2, pp. 53–107, 2011.

[24] Y. Zhang, X. Li, S. Wang et al., "A global North-South division line for portraying urban development," *iScience*, vol. 24, no. 7, Article ID 102729, 2021.

[25] B. Güneralp, M. Reba, B. U. Hales, E. A. Wentz, and K. C. Seto, "Trends in urban land expansion, density, and land transitions from 1970 to 2010: a global synthesis," *Environmental Research Letters*, vol. 15, no. 4, Article ID 044015, 2020.

[26] Y. Yingen, "Urban Agglomeration Formation and Industrial Spatial Distribution," *A Study Based on New Economic Geography*, 2013.

[27] M. Yanji, "The old industrial base in central and southern Liaoning should pay close attention to formulating the development strategy of industrial agglomeration," *National conditions and strength of China*, vol. 000, no. 5, pp. 62–64, 2008.

[28] S. Xiaowu, W. Shiguo, and X. Xin, *The Great Change in the Regional Economy of China under the New Normal (The Great Transformation of China)*, Springer, 2019.

[29] W. Shi-Jun, S. Yang, F. Zhang-Xian, and J. Li-Li, "Pattern and progress of large urban agglomerations and urban flows intensity in northeast China," *Scientia Geographica Sinica*, vol. 31, no. 3, pp. 287–294, 2011.

[30] J. Bo, X. ChunLiang, and C. Cai, "Analysis of urban flow and model explanation on the urban group in middle-south part of liaoning province," *Economic Geography*, vol. 028, no. 5, pp. 853–856, 2008.

[31] T. S. Council, "National Major Function Zone Plan," 2021, http://www.gov.cn/zwgk/2011-06/08/content_1879180.htm.

[32] M. Xiangfei and L. Yumei, "Urban development theory and the choice of innovative development mode of urbanization in Liaoning," *Journal of Northeast University of Finance and Economics*, vol. 5, no. 5, pp. 3–10, 2010.

[33] H. Mo, Z. Xuena, and W. Chen, *Promoting the High-Quality Development of Urban Construction in Liaoning with Urban Renewal. Liaoning Economy*.

[34] J. Zaifang, Y. Xiaoli, and Z. Yuna, "Analysis on the transfer of rural labor force in Liaoning Province - based on the survey of 128 villages in 14 cities in Liaoning Province," *Agricultural Economy*, no. 7, p. 3, 2014.
[35] Z. Yuanqing, "Research on the construction of new urbanization and the willingness of rural residents to move to cities R - Based on the survey data of some districts and counties in Liaoning Province," *Consumer Guide*, vol. 7, p. 164, 2018.
[36] C. Ming, X. Han, and X. Wenhui, "Study on Influencing Factors of construction land scale expansion in Shenyang from 1999 to 2015," *Agricultural resources and Regionalization in China*, vol. 38, no. 7, p. 7, 2017.
[37] W. Yufeng and C. Hongjun, "Temporal and spatial characteristics and driving factors of urban agglomeration expansion from Harbin Changchun to central and southern Liaoning," *Journal of Northeastern University: Natural Science Edition*, vol. 42, no. 5, p. 6, 2021.
[38] Z. Rui, J. Limin, and X. Gang, "The relationship between urban spatial growth and population density change," *Acta Geographica Sinica*, vol. 75, pp. 695–707, 2020.
[39] J. Limin and Z. Xin, "Characterizing Urban Expansion of Main Metropolises in China based on built-up densities in concentric rings," *Resources and Environment in the Yangtze Basin*, vol. 24, no. 10, pp. 1721–1728, 2015.
[40] L. Xiaowen, F. Jingyun, and P. Shilong, "The comparison of spatial characteristics in urban landuse growth among the central and sub-cities in Shanghai Region," *Geographical Research*, vol. 22, no. 6, pp. 769–779, 2003.
[41] G. Xiangdong and W. Wenyu, "Population redistribution and modelling of shanghai city in the 1990s," *Acta Geographica Sinica*, vol. 60, no. 4, pp. 637–644, 2005.
[42] J. Limin, "Urban land density function: a new method to characterize urban expansion," *Landscape and Urban Planning*, vol. 139, pp. 26–39, 2015.
[43] R. C. R Team, *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021.
[44] S. Abdullahi, B. Pradhan, and H. Mojaddadi, "City compactness: assessing the influence of the growth of residential land use," *Journal of Urban Technology*, vol. 25, no. 1, pp. 21–46, 2018.
[45] M. Nadeem, A. Aziz, M. A. Al-Rashid, G. Tesoriere, M. Asim, and T. Campisi, "Scaling the potential of compact city development: the case of lahore, Pakistan," *Sustainability*, vol. 13, no. 9, p. 5257, 2021.
[46] Z. Qianru and B. D. Leibowicz, "Vehicle efficiency improvements, urban form, and energy use impacts," *Cities*, vol. 97, Article ID 102486, 2020.
[47] G. Xu, L. Jiao, J. Liu, Z. Shi, C. Zeng, and Y. Liu, "Understanding urban expansion combining macro patterns and micro dynamics in three Southeast Asian megacities," *The Science of the Total Environment*, vol. 660, pp. 375–383, 2019.
[48] Y. Song, L. Jiao, J. Liu, and G. Xu, "Analysis of the Factors Affecting the Degree of Urban Expansion: Taking Wuhan City as an Example," *Geomatics and Information Science of Wuhan University*, vol. 46, 2021.
[49] W. Shaojian and W. Zehong, "The spatio-temporal differentiation of urban land expansion and influencing factors in the Yangtze River Delta in the economic transition context," *Journal of Natural Resources*, vol. 36, pp. 993–1007, 2021.
[50] M. You, R. Lai, J. Lin, and Z. Zhu, "Quantitative analysis of a spatial distribution and driving factors of the urban heat island effect: a case study of fuzhou central area, China," *International Journal of Environmental Research and Public Health*, vol. 18, no. 24, 2021.
[51] Z. G. Wang Fazeng, Z. Ding, and J. Liu, "Spatial organization of urban system in central plains economic region," *Progress in Geography*, vol. 33, no. 2, pp. 153–168, 2014.
[52] H. You, J. Yang, B. Xue et al., "Spatial evolution of population change in Northeast China during 1992–2018," *Science of The Total Environment*, vol. 776, Article ID 146023, 2021.
[53] C. Yongjun and Q. Guo, "Study on regional differences of population mobility and urbanization in China," *Inquiry into economic issues*, vol. 2013, pp. 36–40, 2013.
[54] S. Fei and Z. Pingyu, "Temporal and spatial evolution characteristics of population distribution of urban agglomeration in central and southern Liaoning," *Progress in Geographical Science*, vol. 29, no. 1, pp. 96–102, 2010.
[55] S. Pingjun and W. Kewen, "Identification and stage division of urban shrinkage in the three provinces of Northeast China," *Acta Geographica Sinica*, vol. 76, pp. 1366–1379, 2021.
[56] J. M. Hassell, M. Begon, M. J. Ward, and E. M. Fèvre, "Urbanization and disease emergence: dynamics at the wildlife-livestock-human interface," *Trends in Ecology & Evolution*, vol. 32, no. 1, pp. 55–67, 2017.
[57] H. Wang, Y. Liu, G. Zhang, Y. Wang, and J. Zhao, "Multi-scenario simulation of urban growth under integrated urban spatial planning: a case study of wuhan, China," *Sustainability*, vol. 13, Article ID 11279, 2021.
[58] J. Yang, H. Yu, T. Li, Y. Jin, and D. Sun, "Morphological and functional polycentric structure assessment of megacity: an integrated approach with spatial distribution and interaction," *Sustainable Cities and Society*, vol. 80, Article ID 103800, 2022.
[59] Y Jun, Y. Haolin, Z. Yuqing, and J. Cui, "Research progress on human settlements: from traditional data to big data," *Progress in Geographical Science*, vol. 39, no. 1, p. 11, 2020.

Hindawi
Complexity
Volume 2022, Article ID 9373911, 20 pages
https://doi.org/10.1155/2022/9373911

Research Article

Technology Topic Identification and Trend Prediction of New Energy Vehicle Using LDA Modeling

Renjie Hu,[1] **Wencong Ma**,[2] **Weiqiang Lin**,[2] **Xiude Chen**,[3] **Zuchang Zhong**,[1] **and Chuhong Zeng**[4]

[1]*Guangdong Research Institute for International Strategies, Guangdong University of Foreign Studies, Guangzhou 510420, China*
[2]*School of Management, Guangdong University of Technology, Guangzhou 510520, China*
[3]*Guangdong Education Big Data Research Center, Guangdong University of Technology, New Engineering Education Research Center of Guangdong University of Technology, Guangzhou, China*
[4]*School of Business, Guangdong University of Foreign Studies, Guangzhou 510006, China*

Correspondence should be addressed to Wencong Ma; mawencong@gdut.edu.cn

Received 24 November 2021; Revised 6 February 2022; Accepted 21 February 2022; Published 23 March 2022

Academic Editor: Zhipeng Tang

As new energy vehicle (NEV) is the future of automobile development, it is of great significance to dig deeper into the technical topics and development trends of new energy vehicles for accurately understanding the technical trends of the new energy vehicle industry, grasping development opportunities, and scientifically formulating strategic plans. This paper takes the patent texts in the field of new energy vehicles from 2000 to 2020 in the patent database of CNKI as the data source, identifies 25 technical topics implied in the patent texts by using the LDA (Latent Dirichlet Allocation) topic model, analyzes the evolution trend of the 25 technical topics in terms of importance and popularity, and predicts the popularity and development trend of each technical topic in new energy vehicles from 2021 to 2025 by constructing the ARIMA model. The popularity and development trend of each technology topic of new energy vehicles in China from 2021 to 2025 are predicted by constructing ARIMA model. Drawing on quantitative evidence, the study found that there are top five technical topics in terms of importance in this field, namely, Topic 8 (Installation and Fixation), Topic 5 (Heat Dissipation), Topic 14 (Vehicle Data Monitoring), Topic 9 (Charging Pile), and Topic 15 (Damping). From 2014 to 2020, the importances of Topic 5 (Heat Dissipation), Topic 8 (Installation and Fixation), Topic 6 (Electric Drive System), Topic 9 (Charging Pile), and Topic 15 (Damping) are gradually rising. In terms of popularity of technical topics, from 2014 to 2020, the first to fifth topics are Topic 20 (Safety), Topic 8 (Installation and Fixation), Topic 3 (Cable Insulation Materials), Topic 15 (Damping), and Topic 10 (Pump Cooling). Based on the prediction of ARIMA model, it is found that the popularity of these five technical topics is steadily increasing from 2021 to 2025, among which the popularity of Topic 20 (Safety) will increase from 63.58 to 113.07, the largest increase in popularity among all technical topics. The paper provides implications for countries dedicated to developing the new energy vehicle industry.

1. Introduction

Countries around the world have different expressions about new energy vehicles, which are called "alternative fuel vehicles" in the United States, "low pollution vehicles" in Japan, and "new energy vehicles" in China. China has used the terms "electric vehicles," "low energy consumption and new energy vehicles," and "new energy vehicles" in its terminology for new energy vehicles. According to the "Entry Management Rules for New Energy Vehicle Manufacturing Enterprises and Products" implemented in China on July 1, 2009, a new energy vehicle refers to a vehicle that uses unconventional automotive fuels as its power source (or uses conventional automotive fuels but with a new vehicle power unit) and integrates advanced technologies in power control and drive of the vehicle to form an advanced technical principle with new technologies and structures [1]. There are five main types of new energy vehicles, namely, fuel cell new energy vehicles, pure electric vehicles, hybrid vehicles, hydrogen engine vehicles, and other new energy vehicles. Since then, the expression "new energy vehicles" has been used uniformly in government documents, enterprises, and academia.

With the air pollution caused by the exhaust emission of fuel vehicles and continuous consumption of nonrenewable resources, it is irreversible to change the dependence on oil and develop clean energy such as water, wind, and light as the driving force of vehicles. In recent years, the world's major automotive countries have strengthened strategic planning and policy support, multinational automotive enterprises to increase investment in research and development and improve the industrial layout; new energy vehicles have become the direction of transformation and development of the global automotive industry and promote the world's sustained economic growth of an important engine. Since 2018, the global output of new energy vehicles has been growing rapidly. In 2019, the global output of new energy vehicles was 2.17 million, and, in 2020, the output was 2.55 million. In 2020, the European new energy vehicles market accounted for 43.06% of the global market, followed by 41.27% in China, 10.12% in the United States, and 0.96% in Japan. In order to implement the State Council's decision to develop strategic emerging industries and strengthen energy conservation and emission reduction and to accelerate the cultivation and development of energy conservation and new energy vehicle industry, the State Council issued the "Energy Conservation and New Energy Vehicle Industry Development Plan (2012–2020)" in 2012. In November 2020, the General Office of the State Council issued the Development Plan for New Energy Vehicle Industry (2021–2035), which calls for the in-depth implementation of the national strategy for developing new energy vehicles, promoting the high-quality and sustainable development of China's new energy vehicle industry, and accelerating the construction of an automobile power.

Since 2012, China has adhered to the strategic orientation of pure electric drive and made great achievements in the development of new energy automobile industry, becoming one of the important forces in the development and transformation of the world automobile industry. According to the China Association of Automobile Manufacturers (CAAM), the production and sales of new energy vehicles in China were 1.366 million and 1.367 million, respectively, in 2020, up 7.5% and 10.9% year-on-year, with production and sales reaching a record high. Meantime, the development of new energy vehicles in China also faces problems such as weak innovation capability of core technologies [2, 3], infrastructure construction still lags [4], the industrial ecology is not yet sound [5, 6], the quality assurance system needs to be improved [7, 8], and the market competition is increasing [9–11]. Therefore, based on patent literature data, identifying the technical topics of new energy vehicles and predicting the development trend of new energy vehicles are important for tracking the technological frontier of new energy vehicles, grasping technological development opportunities, improving R&D (research and development) efficiency, perfecting industrial ecology, and achieving high-quality development.

The rest of this paper is organized as follows. Section 2 sorts out the literature review. Section 3 is devoted to the description of the relevant methodology. The empirical results are presented in Section 4. Section 5 proposes policy implications and concludes the paper.

2. Literature Review

2.1. Technical Topic Identification and Forecast. As the most important output result in science and technology innovation, patent literature carries 90%–95% of technical information worldwide [12]. Data mining and topic identification of patent literature can not only shorten R&D time and save research expenditure but also help to grasp the frontier of technological innovation and predict technological development trend. There are mainly patent classification method, patent citation method, and text mining method for technology topic identification based on patent literature.

2.1.1. Analysis Method Based on Patent Classification. Patent classification is a simple and generic technology classification system provided based on the technical content revealed by the patent [13]. The main technical subject analyses based on patent classification numbers are statistical analysis and coclassification analysis. Jun used international patent classification numbers as the technical topic of literature in a key technology prediction study [14]. Kostoff et al. used word frequency analysis to identify the technical topic of disruptive technologies [15]. Zhang et al. predicted the future development trend of driverless cars based on the statistics and analysis of key technology patents such as automatic braking, cruise control, and lane keeping [16]. Suzuki et al. used the International Patent Classification (IPC) number cooccurrence method to study technical topic identification and trend prediction [17]. Jeong et al. studied the strength of IPC cooccurrence relationship by Jaccard coefficients and analyzed the main types of technology topics [18]. Lee et al. performed link prediction analysis on the cooccurrence network of IPC to predict possible future emerging technology topics and used topic analysis to extract keywords to identify possible future emerging areas [19]. Huang et al. used association rule analysis method to analyze the IPC cooccurrence in two technology areas, namely, information technology and biotechnology, and analyzed the characteristics of technology topics in terms of support, confidence, and lift [20]. By statistically analyzing the number of applications and the number of IPCs for driverless car patents, they concluded that the technological research in this field has entered a stable state to a certain extent [21].

2.1.2. Analysis Method Based on Patent Citation. The patent citation method analyzes the citation relationships among patent literature and between patent literature and scientific literature and analyzes the track of technical topic evolution by constructing a citation relationship network [22]. Kwon et al. comprehensively analyzed technical topics by constructing a patent citation coupling network and a cocitation network to synthesize the distribution of patents [23]. Choi and Park constructed a patent citation network and used the main path analysis algorithm to identify technical topics [24]. Hsueh and Wang combined patent citation time with patent citation relationship to predict the frontier technology in the field of LCD (Liquid Crystal Display) technology [25]. Geum et al. [26] and Zhai et al. [27] used the knowledge flow of citation networks between technology categories to study technology convergence. Kim and Seol fused the similarity of patented technologies and patent cocitation features to achieve a frontier technology identification method based on patent network analysis by constructing a patent cooperation network [28]. Small et al. combined patent direct citation network and cocitation network to identify novel technology topics by community clustering algorithm [29]. Li and Chen proposed an emerging technology identification model based on patent citation coupled clustering and empirically analyzed the field of nanotechnology [30]. Lee et al. collected Google's patents on self-driving cars and used social network analysis methods to identify the company's core patents in this field based on the citation relationships between patents and predicted that Google's R&D capabilities would be focused on hardware control [31].

2.1.3. Analysis Methods Based on the Content of Patent Texts. Patent text mining can discover potential data patterns and internal relations from a large amount of unstructured textual information and is an important method for technical topic evolution analysis. In 2003, Beli [32] first proposed the LDA (Latent Dirichlet Allocation) topic model, which introduces the Dirichlet prior distribution, based on statistical probability level to express the semantic relationship between words and mine document topics. Qin and Le used LDA model to conduct a study on the variation of oncology domain in terms of topic content and intensity by formulating topic association filtering rules [33]. Yan used LDA model to mine the document dataset in library intelligence domain and identified the technical topics in the domain [34]. Yang and Yang used LDA topic model for topic mining of policy text data in the international climate domain [35]. Liao and Le introduced IPC classification numbers to measure technical topic strength based on LDA modeling and realized a study on three aspects of topic strength, topic content, and technical topic strength [36]. Wang et al. identified technical topics in the field of ocean acidification using the LDA model [37]. Fan et al. proposed four characteristic indicators of novelty, innovation, interdisciplinarity, and high interest based on the LDA topic identification results and based on the characteristics of literature in the medical field to identify frontier topics in the medical field [38]. Yi et al. identified technology topic identification and its structural features in the field of graphene based on LDA model and strategic coordinate map [39]. Shen et al. constructed an IPC cooccurrence network by Apriori algorithm, applied Louvain clustering algorithm to divide the network into different technical communities, and discovered the technical topics of each community based on LDA topic model [40]. Ma et al. identified common technologies in the field of new materials based on patent data, using LDA topic model to extract the implied technology topics in the text [41]. Mi et al. dynamically identified technology topics in 3D printing field based on the use of LDA topic identification model combined with time series analysis [42]. Li and Xuan combined LDA topic model with patent value evaluation index and proposed a quantitative method for mining technological innovation topics [43]. Tang and Qiu used LDA model to obtain topic words from multisource heterogeneous texts and carried out fusion analysis to extract emerging technology topics in the field of intelligent network vehicles [44].

2.2. New Energy Vehicle. In recent years, the number of new energy vehicle patents and research literature has grown rapidly [45, 46]. In the patent database of CNKI (China National Knowledge Infrastructure), the patent literature of new energy vehicles can be traced back to as early as 2002. The number of new energy vehicle patents exceeded 100 for the first time in 2013, and the number of new energy vehicle patents has been growing rapidly since then.

Song and Zhu identified the technical topics in the field of artificial intelligence. The research shows that new energy vehicles are the frontier technical topic in the field of artificial intelligence, with a significant upward trend in the next three years [47]. Yang et al. extracted data from the EPO global patent statistics database and classified the technical topics of new energy vehicles by IPC classification number [48]. Xie et al. obtained 1393 pieces of patent data from 2003 to 2012 from the China Intellectual Property Rights Network and analyzed the development trend of new energy vehicles from five aspects: the number of patent applications, the distribution of patent owners, the technology life cycle, the IPC analysis, and the geographical distribution of patent R&D subjects [49]. Liang et al. analyzed the development trend of new energy vehicles by using technology activity analysis and patent technology cooccurrence to analyze the technology research frontier and development trend of new energy vehicles. The new materials, safe electricity, structural parts, and related electric traction are the popularities of R&D in the future period [50]. Wang et al. retrieved 19,610 pieces of patent data from the database of the State Intellectual Property Office and conducted statistics on the annual publication trend of patents, the portfolio structure of patent application subjects, and the distribution of IPC classification numbers, and the statistical results showed that the top three technical fields in patent number are battery, electrical machine, and electric control system [51]. Shen et al. constructed the IPC cooccurrence network of new energy vehicles by Apriori algorithm [40]. Li and Fan conducted a comparative analysis of key technology areas of international and Shanghai Automotive Group Corporation from three dimensions: patent distribution map of major technology areas, topography map of key technology R&D areas, and technology hotspot change map, and extracted the technology areas in which Shanghai Automotive Group

Corporation has comparative advantages [52]. Shi et al. analyzed the overall development of Chinese new energy vehicle technology in the past 30 years from three aspects: number of patents, innovation subjects, and technology categories [53]. Sun et al. took the top 38 automotive manufacturers with the most NEV-related patents published in China as the target of their study and generated a patent cooperation map through a patent citation matrix to further analyze the cooperation behavior among different patent owners [54]. Feng et al. found that the topics of the four most promising ones are mainly about battery arrangement and protection, control systems, framework design, and charging connectors [55]. Miao et al. predicted the sales volume of new energy vehicles in China based on ARIMA model [56]. Wu et al. conducted an exploratory study on the ecosystem construction process of new energy vehicles, taking the NIO new energy vehicle as an example [57]. Fang and Zhang used the global patents of charging pile technology from 1990 to 2018 in the Smart Bud patent database to predict the new energy vehicle charging pile technology based on the life cycle and international patent classification numbers. The results of the study indicate that the charging pile technology in China will enter a mature period in 2020 and a recession period in 2025 [58]. Based on machine learning and empirical mode decomposition method, a technical topic was identified for Chinese new energy vehicles, and the results showed that China has a relatively high technological advantage in the field of battery cooling or maintaining low temperature technology and vehicle battery application technology [59].

From the above discussion, it can be found that the existing literature is mainly based on the analysis method of patent classification and patent citations to study the technological development of new energy vehicles, and although the analysis method of technical topic evolution based on patent classification and patent citations can discover the technological development trend from a macro perspective, it cannot show the specific evolution details of technical topics, and the technical topic identification based on patent texts can make up for this deficiency. Therefore, this paper downloads the patent texts of new energy vehicles from the patent database of CNKI and introduces the LDA topic model into the identification of technical topics in the field of new energy vehicles, which can solve the problems of "time lag" and "multiple meanings of words" of traditional research methods and accurately accurately discover the topics hidden in massive texts. On this basis, this paper predicts the popularity of China's new energy vehicle technology topics from 2021 to 2025 by constructing ARIMA model.

3. Methodology

3.1. LDA Model-Based Technical Topic Identification. The LDA (Latent Dirichlet Allocation) topic model is a typical probability generation model [32] (see Figure 1), a text mining technology that can train large-scale document sets to identify potential topic information. It can realize unsupervised, without setting any topic vocabulary in advance and professional domain knowledge can complete topic crawling. M is the number of documents in the corpus, N is the number of words in the documents, K is the number of topics implied in all documents, θ denotes an $M * K$ document topic distribution matrix, $\vec{\theta}_m$ denotes the topic distribution of the m-th document, ϕ denotes a $K * N$ topic word distribution matrix, $\vec{\phi}_k$ denotes the distribution of topic words numbered k, α is the prior distribution of topics for each document, β is the prior distribution of words for each topic, W represents the observable words, and Z is the potential topic distribution of each observed word.

In this paper, the technical topic matter implied in the patent text is set to obey the distribution as follows:

$$\mathrm{Dir}\left(\theta_d|\alpha\right) = \frac{\Gamma\left(\sum_{k=1}^{K}\alpha_k\right)}{\prod_{k=1}^{K}\Gamma\left(\alpha_k\right)}\prod_{k=1}^{K}\theta_{dk}^{\alpha_k-1}. \quad (1)$$

θ_{dk} denotes the distribution of patent text d in technical topic k. For each technical topic k, the term distribution $\varphi_k \sim \mathrm{Dir}(\beta)$ is generated. For each patent text d, the subject term distribution $\varphi_k \sim \mathrm{Dir}(\beta)$ is generated, and, for the n-th term in each patent text, the subject term $Z_{dn} \sim \mathrm{Multinomial}(\theta_d)$ and the term $w_{dn} \sim \mathrm{Multinomial}(\phi_{z_{dn}})$ are generated.

From the perspective of LDA, the generation process of a document can be decomposed into the following steps: First, the topic distribution $\vec{\theta}_m$ of document d is generated by sampling from the Dirichlet distribution α, where $m \in [1, M]$. Second, the polynomial distribution of topics $\vec{\theta}_m$ is sampled to generate topic $Z_{d,n}$ for the n-th word in document d. Third, the word distribution $\vec{\varphi}_{z\,d,n}$ of topic $Z_{d,n}$ is generated by sampling from the Dirichlet distribution β. Fourth, word $W_{d,n}$ is generated by sampling from the polynomial distribution of words $\vec{\varphi}_{z\,d,n}$.

The above generation process can be represented by the joint distribution of all visible and hidden variables.

$$p\left(\vec{w}_d, \vec{z}_d, \vec{\theta}_d, \phi|\vec{\alpha}, \vec{\beta}\right) = \prod_{n=1}^{N_d} p\left(w_{d,n}|\vec{\phi}_{zd,n}\right)p\left(z_{d,n}|\vec{\theta}_d\right)p\left(\vec{\theta}_d|\vec{\alpha}\right)p(\phi|\vec{\beta}). \quad (2)$$

The LDA topic model uses the Dirichlet distribution as the prior distribution for topic information mining. Compared with other generative probability models, LDA topic model has obvious advantages, which can describe document production mode more accurately, and the analysis results are better than other mixed topic models. The LDA topic model has the feature of discovering potential topics and has obvious advantages in topic research in emerging technology fields.

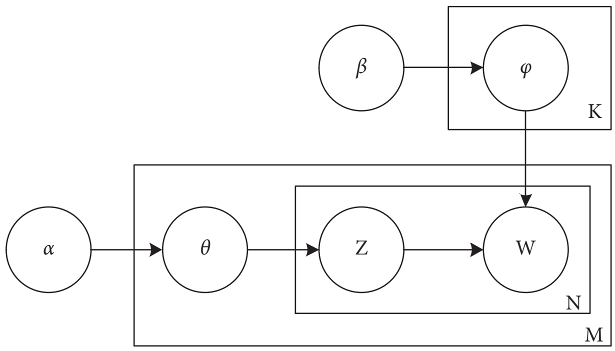


Figure 1: LDA topic model.

The main steps of technology topic mining using LDA model are as follows.

3.1.1. Text Segmentation. Patent text is very different from unstructured information such as news articles, novel biographies, and letters and emails. It records innovative contributions in a specific technical field and involves many specialized vocabulary and technical terms. Therefore, the Viterbi algorithm can be used to split the text of patent data, generate a directed acyclic graph based on the lexicon, and find the shortest path to accurately intercept the patent text according to the exact pattern of jieba splitting. In addition, although the patent abstract has a very standardized structure, including two main parts, introduction and subject introduction, there are still pronouns, articles, conjunctions, and punctuation in the text, which are not very useful in the text subject recognition and must be eliminated.

3.1.2. TF-IDF Weight. Since the new energy vehicle patent text contains specialized words from different disciplines and different words have different effects among themselves in expressing the text topic, this paper invokes the TF-IDF (Term Frequency-Inverse Document Frequency) method to assign weights to words. The main idea of TF-IDF is as follows: two indicators are needed for a word to maximize the representation of the text topic, one is whether the word appears many times in some texts and the other is that the word rarely appears in all texts [60]. Therefore, the value of TF-IDF is equal to the "Term Frequency" (TF) multiplied by the "Inverse Document Frequency" (IDF), which is calculated as follows:

$$\mathrm{TF}_{i,j} = \frac{m_{i,j}}{\sum_k m_{k,j}}. \quad (3)$$

$m_{i,j}$ denotes the number of occurrences of word t_i in text d_j, and $\sum_k m_{k,j}$ denotes the total number of occurrences of all words in text d_j.

$$\mathrm{IDF}_i = \log \frac{|M|}{\left|\left\{j : t_i \in d_j\right\}\right|}. \quad (4)$$

$|M|$ denotes the total number of documents in the corpus, and $|\{j : t_i \in d_j\}|$ denotes the number of documents containing word t_i. When the TF and IDF values of a word are calculated, they are multiplied together to obtain the TF-IDF value of the word. The higher the TF-IDF value of a word is, the more important the word is in the text and the more likely it is a keyword of the text.

$$\mathrm{TF} - \mathrm{IDF}_{i,j} = \mathrm{TF}_{i,j} \times \mathrm{IDF}_i. \quad (5)$$

3.1.3. Model Parameter Setting. The main influence on the results of the LDA topic model is the preprocessing step, which requires more careful processing. For parameter setting, since the LDA topic model is an unsupervised machine learning model, for hyperparameter $\underset{\rightarrow}{\alpha}$ of the document topic prior Dirichlet distribution $\vec{\theta}_m$ and hyperparameter β of the topic word prior Dirichlet distribution $\vec{\varphi}_k$, we generally use the scikit-learn library's default value 1/K, which is because, during the continuous iteration of Gibbs Sampling, these two parameters will keep approximating the true value of the text training set under the training of the LDA topic model. Therefore, the parameter setting of the LDA topic model is mainly focused on the discussion of the number of topics k and the maximum number of iterations.

The selection of the number of topics k is directly related to the training effect of the LDA topic model. There are three common methods for setting the number of topics k: ① Blei et al. used the size of confusion to evaluate the goodness of the model. A smaller value of perplexity means better training result of the training set. ② The Hierarchical Dirichlet Processes (HDP) method based on the Dirichlet process proposed by Teh et al. assumes that the document sets share the same topics before training and the number of topics is unlimited, but the exact number of topics will be determined during the derivation of the Dirichlet parameters [61]. ③ Griffiths et al. proposed the application of a Bayesian model approach to determine the optimal number of topics. Through literature review, it is found that the existing literature mainly uses method ① to determine the number of topics k by the method of perplexity [62], and both method ② and method ③ have the disadvantage of high computational complexity, and this paper uses method ① to calculate the number of topics k. Method ① of determining the number of iterations of the LDA topic model is mainly based on whether the iteration results converge to determine the optimal number of iterations.

3.1.4. Technical Topic Identification and Trend Analysis. Through the above steps, the LDA topic model can be trained, and the corresponding training results can be derived to identify each technical topic, respectively, and the development trend of each technical topic can be calculated by using the document topic distribution matrix.

3.2. Trend Forecasting of Technical Topics Based on ARIMA Model. ARIMA models can be divided into seasonal ARIMA models and nonseasonal ARIMA models. Because the technical topic of this paper is not affected by seasonality and produces obvious periodic changes, the nonseasonal ARIMA model is used in the modeling process. The specific process of model construction is shown in Figure 2 [63].

(1) Sequence stabilization. Usually, the data predicted by ARIMA model is an unsteady temporal dataset, so the first process of construction is to preprocess by observing the changing trend of the original data. Generally, there are three ways to deal with it: difference, seasonal difference, and natural logarithm transformation. In this paper, differential preprocessing is mainly used. Generally, only one or two difference calculations are needed to get the average value of the data close to zero. In addition, in order to get a smooth series, it is also necessary to test the value of the autocorrelation function and the value of the partial correlation function of the data to see whether the value of the function is not significantly different from zero.

(2) Model identification. The parameters of the ARIMA correlation model are determined as shown in Table 1. If the partial autocorrelation function (PACF) of the smooth series is truncated and the autocorrelation function ACF is trailing, it is determined that the series is suitable for constructing the AR(p) model, and the value of p is the order of the truncated partial autocorrelation function; if the autocorrelation function (ACF) of the smooth series is truncated and the partial autocorrelation function (PACF) is trailing, it can be determined that the series is suitable for the MA(q) model, and the value of q is the order of the truncated autocorrelation function; if the autocorrelation function and the partial autocorrelation function of the smooth series are both trailing, it can be decided that the series is suitable for the construction of the ARMA(p, q) model, and the values of p and q need to be calculated by the AIC criterion, SC criterion, and BIC criterion for all possible values, from which the combination of the lowest values of AIC, SC, and BIC is selected as the optimal model parameters.

(3) Model testing. The main test is whether the model is constructed reasonably enough, and this step requires an autocorrelation test on the residual series of the model with the optimal values determined in the previous step to determine whether it is a white noise series [64].

(4) Model prediction. The fitting effect of the ARIMA model is further analyzed by calculating the relative error between the fitted and actual values; in the prediction stage, the first case after the evaluation is set to the specified date, and the future trend is forecasted.

The relative error is calculated by the following formula:

$$\delta = \frac{\triangle}{L} \times 100\%. \tag{6}$$

Thereinto, δ denotes the relative error between the fitted value and the true value, $\triangle$ is the absolute error, which takes a value equal to the difference of the predicted value minus the true value, and L is the true value.

4. Empirical Research

4.1. Data Acquisition. In this paper, the patent database of CNKI was used to search the patent data related to new energy vehicles, the search time was February 1, 2020, the search subject was new energy vehicle, and a total of 18,583 patents related to new energy vehicles were obtained. The latest patent is "a remote monitoring recorder for new energy vehicles," and the patent publication date is November 20, 2020. The latest patent is "a new energy vehicle remote monitoring recorder," with a patent publication date of November 20, 2020, and the application time in the patent document collection spans from October 2000 to December 2020. The results of the patent search mainly focused on topics such as new energy vehicle battery, charging pile, new energy electric vehicle, charging device, power battery, battery box, and preparation method. The information obtained includes Patent Author, Applicant, Title, Country Name, Publication Number (PubNo), Publication Date (PubTime), Summary, and Claims. After removing the data of patents with incomplete information, the total number of remaining patent texts is 17, 816, and the specific number of publications is shown in Figures 3 and 4.

As can be seen from Figure 3, the development of Chinese new energy vehicle patents can be divided into three stages: 2002–2010, the number of new energy vehicle patents issued was small; 2011–2015, the number of new energy vehicle patents issued increased from 44 to 237, and the annual growth rate was 52.3%; 2016–2020, under the combined effect of policy support and market promotion, the number of new energy vehicle patents issued showed a rapid growth trend from 780 to 5888, and the annual growth rate was 65.8%.

4.2. New Energy Vehicle Technology Topic Identification and Evolution Analysis Based on LDA Topic Model

4.2.1. Text Preprocessing. The process of text preprocessing has three main parts: reading data, segmenting sentences, and saving segmentation results. Before the code is run, a proprietary data file named 1.csv, a deactivated words table named stopwords.txt, and a custom dictionary named userdict.txt have been prepared. The first step is to read the data from a file named 1.csv with the read csv( ) method of pandas. The second step is to load the prerecorded custom dictionary, segment the sentences using the jieba( ) method, save them to the sentence seged as an array data type, and delete the words contained in the deactivated word list by traversing each segmented word in the sentence seged. Execute the third step, save the result of the previous steps into a file named fenci.csv, and the effect of document word segmentation is shown in Figure 5.

4.2.2. Calculation of TF-IDF Assignment Weights and Optimal Number of Topics. Firstly, the text feature extraction function CountVectorizer( ) is applied for TF-IDF weight assignment. Secondly, in the LDA topic model parameter

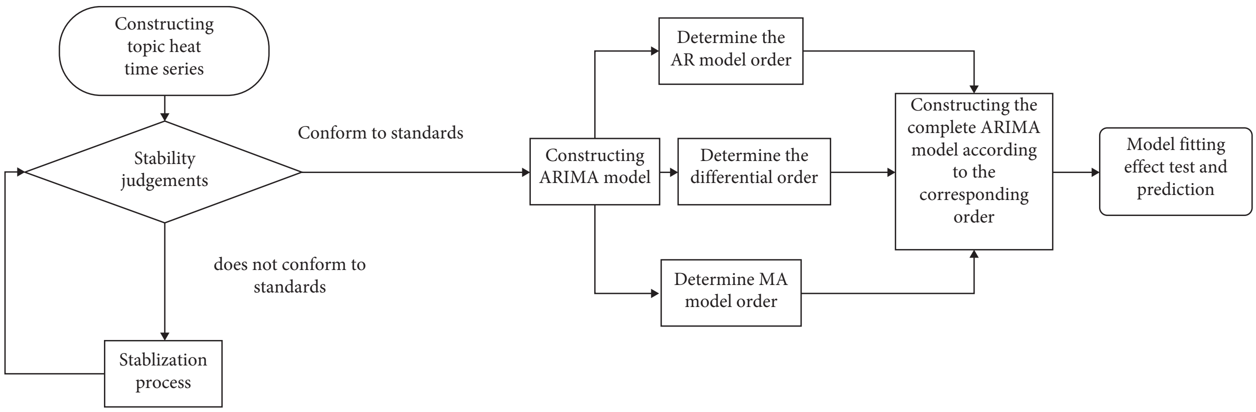


FIGURE 2: ARIMA modeling process.

TABLE 1: Determination of model parameters p and q.

Model	Autocorrelation function (ACF)	Partial autocorrelation function (PACF)
AR(p)	Trailing	p-step posttruncation tail
MA(q)	q-step posttruncation tail	Trailing
ARMA(p, q)	q-step rear trailing	p-step rear trailing

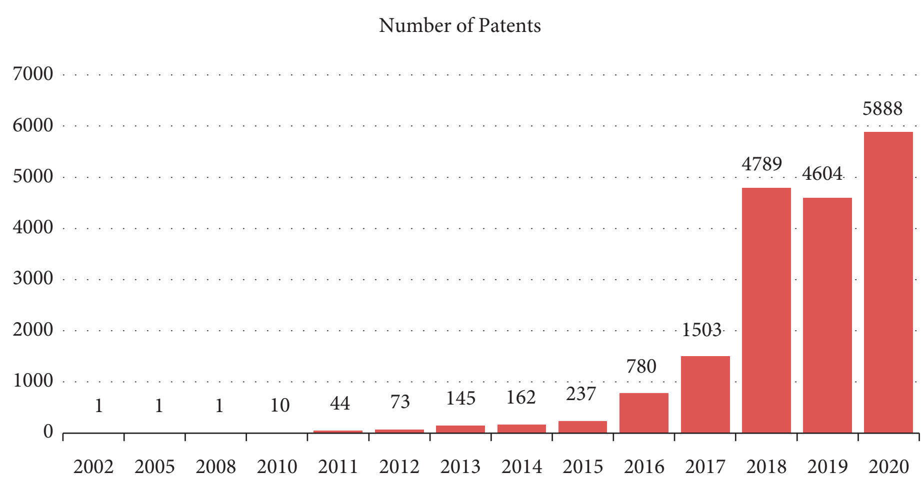


FIGURE 3: Number of patents issued for new energy vehicles over the years.

setting module, the number of iterations is set to 100 times, and the corresponding perplexity values from 2 to 50 topic counts are calculated sequentially. As shown in Figure 6, the curve changes sharply and then slowly. When the document set is divided into 28 topics, the minimum confusion value is 319.18.

4.2.3. Unsupervised Training of LDA Topic Models. The LDA topic model was trained, and the number of iterations was continuously increased until the topics of each word converged to a stable state, and the top 10 technical terms in terms of probability under each topic were derived, and the three topics with little association with the main technology of new energy vehicles were excluded, and the final 25 technical topic contents were determined as shown in Table 2.

Through the division of new energy vehicle technology topics, it can be divided into 25 technology areas, and the topics are named as shown in Table 3.

4.2.4. Technical Topic Evolution Analysis. A 17817 ∗ 25 document topic distribution matrix can be obtained after training by the LDA topic model. The sum of the values of each row of this matrix is 1, and the values in each column are different. The greater the sum of the values in each column, the greater the importance of the topic to the corpus. The top 10 technical topics in the corpus of new energy vehicles in terms of importance are Topic 8: Installation and Fixation, Topic 5: Heat Dissipation and Cooling, Topic 14: Vehicle Data Monitoring, Topic 9: Charging Pile, Topic 15: Vibration Damping, Topic 6: Electric Drive System, Topic 3: Cable Insulation Materials,

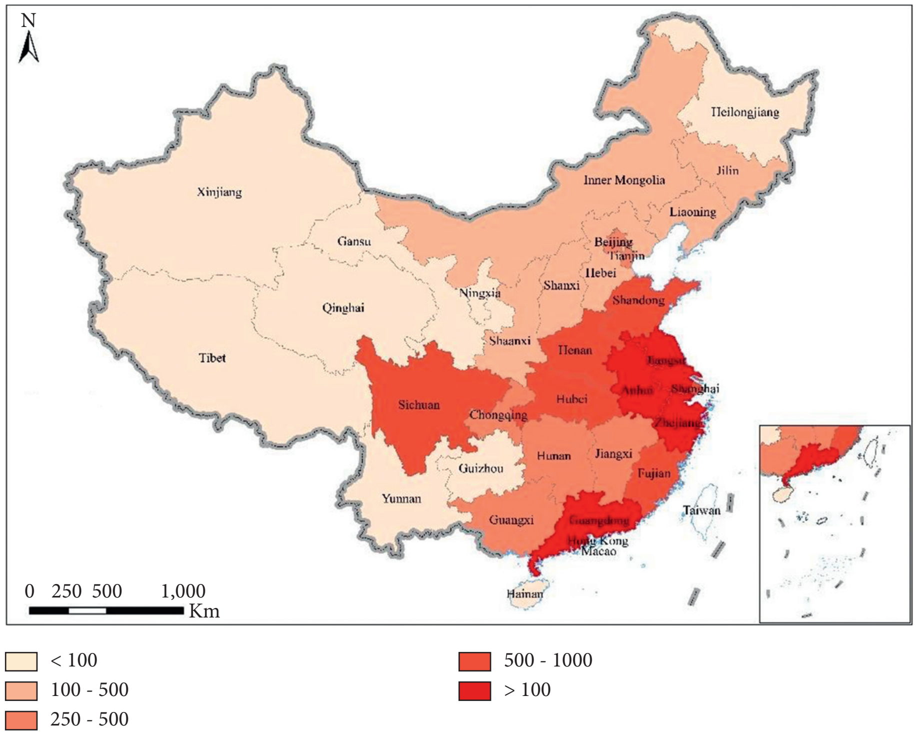


Figure 4: Distribution of new energy vehicle patents by provinces in China.

Topic 13: Production and Processing, Topic 23: Circuit Control, and Topic 21: Wireless Charging.

The time dimension data of patents is introduced to portray the changes of each technology topic with time by taking the year as the unit; see Figure 7. Since 2014, China's new energy vehicles have achieved fruitful results in patent technology and the continuous improvement of the overall technology level. The achievements of Topic 5: Heat Dissipation and Cooling, Topic 8: Mounting Fixation, Topic 6: Electric Drive System, Topic 9: Charging Pile, and Topic 15: Vibration Damping stand out. These technologies are closely related to new energy vehicles and directly determine the performance level of new energy vehicles. This is because the first element that car owners pay attention to when purchasing a new energy vehicle is the driving efficiency of the vehicle, such as maximum mileage, 100 km consumption, power, and other factors. The development of this type of technology can be a great attraction to consumers and can greatly accelerate the speed of the market promotion of new energy vehicles. Topic 7: Chassis, Topic 10: Water Pump, Topic 18: Electric Motor, Topic 24: Module, and other technologies are developing relatively slowly, and new innovations and breakthroughs are urgently needed. Topic 12: Vacuum Pump and Topic 21: Wireless Charging are relatively less important in 2019–2020.

Figure 8 is plotted based on Figure 7, and the value of each box in the figure is equal to the importance of the corresponding topic in that year divided by the overall importance of the technical topic. Through observation, it can be found that the importance of Topic 3: Cable Insulation, Topic 14: Vehicle Data Monitoring, and Topic 23: Circuit Control declined from 2014 to 2020, while Topic 5: Heat Dissipation and Cooling and Topic 8: Mounting and Fixing continued to increase in importance. With the continuous improvement of the drive system of new energy vehicles, this part must require the electric motor to provide higher power with more stable output. Although the electric motor is a low heating component, the heat generated during normal operation is low, but the motor will lead to high coolant temperature after long time and high-power operation. This reflects that, in recent years, with the maturity of new energy vehicle technology, the technical level has been able to meet the needs of consumers, and the most important concern of automobile manufacturers is to develop an efficient process to realize the mass production of new energy vehicles, so the importance of installation and fixing technology is increasing.

4.3. New Energy Vehicle Technology Topic Popularity Screening Based on LDA Model. In the training results of the LDA topic model, the distribution of topics corresponding to each document can be obtained, and the changing trend of technical topic popularity over time is calculated according to the date of publication corresponding to each document. Using each month's record as a cycle, the popularity of each period is connected with a dash line, and the overall trend of changes in 25 technical topics is depicted by a linear regression model, and some of the topics are shown in Figure 9. The overall slope of each technical topic is shown in Table 4, and the top 5 technical topics in terms of the slope of the linear trend were filtered out based on the slope of the linear trend: Topic 20: Safety, Topic 8: Installation Fixing, Topic 3: Cable Insulation, Topic 15: Vibration Reduction, and Topic 10: Pump Cooling.

lithium battery silicone thermal conductivity structure base plate base plate on mounting hole mounting hole inside water tank setting new energy vehicle new energy vehicle charging automatic separation installation plug connection permanent magnet charging pile new energy vehicle more say new energy vehicle battery installation utensil main frame installation control institution leverage installation rectangular new energy vehicle high speed motor bearing rotor dynamic feature analysis method access bearing rotor structure parameter working condition parameter initial value new energy vehicle charging pile installation charging pile charging pile bottom set up inspection mouth inspection mouth inside edge fixed connection locking structure new energy vehicle battery fast replace system new energy vehicle new design solution solve solution restriction new energy vehicle charging time new energy vehicle battery pack magnesium alloy pallet processing sand filling bending welding pallet machine processing neutralization surface anti-corrosive new energy vehicle battery new energy vehicle front end exterior magnet absorption curing type installation hand loading support base permanent magnet iron front end front end skeleton front end exterior trim panels new energy vehicle high intensity bumper vehicle vehicle fixed mounting mounting plate mounting plate bumper bumper mounting slot mounting plate vehicle plastic high pressure connector new energy vehicle on wide range of application connector new energy vehicle domestic capacity constantly increase plastic high pressure vehicle electronic devices simple installation new energy vehicle loudspeaker vocalization device fixing device dustproof device vocalization device basin rack bouncing waves new energy vehicle general PTC heater above number PTC heating strip concave profile strip concave profile article mounting slot PTC heating mobile new energy vehicle charging simulation display device base base setting trapezoid slot trapezoid slot installation slider slider installation percentage meter slider new energy vehicle new energy vehicle motor failure testing installation support part 100 support part part up rotate spindle 200 spindle turntable charging

Figure 5: The effect of document word segmentation.

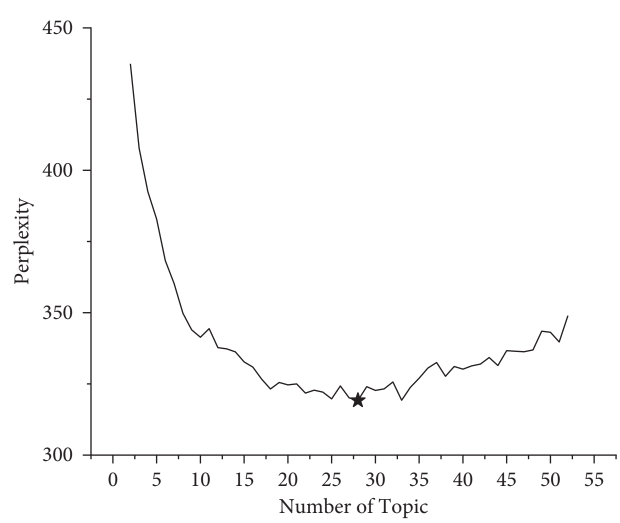


Figure 6: Perplexity of LDA topic models corresponding to different topic number.

Table 2: Topic keywords.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12
Power	Bolster	Cable	Shell	Heat Dissipation	Electrical Machinery	Activity	Installation	Charging Pile	Fixed Block	Base	Battery Pack
High Voltage	Support Plate	Material	Electric Vehicle	Cooling	Drive	Joint Lever	Fixed	Set	Water Tank	Adjust	Terminal
Assembly	Positioning	Insulation	Sealing	Setting	Gear	Swerve	Structure	Base	Mounting Holes	Testing	Wiring
Power Battery	Car body	Wires	Waterproof	Temperature	Spindle	Clamping	Weld	Interior	Water Pump	Battery Pack Lifting and Lowering	Dispose
Engine	Structure	Conductor	Plug	Interior	Transmission	Base	Fixed Plate	Installation	Connection		Collecting
Low Voltage	Beam	Insulation Layer	Connector	Heating	Bearing	Wheels	Setting	Defend	Compress	Clamp	Absorb
Control	Cylinder	Performance	Socket	Effect	Turn	Switch	Mounting Plate	Fixing	Battery Case	Platform	Fixed Ring
Control module	Guidance	Jacket	Hump	Lithium Battery	Rotation	Tire	Bolt	Bottom	Setting	Height	Vacuum Pump
The Whole Car	Turn Plate	Preparation	Protection	Heat	Output Shaft	Workbench	Convenience	Protection	Pressure	Setting	Vacuum
Relay	Frame	Polish	Seal Ring	Air	Setting	Fixture Block	Junction Plate	Display Screen	Connecting Pipe	Seat	Mainframe

Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18	Topic 19	Topic 20	Topic 21	Topic 22	Topic 23	Topic 24	Topic 25
Equipment	Check	Tank	Product	Charging Gun	Power Source	Storage Battery	Organization	Charging	Brake	Module	Element	Battery
Baseboard	Vehicles	Brackets	Design	Charging	Rotor	Setting	Shell	Solar Energy	Vehicles	Controller	Lead	Air Conditioner
Setting	Method	Damping	Appearance Design	Charging Cable	Stator	Car body	Cover	Vehicles	Hub	Circuit	Module	Slip
Process	Information	Setting	Element	Setting	Dc	Cover	Airbag	Equipment	Recycle	Control	End Caps	Compressor
Locking	Data	Battery Box	Mold	Conductivity	Electric Motors	Telescopic Rods	Lower Shell	Cleaning	Energy	Signal	Pipeline	Replacement
Roof	Control	Fixation	Installing	Jack	Direct Current	Electromotion	Interior	Generate Electricity	Electron	Voltage	Engine Housings	Joints
Installation	Monitor	Buffer	Molding	Resistance	Capacitance	Wire Harness	Protection	End Face	Hydraulic Pressure	Input	Setting	Stamping
Side Plate	Sensor	Spring	Shape	Wires	Switch	Annulus	Prevent	Equipment	Slide	Output	Accommodate	Install
Automation	Operation	Interior	Electrical Core	Charging Plug	Docking	Support Frame	Rim	Cleaning	Pedal	Current	Air Pipe	Store
Production	Malfunction	Install	Circuit Board	Avoid	Iron Core	Mounting Bracket	Dustproof	Power Supply	Electronic Control	Control System	Opening	Carriage

Table 3: New energy vehicle technology topics.

Number	Topic	Number	Topic
1	Vehicle System	14	Vehicle Data Monitoring
2	Body Structure	15	Vibration Damping
3	Cable Insulation Materials	16	Product Design
4	Waterproof	17	Charging Gun
5	Heat Radiating and Cooling	18	Electric Motors
6	Electric Drive System	19	Storage Battery
7	Chassis	20	Safety
8	Installation and Fixation	21	Wireless Charging
9	Charging Pile	22	Energy Recovery
10	Cooling Pump	23	Circuit Control
11	Battery Pack	24	Module
12	Vacuum Pump	25	Air Conditioning System
13	Production and Processing		

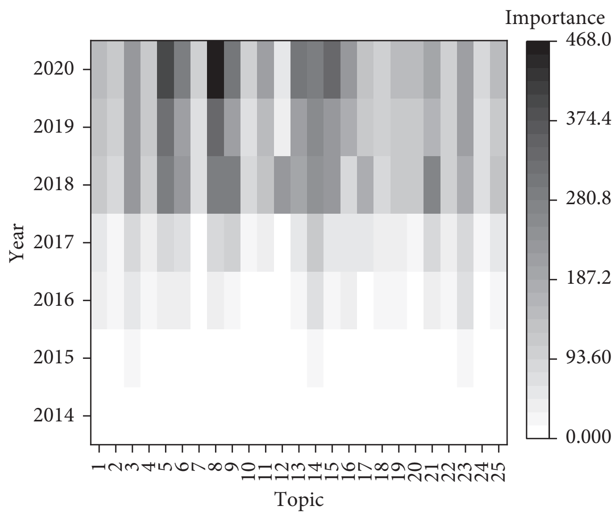


Figure 7: Map of the trend of importance of each technology topic.

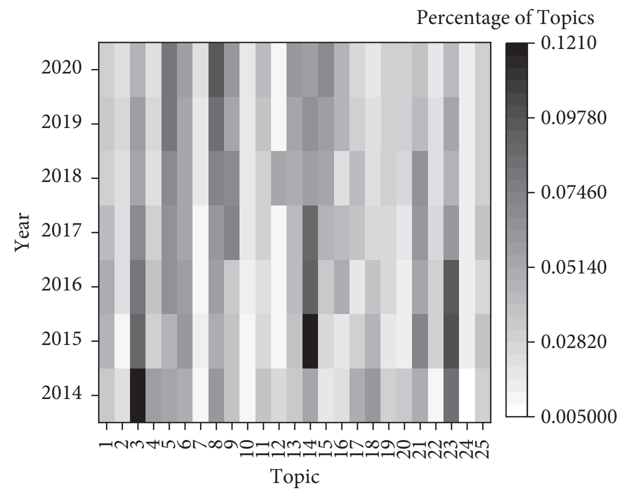


Figure 8: Trend of evolutionary weight of each technology topic, 2014–2020.

4.4. Trend Prediction of Technology Topics Popularity Based on ARIMA Model

4.4.1. Time Series Smoothing Processing. As shown in Figure 10, the popularity of Topic 20 reached its peak in September 2018 and fell to the local lowest point in September 2019, with the overall popularity change trend being constantly rising. Since the ARIMA model requires the time series to be a smooth series, it still needs to perform the difference operation on the original data. The difference operation formula is as follows:

$$\nabla x_t = x_t - x_{t-1}. \quad (7)$$

The timing diagram of sequence $\{\nabla x_t\}$ after the 1st-order differencing process is shown in Figure 11.

The long-term trend of the original series is obtained by 1st-order differencing, and it can be seen from the graph that the values of the series after 1st-order differencing fluctuate mainly around the upper and lower sides of the 0 scale. In addition, the stability is judged by the ADF (Augmented Dickey-Fuller) test, as shown in Table 5. The ADF test value of the original series is −0.125792, whose absolute value is less than the critical absolute value of −1.945389 at 5% significance level, and its corresponding p value is 0.6370, which is greater than 0.05. It shows that the original hypothesis is accepted at 5% significance level and the initial time series has unit root; that is, the initial time series is a nonstationary series. The ADF test value of the time series with 1st-order difference is −10.66079, whose absolute value is greater than the absolute value of −1.945456 at the 5% significance level, and its corresponding $p < 0.05$, indicating that the original hypothesis is rejected at the 5% significance level and the 1st-order difference series does not have a unit root, so the 1st-order difference series is a smooth series.

4.4.2. Model Identification. From Figure 12, it can be found that the autocorrelation diagram after the 1st-order difference shows that the autocorrelation value fluctuates within the basic boundary value range; only the 1st-order and 5th-order autocorrelation values slightly exceed the boundary value, while the autocorrelation value fluctuates within the significant boundary in all other orders. In the partial autocorrelation graph, the 1st-order partial autocorrelation value is on the significant boundary value, while the 5th-order partial

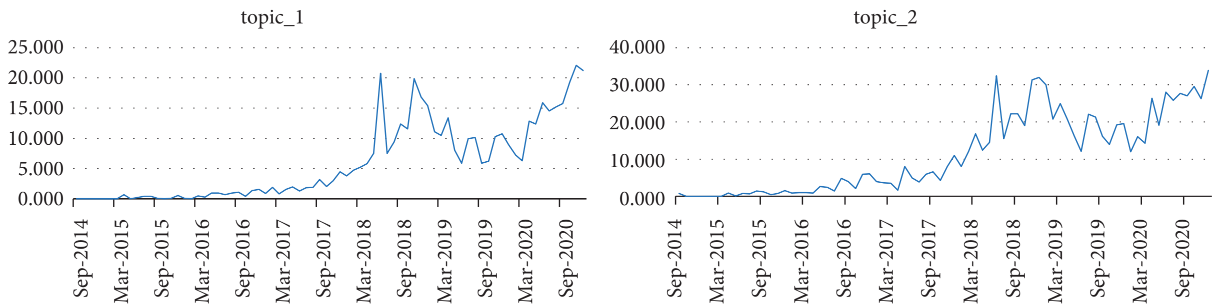


Figure 9: Time series chart of topic popularity in the field of new energy vehicle technology.

Table 4: Overall slope of each technology topic for new energy vehicles.

Topic category	Overall slope	Topic category	Overall slope	Topic category	Overall slope
Topic 1	0.23	Topic 10	0.42	Topic 19	0.26
Topic 2	0.42	Topic 11	0.08	Topic 20	0.74
Topic 3	0.50	Topic 12	0.09	Topic 21	0.15
Topic 4	0.25	Topic 13	0.30	Topic 22	0.18
Topic 5	0.20	Topic 14	0.35	Topic 23	0.20
Topic 6	0.29	Topic 15	0.46	Topic 24	0.38
Topic 7	0.15	Topic 16	0.29	Topic 25	0.27
Topic 8	0.56	Topic 17	0.24		
Topic 9	0.24	Topic 18	0.31		

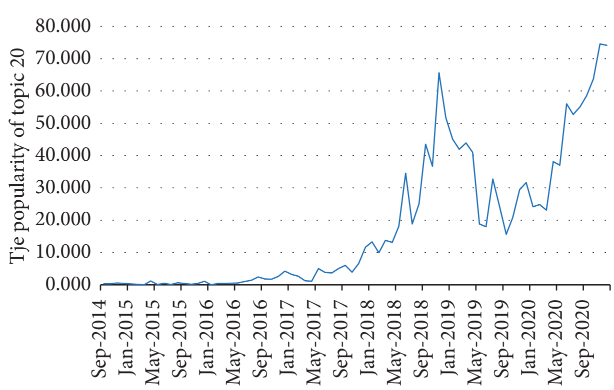


Figure 10: Timing diagram of popularity of Topic 20.

autocorrelation value is slightly beyond the significant boundary, in addition to the other basic within the boundary value; therefore, the autocorrelation and partial autocorrelation can be initially considered as 1st-order or 5th-order trailing. In addition, since SPSS does not have the function of automatic identification of model parameters, it is necessary to calculate all possible parameter combinations, and the BIC indicator is selected as the evaluation indicator among many method indicators, and the optimal parameter combination is determined by the BIC indicator; that is, the parameter combination corresponding to the smallest BIC value is selected as the optimal ARIMA model parameters [64], and the calculation results are shown in Table 6, and the optimal model is ARIMA (0, 1, 1).

4.4.3. Model Testing. After determining the model parameters, the model was then fitted and tested by white noise test. $R^2 = 0.893$, the fitting effect was good, and the p value of the Ljung-Box test was greater than 0.05, so the original hypothesis could not be rejected, and then the series could be considered as white noise. As can be seen from Table 7, the coefficient of MA is 0.282 and the significance level meets the test requirement ($p < 0.05$), so the coefficient is significantly not 0. Also, the autocorrelation and partial autocorrelation of the observed residuals are smooth states (see Figure 13), so it is reasonable to set the ARIMA (0, 1, 1) model.

4.4.4. Trend Prediction. Finally, the model is applied to predict the development trend of China's new energy vehicles from 2021 to 2025 by plotting the time series prediction results of Topic 20: Safety, Topic 8: Installation Fixation, Topic 3: Cable Insulation, Topic 15: Vibration Damping, and Topic 10: Water Pump Cooling (see Figure 14), in which September 2020 is used as the dividing line, the left area shows the effect of model fitting, and the right area shows the trend prediction results. At the same time, by calculating the relative error between the fitted and true values to further judge the fitting effect of the model, the data show that the maximum relative error is less than 17%, which is within the acceptable degree, indicating that the fitting effect of the model is very good and therefore the model is more accurate in predicting the data results.

(1) Topic 20: Development trend of security

According to ARIMA model prediction, from January 2021 to December 2025, Topic 20 (Safety) has a good development prospect, and the popularity will increase from 63.58 to 113.07, with an increase of 49.49, which is the highest increase among all technology topics. There are two types of automotive safety, active safety protection and

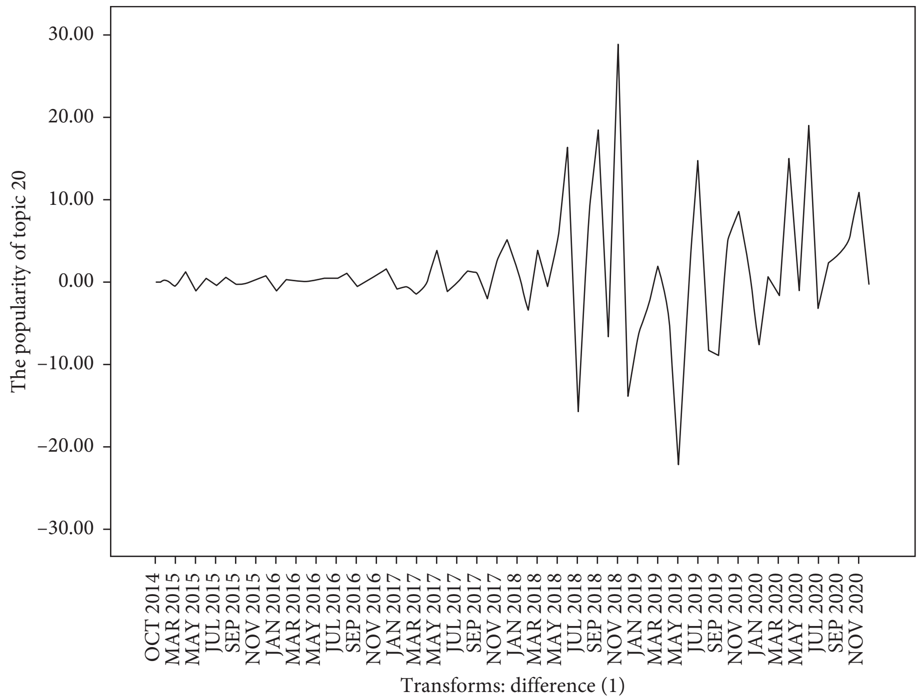


Figure 11: Timing diagram of popularity of Topic 20 after 1st-order differencing.

Table 5: Stability test.

Variable	ADF inspection value	5% statistical significance	*p* value	Stability
ER	−0.125792	−1.945389	0.6370	Unstable
ΔER	−10.66079	−1.945456	<0.0001	Stable

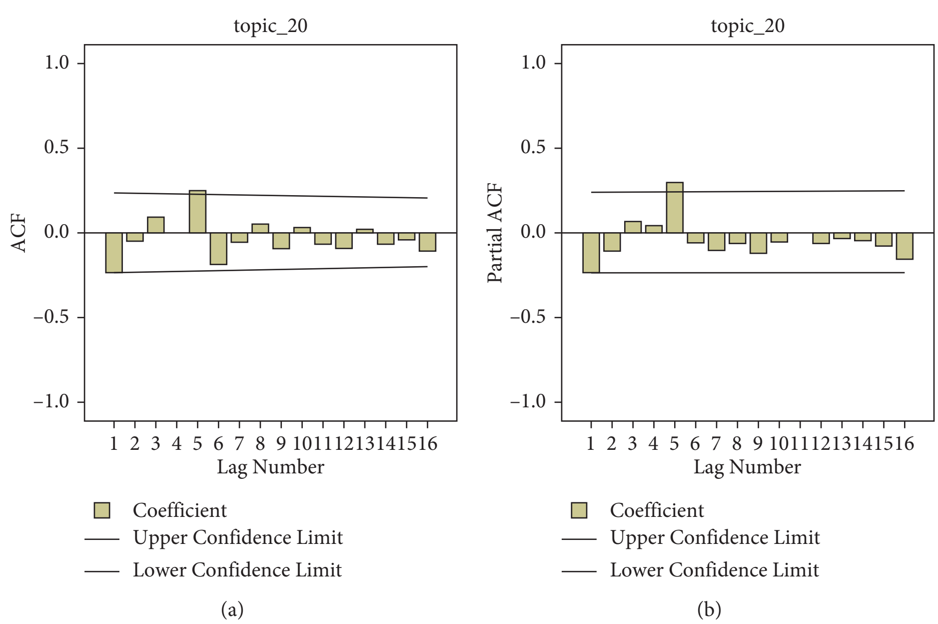


Figure 12: Autocorrelation and partial autocorrelation plots of subject 20 heat after 1st-order differencing. (a) Autocorrelation. (b) Partial autocorrelation.

Table 6: Model parameter determination.

ARIMA (p, d, q)	BIC	ARIMA (p, d, q)	BIC
ARIMA (0, 1, 1)	4.108	ARIMA (1, 1, 1)	4.184
ARIMA (0, 1, 5)	4.331	ARIMA (5, 1, 1)	4.376
ARIMA (1, 1, 0)	4.116	ARIMA (1, 1, 5)	4.366
ARIMA (5, 1, 0)	4.300	ARIMA (5, 1, 5)	4.531

Table 7: ARIMA model parameter estimation results.

Parameter	ARIMA (0, 1, 1)		
	Coefficient	T	p
Constant	−1.601	−1.062	0.292
MA (1)	0.282	2.416	0.018

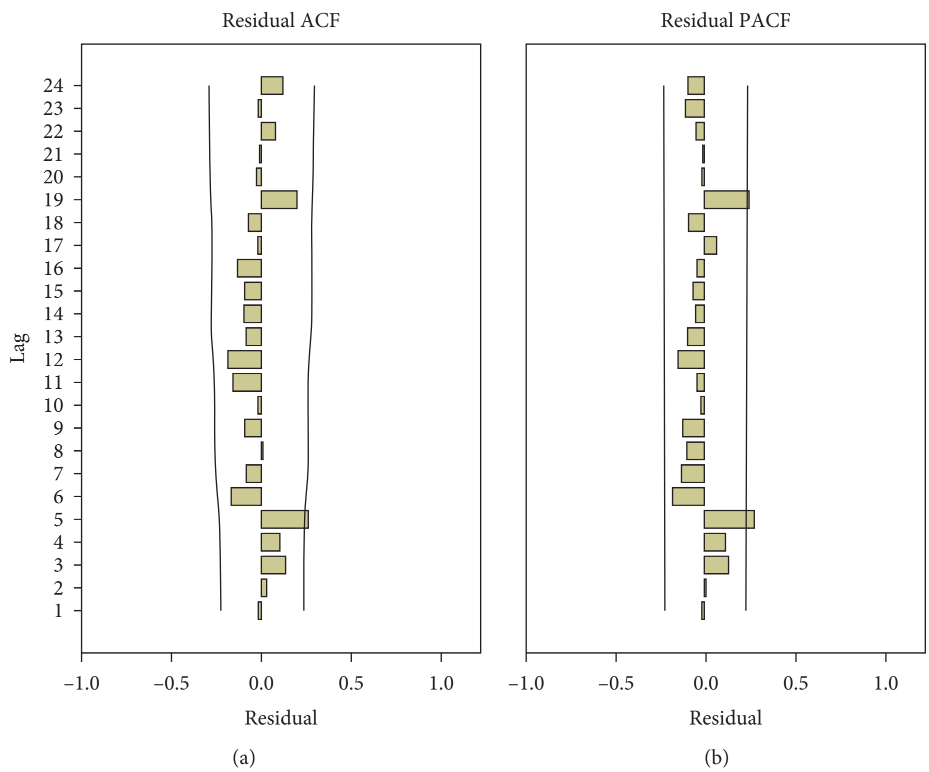


Figure 13: Autocorrelation and partial autocorrelation plots of the residuals. (a) Autocorrelation. (b) Partial autocorrelation.

passive safety protection. Active safety mainly detects the condition of the car, environment, and people and takes active measures to avoid accidents in advance. At the technical level, it is mainly optimized by active suspension system (ASS), antilock brake system (ABS), electronic brake-force distribution (EBD), electronic stability program (ESP), traction control system (TCS), lane departure warning system (LDWS), panoramic surround view (PSW), and so forth. Passive safety is mainly in the car traffic accident through passive equipment to minimize personal injury to passengers or pedestrians in the car; in the technical design it is mainly through airbags, seat belts, head and neck protection devices, crash beams, and a variety of collision energy-absorbing collapse structures, as well as other aspects of improvement. The collision resistance of new energy vehicles is not as good as that of traditional fuel vehicles, and a collision or charging overcurrent may lead to major accidents of battery spontaneous combustion. The transition from fuel vehicles to new energy vehicles has put forward new higher requirements for safety technology. Along with the deep integration of the automotive industry and the Internet information industry, the safety technology of new energy vehicles will be extended to the fields of Internet information security protection. Therefore, new energy vehicle safety technology is one of the main factors affecting its development and solving this type of pain point can greatly promote the popularization of new energy vehicles.

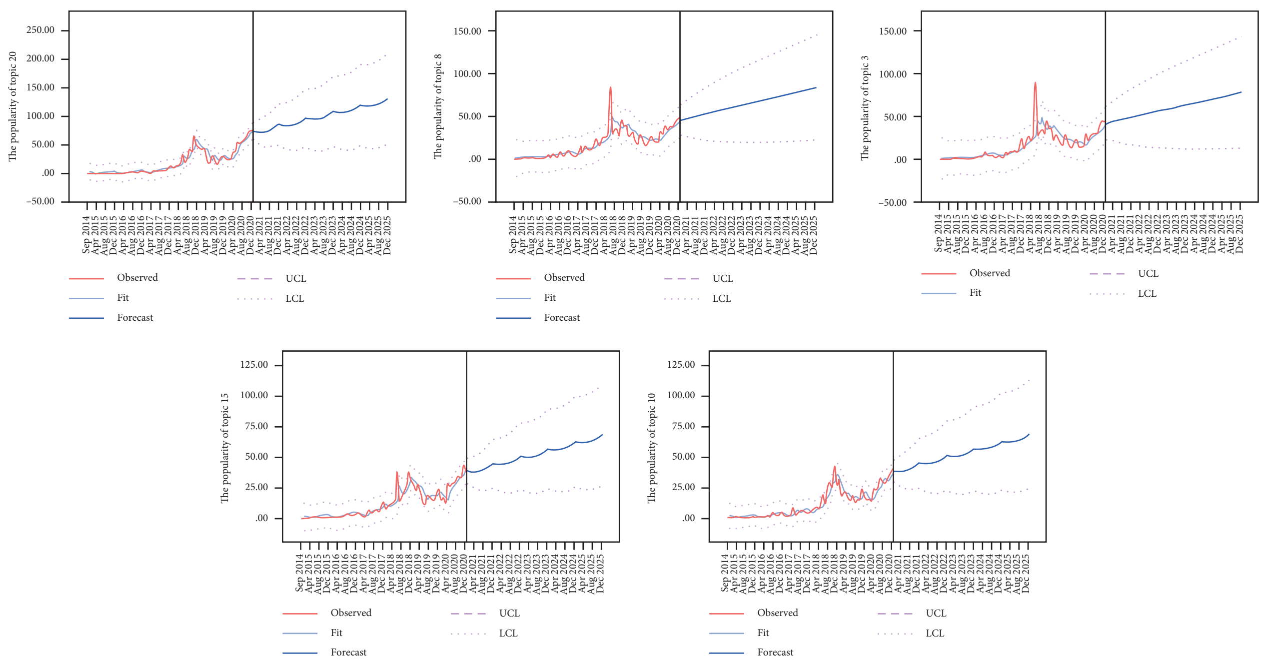


Figure 14: Forecast results of the development trend of new energy vehicle technology topics in China.

(2) Topic 8: Trends in mounting and fixing

Topic 8 (Installation and Fixation) is mainly to improve the efficiency of new energy vehicle parts processing and production, enhance the quality of product production, facilitate future use and maintenance, and transform the processing equipment and parts structure of new energy vehicles. For example, the patent text "a new energy vehicle motor assembly device" proposes improving the efficiency of motor lifting and installation, changing the way of human eye positioning and installation in the past, and setting up a positioning device for the hanging machine, so as to effectively improve the efficiency of production and reduce the work intensity of workers. In the patent text "a new energy vehicle processing parts welding device" for the existing new energy vehicle parts processing welding device has welding angle adjustment difficulties, welding stability is not good, welding combination tools cannot coordinate with the operation and other problems, a new energy vehicle parts processing welding device is designed, and successfully deal with the existing problems, convenient for workers to operate while improving the accuracy of welding. In order to solve the limitation that the punching device cannot adapt to the existing demand in the production process of new energy vehicle chassis, which has low punching accuracy and complicated operation, Gu Guangzhao designed a new energy vehicle chassis positioning punching device equipped with laser displacement sensor, which greatly improves the product production quality. By reforming this kind of fixed-type technology and optimizing the process of product processing, automotive manufacturers can produce products that satisfy consumers more efficiently. According to the ARIMA model, it is predicted that, from January 2021 to December 2025, the popularity of Topic 8 (Installation and Fixation) will continue to rise, from 37.61 to 66.97.

(3) Topic 3: Trends in cable insulation materials

Topic 3 (Cable Insulation Materials) has a pivotal role in the whole technology field of new energy vehicles. It can be seen from Figure 12 that the research heat of Topic 3 decreased slightly in March 2020, but, from the second half of 2020, the research heat of cable insulation materials slowly rose. From January 2021 to December 2025, the research trend for cable insulation materials shows a gentle increase from 31.65 to 56.23. As a highly integrated platform for electrification, cables are as indispensable as the vascular system of the human body. From a broad perspective, cable insulation materials are widely used, which can be used not only as components of the power transmission system in the vehicle but also in charging gun, charging pile, or on-board charging. Since new energy vehicles mainly use unconventional automotive fuels as their power source, the system components are required to transmit currents as high as between 250 A and 450 A when performing their work, which is relatively rare compared to conventional power-driven vehicles with such high current transmission. It is worth mentioning that this cable insulation material with high heat resistance is of great significance to improve the performance of the motor and can obviously reduce the mass-power ratio, namely, kg/kw, in the technical index of the motor. The mass-to-power ratio of electric motors has been reduced from 533 kg/kw in 1900 to 6 kg/kw in 2020. The important reason for this change is the use of insulating materials with high heat resistance, which reduces the consumption of large amounts of metal materials and lowers the cost of motor production. In new energy vehicles, insulation materials account for about half of the cost consumption ratio of electric motors and electrical products, which illustrates the status and role of cable insulation materials research in the new energy vehicle industry. In addition, the development of highly flexible cable insulation materials facilitates the accommodation of more electronic components in a limited space. In the application of charging guns and charging posts, cable insulation materials need to overcome a variety of uncertain and unfavorable factors from sunlight, weathering, humidity, seawater, acid rain, freezing, and so forth and also meet the needs of bending and dragging in use, which puts new demands on cable insulation materials such as high and low temperature resistance, oil resistance, water resistance, flame retardancy, tear resistance, electrical insulation, and UV (ultraviolet) aging resistance performance put forward more and higher requirements.

(4) Topic 15: Development trend of vibration damping

Based on the ARIMA model, it is predicted that, from January 2021 to December 2025, the research enthusiasm of Topic 15 (Vibration Damping) shows a wave-like upward trend from 36.18 to 64.88. Topic 15 (Vibration Damping) seems to be less important and less technical in the automotive industry technology research. However, the reality is that damping technology can avoid damage to car parts by effectively suppressing the reciprocal rebound force generated after spring absorption, by reducing the vibration of the frame and body, and at the same time can effectively improve the smoothness and comfort during the driving process. Expanding to the field of new energy vehicles, the damping technology can have other new applications in addition to the above-mentioned roles. Through a careful reading of the patent text, it is found that the damping technology is generally applied to provide damping devices for batteries. This is because there are various precision electronic parts inside the battery module assembly, which have high environmental requirements in the process of use, especially in shock absorption. The vibration generated by the car driving on bumpy roads can easily affect the battery, and the internal components of the battery are likely to loosen, thus leading to failure and reducing the service life of the battery. The improvement of vibration damping technology can protect the parts of new energy vehicles to the greatest extent, reduce the failure rate of vehicles, ensure the smoothness of driving, and thus improve people's riding experience.

(5) Topic 10: Development trend of water pump cooling

Although Topic 10 (Water Pump Cooling) is not as hot as Topic 20 (Safety), its research popularity follows the development of the overall technology field, as it plays an indispensable role in the field of new energy vehicle technology. Based on ARIMA model prediction, the popularity of topic 10 (Water Pump Cooling) grows from 31.43 to 55.68 from January 2021 to December 2025. Topic 10 (Water Pump Cooling) is mainly used in the process of driving new energy vehicles, the internal components of the motor continuously generate losses during the movement, and these losses are eventually dissipated to the outside through heat. These losses are eventually dissipated to the outside by way of heat, and it is necessary to take away unnecessary heat by effective cooling, so as to ensure the normal and stable operation of the motor in a stable hot and cold circulation system. According to the principle of cooling, it can be divided into three types: natural cooling, air cooling, and water cooling. Natural cooling mainly relies on the heat transfer from the motor body into itself to dissipate the heat generated by the motor. This type of cooling method is inefficient and only applicable to low-power generators. Air cooling is mainly through the coaxial fan to form the cooling air circulation, using the air around the motor as the medium, directly discharging the heat of the motor to the surrounding environment. Water cooling uses coolant as the medium and promotes the continuous circulation of coolant through pipes to take away the heat generated by the motor rotor and stator during operation. Among these three cooling methods, water cooling has the greatest impact on the efficiency improvement of new energy vehicles. At the same time, not only is the application of such technology limited to the cooling of the electric motor, but also it includes the cooling of the battery. During the high load and long-term driving of new energy vehicles, the continuous high temperature caused by the continuous discharge of the battery will inevitably lead to the high temperature of the coolant, which makes the efficiency of the water cooling system drop sharply and leads to the abnormal power supply of the battery at high temperature and also has a negative impact on the service life of the battery. Therefore, the transformation of water cooling technology can provide guarantee for the normal operation of new energy vehicles, which has an important impact on important components such as batteries and motors.

Looking at the above five topics with the fastest growing in research, it can be found that the overall development trend of each technology topic is relatively consistent, the development is relatively slow before the beginning of 2017, and the number of patent issuance is also relatively small. But, with the guidance of policies, the government increased the encouragement of automobile manufacturing enterprises to improve the quality of new energy vehicles through technological innovation, while providing a large number of preferential subsidy policies to consumers, which to a certain extent promoted the rapid development of new energy vehicles; as can be seen from the 5 technology topics after 2018, the technology of new energy vehicles has been rapidly accumulated. The results show that these five technology topics will continue to grow in popularity from 2021 to 2025.

5. Conclusions and Policy Implications

This paper identifies technical topics in the field of new energy vehicles in China and constructs ARIMA model to predict the trend of popularity of technology topics. The main conclusions are as follows.

Firstly, through the topic mining of patent texts in the field of new energy vehicles in China from 2000 to 2020 in the patent database of CNKI, it is found that there are 25 technology topics in the field of new energy vehicles. In terms of importance, the five most important technical topics in the field of new energy vehicles in China are as follows: Topic 8: Installation and Fixation, Topic 5: Heat Dissipation and Cooling, Topic 14: Vehicle Data Monitoring, Topic 9: Charging Piles, and Topic 15: Vibration Damping. Introducing the time dimension, the importance of Topic 5: Thermal Cooling, Topic 8: Installation and Fixation, Topic 6: Electric Drive Systems, Topic 9: Charging Piles, and Topic 15: Vibration Damping grows steadily from 2014 to 2020. In 2014–2020, Topic 7: Chassis, Topic 10: Water Pump, Topic 18: Electric Motor, and Topic 24: Module are relatively slow in development. In 2014–2018, Topic 12: Vacuum Pump and Topic 21: Wireless Charging steadily increase in importance, but, in 2019–2020, the importance decreases.

Second, the LDA topic model is trained unsupervised, and the evolution trend of each technical topic can be obtained by dividing the importance of each technical topic over the years by the overall importance of the technical topic. It was found that the importance of Topic 3: Cable Insulation, Topic 14: Vehicle Data Monitoring, and Topic 23: Circuit Control declined from 2014 to 2020, while the importance of Topic 5: Thermal Cooling and Topic 8: Installation and Fixation continued to increase.

Third, a linear regression model was used to portray the overall trend of changes in the 25 technology topics, and the top 5 technology topics in terms of hotness were screened according to the slope of the linear trend: Topic 20: Safety, Topic 8: Installation and Fixation, Topic 3: Cable Insulation, Topic 15: Vibration Damping, and Topic 10: Water Pump Cooling. Based on ARIMA prediction, it is found that 2021–2025 will be a booming period for new energy vehicles in China, and the popularity of Topic 20: Safety, Topic 8: Installation and Fixation, Topic 3: Cable Insulation Materials, Topic 15: Vibration Reduction, and Topic 10: Water Pump Cooling continues to grow over time. The popularity of Topic 20: Safety, Topic 3: Cable Insulation, and Topic 10: Water Pump Cooling shows a wave upward growth trend. Topic 20: Heat will grow from 63.58 to 113.07 with a value increase of 49.49, which is the fastest growth in heat among all technical topics.

Our findings not only contribute to advancing the study of China's new energy vehicles innovation management and industrial development but also provide inputs for policymakers. Three policy implications are proposed based on the above analysis.

First, strengthen technology research and development in the field of safety to protect the development of the new energy vehicle industry. The study found that, among the 25 technology topics, Topic 20 (Safety) was the most popular. The safety of new energy vehicles is gradually attracting the attention of the industry, which is related to the safety of consumers and even the whole society, as well as the confidence of social groups in the development of new energy vehicle industry, which in turn affects the future development direction and space of new energy vehicle industry. Therefore, it is recommended to do the following: ① The subsidy policy of the new energy vehicle industry should be tilted to the field of technology research and development, focusing on subsidizing the research and development of core components such as batteries, motors, electric controls, and chips to enhance the safety of new energy vehicles. ② The subsidy policy of new energy vehicle industry should be directed to the R&D of battery and its related supporting technologies, support the R&D of graphene application in the field of battery, and improve the safety of new energy vehicle battery. ③ Strengthen the technology research and development of active safety, mainly involving active suspension system (ASS), antilock brake system (ABS), electronic brake-force distribution (EBD), electronic stability program (ESP), traction control system (TCS), lane departure warning system (LDWS), panoramic surround view (PSW), and so forth. ④ Strengthen the technology development of passive safety, mainly involving airbags, seat belts, head and neck protection devices, and so forth. ⑤ Improve the monitoring system of new energy vehicle enterprises and do a good job of after-sales quality tracking and inspection of products. ⑥ Strengthen the supervision of safety from core components to the whole vehicle, and implement the battery catalog system and battery coding system.

Second, accelerate the construction of new energy vehicle infrastructure. With the continuous development of the new energy vehicle industry, the charging problem has gradually become an important factor limiting the development of the new energy vehicle industry, and the lack of convenience of charging facilities has become a common concern for consumers, increasing the uncertainty of market promotion. As of 2020, the number of public charging piles in China was 807,000. Among them, 145,000 were in residential areas, accounting for 18%; the number of private charging piles was 870,000; the number of public charging stations was 63,800; and the number of exchange stations was 555. Despite the promising achievements of China's charging infrastructure in the "13th Five-Year Plan" period, there are still many shortcomings; mainly in the scale of charging facilities there is still a gap, the spatial layout needs to be optimized, charging efficiency still needs to be improved, the utilization rate of public charging facilities is low, and the social environment and synergy mechanism supporting charging facilities have not yet been formed. In order to accelerate the construction of new energy vehicle infrastructure, it is recommended to do the following: ① The financial subsidy policy for new energy vehicles should be gradually tilted toward the construction of supporting infrastructure. ② According to the market demand, appropriately increase the charging piles and other infrastructure, and encourage local governments to build charging spaces in public places such as paid parking spaces, temporary parking strips on the roadside, and parking lots in residential communities according to local conditions, so as to improve the utilization rate of charging spaces. ③ Ensure that residential communities, public institutions, and so forth are in accordance with a certain proportion of charging facilities to properly solve the difficult problem of charging piles into the community.

Third, to strengthen the technology development of Topic 8 (Installation and Fixation), Topic 5 (Heat Dissipation and Cooling), Topic 14 (Vehicle Data Monitoring), Topic 9 (Charging Pile), Topic 15 (Vibration Damping), and Topic 3 (Cable Insulation) to improve the overall performance of new energy vehicles, the study found that the above six topics have high scores in terms of importance and popularity, representing the technical dynamics and development trend of the new energy vehicle technology field. Therefore, new energy vehicle enterprises must strengthen the investment in the above technical fields and strive to gain the initiative to develop in the fierce competition. During the "13th Five-Year Plan" period, China's production and sales of new energy vehicles have grown rapidly, ranking first in the world for five consecutive years since 2015, with a total of more than 4.8 million vehicles promoted, accounting for more than 50% of the world, and new energy vehicles are entering thousands of households. However, we should also see that, compared with traditional fuel cars, the current market share of new energy vehicles is still very small, and to replace traditional fuel cars is a long way off. Historically, new energy vehicles have failed repeatedly in the competition with traditional fuel cars. Therefore, it is crucial to seize the technical dynamics and development trend in the field of new energy vehicle technology, strengthen the research and development of technical topics such as installation and fixation, heat dissipation and cooling, vehicle data monitoring, charging pile, vibration damping, and cable insulation materials, and improve the overall performance of new energy vehicles for the development of new energy vehicle industry.

Data Availability

All data in this paper are available upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The support by Humanities and Social Science Research Projects of Ministry of Education (no. 18YJC630050), Guangdong Province National Natural Science Foundation of China (no. 2018A030313227), Soft science Project of Guangdong Province (no. 2019A101002057), and the National Natural Science Foundation of China (nos. 71874036, U1901222, 72074056, 71673064, and 71974039) is acknowledged.

References

[1] L. Zhang, "Research on business model innovation and path evolution of china's new energy vehicle from the perspective of sociotechnical system," *Beijing Jiaotong University*, vol. 11, no. 6, p. 1711, 2020.

[2] Y. Chen, L. Ni, and K. Liu, "Does China's new energy vehicle industry innovate efficiently? a three-stage dynamic network slacks-based measure approach," *Technological Forecasting and Social Change*, vol. 173, no. 12, Article ID 121161, 2021.

[3] Q. Zhou, X. Gu, L. Zeng, and S. Wu, "The collaboration mechanism of patent implementation in new energy automobile industry," *Studies in Science of Science*, vol. 38, no. 7, pp. 1236–1243+1259, 2020.

[4] G. Zhang, W. Xie, R. Hu, W. Ma, and Y. Liu, *Disruptive Innovation: SNM Perspective*, Science Press, Beijing, China, 2016.

[5] Z. Wu, Q. Shao, Y. Su, and D. Zhang, "A socio-technical transition path for new energy vehicles in China: a multi-level perspective," *Technological Forecasting and Social Change*, vol. 172, no. 11, Article ID 121007, 2021.

[6] B. Sun, X. Xi, and H. Yu, "A study on technology niche's from the perspective of network relationship: empirical analysis of new energy vehicle industry in northeast three province perspective," *Studies in Science of Science*, vol. 31, no. 4, pp. 518–528, 2013.

[7] T. Yang, X. Chen, and X. Li, "Evaluation and analysis of new-energy vehicle industry policies in the context of technical innovation in China," *Journal of Cleaner Production*, vol. 281, no. 1, Article ID 125126, 2021.

[8] R. Wang, X. Zhao, W. Wang, and L. Jiang, "What factors affect the public acceptance of new energy vehicles in underdeveloped regions? A case study of Gansu Province, China," *Journal of Cleaner Production*, vol. 318, no. 10, Article ID 128432, 2021.

[9] L. Liu, T. Zhang, A. P. Avrin, and X. Wang, "Is China's industrial policy effective? An empirical study of the new energy vehicles industry," *Technology in Society*, vol. 63, no. 11, Article ID 101356, 2020.

[10] Z. Zhang and F. Zhao, "The comparative study on the development strategies of the new energy automobile industry between China and America: based on the differences of goal orientation," *Studies in Science of Science*, vol. 32, no. 4, pp. 531–535, 2014.

[11] Y. Huang, Y. Hou, T. Liu, and Z. Gao, "Research on BEV industry competitive profile under patent information field of vision," *Science & Technology Progress and Policy*, vol. 34, no. 4, pp. 72–77, 2017.

[12] Z. Wang, "Characteristics and utilization of patent documentation," *Journal of Modern Information*, vol. 2008, no. 9, pp. 151-152+156, 2008.

[13] H Xiao, *Patent Map Methods and Applications*, Shanghai Jiao Tong University Press, Shanghai, China, 2011.

[14] S. Jun, "Central technology forecasting using social network analysis," in *Computer Applications for Software Engineering, Disaster Recovery, and Business Continuity," Communications in Computer and Information Science*, T. Kim, C. Ramos, H. Kim, A. Kiumi, S. Mohammed, and D. Ślęzak, Eds., vol. 340, Berlin, Heidelberg, Springer, 2012.

[15] R. N. Kostoff, R. Boylan, and G. R. Simons, "Disruptive technology roadmaps," *Technological Forecasting and Social Change*, vol. 71, no. 1-2, pp. 141–159, 2004.
[16] Y. Zhang, W. Zhang, and J. Bi, "Recent advances in driverless car," *Recent Patents on Mechanical Engineering*, vol. 10, no. 1, pp. 30–38, 2008.
[17] K. Suzuki, J. Sakata, and J. Hosoya, "An empirical analysis on progress of technology fusion," in *Proceedings of the 2008 IEEE Third International Conference on Digital Information Management*, no. 11, pp. 937–939, London, United Kingdom, July 2008.
[18] S. Jeong, J.-C. Kim, and J. Y. Choi, "Technology convergence: what developmental stage are we in?" *Scientometrics*, vol. 104, no. 3, pp. 841–871, 2015.
[19] W. S. Lee, E. J. Han, and S. Y. Sohn, "Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents," *Technological Forecasting and Social Change*, vol. 100, no. 11, pp. 317–329, 2015.
[20] B. Huang and L. Huang, "Patent-based association analysis for technological convergence of IT and BT," in *Proceedings of the 2015 International Conference on Logistics, Informatics and Service Sciences*, pp. 1–6, IEEE, 2015.
[21] J. Park and D. Kang, "Trend analysis of unmanned technology using patent information," *Journal of The Korea Society of Computer and Information*, vol. 22, no. 3, pp. 89–96, 2017.
[22] F. Narin, "Patent bibliometrics," *Scientometrics*, vol. 30, no. 1, pp. 147–155, 1994.
[23] O. Kwon, J. Seo, K. Noh, J. Kim, J. S. Kim, and S. Y. Shin, "Categorizing influential patents using bibliometric analysis of patent citations network," *Information-an International Interdisciplinary Journal*, vol. 10, no. 3, pp. 313–326, 2007.
[24] C. Choi and Y. Park, "Monitoring the organic structure of technology based on the patent development paths," *Technological Forecasting and Social Change*, vol. 76, no. 6, pp. 754–768, 2009.
[25] C. C. Hsueh and C. C. Wang, "The use of social network analysis in knowledge diffusion research from patent data," in *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, pp. 393–398, IEEE, Athens, Greece, July 2009.
[26] Y. Geum, C. Kim, S. Lee, and M. S. Kim, "Technological convergence of IT and BT: evidence from patent analysis," *ETRI Journal*, vol. 34, no. 3, pp. 439–449, 2012.
[27] D. Zhai, L. Cai, J. Zhang, and X. Zhang, "The study of patent-based model for identifying technology fusion innovation trajectory illustrated by the case of cloud computing technology," *Journal of the China Society for Scientific and Technical Information*, vol. 34, no. 4, pp. 352–360, 2016.
[28] C. Kim and H. Seol, "On a patent analysis method for identifying core technologies intelligent decision technologies," in *Proceedings of the 4th International Conference on Intelligent Decision Technologies*, pp. 441–448, Springer, Gifu, Japan, May 2012.
[29] H. Small, K. W. Boyack, and R. Klavans, "Identifying emerging topics in science and technology," *Research Policy*, vol. 43, no. 8, pp. 1450–1467, 2014.
[30] B. Li and X. Chen, "Identification of emerging technologies in nanotechnology based on citing coupling clustering of patents," *Journal of Intelligence*, vol. 34, no. 5, pp. 35–40, 2015.
[31] J. Lee, J. Kim, J. Lee, S. Park, and D. Jang, "A study on analysis of R&D intensity based on patent citation information: case study on self-driving car of Google," *Journal of Korean Institute of Intelligent Systems*, vol. 26, no. 4, pp. 327–333, 2016.
[32] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 2003, no. 3, pp. 993–1022, 2003.
[33] X. Qin and X. Le, "Topic evolution research on a certain field based on LDA topic association filter," *Data Analysis and Knowledge Discovery*, vol. 31, no. 3, pp. 18–25, 2015.
[34] E. Yan, "Research dynamics, impact, and dissemination: a topic-level analysis," *Journal of the Association for Information Science and Technology*, vol. 66, no. 11, pp. 2357–2372, 2015.
[35] H. Yang and J. Yang, "Quantitative analysis of policy text merged with LDA model: based on the field of international climate as demonstration," *Journal of Modern Information*, vol. 36, no. 5, pp. 71–81, 2016.
[36] L. Liao and F. Le, "Research on patent technology evolution based on LDA model and classification number," *Journal of Modern Information*, vol. 37, no. 5, pp. 13–18, 2017.
[37] W. Wang and J. Ma, "Topic detection and evolution analysis of research project based on LDA ———a case study of projects on ocean acidification supported by NSF," *Journal of Intelligence*, vol. 36, no. 7, pp. 34–39, 2017.
[38] S. Fan, X. An, G. Yan, and Y. Li, "Study on the recognition method of Frontier topic in the medical field," *Journal of the China Society for Scientific and Technical Information*, vol. 37, no. 7, pp. 686–694, 2018.
[39] H. Yi, H. Wu, and J. Chang, "Technology topics evolution of graphene based on topic life cycle and technology entropy," *Journal of Intelligence*, vol. 38, no. 7, pp. 64–70, 2019.
[40] J. Shen, Q. Yan, Q. Sun, and Z. Wan, "Research on technology fusion recognition and technology opportunity prediction based on patent data mining," *Library Journal*, vol. 38, no. 10, pp. 95–106, 2019.
[41] H. Ma, L. Kong, C. Lin, and X. Yang, "Key generic technology identification based on patent mining," *Journal of the China Society for Scientific and Technical Information*, vol. 39, no. 10, pp. 1093–1103, 2020.
[42] M. Feng, F. Wei, J. Li, and L. Zheng, "Research on topic identification and technology evolution from the perspective of industrial Chain: a case study of 3D-printing," *Journal of Intelligence*, vol. 39, no. 8, pp. 46–52, 2020.
[43] H. Li and H. Xuan, "Multi-attribute mining method for technology innovation subject from the perspective of paten: the case of chip patents," *Library and Information Service*, vol. 64, no. 11, pp. 496–107, 2020.
[44] H. Tang and Y. Qiu, "Emerging technology topic identification based on multi-source information: intelligent connected vehicle as an example," *Journal of Intelligence*, vol. 40, no. 3, pp. 81–88, 2020.
[45] Y. Lianga, Y. Jua, L. Martínezb, and T. Yan, "Sustainable battery supplier evaluation of new energy vehicles using a distributed linguistic outranking method considering bounded rational behavior," *Journal of Energy Storage*, vol. 48, no. 4, Article ID 103901, 2022.
[46] B. Lin and L. Shi, "Do environmental quality and policy changes affect the evolution of consumers' intentions to buy new energy vehicles," *Applied Energy*, vol. 310, no. 3, Article ID 118582, 2022.
[47] K. Song and Y. zhu, "Patent Frontier technology topic identification and trend prediction: a case analysis of artificial intelligence," *Journal of Intelligence*, vol. 40, no. 1, pp. 33–38, 2021.
[48] L. F. Yang, J. H. Xu, and P. Neuhäusler, "Electric vehicle technology in China: an exploratory patent analysis," *World Patent Information*, vol. 35, no. 4, pp. 305–312, 2013.

[49] Z. Xie, Y. Zhang, C. Zheng, and M. Zhang, "Research on patent trend of new energy vehicles industry," *Enterprise Management*, vol. 2015, no. 4, pp. 127–141, 2015.
[50] S. Liang, X. Ji, and Y. Li, "Application of patent scientometrics methods in technology foresight: take the new energy automobile as an example," *Journal of Intelligence*, vol. 34, no. 2, pp. 73–78, 2015.
[51] J. Wang, Y. Liu, and K. Ari, "Research on technology innovation of Chinas NEV industry based on patent analysis," *Journal of Intelligence*, vol. 35, no. 1, pp. 32–38, 2016.
[52] G. Li and X. Fan, "The choice of the key new energy vehicle technology of SAIC basing on the patent analysis," *Journal of Modern Information*, vol. 37, no. 10, pp. 122–127+136, 2017.
[53] X. Shi, R. Jing, G. Zheng, and G. Hou, "Analysis of regional distribution characteristics of new energy vehicle technology innovation in China based on patent data," *Journal of Industrial Technological Economics*, vol. 37, no. 08, pp. 60–67, 2018.
[54] H. Sun, Y. Geng, L. Hu, L. Shi, and T. Xu, "Measuring China's new energy vehicle patents: a social network analysis approach," *Energy*, vol. 153, no. 6, pp. 685–693, 2018.
[55] S. Feng, H. An, H. Li et al., "The technology convergence of electric vehicles: exploring promising and potential technology convergence relationships and topics," *Journal of Cleaner Production*, vol. 260, no. 7, Article ID 120992, 2020.
[56] H. Miao, C. Tang, and L. Luo, "Sales forecast of new energy vehicles based on ARIMA model," *Enterprise technology and development*, vol. 2020, no. 10, pp. 97-98, 2020.
[57] J. Wu, M. Hao, and J. Hang, "The Construction of enterprise innovation ecosystem in the "Internet+" environment: the case study of NIO NEV," *Soft Science*, vol. 35, no. 5, pp. 70–77, 2021.
[58] X. Fang and L. Zhan, "Prediction and analysis on new energy vehicle charging pile technology at home and abroad: based on lifecycle and international patent classification," *Science and Technology Management Research*, vol. 41, no. 8, pp. 148–155, 2021.
[59] X. Xu and M. Gui, "Research on identification and layout of core technologies of new energy vehicles based on machine learning," *Science and Technology Management Research*, vol. 41, no. 9, pp. 96–106, 2021.
[60] X. Tang and J. Liu, "Automatic indexing of questions in Q&A community based on BERT and TF-IDF——taking the CNGOLD Q&A community as an example," *Information Science*, vol. 39, no. 3, pp. 3–10, 2021.
[61] P. Guan and Y. Wang, "Identifying optimal topic numbers from Sci-Tech information with LDA model," *New Technology of Library and Information Service*, vol. 2016, no. 9, pp. 42–50, 2016.
[62] D. Qian and J. Zheng, "Research on subject extraction and evolution of information service literature based on LDA topic model," *Digital Library Forum*, vol. 2019, no. 10, pp. 16–22, 2019.
[63] J. Zhu and L. Hu, "Comparative analysis of RMB exchange rate forecast based on ARIMA and BP neural network: take the exchange rate of US dollar to RMB as an example," *Journal of Chongqing University of Technology (Natural Science)*, vol. 33, no. 5, pp. 207–212, 2019.
[64] J. Qian and Y. wang, "Application comparison of several statistical software to build ARIMA mode," *Statistics and Decision Making*, vol. 34, no. 16, pp. 80–84, 2018.

Hindawi
Complexity
Volume 2022, Article ID 7135969, 15 pages
https://doi.org/10.1155/2022/7135969

Research Article

Analysis of the Spatial Characteristics and Influencing Factors of e-Commerce Industrial Chains from the Perspective of Embeddedness: Taking Xiong'an New Area in China as an Example

Yifan Gu and Jianghui Ding

School of Geographic Sciences, Hebei Normal University, Shijiazhuang 050024, China

Correspondence should be addressed to Jianghui Ding; widjh@163.com

Received 24 December 2021; Accepted 28 February 2022; Published 17 March 2022

Academic Editor: Haitao Ma

e-commerce has promoted the application of the Internet by traditional enterprises. While this phenomenon has deepened industrial embeddedness, it has also profoundly affected the development of traditional manufacturing and retail industries. Based on the embeddedness theory, this paper takes the e-commerce industrial chain in Xiong'an New Area of China as an example, uses location entropy method and local spatial autocorrelation to analyze its spatial characteristics and influencing factors, and also discusses the features of embeddedness from each dimension of industrial chains. The results show the following: ① The supply chain of industry-dependent e-commerce is dominated by production wholesalers, and suppliers with cost advantages and commercial traditions are embedded in local characteristic industries; the e-commerce value chain is dominated by production and processing activities, and the spatial orientation of the business model is affected by regional industrial functions; the e-commerce enterprise chain is dominated by B2B enterprises, the high-level agglomeration of which reflects the absolute scale advantage provided by the e-commercialization of traditional industry. ② The distribution of the e-commerce industrial chains of Xiong'an New Area is characterized by spatial heterogeneity and has obvious spatial agglomeration. The industrial chains are concentrated in Xiong County, followed by Rongcheng County, and Anxin County has the weakest agglomeration. ③ The main factors that cause the uneven distribution of the e-commerce industrial chains in Xiong'an New Area include three aspects: transportation location advantages, traditional industrial foundation, and government policy regulation. Research can provide a scientific reference for accurately identifying the spatial development characteristics of the industrial chains from various perspectives and promoting the sustainable development and e-commercialization of traditional industries.

1. Introduction

With the transformation and upgrading of information and communication technologies (ICTs), e-commerce has not only changed consumers' shopping concepts and shopping behaviors but also greatly changed the traditional commodity-circulating mode [1–3]. The attempts of traditional enterprises to e-commerce have appeared in large numbers in developed and developing countries successively and have brought huge economic benefits [4–6]. The global retail e-commerce sales reached 4,135 billion US dollars in 2020, with an annual growth rate of about 15% [7]. The e-commerce entity is a key node in geographic space and cyberspace, which possesses the dual attributes of reality and virtuality, material flow, and information flow [8]. The rapid expansion of e-commerce scale has enabled various industrial departments to objectively form an e-commerce industrial chain based on technical and economic connections in accordance with a certain logical relationship and spatial organization [9]. Nowadays, the scope of the e-commerce industrial chain has become wider, and its competitive advantages are increasing day by day [10], which has gradually infiltrated into all areas of global business, playing a significant role in optimizing and reconstructing the regional industrial system.

Previous research on e-commerce is mainly based on the perspectives of economics and sociology, and there are relatively few geographical studies. Research of economics mainly starts from the theories of industrial economics, evolutionary economics, and trade economics, focusing on the industrial evolution [11], industrial clusters [12], industrial integration [13], and value chain models [14] of e-commerce. Sociological research is based on the theory of space production and sustainable livelihoods and so on, focusing on social restructuring [15], social transformation [16, 17], and social governance [18, 19] in urban and rural areas driven by e-commerce. Geography researchers mainly start from the perspective of space, analyzing the spatial development law, formation mechanism, regional influence, and spatial effect of e-commerce. The main contents include the spatial organization characteristics of e-commerce entities from different scales [20–22], the formation mechanism and influencing factors of e-commerce industrial clusters [23–25], the relationship between e-commerce enterprises and brick-and-mortar enterprises [26, 27], and the influence of e-commerce development on traditional industries and urban commercial spatial organization [28–30]. There are a few research about the e-commerce industrial chain from geographical perspectives, which have only studied the formation and development of the industrial chain of Taobao villages at the mesomicroscale and the reorganization of the urban supply chain [31–33]. The research reflects the spatial impact of e-commerce industry agglomeration and the importance of cross-border e-commerce in the global value chain [34, 35]. Overall, the study found that, as a kind of knowledge economy, e-commerce has crossed time and space and become the new driving force of national and regional economic development [36]. With the advancement of information technology, the fragmented physical space and industrial elements in various regions are connected to each other, which increases the complexity of the spatial structure of e-commerce entities, reflecting the obvious characteristics of spatial heterogeneity. The main factors that drive and affect the e-commerce industry cluster include regional industrial foundation, economic level, transportation environment, degree of informatization, and social attributes of employees [37]. The expansion of e-commerce information flow, material flow, and capital flow has reconstructed the spatial organization of the supply and demand market and the shopping process, forming the effects of competition, substitution, and complementation with brick-and-mortar retail. Furthermore, it affects the location decision of entity enterprise and drives their agglomeration development. The supporting effect of traditional industrial clusters and the pulling effect of market demand make e-commerce products more embedded in the place, indicating that the importance of traditional geographic factors has not disappeared in cyberspace, which has further promoted the spatial infiltration and integration of virtual and real economy. Due to the differences in online shopping patterns, commodity types, trading platforms, and transaction entities, there will be apparent discrepancies in the functioning manners and intensity of geographical factors in online shopping, which makes e-commerce show different geographical embeddedness.

Our research focuses on the spatial characteristics of the e-commerce industrial chain and the characteristics of its industrial embeddedness. Embeddedness, which is also called "endogeneity," is an important concept in economic geography [38]. It refers to the natural and inherent connection between the development of social and economic activities and their location [39]. It is profoundly affected by a long-term, basic, and intrinsic relationship brought about by the industrial chain or industrial network composed of local resource and affiliated enterprises [40]. Previous studies have integrated the conceptions of embeddedness into the theories of global space economy [41, 42]. Some scholars believe that embeddedness is a dynamic, longitudinal, and adaptive process that evolves with the reshaping of the supply chain, the expansion of industry scale, and the innovation of technology [43]. The embeddedness of the company in the region reflects the sticky relationship and the degree of integration between the two. There are the results of both the company's choice to be embedded in a certain industry or region and a certain industry or region's choice to develop and retain the company. The company's space selection depends on its location decision, and the regional viscidity depends on the influence of the regional economy, society, culture, and so on [44]. The involvement of the Internet increased the product sales channels, which made brick-and-mortar companies accelerate the speed of transferring the knowledge and technology to e-commerce enterprises. Then, the transfer and adaptation of consumer preferences make the e-commerce industry more embedded in the entity industry and the life of consumer groups, which leads to (re)creating new spatial characteristics [45]. Combining the spatial organization of the e-commerce industrial chain, the study of the industrial embeddedness is helpful to judge whether the local traditional industry will be eliminated or not under the boom of e-commercialization and fierce market competition.

All in all, despite the work of spatial development and its influencing factors of e-commerce in different scales and different models, there is a lack of research on the e-commerce industrial chain's spatial distribution characteristics. In addition, the study of e-commerce embeddedness in traditional industries is relatively limited. China's state-level new area, Xiong'an New Area, covers Xiong County, Anxin County, Rongcheng County, and some surrounding areas. After decades of development and accumulation, the counties have formed some traditional industries with a considerable industrial scale, most of which are labor-intensive manufacturing industries, and the overall added product value is low. Compared with the high-end positioning requirements of the development plan of Xiong'an New Area, the nonadaptability characteristics are obvious. With the help of the Internet platform and under the leadership of the new economic model of e-commerce, the way for traditional industries to adapt to the future economic functions and industrial division is a question worth considering. Taking Xiong'an New Area as an example, using spatial autocorrelation and location entropy methods, the authors analyzed the spatial distribution characteristics and influencing factors of different types of e-commerce industry chains at the township scale. The spatial heterogeneity of e-commerce embeddedness was clarified. This work is expected to supplement and improve the research on the spatial organization of regional industrial chain, reveal the survival and development ability of traditional industries embedded by e-commerce, and provide a scientific reference for promoting the integrated development of the digital economy and the substantial economy.

2. Understanding e-Commerce Industrial Chain and Its Embeddedness

2.1. Dimensions of e-Commerce Industrial Chains. In 1958, Hirschman put forward and explained the concept of industrial chain from the perspective of industry's backward-forward linkages in the book "The Strategy of Economic Development" [46]. The industrial chain describes the complex and dynamic relationships at different levels among enterprises, industries, and regions. Western literature seldom uses the traditional concept of "industrial chain" alone and rarely sees the industrial chain as one unit for systematic research. Instead, value chain, supply chain, logistics chain, and commodity chain are more discussed [47–50]. Domestic and foreign perspectives on the industrial chain are different as a result of economic and institutional backgrounds' discrepancies. Wang [51] and Liu [52] all believe that the industrial chain is a network chain with value-added functions that integrates backward-forward linked industries through the flow of commodities based on industrial input and output. Referring to the "point" and "line" structure of the industrial chain, Wu [53] analyzed the industrial chain from the four dimensions of supply and demand chain, value chain, enterprise chain, and space chain. Subsequently, many researchers carried out an in-depth study of the connotation of the industrial chain based on the views of Wu [54–56]. This paper draws on and extends Wu's summary of industrial chain because it covers the supply-demand relationship of the upstream and downstream links in the commodity circulation process, reflects the direction of value flow and the nature of value added, and also considers the spatial connection relationships among enterprises, the main body of production and management.

The e-commerce industrial chain is the result of the integration of the information technology industry and traditional industries relying on the e-commerce platforms, and it is also the innovation of the e-commerce model compared with the original industry [57]. Based on the dimensions of the industrial chain (see Figure 1), this paper explored the e-commerce industrial chain of Xiong'an New Area from the perspective of spatial analysis.

The supply and demand chain represents the connection between "nodes." Given that the data of e-commerce enterprises in this paper are mainly production factors, the products' technology and market demand factors are not considered. The supply chain structure is mainly analyzed.

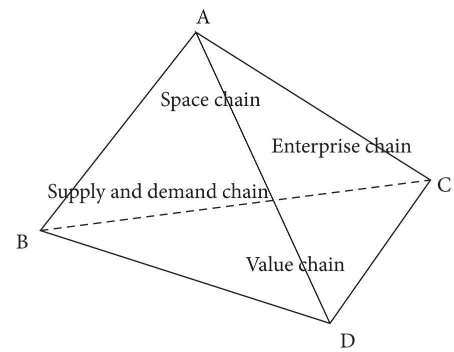


Figure 1: Four dimensions of industrial chain.

The e-commerce supply chain relies on the connection of production factors, logistics, and e-commerce platforms. It is the process of commodities from raw materials' production and processing to wholesale and retail and finally selling to consumers.

The e-commerce enterprise chain represents the connection between "points" and "lines." It is the carrier and the actual manifestation of the industrial chain. It is also the linear connection of all enterprises in different production and sales sections in the same industrial chain. The categories of e-commerce enterprises reflect the complex environment of e-commerce industry clusters. The economic size of enterprises reflects their ability to adapt to market changes and their development status in the e-commerce environment [58].

The e-commerce space chain expresses the connection between "line" and "line," which refers to the distribution of the same type of industrial chain in different regions. If it tends to be spatially concentrated, it will form a spatial agglomeration of the industrial chain, which can be divided into three levels: regional chain, national chain, and global chain.

The e-commerce value chain represents the connection between the "chain" and "chain." Porter was the first to create the value chain theory based on the competitive advantages of enterprises who believed that value chain is the embodiment of enterprise to create and deliver value [59]. Value chain plays an important role in causing the formation and change of the industrial chain. Its changes will first be characterized in the supply chain and further lead to changes in the enterprise chain and space chain. Finally, the various dimensions of the e-commerce industrial chain are intertwined and interact with each other, forming a stable and complex chain of spatiotemporal relationships under multiple factors.

2.2. The Embeddedness of e-Commerce Industrial Chains. e-commerce is inseparable not only from the production, processing, and logistic transport of commodities but also from the transmission of transaction information, capital, and technology. Therefore, the spatial organization of the e-commerce industrial chain has both cyberspace and geographic space forms. When consumers purchase products online, they are generally more inclined to choose products with strong competitive edges; that is to say, the products are unique or at good prices in that place, which leads to the spatial directivity of consumers' online shopping. As far as online sellers are concerned, in order to produce and sell products with competitive edges, they must reduce production costs. Meanwhile, the products sold in related industrial clusters are often more low cost. Information technology embeds local products in cyberspace. As a result, the distribution and agglomeration characteristics of the e-commerce industry will be different, which reflects obvious embeddedness characteristics [60].

In the e-commerce industrial chain, the online e-commerce platform is the central part. In the dimension of the e-commerce supply chain, the starting point of commodity circulation is brick-and-mortar manufacturers and wholesale retailers. Logistics provides a guarantee for the connection of the e-commerce industrial chain and the spatial flow of elements. In the meantime, e-commerce products are mainly embedded in locations, resources, industry, and economy. Location provides development space for the e-commerce industry, and element endowments such as resources provide material conditions for the e-commerce industry to put down roots. The value chain dimension is the chain that realizes the empowerment of e-commerce. Each e-commerce value link carries out similar and interrelated production and operation activities with offline industries from brand R&D and technological upgrading upstream to imports and exports downstream, realizing the continuous value added of the entire e-commerce industrial chain. At the dimension of the enterprise chain, embeddedness means the integration of online and offline departments in one enterprise. The online part is responsible for information transmission and commodity payment, and the offline part is responsible for commodity production and processing, in which social capital (such as labor resources and government orientation) is an important factor to support the sustainable development of e-commerce enterprises in the local area, and the economic output of enterprises is also a manifestation of the differences in industrial embeddedness [40]. At the dimension of the space chain, the externalities of industrial clusters promote the agglomeration of the e-commerce industry. There is also a certain coupling between the e-commerce chain and the traditional industrial chain that it depends on.

3. Materials and Methods

3.1. Study Area. Xiong'an New Area is a state-level new area under the jurisdiction of Hebei Province in China. It is located on the eastern edge of Baoding City, at the junction of Langfang City and Cangzhou City, close to Beijing and Tianjin. Xiong'an New Area consists of Xiong County, Rongcheng County, Anxin County, and some surrounding areas (including the Baiyang Lake). The area has a total of 33 towns (see Figure 2). At the end of 2019, the GDP of Xiong'an New Area reached 21.5 billion RMB yuan. Its industry is dominated by light industry, which is all labor-intensive companies. Leading industries include the paper-plastic packaging industry and latex product industry in Xiong County, clothing industry, toy industry, and luggage manufacturing industry in Rongcheng County, and shoe-making industry and feather product industry in Anxin County. For instance, the paper-plastic packaging and printing industry in Xiong County is a traditional industry that started in the early 1960s. After more than 60 years of original capital accumulation, it has developed three major packaging products (plastic package, paper package, and artificial leather package), four printing technology (intaglio printing, relief printing, screen printing, and planographic printing), and five major production areas (central towns as the comprehensive core area, the production area of paper-plastic package and balloon latex products centered on the Dabu Village Industrial Zone in Longwan Town, the printing area of plastic pipes and silk-screen color balloons with Zangang Town as the center, the artificial leather calendered film production and processing area with the Xiliu Village Industrial Zone in Zhugezhuang Town as the core, and the Dongzhao Village of Daying Town as the core circuit board printing and processing area) [61].

Before 2013, residents in Xiong'an New Area mostly produced and sold products all by themselves or resorted to OEM to operate physical stores or factories as their main source of income. Most kinds of traditional industries had similar development models. The majority of the enterprises were small dispersed individuals, mainly engaged in family workshop's production and sold with large group scale. The enterprises were facing pressure brought by low technological level, high product similarity, and fierce market competition, which caused a negative impact on the development of traditional industries. However, subsequently, under the trend of national e-commerce development, Xiong'an New Area gradually explored methods for the transformation of traditional industries to e-commerce by relying on the advantages of local special industries and leading enterprises. At present, the e-commerce market scale has expanded widely. In order to realize the industrial transformation, all counties in Xiong'an New Area have established e-commerce industrial parks and e-commerce public service centers, upgrading from scattered individuals to intensive groups. As of 2021, there are 25 Taobao villages and 8 Taobao towns in Xiong'an New Area, and the development trend of agglomeration is obvious. The integrated development of e-commerce and physical industries has gradually become an important way for the transformation and upgrading of traditional industries in Xiong'an New Area. Therefore, analyzing the spatial characteristics of the e-commerce industry and its industrial embeddedness is of great significance to the factors optimization of industry-dependent e-commerce and the sustainable development of traditional industries in the future.

3.2. Data Source. Company data from e-commerce platforms: data acquisition software was used to collect information on the homepages and archives of e-commerce enterprises in various industries on "1688.com" (https://www.1688.com/) and "Taobao.com" (https://www.taobao.com/). The "business model" of "1688.com" is the way used

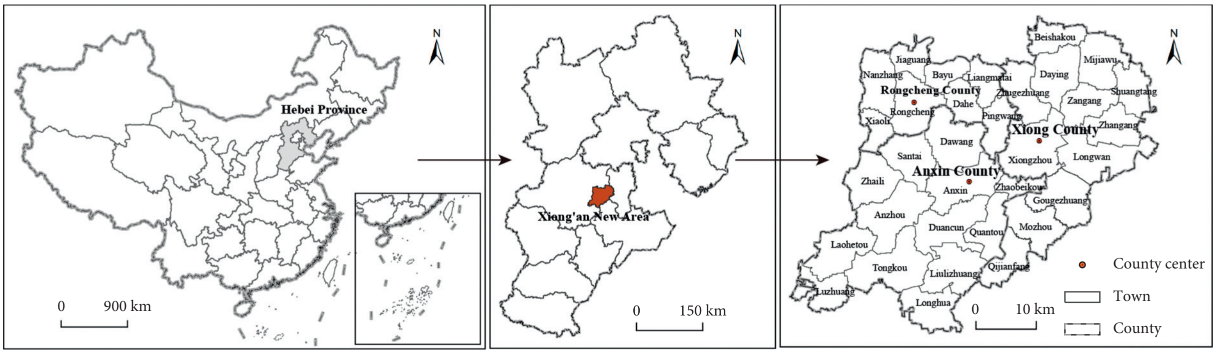


FIGURE 2: Profile of Xiong'an New Area.

by e-commerce enterprises to make profits during the operation, which is divided into four categories: production and processing, distribution and wholesale, investment agency, and business services. "Production and processing" refers to enterprises engaged in production, processing, and OEM work and then wholesale trade; "distribution and wholesale" refers to enterprises that buy and sell large quantities of commodities in the form of intermediaries for profit; the data the authors collected do not include investment agency and business services enterprises. According to the "1688.com" industry market classification, the products can be divided into 19 types, including packaging, rubber plastic chemical, craft decoration, and toys. A total of 1213 valid data were collected from "1688.com." "Commodity retail" refers to the trade activity of selling commodities or services to consumers. This type of data originates from "Taobao.com." A total of 195 valid data are collected. After data sorting, it was found that the first e-commerce company was established in 1998, and the data collection time is from December 15, 2020, to December 31, 2020, with a total of 1408 valid data. The data mainly includes the e-commerce enterprises' name, establishment time, product details, registered address, business model, type of company, annual turnover, and number of employees.

Data from Taobao Village in Xiong'an New Area: the number and location information of Taobao villages in Xiong'an New Area is collected, which is obtained from the list of China Taobao villages released by the Ali Research Institute during the years 2015–2020.

The dimensions of the industrial chain are classified based on the data of e-commerce enterprises. As for the supply chain, according to commodity circulation links, suppliers can be divided into wholesalers (production wholesalers and distribution wholesalers), retailers, and integrated suppliers, among which production wholesalers are enterprises involved in the production and manufacturing sections among wholesalers, distribution wholesalers are enterprises engaged in pure wholesale trade, and integrated suppliers are suppliers that include both wholesale and retail businesses. Value chain dimension can be divided into three categories, production and processing, distribution and wholesale, and commodity retail, according to the value-added ways of commodity circulation of e-commerce enterprises. Enterprise chain dimension can be divided into business to business (B2B), business to consumer (B2C), and business to government (B2G) according to transaction subjects. Because the enterprises are online, almost all the corresponding production and sales sections from the original industrial space can be found on the Internet [62]. Therefore, combined with the connotation of the industrial chain, the embeddedness connections between different e-commerce industrial chain links and industrial space are shown in Figure 3. Due to the lack of horizontal connection of data, this paper does not analyze the characteristics of the e-commerce space chain.

3.3. Methods

3.3.1. Location Entropy. Location entropy refers to the specialization and concentration level of an industry in a region [63]. In this paper, location entropy is used to calculate the regional orientation and dominance of e-commerce enterprises of different product categories in Xiong'an New Area. The formula used is as follows:

$$Q_j^i = \frac{N_j^i / N^i}{N_j / N}. \quad (1)$$

Here, Q_j^i is the location entropy of product j in county i, N_j^i is the number of that kind of e-commerce enterprises, N^i is the total number of the e-commerce enterprises in county i, and N_j is the total number of the e-commerce enterprises that produce product j. When $Q_j^i < 1$, e-commerce enterprises of this kind product in this region do not have obvious professional advantages compared with Xiong'an New Area as a whole. When $Q_j^i > 1$, the comparative advantages do exist, and the larger the value, the higher the degree of advantage. Combined with the result of location entropy, the local embeddedness of e-commerce enterprises can be further judged by comparing the dominant e-commerce products with the local leading industries.

3.3.2. Spatial Visualization and Spatial Analysis. Local spatial autocorrelation method is used to visualize the spatial organization of the e-commerce industrial chains in Xiong'an New Area by ArcGIS. Specifically, Getis-Ord Gi $*$ is

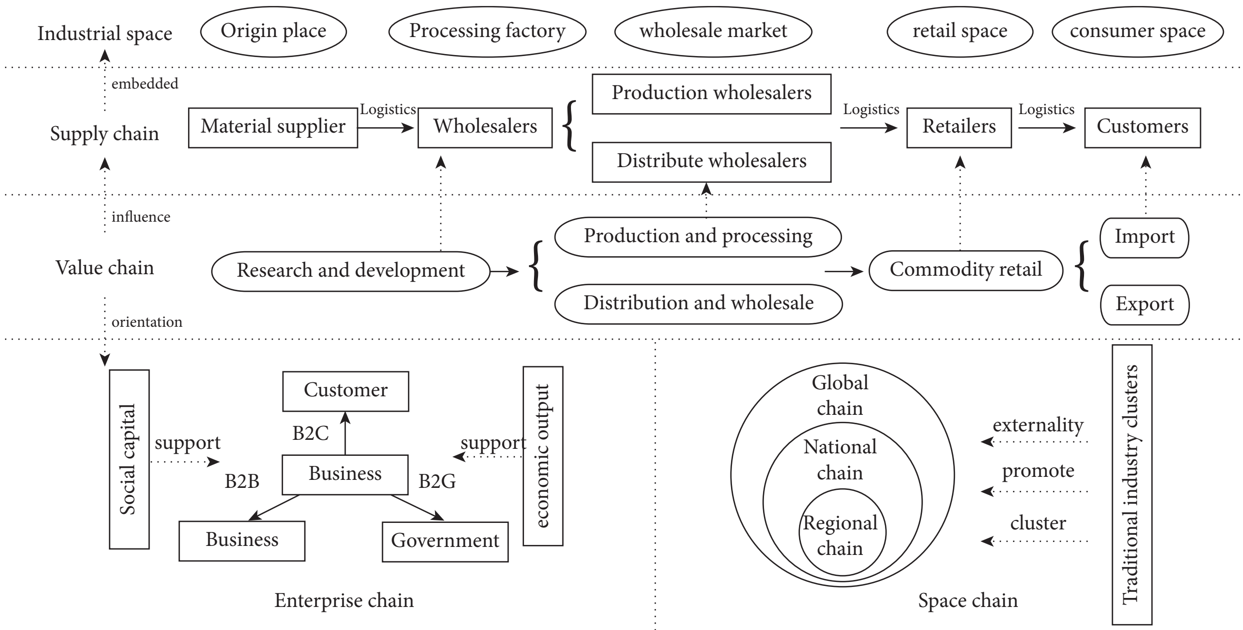


Figure 3: The embeddedness relationship between e-commerce industrial chains and industry space.

used to calculate and estimate the cold and hot spots of different types of enterprises in the e-commerce industries chain in Xiong'an New Area so as to explore its spatial organization mode. The formula is omitted. Finally, based on the spatial characteristics and combined with the actual situation of product types and sales types of e-commerce enterprises, the authors analyzed the embeddedness characteristics of e-commerce industry chains.

3.3.3. Transport Superiority. With reference to the research by Jin and others [64, 65], the degree of transport superiority is mainly composed of traffic network density, traffic line influence, and location superiority. It embodies the scale and the agglomeration capacity of transportation facilities and the accessibility status of key nodes in the region and is used to describe the advantages and disadvantages of regional traffic as an integrated indicator. The transport superiority reflects the advantages of quantity, quality, and superiority of regional traffic. Previous studies generally believe that these three aspects are of the same importance to the degree of transport superiority. In order to understand the relationship between e-commerce development and transportation location in Xiong'an New Area and further clarify the influence of transportation endowment on e-commerce, this paper revised the existing evaluation system of transport superiority, combined with the actual situation of e-commerce enterprises, and identified five secondary indexes, including the density of traffic lines in each township, the number of traffic nodes, the number of traffic lines, the shortest distance from e-commerce enterprises to the traffic line, and the distance between township administrative centers and county-level administrative centers (see Table 1). To calculate the transport superiority of Xiong'an New Area, the data in the density of traffic lines, traffic line influence on the region, and location superiority were standardized, and all dimensions were weighted uniformly. The weight is set to 1/3, and then the processed elements are integrated. Finally, the transport superiority of each township was obtained.

4. The Spatial Characteristics of e-Commerce Industrial Chain and the Influencing Factors

4.1. Spatial Distribution Characteristics of e-Commerce Industrial Chain in Different Dimensions

4.1.1. Analysis of e-Commerce Supply Chain. e-commerce products circulate through the supply chain to connect upstream and downstream related enterprises and customers. The core feature is transaction orientation and linear connection. In terms of quantity, the types of products sold by suppliers are quite different. Production wholesalers (63.92%) mostly sell packaging and rubber plastic chemical products. Most of them are raw materials or semifinished products such as plastic pellets, plastic bags, and chemical supplies. Most processing enterprises need a large amount of raw material for further working in local factories, such as pipe manufacturing, bag making, printing, brand customization, and other primary processing or deep processing activities. Comprehensive suppliers (16.55%) mainly produce clothing textiles and toys. It covers a wide range, and the companies play the role of multiple suppliers. Retailers' (13.85%) product types are mainly shoe bag accessories and craft decoration products, all of which are finished products. Distribution wholesalers' (5.68%) product types are mainly composed of rubber plastic chemicals, toys, and craft decoration products. Among them, craft decorations mainly include balloons and wreaths, and rubber plastic chemical products are mainly semifinished or finished products of plastic pipes. The processing level is relatively high, and it is

Table 1: Evaluation index system of transport superiority of Xiong'an New Area.

Indices	Indicator	Statistical method
Density of traffic lines	Traffic nodes' number	Calculation of traffic nodes' number.
Traffic line influence on the region	Traffic lines' density	The ratio of the traffic line length of each township to the administrative area of the township.
Location superiority	Traffic lines' number	Number of traffic routes covered by each town.
	Shortest distance from e-commerce enterprises to the traffic line	The natural breakpoint method by ArcGIS is used to divide the shortest distance between e-commerce enterprises and transportation lines into 5 categories. The closer the distance, the easier the transportation. Therefore, the weights of 5, 4, 3, 2, and 1 are set in order from the nearest to the farthest distance. The number of each distance level is counted, and each town's weighted value is summed.
	Distance between township administrative centers and county-level administrative centers	Calculation of Euclidean distance between the town government and the county government.

only used for transit and wholesale sales. There are not many such enterprises. From this point of view, the supply and marketing mode of the same type of commodities determines the degree of processing.

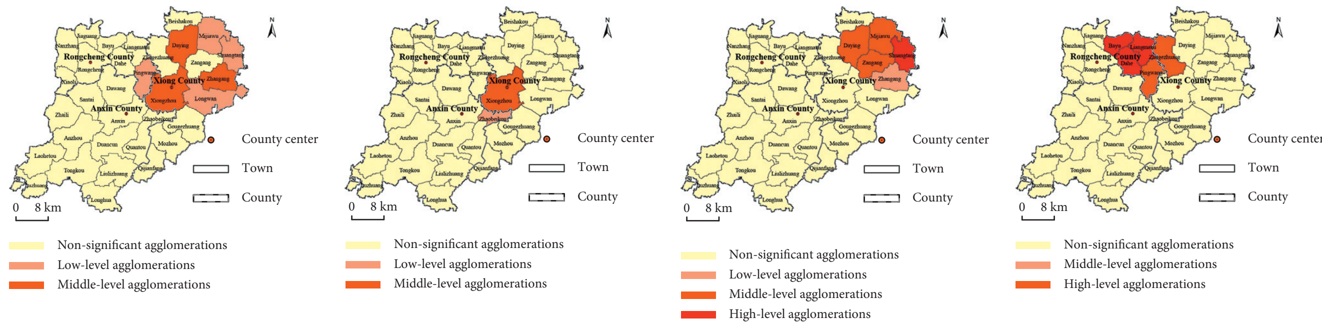
In order to explore the spatial distribution law of e-commerce enterprises' supply chain, local spatial autocorrelation analysis was used to identify the clustering characteristics of the number of suppliers. The results show that hot spots with a Z-score greater than 1.65 and a 90% confidence level ($P < 0.1$ for probability likelihood) are mainly distributed in Xiong County and Rongcheng County. Anxin County is in a large area with nonsignificant agglomeration areas; meanwhile, the overall hot spot coverage rate is low.

The hot spots for production wholesalers (in Figure 4(a)) and distribution wholesalers (in Figure 4(c)) are mainly distributed in towns of Xiong County. The product types in the hot spots are mainly rubber plastic chemical and packaging, respectively. Self-produced products and dealers' products are mostly from the local village or nearby rural enterprises. These two types of wholesalers are usually close to the industrial clusters in order to minimize the cost between production and sales. In the meantime, their hot spots just cover the five major production areas of Xiong County's paper-plastic packaging and printing industry. The hot spots for retailers (in Figure 4(b)) are in Xiongzhou Town and Zhaobeikou Town, with the smallest middle-level clusters, in which the products are mainly packaging. The physical size of retailers is relatively small. In order to obtain price advantages, they retail through specialized markets nearby or out of town. As for comprehensive suppliers (in Figure 4(d)), zones with $Z > 2.98$ are mostly concentrated in the east of Rongcheng and mainly based on the toy products of the local characteristic industry. The cluttering zone not only includes specialized retail markets but also has a large-scale and highly specialized toy industry cluster.

Rongcheng County's e-commerce supply chain is relatively comprehensive. Xiong County's wholesalers and retailers are more concentrated. Anxin County's supply chain is not yet mature, and characteristic industries still have great development potential. The supply chain mainly connects various suppliers with the flow of commodities. Most of the traditional industrial clusters in Xiong'an New Area have a history of decades of development. The most obvious competitive advantage stems from the low cost of commodity manufacturing, and the tradition, culture, and invisible knowledge and experience of local business have also become the foundation of e-commerce. Therefore, in general, the product types in the hot spots are based on local traditional industrial clusters, and e-commerce is embedded on the basis of geographical proximity, sharing economic resources such as production materials and labor with traditional industries [66].

4.1.2. Analysis of e-Commerce Value Chain. The value-added process of e-commerce products in circulation is shown in Figure 3. The value-added activities of the e-commerce value chain in Xiong'an New Area are mainly production and processing. This type of enterprise accounts for 80.04% of the total, which has a great advantage in terms of quantity, followed by commodity retail enterprises, accounting for 13.85%. The number of distribution and wholesale enterprises is very small, only accounting for 6.11%.

The value chain distribution and spatial orientation of different product types are shown in Table 2. To avoid the lack of statistical significance, Table 2 only counted the top 10 product types with the largest number of e-commerce enterprises included. The components of the value chain of all types of products are roughly the same, with production and processing as the main focus and distribution and wholesale as the least. Except for toys and rubber plastic chemical products, which have more distribution and wholesale enterprises than retailers, the rest of the enterprises are mainly concentrated in the production and processing and retail links, that is, the upstream and downstream of the e-commerce industrial chain, leading to the "polarization" characteristic of the chain. Now that mixed operations have become the mainstream, high-intensity links in the value chain, such as production and processing (production and wholesale) and retail (export and import), can bring more enterprise profits [3], which causes less pure wholesale trade.



(a) (b) (c) (d)

FIGURE 4: Hot spot distribution of supply chain e-commerce enterprises in Xiong'an New Area. (a) Production wholesalers. (b) Retailers. (c) Distribution wholesalers. (d) Comprehensive suppliers.

TABLE 2: Value chain composition and main distribution areas of major e-commerce product types in Xiong'an New Area.

Type of product	Number of enterprises and spatial orientation	Production and processing	Commodity retail	Distribution and wholesale
Packaging	Number	221	41	2
	Orientation	Xiongzhou Town	Xiongzhou Town	Xiongzhou Town
Rubber plastic chemicals	Number	228	14	20
	Orientation	Zangang Town	Xiongzhou Town	Zangang Town
Craft decorations	Number	175	51	14
	Orientation	Longwan Town	Longwan Town	Mijiawu Town
Toys	Number	124	5	13
	Orientation	Liangmatai Town	Rongcheng Town	Liangmatai Town
Shoe bag accessories	Number	108	24	10
	Orientation	Zhugezhuang Town	Daying Town	Santai Town
Clothing textiles	Number	116	9	7
	Orientation	Rongcheng Town and Anxin Town	Rongcheng Town	Liangmatai Town and Longhua Town
Household supplies	Number	61	7	7
	Orientation	Luzhuang Town	Xiongzhou Town	Anxin Town
Machinery and equipment	Number	20	9	2
	Orientation	Xiongzhou Town	Rongcheng Town	Zhugezhuang Town and Liulizhuang Town
Hardware tools	Number	18	6	4
	Orientation	Pingwang Town	Xiongzhou Town and Jiaguang Town	Rongcheng Town
Automotive supplies	Number	14	11	2
	Orientation	Pingwang Town	Xiongzhou Town and Rongcheng Town	Rongcheng Town and Daying Town

In terms of spatial distribution, the overall value chain tends to be centered on Xiongzhou Town, Zangang Town, Longwan Town in Xiong County and Liangmatai Town and Rongcheng Town in Rongcheng County. The development of an industry not only is based on resource endowments but also depends on the regional preference of its industrial functions. The value-added business model will be more inclined to areas with strong industrial functions and then realize the specialization and scale operation of this type of e-commerce product. For example, although the traditional industries of shoe bag accessories and clothing textiles are leading industries in Anxin County, the production processing and retail links are concentrated in Zhugezhuang Town, Daying Town, and Rongcheng Town in Rongcheng County. Locally, labor-intensive manufacturing activities are highly specialized; the production and processing and distribution and wholesale links of household supplies enterprises are concentrated in Luzhuang Town and Anxin Town, Anxin County. Originally, this type of product had the greatest advantage in the traditional industries of Anxin County. However, its e-commerce retail links point to Xiongzhou Town, Xiong County, which has a higher economic level and richer market resources. The distribution of the value chain of the above categories of products has shifted from its advantageous regions, which not only reflects the social division of labor of the same type of products in different regions but also proves the impact of the specialization of regional industrial functions on the value chain [67].

4.1.3. Analysis of e-Commerce Enterprise Chain. Divided by transaction entities, B2B enterprises are the main component of the Xiong'an New Area enterprise chain (74.29%), followed by B2C (13.85%), and B2G enterprises (11.86%) are the least, which is in line with the scale of national e-commerce [68]. B2B enterprises are mainly production wholesalers (72.37%), which is consistent with the characteristic that the traditional industries in Xiong'an New Area are basically labor-intensive manufacturing. They produce and sell finished products and semifinished products, such as plastic bags, paper packaging bags, balloons, and funeral supplies. Comprehensive suppliers (21.13%) are the second, which mainly produce toys and clothing. Distribution and wholesalers (6.50%) are the least, mainly to wholesale downstream enterprises toys, clothing, plastic pipes and balloons, and other completed products. B2C enterprises are retailers, mainly selling plastic bags, cartons, balloons, funeral supplies, and other finished products to customers. B2G enterprises are also mainly production wholesalers (85.62%), aiming to serve the government's municipal works, water conservancy, and building works by manufacturing and selling plastic pipes. Distribution and wholesalers (7.19%) and integrated suppliers (7.19%) are equal, mainly supplying plastic pipes and shoe bags.

During the rapid expansion of e-commerce in space, the scale of enterprises is dominated by small, medium, and microenterprises, among which all B2C enterprises are this kind of enterprise. e-commerce enterprises with 11–50 employees have become the largest type (41.55%) in the enterprise chain of Xiong'an New Area. A large number of small-scale e-commerce enterprises means that the e-commerce market is sinking and refined, and at the same time, it also reflects the strong driving force of e-commerce to transform traditional small and microenterprises. e-commerce enterprises with an annual turnover of 7–10 million yuan accounted for the highest proportion (14.91%), mostly pointing to areas with industrial foundations such as Xiongzhou Town, Longwan Town, and Zangang Town in Xiong County. e-commerce enterprises with an annual turnover exceeding 100 million yuan are also concentrated in Zangang Town, Xiong County, which shows that the externalities of the traditional industrial economy have in turn enhanced the vitality of e-commerce development.

Using the hot spot analysis method to explore the spatial correlation patterns of various e-commerce companies, this paper observed the spatial clustering diagram of the enterprise chains in Xiong'an New Area. On the whole, hot spots with a Z-score greater than 1.96 and a 95% confidence level ($P < 0.05$ for probability likelihood) of B2B, B2C, and B2G are basically concentrated in Xiong County and have strong continuity, while nonaggregating areas and cold spot areas are concentrated in Anxin County and Rongcheng County. The hot spots for B2B enterprises (in Figure 5(a)) are gathered in Pingwang Township, Rongcheng County, Xiongzhou Town, and Longwan Town, Xiong County, showing a linear distribution. The cold spot area is located in Mozhou Town of Anxin County. B2C enterprises hot spots (in Figure 5(b)) are located in Xiongzhou Town, Xiong County, and Zhaobeikou Town, Anxin County, with a small distribution area, and the rest are nonsignificant areas. As for B2G enterprises, areas with $Z > 3$ (in Figure 5(c)) are concentrated in Zhanggang Town, Shuangtang Town, Zangang Town, Mijiawu Town, and Daying Town, Xiong County, covering the Xiongdong area, presenting a block distribution. Affected by the construction of the new area, e-commerce enterprises' aggregation provides a large amount of materials needed for urban infrastructure.

The number of hot spots of B2G companies is significantly higher than that of B2B and B2C enterprises, indicating that B2G companies have a higher degree of agglomeration. It can be seen that B2B, B2C, and B2G enterprises have significant spatial agglomeration characteristics. Xiong County has always been a high-level cluster of various enterprises, followed by Rongcheng County, and Anxin County has not seen an obvious spatial agglomeration. This high-level agglomeration is embedded in the strong industrial foundation of Xiong County, and the base of e-commerce scale is relatively large, so it has an absolute scale advantage provided by the e-commercialization of traditional industry [69]. e-commerce enterprises in low-level agglomeration areas are sparse, and the development of traditional industries is relatively weak. The main products of local industries are not prominent. It is necessary to promote the transformation of traditional industries by improving the level of e-commerce development or developing "blue ocean" products.

4.2. Influencing Factors of the Spatial Distribution of e-Commerce Industrial Chain

4.2.1. Transport Superiority Promotes the Agglomeration of e-Commerce Industry. Although e-commerce can cross the constraints of spatial distance, its development is still based on regional embeddedness in reality, and the transportation and logistics foundation has profoundly affected the agglomeration level of the regional e-commerce industry [22]. Visualizing and spatially superimposing the e-commerce enterprises in the Xiong'an New Area with highways and railway traffic lines shows that there is a very high correlation between the two layouts. On the whole, the northern part of Xiong'an New Area, namely, Xiong and Rongcheng County, has more and dense traffic routes and hubs than Anxin County and has a higher traffic location advantage. There are obvious clusters and strips of e-commerce enterprises around the traffic lines. This feature is extremely prominent in Xiongzhou Town and Rongcheng Town. Figure 6 shows that the distribution of e-commerce enterprises and transport superiority in Xiong'an New Area is basically the same. Pearson correlation analysis and verification are carried out on the number of e-commerce enterprises and the degree of transport superiority. The correlation coefficient is 0.804, which means that it has a strong correlation. The significance level is 0.00 ($p < 0.01$). It has an extremely significant statistical effect; that is, the greater the degree of transport superiority, the greater the number of e-commerce enterprises in the town.

Figure 5: Hot spot distribution of enterprises chain e-commerce enterprises in Xiong'an New Area. (a) B2B enterprises hot spots. (b) B2C enterprises hot spots. (c) B2G enterprises hot spots.

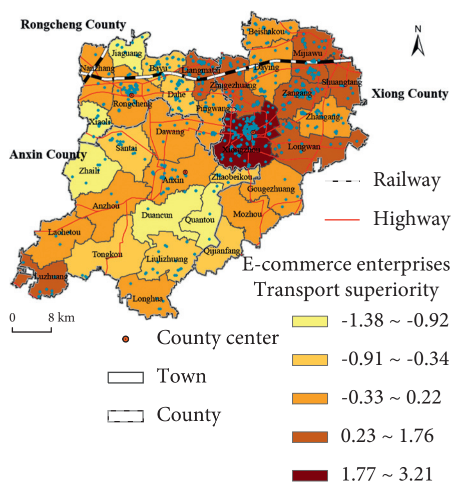


Figure 6: Distribution of transport superiority in Xiong'an New Area.

4.2.2. Traditional Industries Support the Diffusion of the e-Commerce Industrial Chains. With the popularization of e-commerce, a large number of traditional industries in Xiong'an New Area have formed the sales network. Relying on the guarantee of the government and major e-commerce platforms, the originally relatively scattered but highly local advantage industries and products have begun to connect to the Internet, and they gradually turned into an online e-commerce industry cluster [58]. Use the location entropy method to explore the correlation between the location orientation of e-commerce companies and local industrial clusters (see Table 3).

It can be seen that the distribution of location entropy of the e-commerce products in Xiong'an New Area has a strong correlation with its leading industries. The industrial chain of packaging, rubber plastic chemical, craft decoration, electrical electronics, and medical supplies has significant advantages in Xiong County. Toys, machinery and equipment, hardware tools, automotive supplies, office supplies, and pets supplies tend to be distributed in Rongcheng County. Shoe bag accessories, clothing and textiles, household supplies, food and beverages, home building materials, gardening supplies, sports and fitness, and metal processing products are more specialized in Anxin County. The paper-plastic packaging industry is one of the leading industries in Xiong County. The location entropy of corresponding e-commerce enterprises such as packaging (plastic bags, paper bags, etc.), rubber plastic chemicals (pipes, pipe fittings, and other plastic material products), and craft decorations (latex balloons and accessories) is relatively large, and the e-commerce industrial chain is concentrated. One of the leading industries in Rongcheng County is the toy industry. The location entropy of the corresponding e-commerce products is as high as 4.35. There are 137 enterprises, accounting for 43.91% of the total number of e-commerce enterprises in Rongcheng, which indicates that the e-commerce industrial chain involves many links. The leading industries in Anxin County include the shoemaking industry and feather product industry, the corresponding e-commerce products are shoe bag accessories (shoes, soles, and other products) and clothing textiles (down products), and this category of e-commerce enterprises has the largest location entropy.

Table 3: Location entropy distribution of different types of e-commerce products in Xiong'an New Area.

Type of commodity	Xiong County	Rongcheng County	Anxin County
Packaging	**1.33**	0.39	0.32
Rubber plastic chemicals	**1.51**	0.02	0
Craft decorations	**1.47**	0.06	0.14
Toys	0.03	**4.35**	0.12
Shoe bag accessories	0.99	0.48	**2.02**
Clothing textiles	0.21	2.50	**2.62**
Household supplies	0.48	0.42	**4.95**
Machinery and equipment	0.68	**1.75**	1.36
Hardware tools	0.70	**1.93**	0.90
Automotive supplies	0.62	**2.67**	0
Food and beverages	0.45	1.06	**4.96**
Home building materials	0.80	0.95	**1.33**
Gardening supplies	0.51	0.38	**4.92**
Electrical electronics	**1.52**	0	0
Sports and fitness	0.38	1.13	**4.22**
Office supplies	0.51	**3.01**	0
Metal processing	0	0	**8.43**
Medical supplies	**1.52**	0	0
Pets supplies	0	**4.51**	0

The bold values are the highest location entropy value of each type of product, which means the highest regional specialization degree.

4.2.3. Government Policy Guides the Development of e-Commerce Industrial Chains. The agglomeration and development of the e-commerce industrial chains in Xiong'an New Area cannot be separated from the adjustment and policy guidance of governments. In recent years, Hebei Province has successively issued "Implementation Opinions on Promoting Full Coverage of Rural e-commerce," "Three-Year Promotion Plan for e-commerce Development in Hebei Province (2014–2016)," "Action Plan for Accelerating e-Commerce Development in Hebei Province (2018–2020)," and other policy documents. In 2020, the State Council approved the establishment of a national cross-border e-commerce comprehensive pilot zone in Xiong'an New Area. In 2017, the Xiong'an New Area Management Committee began to study the ideas for the construction of "digital Xiong'an" based on big data, which will promote the innovative development of e-commerce transformation and new retail. In 2018, the Xiong County Bureau of Commerce signed an e-commerce talent training strategy with Taobao University to promote the transformation and upgrading of Xiong County's traditional industries to apply "Internet +." In 2019, Rongcheng County plush toys association achieved strategic cooperation with Quyang County to jointly build an e-commerce training base. In 2020, the country suffered the impact of the COVID-19 epidemic, and countless offline physical industries turned to online. Especially with the help of the government, Rongcheng County formulated the "Nine Measures to Support the Transformation and Upgrading of Traditional Industries in Rongcheng County (Temporary)" to strongly support the local enterprises such as Jihongxing Apparel and Aosen Apparel. These companies transition to the production of antiepidemic materials and live-streaming e-commerce to fill the gaps in the New Area. In general, the Xiong'an New Area will focus on key urban areas and town centers, and policies will internally support the development of e-commerce enterprises, accelerate the cultivation of specialized villages, and plan and deploy e-commerce from point to point. e-commerce industry collaboration relationships across administrative regions should be established while strengthening front-to-back industry linkages through Internet technology and platforms. Finally, the e-commerce industry will realize the upgrade of the industrial chain.

4.2.4. Influence Mechanism. The influence mechanism of the e-commerce industry chain development in Xiong'an New Area is shown in Figure 7. First of all, the vast majority of e-commerce enterprises in Xiong'an New Area sell tangible products, so the products' flow path cannot be separated from the production processing and logistics transportation. In order to produce and sell goods with cost advantages, the location of e-commerce enterprises must be close to factories and markets to reduce transportation costs, which will promote the integrity and efficiency of the e-commerce supply chain. Therefore, the advantage of transportation and location is the primary condition and the basis for the formation and agglomeration of the e-commerce industry. Further, in order to reduce production costs to a greater extent and increase the competitive advantage of commodities, online and offline industries will replicate and migrate the value chain, and the layout of e-commerce enterprises will point to areas with relevant industrial clusters. Then, the advantages of traditional industries promote the diffusion of the e-commerce industry, making the e-commerce in Xiong'an New Area form an industry-dependent development path. Finally, the guidance and support capabilities of government policies support the coordinated development of multiple places and institutions from a macroperspective and ensure the development of e-commerce and infrastructure construction within the county on a microlevel. The government measures provide

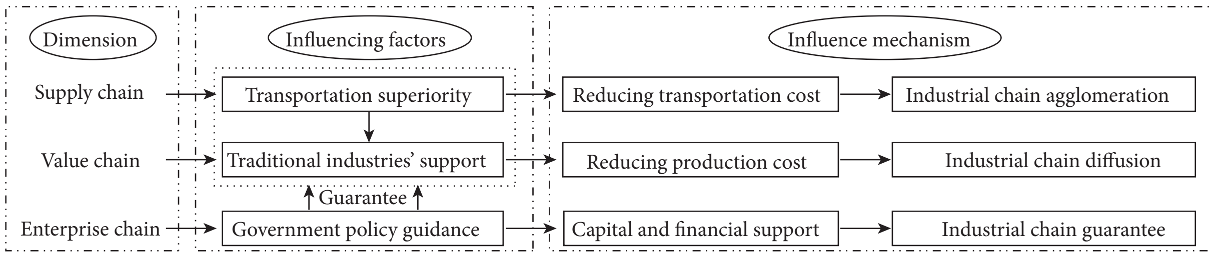


Figure 7: Influence mechanism.

social capital and economic support for the e-commerce enterprise chain and also provide a suitable institutional environment for the industrial chain.

5. Conclusion and Discussion

Based on 1408 e-commerce enterprises' data in Xiong'an New Area, this paper revealed the spatial characteristics and influencing factors of e-commerce industrial chains and discussed the characteristics of industrial embeddedness based on various aspects of the e-commerce industrial chain.

The e-commerce supply chain in Xiong'an New Area is dominated by production wholesalers, followed by integrated suppliers, retailers, and distribution wholesalers. Rongcheng County's e-commerce supply chain is more comprehensive, Xiong County's wholesalers and retailers are more concentrated, and Anxin County's supply agglomeration is the weakest. Commodities can characterize the industrial embeddedness of the supply chain, and supply links' differences can cause discrepancies in the degree of product completion. The cost advantage of product manufacturing makes traditional industries the basis for the development of e-commerce.

The value chain dimension is dominated by production and processing activities, followed by commodity retailing and distribution and wholesale. Value links spatially inclined to Xiongzhou Town, Zangang Town, and Longwan Town in Xiong County and Liangmatai Town and Rongcheng Town in Rongcheng County. The intensity of value information promotes the agglomeration of the value chain sections, while the production and sales links deviate from the industrial advantage regions. This cross-township and even cross-regional geographical division of labor structure reflect the impact of the specialization of regional industrial functions on the value chain.

At the enterprise chain, B2B enterprises are the majority, followed by B2C enterprises and B2G enterprises. The scale of employees and the annual turnover of all kinds of enterprises have an obvious trend of miniaturization. Local industrial clusters bring economic vitality to e-commerce enterprises. Spatially, the high-level clusters of the three types of e-commerce companies are all located in Xiong County. The agglomeration of Rongcheng County is the second, and Anxin County has no obvious agglomeration. The high-level agglomeration stems from the strong traditional industrial foundation, which brings scale advantages for the transformation of e-commerce.

Exploring the factors influencing the spatial distribution of the e-commerce industrial chains in Xiong'an New Area, it is found that transportation location advantages are the basis for e-commerce enterprises and their chains to agglomerate and also provide conditions for e-commerce supply chains to be rooted in traditional industries. Relying on traditional industries, different types of e-commerce commodities have achieved large-scale operations in their industrial advantage regions. Government policy regulation provides a guarantee for the agglomeration and expansion of the e-commerce industrial chains.

The study found that the supply chain of industry-dependent e-commerce still relies on the commodity circulation path of traditional industries, especially the e-commerce derived from the strong industrial foundation of Xiong'an New Area, whose products have more local embeddedness. This phenomenon indicates that traditional geographical factors are still very important in cyberspace [60, 70, 71]. Under the embeddedness effect of e-commerce, traditional industries will not be completely eliminated. Enterprises that already have the scale of the supply chain, are rich in value links, and have vigorous enterprises are more likely to survive in the future development of Xiong'an New Area. Based on the above conclusions, with the help of a multidimensional industrial chain formed by the connection of e-commerce enterprises in Xiong'an New Area, the local government should pay special attention to the improvement of disadvantaged areas and use e-commerce advantages and positive externality of traditional industries to radiate to underdeveloped areas in order to drive the application of e-commerce by business entities in Anxin County. On the other hand, focusing on improving the quality and efficiency of e-commerce in areas where the e-commerce industrial chain is concentrated and under the premise of green development, Xiong'an New Area should promote the digitalization and intelligentization of industries to tackle the "low-locked" and element loss problem of traditional industries with low added value, weak R&D capabilities. Then thereby, the elimination rate of enterprises in the industrial clusters of Xiong County and Rongcheng County could be reduced. Finally, the embeddedness characteristics of the e-commerce industrial chain studied in this paper can be used as proof to analyze the interaction between e-commerce and traditional industries, focusing on the adjustment of the traditional industrial structure under the influence of e-commerce, provide a theoretical reference for the entry and exit of enterprises, and point out the direction of the sustainable development of traditional industries.

Previous research always makes further discussion about the transformation and reconstruction of physical space caused by the embedding of e-commerce into local economic space after analyzing the e-commerce characteristics and influencing factors. This paper mainly revealed how e-commerce was embedded in the local industry and area from a different perspective of industrial chains and discussed the survival and development of traditional industries. Limited by data, the research in this paper is limited to the qualitative discussion of the embeddedness characteristics of the e-commerce industrial chain. In fact, based on the embeddedness connotation, it should also cover social, cultural, political, and other aspects. It needs to be further investigated through field research and data collection. In short, the research on the spatiality and embeddedness of the e-commerce industrial chain plays an important role in the transformation and upgrading of traditional industries, as well as the integration and development of the digital economy and the real economy, and more empirical cases are urgently needed for systematic research on it.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (no. 41671127) and the Hebei Province Soft Science Research Project of China (no. 21557624D).

References

[1] T. Garín-Muñoz, R. López, T. Pérez-Amaral, I. Herguera, and A. Valarezo, "Models for individual adoption of eCommerce, eBanking and eGovernment in Spain," *Telecommunications Policy*, vol. 43, no. 1, pp. 100–111, 2019.

[2] Á. Valarezo, T. Pérez-Amaral, T. Garín-Muñoz, I. Herguera García, and R. López, "Drivers and barriers to cross-border e-commerce: evidence from Spanish individual behavior," *Telecommunications Policy*, vol. 42, no. 6, pp. 464–473, 2018.

[3] G. Saridakis, Y. Lai, A.-M. Mohammed, and J. M. Hansen, "Industry characteristics, stages of E-commerce communications, and entrepreneurs and SMEs revenue growth," *Technological Forecasting and Social Change*, vol. 128, pp. 56–66, 2018.

[4] N. Wrigley and A. Currah, "Globalizing retail and the 'new e-conomy': the organizational challenge of e-commerce for the retail TNCs," *Geoforum*, vol. 37, no. 3, pp. 340–351, 2006.

[5] D. Zhang, P. Zhu, and Y. Ye, "The effects of E-commerce on the demand for commercial real estate," *Cities*, vol. 51, pp. 106–120, 2016.

[6] G. Clarke, C. Thompson, and M. Birkin, "The emerging geography of e-commerce in British retailing," *Regional Studies, Regional Science*, vol. 2, no. 1, pp. 371–391, 2015.

[7] A. H. Mohamad, G. F. Hassan, and A. S. A. Elrahman, "Impacts of e-commerce on planning and designing commercial activities centers: a developed approach," *Ain Shams Engineering Journal*, vol. 13, no. 4, Article ID 101634, 2022.

[8] B. Parker and R. Weber, "Second-hand spaces: restructuring retail geographies in an era of e-commerce," *Urban Geography*, vol. 34, no. 8, pp. 1096–1118, 2013.

[9] J. H. Ding, Z. Y. Song, and W. D. Liu, "An empirical study on the spatial organization variation with the application of Information and Communication Technologies(ICTs): a case study of textile and apparel enterprises in China," *Geographical Research*, vol. 28, no. 4, pp. 883–892, 2009.

[10] J. Zhao and C. Wang, "Size distribution of credit rating of C2C clothing stores and its influencing factors: a case study of Jiangsu, Zhejiang, and Shanghai," *Tropical Geography*, vol. 39, no. 5, pp. 790–798, 2019.

[11] Y. J. Liu and X. M. Chu, "The industrial evolution of TAOBAO villages," *China Soft Science*, vol. 2, pp. 29–36, 2017.

[12] Z. Xue and X. H. Geng, "E-commerce platform, acquaintance society and rural distinctive industrial clusters--case study of 'Taobao Village' in Shaji Town," *Journal of Northwest A&F University(Social Science Edition)*, vol. 18, no. 5, pp. 46–54, 2018.

[13] X. Z. Li and R. Huang, "Research on the competitiveness of China's textile industry in the context of industry 4.0 based on the integration of textile industry and electronic information industry," *China Soft Science*, vol. 2, pp. 21–31, 2018.

[14] Z. R. Zhang and J. D. Yang, "How to optimize the path of industrial goods downlink by rural e-commerce from the perspective of rural revitalization: based on grounded theory research of value chains coupling mechanism," *Issues in Agricultural Economy*, vol. 4, pp. 118–129, 2019.

[15] Y. N. Zhang, H. L. Long, S. S. Tu, Y. R. Li, L. Ma, and D. Z. Ge, "A multidimensional analysis of rural restructuring driven by e-commerce: a case of Xiaying village in central China," *Scientia Geographica Sinica*, vol. 39, no. 6, pp. 947–956, 2019.

[16] J. Q. Zhang and Q. L. Qian, "Research of the spatial transformation of 'taobao village' in information age," *Urban Development Studies*, vol. 22, no. 10, pp. 81–101, 2015.

[17] J. Zhou, L. Yu, and C. L. Choguill, "Co-evolution of technology and rural society: the blossoming of Taobao villages in the information era, China," *Journal of Rural Studies*, vol. 83, pp. 81–87, 2021.

[18] C. C. Wang, J. T. Miao, N. A. Phelps, and J. Zhang, "E-commerce and the transformation of the rural: the Taobao village phenomenon in Zhejiang Province, China," *Journal of Rural Studies*, vol. 81, pp. 159–169, 2020.

[19] F. F. Chen, Z. D. Luo, and H. M. He, "Study on the pluralistic reconstitution of rural governance driven by e-commerce: a case study of Daji town of Cao County, Shandong Province," *Modern Urban Research*, vol. 10, pp. 22–29, 2016.

[20] Z. W. Ding, M. L. Han, G. S. Zhang, and Z. H. Jian, "Spatial differentiation and influencing factors of the service quality of Taobao online C2C stores in central plains urban agglomeration at county level," *Economic Geography*, vol. 39, no. 5, pp. 143–154, 2019.

[21] J. G. Yu, L. H. Wang, and N. Li, "E-shops spatial distribution rule: a case study of Taobao net," *Economic Geography*, vol. 30, no. 8, pp. 1248–1253, 2010.

[22] M. F. Wang and S. Lu, "Exploring the spatial organization of online retailing: a case study of Dangdang. com," *Geographical Research*, vol. 30, no. 6, pp. 965–976, 2011.

[23] Z. Lu, X. N. Li, L. H. Yang, D. Yang, and L. L. Deng, "Location orientation of E-shops in China's major cities based on neighborhood facilities: case studies of Shanghai, Shenzhen,

Tianjin and Beijing," *Acta Geographica Sinica*, vol. 66, no. 6, pp. 813–820, 2011.
[24] G. F. Gu and Y. H. Xu, "Influencing factors and spatial pattern of electronic commerce development level in China at the prefecture-level city," *Economic Geography*, vol. 39, no. 10, pp. 123–129, 2019.
[25] Y. W. Zeng, D. M. Qiu, Y. T. Shen, and H. D. Guo, "Study on the formation of taobao village: taking dongfeng village and junpu village as examples," *Economic Geography*, vol. 35, no. 12, pp. 90–97, 2015.
[26] K. B. Shi, Y. C. Yang, S. Bai, E. L. Li, and D. Chen, "Innovation diffusion hypothesis or efficiency hypothesis: spatial penetration of online-to-offline e-commerce in China based on Meituan.com," *Geographical Research*, vol. 37, no. 4, pp. 783–796, 2018.
[27] G. Yeung and K. L. Ang, "Online fashion retailing and retail geography: the blogshop phenomenon in Singapore," *Tijdschrift voor Economische en Sociale Geografie*, vol. 107, no. 1, pp. 81–99, 2016.
[28] M. F. Wang, S. Lu, and J. Qiu, "Exploring the spatial organization of online retailing: a case study of Dangdang•com," *Geographical Research*, vol. 30, no. 11, pp. 1835–1840, 2010.
[29] X. Liu, F. Zhen, M. Zhang, and G. L. Xi, "Research review of online shopping impact on personal travel and urban retail space and implications," *Progress in Geography*, vol. 34, no. 1, pp. 48–54, 2015.
[30] W. Tang and J. Zhu, "Informality and rural industry: rethinking the impacts of E-Commerce on rural development in China," *Journal of Rural Studies*, vol. 75, pp. 20–29, 2020.
[31] J. Qian and Z. F. Zheng, "E-commerce clusters based on taobao industrial chain: a case study on Qingyanliu Village in Yiwu City," *City Planning Review*, vol. 11, pp. 79–83, 2013.
[32] C. S. Zhu, X. Wang, D. Wang, T. Zhang, and W. J. Zhong, "Spatial connection of Taobao Village in Haining based on the operation process," *Shanghai Urban Planning Review*, vol. 2, pp. 36–41, 2021.
[33] Z. P. Xiao, Q. Yuan, Y. H. Sun, and X. X. Sun, "New paradigm of logistics space reorganization: E-commerce, land use, and supply chain management," *Transportation Research Interdisciplinary Perspectives*, vol. 9, Article ID 100300, 2021.
[34] G. H. Gessner and C. R. Snodgrass, "Designing e-commerce cross-border distribution networks for small and medium-size enterprises incorporating Canadian and U.S. trade incentive programs," *Research in Transportation Business & Management*, vol. 16, no. 16, pp. 84–94, 2015.
[35] Y. Xu, H. Gui, J. Zhang, and Y. Wei, "Supply chain analysis of cross border importing e-commerce considering with bonded warehouse and direct mailing," *Sustainability*, vol. 11, no. 7, pp. 1909–1921, 2019.
[36] B. Moriset, "e-Business and e-Commerce," *International Encyclopedia of Human Geography*, Elsevier, Amsterdam, Netherlands, Second edition, 2020.
[37] M. Liu, Q. Zhang, S. Gao, and J. Huang, "The spatial aggregation of rural e-commerce in China: an empirical investigation into Taobao Villages," *Journal of Rural Studies*, vol. 80, pp. 403–417, 2020.
[38] M. Hess, "'Spatial' relationships? Towards a reconceptualization of embedded ness," *Progress in Human Geography*, vol. 28, no. 2, pp. 165–186, 2004.
[39] A. Jones, "Beyond embeddedness: economic practices and the invisible dimensions of transnational business activity," *Progress in Human Geography*, vol. 32, no. 1, pp. 71–88, 2008.
[40] X. D. Fu and J. S. Fu, "The theory origin and enlightement of leading industrial embeddedness," *Regional Economic Review*, vol. 1, pp. 26–32, 2017.
[41] S. Burt, U. Johansson, and J. Dawson, "International retailing as embedded business models," *Journal of Economic Geography*, vol. 16, no. 3, pp. 715–747, 2016.
[42] S. Wood, N. M. Coe, and N. Wrigley, "Multi-scalar localization and capability transference: exploring embeddedness in the Asian retail expansion of Tesco," *Regional Studies*, vol. 50, no. 3, pp. 475–495, 2016.
[43] N. M. Coe and Y.-S. Lee, "'We've learnt how to be local': the deepening territorial embeddedness of Samsung-Tesco in South Korea," *Journal of Economic Geography*, vol. 13, no. 2, pp. 327–356, 2013.
[44] H. Y. Zhu, Z. X. Zhu, and L. Zhang, "Theoretical framework of the primary features of industrial agglomeration and regional development," *Economic Geography*, vol. 38, no. 10, pp. 111–117, 2018.
[45] A. Appel, "Embeddedness and the (Re)making of retail space in the realm of multichannel retailing: the case of migros sanal market in Turkey," *Geografiska Annaler Series B Human Geography*, vol. 98, no. 1, pp. 55–69, 2016.
[46] A. O. Hirschman, *The Strategy of Economic Development*, Yale University Press, New Haven, CT, USA, 1958.
[47] A. Flynn, K. W. Chan, Z. H. Zhu, and L. Yu, "Sustainability, space and supply chains: the role of bamboo in Anji County, China," *Journal of Rural Studies*, vol. 49, pp. 128–139, 2017.
[48] R. S. Piao, L. Fonseca, E. D. C. Januário, M. S. Macchione Saes, and L. Florencio de Almeida, "The adoption of voluntary sustainability standards (VSS) and value chain upgrading in the Brazilian coffee production context," *Journal of Rural Studies*, vol. 71, pp. 13–22, 2019.
[49] T. Styliadis and C. Chlomoudis, "Analyzing the evolution of concentration within containerized transport chains through a circuitist approach: the role of innovations in accelerating the circuits of liner and container terminal operators," *The Asian Journal of Shipping and Logistics*, vol. 37, no. 4, pp. 321–328, 2021.
[50] J. B. Huang, Q. Ding, Y. Wang, H. J. Hong, and H. W. Zhang, "The evolution and influencing factors of international tungsten competition from the industrial chain perspective," *Resources Policy*, vol. 73, Article ID 102185, 2021.
[51] J. C. Wang, "Some related concepts in China's research of industry cluster," *Acta Geographica Sinica*, vol. 59, no. S1, pp. 47–52, 2004.
[52] G. F. Liu and Y. C. Zhao, "Content'characteristics and formats of industrial chains," *The Theory and Practice of Finance and Economics*, vol. 27, no. 141, pp. 114–117, 2006.
[53] J. M. Wu and C. Shao, "Research on formation mechanism of industrial chain: '4+4+4' model," *China Industrial Economy*, vol. 4, pp. 36–43, 2006.
[54] Z. L. Fei, L. Zhu, and S. L. Zhao, "The governance of regional industry chain: connotation and measures," *Economic Geography*, vol. 30, no. 10, pp. 1688–1692, 2010.
[55] W. Wang, "The evolution, government and upgrading of the resource-based industrial Chain: taking the copper industry chain of Tongling City as a case," *Economic Geography*, vol. 37, no. 3, pp. 113–120, 2017.
[56] H. Z. Zhang, L. Lu, and Y. N. He, "Industry chain tourism: concept and cases," *World Regional Studies*, vol. 29, no. 5, pp. 1006–1016, 2020.
[57] S. J. Li and Q. Li, "Research on the growth mechanism of e-commerce platform enterprises from the perspective of

industry chain integration," *China Business And Market*, vol. 33, no. 9, pp. 83–92, 2019.
[58] Q. L. Qian, Y. B. Chen, S. X. Liu, J. X. Zhang, and Z. X. Chen, "Development characteristics and formation mechanism of taobao town: taking xintang town in guangzhou as an example," *Scientia Geographica Sinica*, vol. 37, no. 7, pp. 1040–1048, 2017.
[59] M. Porter, *Competitive Advantage*, Free Press, New York, NY, USA, 1985.
[60] D. D. Bai and Z. W. Sun, "The spatial organization and geographic embeddedness of Taobao Village in China," *World Regional Studies*, vol. 28, no. 1, pp. 121–129, 2019.
[61] Y. Qin, "Function orientation and transformation and upgrading of the traditional industries in the xiong'an new area," *Reforma*, vol. 1, pp. 77–86, 2019.
[62] Z. P. Xiao, J. X. Wang, and Y. H. Sun, "The geographical impact of e-retailing and supply chain reconfiguration," *Economic Geography*, vol. 35, no. 12, pp. 98–104, 2015.
[63] X. F. Fan and X. Q. Kang, "Agglomeration level measurement of manufacturing in Shaanxi Province and its influencing factors empirical analysis," *Economic Geography*, vol. 33, no. 9, pp. 115–119, 2013.
[64] F. J. Jin, C. J. Wang, and X. W. Li, "Discrimination method and its application analysis of regional transport Superiority," *Acta Geographica Sinica*, vol. 63, no. 8, pp. 787–798, 2008.
[65] X. G. Cui, C. L. Fang, and Q. Zhang, "Spatial relationship between high-speed transport superiority degree and land-use efficiency in Shandong Peninsula urban agglomeration," *Acta Geographica Sinica*, vol. 76, no. 6, pp. 1149–1161, 2018.
[66] G. Y. Chang, "Social and geographical embeddedness of SMEs in Lanzhou," *Urban Development Studies*, vol. 18, no. 1, pp. 114–119, 2011.
[67] K. R. Sheng, L. P. Wang, and W. Sun, "Patterns and determinants of functional division of cities across product value chain in China," *Geographical Research*, vol. 39, no. 12, pp. 2763–2778, 2020.
[68] Z. Yang, X. Wang, and Z. D. Luo, "The spatial pattern of the Yangtze River Delta Region based on the B2B e-commerce enterprise network," *Urban Planning Forum*, vol. 4, pp. 37–42, 2020.
[69] Y. L. Rao and Z. D. Luo, "Digital transformation and deep diffusion: the evolution of industrial space in the Yangtze River Delta Region based on B2B E-commerce enterprises," *Urban Planning Forum*, vol. 3, pp. 82–89, 2021.
[70] Z. Y. Song and T. Xiao, "Dose spatial distance factor still matter in online shopping behavior: a study of undergraduates'online shopping in Guangzhou," *Economic Geography*, vol. 40, no. 8, pp. 18–26, 2020.
[71] J. Peng and J. He, "Spatial differentiation and influencing factors of fan economy in China: taking TikTok livestreaming commerce host as an example," *Progress in Geography*, vol. 40, no. 7, pp. 1098–1112, 2021.

Hindawi
Complexity
Volume 2022, Article ID 8240806, 14 pages
https://doi.org/10.1155/2022/8240806

Research Article

Digital Financial Inclusion, Spatial Spillover, and Household Consumption: Evidence from China

Yao Li, Haiming Long, and Jiajun Ouyang

College of Finance and Statistics, Hunan University, Yuelu District, Changsha, Hunan, 410006, China

Correspondence should be addressed to Jiajun Ouyang; tahoyou@hnu.edu.cn

Received 18 June 2021; Revised 15 November 2021; Accepted 7 January 2022; Published 25 February 2022

Academic Editor: Haitao Ma

Financial development is often considered one of the main drivers promoting household consumption. As a form of financial development, whether digital financial inclusion can promote household consumption has been a concern for researchers and policymakers. Considering geographical connectivity characteristics, we examine the effects of digital financial inclusion on household consumption by applying spatial econometric models and using data from 31 provinces in China from 2013 to 2018. The impact of digital financial inclusion is further disaggregated into direct, indirect, and total impacts. The results show that if digital financial inclusion is improved by 1%, household consumption will correspondingly increase by 0.2207%. The spatial spillover effect on neighboring areas is negative: a 1% increase in the level of digital financial inclusion of nearby provinces leads to a 0.1289% decrease in household consumption in the local province. For policymakers, it is necessary to balance the development of digital financial inclusion and view different areas as a whole when making policies to promote consumption. Further analysis based on subsamples finds that the effect of digital financial inclusion on household consumption is more considerable for rural households than for urban households.

1. Introduction

Consumption, investment, and export have been described as the "three drivers" of economic growth in China for a long time. However, under the circumstances of slow global economic growth and complex international relationships due to the prevalence of COVID-19, China's net exports show signs of decline. China's net export was RMB 15057.1 billion in 2010 and RMB 11397.9 billion in 2019, a decrease of 24.30%. During the same period, investment, which has long been the largest engine of economic growth in China, also shows a slowing growth rate, with its contribution to GDP growth falling from 47% in 2010 to 43.1% in 2019. As investment and export hit bottlenecks in China, consumption is playing an increasingly fundamental role in economic growth. Its contribution to economic growth rose from 49.3% in 2010 to 55.8% in 2019. This indicates that China's economy has transformed from investment-driven growth and export-driven growth into consumption-driven growth, and consumption has become the most considerable driving force. The Fifth Plenary Session of the 19th Central Committee proposed emphasizing the "domestic cycle", which means stimulating domestic demand and making consumption the main engine of economic development. Therefore, under the new normal of China's economy, expanding domestic demand and promoting consumption are of great significance to China's sustainable development.

Thanks to the development of innovative technologies, such as big data, the Internet, and cloud computing, digital financial inclusion has experienced exponential growth since the late 2000s and early 2010s. According to the digital financial inclusion index, the index increases from 40 in 2011 to 300 in 2018. Digital financial inclusion provides various services to residents, including mobile payments, online loans, online insurance, and Internet finance. By the end of 2020, mobile payment had 854 million users, accounting for 86.4% of the total number of netizens in China, and its transaction value had reached RMB 200 trillion. "Huabei" and "Jiebei", online consumer loan services launched by AliFinance, had offered loans up to USD 95 billion by the first quarter of 2018, nearly 3.7 times the amount of loans launched by the China Construction Bank [1]. In recent years, 12.29 billion new Internet insurance policies have been issued, accounting for 71% of the number of new insurance policies. Yu'E Bao, the country's largest online money market fund, was launched in June 2013 by Alipay. By the first quarter of 2014, it had more than 15 million users and assets reaching RMB 250 billion.

Financial development plays a pivotal role in boosting household consumption. It can promote efficient resource allocation and smooth intertemporal consumption, thus relieving residents from liquidity constraints and increasing consumption demand [2]. As a form of financial development, digital financial inclusion influences household consumption in an irreplaceable way. Some studies have studied the relationship between digital financial inclusion and household consumption, but few take spatial links into account. However, according to the first law of geography, everything is alike and correlated to each other [3]. First, because of the large territory and great economic and geographic heterogeneity, there are tremendous differences between provinces. As a result, geographical disparities in household consumption and digital financial inclusion are apparent. However, with the improvement in infrastructure, the popularization of Internet technology, and the issuance of interprovincial cooperation policies, communication across regions is becoming closer. Factors of production, such as talent and technology, thus flow more frequently. Consequently, household consumption and the development of digital financial inclusion show spatial correlation characteristics. Additionally, as provinces become more closely linked, the socioeconomic situation of one province might exhibit a remarkable demonstration effect [4] or crowding-out effect [5] on neighboring provinces immediately or within a short period [6]. Therefore, identifying the spatial distribution and taking spatial correlation into account are meaningful when exploring the relationship between household consumption and digital financial inclusion.

This paper contributes to the relevant research in three dimensions. First, the attributes of entities that are adjacent to each other are prone to be similar. Spatial econometrics takes spatial agglomeration and the evolution of spatial-temporal patterns into account. This approach is chiefly used to investigate environmental economics and is hardly applied to studies on digital financial inclusion and household consumption. In this paper, we use the global Moran's I index and local Moran's I index to describe their spatial autocorrelation characteristics in China. Second, the existing research considering the relationship between digital financial inclusion and household consumption assumes that spatial objects with adjacent geographical locations are not related. Based on SDM, we analyze the spatial spillover effects in China, taking the effect of neighboring provinces into account. This method is particularly useful for providing suggestions to policymakers about how to optimize household consumption among different areas to create win-win situations. Third, recent studies tend to regard the effect as a whole and seldom consider the difference between disadvantaged people in rural districts and the main groups in urban districts. However, by comparing the empirical results of these two groups, we find that digital financial inclusion has significantly distinct impacts on household consumption in China.

The rest of this paper is organized as follows: Section 2 presents an overview of the related literature and proposes research hypotheses. Section 3 introduces the adopted data, variables, and empirical model. Section 4 exhibits the empirical results, associated interpretations, endogeneity tests, and robustness tests. Section 5 concludes.

2. Literature Review and Research Hypotheses

Financial inclusion can ease the exclusiveness of traditional finance. The development of digital financial inclusion can improve financial services' accessibility, coverage, and efficiency [7]. Especially for unbanked residents, digital financial inclusion expands the channels by which they can obtain access to and actively use affordable financial services that can meet their various financial needs [8]. Zhong and Jiang (2020) empirically demonstrated this viewpoint by using daily peer-to-peer (P2P) lending data from multiple platforms in China. They found that digital financial inclusion can reduce the asymmetry in traditional financial markets and thus weaken the exclusiveness of traditional finance [9]. As long as consumers have digital devices that can access the Internet, they can experience services offered by digital financial inclusion [10].

Digital financial inclusion has a significantly positive impact on household consumption [11, 12]. More precisely, digital financial inclusion is expected to influence household consumption in several aspects by providing different kinds of services. First, mobile payments make it easier for consumers to make purchases. It can avoid the situation in which residents do not have enough cash, thus accelerating the process of consumption decisions and increasing the likelihood of consumption [12]. In addition, mobile payments can significantly stimulate household consumption through a "mental account". The theory argues that the psychological loss caused by digital transactions is lower than that caused by cash transactions [13]. Moreover, by lowering the threshold of entry and reducing the impacts of distance on trade, mobile payments promote the emergence of e-commerce. E-commerce can not only enrich the diversity of commodities but also reach disadvantaged consumers who were previously not served [14]. Second, the life and transaction data of residents can be accumulated through big data, thus improving information asymmetry and helping individuals obtain credit easily and quickly. As a result, their liquidity constraints can be relieved [15–17]. Third, on the one hand, the development of online insurance can intensify competition among insurance companies and facilitate the upgrading of the traditional insurance industry. On the other hand, it can also enrich the types of products and improve insurance accessibility, which encourages consumers to buy insurance, reduces their uncertainty losses, and improves their sense of security [18]. Fourth, the emergence of Internet finance lowers the threshold of investment services, improves investment convenience, and provides subsidy channels [19]. For example, by expanding the approaches to investments with a small amount of money, Yu'E Bao increases investment returns and household wealth.

Based on the analysis, this paper proposes the following hypothesis:

Hypothesis 1. (H1). *The development of digital financial inclusion can promote household consumption.*

In theory, digital finance does not need to rely on physical outlets. As a result, it has greater geographical penetration and can save costs. However, as a new form of finance, digital finance still needs to follow the basic rules of financial development. On the one hand, its development still depends on the real economy and traditional finance. Digital finance evolves as the innovative superposition of traditional finance, and the two are mutually integrated and complementary [20]. On the other hand, the factors that influence the development of digital finance may have much in common with those that affect the real economy and traditional finance. These factors are usually spatially related, such as GDP, the Internet, industrial structure, population density, and openness [21]. Moreover, the promotion and diffusion of digital finance depend on geography. For example, Ant Financial is headquartered in Hangzhou. If the company wants to promote business in other places, the level of difficulty is positively related to the distance from Hangzhou. This rule has been repeatedly emphasized in traditional economic geography, and digital finance can hardly escape it. In addition, digital finance originated from online shopping and online payment. Due to express delivery and other factors, online shopping is highly geographically dependent, as shown taking "free delivery" in Jiangsu, Zhejiang, and Shanghai as an example [22].

Second, with the improvement in the transportation system and the decrease in communication costs, communication between different provinces is becoming closer in China, and production factors can easily flow from one region to another [14, 23]. This situation is beneficial to the exchange of resources and the sharing of information between adjacent provinces regarding the development of digital financial inclusion [21].

Third, the mode of governmental guidance "from points to areas; parts pushing the whole" makes spatial agglomeration the characteristic of the level of digital financial inclusion [24]. In addition, governments tend to learn from and imitate each other when making policies, thus forming province clusters and further promoting the spatial agglomeration of the level of digital financial inclusion.

Therefore, although digital financial inclusion can have some hypergeographic characteristics, in theory, it cannot eliminate geographical restrictions in reality and shows strong spatial agglomeration [25–28].

Regarding the spatial effect of digital financial inclusion on household consumption, on the one hand, the development of digital financial inclusion can increase household consumption, and it presents spatial correlation. The development of digital financial inclusion in nearby places can affect the development of digital financial inclusion in a province, thus indirectly increasing household consumption in the area [29]. Therefore, the development of digital financial inclusion in nearby places might contribute to the growth of household consumption in a particular place. On the other hand, unbalanced digital financial inclusion development may intensify interprovincial competition in resource distribution. As a result, the economy of provinces with a higher level of digital financial inclusion prospers, siphoning increasingly more resources from less-developed provinces and impeding their development, thus negatively affecting household consumption [24]. Therefore, the development of digital financial inclusion in nearby places might bring about a crowding-out effect and decrease household consumption in a particular district [5].

Based on the analysis, this paper proposes the following hypothesis:

Hypothesis 2. (H2). *The development of digital financial inclusion has a spatial effect on household consumption.*

3. Methodology

3.1. Data. The household data and control variables are from the China Statistical Yearbook released by the National Bureau of Statistics. The calculated calibers of household consumption before 2013 and after 2013 are different. After considering the comparability of the data, we decide to use the data after 2013.

The digital financial inclusion index was compiled by Alipay and the Internet Finance Institute of Peking University from 2011 to 2018 to reflect the level of digital financial inclusion in China. Table 1 shows the components of the comprehensive index. The index consists of three dimensions: coverage, depth, and digital support services. There are three levels of indexes available: provincial, municipal, and county.

Based on the principles of availability and consistency, we use data from 31 provinces in China from 2013 to 2018.

3.2. Spatial Econometric Model. The existing literature illustrates various factors that impact household consumption [30–32]. Based on these, we choose the following control variables: household income, human capital, transport infrastructure, fiscal expenditure, industrial structure, openness, and population density. Table 2 shows how those variables are defined.

To empirically investigate how digital financial inclusion affects household consumption, we exhibit a baseline regression model as follows:

$$\begin{aligned}\ln \mathrm{Con}_{it} = {} & \beta_0 + \beta_1 \ln \mathrm{DFI}_{it} + \beta_2 \ln \mathrm{Incom}_{it} + \beta_3 \ln \mathrm{Hum}_{it} + \beta_4 \mathrm{Trans}_{it} \\ & + \beta_5 \mathrm{Fisc}_{it} + \beta_6 \mathrm{Indsu}_{it} + \beta_7 \mathrm{open}_{it} + \beta_8 \mathrm{Pop}_{it} + u_i + \varepsilon_{it},\end{aligned} \tag{1}$$

Table 1: Components of the digital financial inclusion index.

	Level 2 indexes	Level 3 indexes
Digital financial inclusion index	Coverage	Coverage of account
	Coverage depth	Coverage of account payment
	Depth digital support services	Payment insurance
		Insurance investment
		Investment monetary fund
		Monetary fund credit
		Credit loan
		Loan convenience
	Digital support services	Convenience cost

Table 2: Variable definition.

Variables	Symbol	Definition
Household consumption	Con	The aggregate household expenditure
Digital financial inclusion	DFI	Digital financial inclusion index compiled by Alipay and Internet finance institute of Peking university
Household income	Incom	The aggregate household income
Human capital	Hum	The number of students enrolled in colleges and universities
Transport infrastructure	Trans	The sum of paved roads and railways/total area of province
Fiscal expenditure	Fisc	The aggregate of government expenditure/GDP
Industrial structure	Indus	The added value of the tertiary industry/GDP
Openness	Open	The total value of imports and exports/GDP
Population density	Pop	The total population/total area of province

where β_k is the kth parameter to be estimated; i represents the province, and t represents the period; u_t represents the time fixed effect; and ε_{it} is the residual error.

Table 3 presents the descriptive statistics of the independent variable, dependent variables, and control variables. As Table 3 shows, the average household consumption is RMB 16,443.12, and the average digital financial inclusion index is 226.2849. The maximum value of household consumption is RMB 43351.3, which is more than six times the minimum value (RMB 6306.79). This implies that there is a significant disparity in consumption between different provinces. The maximum value of digital financial inclusion is 377.7337, while the minimum value is 115.1, indicating substantial inequality in the development of digital financial inclusion in China.

Equation (1) does not consider the spatial dependence between entities. However, the household consumption of adjacent provinces may affect the household consumption of a province, as may the development of neighboring areas. This is an inherent shortcoming of traditional econometric models, which consider cross-sectional units to be unrelated. Spatial econometrics is an effective method to evaluate geographical dependency. Next, we introduce spatial econometric models to take geographical dependency into consideration. The spatial Durbin model below is specified for this purpose:

$$\begin{aligned} \ln \text{Con}_{it} = {} & \beta_0 + \rho \sum_{j=1}^{N} w_{ij} \ln \text{Con}_{jt} + \beta_1 \text{DFI}_{it} + \beta_2 \ln \text{Incom}_{it} + \beta_3 \ln \text{Hum}_{it} + \beta_4 \text{Trans}_{it} \\ & + \beta_5 \text{Fisc}_{it} + \beta_6 \text{Indsu}_{it} + \beta_7 \text{open}_{it} + \beta_8 \text{Pop}_{it} \\ & + \theta_2 \sum_{j=1}^{N} w_{ij} \ln \text{DFI}_{jt} + \theta_2 \sum_{j=1}^{N} w_{ij} \ln \text{Incom}_{jt} + \theta_3 \sum_{j=1}^{N} w_{ij} \ln \text{Hum}_{jt} \\ & + \theta_4 \sum_{j=1}^{N} w_{ij} \text{Trans}_{jt} + \theta_5 \sum_{j=1}^{N} w_{ij} \text{Fisc}_{jt} + \theta_6 \sum_{j=1}^{N} w_{ij} \textstyle\sum_{j=1}^{N} w_{ij} \text{Indus}_{jt} \\ & + \theta_7 \sum_{j=1}^{N} w_{ij} \text{Open}_{jt} + \theta_8 \sum_{j=1}^{N} w_{ij} \text{Pop}_{jt} + \varepsilon_{it}, \end{aligned} \tag{2}$$

$$\varepsilon_{it} = \lambda w \varepsilon_{it} + \mu_{it}, \tag{3}$$

Table 3: Descriptive statistics.

Variables	Mean	Std. Dev.	Min	Max	Obs
Con	16443.12	6715.299	6306.79	43351.3	186
DFI	226.2849	55.47315	115.1	377.7337	186
Incom	22914.17	10036.82	9740.43	64182.65	186
Hum	855996.5	526122.5	33474	2140780	186
Trans	9.426809	5.299249	0.5916085	21.74448	186
Fisc	0.2860584	0.2115872	0.1207761	1.379157	186
Indus	0.0735436	0.0131422	0.0560575	0.1273875	186
Open	0.0735436	0.0131422	0.0560575	0.1273875	186
Pop	0.0452295	0.0692612	0.0002595	0.3826197	186

where w_{ij} is the spatial weight matrix; ρ is the spatial autoregression coefficient, which reflects the spatial spillover effect of the dependent variable; θ is the spatial lag coefficient, which reflects the spatial spillover effect of the independent variables; ε_{it} is the residual error; λ is the spatial autoregression coefficient of the residual error; and u_{it} is the specific effect.

Equation (2) is the spatial Durbin model (SDM) and can be simplified into the spatial autoregression model (SAR) and spatial error model (SEM). If $\theta = 0$, the SDM can be transformed into an SAR. If $\theta + \rho\beta = 0$, the SDM can be transformed into an SEM. The Wald test and likelihood ratio (LR) test examine whether $\theta = 0$ and $\theta + \rho\beta = 0$, respectively. These tests are performed below.

3.3. Spatial Weight Matrix. The selection of the spatial weight matrix is pivotal when constructing spatial models because a matched matrix can support the validity of the judgment and estimation. The most commonly used matrices are the adjacent spatial weight matrix, geographical distance spatial weight matrix, economic distance weight matrix, and relative threshold weight matrix [33]. Because of the unique structure of different provinces in China, applying the geographical distance spatial weight matrix, economic distance weight matrix, or relative threshold weight matrix is unreasonable [34]. Accordingly, the adjacent spatial weight matrix remains the only choice.

An adjacent spatial weight matrix is set based on the criteria of contiguity. $w_{ij} = 1$ means *i* and *j* share a boundary; otherwise, $w_{ij} = 0$:

$$w_{ij} = \begin{cases} 1, & i \text{ is adjacent to } j, \\ 0, & i \text{ is not adjcent to } j. \end{cases} \tag{4}$$

4. Empirical Analysis

4.1. Spatial Distribution Analysis. To better understand the spatial distribution characteristics of household consumption and the level of digital financial inclusion in China, we draw spatial distribution maps of 31 provinces in mainland China in 2013 and 2018 (Figures 1 and 2).

Household consumption is shown in Figure 1. In 2013, the five provinces with the most household consumption were Shanghai, Beijing, Zhejiang, Tianjin, and Jiangsu, which are all located in the eastern areas, while the five provinces with the least household consumption were Tibet, Guizhou, Yunnan, Gansu, and Guangxi, which are all in the west. The household consumption of Jilin and Tibet is the highest among provinces in the central and western districts, respectively, but the average household consumption of eastern districts (RMB 18276.33091) is almost twice as high as that of Jilin (RMB 12054.34) and nearly three times that of Tibet (RMB 6306.79).

In 2018, household consumption increased from RMB 13307.11226 to RMB 19787.75613. The top five provinces and the last five provinces in terms of household consumption are the same as those in 2013. Households living in Hubei consume the most (RMB 19537.79) among central provinces, but the number is still slightly smaller than the average number of households in the whole country. Inner Mongolia has the highest household consumption (RMB 19665.22) among provinces in the west, but it is still less than the national average.

On the whole, household consumption in China is undoubtedly unbalanced. There is a gradient transfer characteristic from east to west, indicating that household consumption is high in the east but low in the west. Although the sum of household consumption of each province increases during the corresponding sample period, the pattern of distribution is relatively stable. This phenomenon is partly because of the geographical location and economic development plan of China. Due to its vast territory, China is usually separated into three parts: the western region, the central region, and the eastern region[1]. The eastern areas have long coastlines, which is the most prominent difference from provinces in the western region, while the provinces in the west are all inland and far from the coast [35]. Because of their geographical advantage, eastern provinces opened up earlier and thus attracted more foreign investment than other provinces. The government's preference for supporting eastern provinces further intensifies the imbalance of economic development among different areas. In the later period of China's reform and opening up, China gradually opened up to foreign investors and international markets, from the coastal to the middle and inland areas. Geographical features and governmental policies contribute to disproportionate financial development, with the eastern provinces developing much faster than the central and western provinces. Accordingly, families living in the east are wealthier than those in the west, thus consuming more.

The development of digital financial inclusion is shown in Figure 2. In 2013, the five provinces with the highest level of digital financial inclusion were Shanghai, Beijing, Zhejiang, Fujian, and Guangdong, which are all located around Hangzhou (where the headquarters of Ant Financial are located). The five provinces with the lowest levels were Sichuan, Tibet, Gansu, Qinghai, and Ningxia, which are far from the coast. In 2018, the level of digital financial inclusion increased from 155.75 to 300.21. The top five provinces in terms of the level of digital financial inclusion were Shanghai, Beijing, Zhejiang, Fujian, and Jiangsu, while the last five were Inner Mongolia, Gansu, Qinghai, Ningxia, and Xinjiang.

On the whole, the development of digital financial inclusion presents a characteristic of spatial heterogeneity. The eastern coastland, where Hangzhou is located, gathers

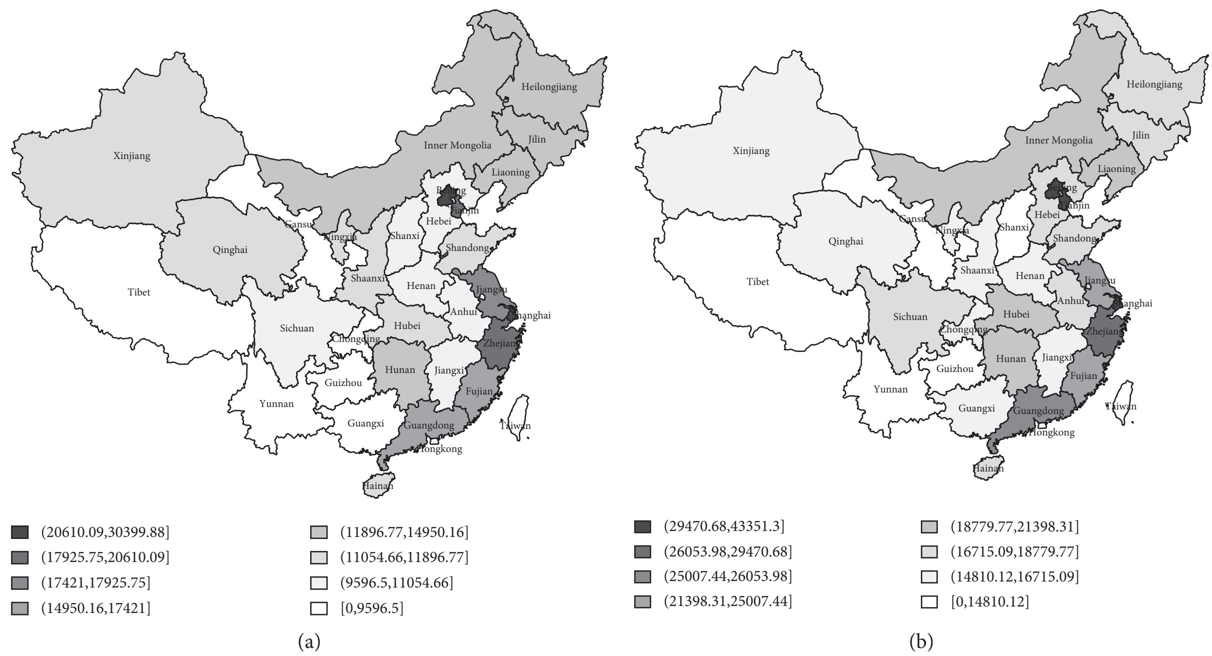


Figure 1: Spatial distribution of household consumption in 2013 (a) and 2018 (b).

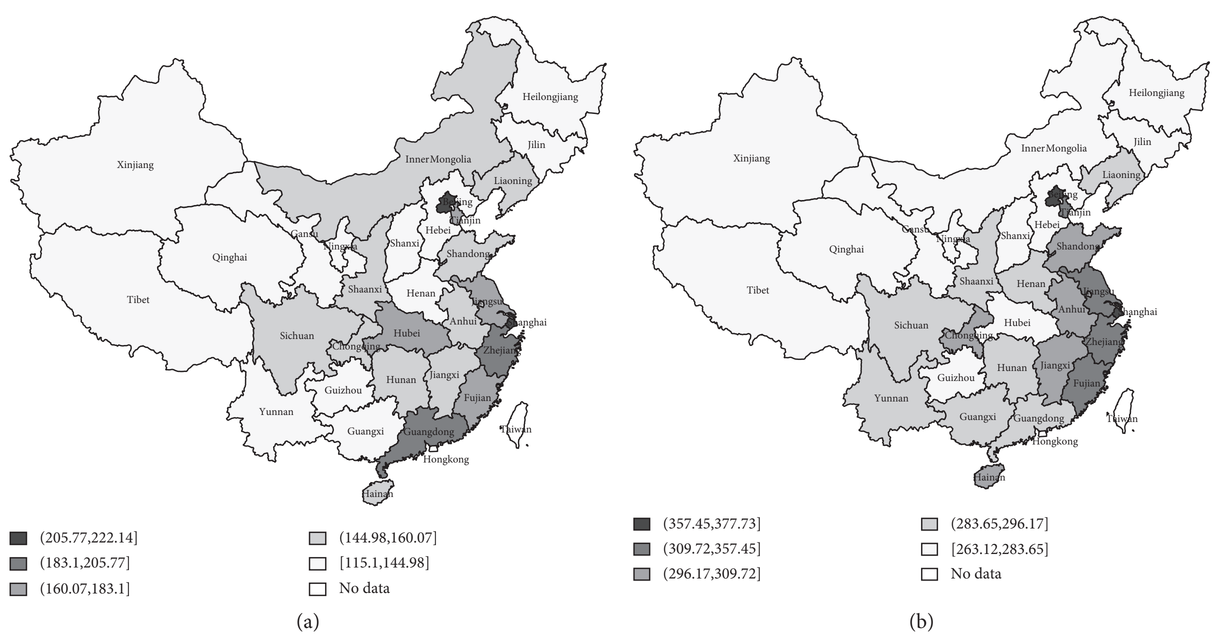


Figure 2: Spatial distribution of the development of digital financial inclusion in 2013 (a) and 2018 (b).

provinces with advanced digital financial inclusion, while high-level provinces are relatively scarce in the central and western regions. This phenomenon implies that the development of digital financial inclusion still depends on geography. That is, the closer a province is to Hangzhou, the higher its development of digital financial inclusion. This is partly because the promotion and diffusion of digital finance depend on geography.

Based on the analyses above, we conclude that the spatial distribution of China's household consumption presents a ladder distribution from coastal to inland areas and has a steady-state characteristic. The levels of digital financial inclusion in China basically decline from Hangzhou to its neighbors.

4.2. Spatial Autocorrelation Analysis. Spatial autocorrelation measures aim to investigate the spatial relationship between entities with adjacent geographical locations [36]. The Moran's I index, Geary coefficient, and Getis index can be used to examine spatial autocorrelation and spatial agglomeration. Among them, the Moran's I index is the most frequently used. If an entity is affected by others around, spatial correlation exists. This is not in accordance with the basic premise of traditional statistics and makes most statistical tests invalid, so the tests are essential. The Moran's I index consists of the global Moran's I index and local Moran's I index.

The global Moran's I test measures how similar an entity is to others that are geographically nearby [37, 38]. Similar to the correlation coefficient, the global Moran's I index is distributed in [−1, 1]. The coefficient is high when the absolute value of the autocorrelation coefficient is high. When the global Moran's I index varies between [0, 1], it indicates that the entity is positively correlated with others. When the index varies between [−1, 0], it means that the correlation is negative. A value of 0 implies no correlation.

Based on the adjacent spatial weight matrix, we use the global Moran's I to test the spatial autocorrelation of household consumption and the level of digital financial inclusion of 31 provinces in China from 2013–2018. Table 4 displays the results. As shown in Table 4, the global Moran's I indexes are between [0, 1], and all are significant at the 1% level, implying that household consumption and the level of digital financial inclusion in China are positively autocorrelated in space. In addition, the global Moran's I index basically increases over time, especially after 2015, which implies that the spatial autocorrelation strengthens with continuous socioeconomic development [39].

The results above imply that provinces with similar levels of household consumption and digital financial inclusion are prone to concentrate geographically. To further explore the spatial effects at the provincial level, scatter plots of the local Moran's I index are drawn (Figures 3 and 4). Local Moran's I scatter plots can divide household consumption into four quadrants that denote four kinds of agglomerations. Quadrant I indicates high-high (HH) agglomeration, which means that provinces with high household consumption are near to each other. Quadrant III reflects low-low (LL) agglomeration; that is, if a province with low household consumption falls into quadrant III, the provinces nearby also have lower consumption. Quadrants II and IV represent low-high (LH) agglomeration and high-low (HL) agglomeration, respectively, which means that a province with low household consumption is surrounded by provinces with high household consumption, and vice versa.

For the results of household consumption shown in Figure 3, most of the provinces are located in quadrant I and quadrant III, which means that most provinces have similar characteristics to their neighbors, in line with the spatial cluster characteristic detected by the global Moran's I index. In 2013, 6 provinces fall into quadrant I and 16 provinces into quadrant III, occupying 70.97% of the total number of samples. The number of provinces located in quadrant I and quadrant III is the same in 2014 and 2013. In 2015, 6 provinces are in quadrant I, and 17 provinces are in quadrant III, with a proportion of 74.19%. The situation in 2015 persists into 2016 and 2017. In 2018, 6 provinces are located in quadrant I, and 19 provinces are in quadrant III, accounting for 80.65%.

For the level of digital financial inclusion, the same characteristics of positive spatial agglomeration can be found in Figure 4. Most of the provinces are located in quadrant I and quadrant III, representing 70.97% and 77.42% of the total number in 2013 and 2018, respectively. The provinces falling into quadrant I are coastal provinces, while those falling into quadrant III are mainly inland provinces.

Table 4: Global Moran's I indexes (2013–2018).

Year	Global Moran's I index	
	Household consumption	Digital financial inclusion
2013	0.371***	0.456***
2014	0.373***	0.453***
2015	0.372***	0.416***
2016	0.362***	0.438***
2017	0.369***	0.498***
2018	0.377***	0.544***

Note. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The results show that the geographical effect cannot be ignored in the study of digital financial inclusion and household consumption. Otherwise, the estimation would be invalid.

4.3. Spatial Effect Test

4.3.1. Lagrange Multiplier (LM) Test. Before building the SDM model, we first use LM tests to test the existence of the spatial effect. The tests determine whether SAR or SEM can describe the features of the data better than the OLS model, which does not consider spatial interaction. We make judgments by using the LM test (LM_{SAR}, LM_{SEM}) designed by Anselin (1988) and the robust LM test (Robust LM_{SAR}, Robust LM_{SEM}) developed by Anselin et al. (1996) [33, 40]. If the results of LM_{SAR} and LM_{SEM} are nonsignificant, the OLS model should be selected. However, if either of the statistics are significant, spatial econometric models should be built. We should select the model with the larger LM statistic under the same condition of significance. If we are not able to conclude merely according to the LM statistic, the robust LM statistics are needed, and the model with the larger robust LM statistic under the same significance is considered more convincing.

Table 5 shows the results of the LM test and the robust LM test. Only the statistics of the SAR model (LM_{SAR} and Robust LM_{SEM}) pass the significance level test of 1%, demonstrating the existence of spatiality. Thus, it is more appropriate to select SAR than SEM.

4.3.2. Likelihood Ratio Tests (LR) and Wald Test. Elhorst (2010) pointed out that if the OLS model is not accepted by the LM tests, then we should examine whether the SDM can be transformed into the SAM or SEM based on the results of the LR test and the Wald test [41].

The LR test and the Wald test are built to test the hypotheses H_0: $\theta = 0$ and H_0: $\theta + \rho\beta = 0$ (the letters are defined in equation (2)). The results of the LR test and the Wald test are presented in Table 6. As Table 6 indicates, the

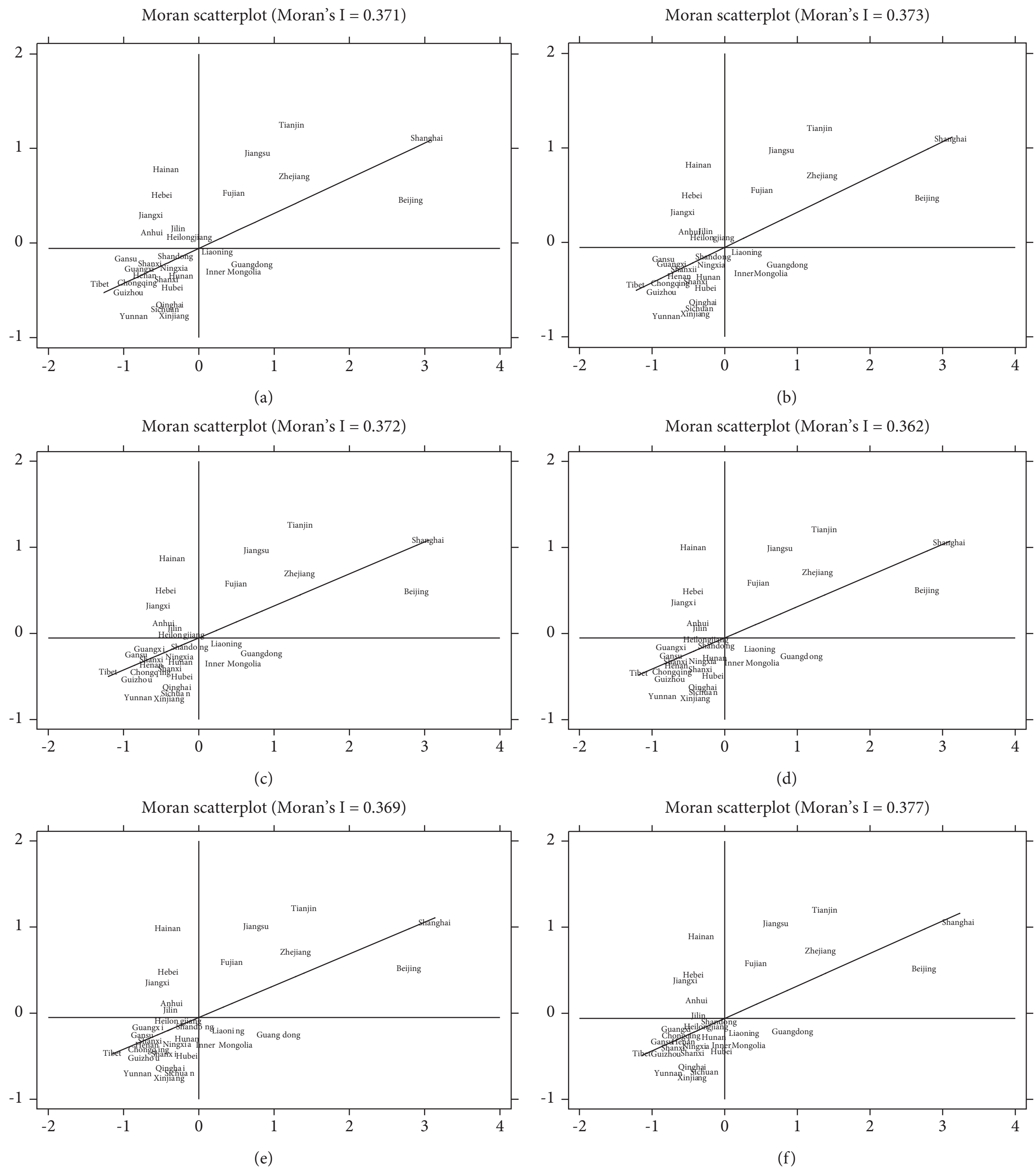
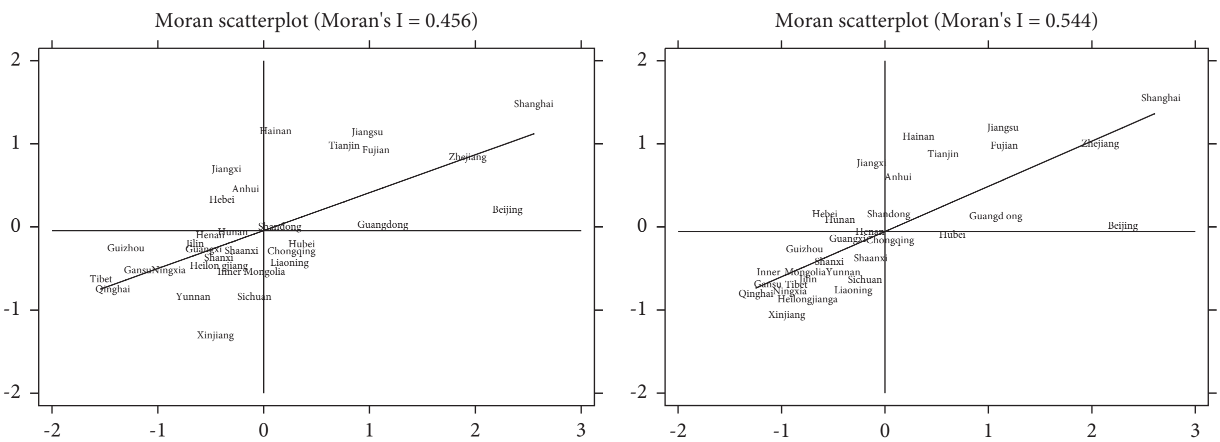


Figure 3: Scatter plots of the local Moran's I index for household consumption (2013–2018). (a) Household consumption of 2013. (b) Household consumption of 2014. (c) Household consumption of 2016. (d) Household consumption of 2017. (e) Household consumption of 2018.

statistics are all significant at the 1% level, proving that the SDM cannot be simplified, and the SDM is the final choice.

4.3.3. Hausman Test. Before model construction, the Hausman test is needed to determine whether random effects or fixed effects are more appropriate. According to the result, the statistic value is 227.87 and is significant at the 1% level, so we choose fixed effects instead of random effects.

4.4. The Effect of Digital Financial Inclusion on Household Consumption. We report the empirical results of the OLS model and spatial model in column (1) and column (2) in



(a) (b)

Figure 4: Scatter plots of the local Moran's I index for digital financial inclusion (2013 and 2018). (a) Digital Financial Inclusion Index in 2013. (b) Digital Financial Inclusion Index in 2018.

Table 7, respectively. For the results of the spatial model in column (2), researchers pointed out that the coefficients should be decomposed to prevent estimation bias caused by the feedback loop effect. The decomposition results are shown in Table 8. The coefficient of digital financial inclusion in the OLS model is not significant, while the coefficients in the decomposition in Table 8 are all significant. This difference implies that the OLS result might be contaminated by ignoring variable bias [42]. That is, the nonsignificance of the OLS model can be attributed to endogeneity caused by variable omission. According to the tests presented above and the fact that production factors flow across boundaries with increasing frequency, we know that provinces are increasingly correlated. Therefore, if we fail to consider spatial factors, the estimation will be invalid. Overall, the OLS model fails to consider the important effects of variables from surrounding areas, and spatial models can partly solve this problem.

The parameter of spatial autoregression ρ is −0.6400 and is significant at the 1% level. This phenomenon might be interpreted as follows. First, in general, household consumption among regions presents a competitive relationship because resources are limited as a whole. Resources are usually allocated according to demand, which contributes to a spatial crowding-out effect [5]. This indicates that higher household consumption demand in one region may increase the consumer goods transmission from its neighbors and accordingly inhibit household consumption in those provinces. Second, the increasing development of transportation systems has significantly facilitated population movement between different areas. People tend to move to regions with higher living standards and more kinds of consumer goods, which accelerates regional differences in household consumption. From this viewpoint, it is imperative to develop policies that strengthen cooperation among provinces, promote effective resource allocation, and increase the aggregate supply of goods to form a balanced consumption pattern and a win-win situation.

Table 5: Results of the LM test and robust LM test.

Test method	Statistic	p-value
LM_{SAR}	12.782***	≤0.01
Robust LM_{SAR}	12.877***	≤0.01
LM_{SEM}	0.953	0.33
Robust LM_{SEM}	1.048	0.31

Note. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

4.5. Decomposition of the Spatial Spillover Effect. In the ordinary model, the coefficient β can directly reflect the impact of independent variables on the dependent variable. However, in the case of spatial econometric models, the explanation of coefficients is more complex [43]. Anselin and LeGallo (2006) and Kelejian and Prucha (2007) pointed out that for SEMs with the spatial autoregression term, the coefficients should be particularly explained to prevent estimation bias caused by the feedback loop effect [44, 45]. Thus, we decompose the spatial spillover effect into the direct effect, the indirect effect, and the total effect. The decomposition results of the SDM are listed in Table 8.

For the effect of digital financial inclusion, a 1% improvement in digital financial inclusion directly increases household consumption by 0.2207% but indirectly decreases it by 0.1289%, and the net effect is a total increase of 0.0917%. By providing diversified financial services, such as mobile payments, online loans, Internet insurance, and Internet finance, digital financial inclusion may affect household consumption in four dimensions. Digital payment platforms give rise to online shopping, which can improve information asymmetry and diversify commodities in the consumer market, thus boosting household consumption. Digital credit expands the way residents obtain loans, eases their liquidity constraints, and, as a result, promotes household consumption. Online insurance makes it easier for residents to obtain access to insurance services, thus smoothing consumption when individuals face adverse shocks. Online investment provides households with accessible investment services and increases their investment income, accordingly

Table 6: Results of the LR test and the Wald test.

LR/Wald	Statistic	p-value
LR test (SDM vs. SAR)	91.32***	≤0.01
LR test (SDM vs. SEM)	72.29***	≤0.01
Wald test	88.34***	≤0.01

Note. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 7: Results of the nonspatial model and SDM.

Variables	(1) panel	(2) SDM
Ln(DFI)	0.0509(1.14)	0.1956***(3.96)
Ln(Incom)	0.7772***(11.67)	0.6141***(5.78)
Ln(Hum)	0.0792**(1.98)	0.0357(1.11)
Trans	0.0254***(5.17)	0.0282***(8.17)
Fisc	0.1457*(1.65)	0.2743***(3.89)
Indus	−0.4043(−0.84)	0.6768*(1.70)
Open	0.3203(0.49)	−0.3720(−0.81)
Pop	−0.3142(−0.11)	5.7158***(2.64)
w*Ln(DFI)		−0.0421(−0.53)
w*ln(Incom)		0.4967**(2.21)
w*ln(Hum)		0.5457***(8.58)
w*(Trans)		0.0439***(4.82)
w*(Fisc)		0.0535(0.41)
w*(Indus)		0.9989(1.20)
w*(Open)		−0.4187(−0.46)
w*(Pop)		−0.6529(−0.17)
ρ		−0.6400***(−5.80)
θ		0.0002***(9.20)
R-square	0.7840	0.4875

Note. z-values are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 8: Decomposition of the spatial spillover effect at the national level.

Variables	Direct effects	Indirect effects	Total effects
Ln(DFI)	0.2207***(3.63)	−0.1289*(−1.72)	0.0917**(2.10)
Ln(Incom)	0.5860***(4.90)	0.0879(0.49)	0.6738***(5.69)
Ln(Hum)	−0.0373(−0.99)	0.3963***(7.18)	0.3589***(8.19)
Trans	0.0248***(5.47)	0.0190***(2.58)	0.0440***(7.54)
Fisc	0.2756***(2.90)	−0.0823(−0.72)	0.1933**(2.19)
Indus	0.6050(1.52)	0.4900(0.84)	1.0952*(1.93)
Open	−0.3343(−0.64)	−0.1844(−0.25)	−0.5187(−0.81)
Pop	5.8262**(2.43)	−2.6660(−0.86)	3.1602(1.50)

Note. z-values are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

improving their level of consumption. As a whole, by providing a variety of financial services, digital financial inclusion increases the actual and expected income of residents, thus influencing their purchasing decisions. For the negative indirect effect, it seems that the negative spatial spillover effect outweighs the positive effect. The negative indirect effect suggests a negative spatial dependency in determining household consumption in neighboring provinces. As mentioned above, the most plausible explanation may be the interprovincial movements of capital, knowledge, high-skilled laborers, and even innovative enterprises, which are increasingly regarded as local growth drivers. The movements can mainly be attributed to easier financial service access induced by the advanced digital financial inclusion of neighboring areas. Taking small and medium-sized enterprises (SMEs) as an example, SMEs in China have long faced financial restrictions linked to traditional finance, and digital financial inclusion can ease this problem [46]. Therefore, these enterprises are more likely to move from provinces with a low level of digital financial inclusion to those with a high level. The majority of SMEs are innovative companies with much potential in economic terms, and thus, the unbalanced pattern of economic development will worsen. Overall, provinces with advanced digital financial inclusion can develop better and attract more resources and residents from a province with a lower level of digital financial inclusion, thereby negatively influencing economic development and household consumption in the province. The vampire effect is obvious [34]. For policymakers, it is essential to balance the regional development of digital financial inclusion in terms of quantity and coverage. By doing so, digital financial inclusion can have an immeasurable influence on household consumption promotion at the national level in the future.

For the impact of household income, the direct, indirect, and total effects are all positive. The direct and total effects are significant at the 1% level, while the indirect effect is not significant. A 1% increase in household income will lead to a growth of 0.5860% in household consumption, indicating that household income is one of the main engines of consumption. This result is in accordance with Keynes (1936), Friedman (1957), and Grunberg and Modigliani (1954) [47–49]. They pointed out that deposits and consumption depend on income. They believed that consumers would consume according to their income level, and as their income increases, people spend more on consumption. The indirect effect indicates that a 1% growth in the income of surrounding provinces drives the growth of consumption in the local province by 0.0879%, showing mutual promotion and synergy in terms of the effect of household income. Although the impact of household income on household consumption in the sample is not very obvious, it is still meaningful for helping China stimulate household consumption.

Transport infrastructure has significantly positive direct, indirect, and total effects on household consumption, with values of 0.0248, 0.0190, and 0.0440, respectively. This result is in accordance with that of Zhang and Chen (2003). They held that modernized traffic networks can attract capital, technology, and talent resources and thus are conducive to economic development and consumption growth [50]. China's transportation system usually links a province to its surroundings. As a result, the effect of transportation infrastructure in China presents a significant agglomeration effect of mutual promotion and a spatial spillover effect. As transportation infrastructure is further improved, these agglomeration and spillover effects are increasingly enhanced. However, in recent years, the existing transportation systems have not met the ever-growing demand of population explosion brought about by urbanization and the two-child policy. Economic development is also restricted by traffic congestion and the shortage of transportation resources. Accordingly, it is essential to invest more in road expansion and transportation construction and to better support the driving role of transportation in household consumption [34].

From the perspective of human capital, the indirect and total effects are positive at the 1% significance level, while the direct effect is exceptional, being negative and nonsignificant. The positive indirect effect may be explained by imitative behavior: to keep up with the economic development of their neighbors, local governments are inclined to observe and imitate the policies of their neighbors. An advanced education environment provides a strong impetus for economic development and household consumption. Hence, the education level of one province can be conditional on that of its neighbors, resulting in positive spillover effects of the education environment among these regions.

Regarding fiscal expenditure, the local and total impacts are significantly positive, but the indirect effect shows a reserved result, as it is a negative externality and is not significant. According to the precautionary savings hypothesis, when the expected income in the future is uncertain, consumers will become cautious. They will increase savings to deal with the possible risks caused by the uncertainty of income [18]. Researchers name this kind of savings caused by the uncertainty of future income precautionary savings. Carroll (1994) supported the conclusion that the primary motivation for savings is to protect against accidents [30]. Based on the theory, as government spending increases and the social security system improves, the precautionary saving motive becomes less crucial [51]. In addition, macroeconomics considers fiscal expenditure a form of income redistribution that can increase the household income of the poor and alleviate income imbalance. Compared with high-income households, low-income households have more potential for consumption as income increases [52]. As a result, fiscal expenditure can promote household consumption as a whole [53]. Moreover, government policies, such as offering subsidies for countryside inhabitants to buy home appliances, can directly boost household consumption. The negative indirect effect means that increased government spending can reduce consumption activities in neighboring cities, and it can be interpreted by the crowding-out theory. The residents living in a province may be attracted by the better financial, sanitation, and living conditions of nearby places. These results remind policymakers to comprehensively consider the situations of the local province and its neighbors when making full use of "the visible hand".

The direct and indirect effects of the industrial structure are positive but statistically nonsignificant. As a less energy-intensive industry, service sector is expected to greatly encourage economic development and promote consumption when expands. Upgrading of the industrial structure can also make production more effective, thus lowering the price of products. Moreover, the prosperity of the tertiary industry diversifies goods and provides households with more options. Although the coefficients of the direct and indirect effects are statistically nonsignificant, they do not suggest that industrial structure is of no significance for household consumption. We tend to interpret that the tertiary industry has a low degree of development, and the current scale of development is not enough to have effects on household consumption. Thus, to increase household consumption, it is essential to promote the upgrading of the industrial structure. The positive but nonsignificant indirect effect implies that the optimization of the industrial structure in nearby provinces can boost overall consumption, implying positive contagion impacts from industrial upgrading. The total effect is significantly positive, which implies that a motivating function on consumption exists, although the direct and indirect effects have not yet become significant.

The coefficients of the direct, indirect, and total effects of openness are −0.3343, −0.1844, and −0.5187, respectively, confirming that the degree of openness hinders the consumption behavior of households. This superficially counterintuitive result is mainly attributed to the adverse effects brought about by the international commodity trade. Import and export trade crowds the domestic market, causing consumer prices to rise and thus restricting the growth of regional household consumption. The negative coefficient of the indirect effect proves an external spillover effect on consumption activities in China. Promising import and export projects are usually rare. An increase in import and export projects in a region encourages foreign investors to seek more profitable opportunities in neighboring regions, thus influencing the neighbors' consumption. The negative effects indicate that the fair taxing and reasonable pricing policies of import goods are instrumental in promoting household consumption. However, none of the three coefficients are significant, proving that the effect of openness is relatively limited. This result may be because China has been opening up for a long period, and the effect of opening up has levelled off.

The last point relates to population density. The direct impact is significantly positive, but the indirect effect is nonsignificantly negative. The growth of the population encourages household consumption activities, but the increase in the population in surrounding areas decreases such activities, which may result from a complete consumer goods transmission system. When the number of people living in nearby provinces increases, their demands on the local province increase, and the supply they provide for the local province decreases. Overall, it is beneficial to consider the population density when making consumption promotion policies for an area.

4.6. Endogeneity Test. There may be endogeneity problems caused by reverse causality, variable omission, etc. To further address this problem, appropriate instrumental variables are required. Based on Zhang and Hu [54], we use the one-period lag of the digital financial inclusion index as its instrumental variable to perform the endogeneity test. Because the level of digital financial inclusion with a one-period lag is related to its current level, and the past level of digital financial inclusion cannot be affected by the current household consumption. Table 9 shows the results. According to the decomposition results, the coefficient of digital financial inclusion is significantly positive, and the coefficient of the

Table 9: Decomposition results of the endogeneity test.

Variables	Direct effects	Indirect effects	Total effects
Ln(DFI)	0.2112***(3.35)	−0.0951*(−1.67)	0.1162**(1.98)
Control	YES	YES	YES

Note. z-values are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

spatial lag term is significantly negative, which is consistent with the baseline regression.

4.7. Robustness Analysis. To ensure the robustness of the results, we re-estimate the SDM using a geographic distance spatial weight matrix (calculated by the reciprocal of the longitude and latitude distances between two provinces). Table 10 presents the robustness results of spatial effects based on the geographic distance spatial weight. Although there may be some inconsistency in amplitude and significance level, the direction of the impact and the main conclusions of the core independent variable remain unchanged. The necessity of adopting spatial models is proven again. The improvement in digital financial inclusion will promote household consumption in the local province and reduce household consumption in neighboring provinces.

4.8. Comparison of Spatial Effects between Urban and Rural Regions. There is an urban-rural dual structure in China, which is mainly manifested in the mode of economic development. The urban economy is based on modern large-scale industrial production, while the rural economy is based on small-scale farming. The difference contributes to the advanced infrastructure of urban areas and the consumption gap between cities and the countryside. As a form of financial inclusion, digital financial inclusion can reach consumers in remote areas who otherwise would not be covered by traditional financial institutions. Consequently, it can narrow the consumption gap and improve the urban-rural dual structural system. This section compares the spatial effect of urban and rural regions, and the results of the core in dependent variable are shown in Table 11.

As shown in the results of the subregions in Table 11, the direct effects of digital financial inclusion on the household consumption of urban and rural regions are all positively significant. The direct effect of urban areas is 0.2657, which is lower than the number of rural areas (0.3864), proving that digital financial inclusion has a smaller promoting effect on the consumption of households living in cities. A possible interpretation is that digital financial inclusion can eliminate the restrictions of financial thresholds. Owing to financial repression, the financial threshold brings about disparities in financial availability. High-income groups living in cities have substantial financial resources. They can enjoy convenient financial services because of their advantages in geography and capital accumulation. However, due to geographical disadvantages and because their capital accumulation restrictions cannot reach the wealth threshold level, low-income groups living in less-developed regions are not able to easily obtain access to financial services. They face more

Table 10: Robustness test.

Variables	Direct effects	Indirect effects	Total effects
Ln(DFI)	0.3522***(4.59)	−0.3098**(−2.29)	0.0424(0.49)
Ln(Incom)	0.7767***(5.64)	−0.4626*(−1.73)	0.3141(1.47)
Ln(Hum)	−0.0236(−0.61)	0.2979***(3.33)	0.2745***(3.26)
Trans	0.0247***(5.22)	0.0317***(3.36)	0.0564***(5.38)
Fisc	0.2292**(2.24)	−0.0451(−0.23)	0.1841(1.04)
Indus	0.3371(0.76)	0.6041(0.79)	0.9412(1.08)
Open	−0.3534(−0.62)	0.0192(0.02)	−0.3342(−0.32)
Pop	6.3964**(2.09)	−6.531*(−1.67)	−0.1348(−0.05)

Note. z-values are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 11: The effect of digital financial inclusion on urban and rural regions.

Regions	Direct effects	Indirect effects	Total effects
Urban regions	0.2657***(2.86)	−0.1818(−1.02)	0.0838(0.47)
Rural regions	0.3864***(3.76)	−0.2502(−1.59)	0.1361(1.02)

Note. z-values are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

liquidity constraints and are more impacted by the marginal effect of digital finance on consumption. Overall, digital financial inclusion can dramatically lower the cost of financial services and make financial resources more available. Therefore, low-income people can have broader and easier access to credit, and their credit restrictions can be eased, which is beneficial to promote their consumption.

However, the total effects of both areas are not significant. This may be because digital financial inclusion is still underdeveloped, so it cannot make a difference when affecting the disadvantaged people and the main public. Although digital financial inclusion has developed rapidly in recent years, it is still far from developed. Policymakers should create a friendly environment to encourage the development of digital financial inclusion and make it easier for disadvantaged people to gain access to financial services, which is the primary goal of financial inclusion.

5. Conclusion

Based on panel data for 31 provinces from 2013 to 2018 in China, we study the impact of digital financial inclusion on household consumption. We find that household consumption and the development of digital financial inclusion show geographical agglomeration. The spatial distribution of China's household consumption presents a ladder distribution from coastal to inland areas because of the unbalanced pattern of economic development. The levels of digital financial inclusion in China decline with the distance from Hangzhou, partly because of the difficulties of promoting and diffusing digital finance from Hangzhou. By constructing spatial panel models and disaggregating spatial spillover effects into the direct, indirect, and total effects, we find that digital financial inclusion has a positive direct impact and an adverse indirect impact on household consumption at the national level. Further analysis based on subsamples shows that the influence of digital financial inclusion on household consumption is more significant for rural residents than for urban residents. However, the difference is not significant because of the underdevelopment of digital financial inclusion.

Our findings have the following meaningful implications for policymakers. First, because the development of digital financial inclusion can boost household consumption and improve the dual urban-rural system, policymakers should focus on its driving impact and make no effort to promote it. However, it is not enough to consider only local conditions because provinces are increasingly related. When making policies, it is crucial to assess the condition of peripheral regions and view provinces as a whole. Third, provinces are expected to develop in a coordinated way by cooperating with surrounding provinces and realizing effective source distribution to form a balanced pattern among different provinces, resulting in a win-win situation.

The research in this paper has certain limitations. First, the data are at the provincial level, so they may have specific sample deviations and cannot reflect individual features. If micro data could be accessed in the future, researchers could compare the new outcomes using micro data with the conclusions drawn from this paper. Second, this paper uses data from 2013 to 2018, which is not very recent. A policy for making data publicly available in China is urgent. Third, as this paper involves data only on aggregate household consumption, it is advised to comprehensively investigate different types of consumer goods that households buy.

Data Availability

The data of the Digital Financial Inclusion Index used to support the findings of this study may be released upon application to the Institute of Digital Finance, Peking University, which can be contacted at guofengsfi@163.com. The household consumption data and the control variables used to support the findings of this study are released by the National Bureau of Statistics.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. T. Lai, I. K. M. Yan, X. Yi, and H. Zhang, "Digital financial inclusion and consumption smoothing in China," *China and World Economy*, vol. 28, no. 1, pp. 64–93, 2020.

[2] A. A. Levchenko, "Financial liberalization and consumption volatility in developing countries," *IMF Staff Papers*, vol. 52, pp. 237–259, 2005.

[3] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Economic Geography*, vol. 46, pp. 234–240, 1970.

[4] J. S. Duesenberry, "Income, saving, and the theory of consumer behavior," *The Review of Economics and Statistics*, vol. 33, p. 111, 1949.

[5] Y. Funashima and Y. Ohtsuka, "Spatial crowding-out and crowding-in effects of government spending on the private sector in Japan," *Regional Science and Urban Economics*, vol. 75, pp. 35–48, 2019.

[6] Y. Lv, W. Chen, and J. Cheng, "Modelling dynamic impacts of urbanization on disaggregated energy consumption in China: a spatial Durbin modelling and decomposition approach," *Energy Policy*, vol. 133, Article ID 110841, 2019.

[7] L. Klapper, D. Singer, S. Ansar, and J. Hess, *Financial Risk Management In Agriculture: Analyzing Data From A New Module Of The Global Findex Database*, The World Bank, Washington, DC, USA, 2019.

[8] D. H. Vo, N. T. Nguyen, and L. Thi-Hong Van, "Financial inclusion and stability in the Asian region using bank-level data," *Borsa Istanbul Review*, vol. 21, no. 1, pp. 36–43, 2021.

[9] W. Zhong and T. Jiang, "Can internet finance alleviate the exclusiveness of traditional finance? evidence from Chinese P2P lending markets," *Finance Research Letters*, vol. 40, Article ID 101731, 2020.

[10] P. Gomber, J.-A. Koch, and M. Siering, "Digital finance and fintech: current research and future research directions," *Journal of Business Economics*, vol. 87, no. 5, pp. 537–580, 2017.

[11] Q. Song, J. Li, Y. Wu, and Z. Yin, "Accessibility of financial services and household consumption in China: evidence from micro data," *The North American Journal of Economics and Finance*, vol. 53, Article ID 101213, 2020.

[12] J. Li, Y. Wu, and J. J. Xiao, "The impact of digital finance on household consumption: evidence from China," *Economic Modelling*, vol. 86, pp. 317–326, 2020.

[13] R. Thaler, "Toward a positive theory of consumer choice," *Journal of Economic Behavior & Organization*, vol. 1, no. 1, pp. 39–60, 1980.

[14] J. Fan, L. Tang, W. Zhu, and B. Zou, "The alibaba effect: spatial consumption inequality and the welfare gains from e-commerce," *Journal of International Economics*, vol. 114, pp. 203–220, 2018.

[15] S. P. Zeldes, "Consumption and liquidity constraints: an empirical investigation," *Journal of Political Economy*, vol. 97, no. 2, pp. 305–346, 1989.

[16] A. Deaton, *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, The World Bank, Washington, DC, USA, 1997.

[17] Y. Pierrakis and L. C. Collins, *A New Innovative Model Of Providing Funding To Projects And Businesses*, Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2013.

[18] H. E. Leland, "Saving and uncertainty: the precautionary demand for saving," *Uncertainty in Economics*, Elsevier, vol. 82, no. 3, , pp. 465–473, 1968.

[19] J. Grossman and M. Tarazi, *Serving Smallholder Farmers: Recent Developments in Digital Finance*, CGAP, Washington, DC, USA, 2014.

[20] F. Guo, S. T. Kong, and J. Wang, "General patterns and regional disparity of internet finance development in China: evidence from the peking university internet finance development index," *China Economic Journal*, vol. 9, no. 3, pp. 253–271, 2016.

[21] X. Liu, J. Zhu, J. Guo, and C. Cui, "Spatial association and explanation of China's digital financial inclusion development based on the network analysis method," *Complexity*, vol. 2021, Article ID 6649894, 13 pages, 2021.

[22] G. Feng, K. Tao, and W. Jingyi, "The spatial agglomeration effect of internet finance-evidence from the financial development index," *Studies of International Finance*, vol. 8, 2017.

[23] B. Zhu, J. He, and S. Zhai, "Does financial inclusion create a spatial spillover effect between regions? Evidence from China," *Emerging Markets Finance and Trade*, vol. 55, no. 5, pp. 980–997, 2019.

[24] Y. Su, Z. Li, and C. Yang, “Spatial interaction spillover effects between digital financial technology and urban ecological efficiency in China: an empirical study based on spatial simultaneous equations,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 16, p. 8535, 2021.

[25] X. Wang and J. Guan, “Financial inclusion: measurement, spatial effects and influencing factors,” *Applied Economics*, vol. 49, no. 18, pp. 1751–1762, 2017.

[26] J. Huang, W. Zhang, and W. Ruan, “Spatial spillover and impact factors of the internet finance development in China,” *Physica A: Statistical Mechanics and Its Applications*, vol. 527, Article ID 121390, 2019.

[27] G. Feng, W. Jingyi, W. Fang, K. Tao, Z. Xun, and C. Zhiyun, “Measuring China’s digital financial inclusion: index compilation and spatial characteristics,” *China Economic Quarterly*, vol. 19, pp. 1401–1418, 2020.

[28] Y. Shen, C. J. Hueng, and W. Hu, “Measurement and spillover effect of digital financial inclusion: a cross-country analysis,” *Applied Economics Letters*, vol. 28, pp. 1–6, 2020.

[29] J. Álvarez-Gamboa, P. Cabrera-Barona, and H. Jácome-Estrella, “Financial inclusion and multidimensional poverty in Ecuador: a spatial approach,” *World Development Perspectives*, vol. 22, Article ID 100311, 2021.

[30] C. D. Carroll, “How does future income affect current consumption?” *Quarterly Journal of Economics*, vol. 109, no. 1, pp. 111–147, 1994.

[31] O. P. Attanasio and G. Weber, “Is consumption growth consistent with intertemporal optimization? evidence from the consumer expenditure survey,” *Journal of Political Economy*, vol. 103, no. 6, pp. 1121–1157, 1995.

[32] W. Zhao, “Does health insurance promote people’s consumption? new evidence from China,” *China Economic Review*, vol. 53, pp. 65–86, 2019.

[33] L. Anselin, “Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity,” *Geographical Analysis*, vol. 20, pp. 1–17, 1988.

[34] G. Qi, W. Shi, K.-C. Lin, K. F. Yuen, and Y. Xiao, “Spatial spillover effects of logistics infrastructure on regional development: evidence from China,” *Transportation Research Part A: Policy and Practice*, vol. 135, pp. 96–114, 2020.

[35] J. A. Donaldson, *China’s Administrative Hierarchy: The Balance of Power and Winners and Losers within China’s Levels of Government*, Routledge, Oxfordshire, Uk, 2017.

[36] A. D. Cliff and J. K. Ord, *Spatial Processes: Models & Applications*, Taylor & Francis, Oxfordshire, Uk, 1981.

[37] P. A. P. Moran, “Notes on continuous stochastic phenomena,” *Biometrika*, vol. 37, no. 1-2, pp. 17–23, 1950.

[38] H. Li, C. A. Calder, and N. Cressie, “Beyond Moran’s I: testing for spatial dependence based on the spatial autoregressive model,” *Geographical Analysis*, vol. 39, no. 4, pp. 357–375, 2007.

[39] H. Wang, L. Ding, R. Guan, and Y. Xia, “Effects of advancing internet technology on Chinese employment: a spatial study of inter-industry spillovers,” *Technological Forecasting and Social Change*, vol. 161, Article ID 120259, 2020.

[40] L. Anselin, A. K. Bera, R. Florax, and M. J. Yoon, “Simple diagnostic tests for spatial dependence,” *Regional Science and Urban Economics*, vol. 26, no. 1, pp. 77–104, 1996.

[41] J. P. Elhorst, “Applied spatial econometrics: raising the bar,” *Spatial Economic Analysis*, vol. 5, no. 1, pp. 9–28, 2010.

[42] C. Cheng, Y. Hua, and D. Tan, “Spatial dynamics and determinants of sustainable finance: evidence from venture capital investment in China,” *Journal of Cleaner Production*, vol. 232, pp. 1148–1157, 2019.

[43] R. K. Pace and J. P. LeSage, “A sampling approach to estimate the log determinant used in spatial likelihood problems,” *Journal of Geographical Systems*, vol. 11, no. 3, pp. 209–225, 2009.

[44] L. Anselin and J. Le Gallo, “Interpolation of air quality measures in hedonic house price models: spatial aspects,” *Spatial Economic Analysis*, vol. 1, no. 1, pp. 31–52, 2006.

[45] H. H. Kelejian and I. R. Prucha, “HAC estimation in a spatial framework,” *Journal of Econometrics*, vol. 140, no. 1, pp. 131–154, 2007.

[46] L. Yang and Y. Zhang, “Digital financial inclusion and sustainable growth of small and micro enterprises-evidence based on China’s new third board market listed companies,” *Sustainability*, vol. 12, no. 9, p. 3733, 2020.

[47] J. M. Keynes, “The general theory of employment,” *Quarterly Journal of Economics*, vol. 51, no. 2, pp. 209–223, 1937.

[48] M. Friedman, “The permanent income hypothesis,” in *A Theory of the Consumption Function*, pp. 20–37, Princeton University Press, Princeton, NJ, USA, 1957.

[49] E. Grunberg and F. Modigliani, “The predictability of social events,” *Journal of Political Economy*, vol. 62, no. 6, pp. 465–478, 1954.

[50] A. Zhang and H. Chen, “Evolution of China’s air transport development and policy towards international liberalization,” *Transportation Journal*, vol. 42, no. 3, pp. 31–49, 2003.

[51] M. Browning and A. Lusardi, “Household saving: micro theories and micro facts,” *Journal of Economic Literature*, vol. 34, pp. 1797–1855, 1996.

[52] C. Binkai, “Income inequality and consumption demand: theory and evidence from China,” *Nankai Economic Studies*, vol. 1, 2012.

[53] J. Galí, J. D. López-Salido, and J. Vallés, “Understanding the effects of government spending on consumption,” *Journal of the European Economic Association*, vol. 5, pp. 227–270, 2007.

[54] W. Zhang and Y. Hu, “The effect of innovative human capital on green total factor productivity in the Yangtze River Delta,” *China population, resources and environment*, vol. 30, pp. 106–120, 2020.

Hindawi
Complexity
Volume 2022, Article ID 7353462, 14 pages
https://doi.org/10.1155/2022/7353462

Research Article

The Evolution of China's High-Level Talent Mobility Network: A Comparative Analysis Based on School and Work Stage

Fangfang Zhang,[1,2,3,4] **Hui Liu**,[1,2,3] **Juntao Zhang**,[5] **and Yi Cheng**[1,2,3]

[1]*Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*
[2]*Key Laboratory of Regional Sustainable Development Modeling, Chinese Academy of Sciences, Beijing 100101, China*
[3]*College of Resource and Environment, University of Chinese Academy of Sciences, Beijing 100049, China*
[4]*Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*
[5]*School of Computer Science, Wuhan University, Wuhan 430072, China*

Correspondence should be addressed to Hui Liu; liuh@igsnrr.ac.cn

Received 24 September 2021; Revised 8 December 2021; Accepted 21 December 2021; Published 24 January 2022

Academic Editor: Giacomo Fiumara

The flow of high-level talents in China is becoming more and more active and has a profound impact on the innovation and development of cities. Based on the research on the spatial distribution of the school and work stages of high-level talent in China, we study the structure and evolution of China's high-level talent mobility network by using the data on the flow trajectory of talent among cities in China. The results show that: (1) the spatial distribution of talents in the two stages has greater consistency, but there is a change characteristic of "large dispersion, small agglomeration." (2) Compared with the school stage, the network in the work stage has much more connections; the centrality gap of cities has narrowed. (3) "Core-periphery" structure of two networks is relatively significant and similar. The core layer in the two networks has not changed, but the subcore layer and the subedge layer have undergone changes. In addition, talents mainly flow to core and subcore cities. (4) The space structure of network is more complex in the work stage, and it is expanding to the Guangdong-Hong Kong-Macao Greater Bay Area in southeastern China.

1. Introduction

Talent has become a crucial strategic resource in China's economic and social development [1]. The spatial layout and flow efficiency of high-level talents (HLT) to a large extent characterize the regional economic development status [2] and represent the distribution of the engine of the science and technology industry [3]. Large-scale talent flow is bound to have a significant impact on China's regional innovation capabilities, economic system structure, industrial layout, population distribution structure, and social and cultural environment [4, 5]. Therefore, studying the state of HLT is crucial to the overall understanding of innovation and economic development [3, 6].

At present, China's rapid technological innovation, increasing progress in transportation technology, fast development of information and communication technology, and the combined effect of multiple factors have provided great convenience for the flow of HLT across the country. At the same time, local governments have introduced a series of policies and measures to attract talents, which accelerate the flow of HLT. However, the unhealthy competition for talent introduction among cities has aggravated the irrational flow of talents, leading to an uneven regional distribution of talents. Issues such as "Peacocks Flying Southeast," "Crisis in the Midwest," and "Dilemma in the Northeast" have emerged in China [7]. In April 2017, the Ministry of Science and Technology issued the "Thirteenth Five-Year Plan for the Development of Scientific and Technological Talents" (Ministry of Science and Technology, PRC. The 13th Five-Year Plan for The Development of Scientific and technological Talents, 2017.

Available at: http://www.casec.org.cn/article/show%20article.%20asp?articleid=728), proposing to remove obstacles to the flow of talents and allow free flow of scientific and technological talents in accordance with market laws. In September 2021, General Secretary Xi Jinping pointed out at the Central Conference on Talent work (http://www.gov.cn/xinwen/2021-09/28/content_5639868.htm?version=2.5.40020.452&platform=win) that Beijing, Shanghai, and the Guangdong-Hong Kong-Macao Greater Bay Area could be built into HLT hubs, and some central cities with HLT concentration should also focus on building platforms to attract and gather talent. Factors such as regional development differences, unhealthy talent competition among cities, and government policy guidance will inevitably have an important impact on the flow of HLT in China. In this context, there is an urgent need to thoroughly study the spatial distribution and flow characteristics of China's HLT, so as to provide support for the government to formulate scientific policies for attracting talents.

In recent years, the study of talent flow in China has become a hot topic. Relevant researches mainly focus on the current situation of talent flow [8–10], model of talent flow [11, 12], influencing factors of talent flow [13], and the impact of talent flow on scientific research output [14]. Although scholars have studied talent flow in China, there are still some problems that need to be remedied. First, there is a lack of quantitative research on the domestic flow of HLT in China. Due to the limitation of data acquisition, the study of HLT flow in China has always been a difficult problem. Most of the existing quantitative studies only studied the geographical flow characteristics of a specific type of talents (academicians, outstanding young scholars, or Changjiang scholars) [8, 9, 15]. These studies have great limitations, and the results cannot represent the flow of HLT in China in a broader sense [9, 15, 16]. Second, few studies analyzed and compared talent flow from the growth stage of talent [17]. Due to the distribution of national education and work resources, regional development, and individual factors of talents, the flow of talents in the stage of education and work has different characteristics, and the reasons and mechanisms of the flow are also different. Therefore, it is of great practical significance to study the characteristics of the flow of HLT among cities. It can also enrich the relevant researches on the flow of talents in China and provide reference for the formulation of talent attraction policies for cities in China.

This article aims to compare and analyze the spatial distribution, flow characteristics, and differences of China's HLT in the school and work stage. The article has six parts. The first part is introduction. The second part reviews the relevant research on the flow of HLT in China and discusses how to study the flow of HLT based on the theory of networks and mobility spaces. The third part introduces the data sources and research methods. The fourth part is the result analysis, which is elaborated from four aspects: the spatial distribution characteristics of HLT, the changes in the basic attributes of the talent flow network, and the changes in the "core-edge" structure and spatial structure of the talent flow network. The fifth part further discusses the changing characteristics of the spatial distribution of talents, the characteristics of talent mobility (bidirectional, stickiness, and Matthew effect), and the spatial heterogeneity of talent mobility networks. Finally, the main conclusions and contributions of this article are briefly summarized.

2. Space of Flows, Network, and Talent Flow

The concept of "flow" has emerged in the context of economic globalization and informatization, which are reshaping the organization structure of the world—what Castells calls "the rise of network society" [18]. The salient feature of this new organizational structure is the dominant position of various "flows," such as population flow, information flow, capital flow, and technology flow. "Flow" and "space" are inseparable. Manuel Castells proposed that "space of flows" means "physical organization of social practices that share time without geographical adjacency" [19]. Its structure consists of four parts: "(1) It is made up of a technological infrastructure of information systems, telecommunications, and transportation lines. (2) The space of flows is made up of nodes and hubs. (3) The space of flows is also made of habitats for the social actors that operate the networks. (4) The space of flows comprises electronic spaces" [20].

Ma summarized and proposed that "flow" in various forms between cities is a "soft network" (which can be divided into economic, political, social, and cultural connection functions) based on the "hard network" (such as highway, railway, airport, telecommunications, and other infrastructure) between cities, which are expressed in multiple forms of connection among human, materials, and information [21–23]. Talent mobility network between cities can be regarded as "soft network" based on "hard network." Talent flow is always associated with specific cities [24]. As the main place of knowledge innovation and technology diffusion, city is one of the important network nodes, which belongs to the second part of the structure of "space of flows."

At the end of the 20th century, the complex network theory based on mathematical graph theory and statistical physics provided a theoretical basis for studying the complexity of network systems [25–27]. Complex network theory and methods have been used in many fields such as disease transmission [28], transportation networks [29], scientific research cooperation networks [30], and urban networks [31, 32]. This method also provides a new perspective for the identification of talent flow network structure. It usually abstracts the overall flow of talents into a network and applies network analysis methods to carry out related research. With the development of big data technology, the data of "talent flow" have become available. Scholars have begun to try to collect information on the flow of talents online, mining trajectory data of talents from the resume, published papers, patents, and other information at various stages of learning, work, etc. [33]. On the basis of big data mining, scholars usually apply various characteristics and indicators of complex networks to the relevant analysis of talent flow networks. For example, there is a network of the talent flow composed of scientific research leaders [27], scientists [34], corporate talents [35, 36], overseas students [37], and returned talents [38].

There have been studies using the talent flow network to explore the spatial characteristics and trends of talent flow, and to predict and simulate the trend of talent flow. Fagiolo and Mastrorillo obtained relevant conclusions about the trends and characteristics of global migration in the world today by studying international migration networks and predicted the trend of international migration in the future [39]. Wang et al. [40, 41] constructed a global technology leader talent flow network model to measure and analyze the characteristics of the global technology leader talent flow in major countries (regions) and the world as a whole. Zhu Pengcheng [36] conducted research on the geographical network of talent flow in the Yangtze River Delta and explored the direction and activity of talent flow in the network. Shi Wentian [34] found that Beijing and Shanghai located in the core area of the scientific research cooperation network of Chinese scientists.

In addition, scholars have also begun to portray different types of urban networks or innovation networks in terms of talent flow. In the era of knowledge economy, the importance of human capital in economic and social development has become increasingly prominent, and the flow and agglomeration of talents has become an important driving force for the evolution of the world's urban network system and the development of global cities. In the study of the world city network based on the flow of "highly skilled international migrants" [42, 43], scholars believe that the processes of talent management and global staffing are nourishing the knowledge base of world cities. It provides a continuous flow of talents for these places, ensuring that they achieve a high ranking in the global city hierarchy [43]. Moreover, using talent mobility to build innovative relationships between cities can better reflect the connection of tacit knowledge than papers and patents, and HLT can reflect the high-level part of tacit knowledge, which has an important value for simulating the flow of high-level knowledge and innovative connections between cities [21]. Relevant studies have found that HLT can also establish important knowledge connections for cities where they live, study, or work for a long period of time, and built innovation network between cities by using China's "Thousand Talents Plan" talents and important entrepreneurs' migration [38].

Network science and methods provide a strong support for the research on the structural characteristics and evolution of talent regional flow network. This paper tries to use the talent flow network to answer the following three core questions: first, compare the characteristics of node attributes in the talent flow network and analyze the characteristics of talent flow between cities. Second, by comparing the core-edge structure of the network, this paper discusses whether there is Matthew effect in the current flow of HLT in China. Third, the spatial heterogeneity of talent flow network is analyzed by comparing its spatial structure.

3. Data and Methods

3.1. Definition of High-Level Talents and Data Sources. According to the definition of talents in talent science, talents are people who have certain knowledge and skills under certain historical conditions and use their creative work to make greater contributions to society or a certain aspect of society [44]. HLT usually refers to a small number of academic elites who produce knowledge, enjoy a high reputation, and lead the development of disciplines [45]. Chinese city governments have slightly different identification systems for HLT, but they generally divide talents into three categories: domestic and foreign top talents, national-level leading talents, and local-level leading talents. At present, a number of high-level academic groups have formed, including academicians of the Chinese Academy of Sciences and Chinese Academy of Sciences, Changjiang Scholars, Outstanding young people, and so on. Considering the availability of data, the HLT in this article refers to the academicians of the Chinese Academy of Sciences (ACAS), the academicians of the Chinese Academy of Engineering (ACAE), the Chinese academicians of the International Eurasian Academicians (IAEA), the Chinese academicians of the Academicians of Developing Countries (TWAS), the scholar of the Changjiang Scholars Program (CJSP), and the scholar of the National Outstanding Youth Fund Project (NOYFPS).

The election of the academicians of the CAS began in 1955. The earliest year of the academicians still living in the datais 1980. The election of the CAE began in 1994, and the earliest year of the living academicians is 1994. The earliest Chinese academicians of IAEA and TWAS, still alivingis 1994 and 1983, respectively. The Changjiang Scholars Program was established in 1998, and the National Science Fund for Outstanding Youths was established in 1994. Therefore, the research period of this article is 1980–2020.

The talent data in this paper are all obtained from the official website. Among them, the data of ACAS are obtained from the official website of the Chinese Academy of Sciences (http://casad.cas.cn/ysxx2017/ysmdyjj/qtysmd_124280/), with a total of 812 talents. The data of ACAE are obtained from the official website of the Chinese Academy of Engineering (https://www.cae.cn/cae/html/main/col48/column_48_1.html), with a total of 911 talents. The IAEA and TWAS were obtained from the official website (http://bj.ieaschina.org/expert/jieshao.html; https://www.cas.cn/zt/hyzt/twas23/abouttwas/201208/t20120808_36%2025109.html), respectively, with a total of 378 talents. The source of the information on the CJSP is with the Ministry of Education of the People's Republic of China (http://www.moe.gov.cn/jyb_xxgk/s5743/s5745/), with a total of 3143 talents. The data of NOYFPS came from the National Natural Science Foundation of China (https://www.nsfc.gov.cn/), with a total of 4303 talents. Because some talents have multiple titles at the same time, this article merges the data of such talents into one sample. For example, if a scientist has successively obtained the titles of CJSP, TWAS, and ACAS, then we will retain the last title of ACAS for this scientist, but his data information contains the titles of CJSP and TWAS. In addition, this article does not include foreign talents from 1980 to 2020. In the end, this article collected information on 7734 talents.

3.2. Talent Flow Trajectory Data Processing. The research area of this article is China (as some talents have study or work experience in Hong Kong, Macau, Taiwan, and other regions, the above regions are also included in this research), and the basic research unit is prefecture-level cities. The talent flow trajectory data are mainly judged by the change of the working city. If the working city of the talent changes, we judge that the talent flow has occurred [8, 41]. The specific steps are as follows: (1) Define the city of the talent's first school as the "first city of school" and then count the city where the school is located in chronological order. When the city changes, it is defined as the "second, third, . . ., school city" of the talent. On this basis, determine the talent flow trajectory based on prefecture-level cities (first school city ⟶ second school city ⟶ third school city, etc.). If the city where the talent school is located has not changed, it will not be included in the flow trajectory. (2) Define the city of the talent's first workplace as the "first city of workplace" and then count the city where the workplace is located in chronological order. When the city changes, it is defined as the "second, third, ... city of workplace" of the talent. On this basis, determine the talent flow trajectory based on prefecture-level cities (first city of workplace ⟶ second city of workplace ⟶ third city of workplace, etc.).

3.3. Network Construction. Drawing lessons from network graphics and matrix expression methods, create the talent flow network of this article [27, 46]. The talent flow network is a directed complex network. In this paper, cities are regarded as nodes in the network, and the direction of talent flow is regarded as edges. To build a networked model, talent flow is along the direction of the edges between nodes (Figure 1). To construct a network matrix of talent flow between cities, the ordinate of the matrix is set as the outflow city, and the inflow city is set as the abscissa of the matrix. The value of the element in the matrix is the number of high-level talents flowing from the out-flow city to the in-flow city.

3.4. Centrality Analysis of Network Nodes

3.4.1. Degree Centrality. This refers to the number of other nodes directly connected to the target node in the network and represents the degree of node connection [46]. In this article, the degree centrality of a node indicates the number of cities that have a talent flow connection with the city. Degree centrality is divided into in-degree centrality and out-degree centrality. Among them, the in-degree centrality can reflect the centrality of a city as a city for attracting talents in the talent flow network. The out-degree centrality reflects the centrality of a city as a talent outflow city in the talent flow network [47]. It is calculated as follows:

$$C_D(i) = \sum_{j=1, j \neq i}^{n} k_{ij} + \sum_{j=1, j \neq i}^{n} k_{ji}. \quad (1)$$

Here, $C_D(i)$ is the degree centrality of the node, which refers to the number of other nodes connected to the node i, k_{ij} represents the outflow of the node, and k_{ji} represents the inflow of the node.

3.4.2. Strength Centrality. This refers to the sum of the weights of the edges formed by all other nodes associated with a node in the network [48, 49]. It represents the connection scale of nodes in the network. In this paper, strength centrality is the sum of the flow of talents between a certain city and all other cities:

$$S_i = \sum_{j=1, j \neq i}^{N} C_{ij}. \quad (2)$$

Here, S_i is the strength centrality of city i, that is, the sum of the flow of city i and other cities; j is the other cities connected to city i; N is the number of cities; and C_{ij} is the flow quantity of talent between city i and city j.

3.4.3. Betweenness Centrality. This measures how much a point lies "in the middle" of other points in the graph. If a point is on the shortcut of other point pairs, we say that the point has a high intermediate centrality [46]. The calculation formula for the centrality of the directed network is as follows:

$$C_B(i) = \frac{1}{2} \sum_{j}^{n} \sum_{k}^{n} \frac{g_{jk}(i)}{g_{jk}}, \quad j \neq k \neq i, \text{ and } j < k, \quad (3)$$

where g_{jk} is the number of shortest paths used to connect point j and point k and $g_{jk}(i)$ is the number of shortest paths through point i between point j and point k.

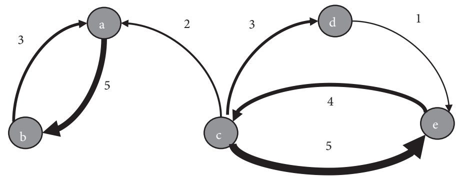
3.4.4. Comprehensive Centrality [30]. The first three node centrality indicators in the article all measure the rights of nodes in the network from a single aspect and lack the comprehensive centrality measurement of the connections between nodes. Therefore, this article draws on the comprehensive centrality model to comprehensively evaluate each node, which is calculated as follows:

$$z_i = \rho_1 \times \text{Stand}(C_D) + \rho_2 \times \text{Stand}(S_i) + \rho_3 \times \text{Stand}(C_B), \quad (4)$$

where z_i is the comprehensive centrality and $\text{Stand}(C_D)$, $\text{Stand}(S_i)$, and $\text{Stand}(C_B)$ are the standardized values of degree centrality, strength centrality, and betweenness centrality, respectively. ρ_1, ρ_2, and ρ_3 are the weights of the three indicators, respectively. This article believes that the three indicators are equally important, so the weights are all set to 1/3.

3.5. Network Visualization Method

3.5.1. Visualization of Core-Edge Structure. Based on the value of comprehensive centrality, this paper uses the Jenks method (a method of classification) to divide the network



Out-flow \ In-flow	a	b	c	d	e
a	0	5	0	0	0
b	3	0	0	0	0
c	2	0	0	3	5
d	0	0	0	0	1
e	0	0	4	0	0

Multivalued directed network matrix

FIGURE 1: Multivalued network graph and multivalued network matrix.

into four levels of "core, subcore, subedge, and edge" and draw it with the help of Ucinet, Pajek. The color of the city nodes in Figure 2 represents the different levels of comprehensive centrality. Because there are too many city nodes, the city structure diagram only shows the core layer, the subcore layer, and the subedge layer.

3.5.2. Visualization of Space Structure. Taking the flow trajectory of talent between cities as the core connection of the provincial talent mobility network, this paper draws the spatial structure of the talent mobility network by using ArcGIS. The size of the node in Figure 3 indicates the number of talent flowing through the city. The larger the node, the more mobile talents. The thickness of the edge in Figure 3 indicates the amount of mobility between the two cities. The thicker the edge, the more the amount of mobility.

4. Results

4.1. Spatial Distribution and Flow of HLT. The distribution of talents in China has always shown a strong spatial imbalance [50]. At the school stage, the HLT are distributed in 141 cities (Figure 4), with an average of 97 talents in each city. There are 22 cities above average, and the top 10 cities are Beijing, Shanghai, Nanjing, Wuhan, Xi'an, Hefei, Hangzhou, Harbin, Changchun, and Tianjin. There are 212 cities where HLT in China are located in Figure 4, with an average of 63 talents in each city. There are 24 cities exceeding the average, and the top 10 cities are Beijing, Shanghai, Nanjing, Wuhan, Guangzhou, Hong Kong, Xi'an, Hangzhou, Hefei, and Tianjin. Cities that attract HLT to choose jobs are mainly located in the east of China, followed by the central region, and very few in the west.

Comparing the spatial distribution of talents, the spatial distribution of talents in the work stage has been significantly expanded. However, the spatial distribution of talents in the two stages is still relatively consistent. They are mainly distributed in the eastern part of China, along the Yangtze River valley, and central and western provincial capital cities. These cities are also the main locations of China's 985 and 211 universities (Figure 4). It can be seen that this spatial distribution of HLT is closely related to the spatial distribution of national universities, laboratories, and innovation platforms.

The data show that at the school stage, 2047 talents (26.47%) moved once, that is, they changed to a master's or doctoral school; 303 people (3.92%) moved twice, that is, both master and doctoral schools changed. In comparison, a total of 2,348 people moved during the work stage. The highest number of flows reached 9 times. Among them, 2914 people (22.12%) moved 1 time; 1365 people (17.65%) moved twice; and 566 people moved 3 times or more (37.68%). From the school stage to the work stage, the number of talent flows increased from 2,653 to 4341, and the flow of talents became more active. However, cities with a large amount of talent flow are concentrated in a few cities (Table 1). The results show that the top five city pairs for the flow of HLT during the school stage are Wuhan ⟶ Beijing, Nanjing ⟶ Beijing, Xi'an ⟶ Beijing, Hefei ⟶ Beijing, and Shanghai ⟶ Beijing. The top five cities for the flow of high-level talents at the work stage are Beijing ⟶ Shanghai, Beijing ⟶ Hong Kong, Shanghai ⟶ Beijing, Hong Kong ⟶ Beijing, and Beijing ⟶ Wuhan.

4.2. Node Centrality of the Network

4.2.1. Degree Centrality. In the school stage, Beijing has the highest indegree centrality, followed by Shanghai. The indegrees of the top 15 cities are all greater than the outdegree (Figure 5). These cities have high-quality educational resources and are favored by talent. At the work stage, Beijing's indegree is still the highest, but Beijing's outdegree and indegree are almost the same. Beijing's high housing prices and competitive pressure may have contributed to the outflow of some young talents. In addition, some talents work part time in areas outside Beijing, which makes Beijing's outdegree relatively high. The outdegree of Changsha and Harbin is higher than the indegree, and they are the cities with more serious brain drain in China. In contrast, first-tier cities such as Beijing, Shanghai, Guangzhou, Nanjing, Wuhan, and Hangzhou all have higher indegree than outdegree and are key cities for absorbing talent.

4.2.2. Strength Centrality. The results (Figure 6) show that in the school stage, the strength centers of cities such as Beijing, Shanghai, Nanjing, and Wuhan are at the forefront, indicating that these cities are the preferred cities for talent in the process of studying. Moreover, these cities are regions with extremely rich educational resources, which promotes the frequent flow of talents between these cities. Although

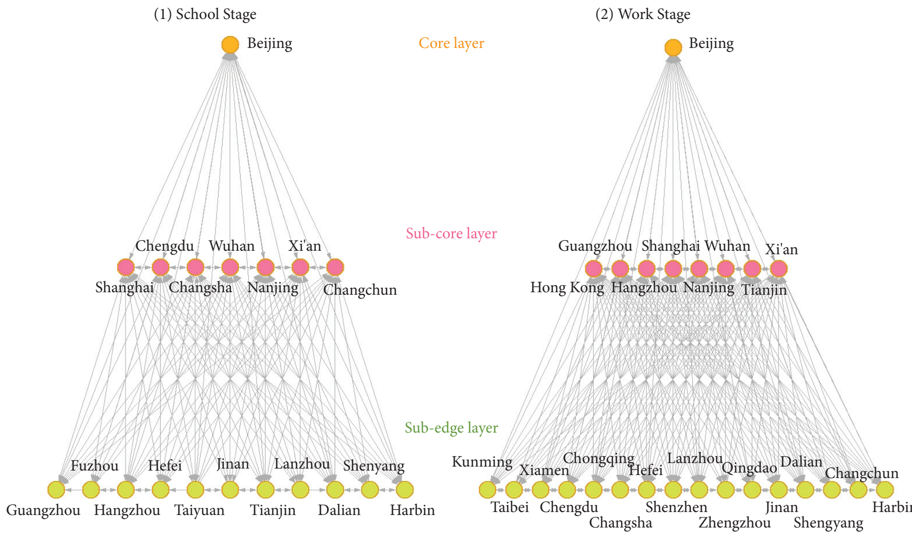


Figure 2: Hierarchical structure diagram of the talent mobility network.

Table 1: Top 10 city pairs of the number of talent flows.

School stage			Work stage		
Outflow city	Inflow city	The number of talent flows	Outflow city	Inflow city	The number of talent flows
Wuhan	Beijing	115	Beijing	Shanghai	121
Nanjing	Beijing	83	Beijing	Hong Kong	85
Xi'an	Beijing	62	Shanghai	Beijing	65
Hefei	Beijing	60	Hong Kong	Beijing	62
Shanghai	Beijing	59	Beijing	Wuhan	57
Changchun	Beijing	52	Beijing	Guangzhou	56
Chengdu	Beijing	51	Beijing	Nanjing	56
Jinan	Beijing	51	Wuhan	Beijing	56
Lanzhou	Beijing	44	Beijing	Hefei	51
Nanjing	Shanghai	43	Nanjing	Beijing	50

Guangzhou has a high degree of centrality, its strength centrality is low. Compared with cities such as Beijing, Shanghai, Wuhan, and Nanjing, Guangzhou's educational resources are relatively weak. In the work stage, Beijing's strength centrality is still the highest, and the number of connections in Beijing is almost the total number of connections in Shanghai, Nanjing, Wuhan, Xi'an, and Hangzhou. In addition, despite the massive increase in the number of nodes in the network during the work phase, there are only 21 cities with a strength centrality higher than 100, which shows that the destination of talent flow has a strong agglomeration.

4.2.3. Betweenness Centrality. In the school stage, the ranking of the centrality of cities is quite different. Cities such as Beijing, Shanghai, Nanjing, Wuhan, and Xi'an have relatively high betweenness centrality, and they have a strong ability to control the resources in the network. As an important gathering place for educational resources, these nodes play an intermediary role. During the work stage, Beijing's centrality is extremely high, indicating that no matter whether it outflows or inflows, a larger part of the work place of talent includes Beijing.

From the stage of school to the stage of work, the scale of the network has expanded greatly. The number of city nodes in the network has increased from 142 to 216. From the perspective of the changes in node centrality (Figure 6), the nodes and their projections are relatively scattered, and the centrality of each node has a large gap during the school stage. During the work stage, the centrality of Beijing and Shanghai is still much higher than that of other cities. However, other city nodes have become more concentrated, the centrality gap between nodes has narrowed, indicating that the network is more mature in the work stage, and the role of city nodes is more prominent.

4.3. Core-Edge Structure of Network

4.3.1. School Stage (Figure 2). (1) Core layer: Beijing. As a national science and technology innovation center, Beijing has the largest number of universities and research institutions in China, attracting a large number of high-level talents, becoming the core of the network, and occupying a dominant position in the network. And the resources of the edge layer continue to flow to the core, which will further strengthen the position of the core node. (2) Subcore layer:

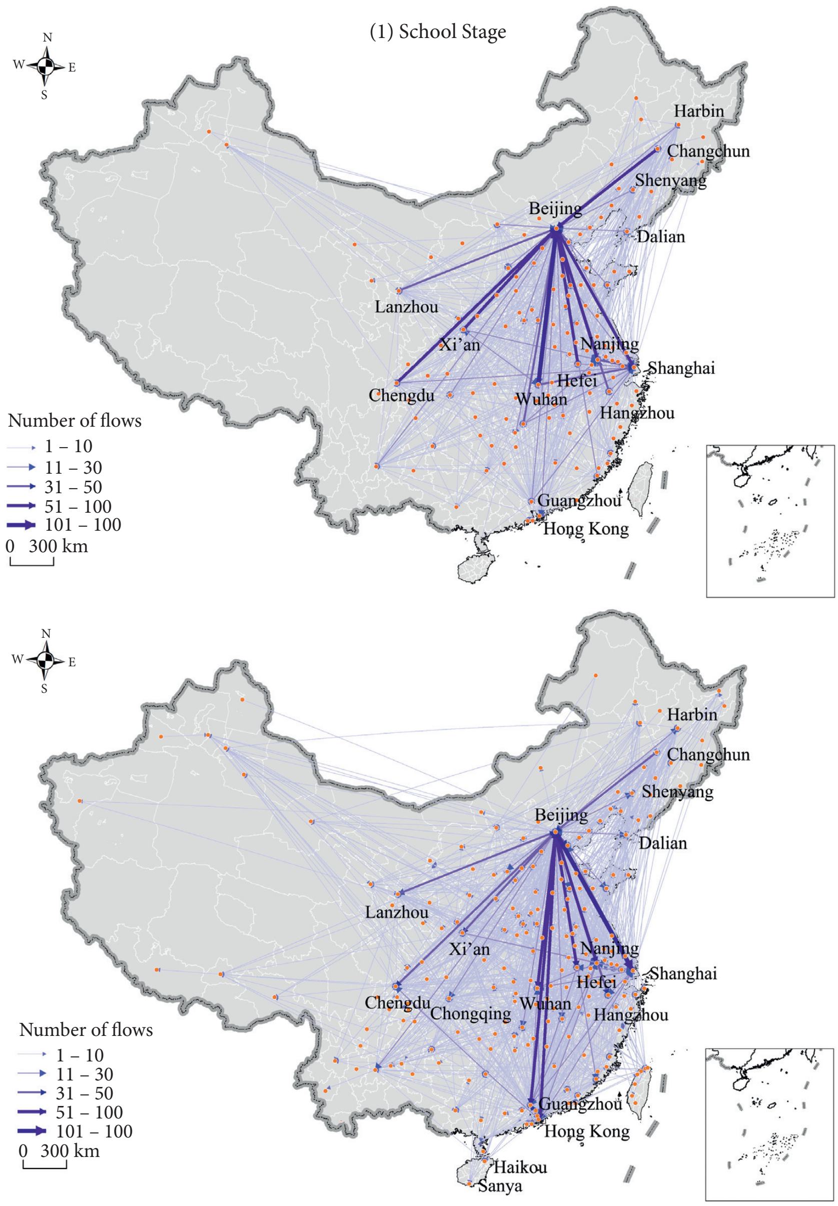


Figure 3: Changes in the spatial structure of the talent mobility network. (a) School stage. (b) Work stage.

Shanghai, Nanjing, Chengdu, Wuhan, Xi'an, Changchun, and Changsha. As the secondary core, cities in this level are closely connected to other cities. (3) Subedge layer: Lanzhou, Hefei, Harbin, Dalian, Tianjin, Taiyuan, Guangzhou, Hangzhou, Shenyang, Jinan, and Fuzhou. Although the average comprehensive centrality of this level of nodes is not as high as that of the first two levels, it is also a relatively important node for leading the connections between cities of different levels. (4) Edge layer: Shangrao, Dongying, Linfen, Linyi, and Urumqi and other 123 cities. The nodes in this level have low connections with other nodes and are located at the edge of the network, which forms an extremely unbalanced development pattern on the core.

4.3.2. Work Stage (Figure 2). (1) Core layer: Beijing. Beijing is firmly at the core and highest level of the network. As an important economic, educational, and cultural center of China, Beijing is very attractive to high-level talents. (2) Subcore layer: Shanghai, Nanjing, Tianjin, Guangzhou, Hangzhou, Wuhan, Xi'an, and Hong Kong. The subcore cities are the more important nodes in the network. Most of these cities are located in coastal areas, with developed economy, high education level, superior working environment, and suitable natural environment. These cities have created good conditions for the influx of talent and become the main influx destination of talents. Although Wuhan and Xi'an are located in the central and western regions,

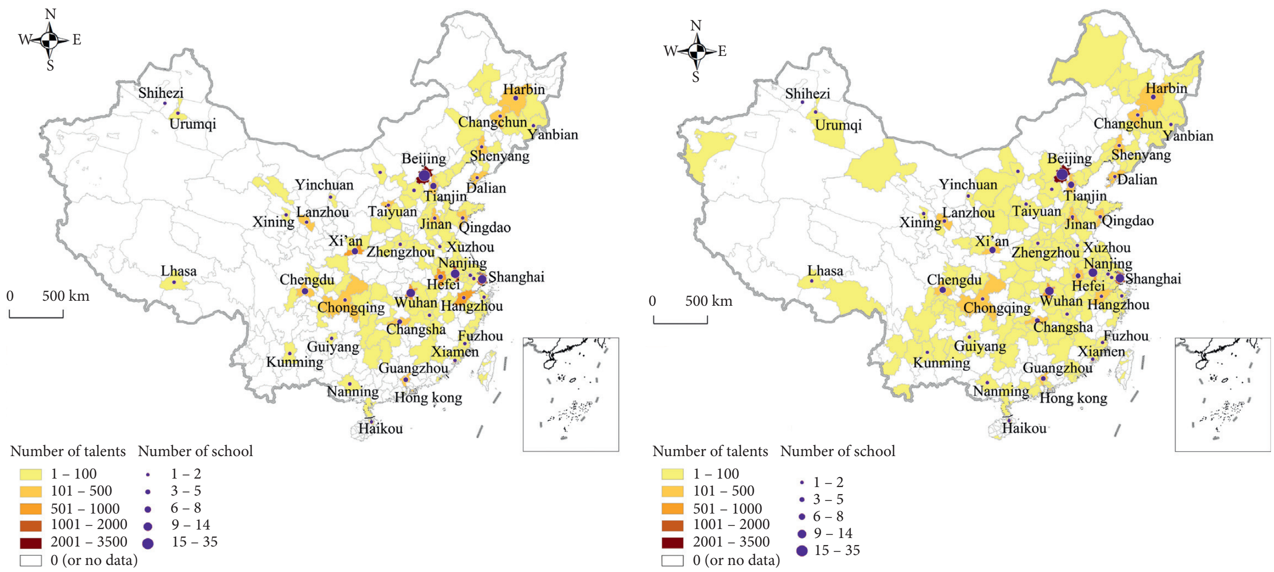


Figure 4: The spatial distribution map of China's high-level talents 985 and 211 universities.

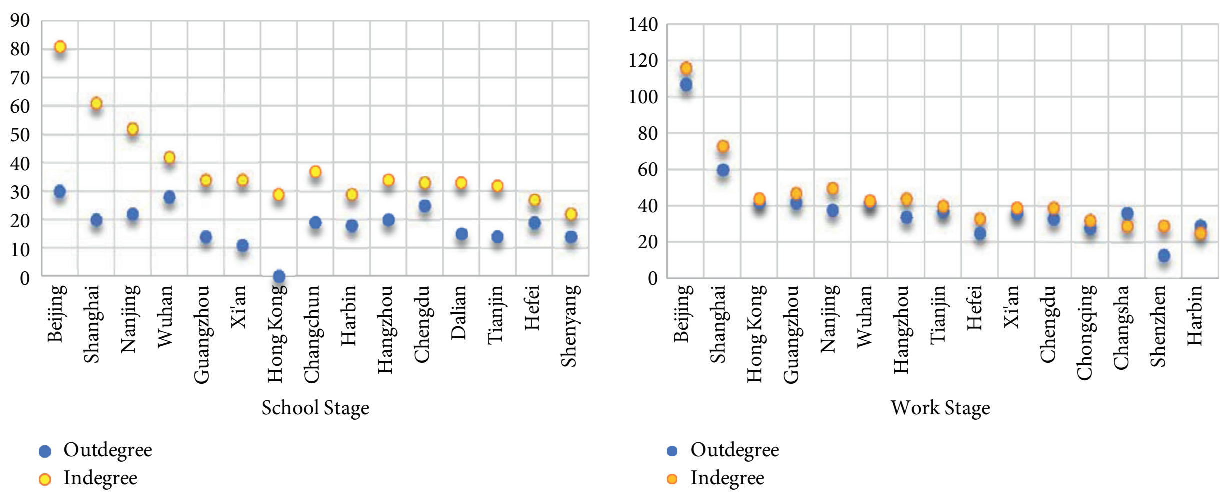


Figure 5: Node degree centrality at different stages. Note: only the top 15 nodes with degree centrality are displayed.

respectively, the two cities are not only provincial capitals but also gateway cities in the central and western regions. They also have a large number of scientific research platforms and profound cultural heritage. Therefore, it has also attracted a large influx of talents and has become a secondary core. (3) Subedge layer: Lanzhou, Xiamen, Taipei, Hefei, Harbin, Dalian, Chengdu, Kunming, Shenyang, Jinan, Shenzhen, Zhengzhou, Chongqing, Changchun, Changsha, and Qingdao. The flow of talent at this level has been significantly reduced, and the number of cities has increased significantly, and most of them belong to provincial capital cities or provincial central cities. Among them, the status of Shenzhen in the network has been greatly improved compared to the status of education. Shenzhen's low status in the talent mobility network may be due to the fact that it is an emerging innovation center. The number of top universities, scientific research institutions, and national laboratories in Shenzhen in the early days was small, and the influx of talent was little. With the accelerated flow and integration of innovative elements such as high-tech industries, capital, and technology, Shenzhen is gradually building into an international scientific and technological innovation center, and it will certainly attract a steady flow of high-level talent. (4) Edge layer: Sanya, Sanming, Shangrao, Dongguan, Dongying, and other 188 cities. This level of network node has the lowest status, and a large number of talents flow to the core and subcore layers.

From the perspective of change, the core of China's high-level talent mobility network is always Beijing. The subcore level has changed a lot. During the education stage, Chengdu, Changchun, and Changsha were in the subcore level, but in the working stage, these three cities have already withdrawn from this level. Tianjin, Guangzhou, Hangzhou, and Hong Kong have been upgraded from the original

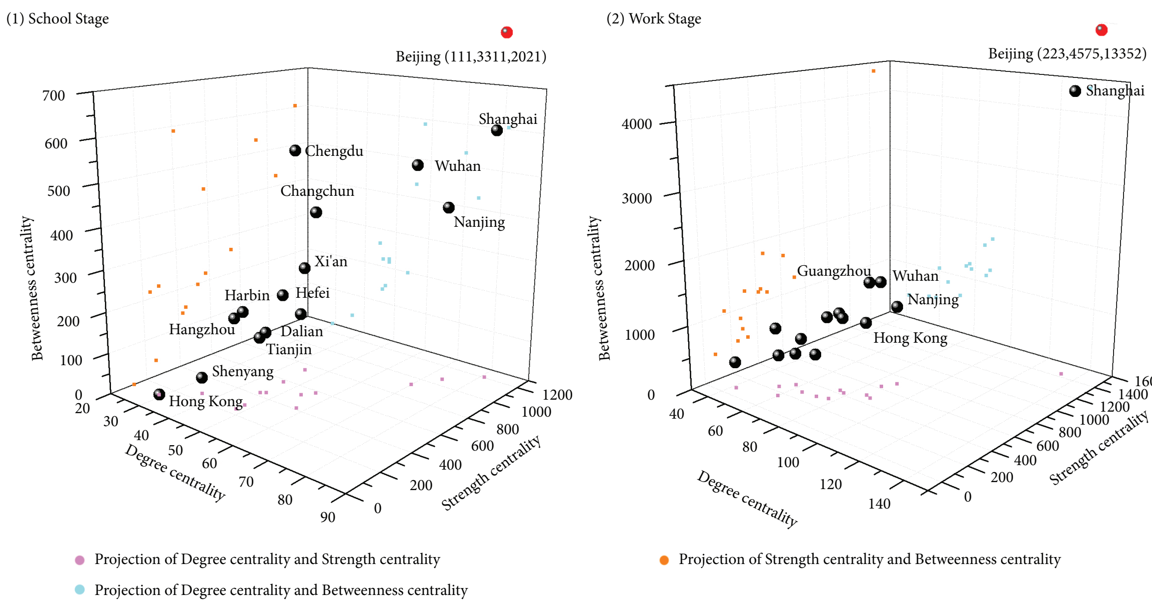


Figure 6: Centrality of network nodes. Note: only the top 15 nodes of comprehensive centrality are shown in the figure. Because Beijing has a high centrality and affects the visualization of the results, it is listed separately. (a) School stage. (b) Work stage.

subedge level to the subcore level. This situation may have a great relationship with the distribution of educational resources in China. The overall subedge layer has not changed much, while the number of nodes in the edge layer has increased.

4.4. Spatial Structure of the Network. At the stage of school, the network space structure showed a "small quadrangular pyramid structure" with Beijing as the apex and Xi'an, Wuhan, Shanghai, and Nanjing as the base (Figure 3). This structure is not only related to the distribution of national educational resources but also related to the region itself. For example, Xi'an is the center of the western region, and talents in the western region tend to flow to Xi'an. The talent flow in the network is mainly concentrated on Shanghai, Nanjing, Wuhan, Xi'an, Hefei, and other cities. Beijing, Shanghai, and Guangzhou are major cities with both inflow and outflow of talent, and in general the inflow is greater than the outflow. Cities such as Wuhan, Nanjing, Hefei, Xi'an, Changchun, and Harbin are dominated by the outflow of talent. The city pairs with the largest flows are as follows: Wuhan-Beijing, Nanjing-Beijing, Xi'an-Beijing, Hefei-Beijing, Beijing-Shanghai, and Changchun-Beijing.

During the work stage, the network space structure was upgraded and expanded, showing a "large quadrangular pyramid structure" with Beijing as the apex and Xi'an, Chengdu, Guangzhou (Hong Kong), and the Yangtze River Delta region (Shanghai, Nanjing) as the base (Figure 3). The entire spatial structure expands to the Guangdong, Hong Kong, and Macao regions in southeastern China. The city pairs with the largest flows are as follows: Beijing-Shanghai, Beijing-Hong Kong, Shanghai-Beijing, Hong Kong-Beijing, Beijing-Wuhan, and Beijing-Guangzhou. Talents' choice of workplace is related to the city in which they were educated. However, the scope and number of talents flowing during the work stage is still expanding to the coastal cities in the southeast.

Comparing the changes in the structure of the talent mobility network in the two stages, it is found that the network is expanding to the southeast, and the status of the city nodes in the southern part of the network is constantly increasing. Take Guangzhou as an example. In the education stage, Beijing and Guangzhou are less connected, while the flow of talent between Beijing and Guangzhou at the work stage has greatly increased. In recent years, Guangzhou has continuously introduced policies to attract talent. From 2016 to 2020 alone, Guangzhou has invested a total of 3.5 billion yuan in finance to focus on cultivating and introducing 500 leading talents in innovation and entrepreneurship. In addition, 500,000 to 5 million, up to 30 million salary subsidies will be given to outstanding industrial talent and high-level talent, and the team of leading entrepreneurs and talents will be supported in various ways. The superior talent policy attracts the continuous inflow of talent. According to the data in this paper, the current high-level talent in Guangzhou has reached 291. The influx of talent can enhance Guangzhou's position in the talent mobility network. With the exchange, diffusion, and accumulation of knowledge, technology, and information among talents, Guangzhou's innovation cooperation and innovation capabilities will continue to be strengthened, and it can also enhance the city's innovation network influence. While the network is extending to the southeast, the connection structure between cities in the Northeastern region and Beijing has been weakening. In particular, Beijing's ties with

Changchun have decreased the most. It can be seen that the Northeast region is unable to retain talent well and is in a disadvantaged position in the talent mobility network as an important talent training area.

5. Discussion

Based on China's HLT flow data, this paper constructs HLT flow networks in the two stages of school and work, abstracts the actual flow of talent into an analyzable and operable digital matrix, and uses complex network analysis methods to carry out related research and analysis. Comparative analysis found that the spatial distribution of HLT in China has a trend of "large dispersion and small aggregation." The gap between the centrality of nodes in the network has narrowed.The flow of talents has become more frequent and complicated. The core-peripheral structure of the network has not changed much, and the Matthew effect of the flow of talents still exists. From a spatial perspective, the talent flow network has significant spatial heterogeneity and continues to expand southward.

5.1. Changing Characteristics of the Spatial Distribution of Talents. As talents moving from the education stage to the work stage, their spatial distribution shows changed characteristics of "large dispersion and small gathering." "Large dispersion" is mainly manifested in the obvious increase in the number of cities where talents are distributed during the work stage. Some non-first-tier and nonprovincial capital cities in the central and western regions are also destinations for talent choices. HLT are distributed in second- and third-tier small cities due to the following reasons: first, it is related to the geographical distribution of work platforms. The establishment of some of the top domestic colleges and universities is affected by historical accumulation, administrative planning, and other nonmarket factors. They are located in the central and western regions. Second, it has a greater relationship with national and local policies. When implementing the policy of attracting talents, the country will also take efficiency and fairness into account and give full consideration of the coordinated development of the region. The "Outline of the National Medium and Long-term Talent Development Plan (2010–2020)" proposes to establish and improve a regional talent exchange and cooperation mechanism that is compatible with the development of the western region, the revitalization of old industrial bases in the northeast, the rise of the central region, and the development strategy of the eastern region. In the process of implementation, localities also attract talents through flexible methods such as building academician workstations and setting up part-time professors and lecture professors. The third is the balance between life pressure and personal development space [3]. High housing prices or living costs are an important obstacle for talents to stay in developed cities for a long time. Therefore, high-level talents have shown a trend of spreading across the country.

"Small gathering" is mainly reflected in the fact that talents are still gathering in a few eastern first-tier cities and provincial capitals [51–53]. Beijing, Tianjin, and Shanghai are hot spots for talent distribution [50]. This is mainly due to the rapid economic development in the eastern part of China, as well as the better conditions in scientific research, education, life, job opportunities, platforms, and prospects [8, 50, 54, 55]. Provincial capital cities such as Wuhan and Hefei in the central part also have a number of first-class universities and scientific research institutes, and there are more talents who choose to work in central cities [56, 57]. The western region is far lower than the eastern region in terms of economic development, employment environment, and talent introduction policies, which resulted in fewer talents to work. This phenomenon also reflects that the distribution of human resources does not completely overlap with educational resources. For example, Lanzhou, as a top talent training city, has 319 talents who choose to go to Lanzhou University for education, but only 150 talents choose to work in Lanzhou.

5.2. Bidirectional, Sticky, and Matthew Effect of HLT Flow. Through comparative research, it is found that the flow of HLT between cities in the school stage is agglomerating, mainly in a few cities such as Beijing and Shanghai. The flow of talents at the work stage has a significant bidirectional characteristic. That is, some cities are not only the places where talents flow in but also the places where talents flow out. This finding is more consistent with the research results of Refs. [8, 9], and [15]. From the spatial flow trajectory of China's HLT workplaces, we find that the flow of talents is mainly among Beijing, Shanghai, Hong Kong, and provincial capital cities. These cities are both strong inflow centers and strong outflow centers. Moreover, these cities are often developed cities with technology and economy.

The flow of HLT in China has a strong stickiness. That is, where the talents go to school and where they are employed have a strong consistency. From the talent flow data, it is known that 1,822 (23.56%) talents' school locations and work locations have never changed; 2,569 (33.22%) talents finally returned to the city where the school is located to work after experiencing a series of flows. This result is similar to the research results of Refs. [51, 58]. Although the types of talents studied are different, they show a relatively similar flow trend. That is, when choosing a working city, it is easier for talents to choose a city with academic relationships. The research of Refs. [6, 17] also found that factors such as interpersonal relationships established during school and a high degree of familiarity with the surroundings enhanced their tendency to find employment near the place of study. Of course, there are also significant regional differences in the stickiness of high-level talent flow. Cities with strong stickiness are mainly located in the southeast coastal area, the middle and lower reaches of the Yangtze River, and provincial capital cities. This result shows the significance of the overall stickiness of the superior area to the formation of talent highlands [51].

The Matthew effect still exists in the flow of HLT in China. That is, the more talented places are, the more attractive to talents, and the more talents they can get; the less talented places, the less attractive to talents, they will lose more talents, where are at a disadvantage in the talent competition [59]. In this study, this effect is mainly manifested in the talent flow network controlled by a small number of core nodes, and most of the talent flow is still concentrated in a few core cities. By comparison, it is found that the core-edge structure of the network in the school stage and the work stage is very significant and similar. From education to work, the network nodes at the subcore layer and the subedge layer have changed significantly. In the network, the core city Beijing will continue to strengthen and consolidate its core position by attracting resources from other levels of the network. In the subcore layer, cities such as Chengdu, Changsha, and Changchun withdrew, while cities such as Hangzhou, Tianjin, and Hong Kong entered. It not only reflects the difference between talents' preference for education cities and work cities but also reflects cities at the subcore and subperipheral levels can continuously strengthen their own investment, enhance their economic and innovative capabilities, and achieve a leap to a higher level. The marginal cities are still in a state of transporting talents to other levels for a long time due to their weak strength.

5.3. Spatial Heterogeneity of the HLT Flow Network. Through comparative analysis, it is found that the spatial heterogeneity of China's HLT flow network at the work stage is more significant. At the school stage, the flow of talents is mainly among cities with abundant education resources. At the work stage, the trend of regional concentration of talent mobility in China is more obvious. Talents mainly gather in the Beijing-Tianjin-Hebei, Yangtze River Delta, Wuhan metropolitan area, Chengdu-Chongqing urban agglomeration, and the Guangdong-Hong Kong-Macao Greater Bay Area. Among them, the talent flow network is expanding to the Guangdong-Hong Kong-Macao Greater Bay Area in southeastern China. With the continuous implementation of regional development strategies such as the Guangdong-Hong Kong-Macao Greater Bay Area, it may stimulate the continued flow of talents to the region. Cities such as Guangzhou, Shenzhen, and Hong Kong will continue rising in their roles as talent-absorbing places. Moreover, the heterogeneity of China's HLT flow at different stages is also greatly related to the distribution of the country's education and science centers. It can be said that the formation and transfer of education centers and technology centers will inevitably be accompanied by the flow of talents. On November 8, 2020, the Chinese Academy of Social Sciences and China Times jointly released the "2020 National Central City Index" report (https://baijiahao.baidu.com/s?id=1682758480360564512&wfr=spider&for=pc). The report ranks China's education center cities and identifies Beijing, Shanghai, Nanjing, Xi'an, Wuhan, Guangzhou, Chengdu, and Changsha as the country's education center cities. This ranking is almost the same as the ranking of cities with HLT in the school stage, which further proves that high-level talents are more inclined to flow into cities with rich educational resources during the education stage. In March 2021, the 14th Five-Year Plan for the National Economic and Social Development of the People's Republic of China and the 2035 long-term goal outline (http://www.gov.cn/xinwen/2021-03/14/content_5592884.htm) were released, pointing out to support Beijing, Shanghai, and the Guangdong-Hong Kong-Macao Greater Bay Area to form an international science and technology innovation center, and to construct comprehensive national science centers in Beijing-Huairou and Shanghai-Zhangjiang, the Greater Bay Area, and Anhui-Hefei as well as to support the construction of regional science and technology innovation centers in places where conditions permit. The spatial distribution of these technological innovation centers is very consistent with the regional concentration of talent flow during the work phase, which shows that there is a close relationship between the two.

6. Conclusion

The flow of talents has always been highly concerned by scholars and the government. At present, China is in the most active period of cross-regional flow of HLT. This research conducted a comparative study on the current status of China's HLT flow and its network structure during the school and work stages, and the following conclusions were reached.

First, comparative analysis found that the spatial distribution of HLT in China has a trend of "large dispersion, small agglomeration." The distribution of talents in the school stage is more consistent with the spatial distribution characteristics of the country's educational resources, while the work stage is mainly concentrated in cities with better economic development and clusters of innovative platforms and does not completely coincide with the distribution of educational resources.

Second, from school to work stage, the flow of talents becomes more active and the scope of mobility continues to expand. At the school stage, China's HLT mainly gather in cities with rich educational resources such as Beijing and Shanghai. At the work stage, talents are more inclined to flow into eastern coastal cities and provincial capitals. In addition, the flow of talents at the work stage has a significant bidirectional characteristic of flow between first-tier cities or provincial capital cities. At the same time, some advantageous regions also have strong stickiness to talents.

Third, the core-edge structure of the network in the two stages is significant. The core layer is Beijing, and the subcore and subedge layer cities have changed. The core-edge structure of the two stages shows relatively similar characteristics. That is, cities with high economic levels, rich educational resources, complete innovation infrastructure, and high-level technology are at the core of the talent flow network. Most non-first-tier and nonprovincial capital cities are at the marginal level. Moreover, the Matthew effect of talent flow has always existed.

Fourth, the spatial structure of the network has been upgraded and expanded, from the "small quadrangular pyramid" structure with Beijing as its apex and Xi'an, Wuhan, Shanghai, and Nanjing as its base, to Beijing as its apex, Xi'an, Chengdu, and Guangzhou (Hong Kong), the Yangtze River Delta (Shanghai, Nanjing) as the base of the "large quadrangular pyramid" structure. At the work stage, the flow of talents shows obvious spatial heterogeneity, and the trend of regional concentration of talents is more obvious.

From the perspective of urban development, the higher the position of a city in the talent flow network, the more talent resources it can attract from the network, as well as the capital, knowledge, and technology resources that talents can bring. The city's high-level talent matching, overall creative ability and level, high-end level, and innovation benefits will directly determine the city's status in national competition and industrial division of labor. However, by comparing and analyzing the characteristics of the talent flow in the two stages, we find that in cities with weak economic strength, the educational advantage cannot be transformed into a technological advantage and an economic advantage, which will lead to brain drain. In this case, the government needs to consider reducing the barriers to talent introduction through more flexible methods such as precise introduction of talents, flexible introduction of talents (intellectual introduction, academic part-time job, and talent dispatch), etc., when its own human resources cannot be better reserved. By strengthening the connection with the human resources of other cities, forming a cross-regional cooperation network, and using the resource exchange in the network, these are also ways to minimize the disadvantages caused by the brain drain.

Based on the growth stage of talents, this paper studies the spatial distribution and talent flow network of China's HLT, trying to enrich the current research on talent mobility and even talent geography in China. However, this article still has some shortcomings. In the future, we will continue to improve relevant theoretical analysis, try to build a multiscale, multiconnotation regional talent mobility network, and explore the impact of regional talent mobility and the formation of the network on regional economic and technological development, so as to provide targeted and operable policy recommendations for national or regional talent policies and innovative development.

Data Availability

The data used to support the study can be available upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the "Strategic Priority Research Program of the Chinese Academy of Sciences" (No. XDA20010103).

References

[1] J. Xi, "Speech at the teacher-student symposium of Peking University," 2018, http://www.chinadaily.com.cn/a/201805/08/WS5af0f499a3105cdcf651c7%20fc.html.

[2] A. Geuna, *Global Mobility of Research Scientists*, Elsevier Monograph, Amsterdam, Netherlands, 2015.

[3] China Science Coordination Research Propaganda Department, *The Total Amount, Structure And Flow Of Scientific Research Personnel Of Scientific And Technological Human Resources*, vol. 6, Tsinghua University Press, Beijing, China, 2021.

[4] Y. Wang, H. Luo, and G. Yang, "Global features and evolution of talent mobility with complex network between provinces in China," *Science Research Management*, Available at: https://kns.cnki.net/kcms/detail/11.%201567.G3.20210914.1328.005.html, 2021.

[5] Y. Li and J. Liu, "The practice and prospect of the construction of scientific and technological talent team in China since the reform and opening-up," *Human Resource Development of China*, vol. 35, no. 11, pp. 30–43, 2018, (in Chinese).

[6] Y. Jiang, X. Wang, and R. Ma, "Theoretical exploration of innovative talent agglomeration: from the perspective of city selection of global talent flow," *Scientia Geographica Sinica*, vol. 41, no. 10, pp. 1802–1811, 2021, (in Chinese).

[7] H. Huang and Y. Qu, "Peacock flying to Southeast: economic transformation and star scientist mobility," *Journal of Huazhong University of Science and Technology (Nature Science Edition)*, vol. 33, no. 3, pp. 63–72, 2019, (in Chinese).

[8] Y. Zhou, Y. Guo, and Y. Liu, "High-level talent flow and its influence on regional unbalanced development in China," *Applied Geography*, vol. 91, no. 91, pp. 89–98, 2018.

[9] L. Zhou and Y. Zhang, "The spatial distribution and flow trends of Chinese top academic talents: based on the case study of academicians of Chinese academy of sciences," *Geographical Research*, vol. 38, no. 7, pp. 1749–1763, 2019, (in Chinese).

[10] Y. Liu, J. Shen, and W. Xu, "From school to university to work: migration of highly educated youths in China," *The Annals of Regional Science*, vol. 59, no. 3, pp. 651–676, 2017.

[11] H. Gu, X. Meng, T. Y. Shen, and L. Wen, "China 's highly educated talents in 2015: patterns, determinants and spatial spillover effects," *Applied Spatial Analysis and Policy*, vol. 13, no. (2), 2020.

[12] H. Gu, F. Rowe, Y. Liu, and T. Shen, "Geography of talent in China during 2000–2015: an eigenvector spatial filtering negative binomial approach," *Chinese Geographical Science*, vol. 31, no. 2, pp. 297–312, 2021.

[13] Y. Liu and J. Shen, "Modelling skilled and less-skilled interregional migrations in China, 2000–2005," *Population, Space and Place*, vol. 23, no. 4, pp. e2027.1–e2027.17, 2017.

[14] H. Huang and J. Lian, "Does job mobility improve scientists' academic productive?" *Tsinghua Journal of Education*, vol. 41, no. 5, pp. 127–135, 2020, (in Chinese).

[15] Z. Gao, C. Zhong, C. Wang, and Y. Liu, "Analysis on the flow of high-level scientific and technological talents in China: taking the sponsor of national science fund for distinguished young scholars as example," *Bulletin of National Natural Science Foundation of China*, vol. 33, no. 4, pp. 363–366, 2019, (in Chinese).

[16] D. Shi and Z. Zhang, "Talent programs and scientists' professional mobility: a case study of China jiang scholars," *Document, Information & Knowledge*, no. 2, pp. 42–50, 2020, (in Chinese).

[17] L. Rui, D. Wu, J. Bao, Y. Qiu, and W. Wang, "Evolution of temporal and spatial pattern and driving mechanisms of agglomerative growth of senior science talents a case study of academicians of CAS," *Progress in Geography*, vol. 32, no. 7, pp. 1123–1138, 2013, (in Chinese).
[18] M. Castells, *The Rise of the Network Society*, Wiley-Blackwell, Hoboken, NJ, USA, 2000.
[19] M. Castells, *The Informational City: Information Technology, Economic Restructuring and the Urban-Regional Progress*, Blackwell, Oxford, UK, 1989.
[20] M. Castells, "Grassrooting the space of flows," *Urban Geography*, vol. 20, no. 4, pp. 294–302, 1999.
[21] H. Ma, "The theoretical construction and network simulation of intercity innovative relationships in knowledge flow space," *Acta Geographica Sinica*, vol. 75, no. 4, pp. 708–721, 2020, (in Chinese).
[22] P. J. Taylor, *World Cities Network: A Global Urban Analysis*, Routledge, London, UK, 2004.
[23] E. J. Malecki, "Hard and soft networks for urban competitiveness," *Urban Studies*, vol. 39, no. 5, pp. 929–945, 2002.
[24] R. Florida, C. Mellander, K. Stolarick, and A. Ross, "Cities, skills and wages," *Journal of Economic Geography*, vol. 12, no. 2, pp. 355–377, 2012.
[25] D. J. Watts and S. H. Strogats, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
[26] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
[27] Y. Wang, H. Luo, and Z. Li, "The complex international migration network for top scientific and technological talents," *International Journal of Humanities and Social Science*, vol. 7, no. 10, pp. 133–142, 2017.
[28] R. Pastor Satorras and A. Vespignani, *Handbook of graphs and networks: from the genome to the Internet, Chapter Epidemics and Immunization in Scale-Free Networks*, Wiley-VCH Verlag GmbH & Co., KGaA, Weinheim, Germany, 2002.
[29] F. Jin, "A study on network of domestic air passenger flow in China," *Geographical Research*, vol. 20, no. 1, pp. 31–39, 2001, (in Chinese).
[30] W. Gu and H. Liu, "Spatial structure, hierarchy and formation mechanisms of scientific collaboration networks: evidence of the belt and road regions," *Chinese Geographical Science*, vol. 30, no. 6, pp. 959–975, 2020.
[31] M. Chen, X. Luo, M. Hu, and X. Zhang, "Club characteristic and the proximity effect of travel links between cities in China," *Complexity*, vol. 2020, Article ID 5786384, 13 pages, 2020.
[32] X. Luo and M. Chen, "Urban network of China from the perspective of population mobility: three-dimensional co-o44ccurrence of nodes and links," *Environment & Planning A*, vol. 53, no. (5), pp. 887–889, 2021.
[33] Y. Zhu, X. Zheng, and Y. Zhang, "Analysis on the flowing rules of creative talents in science and technology in China- an empirical research based on leading talents in science and technology of national ten thousand talents program," *Forum on Science and Technology in China*, no. 3, pp. 166–173, 2020, (in Chinese).
[34] W. Shi, W. Yang, and D. Du, "The scientific cooperation network of Chinese scientists and its proximity mechanism," *Suatainability*, vol. 12, p. 660, 2020.
[35] D. Duan, D. Du, Q. Gui, and L. H. Duan, "The geography of Chinese entrepreneurial development," *Human Geography*, vol. 33, no. 4, pp. 102–112, 2018, (in Chinese).
[36] P. Zhu, Y. Zhang, W. Cao, Y. Ren, and Y. Jiang, "Spatial distribution and geographical mobile network of enterprise management talents in Yangtze river delta: based on the data of senior executive director team of listed company," *Human Geography*, vol. 35, no. 4, pp. 121–129, 2020, (in Chinese).
[37] E. M. Vogtle and M. Windzio, "Networks of international student mobility: enlargement and consolidation of the European transnational education space?" *Higher Education*, vol. 72, no. 276, pp. 1–19, 2015.
[38] H. Ma, F. Zhang, and Y. Liu, "Transnational elites enhance the connectivity of Chinese cities in the world city network," *Environment & Planning A*, vol. 50, no. 4, pp. 749–751, 2018.
[39] G. Fagiolo and M. Mastrorillo, "International migration network: topology and modeling," *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, vol. 88, no. 1, Article ID 012812, 2013.
[40] Y. Wang, H. Luo, and Z. Li, "A study on the network of global mobility of scientific top talents - based on bibliometrics and complex network," *Technology and Innovation Management*, vol. 39, no. 5, pp. 521–529, 2018, (in Chinese).
[41] Y. Wang and H. Luo, "Global talent mobility from 1990-2012 according to patent data," *China Soft Science Collection 2019 China Soft Science Essays*, (in Chinese), p. 10, China Soft Science Research Association, Beijing, China, China Soft Science Research Association, 2020.
[42] J. V. Beaverstock, "Skilled international migration: an analysis of the geography of international secondments within large accountancy firms," *Environment & Planning A*, vol. 23, no. 8, pp. 1133–1146, 1991.
[43] J. V. Beaverstock, "The spatial mobility of corporate knowledge," *Expatriation, Global Talent, and the World City// Mobilities of Knowledge*, Springer International Publishing, New York, NY, USA, 2017.
[44] S. Chen, *Introduction to Innovative Talents*, Sun Yat-sen University press, Guangzhou, China, (in Chinese), 2015.
[45] J. Qiao and X. Chen, "Dose mobility of high-lecel talents affect academic production? taking the field of life sciences as an example," *Universities and Disciplines*, vol. 2, no. 3, pp. 91–107, 2021, (in Chinese).
[46] J. Liu, "Whole Network Analysis," *UCINET Software Practical Guide*, Gezhi Publishing House, Shanghai People's Publishing House, Shanghai, China, (in Chinese), 3rd edition, 2019.
[47] C. Hou, D. Du, C. Liu, Q. Gui, S. Liu, and X. Qin, "Spatio-temporal evolution and factors influencing international student mobility networks in the world," *Acta Geographica Sinica*, vol. 75, no. 4, pp. 681–694, 2020, (in Chinese).
[48] C. Liu, Q. Gui, D. Duan, and M. Yin, "Structural heterogeneity and proximity mechanism of global scientific collaboration network based on co-authored papers," *Acta Geographica Sinica*, vol. 72, no. 4, pp. 737–752, 2017, (in Chinese).
[49] C. Hou, D. Du, C. Liu, and C. Zhai, "Spatiotemporal evolution of global talent mobility network based on the data of international student mobility," *Geographical Research*, vol. 38, no. 8, pp. 1862–1876, 2019, (in Chinese).
[50] R. Wang, H. Xu, and D. Xue, "Spatio-temporal change of facaulty members of higher education institute and its influential factors in China in 2005-2015," *Scientia Geographica Sinica*, vol. 39, no. 8, pp. 1199–1207, 2019, (in Chinese).
[51] J. Nie and H. Liu, "Spatial pattern and the resulting characteristics of talent flows in China," *Scientia Geographica Sinica*, vol. 38, no. 12, pp. 1979–1987, 2018, (in Chinese).
[52] P. Zhang, Y. Zhao, and Y. Ma, "Intellectual resources development and regional competitiveness of northeast China,"

Scientia Geographica Sinica, vol. 23, no. 5, pp. 513–518, 2003, (in Chinese).
[53] J. Zhang, D. B. Du, and H. N. Jiang, "Regional differences of S&T talents in jiangsu province," *Scientia Geographica Sinica*, vol. 31, no. 3, pp. 378–384, 2011, (in Chinese).
[54] Y. Liu, J. Shen, and Y. Liu, "Transnational mobility of the highly skilled: a review," *Human Geography*, vol. 28, no. 2, pp. 7–12, 2013, (in Chinese).
[55] H. Ma and F. Zhang, "Literature review on the impetus and influence of talents mobility between countries and regions," *Economic Geography*, vol. 39, no. 2, pp. 40–47, 2019, (in Chinese).
[56] X. Jiang, W. Fu, and G. Li, "Can the improvement of living environment stimulate urban innovation? - analysis of high-quality innovative talents and foreign direct investment spillover effect mechanism," *Journal of Cleaner Production*, vol. 255, pp. 120–138, 2020.
[57] R. Ma, "Constructing the creative space and evolving perspectives on cities," *Urban Planning Forum*, vol. 54, no. 6, pp. 109–118, 2010, (in Chinese).
[58] Y. Wang, C. Cui, Q. Wang, Y. Ning, and Z. Yang, "Migration of human capital in the context of vying for talent competition: a case study of China's "first-class" university graduates," *Geographical Research*, vol. 40, no. 3, pp. 743–761, 2021.
[59] X. Yuan, *Study on Behavior and Effects of Creative Talents' Agglomeration from the View of Complex Adaptive System*, Donghua University, Shanghai, China, 2017.

Hindawi
Complexity
Volume 2021, Article ID 2574025, 10 pages
https://doi.org/10.1155/2021/2574025

Research Article

Delineating the Spatial Boundaries of Megaregions in China: A City Network Perspective

Wei Chen

Key Laboratory of Regional Sustainable Development Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Correspondence should be addressed to Wei Chen; chenw@igsnrr.ac.cn

Received 8 September 2021; Accepted 18 December 2021; Published 30 December 2021

Academic Editor: Hang-Hyun Jo

Globalization and informatization have significantly reshaped the map of the global economy. Mega cities and regions have become the battlegrounds in the interplay between globalization and localization, with megaregions becoming the most globally significant spatial configurations in this regard. However, academics and government departments disagree on how to define the spatial boundaries of megaregions. In this study, on the basis of highway traffic flow data between cities, we integrate the community detection and core-periphery profile algorithms to characterize the city networks in China and identify the city groups and delineate the core structures of city groups, which are the underlying megaregional structures in China. Based on this, we identify 21 megaregions among city groups in China, including the Yangtze River Delta, Pearl River Delta, Beijing-Tianjin-Hebei, and Chengdu-Chongqing megaregions, and preliminarily delineate their spatial boundaries. On the whole, there are spatial differences among China's megaregions to a certain extent. Central and eastern China have numerous, large, and a high density of megaregions, while the western region has relatively few megaregions. The latter also differs notably from mature megaregions in terms of rank sizes, urban systems, and functional divisions of labor. Generally, this study develops a novel analytical framework for identifying the functional regions of megaregional space in China from a perspective of relational geography, with methodological implications for other fields of inquiry.

1. Introduction

The past two decades have seen globalization and informatization that greatly reshape global economic geography and the rise of a network society, and mega cities and regions have become the spatial units that host the fiercest interplay between globalization and localization [1–3]. As globalization and urbanization have progressed, urban competition has gone beyond individual cities, and it is increasingly about position and functional connections in divisions of labor, especially competition, and cooperation in urban networks. Because of this, megaregions have become globally significant spatial configurations [4–6]. Since China's 11th Five-Year Plan period (2006–2010), the development of megaregions has been elevated to the status of a national strategy, and they have been seen as the primary entities for promoting China's "new type of urbanization." Megaregions are currently a popular topic in both academia and governments. The concept repeatedly crops up in various settings, and it has become an important development strategy in regional economic and spatial planning in China [7]. As an objective geographical phenomenon, megaregions have always been an important topic of research in urban geography and urban planning, but for a long time the term has had ambiguous connotations in academic research, as the academic community has failed to reach a consensus on its definition and spatial boundary. This, in turn, has led to the ambiguity surrounding megaregions in practice [8, 9].

Defining the spatial boundary of a megaregion has always been a fundamental task of research on these spatial entities, and it is a prerequisite for understanding their formation, development, and evolution. Given that megaregions are very large, complex, dynamic, and open systems

[10], their spatial boundaries can be ambiguous and changing. Defining and understanding the spatial boundaries of megaregions can also be tedious and difficult work for both geographers and planners. Based on the definition of megalopolis [10], combined with criteria for distinguishing metropolitan areas and regions, scholars have proposed criteria for identifying the spatial boundaries of megaregions (or megalopolises) from various perspectives [7, 11, 12]. Along with the "relational turn" in economic geography [13, 14], many scholars have emphasized the importance of the perspectives of relationality and connectivity in understanding the regional delineation of bounded geographic areas [15, 16]. Recently, Nelson and Rae used a large data set of commuter flows to identify megaregions in the United States [17]. Others integrate the big data with emerging methods, shedding light on the division of multilevel functional regions as a geographic issue through visualization and spatial analysis [18–20].

With regards to identifying megaregions in China, substantial changes have taken place in both research data and methods, especially along with the advancement of modern geographic information technology and the advent of the era of big data. In addition to using traditional urban area-based spatial identification methods, some scholars have begun to try new methods of geographic information analysis, such as gravity model [21] and network analysis [22], to explore emerging data types, including night-time light [23] and point of interest (POI) [24], or combining qualitative and quantitative analyses to develop frameworks for recognizing the spatial boundaries of megaregions from a multidimensional perspective and defining the spatial boundaries of China's megaregions or a single megaregion [25, 26]. Depending on their research approaches, scholars have defined and divided megaregions into different numbers, types, and ranges, and they have produced various megaregion boundary schemes [27].

In this era of globalization and information, cities are constantly becoming connected to the global production networks as locations and nodes. Cities and regions are being linked through various flows, networks, and relationships, forming urban functional region systems of various spatial scales [28, 29]. As large-scale urbanized regional landscapes, megaregions are huge complex systems that integrate economic, social, political, cultural, and other elements [1, 30]. Economic connections and functional divisions of labor are the most important spatial characteristics of these functional region systems, and they are key indicators that reflect their essence. As a result, defining the spatial boundaries of megaregions requires consideration from the perspective of relational geography.

The field of network science has shown that real-world networks tend to consist of various mesoscale structures, such as community structures [31] or core-periphery structures [32]. Thus, the national-scale city networks also have mesoscale structures, such as community and core-periphery structures, within the overall network. These mesoscale structures provide potential and new perspectives for a comprehensive and scientific understanding of urban and regional structural systems within the national urban system. City networks, which are rooted in geographical spatial relationships, have typical mesoscale structural foundations that can provide a potential methodology for empirically classifying and identifying the spatial boundaries of megaregional space. Highway passenger transportation mainly covers short distances, with significant spatial dependence and distance decay, which has proved useful in analyzing intercity functional relationships at the city and regional scales. For these reasons, intercity highway flows are crucial indicators for analyzing regional economic systems at the megaregional scale.

Against this background, on the basis of the city network perspective, we integrate the community detection and core-periphery profile algorithms to identify the city groups of the city networks and delineate the megaregions among the city groups in China so as to provide a methodological framework based on the relational geography for determining the spatial boundaries of megaregions.

2. Methodology and Data

2.1. Analytical Framework. As for the spatial characteristics of megaregions, we adopt a city network perspective to identify the spatial boundaries of megaregions. Specifically, we integrate the community detection and core-periphery profile algorithms to build an analytical framework and then spatially recognize the boundaries of megaregions in China based on the intercity connectivity. First, using highway flow data among cities at or above the prefectural level, we characterize the characteristics of the city network structures in China and then employ the community detection algorithm to divide China's city networks and identify the city groups in functional region systems. Second, based on the above city groups, we further use the core-periphery profile algorithm to identify the core structures of the city groups and extract the core nodes, which are the most influential and most closely connected groups of nodes and correspond to the most cohesive agglomeration of cities in the city groups, which are believed as the "megaregional" component of the city groups. Third, taking into consideration natural surface conditions and socioeconomic development backgrounds, we finally delineate the spatial boundaries of megaregions in China.

2.2. Methods

2.2.1. Community Detection. In the field of network science, a community refers to a subset of a network. The nodes of a network can be grouped into sets of nodes so that each community is closely connected internally with sparser connections between groups. The identification of densely connected groups, based on network attributes, is referred to as community detection [33]. Community detection, which is crucial for understanding group structures in networks, has long been one of the most important issues in network science. To identify the community structures in the real-world networks, many efficient algorithms are created, such as Girvan and Newman [34], Walktrap [35], Fast-greedy [36], and Infomap algorithms [31].

As in real-world networks, node weight, edge weight, and linkage direction are important features for illustrating community structures of networks. Although the community detection methods could be conducted in different algorithms, many of the popular algorithms are currently unable to take directed and weighted networks into consideration. To fill this gap, the Infomap algorithm, currently one of the most robust algorithms, can consider topological properties including node and edge weights and network directions, which exhibits remarkable robustness and adaptability [5, 37, 38]. Therefore, based on the intercity highway passenger flows, we employ the Infomap algorithm of community detection to identify the city groups from the city networks in China.

In brief, the Infomap algorithm identifies communities within directed and weighted networks via the combined use of random walks and compression principles [31]. According to Shannon's source coding theorem, if we use n codewords to describe the n states of a random variable X that occur with frequencies p_{i}, the average length of a codeword can be no less than the entropy of the random variable X itself:

$$H(X) = -\sum_{1}^{n} p_i \log(p_i). \tag{1}$$

This expression provides a lower bound on the average length of codewords in each codebook. To calculate the average length of the code describing a step of the random walk, we need only to weight the average length of codewords from the index codebook and the module codebooks based on their rates of use. This is the map equation:

$$L(M) = qH(Q) + \sum_{i=1}^{m} p_i H(P_I). \tag{2}$$

In this expression, $L(M)$ denotes the expectation of average code length that the random walk spends inside and outside communities; q is the probability to exit module i; $H(Q)$ is the frequency-weighted average length of codewords in the index codebook; p_{i} is the probability to visit any node that the random walker spends in module i; and $H(P_i)$ is the frequency-weighted average length of codewords in module codebook i.

2.2.2. Core-Periphery Profile. The portrait of a network as divided into a dense core and a sparse periphery, referred to as a core-periphery structure, originated from scholars in social sciences in the 1990s, and now this paradigm has been extended to other disciplines [32]. To identify the core-periphery structures in networks, some algorithms were proposed successively, such as the block-modeling [39], k-shell decomposition [40], and centrality [41]. However, most of the proposed algorithms are unable to deal with the weighted networks, and their robustness still needs to be verified. Against this background, Della et al. recently proposed the algorithm of core-periphery profile [42], disclosing the overall network structures and the peculiar roles of specific nodes.

In a network with an ideal core-periphery structure, peripheral nodes (p-nodes) are allowed to link to core nodes only; namely, no connectivity exists among p-nodes. In most real-world networks, however, the structure is not ideal although the core-periphery structure is evident: a weak (but not null) connectivity exists among the peripheral nodes. This calls for the generalized definition of α-periphery, which denotes the largest subnetwork S with the persistence probability $\alpha_S \le \alpha$.

We define the core-periphery profile α_k, $k = 1, 2, \ldots, n$, of the network by the following algorithm. This is the equation:

$$\begin{aligned} \alpha_k &= \min_{h \in N/P_{k-1}} \frac{\sum_{i,j \in P_{k-1} \cup \{h\}} \pi_i m_{ij}}{\sum_{i \in P_{k-1} \cup \{h\}} \pi_i}, \\ &= \min_{h \in N/P_{k-1}} \frac{\sum_{i,j \in P_{k-1}} \pi_i m_{ij} + \sum_{i \in P_{k-1}} \left(\pi_i m_{ih} + \pi_h m_{hi}\right)}{\sum_{i \in P_{k-1}} \pi_i + \pi_h}. \end{aligned} \tag{3}$$

We start by the node i with the weakest connectivity and generate a sequence of sets $\{1\} = P_1 \subset P_2 \subset \ldots P_n = N$ by adding, at each step, the node attaining the minimal increase in the persistence probability. Correspondingly, we obtain the core-periphery profile, that is, the sequence $0 = \alpha_1 \le \alpha_2 \le \cdots \le \alpha_n = 1$ of the persistence probabilities of the sets P_k.

The above algorithm provides, as byproducts, two other important tools of analysis including the centralization and coreness. We define the centralization C for a core-periphery profile α_k as the complement to 1 of the normalized area; namely,

$$C = 1 - \frac{2}{n-2} \sum_{k=1}^{n-1} \alpha_k. \tag{4}$$

We can therefore quantify such a similarity by measuring the area between the α_k-curve of a given network and that of the star network and normalizing to assign $C = 1$ to the star network itself (maximal centralization) and $C = 0$ to the complete network (no centralization). If a network displays a definite core-periphery structure (large C), then the sequence α_k naturally provides a measure of coreness of each node. We have $\alpha_k = 0$ for all p-nodes (the periphery in the strict sense), whereas the coreness of the last inserted node is maximal and equal to $\alpha_n = 0$.

2.3. Data Sources. This study uses original data on highway flows between 289 administrative units above the prefectural level in China. The data structure is asymmetric 289×289 matrix, characterizing the strength of spatial connectivity among cities in China based on the highway flows. We primarily used a vehicle services website (checi.cn) for web page retrieval and then extracted intercity highway flow data by circular queries. In addition, we combined the websites of the commercial services, such as China Highway Ticket Network and provincial and municipal highway ticket networks, Ctrip.com and Changtu.com, to verify and correct data manually based on cross-checks to ensure the completeness and accuracy of the data. This included data capture carried out using Microsoft's C# programming language. The data collection date was June 2017.

3. Results

3.1. Identifying the City Groups in China. Based on the highway flow data between 289 cities at or above the prefectural level in China, we employ the Infomap community detection algorithm [31] to divide the whole cities into groups of cities. And then, we identify 19 city groups in total. Figure 1 shows the spatial network patterns of city groups in China based on the intercity highway flow data.

In terms of spatial patterns, these 19 city groups exhibit strong spatial dependence and diverse spatial organization patterns, with an apparent multilevel, networked regional configuration. In terms of spatial form, there are obvious hierarchical structures of city networks developing within city groups, which reflect the spatial imbalance toward core cities. Most city groups show varying degrees of core-periphery structures, with core cities occupying dominant positions within the network structures. Peripheral cities have a relatively weak connection with the network structures. This is the major reason that we use the core-periphery structures in city groups to identify megaregions in this study.

The 19 city groups have different spatial compositions. Within the city groups, cities have obvious differences in rank size and distribution, with a general trend of a slightly decreasing gradient. The specific parameters are shown in Figure 2. City Group 1, City Group 2, and City Group 3 are the three largest city groups, corresponding to the Jiangsu-Zhejiang-Shanghai city group in the Yangtze River Delta, which contains 25 cities and has a total of 8,357 connections; the Guangdong-Guangxi city group, which contains 37 cities and has a total of 9,519 connections; and the Beijing-Tianjin-Hebei-Shandong city group, which contains 32 cities and has a total of 5,993 connections. These city groups far exceed the other city groups in terms of the number of cities they incorporate and their number of connections, indicating significant agglomerative economic effects.

3.2. Extracting the Core Structures of City Groups. Based on the above city groups, we employ the core-periphery profile algorithm [42] on the basis of random walkers to extract the most cohesive structures of city groups, which are embryonic structures of megaregions among China's regional economies. Because City Group 19 has only two prefectural-level cities—Urumqi and Karamay—that are geographically distant and have relatively weak economic connections, they do not have the basic conditions to develop city networks and economic integration. We do not believe that Xinjiang currently has the natural and economic conditions for a megaregion to develop, so only the other 18 city groups are used as the basis of this study.

Using the core-periphery profile algorithm mentioned previously, we calculate the polarization effects of the core-periphery structures as well as the position and importance of nodes in the networks. The centralization results are shown in Figure 3. From the core-periphery centralization, most of city groups in China have values greater than 0.5, indicating that they have obvious core-periphery structures, core cities have a clear agglomerative effect within the city groups, and some nodes have important positions and roles in the networks. The centralization coefficients of City Group 1 and City Group 16 are less than 0.5, which means the core-periphery structures of those two groups are not as strong as the others, rather than showing that they do not have core-periphery structures. Specifically, City Group 1 (Jiangsu-Zhejiang-Shanghai) has a relatively balanced network structure, so the polarization effect of the core-periphery structure is less pronounced than in other city groups. City Group 16 (central Inner Mongolia) has a low overall network density and no node cities with obvious advantages. The gap in strength between the cities is not notable, and the city network has relatively low connectivity. The city group does not have obvious core and periphery components, so its centralization coefficient is relatively low.

Figure 4 compares the "coreness" of the core-periphery structures of city groups in China. The x-axis ranks the city groups according to their coreness, and the y-axis is the coreness value. To investigate core-periphery structure features of each city group, we set the node order and coreness of the city groups to the same level, assuming that the number of nodes is 100, which allows comparison of the core-periphery structures of different city groups. Overall, the core-periphery structures of city groups in China are roughly similar. The number of core nodes in the network is generally greater than the number of periphery nodes, with the ratio of core to periphery nodes relatively stable at around 6 : 4. The curves in Figure 4 reveal that the watershed between core and periphery nodes is at approximately the point of 40%. After that, the coreness of the city groups begins to diverge, displaying four types.

The first type is the city groups with centralization coefficients greater than 0.8, which have notable features of core-periphery structures. Their core structures are situated after the 50% mark. This type includes City Group 15 (Guizhou), City Group 17 (Gansu, Ningxia, Qinghai, and Tibet), and City Group 18 (Yunnan). The second type is the city groups with centralization coefficients between 0.6 and 0.8, which have clear features of core-periphery structures. They are the most numerous types of city groups, including the City Groups 2, 4, 5, 6, 8, 9, 11, 12, 13, and 14. The core structures of this type are situated at over 40%. The third type is the city groups with centralization coefficients between 0.4 and 0.6, with core nodes at or above 30%. This type includes the City Groups 1 (Jiangsu-Zhejiang-Shanghai), 3 (Beijing-Tianjin-Hebei-Shandong), 7 (Fujian), and 10 (Jiangxi). The fourth type is the City Group 16 (central Inner Mongolia), which has a centralization coefficient of 0.32 and a relatively weak core-periphery structure, with little distinction between core and periphery nodes.

Based on the above descriptions of coreness and node order, we separate the core structures and periphery structures of the city groups and extract the core-

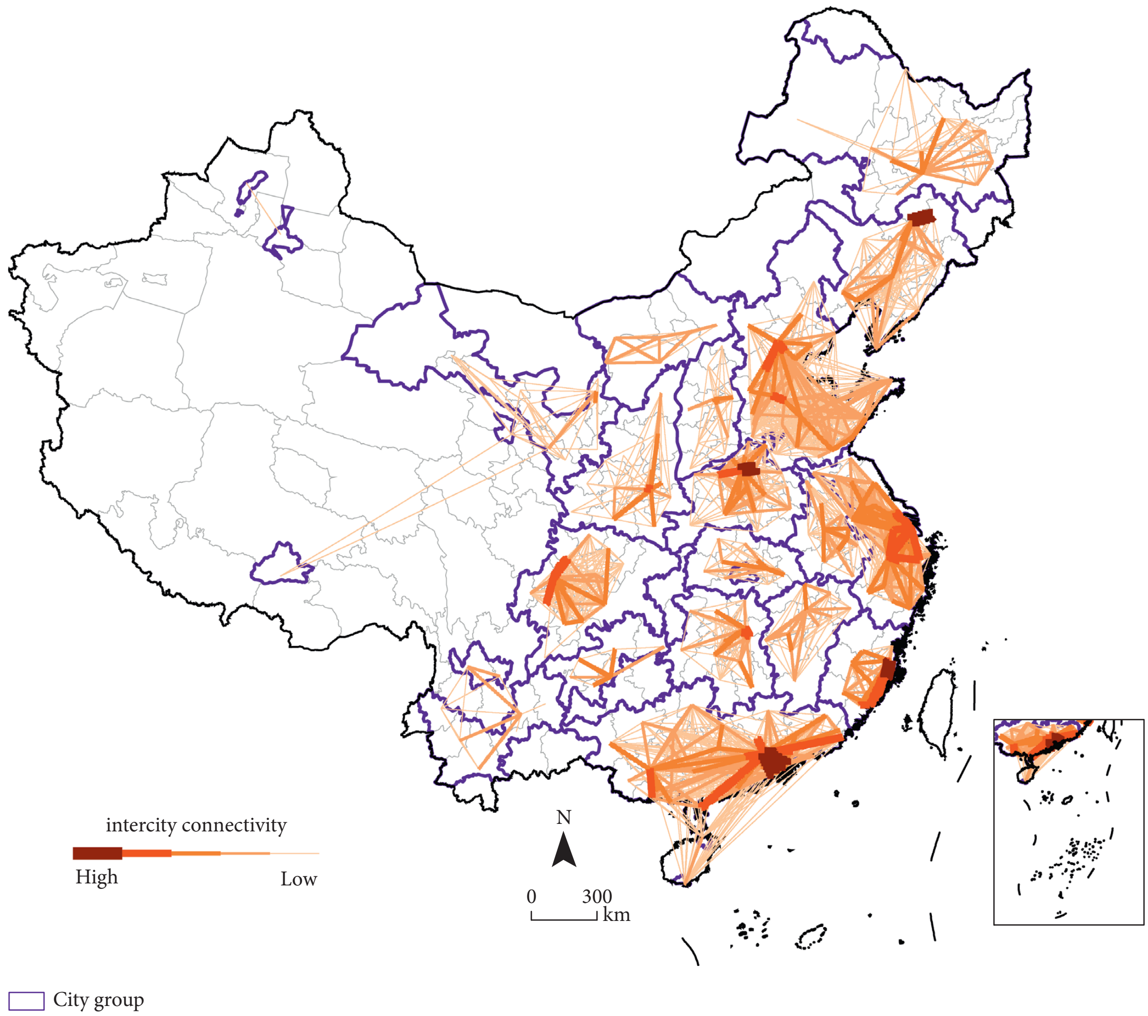


FIGURE 1: Spatial networks of city groups in China.

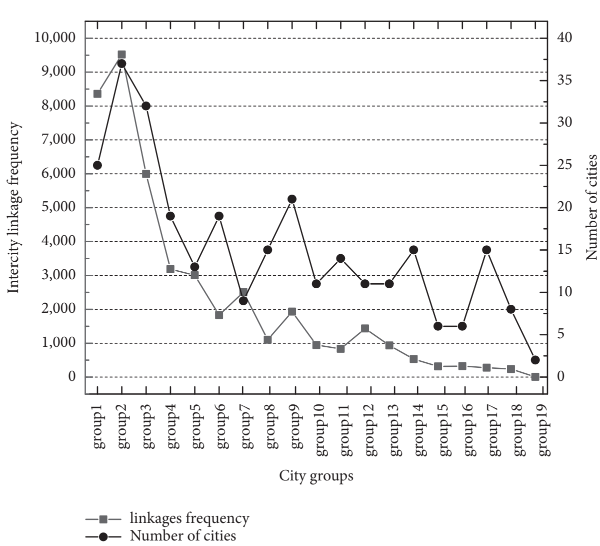


FIGURE 2: Structural attributes of city groups in China.

periphery structures of city groups. The core structures are the embryonic structures of megaregions, so they provide a foundation for identifying the spatial boundaries of megaregions in China. The results are shown in Figure 5.

On the whole, the city groups in China exhibit the core-periphery structures embedded in the city networks. They essentially form ringed nested structures, with regional central cities at the core and regional peripheral cities at the edge. The core structures are larger than the periphery structures. In addition, the core structures have a variety of irregular spatial patterns. The different-color core structures in Figure 5 represent the core structures of corresponding city groups. For example, Core Structure 1 corresponds to City Group 1, and the other core structures are similarly shown. From the composition of spatial structures, the core structures of China's city groups are mainly based around a regional administrative or economic center that forms a spatially adjacent and compact urban agglomeration with surrounding cities. Periphery structures, on the other hand, are mainly composed of cities located on the edges of city groups and a small number of cities in city groups that lack connections to other cities in the region. The ratios of the core structures to the periphery structures of city groups differ, but the number of cities in core structures is mostly greater than the number of cities in corresponding periphery

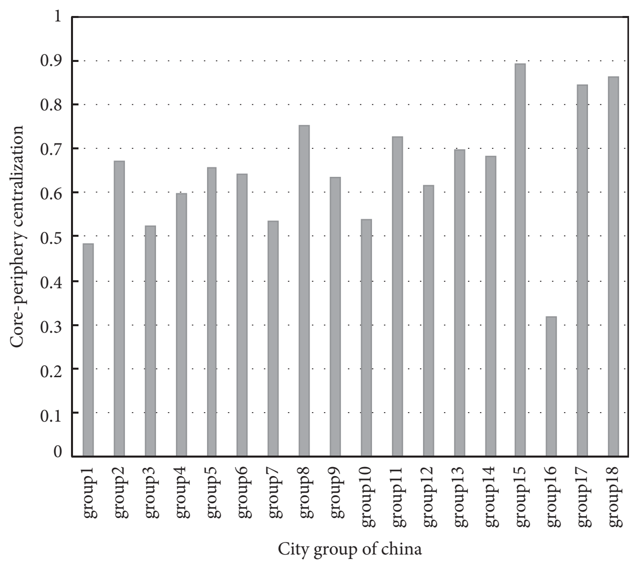


Figure 3: Centralization coefficients of the core-periphery structures of city groups in China.

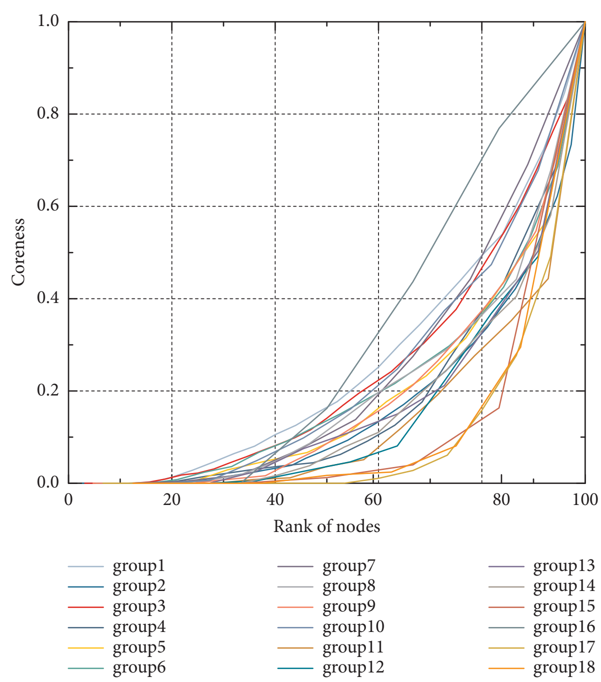


Figure 4: Coreness distribution of city groups in China.

structures. Of course, there is also a small number of periphery structures that contain more cities than corresponding core structures, which is determined by the inner structure of city networks of city groups.

Moreover, the core structures of the various city groups have a variety of irregular spatial distribution patterns, which can be roughly divided into three types: spatially adjacent and compact clustered structure; spatially separate or discontinuous multigroup structure; and belt structure with obvious directionality. Due to the spatial interactions and superpositions during the development of city networks [5], the core-periphery structures of some city groups display notable spatial fragmentation. Core structures of city groups are not affected by geographical distance decay that exhibit in the spatially adjacent and compact clustered type, but they are more common in the spatially separate or discontinuous multigroup type. This is due to the nature of "space of flows" and city networks.

3.3. Delineating the Spatial Boundaries of Megaregions. Based on the above analysis, we separate the core and periphery structures in the city networks of city groups and extract the core structures of the city groups. These core structures are the embryonic structures of megaregions, so they provide a foundation upon which to delineate the spatial boundaries of megaregions. The results show that the core structures of city groups are largely spatially coupled with the distribution of megaregions in China, which provides a new approach and understanding for comprehensively defining the spatial boundaries of megaregions.

With reference to the basic conditions for the development of megaregions and the above analysis of the core-periphery structures of city groups, we initially identify 21 underlying megaregions and delineate their corresponding spatial boundaries. The 21 megaregions are the Yangtze River Delta (YRD), Pearl River Delta (PRD), Beijing-Tianjin-Hebei (BTH), Chengdu-Chongqing (CCQ), Shandong Peninsula (SDP), Central Plains (CPL), Central and Southern Liaoning (LNP), Western Taiwan Strait (WTS), Guanzhong Plain (GZP), Changsha-Zhuzhou-Xiangtan (CZT), Hohhot-Baotou-Ordos-Yulin (CIM), Lanzhou-Xining (LXN), Central Shanxi (CSX), Wuhan (WUH), Central Anhui (CAH), Northern Jiangxi (NJX), Southern Guangxi (SGX), Central Guizhou (CGZ), Central Yunnan (CYN), Harbin-Daqing-Qiqihar (HAB), and Central Jilin (CJL) megaregions. Their specific geographic locations and boundaries are shown in Figure 6.

In terms of the foundations and current state of socioeconomic development of megaregions in China, the central and eastern megaregions have high population density, large economies, and deepening production networks, which promote functional divisions of labor and collaborative development between cities. They have good foundations for development in terms of their size, flows of factors of production, and city networks, which provide excellent conditions for the formation and development of megaregions. There are relatively few megaregions in western China, however. With the exception of the Chengdu-Chongqing megaregion, megaregions in the western region are smaller, with issues such as lower population density and relatively small economies. Moreover, the cities in some megaregions are scattered and spatially separated, so the intercity connections require crossing sizable geographic distances, which restricts the development of regional city networks. The megaregions in western China also suffer from uneven development. Most of their

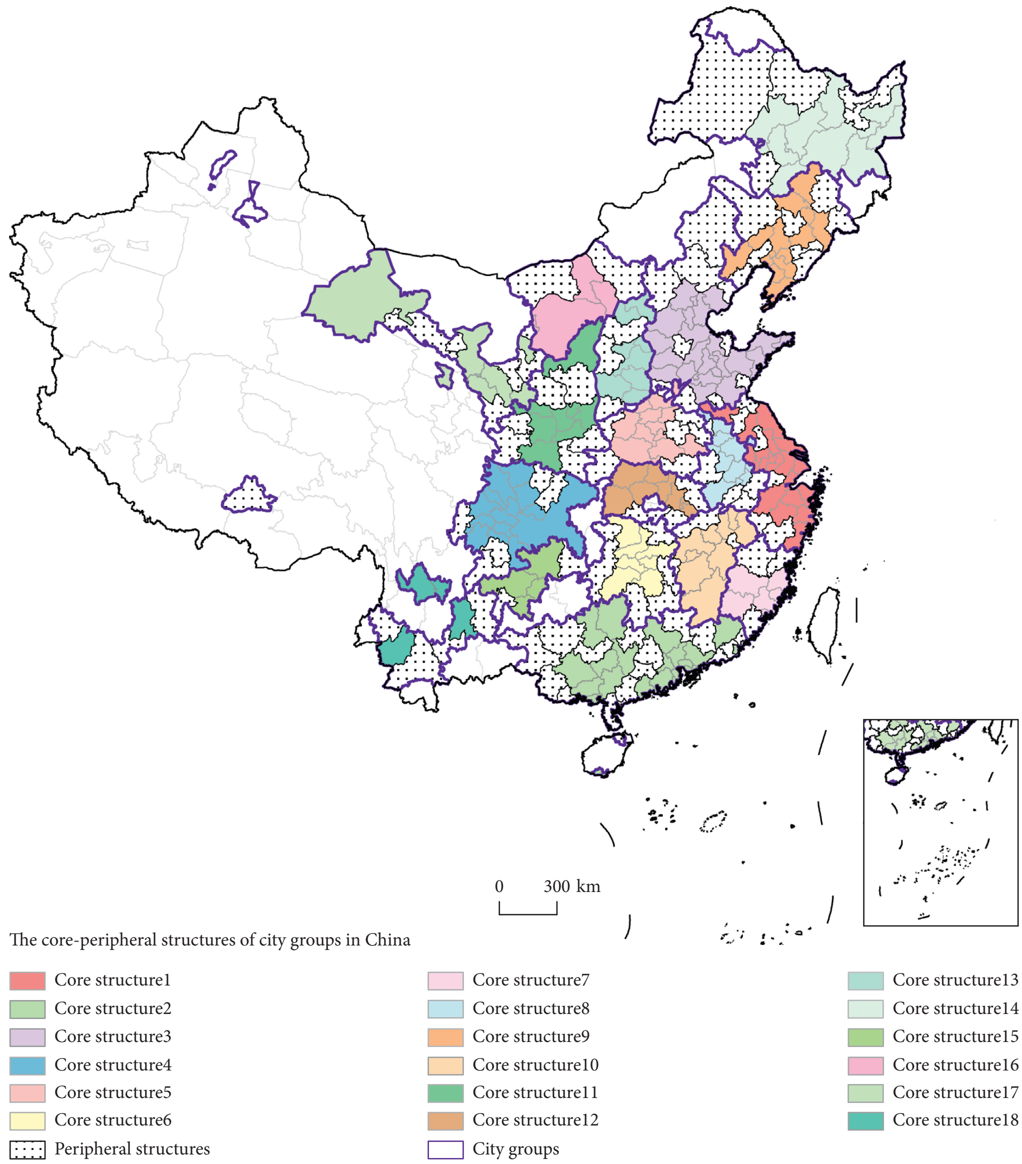


FIGURE 5: The core and periphery structures of city groups in China.

core cities are in the agglomeration stage of development, with a high degree of urban primacy, and significant polarization of central cities is relatively common. The megaregions in western China are also still developing in terms of their spatial structures, organizational patterns, and functional divisions of labor.

4. Discussion

Overall, the results reveal the spatial differentiation among megaregions in China. Central and eastern China have numerous, large, and a high density of megaregions, while the western region has relatively few. The latter also differs notably from mature megaregions in terms of rank sizes, urban systems, and functional divisions of labor, with some currently not deserving being megaregions. Some regions are disadvantaged by their basic conditions, such as geographical location, carrying capacity on resource and environment, and socioeconomic development, thus hampering the development of megaregions.

Compared with the 19 megaregions proposed in the 13th Five-Year Plan (2016–2020) of China, the megaregions identified in this study both overlap and differ to some extent. The main differences are the inclusion of the Middle Yangtze River, Harbin-Changchun, Northern Ningxia, and Northern Xinjiang megaregions in the 13th Five-Year Plan and their exclusion from this study, and the inclusion of the Central Anhui megaregion in this study and its exclusion from the 13th Five-Year plan.

Regarding the Middle Yangtze River megaregion, as we know, there are natural geographical obstacles between the three provinces of Hubei, Hunan, and Jiangxi, including the Mufu, Jiuling, and Luoxiao mountains range. Thus, the three provinces have so far failed to achieve functional integration of their urban areas. Moreover, the megaregions of the three provinces have tended to develop spatial connections of city

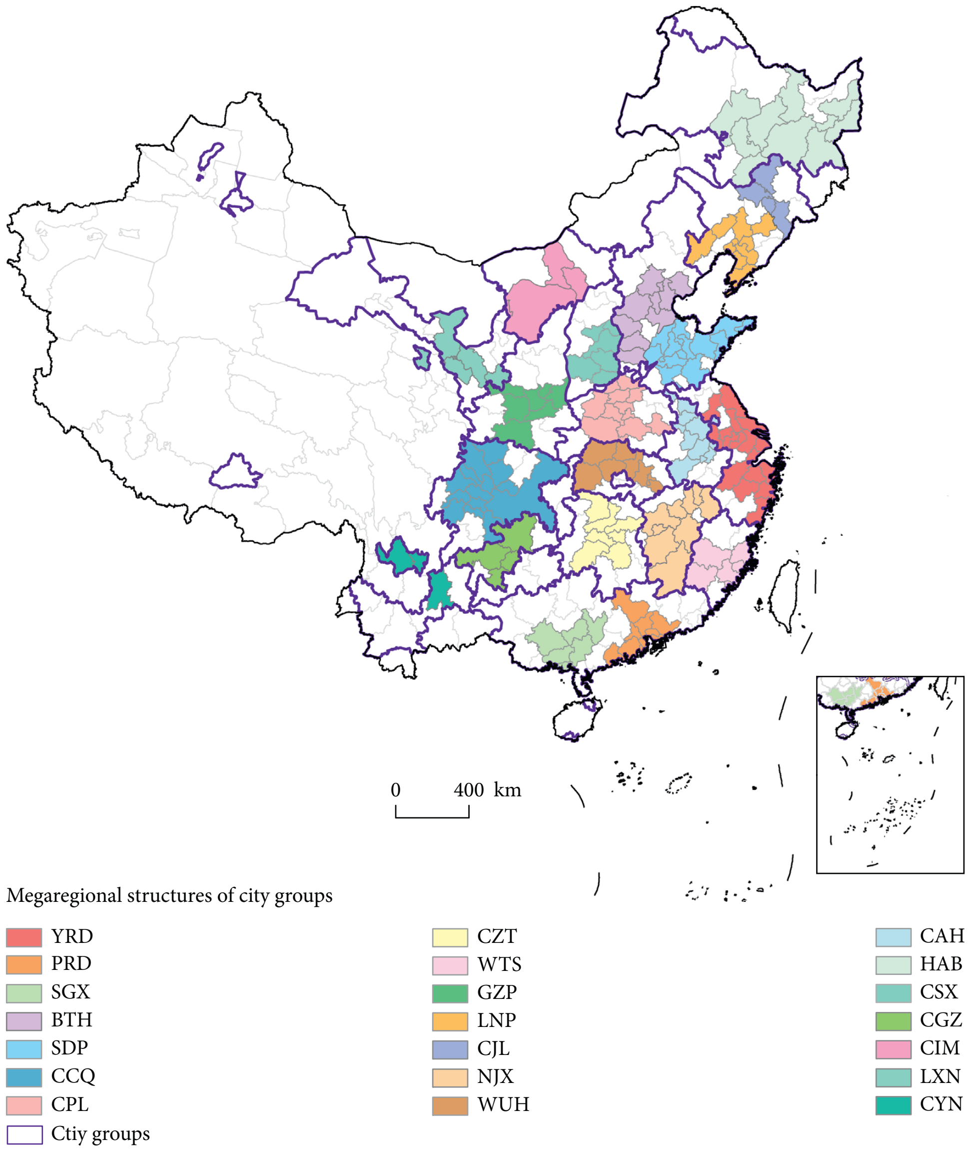


FIGURE 6: Megaregional structures of city groups in China.

networks in isolation and produce three independent city groups in the functional region system. The Harbin-Changchun megaregion also has obvious spatial separation and scattered distribution. In the process of dividing the city groups, several cities in central Jilin showed closer spatial connections with Liaoning Province, while the Harbin-Changchun megaregion does not show the tendency of integrated development. As for the Northern Ningxia megaregion, some cities in Ningxia are divided into groups with Gansu, Qinghai, and Tibet during the division of city groups, which means that the Northern Ningxia megaregion has not been independent. For the Northern Xinjiang megaregion, there are only two prefectural-level cities that are spatially distanced, so we could not identify a basis for the development of an independent megaregion. The Central Anhui megaregion identified in this article is not mentioned in the 13th Five-Year Plan.

5. Conclusions

As large-scale agglomerative landscapes of urbanized regions, megaregions are huge and complex systems that integrate multiple elements. From the perspective of spatial connotation, megaregions have been an important type of functional region system. The economic connections and functional divisions of labor, as the most spatial characteristics that reflect the nature of megaregions, are the most crucial foundations for the development of megaregions. Therefore, on the basis of the city network perspective, we attempt to integrate the community detection and core-periphery profile algorithms to identify the city groups of city networks in China and delineate the megaregions among city groups in China so as to provide a methodological framework based on relational geography and functional connections for determining the spatial boundaries of megaregions.

In the end, we identify the following 21 megaregions among city groups in China and tentatively delineate the corresponding spatial boundaries. Specifically, they are the Yangtze River Delta, Pearl River Delta, Beijing-Tianjin-Hebei, Chengdu-Chongqing, Shandong Peninsula, Central Plains, Central and Southern Liaoning, Western Taiwan Strait, Guanzhong Plain, Changsha-Zhuzhou-Xiangtan, Hohhot-Baotou-Ordos-Yulin, Lanzhou-Xining, Central Shanxi, Wuhan, Central Anhui, Northern Jiangxi, Southern Guangxi, Central Guizhou, Central Yunnan, Harbin-Daqing-Qiqihar, and Central Jilin megaregions.

As complex systems of urban-regional development, megaregions are the functional regions of large-scale urbanized agglomerative areas. Therefore, the delineation of the spatial boundaries of megaregions requires a comprehensive evaluation of multiple dimensions, rather than relying solely on one certain aspect. In this study, we introduce network science to the delineation of spatial boundaries of megaregions from the city network perspective. This analytical framework shows great potential in the identification of functional regions, which could also be one of the contributions of this article.

Although we use highway traffic flows to delineate the spatial boundaries of megaregional regions in China at the scale of the prefecture level and above the city, the intercity traffic flow used in this study just represents one type of intercity flows and the results are still preliminary. In the future, we still need to combine multiple-source data of networks and flows for complement and cross-checking [43, 44] and then comprehensively evaluate the spatial boundaries of megaregions based on the knowledge of the regional background.

Data Availability

All related traffic flow data in this paper are available upon request.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grants Nos. 41901154 and 42130508), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA20010102), and the Second Tibetan Plateau Scientific Expedition and Research Program (Grant No. 2019QZKK1007).

References

[1] J. Harrison and M. Hoyler, *Megaregions: Globalization's New Urban Form?*, Edward Elgar, Cheltenham, UK, 2015.

[2] S. Wheeler, "Regions, megaregions, and sustainability," *Regional Studies*, vol. 43, no. 6, pp. 863–876, 2009.

[3] J. Marull, V. Galletto, E. Domene, and J. Trullén, "Emerging megaregions: a new spatial scale to explore urban sustainability," *Land Use Policy*, vol. 34, pp. 353–366, 2013.

[4] H. Taubenböck and M. Wiesner, "The spatial network of megaregions-Types of connectivity between cities based on settlement patterns derived from EO-data," *Computers, Environment and Urban Systems*, vol. 54, pp. 165–180, 2015.

[5] W. Chen, W. Liu, W. Ke, and N. Wang, "Understanding spatial structures and organizational patterns of city networks in China: a highway passenger flow perspective," *Journal of Geographical Sciences*, vol. 28, no. 4, pp. 477–494, 2018.

[6] W. Chen, O. Golubchikov, and Z. G. Liu, "Measuring polycentric structures of megaregions in China: linking morphological and functional dimensions," *Environment and Planning B: Urban Analytics and City Science*, vol. 48, no. 8, pp. 2272–2288, 2020.

[7] Y. M. Ning, "The definition and role of the large urban agglomerations: role of large urban agglomeration in regional development," *Scientia Geographica Sinica*, vol. 31, no. 3, pp. 257–263, 2011.

[8] C. L. Fang, "Progress and the future direction of research into urban agglomeration in China," *Acta Geographica Sinica*, vol. 69, no. 8, pp. 1130–1144, 2014.

[9] W. Chen and C. L. Xiu, "A rethinking of the theoretical connotation of megaregion in the new era," *Progress in grography*, vol. 40, no. 5, pp. 848–857, 2021.

[10] J. Gottmann, *Megalopolis: The Urbanized Northeastern Seaboard of the United States*, Twentieth Century Fund, New York, 1961.

[11] M. Dewar and D. Epstein, "Planning for 'megaregions' in the United States," *Journal of Planning Literature*, vol. 22, pp. 108–124, 2007.

[12] J. E. Innes, D. E. Booher, and S. Di Vittorio, "Strategies for megaregion governance," *Journal of the American Planning Association*, vol. 77, pp. 55–67, 2010.

[13] H. W. Yeung, "Rethinking relational economic geography," *Transactions of the Institute of British Geographers*, vol. 30, no. 1, pp. 37–51, 2005.

[14] P. Sunley, "Relational economic geography: a partial understanding or a new paradigm?" *Economic Geography*, vol. 84, no. 1, pp. 1–26, 2008.

[15] A. Amin, "Regions unbound: towards a new politics of place," *Geografiska Annaler Series B Human Geography*, vol. 86, pp. 33–44, 2004.

[16] M. Jones, "Phase space: geography, relational thinking, and beyond," *Progress in Human Geography*, vol. 33, pp. 487–506, 2009.

[17] G. D. Nelson and A. Rae, "An economic geography of the United States: from commutes to megaregions," *PloS one*, vol. 11, no. 11, p. e0166083, 2016.

[18] S. Sobolevsky, M. Szell, R. Campari, T. Couronné, Z. Smoreda, and C. Ratti, "Delineating geographical regions with networks of human interactions in an extensive set of countries," *PLoS noe*, vol. 8, p. e81707, 2013.

[19] C. Karlsson and M. Olsson, "The identification of functional regions: theory, methods, and applications," *The Annals of Regional Science*, vol. 40, no. 1, pp. 1–18, 2006.

[20] S. Gao, K. Janowicz, and H. Couclelis, "Extracting urban functional regions from points of interest and human activities on location-based social networks," *Transactions in GIS*, vol. 21, no. 3, pp. 446–467, 2017.

[21] Li. Wang, Y. Deng, and W. Y. Niu, "The definition and identification of urban agglomerations," *Acta Geographica Sinica*, vol. 68, no. 8, pp. 1059–1070, 2013.
[22] J. H. Pan and W. L. Dai, "Identification of urban hinterlands and urban agglomerations in China based on network analysis," *Progress in Geography*, vol. 36, no. 6, pp. 667–676, 2017.
[23] L. Zhou, Q. Zhao, and F. Yang, "Identification of urban agglomeration boundary based on POI and NPP/VIIRS night light data," *Progress in Geography*, vol. 38, no. 8, pp. 840–850, 2019.
[24] X. Gao, Z. Xu, F. Niu, and Y. Long, "An evaluation of China's urban agglomeration development from the spatial perspective," *Spatial Statistics*, vol. 21, pp. 475–491, 2017.
[25] Q. Zhang, Y. F. Hu, J. Y. Liu, Y. Liu, W. Ren, and J. Li, "Identification of urban clusters in China based on assessment of transportation accessibility and socio-economic indicators," *Acta Geographica Sinica*, vol. 66, no. 6, pp. 761–770, 2011.
[26] W. Chen, W. Liu, and Z. Liu, "Integrating land surface conditions and transport networks to quantify the spatial accessibility of cities in China," *Journal of Maps*, vol. 16, no. 1, pp. 6–12, 2020.
[27] S. Su, Z. Liu, Y. Xu, J. Li, J. Pi, and M. Weng, "China's megaregion policy: performance evaluation framework, empirical findings and implications for spatial polycentric governance," *Land Use Policy*, vol. 63, pp. 1–19, 2017.
[28] P. Khanna, *Connectography: Mapping the Future of Global Civilization*, Random House, 2016.
[29] W. Ke, W. Chen, and Z. Yu, "Uncovering spatial structures of regional city networks from expressway traffic flow data: a case study from Jiangsu Province, China," *Sustainability*, vol. 9, no. 9, p. 1541, 2017.
[30] J. Harrison and H. Gu, "Planning megaregional futures: spatial imaginaries and megaregion formation in China," *Regional Studies*, vol. 55, no. 1, pp. 77–89, 2021.
[31] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, 2008.
[32] J. P. Boyd, W. J. Fitzgerald, and R. J. Beck, "Computing core/periphery structures and permutation tests for social relations data," *Social Networks*, vol. 28, no. 2, pp. 165–178, 2006.
[33] G. Csardi and T. Nepusz, "The igraph software package for complex network research. InterJournal," *Complex Systems*, vol. 1695, no. 5, pp. 1–9, 2006.
[34] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
[35] P. Pons and M. Latapy, "Computing communities in large networks using random walks," *Journal of Graph Algorithms and Applications*, vol. 10, no. 2, pp. 191–218, 2006.
[36] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review A*, vol. 70, no. 6, p. 066111, 2004.
[37] A. Lancichinetti and S. Fortunato, "Community detection algorithms: a comparative analysis," *Physical Review A*, vol. 80, no. 5, p. 056117, 2009.
[38] C. Zhong, S. M. Arisona, X. Huang, M. Batty, and G. Schmitt, "Detecting the dynamics of urban structure through spatial network analysis," *International Journal of Geographical Information Science*, vol. 28, no. 11, pp. 2178–2199, 2014.
[39] S. P. Borgatti and M. G. Everett, "Models of core/periphery structures," *Social Networks*, vol. 21, no. 4, pp. 375–395, 2000.
[40] P. Holme, "Core-periphery organization of complex networks," *Physical Review A*, vol. 72, 2005.
[41] M. R. Da Silva, H. Ma, and A. P. Zeng, "Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks," *Proceedings of the IEEE*, vol. 96, no. 8, pp. 1411–1420, 2008.
[42] R. F. Della, F. Dercole, and C. Piccardi, "Profiling core-periphery network structure by random walkers," *Scientific Reports*, vol. 3, no. 3, pp. 1–8, 2013.
[43] W. Chen, C. L. Xiu, W. Q. Ke, and Z. Yu, "Hierarchical structures of China's city network from the perspective of multiple traffic flows," *Geographical Research*, vol. 34, no. 11, pp. 2073–2083, 2015.
[44] W. Chen, N. Wang, and Z. Liu, "Mapping disciplinary differences of knowledge production and collaboration networks in China," *Regional Studies, Regional Science*, vol. 7, no. 1, pp. 210–213, 2020.

Hindawi
Complexity
Volume 2021, Article ID 7082162, 13 pages
https://doi.org/10.1155/2021/7082162

Research Article

Spatial Patterns of the Urban Agglomeration of the Yellow River Ji-Shaped Bend Based on Space of Multiple Flows

Mengyao Guo,[1] **Fang Wang**,[1] **Ruixi Zhao**,[2] **and Xi Guo**[1]

[1]*School of Public Administration, Inner Mongolia University, Hohhot 010020, China*
[2]*Business College, Hebei Normal University, Shijiazhuang 050024, China*

Correspondence should be addressed to Fang Wang; wangf741@163.com and Xi Guo; 13327118080@163.com

Received 23 September 2021; Accepted 8 December 2021; Published 28 December 2021

Academic Editor: Jun Li

Previous studies on the space of flows have mainly focused on a single flow element and have given less consideration to the joint effect of complex correlations among multiple flows. To fill this gap, in the context of the coordinated development of the urban agglomeration of the Yellow River Ji-shaped bend (UAYB), in this study, three flows, namely population flow, logistics flow, and information flow, are selected to research the spatial patterns of the UAYB. The results show the following: (1) The information flow among cities in the UAYB is the strongest, followed by logistics flow and population flow. (2) Hohhot, Ordos, Yinchuan, and Taiyuan are the core cities and have attracted more flows to converge there; different flows have formed different central features. (3) Regarding the correlation of the three flows, information flow has a strong correlation with the other two flows. (4) From the perspective of the joint effects of the three flows considered, the hierarchy of UAYB is dominated by Hohhot and Taiyuan and sub-dominated by Yinchuan and Ordos; four prominent city groups have been formed with these four cities as the center. On the basis of our results, we put forward some recommendations on the integrated development of cities at various levels within the UAYB to provide a reference for its spatial optimization strategy.

1. Introduction

The urban agglomeration of the Yellow River Ji-shaped bend (UAYB) is a concept that was proposed by the Sixth Meeting of China's Central Finance and Economics Commission on 3 January 2020. "Ji" is a Chinese character that has a shape that is basically the same as the letter "n" in the English alphabet, and its shape is very similar to the shape of the middle reaches of the Yellow River. The UAYB is located in the middle "Ji"-shaped part of the Yellow River [1]. As an urban agglomeration, the UAYB is characterized by proximity and similar culture. These characteristics make the UAYB closely connected, with close geopolitical relations, a large population, goods, and information. The flow of these elements and the economic development are interactive processes [2]. In recent years, with the rapid development of the economy, much more closer connections have gradually been formed within cities of the UAYB, and the space of multiple flows are becoming increasingly complicated and prominent. At the same time, the UAYB is a typical resource-rich area, an important energy and chemical production base, and a basic industrial base in China, as well as an important ecological barrier in the Yellow River basin. Thus, with the development of the economy and the increasing pressure on resources and the environment, the cities of UAYB are also facing the urgent problem of transformation to promote high-quality synergistic development.

Multiple flows are the primary driving forces for regional development, and in the process of these flows, complex networks of different densities are formed within an urban agglomeration [3–5]. In the past, mainstream studies on the spatial structure of urban agglomerations have been based on place space, which ignored factors such as the dynamic nature of flows and the multiple spatial interactions [6–8]. After 2000, a group of researchers conducted a large number of studies on city networks by combining the theory of "space of flows" with the theory of "world cities" [9–11].

Since then, the space of flows perspective has gained more and more attention, and "flow thinking" has been invoked to understand the position of cities in an urban agglomeration. There are three ways to examine the traditional space of flows [12]: The first way is to reflect the connection between cities through transportation or communication infrastructure [13, 14]. The second way is to analyze the urban network through the corporate spatial organization and location strategy of producer service companies or multinational companies [15]. The third way is to measure the urban network through social organizations or social groups (such as skilled immigrants and elites) [16]. Although a combination of multiple approaches can consider urban agglomeration from a static to a dynamic perspective, the lack of relational data is still a key issue that restricts empirical research on the spatial patterns of urban agglomerations.

In recent years, increased dynamic data availability has enabled researchers to identify more segmented patterns of space of flows [17, 18], and therefore more and more analyses of space of various flows have emerged. They have mainly included population flow [19–23], traffic flow [24–28], and information flow [29, 30]. Most of these previous studies have constructed original-destination matrices based on the flow of a single element to construct urban networks for analysis. Meanwhile, some new methodologies on the space of flows have been developed. For example, Teixeira and Derudder created a flexible R package called SKYNET for the processing and transformation of airflow data [31]. Shu et al. proposed L-functions to detect the scale of flow aggregation at multiple scales [32]. In addition, new research fields have emerged regarding the space of flows. An example is the application of space of flows in the field of border enforcement [33] and industrial collaboration directions [34].

In summary, studies on space of flows have mainly focused on the flow of a single element and on reconstructing the regional spatial patterns through traffic flow and population flow. Furthermore, studies have mainly focused on the current situation of each metropolitan area, and the analysis of the integration and development of cities at various levels within urban agglomerations remains at an introductory level. At the same time, less consideration has been given to the joint effect of complex correlations among multiple flows on an urban spatial structure. In addition, with the deepening influence of both globalization and decentralization forces on urban agglomerations and the rapid economic growth in midwestern China, the network patterns of cities and urban agglomerations in the midwestern China need to be studied [35]; however, there have been few studies conducted on this region.

In this study, we take the UAYB as the research area and use a combination of the theory of space of flows and open big data. In order to avoid the limitations of the flow of a single element, in this study, we select multiple flows which contain three elements, i.e., human flow, logistics flow, and information flow, which are the most representative among cities. In addition, the linkage flow strength and spatial correlation of each flow are analyzed. By dividing the spatial hierarchy in the region, we explore the status and role of each city in the UAYB as well as the regional development organization model. This is expected to provide a basis for the spatial optimization strategy and development policy formulation of the UAYB integration.

2. Study Area and Methodology

2.1. Study Area. The UAYB is located in the "Ji-shaped" bend at the top of the Yellow River basin. The length of the Yellow River in this section is 921 kilometers, which accounts for about one-sixth of the total length of the Yellow River. The UAYB starts from Baiyin city in Gansu province in the west and flows through Ningxia province, Inner Mongolia, and Shaanxi province to Linfen city in Shanxi province. It includes the 3 provincial capital cities of Taiyuan, Hohhot, and Yinchuan; Wuzhong and Zhongwei of Ningxia province; Wuhai, Bayannur, Baotou, and Ordos of Inner Mongolia; Yulin of Shanxi province; and Shuozhou, Xinzhou, and Luliang of Shanxi province. In total, there are 21 cities, which covered an area of nearly 557,000 square kilometers [1] (Figure 1).

2.2. Data Source. In this study, 21 cities of the UAYB were treated as nodes and a 21×21 multivalued network matrix was constructed, which mainly included three sets of values, namely, population flow, logistics flow, and information flow. The data of population flow were obtained from the "Baidu Migration Map" (https://qianxi.baidu.com) from 22 September 2020 to 28 January 2021. The data mapped the population flow trajectory through the information of cell phone users and showed the real-time, dynamic, and intuitive daily population flow between cities. The logistics data were obtained from 10 main logistics companies in 21 cities in the study area, namely, Yuantong, Zhongtong, Best, Shentong, Yunda, SF, ZJS, Deppon, Anneng, and Tiantian. A total of 5687 logistics sites of these 10 companies were obtained from the Kuaidi 100 website on 10 July 2021 (https://www.kuaidi100.com). The information flow data was obtained from the Baidu Index (https://index.baidu.com), which is a data analysis platform based on Baidu's massive Internet user behavior data. Through the Baidu Index, we can get how big the search scale of the cities of UAYB in Baidu platform. This part of the data was collected in July 2021. In addition, the space of traffic flow was used to test the correlation of the above three flows in the paper, since it has partial overlapping with both population flow and logistics flow. The traffic flow was represented by railway traffic data, which were derived from the official website of the China National Railway Group Co., Ltd. (Beijing, China; https://www.12306.cn/index/, accessed on 10 July 2021), including all passenger train routes and times in the UAYB.

Regarding the construction of space of multiple flows, the numerical calculations and processing of space of different flows are different. The population space of flow takes the Baidu Migration big data, which mainly includes two parts: the total in-migration index and the total out-migration index. The total migration index refers to the sum of the ratio

Figure 1: The study area.

of the population migrating from each source to the current region to the total population migrating to the current region. The total out-migration index is the sum of the ratio of the population moving out from the current region to the different in-migration locations to the total population moving out from the current region. The numerical representation of the information space of flow is the average value of the Baidu search index between two cities. The logistics space of flow is the average value of the number of logistics sites between the two cities.

2.3. Methodology

2.3.1. Social Network Analysis Methods. Social network analysis methods are widely used in the study of complex network structures. In this study, we used the overall network density, the degree centrality, and the quadratic assignment procedure (QAP) to analyze multidimensional networks in terms of nodes, hierarchies, and networks.

(1) Overall Network Density. The overall network density is an indicator that reflects the overall density of intercity connections in an urban agglomeration. The larger the network density value is, the more closely connected the cities are in the urban network. The overall network density value theoretically lies within the interval 0–1, but in the actual urban agglomeration, the network with a density of 1 basically does not exist, and there are very few urban agglomerations that have a density value greater than 0.5. The calculation of (1) is as follows:

$$D = \frac{2L}{n \times (n-1)}, \quad (1)$$

where D is the overall network density, L is the actual number of connections in the network, and n is the number of nodes in the network.

(2) Degree Centrality. The degree centrality is the most direct metric to portray node centrality in network analysis. The higher the degree of a node, the higher the degree centrality of the node is, which indicates the node importance in the network. Formula (2) is as follows:

$$C_D(n_i) = \sum_{j=1}^{n} X_{ji}, \quad (2)$$

where $C_D(n_i)$ represents the degree centrality and X_{ji} means the connection strength between nodes.

(3) Quadratic Assignment Procedure. The quadratic assignment procedure (QAP) is a common method to research the relationship between different types of networks. It compares the values of each element of the network matrix

based on data permutation to obtain the correlation and regression coefficients between matrices. It also performs a nonparametric test on the coefficients, which helps to avoid problems of the collinearity arising from relational data regression. In this study, the QAP was mainly used to compare the structure similarity of different space of flows.

2.3.2. Spatial Structure Index. The spatial structure index is an improved algorithm based on the study of regional central structure, which was proposed by Hanssens in 2013 [36]. The index is in the range from 0 to 1, where 0 indicates that the regional spatial structure shows a significant unipolar development, and 1 indicates that the regional spatial structure is significantly multipolar in character. The index can be used to evaluate the degree of dispersion of regional spatial structure of the following formula:

$$\mathrm{SSI} = \begin{cases} \left(2 - \dfrac{\mathrm{SD}}{\mathrm{SD}_r}\right) \times 2^{-1}, & \mathrm{SD} < \mathrm{SD}_r \\ \\ \dfrac{\mathrm{SD}_r}{2\mathrm{SD}}, & \mathrm{SD} > \mathrm{SD}_r, \end{cases} \tag{3}$$

where SD is the standard deviation value of the flow between the nodes, SD_r is the standard deviation of the serial number of all nodes after ranking, and SSI is the spatial structure index.

2.3.3. Integrated Data Processing. The three flows of the matrix data are formed into a comprehensive network for normalizing the original data with the following formula:

$$S = \frac{x - \mathrm{Min}}{\mathrm{Max} - \mathrm{Min}}, \tag{4}$$

where x is the original data, Min is the minimum value of the matrix data, Max is the maximum value of the matrix data, and S is the normalized data.

After the normalization, S lies in the range of 0~1; 1 indicates that the two cities are most closely connected, and 0 indicates that the two cities are most distant. The normalized data are summed to characterize the linkage flow strength of cities by the following formula:

$$R_i = \sum_j R_{ij}, \tag{5}$$

where R_i denotes the intensity of flow of city I, and R_{ij} is the intensity of the network connectivity of city i with city j. In the study of the spatial structure of integrated flow of urban agglomerations, by Wang et al., the flow of each element is regarded as equally important, with a weight of 0.33 [37].

2.3.4. Dominant Flow Analysis. Nodal regions were proposed by Nystuen and Dacey in 1961 based on graph theory [38, 39]. The central idea is to distill the hierarchical structure and organizational relationships of the major core cities and hinterland regions from the complex urban network through the dominant flow filtering method. This method follows the following principles: ① The largest flow of the dominant city flows to a relatively smaller city, ② the largest flow of the secondary dominant city flows to the dominant city, ③ the largest flow of the subordinate city flows to a relatively larger city, and ④ the subordination is transitive. If city A is subordinate to city B, and city B is subordinate to city C, then city A is also subordinate to city C; that is, A is also the hinterland of C.

Table 1: Overall network density of the three flows.

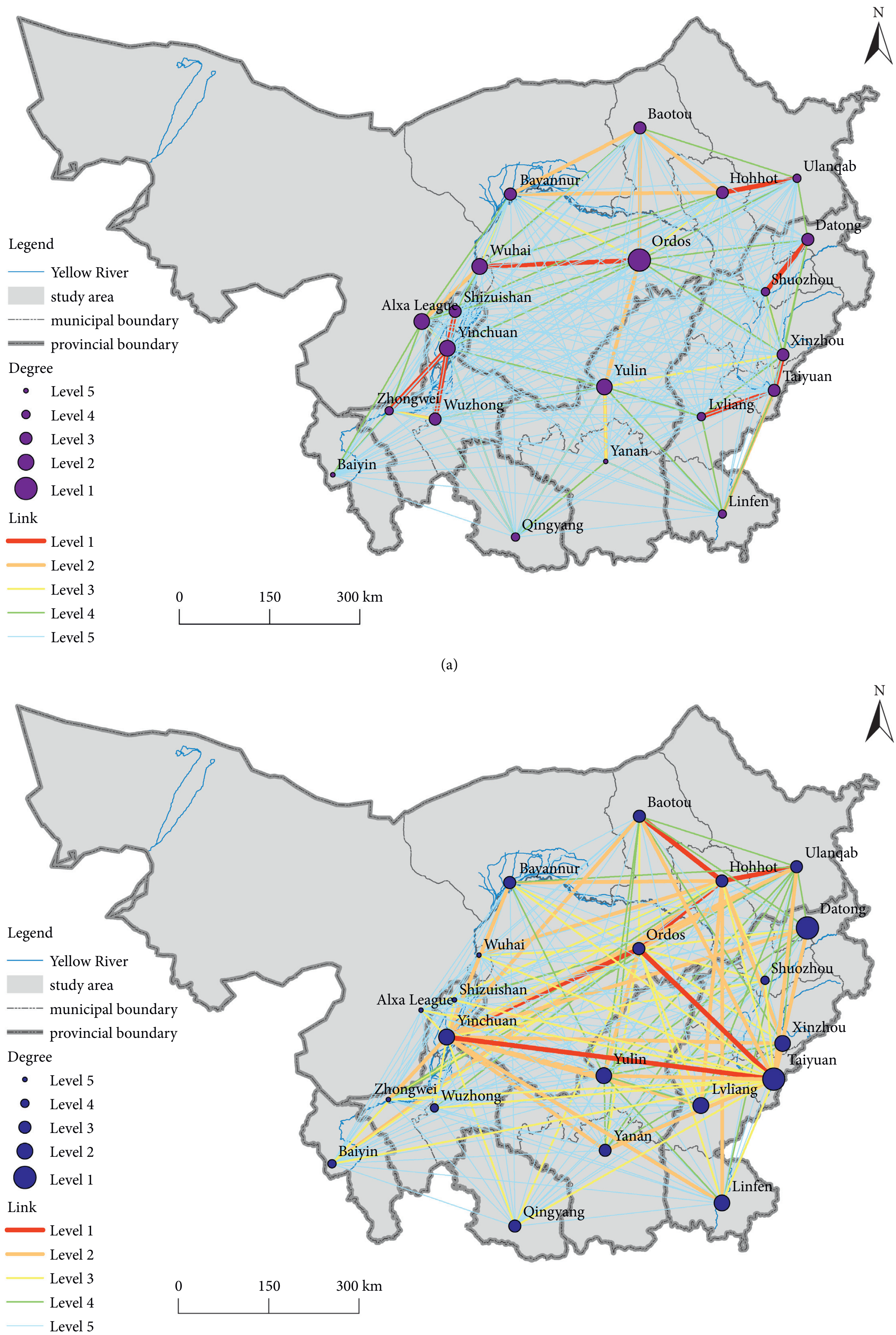
Category	Density
Information flow	0.193
Logistics flow	0.186
Population flow	0.113

3. Results

3.1. Multiple Linkages between Cities in the UAYB

3.1.1. Intensity Analysis of Multiple Linkages between Cities. The overall network density describes how closely the nodes in an urban agglomeration network are connected to each other. According to (1), we calculate the overall network density of the linkages among cities. As illustrated in Table 1, there are three different flows that differentiate the sparseness and strength of intercity connections. The information flows are a full-coverage urban network, and there are either large or small information flows between every two cities. Therefore, the network density of information flow (0.193) is also the largest. The second largest is the logistics network with a density value of 0.186. This is because the national logistics network has become more and more sophisticated with the personalization trend of increasing consumers and the significant potential development in the logistics industry in China. The overall network density has reached a high level, and logistics services basically cover all prefecture-level cities in the UAYB. Furthermore, the population flow data show that 91.43% of the cities in the UAYB have direct population linkages, but the network density value is not as high as the density values of information linkage and logistics linkage. This may be due to the low flow of population between small cities.

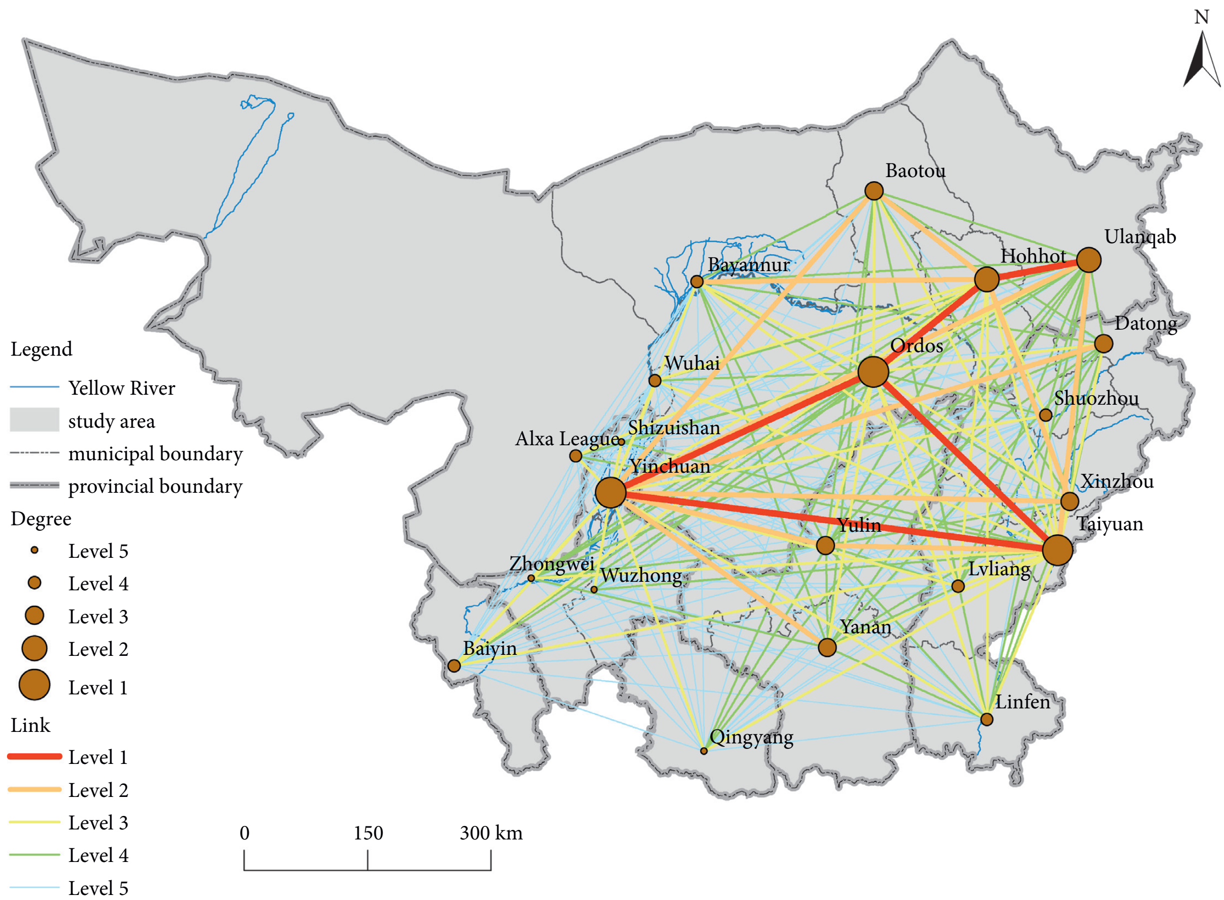
Then, we analyzed and compared the pattern of the three different flows. The degree centrality of cities was assigned to nodes, the strength of intercity linkages was assigned to lines, and they were all divided into five levels according to the natural break classification. There are both similarities and differences between the three flows (Figure 2). In terms of similarities, all cities in the UAYB are closely connected with each other, forming a complex network of multiple flows in the region. In addition, the cities with higher values of degree centrality in the three flows are Taiyuan, Ordos, and Yinchuan, which are provincial capitals or cities with higher levels of economic development and which play a crucial role in this region. In terms of differences, the distribution of the population flow is significantly different as compared with the logistics and information flows. We found that the larger the value of population flow, the more



(a)

(b)

FIGURE 2: Continued.



(c)

FIGURE 2: Functional linkage intensities of the three different flows: (a) population flow; (b) logistics flow; (c) information flow.

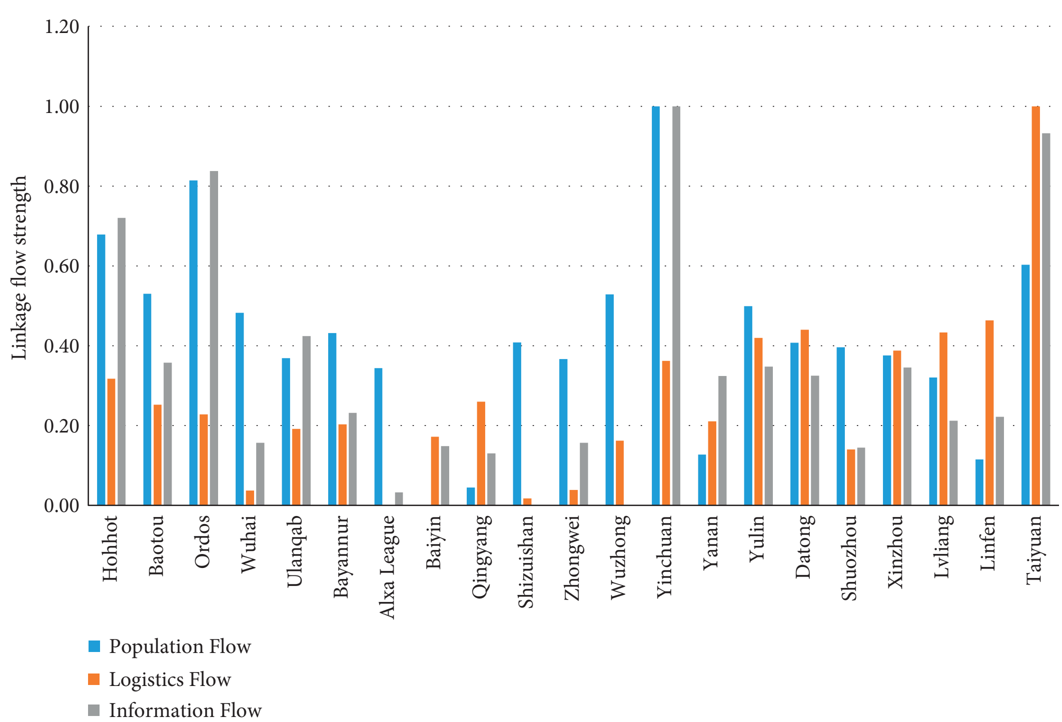


FIGURE 3: The linkage flow strengths of the three flows for each city in the UAYB.

dispersed it is, and no spatial polar core interaction occurs. This also confirms that population flows are obviously influenced by geographical distance.

Specifically, in terms of population flow (Figure 2(a)), Ordos, as a core node, generates the most population flow with surrounding cities. Among them, the population flow with Wuhai is particularly frequent. With strong economic advantages and convenient transportation, Ordos becomes the leading city in the population space of flow. It has the strongest population attraction in the region. It is followed by Yinchuan, Taiyuan, and Hohhot. The cities have all formed large population flows with their neighboring cities due to their attractiveness generated by their advantages in administration, economy, transportation, infrastructure, and employment. In addition, the population flows among fringe regions in the UAYB are also very obvious. For example, the intensity of population flows between Hohhot and Ulanqab, and Datong and Shuozhou, as well as Taiyuan and Xinzhou, is relatively high. This is due to the fact that the cities have similar cultural background and living customs and there are numerous transportation options among these cities, which has led to a large number of short distance population flows among them.

The logistics flow network of UAYB is maturely developed (Figure 2(b)). This is in line with the current background of rapid development of the logistics industry in China. It is found that the cities with high degree centrality are concentrated in the eastern part of the whole UAYB, especially in Shanxi province. There are three reasons for this. Firstly, Shanxi province was once a major coal resource province in China and has an extremely developed transportation system. The original transportation system has become the basis for the rapid development of its logistics space of flow. Secondly, Shanxi province is located in the east of the UAYB, where it is radiated by nearby developed regions, such as Beijing, Tianjin, and Shijiazhuang. Finally, Taiyuan belongs to the national logistics distribution point and is an important logistics hub in North China. In terms of the intensity of logistics flow, it basically conforms to the core interaction type; i.e., the high-intensity logistics flow is concentrated between the core cities, such as Yinchuan, Ordos, Taiyuan, Hohhot, and Baotou.

The information flow network is very similar to the logistics flow network (Figure 2(c)), in that the flows with higher spatial linkage intensity both occur in the central triangle of the region. In addition, except for Taiyuan, there are obvious differences; i.e., the core node cities within these two networks are different. The information space of flow has formed a complete regional network consisting of three core cities and two sub-core cities. The three leading cities are Ordos, Yinchuan, and Taiyuan, forming a spatial "triangle" to lead the information flows of cities in the UAYB. Furthermore, Hohhot and Ulanqab are two sub-core secondary cities, and the intensity of the information flow between them is relatively high. In addition, the information flow between these two cities and Ordos is also relatively frequent. In this way, the sub-core cities also realize real-time information linkage with the core "triangle." The information flow connects the entire region, embodies the strong cross-territory and cross-level linkages of the information network, and enables more synergistic development among the central cities.

Table 2: Spatial structure index (SSI) of the three types of flows.

Category	SSI
Population flow	0.883
Logistics flow	0.135
Information flow	0.535

Finally, comparative analyses of the linkage flow strengths of the three different flows were performed for each city (Figure 3), and the spatial structure index (SSI) of the three linkages was calculated separately (Table 2). The linkage flow strength is obtained by summing and normalizing the matrices of each flow separately ((4) and (5)). Its horizontal coordinates are the cities of the UAYB, and its vertical coordinates are the normalized linkage flow intensities (Figure 3). It can be seen that the overall spatial patterns of population flow, logistics flow, and information flow are similar in the UAYB. The overall spatial pattern of the population space of flow is multipolar with an SSI value of 0.883. In addition, it is found that Yinchuan, Taiyuan, Ordos, and Hohhot have larger population movements. The overall spatial pattern of the information space of flow also tends to be multipolar with an SSI value of 0.535. In contrast, the spatial pattern of the logistics space of flow tends to be relatively unipolar with an SSI value of 0.135; in particular, Taiyuan has an absolute advantage in the UAYB region in terms of the strength of logistics flow.

3.1.2. Correlation Analysis of Multiple Linkages between Cities. In order to verify the correlation between different flows, we calculate the QAP values. The details are illustrated in Table 3. It can be seen that the information space of flow has a strong correlation with both population and logistics space of flows, with QAP correlation coefficients of 0.351 and 0.544, respectively. This is because the information space of flow is a full-coverage network, with more or less connectivity between every two cities. The spatial correlation between logistics space and population space of flow is low with a value of 0.231. Thus, traffic space of flow is introduced to further verify the correlation between these three flows. After validation with the traffic flow separately, it can be seen that the correlation between the population flow and the traffic flow is high. This is because the traffic data used in this study is that of the passenger trains between cities, and trains are the most frequent way for residents to travel. At the same time, population and goods usually take very different modes of transportation. Therefore, the correlation between traffic space of flow and logistics space of flow is low. In addition, it is obvious that information flow is also less dependent on transportation. This illustrates that information flow relies more on the Internet than on face-to-face communications.

Table 3: Quadratic Assignment Procedure (QAP) of the space of flows.

Category	Traffic flow	Population flow	Information flow	Logistics flow
Traffic flow	—	0.436 (0.000)	0.404 (0.000)	0.211 (0.010)
Population flow	0.436 (0.000)	—	0.351 (0.000)	0.231 (0.000)
Information flow	0.404 (0.000)	0.351 (0.000)	—	0.544 (0.000)
Logistics flow	0.211 (0.010)	0.231 (0.000)	0.544 (0.000)	—

Note: "—" means no data, and the values in parentheses are the significance levels of the tests.

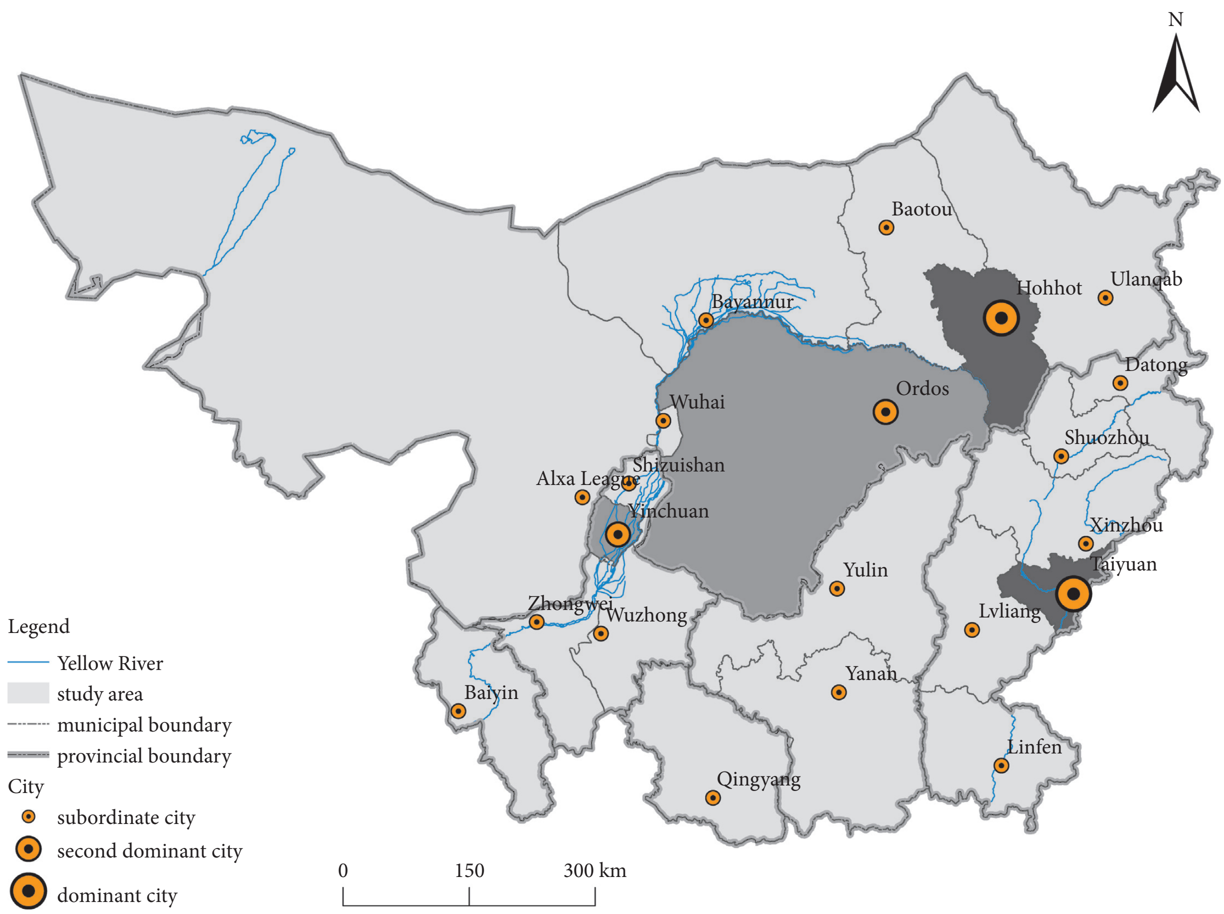


Figure 4: Spatial hierarchy of the UAYB by dominant flow analysis.

3.2. The Spatial Patterns of the UAYB

3.2.1. The Spatial Hierarchy of the UAYB. The analysis of different types of flows can only represent the spatial structure under a certain level of the UAYB, but it cannot fully reflect the spatial characteristics of each city in the study area. According to the results of the analysis in the previous section, the flows of population, logistics, and information are integrated to obtain the integrated flow ((4) and (5)). The network density of the integrated flow is 0.252, and the network connectivity is the strongest.

The spatial hierarchy of cities in the UAYB is analyzed using the dominant flow method (Figure 4). The flow direction of the dominant flow indicates that the dominant cities in the region are Hohhot and Taiyuan, which are the capital cities of Inner Mongolia Autonomous Region and Shanxi province, respectively, and also the hubs of economy, politics, and culture. Therefore, they have strong comprehensive attractiveness and have the largest radiation range in the UAYB. The second dominant cities are Ordos and Yinchuan, which are directly attracted by the dominant cities. Spatially, the dominant and second dominant cities are basically located in the middle of the entire UAYB. The remaining subordinate cities are distributed among these four dominant cities, making a complex network of urban agglomeration throughout the region.

3.2.2. The Identification of Urban Groups in the UAYB. The visualization of the first dominant flows of the integrated flow reveals that the UAYB has formed two dominant urban groups and two subdominant urban groups. The two dominant urban groups are centered in Hohhot and Taiyuan, respectively, bring together the largest dominant flows of many cities, and form a large radiation area (Figure 5).

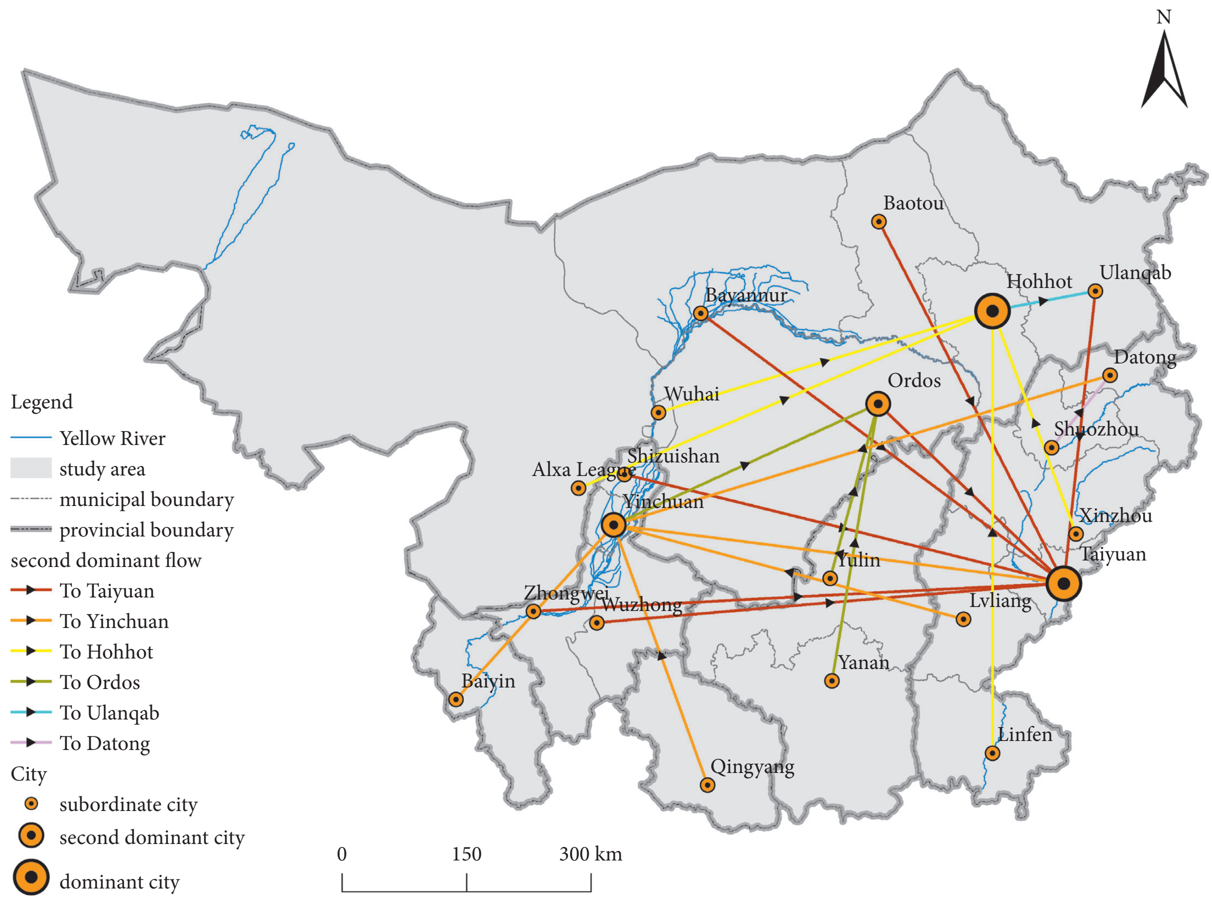
Figure 5: Spatial distribution of the first dominant flows in the UAYB.

The dominant urban group, with Hohhot as its core, is centered in the northeast corner of the UAYB. It has strong linkages with several other cities in the Inner Mongolia Autonomous Region. The three flows in the Inner Mongolia are not strongly linked within provinces, especially the population flow. The dominant urban group, with Taiyuan as the core, receives the largest number of dominant flows and has the widest radiation range, which basically covers the eastern and southern parts of the UAYB. Taiyuan is an extremely important node among the three flows, especially occupying the position of the dominant city in the logistics space of flow and information space of flow. At the same time, the subdominant city groups are centered in Ordos and Yinchuan, which receive relatively fewer dominant flows and form a radiation limited to the neighborhood areas with less radiation capacity. The number of cities of the Ordos urban group is relatively small. The main reason for the formation of this urban group is that the economic development level of Ordos is good, and its geographical location is in the center of the UAYB.

Furthermore, the flows of population and information are relatively more pronounced within the group centered in Ordos. The development of the group centered in Yinchuan has benefited from the proximity of the nearby cities, as well as the similar cultural background and folk customs, which has led to the strong flow of elements. Hence, a small interactive space of flow centered in Yinchuan has been formed. It is noteworthy that although Yinchuan received a high number of first dominant flows, it formed only a second dominant city group. The more immature development of the Yinchuan city group is due to the fact that this urban group is far away from the other three urban groups and the interaction among them is not strong.

In this study, we also mapped the second and third dominant flows to research the interaction between cities in the UAYB. The details are shown in Figure 6. In this part, the details of the three flows in the previous section are represented. For instance, a large population flow is generated between Datong and Shuozhou. Except for some changes in the more connected neighborhoods, the dominant cities are Hohhot, Taiyuan, Ordos, and Yinchuan, which continue to receive most of the second dominant flows and the third dominant flows. Such a phenomenon indicates that the dominant cities in the UAYB have stronger control and attraction than the nondominant cities, while the linkages among the subordinate cities are weak and the development of the subordinate cities faces difficulties. Regional development integration is urgently needed.

Finally, the spatial structure index of the integrated flows was calculated to determine the degree of spatial dispersion of the UAYB, and the result was 0.606. This result indicates that the development dynamics of the UAYB is slightly



(a)

(b)

Figure 6: Distribution of the second and third dominant flows in the UAYB: (a) the second dominant flow; (b) the third dominant flow.

multipolar in a balanced manner. There is a need to promote the development of the dominant cities and play a more driving role in the region. Deepening the intraregional cycle will enable the development of the subordinate cities (e.g., less developed cities) and will continue to reduce the gap between the rich and the poor; therefore, it can also promote the process of regional integration in the UAYB.

4. Conclusions and Discussion

This study provides an in-depth analysis of the UAYB city agglomeration in terms of multiple linkages and regional spatial patterns, respectively. The main conclusions are as follows: (1) In terms of overall network density, the three flows all have strong development potential. Information flow received the highest score of overall network density, followed by logistics space of flow and population flow. (2) In terms of node centrality and linkage intensity, Hohhot, Ordos, Yinchuan, and Taiyuan have become the core cities and attract more types of flows in the UAYB to converge there. In contrast, the cities in the south and northwest of the UAYB are less centralized, and most of their population, materials, and technology flow to the leading cities. Regarding the population flow, Ordos, in the central part of the UAYB, has a prominent position. Taiyuan and Datong, two cities in Shanxi province, play a pivotal role in the logistics space of flow. A spatial triangle is formed in the information space of flow. The apexes are Ordos, Yinchuan, and Taiyuan, and information flows frequently among the three. (3) Regarding the correlation of the three flows, the information space of flow has a strong correlation with the other two flows. (4) The hierarchy of the UAYB is dominated by Hohhot and Taiyuan and subdominated by Yinchuan and Ordos. Four prominent city groups are formed with these four cities as the center. However, the regional comprehensive structure of the UAYB presents a multicentered, flat, and balanced situation; only the logistics space of flow shows a unipolar trend.

In order to strengthen the UAYB's overall radiation capacity and spatial linkages, as well as promote the rapid, coordinated, and comprehensive development of urban agglomerations, the following policy implications are recommended based on the results of this study: (1) Internal growth poles should be cultivated, and urban growth poles should be created. The special industries in the growth pole area, such as grassland agriculture and animal husbandry, tourism, and rare earth industry, should be taken advantage of. The transferred industries from the developed eastern regions should be actively undertaken and combined with the UAYB's own situation to achieve complementary advantages. The formation of the growth pole model should be accelerated to promote rapid development of the whole region. (2) While vigorously promoting the development of the two dominant cities of Hohhot and Taiyuan, we should also pay attention to the development and cultivation of the two second dominant cities with obvious advantages, so as to play a radiating and leading role in surrounding cities and enhance the overall concentration and radiation capacity of the UAYB. (3) On the basis of traditional resource-based industries and basic manufacturing industries, cities in the UAYB should improve their independent innovation capabilities through introducing talents and technology and providing policy encouragement, as well as vigorously developing modern service industries and high-tech industries. In particular, cities such as Hohhot and Taiyuan should give full attention to the scientific and educational advantages of their universities, accelerate the construction of high-tech industrial parks and innovation demonstration zones, transform scientific and technological advantages into industrial advantages, promote the optimization and upgrading of urban industrial structure, and reduce the environmental pressure. (4) Hohhot and Taiyuan should make full use of their advantages in resources, technology, and information to enhance their comprehensive economic strength and realize their radiating and leading role in regional development. (5) An effort should be made to break through the limitation of the scattered regional structure in the UAYB, by speeding up the construction of intercity transportation, communication, logistics, and other networks centered in Hohhot, Taiyuan, Yinchuan, and Ordos and promoting the integrated construction of infrastructure in the UAYB. The five provinces, namely, Shanxi, Gansu, Shaanxi, Inner Mongolia Autonomous Region, and Ningxia Hui Autonomous Region, should break down administrative barriers, promote the free flow of elements, and optimize the allocation of resources. In addition, we should actively promote the sharing of public resources such as tourism, culture, and health, thus strengthening economic links among cities and promoting the integration of urban agglomerations. (6) New models to align the UAYB's development with China's major national strategies should be actively explored, and the synergistic development of the UAYB and various urban clusters in China should be developed. In addition, we should actively dovetail with the "one belt and one road" strategy to gradually enhance the UAYB's opening capacity to the outside world. The UAYB will become the frontline of opening up in northern China.

The limitations and future works of this study are as follows: Firstly, the research is limited to the UAYB and does not consider the linkage between the UAYB and external areas. Future research should consider the flows of various elements from the perspective of a larger region and even the whole country to promote the interconnectedness of the UAYB across the country. Secondly, the three flows (people, logistics, and information) selected in this study cannot cover all the flows of elements between cities. Elements such as capital flow and innovation flow should also be included in the overall research scope of multiple element flows. Thirdly, the next step can be based on a multidimensional perspective to analyze space of flows. If the time scale can be integrated into existing research, it can be combined with historical data for longitudinal comparison, thus obtaining the development history of the regional network and making existing research better, such as effectively combining policies and proposing more suitable optimization strategies.

Data Availability

This paper uses the "Baidu Migration" platform to obtain population migration data on Baidu Maps (https://qianxi.baidu.com/). Train routes data are from https://www.12306.cn/index/. Logistics sites data are from https://www.kuaidi100.com.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (grant nos. 41801149 and 71864025), the Young Talents of Science and Technology in Universities of Inner Mongolia (grant no. NJTY-20-B09), the Hebei Provincial Department of Human Resource and Social Security (grant no. C20210314), and the Hebei Social Science Development Research project (grant no. C20210301048).

References

[1] F. Wang, M. Guo, X. Guo, and F. Niu, "Research on the hierarchical spatial structure of the urban agglomeration of the Yellow River Ji-Shaped Bend," *Complexity*, vol. 2021, Article ID 2293524, 13 pages, 2021.
[2] H. Zhao, W. Zhang, and T. Ha, "Network characteristics of urban population migration," *Journal of Northeastern University*, vol. 27, no. 2, pp. 169–172, 2006, in Chinese.
[3] M. Castells, *The Rise of the Network Society*, Wiley Blackwell, Hoboken, NJ, USA, 2011.
[4] E. Schoenberger and R. A. Walker, "Beyond exchange and agglomeration: resource flows and city environments as wellsprings of urban growth," *Journal of Economic Geography*, vol. 17, no. 5, pp. 935–958, 2017.
[5] L. Dai, B. Derudder, and X. J. Liu, "Generative network models for simulating urban networks, the case of inter-city transport network in Southeast Asia," *Cybergeo*, 2016.
[6] B. Jin, W. Yang, X. Li, J. Sha, and X. Wang, "A literature review on the space of flows," *Arabian Journal of Geosciences*, vol. 14, no. 13, 2021.
[7] J. W. Li, J. Qian, and Y. L. Liu, "A novel analysis method of geographical centrality based on space of flows," *ISPRS International Journal of Geo-Information*, vol. 6, no. 5, p. 15, 2017.
[8] F. Q. Niu, X. Y. Yang, and F. Wang, "Urban agglomeration formation and its spatiotemporal expansion process in China: from the perspective of industrial evolution," *Chinese Geographical Science*, vol. 30, no. 3, pp. 156–167, 2020.
[9] P. J. Taylor and B. Derudder, *World City Network: A Global Urban Analysis*, Routledge, Oxfordshire, UK, 2003.
[10] M. Burger and E. Meijers, "Form follows function? Linking morphological and functional polycentricity," *Urban Studies*, vol. 49, no. 5, pp. 1127–1149, 2012.
[11] P. J. Taylor, G. Catalano, and D. R. F. Walker, "Measurement of the world city network," *Urban Studies*, vol. 39, no. 13, pp. 2367–2376, 2002.
[12] Y. Wei, Y. Jin, D. Ma, and C. Xiu, "Impact of colored motif characteristics on the survivability of passenger airline networks in China," *Physica A: Statistical Mechanics and Its Applications*, vol. 556, Article ID 125658, 2021.
[13] C. Dong, C. L. Xiu, and Y. Wei, "Network structure of 'space of flows' in Jilin Province based on telecommunication flows," *Acta Geographica Sinica*, vol. 69, no. 4, pp. 510–519, 2014, in Chinese.
[14] B. Derudder and F. Witlox, "Mapping world city networks through airline flows: context, relevance, and problems," *Journal of Transport Geography*, vol. 16, no. 5, pp. 305–312, 2008.
[15] J. Wang, J. Li, J. Zhao, and G. Zhang, "A city network analysis based on spatial distribution of international hotel groups in China," *Human Geography*, vol. 30, no. 1, pp. 148–153, 2015.
[16] S. Scott, "Transnational exchanges amongst skilled British migrants in Paris," *Population, Space and Place*, vol. 10, no. 5, pp. 391–410, 2004.
[17] J. Raun, R. Ahas, and M. Tiru, "Measuring tourism destinations using mobile tracking data," *Tourism Management*, vol. 57, pp. 202–212, 2016.
[18] Y. J. Yang, D. Yin, Z. W. Liu, Q. X. Huang, C. Y. He, and K. Wu, "Research progress on the space of flow using big data," *Progress in Geography*, vol. 39, no. 8, pp. 1397–1411, 2020, in Chinese.
[19] J. B. Lai and J. H. Pan, "China's city network structural characteristics based on population flow during Spring Festival travel rush: empirical analysis of Tencent Migration big data," *Journal of Urban Planning and Development*, vol. 146, no. 2, 2020.
[20] J. Pan and J. Lai, "Spatial pattern of population mobility among cities in China: case study of the National Day plus Mid-Autumn Festival based on Tencent migration data," *Cities*, vol. 94, pp. 55–69, 2019.
[21] Y. Shen, "Segregation through space: a scope of the flow-based spatial interaction model," *Journal of Transport Geography*, vol. 76, pp. 10–23, 2019.
[22] S. Z. X. Luo, F. Zhen, and S. Q. Zhang, "A conceptual model and methodological framework for examining urban people flow space based on complex network perspective," *Geographical Research*, vol. 40, no. 4, pp. 1195–1208, 2021, in Chinese.
[23] F. Xue, M. Y. Li, and A. R. Dang, "Centrality and symmetry of people flow network structure of the Yangtze River delta urban agglomeration at multi-spatial scales," *Economic Geography*, vol. 40, no. 8, pp. 49–58, 2020, in Chinese.
[24] D. He, Z. X. Chen, T. Pei, and J. Zhou, "The regional and local scale evolution of the spatial structure of high-speed railway networks-A case study focused on beijing-tianjin-hebei urban agglomeration," *ISPRS International Journal of Geo-Information*, vol. 10, no. 8, 2021.
[25] W. X. Chen, W. C. Liu, and X. J. Duan, "Spatial organization evolution of railway passenger transportation in the perspective of space of flow: a case study of the Yangtze River Delta urban agglomeration," *Geographical Research*, vol. 39, no. 10, pp. 2330–2344, 2020, in Chinese.
[26] Y. J. Li, Q. D. Wu, Y. L. Zhang, H. G. Zhang, S. Fang, and S. Q. Jin, "Spatial structure and formation mechanism of expressway traffic flow network based on space of flows: a case study of Guangdong province," *Geographical Research*, vol. 40, no. 8, pp. 2204–2219, 2021, in Chinese.
[27] Q. Y. Liu, Q. M. Zhan, W. Liu, and C. Yang, "Research on the characteristics of urban network assocoation and spatial organization structure based on railway passenger flow in Hubei province," *Journal of Geo-Information Science*, vol. 22, no. 5, pp. 1008–1022, 2020, in Chinese.
[28] Y. Y. You, H. R. Yang, and J. E. Wang, "The structure evolution of China's urban networks from the perspective of HSR

flows," *World Regional Studies*, vol. 29, no. 4, pp. 773–780, 2020, in Chinese.
[29] X. Liu, J. Huang, J. Lai, J. Zhang, A. M. Senousi, and P. Zhao, "Analysis of urban agglomeration structure through spatial network and mobile phone data," *Transactions in GIS*, vol. 25, 2021.
[30] Y. B. Liu and M. Sun, "Comparison on the characteristics of city network connections between workdays and holidays based on information flow: a case study of the urban agglomeration in the middle reaches of the Yangtze River," *Economic Geography*, vol. 41, no. 5, pp. 75–84, 2021, in Chinese.
[31] F. Teixeira and B. Derudder, "SKYNET: an R package for generating air passenger networks for urban studies," *Urban Studies*, vol. 56, no. 14, pp. 3030–3044, 2019.
[32] H. Shu, T. Pei, C. Song, X. Chen, S. Guo, and Y. Liu, "L-function of geographical flows," *International Journal of Geographical Information Science*, vol. 35, no. 4, pp. 689–716, 2021.
[33] L. Weber, S. B. Mohn, F. Vecchio, and A. Fili, "Beyond deportation: researching the control of outward mobility using a space of flows logic," *Global Networks-A Journal of Transnational Affairs*, vol. 20, no. 1, pp. 65–84, 2020.
[34] C. A. Tang, J. W. Qiu, L. J. Zhang, and H. Y. Li, "Spatial econometric analysis on the influence of elements flow and industrial collaborative agglomeration on regional economic growth:based on manufacturing and producer services," *Economic Geography*, vol. 41, no. 7, pp. 146–154, 2021, in Chinese.
[35] F. Zhang, Y. M. Ning, and X. Y. Lou, "The evolutionary mechanism of China's urban network from 1997 to 2015: an analysis of air passenger flows," *Cities*, vol. 109, 2021.
[36] H. Hanssens, B. Derudder, S. van Aelst, and F. Witlox, "Assessing the functional polycentricity of the mega-city-region of central Bel-gium based on advanced producer service transaction Links," *Regional Studies*, vol. 4848, no. 12, pp. 1939–1953, 2013.
[37] S. J. Wang, S. Gao, and Y. Q. Wang, "Spatial structure of the urban agglomeration based on space of flows: the study of the Pearl River Delta," *Geographical Research*, vol. 38, no. 8, pp. 1849–1861, 2019, in Chinese.
[38] J. D. Nystune and M. F. Dacey, "A graph theory interpretation of nodal regions," *Papers in Regional Science*, vol. 7, no. 1, pp. 29–42, 1961.
[39] D. Dail and X. Liu, "Transport network backbone extraction: a comparison of techniques," *Journal of Transport Geography*, vol. 69, pp. 271–281, 2018.

Hindawi
Complexity
Volume 2021, Article ID 8716737, 18 pages
https://doi.org/10.1155/2021/8716737

Research Article

Competitive Strategy of Firms' Participation in the Global Value Chains and Labor Income Share

Zhaoji Sun,[1,2] Danling Tang,[2] and Qing Li[1]

[1]*Guangdong Institute for International Strategies, Guangdong University of Foreign Studies, Guangzhou 510420, China*
[2]*Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangdong Key Laboratory of Ocean Remote Sensing, State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 511458, China*

Correspondence should be addressed to Zhaoji Sun; szjszl1988@163.com

Received 24 September 2021; Revised 14 November 2021; Accepted 19 November 2021; Published 7 December 2021

Academic Editor: Zhipeng Tang

The division of labor in the global value chain (GVC) has reshaped the competitive advantage of enterprises participating in the international market and has a significant influence on the distribution of their factor income. Based on the perspective of market choice, this paper uses China's industrial enterprises' data, Customs Statistical Data on Import and Export, Word Input and Output Database (WIOD), and BACI database from 2000 to 2007 to analyze the effect of competitive strategy of a firm's GVC participation on its labor income share. Herein, the competition strategy is depicted by the quality and price of export goods. The empirical results show the following. (1) Quality competition has a larger effect on labor income share than price competition. High-quality and high-price or high-quality and low-price strategies tend to have a significant negative effect on labor income share, while low-quality and low-price and low-quality and high-price strategies have a positive effect. (2) The higher the target market GVC status, the lower the labor income share of exporting firms, and the target market GVC status amplifies that the high-quality and high-price strategies on firms' labor income share the negative effect of high-quality and low-price strategies on exporters' labor income share. (3) High-quality and low-price strategies have a significant negative effect on exporters' labor income share when competing with developing countries in developed country markets or with developed countries in developing country markets. However, the positive effect of low-quality and low-price strategies and low-quality and high-price strategies kept unchanged. The findings remain robust after controlling for endogeneity and accounting for the effects of firm heterogeneity, indicator measurement, and sample variation.

1. Introduction

Letting some people get rich first and eventually achieving common prosperity is an important path for China to improve people's living standards through reform and opening up. Achieving common prosperity includes two dimensions: the increase of overall prosperity and the sharing of development achievements. The continuous rapid growth of economic output over the past 40 years has significantly increased China's overall prosperity, while the sharing of development achievements needs further improvement. The report of the 19th National Congress of the Communist Party of China proposed that "socialism with Chinese characteristics has entered a new era, and the main contradiction in our society has been transformed into the contradiction between people's growing need for a better life and unbalanced and insufficient development." The unbalanced development is mainly reflected in the differences among people, regions, and urban and rural areas, and the differences among people have become an important aspect that hinders the sharing of development achievements. Since the mid-1990s, the share of labor income in China has been declining [1], and the continuous decline in the share of labor income is an important reason affecting the distribution of residents' income which has attracted widespread concern from all sectors of society. In response, the report of the 18th National Congress points out that "the share of labor compensation in the initial distribution should be increased, and the mechanism of initial distribution should be improved so that factors are distributed according to their contribution."

In an open economy, according to the Stolper-Samuelson theorem (SS Theorem), the share of labor income in China should increase with the expansion of international trade, but the opposite is true. The period of declining labor income share is the period when China actively participates in the global value chain (GVC) competition, which provides a stimulus for the research in this paper. Based on this, this paper investigates how the competitive strategies adopted by firms when facing different competitors in different markets affect the labor income share of firms? How does this impact relate to the GVC status of the target market? The research in this paper contributes to improving the initial factor income distribution to achieve common prosperity in an open economy and also provides empirical evidence from Chinese manufacturing industries to study the impact of competitive strategies on factor income distribution.

2. Literature Review

The decline in many countries' labor income share since the 1980s has driven scholars to rethink the changes in it. They have explained the variation from domestic perspectives of industrial structural transformation, policy and system reform, biased technical progress, and labor markets. Empirical research by Lu and Tian found that industrial concentration has a significant negative effect on the share of industrial labor income, and the main reason for the simultaneous decline in the labor income share of various industries is the division of enterprise scale within the industry [2]. The "VAT reform" has increased the labor income share of China's service industry export enterprises [3]; the minimum wage regulation has allowed enterprises to change the allocation of factors and reduce the labor income share of China's industrial enterprises [4]. The shareholding reform has significantly reduced the labor income share of listed companies in China, with a significant negative effect especially on state-owned labor-intensive enterprises [5]. The capital-labor elasticity of substitution in China is significantly greater than 1, and the decline in the relative price of labor leads to a decrease in labor income share [6].

Under open economy conditions, international trade affected labor income shares through the following channels such as endowment differences [7], labor-management bargaining power, and technological progress bias [8, 9]. According to neoclassical trade theory, developed countries have larger advantages in capital endowment than labor compared with developing countries. International trade has lowered the labor income share in developed countries, while it has raised that of developing countries. In fact, the decline in labor income share is a global phenomenon [10]. Decreuse and Maarek used the trade dataset for 8 OECD countries over the period 1970–2005 to conduct analysis and found that other countries' trade intensity with China explained between 30% and 60% of their labor share decline [11]. Elsby et al. also found that it was the extensive imports of Chinese products that drove the US labor share decline [12]. International trade has lowered the bargaining strength of the above countries' laborers and led to the decline in their labor share. Labor bargaining power is significantly positively correlated with its income share [13]. The stability of trade policy has a positive effect on labor bargaining power, so more stable trade policies can help increase the labor income shares [14].

From the existing studies, the relationship between exports and labor income share in China has not been consistently concluded. Applying the income distribution theory of the Post-Keynesian conflicting claims model, Tang used the provincial panel data over the period of 1995–2007 to conclude that exports have significantly improved China's labor share [15]. Using China's industrial enterprises' data from 1998 to 2007, Zhou et al. found a similar relationship for enterprises [16]. Using Chinese industrial enterprise data from 2004 to 2009, Wu and Shao find that export trade increases the labor income share of manufacturing firms, especially more significantly for labor-intensive SMEs. [17]. However, the empirical analysis from Zhang et al. has shown exports squeezing the increase of China's manufacturing enterprises' labor share [18]. Zhang et al. constructed an equation that described the decision of factor income distribution and used the cross-country data from 1980 to 2007 to conduct analysis [19]. Their conclusions showed that international trade led to a bias toward technical progress and thus lowered the labor income share. Findings from Xiao and Zhou showed that it was mainly due to the transformation of exports trade mode that China's labor income share has seen a trend of rise first and then fall during the period of 1993 and 2007 [20]. Peng and Zhang constructed the international division of labor income model under different trade types and found that Chinese processing trade and general trade income both showed a trend of first rising and then falling [21].

Besides exports, some scholars tried to explain the relationship between international trade and labor income share from the perspective of trade liberalization. Trade liberalization affected China's labor share via factor price and technical progress [22]. However, there have been great disparities in empirical conclusions. Böckerman and Maliranta conducted research based on microenterprises, and they held that the negative effect of trade liberalization on industry labor share emerges through existing plants, i.e., those with low productivity [23]. Trade liberalization has a positive effect on labor share within Chinese enterprises and intermediates trade liberalization lowered labor income share in Chinese enterprises [24, 25]. However, the relationship between trade liberalization and labor income share in India, which is also a developing country, is uncertain and is related to the size of the firm and whether it is labor-intensive [26]. With technical progress and a decrease in circulation costs, exports categories have shifted from final consumer goods to intermediate inputs. International specialization has transformed from intraproduct and horizontal interindustries specialization to that among factors and within value chains. The global value chain is formed therein. Multinational companies in developed countries have reconstructed their value chains based on their core competencies, and local Chinese enterprises mostly participate in global value chains by means of OEM and OEM production [27]. In this way, Chinese enterprises have seen rapid development, but they have gained slimmer margins in the GVC specialization system [28]. Sui et al.'s research found that the degree of embeddedness and upstream degree of the global value chain significantly affects the share of corporate labor income [29, 30]. The lower-end Embed method is reasonable in certain stages, while the high cost of labor factor cannot be ignored. The demographical dividend is disappearing, and other developing countries such as India and Vietnam are engaging in GVC competition. Under this condition, Chinese export enterprises need to seek changes in their competitive strategies in GVC participation [31].

Using Chinese customs statistics from 2000 to 2006, Li et al. find that Chinese exporters mainly adopt the "low-quality and low-price" competition strategy, especially private enterprises that newly enter the export market and enterprises in the eastern region are more inclined to compete at low prices [32]. Liu and Ding used the United Nations trade database and conducted a triple-dimension marginal decomposition and found that the driving forces of export increase have been shifted from quantitative types with low-quality and low-price to price driving forces with high-quality and high-price [33]. Based on analysis of BACI national data, Han and Yu held that since China joined WTO, it has seen a consistent decrease in the unit price of exports and improvement in exports quality. Enterprises tend to high-quality and low-price strategies [34]. However, compared with those of countries with high income, the quality of Chinese exports is still low, and export expansion is more reliable on price index decrease [35]. Fan and Guo inspected enterprises' pricing strategy using China's industrial enterprises' database and China customs database over the period 2000–2006. They found that when their export products are homogeneous, enterprises with high productivity have lower export prices, while the opposite is true under the heterogeneous framework [36]. An empirical study based on Greek export data found that reduced credit constraints are more likely to make companies choose high-quality and high-price competitive strategies [37].

Exporters' choice of competitive strategy varies depending on the global value chain in which the export market is located and the competitors. Different markets have a different extent of sensitivity to price and quality. In detail, countries lying in the lower-end in GVC have a low requirement for quality. In such countries, enterprises resort to price competition by factor price advantages. However, enterprises exporting to those countries lying in the higher-end in GVC tend to adopt a quality competition strategy [38]. Luo and Wei applied United Nations trade data over the period 2007–2012 and took the machinery and transportation equipment industry as an example to inspect the exporting price competition effect between China and medium- and high-income countries [39]. Findings showed that there exists a significant price competition relationship with different intensities in different markets. Exporting enterprises adjusted their competitive strategies facing different competitors. It is due to this that competition with high-wage countries led to improvement in Chinese enterprises' exporting goods [40]. However, competition within Chinese industries has driven enterprises into taking low-price strategies [41]. External shocks cause firms to produce heterogeneous responses of endogenous quality and pricing strategies [42]. Zhong and Yu studied the impact of external demand shocks and found that multiproduct enterprises that adopt quality competition strategies encounter external demand increases, the export price of core products will increase, while the reaction of multiproduct companies that adopt cost-competitive strategies is the opposite [43].

With the deepening specialization of global production, different countries have participated in the production processes, and intermediates have flowed across countries many times. Traditional trade statistics record trade flows on a gross basis, leading to a serious double-counting problem. The trade scale under the traditional trade statistic framework cannot truly reflect the benefit distribution, that is to say, what you see is no more than what you get [44]. Taking into account the difference in value-added capabilities between production processes, by constructing nonequidistant industry upstream indicators, it is found that the smiling curve is not universal, and FDI makes territorial export value added cannot truly reflect the factor income structure of foreign-funded enterprises [45, 46]. So trade scale cannot reflect the accurate effect of international trade on labor income share, and we have analyzed the effect of enterprises' GVC participation strategies on their labor income share. The marginal contributions in this paper are mainly in the following three areas. Firstly, under the background of economic globalization, current works of literature have analyzed how international trade affected labor income share from perspectives of import and export behaviors and trade liberalization [24, 26]. Among the rare researches into labor share based on GVC, existing papers mainly conducted macrolevel empirical analysis of GVC participation and GVC position [47, 48]. In this paper, we have made an analysis of the effect of exporting enterprises' competitive strategy choices in GVC on their labor income share from the view of market choice, hoping to provide us with a comprehensive understanding of how international trade affects labor income share.

Secondly, the existing literature tends to adopt the unit value method to measure product quality [49, 50]. This system usually assumes that high price reflects high quality, while low price reflects low quality without disentangling quality from price [32, 33]. The unit value method uses unit product price to directly reflect its quality and assumes that products' quality is totally reflected in their price, resulting in serious errors. Another method based on demand-side information to measure product quality has disentangled quality from products price and stated that on condition of price, imports with higher market shares are assigned higher quality. This method has considered consumers' preference for quality but has ignored the fact that enterprises themselves have a decisive role in the products' quality [51–53].

Base on this method, the latest method has utilized both demand-side and supply-side information to measure products' quality, which is more accurate [54, 55]. Li and Jiang identified national competitive strategies using relative price and commodities' quality [56]. Based on the researches of Feenstra and Romalis, and Yu and Zhang, we have adopted demand-side and supply-side information to infer the exporting goods' relative quality from the macronational level and microenterprise level and to construct the enterprises' relative quality compared with the world's average level. In this way, we used China Customs statistic data and the BACI database to measure the relative price at the national and enterprise levels and to construct the enterprises' relative prices. Applying the method of Li and Jiang, we identified enterprises' exporting strategies using their relative quality and relative price. This competitive strategy responds to the firm's position in the world average quality and price levels, not just the firm's position among all firms in the country, and can reflect both firm and country characteristics.

Thirdly, current researches usually focus on the GVC position of domestic countries [57, 58], ignoring that of the aim markets. In this paper, we used the improved GVC position index to measure the GVC position of the enterprise's aim market, combining the GVC position with the competitive strategy. GVC position index constructed by Koopman et al. did not consider the returned domestic value added (RDV) and did not eliminate pure double-counting of foreign value-added in a country's exports (FDC), either. However, FDC can only indicate GVC participation and is not very helpful in explaining the location of GVCs [59, 60], so this article will improve on this basis.

3. Competition Strategy and Measurement of Target Market's GVC Position

3.1. Identification of Competition Strategy

3.1.1. Measurement of Enterprises' Export Product Quality. Drawing on the method of Yu and Zhang, the supply-demand information extrapolation method is used to measure the quality of enterprise export products [55]. Due to large amounts of import intermediate inputs used in the processing trade, the quality of exporting products cannot be measured accurately. What's more, enterprises undertaking processing trade can only choose limited competition strategies. To make it feasible, we only conduct analysis of general trade situations. Structural parameters including exporters' FOB unit price, enterprises' productivity, and input cost are constructed to measure the quality of the exporting products. Exporters' FOB unit price uv_{ijgt} was calculated using China Customs statistical data over 2000–2007. uv_{ijgt} denotes the FOB unit price of product g produced by enterprise i exporting to country j.

$$uv_{ijgt} = \frac{\text{value}_{ijgt}}{\text{quantity}_{ijgt}}. \quad (1)$$

In equation (1), we define value_{ijgt} as the FOB value of enterprise i's product g exporting to country j in the year t. quantity_{ijgt} represents the exporting amount. We aggregated from the 8-digit HS to the 6-digit HS and uniformed HS 1996 and HS 2007 into HS 2002.

Enterprise's productivity φ_{it} was measured using the OP method based on Industrial Enterprise's Database over 2000–2007. According to the method of Brandt et al. [24], we calculated deflators and made unification of industry code consistent with the year 2002. Total factor productivity was estimated based on 4-digit CIC (Chinese Industry Classification).

Inputs include labor, capital, and intermediates. Total input costs were based on these three factors' costs, and the average industrial input cost w_t can be obtained by the following equation:

$$\ln(w_{it}) = \alpha' \ln\left(w_t^L\right) + \beta' \ln\left(w_t^K\right) + \gamma' \ln\left(w_t^M\right). \quad (2)$$

In the above equation, w_t^L, w_t^K, w_t^M and α', β', γ' represent the unit input cost of labor, capital, and intermediate, and their respective share. Since intermediates used to produce general exporting goods are mainly from domestic countries, the unit input cost equals 1 at the equilibrium. And we obtain the actual average industrial input cost w_t through equation

$$\ln(w_{it}) = \alpha' \ln\left(w_t^L\right) + \beta' \ln\left(w_t^K\right). \quad (3)$$

In this equation, $w_t^L = (\text{wage}_t + \text{compensation}_t)/\text{employee}_t$ and $w_t^K = \text{depreciation}_t/\text{capital}_t$. wage_t and compensation_t refer to the enterprise's total payable wages and total welfare funds, respectively. employee_t is the employment. Adding wages and welfare funds over 2-digit CIC and then deflating, we can get the labor cost. We denote depreciation_t and capital_t as enterprise's depreciating value of fixed capital and total fixed capital, respectively. In the C-D production equation, $\alpha' = \alpha/(\alpha + \beta + \gamma)$ and $\beta' = \beta/(\alpha + \beta + \gamma)$. While estimating the total factor productivity, we can get the output elasticity α, β, γ at the 2-digit CIC.

Adopting the structural parameters of 4-digit SITC Revision 2 of all countries in each year proposed by Feenstra and Romalis [54], we used the concordance file of SITC and 6-digit HS to create the structural parameter θ_g and κ_{jgt} of the 6-digit HS goods.

Matching all the above parameters, the quality $\ln(z_{ijgt})$ of certain enterprise's products exporting to destinations in the year can be obtained by

$$\ln\left(z_{ijgt}\right) = \theta_g\left[\ln\left(\kappa_{jgt}\right) + \ln\left(uv_{ijgt}\right) + \ln(\varphi_{it}) - \ln(w_t)\right]. \quad (4)$$

To avoid the influence of extreme values, the extreme values of 1% before and after are removed under HS6 codes. The obtained product quality $\ln(z_{ijgt})$ can be compared within one good across the country and across time. In order to get the product quality at the enterprise level, the product quality is standardized as follows.

$$\text{qual}_{ijgt} = \ln(z_{ijgt}) - \ln(z_{10\%-g}). \quad (5)$$

In equation (5), qual_{ijgt} is the normalized products quality index, and $\ln(z_{10\%-g})$ is the 10% quartile of the same 6-digit HS category product's quality. Based on the data proposed by Feenstra and Romalis [54], we utilized both demand-side and supply-side information to obtain the quality of 4-digit SITC exporting goods and then got the relative quality rq_{ijgt} of China's exporting enterprises based on the concordance file of SITC and 6-digit HS. Details are as follows:

$$rq_{ijgt} = \text{qual}_{ijgt} * \left(\frac{\text{cqual}_{jgt}}{\text{wqual}_{jgt}}\right), \quad (6)$$

$\text{cqual}_{jgt}/\text{wqual}_{jgt}$ is the ratio of the quality of product g exporting to country j from China in the year t and the average quality of product g exporting to country j from all over the world. The relative quality of enterprise i's product g exporting to country j can be obtained by multiplying the relative product quality of enterprise i compared to that of all domestic enterprises by that of China compared to the world. Adding by the exporting volume weight based on the relationship between exporting goods and exporting destinations, we can get the relative exporting quality of enterprise i in the year t and denoted as rq_{it}.

3.1.2. Measurement of Enterprise's Exporting Product Price. In light of the method from Li et al. and Li [32, 40], we measured the enterprises' relative exporting price rp_{ijgt} based on China Customs Statistic Database and BACI database. The specific calculation formula is as follows:

$$rp_{ijgt} = \left(\frac{uv_{ijgt}}{uv_{jgt}}\right) * \left(\frac{cuv_{jgt}}{wuv_{jgt}}\right), \quad (7)$$

where uv_{ijgt}/uv_{jgt} is the ratio of FOB unit price of enterprise i's product g exporting to country j in the year t and that of all enterprises' product g exporting to country j. We applied Customs Statistic Database over the period of 2000–2007 to calculate. We constructed cuv_{jgt}/wuv_{jgt} which denotes the ratio of the unit price of product g exporting to country j from China in the year t and that of product g exporting to country j from all over the world. BACI database of 2000–2007 is used to measure the ratio. By making unification of industry code into 2002 version of 6-digit HS, we can obtain the relative price of enterprise i's product g exporting to country j and that of all countries' product g exporting to country j. (Prices in the BACI database were obtained by deducting trading costs such as transporting cost and insurance and unifying differences of all countries' product quantity based on United Nations trade data. Besides, due to the large share of transshipment trade of Hong Kong and Singapore, we dropped samples of Hong Kong and Singapore when measuring enterprises' relative exporting prices.) Adding by the exporting volume weight based on the relationship between exporting goods and exporting destinations, we can get the relative exporting price of enterprise i in the year t and denoted as rp_{it}.

3.1.3. Exporting Competition Strategy. Using relative quality rq_{it} and relative price rp_{it}, we applied the method of Li and Jiang to construct enterprises' competition strategies [56]. An enterprise can choose a low-price strategy or a high-price strategy. We define the low-price competition strategy as an enterprise choosing the price level lower than the average price of all enterprises; otherwise, it is the high-price competition strategy. Regarding quality, an enterprise can also choose a low-quality competition strategy or a high-quality competition strategy. Similarly, a low-quality strategy comes when an enterprise's exporting quality is lower than the average level, otherwise forming the high-quality strategy. The competitive strategies of firms are defined as shown in Figure 1.

3.2. GVC Position of Target Markets. Koopman et al. have defined GVC position indexes [60]. Based on the decomposition framework proposed by Wang et al. [61], we have improved the reimport index to contain returned domestic value-added (RDV) part and dropped the pure double counting (FDC) of foreign value-added (FV). The modified index is stated as follows:

$$\text{GVC\_Position}_k = Ln\left(1 + \frac{\text{IV}_i + \text{RDV}_k}{E_k}\right) - Ln\left(1 + \frac{\text{FVA}_k}{E_k}\right). \quad (8)$$

In formula (8), IV represents the sum of domestic value added contained in the intermediates' export absorbed by direct importers and that contained in the intermediates produced by direct importers exporting to the third country and absorbed by the third one. FVA refers to the foreign value added contained in domestic exports and absorbed by other countries. E is the total exports, with the subscript k representing countries. The WIOD database is used to measure the GVC status of each country, and Figure 2 shows the GVC status of major countries and the comparison before and after modification.

From Figure 2, we can see that developed countries' GVC positions have been underestimated to a different extent, especially those of the United States and Germany. The decomposition results tell us that the sum of RDV and FDC accounts for 11.0%, 7.7%, and 4.7% of reimports in exports in the United States, Germany, and China, respectively, due to which that the GVC positions of the United States and Germany have been underestimated. For the USA, its underestimation is mainly attributed to the large share of RDV contained in the reimports of exports, while for Germany, FDC accounts for more. Major resources exporters like Russia and Brazil have high GVC positions without much change after modifying the index. The GVC position of India and Germany has shown the most obvious declining trend during this period, especially that of India since 2003. Considering the large share of China's exports to developed countries, our modification of the index will lead to a more accurate measurement of target markets' GVC position.

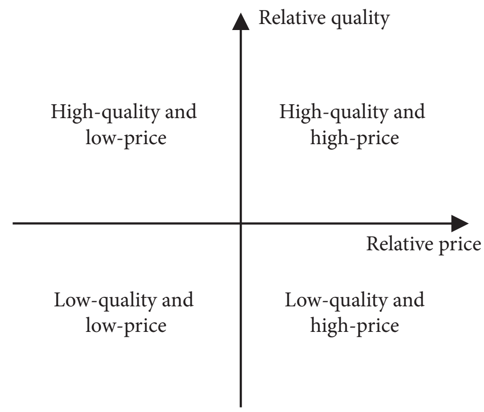


FIGURE 1: Enterprise's exporting competition strategy.

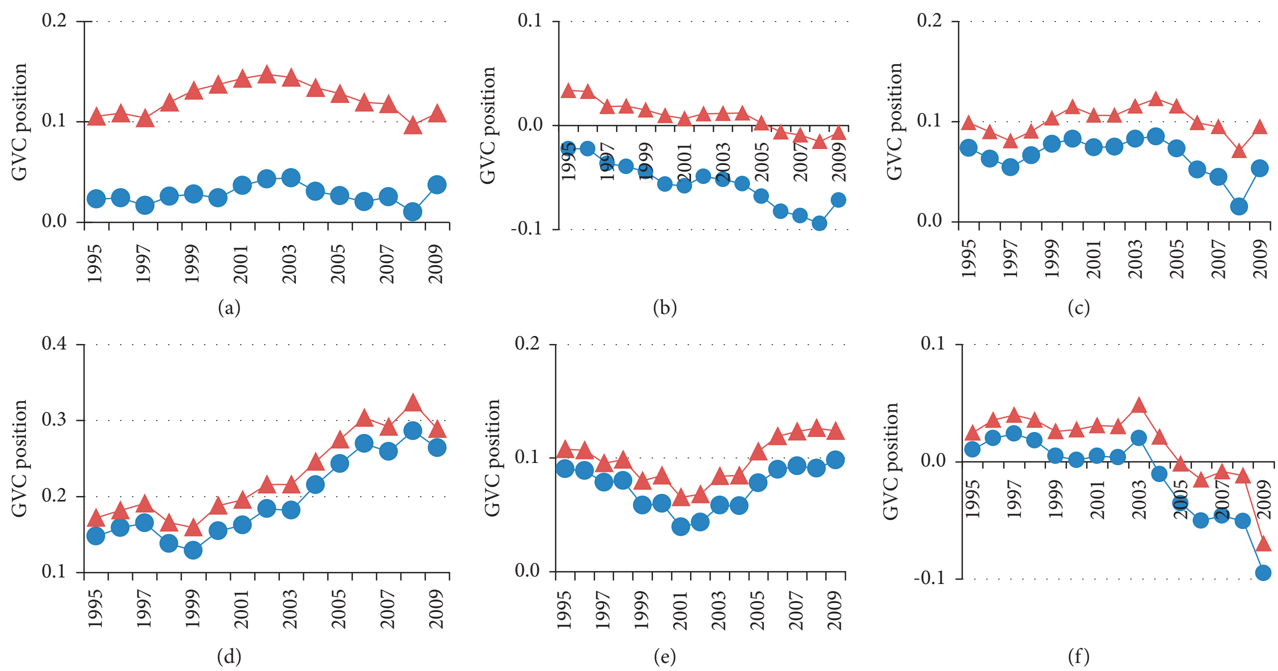


FIGURE 2: Major countries' GVC position over the period of 1995–2009. *Note.* The blue line is the GVC position before the improvement, and the red line is the GVC position after the improvement. (a) USA. (b) Germany. (c) Japan. (d) Russia. (e) Brazil. (f) India.

3.3. Statistical Analysis. We adopted the income method to measure enterprises' labor income share. Based on the data acquired, the work remuneration is measured by total payable wages and total payable welfare funds, and capital income refers to business surplus and depreciation. The formula of labor income share is described as follows:

$$\text{Labor Income Share} = \frac{\text{total payable wages} + \text{total payable welfare funds}}{\text{total payable wages} + \text{total payable welfare funds} + \text{business surplus} + \text{depreciation}}. \quad (9)$$

Table 1 depicts the competition strategy, target markets' GVC position, and labor income share of China's manufacturing exporting enterprises. The proportion of low-quality and low-price competitive strategies among the full sample of enterprises is 46%, indicating that low-quality and low-price strategy is the most important competitive strategy of Chinese manufacturing export enterprises. Seeing from the exporting destination, for the high-quality and high-price strategy adopters, developed countries are their preference compared with nondeveloped ones, while the opposite applies to the low-quality and low-price adopters. Enterprises exporting to developed countries

Table 1: Description of China's manufacturing exporting enterprises.

	Competition strategy				GVC position of target markets	Labor income share	The number of samples
	High-quality and high-price	Low-quality and low-price	High-quality and low-price	Low-quality and high-price			
Complete sample	0.21	0.46	0.21	0.12	0.030	0.45	185400
Exporting destination							
Developed countries	0.23	0.44	0.22	0.11	0.048	0.47	124814
Nondeveloped countries	0.17	0.51	0.21	0.11	-0.008	0.42	60586
Competitors							
Developed countries	0.24	0.42	0.25	0.09	0.019	0.41	91722
Developing countries	0.18	0.47	0.20	0.15	0.036	0.45	19607
Domestic enterprises	0.18	0.51	0.17	0.14	0.041	0.50	74071
Exporting products type							
Intermediates	0.21	0.46	0.25	0.08	0.020	0.40	91108
Capital goods	0.19	0.48	0.17	0.16	0.044	0.52	75339
Consumer goods	0.33	0.32	0.26	0.09	0.019	0.44	18864
Whether imports							
Imports	0.29	0.36	0.25	0.10	0.034	0.42	75499
Nonimports	0.16	0.53	0.19	0.12	0.027	0.48	109901
Ownership							
Foreign-funded enterprises	0.26	0.39	0.23	0.12	0.037	0.46	72121
Nonforeign-funded enterprises	0.18	0.50	0.20	0.12	0.025	0.45	113279
Productivity							
High productivity	0.30	0.24	0.43	0.03	0.027	0.35	61800
Medium productivity	0.22	0.50	0.17	0.11	0.030	0.47	61800
Low productivity	0.11	0.63	0.05	0.21	0.032	0.54	61800

Data resource: calculated based on the above method and data, the data of competition strategy and labor income share represents the respective share, and their classification is stated in the following part.

have higher target markets GVC position and enjoy higher labor income share. Such phenomenon can be naturally attributed to the fact that developed countries are more sensitive to quality while developing countries are more sensitive to price, and developed counties have higher GVC positions than developing ones. Competitor investigation shows that China's manufacturing exporting enterprises are more inclined to adopt a high-quality and high-price strategy facing competitors from developed countries compared with facing other competitors. And low-quality and low-price strategy appears more often in the competition among domestic enterprises. With regard to exporting products type, enterprises exporting consumer goods tend to choose high-quality and high-price strategy compared with those exporting intermediates and capital goods. Business scope also affects the enterprise's strategy. For enterprises that only export but not import, low-quality and low-price strategy is their preference, and they have a higher labor income share than those that both export and import. By firm ownership and productivity status, foreign-owned firms and high-productivity firms have a larger share of exporting using a high-quality, high-price competitive strategy than nonforeign-owned, low-productivity firms.

4. Theoretic Mechanism and Model Construction

4.1. Analysis of Theoretic Mechanism. Suppose that there are two factors of production labor (L) and capital (K), and the factor prices are the wage rate (w) and the interest rate (r), respectively. Without considering government tax subsidies, the labor income share (ls) is $\text{ls} = wL/(wL + rK)$, which is deformed to give

$$\text{ls} = \frac{1}{1 + [(K/L)/(w/r)]}. \tag{10}$$

From the above equation, it is clear that the labor income share depends on the factor use ratio (K/L) and the relative factor price (w/r).

The effect of the competitive strategy of the firm's participation in the target market GVC on the labor income share depends on the change in the factor use ratio and the relative factor price of the exported product. Specifically, there are two paths of action as follows: first, the factor intensity effect. Assuming that there is quality heterogeneity in the products exported by firms in country A to country B of the target market, the factor use ratio (K/L) required for high-quality products is usually greater than that for low-quality products, and the adoption of quality competition strategy increases the capital-labor ratio of firms and thus reduces the labor income share. Assume that there are differences in consumer preferences in target market B. Target markets with high GVC status pay more attention to product quality, while target markets with low GVC status are more sensitive to the product price. Export product quality affects the factor use ratio of firms, while export product prices affect the relative factor prices through the price transmission mechanism. For target markets with different GVC statuses, there are differences in the extent to which firms focus on the quality and price of export products. Therefore, the impact of firms' competitive strategy choice on labor income share is related to the target market GVC status. Second, factor substitution effect, competitive strategy choice, and adjustment change the proportion of factors used by firms in country A, and the relative prices of factors change accordingly. According to Bai and Qian [62], the rise in relative factor prices (w/r) prompts firms to use capital instead of labor, which in turn leads to an increase in the factor use ratio (K/L). At this point, judging the direction of change in the labor income share requires comparing which grows faster, the factor use ratio or the relative price. When the factor substitution elasticity is greater than one, the factor use ratio grows faster than the relative price of factors and the labor income share decreases; when the factor substitution elasticity is less than one, the labor income share increases.

According to Antoniades' endogenous decision model of a firm's export quality [63], the quality of a firm's export products in country A is related to the degree of competition in the target market. As shown in Figure 3, firms choose the optimal export product quality and marginal cost (inversely proportional to productivity), and firms with higher productivity have higher export product quality. When the degree of competition in the target market increases, the quality ladder shifts to the left, and the product quality varies more between firms. Firms with higher productivity respond to market competition by increasing product quality and price (region D1), firms with lower productivity respond by decreasing product quality and price (region D2), and firms with the lowest productivity exit the market (region D3). Firms adopt different competitive strategies (quality competition and price competition) depending on the target market GVC status, which has an impact on the labor income share. When the degree of competition in the target market changes, the higher the target market GVC status, the more likely the firm will adjust to a quality competition strategy; the lower the target market GVC status, the more likely the firm will adjust to a price competition strategy.

4.2. Econometric Model and Variables. Based on the above theoretic analysis, we built the following basic regression model to inspect the effect of the competition strategy an enterprise adopts for GVC participation on its labor income share.

$$ls_{it} = \beta_0 + \beta_1 cs_{it} + \beta_2 gvc_{it} + \beta_3 x_{it} + \varepsilon_{it}. \tag{11}$$

In this model, ls_{it} denotes labor income share, cs_{it} is the competition strategy, gvc_{it} represents the GVC position of enterprises' exporting target market, and x_{it} refers to a series of control variables including financing constraints fc_{it}, export intensity es_{it}, the size of enterprise size_{it}, and the age of enterprises age_{it}. ε_{it} is the error term with subscript i and t being enterprise and the year, respectively.

Being the explained variable, exporting enterprises' labor income share (ls_{it}) can be measured as the above. Since labor income share is valued between 0 and 1 (the labor income share of enterprises in loss may be larger than 1 or less than 0; we did not take these enterprises into consideration), such limitation will bring bias to the statistical results. To avoid bias, we borrowed the method of Li et al. and Wei et al. to conduct Logist transformation to transform labor income share ls_{it} into $ls_{it}/(1 - ls_{it})$ and took the natural logarithm of it [1, 64].

The core explanatory variables include enterprises' competition strategy (cs_{it}) and the target market's GVC position (gvc_{it}). Competition strategy can be divided into high-quality and high-price strategy $cs0_{it}$, low-quality and low-price strategy $cs1_{it}$, high-quality and high-price strategy cs2_{it}, and low-quality and high-price strategy $cs3_{it}$ based on the relative quality rq_{it} and relative price rp_{it} of enterprises' exporting goods. The target market's GVC position is measured in the method said before.

We set four control variables. Among them, financing constraints (fc_{it}) facing enterprises are measured by debt-to-assets ratio, details of which can be searched in Luo and Chen [65], export intensity (es_{it}) is measured by enterprises' export volume-to-sales ratio, enterprise size (size_{it}) is depicted by its total assets, and enterprise's age (age_{it}) is calculated by subtracting its establishing year from the current year and then adding 1.

For the division of export destinations of enterprises, according to the number of enterprise exports, the proportion of exports to developed countries exceeds 50% is defined as mainly exporting to developed countries, otherwise mainly exporting to nondeveloped countries (according to the classification rule of United Nations Developed Program, Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, Germany, France, Greece, Hungary, Iceland, Ireland, Italy, Japan, South Korea, Luxemburg, Netherland, New Zealand, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, the United States, Great Britain, Andorra, Bahrain, Barbados, Brunei, Cyprus, Estonia, Hong Kong, Israel, Liechtenstein, Malta, Monaco, Qatar, Slovenia, Singapore, and United Arab Emirates belong to developed countries and regions, others belonging to developing countries and regions). Applying the BACI database, the respective share of certain products exporting to one country from developed countries and developing countries except China and China can be obtained. Based on the China Customs database, we can calculate the share of the product exporting to the country from different enterprises. Matching products with its exporting destination, we can get the share of the product exporting to the country from enterprises of developed countries, developing countries except China, and other Chinese

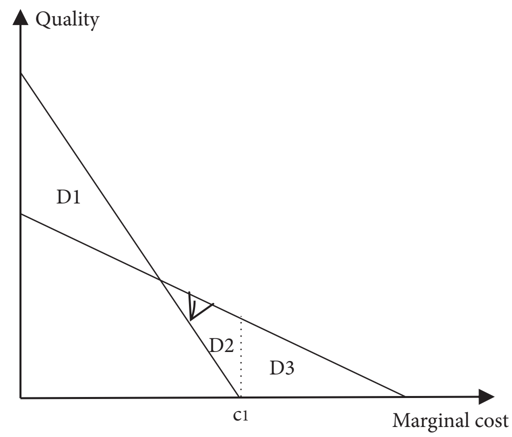


Figure 3: Competitive strategy and degree of competition in the target market (source: [63]).

enterprises. According to BEC, we classified enterprises' exporting products into intermediates, capital goods, and consumer goods, dropping the few other categories. The enterprise location is divided into the eastern region, central region, and western region. Provinces and cities including Beijing, Tianjin, Hebei province, Liaoning province, Shanghai, Jiangsu province, Zhejiang province, Fujian province, Shandong province, and Guangdong province belong to the eastern region. The central region covers Shanxi province, Jilin province, Heilongjiang province, Anhui province, Jiangxi province, Henan province, Hubei province, and Hunan province, other provinces being the western region.

Data in this paper are collected from China Industrial Enterprise Database over the period of 2000–2007, China Customs Statistic Database, BACI database, WIOD Database over 1995–2011, and data supplied by Brandt et al. and Feenstra and Romalis [24, 54]. The industry concordance of HS code in each year, SITC industrial classification, and BEC product category were borrowed from the United Nations Database, and the RMB exchange rate against the US dollar is searched from the Chinese Statistical Yearbook. We constructed a harmonized classification that groups China's Industrial Enterprises Database and China Customs Statistical Database. In length, we used the name of enterprises to match these two databases and then matched the postal code of the enterprise's location with the last seven-digit number of the enterprise's telephone number. Next, we matched the simplified enterprise's name without address and ownership with the postal code. Finally, we retain enterprises successfully matched with at least one of the above methods as our samples. These data were processed using Feenstra et al. methodology [66]. We excluded observations showing missing, zero, or negative values of total assets, net fixed assets, total wage bills, industrial sales, and the number of employees and further dropped all firms with less than 8 employees. We treat observations with more net fixed assets than total assets or without enterprise code similarly and those with no or less than zero of paid-in capital. We also dropped the largest and smallest 1 percent of observations.

5. Empirical Analysis

The panel dataset of China's exporting enterprises conducting general trade over 2000–2007 is applied to inspect the effect of enterprises' GVC participation competition strategy on its labor income share. The whole process is conducted as follows. First, a two-way fixed effects model is used for estimation and the basic results are analyzed to draw basic conclusions. Second, in order to exclude possible sample selection bias, endogeneity problems, etc., indicator substitution, subsampling, and different estimation methods are used for robustness tests, respectively; finally, interaction terms are added to the regressions to further test the effects of interactions between different variables on labor income shares.

5.1. Benchmark Estimates. We first conduct OLS regression for all samples. Results show the maximum variance inflation factor is 1.77, far below the threshold value of 10, excluding severe multicollinearity among variables. To see the influence of exporting enterprises' competition strategy on its labor income share, we used the relative exporting price and relative quality as explaining variables to make regression. Seeing from column (1) and column (2) in Table 2, the elasticity coefficient of relative exporting price is 0.021, which is significant at the 1 percent level, showing the positive effect of exporting price on labor income share. With other conditions being the same, higher exporting price incurring larger labor share may be due to the fact that labor factor price is more sensitive to exporting product price than capital factor price did. Statistical results show that about half of Chinese enterprises adopt low-quality and low-price strategy, and the export is more reliable on labor than on other factors, to some extent that explains why labor income share is greatly affected. The elasticity coefficient of relative exporting quality is -0.023 significant at the 1 percent level, presenting that relative exporting quality significantly lowered labor income share. This may be attributed to the positive relationship between export quality and capital intensity which was found in Fan and Guo [36]. With other conditions being the same, enterprises with higher exporting quality tend to have a lower labor share. Comparing the estimate coefficients, we can see that the quality competition showed larger effect on labor income share than price competition did. To make deeper exploitation, we check the effect of four bundles of competition strategies on labor income share one by one and in conjunction in the following part of this paper. (By setting the variable of certain competition strategy as 1 and others as 0, we check the effect of this strategy. To avoid complete collinearity, we set high-quality and high-price strategy as contract term when taking all competition strategies into the regression model.)

Column (3) and column (5) in Table 2 show the significant negative effect of high-quality and high-price strategy as well as high-quality and low-price strategy on labor income share, while results in columns (4) and (6) tell the significant positive effect of low-quality and low-price strategy and low-quality and high-price strategy on the labor share. Both verified the major role quality strategy played in labor income share, which is

Table 2: Benchmark estimate results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rp	0.021***	0.021***					
	(0.0026)	(0.0026)					
Rq	−0.230***	−0.217***					
	(0.0048)	(0.0048)					
*cs*0			−0.096***				
			(0.0078)				
*cs*1				0.138***			0.214***
				(0.0066)			(0.0093)
*cs*2					−0.153***		−0.026***
					(0.0070)		(0.0087)
*cs*3						0.114***	0.241***
						(0.0085)	(0.0106)
gvc		−0.166***	−0.197***	−0.190***	−0.192***	−0.197***	−0.162***
		(0.0570)	(0.0574)	(0.0574)	(0.0574)	(0.0574)	(0.0572)
fc		0.617***	0.634***	0.632***	0.633***	0.634***	0.629***
		(0.0176)	(0.0178)	(0.0177)	(0.0177)	(0.0178)	(0.0177)
es		0.111***	0.143***	0.136***	0.139***	0.145***	0.126***
		(0.0138)	(0.0139)	(0.0139)	(0.0139)	(0.0139)	(0.0139)
Size		−0.207***	−0.245***	−0.245***	−0.244***	−0.244***	−0.239***
		(0.0064)	(0.0064)	(0.0064)	(0.0064)	(0.0064)	(0.0064)
Age		0.015***	0.012***	0.012***	0.012***	0.012***	0.012***
		(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Observations	185400	185400	185400	185400	185400	185400	185400
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.020	0.045	0.029	0.031	0.032	0.029	0.038

***, **, and *respectively, indicate significance at 1%, 5%, and 10%. The values in the bracket indicate the standard error, the same below.

consistent with the estimates of respective checks for price competition and quality competition. China's exporting enterprises mainly adopt the low-quality and low-price strategy, leading to a relatively large labor income share. However, with the increase in export quality, labor income share goes down. From column (7), compared with the high-quality and high-price strategy, low-quality and low-price strategy and low-quality and high-price strategy incurred higher labor share, while the high-quality and low-price strategy tends to lower the labor income share of exporting enterprises. Concluded from this, with export quality being the same, lower exporting price results in lower labor income share, while with the same exporting price, lower quality leads to higher labor share. Considering the both, quality competition plays the major role.

The estimated coefficient of the target market's GVC position is negative, indicating that exporting to markets with higher GVC position tend to lower the enterprises' labor income share, consistent with our expectation. Countries with high GVC positions are more possible to conduct capital-intensive production while relocating labor-intensive procedures to other countries such as China. However, such procedures are more likely to be labor-intensive for Chinese producers. What's more, the choice of target market may be related to enterprises' competition strategy, which will be further illustrated in the following sections.

For the estimation of other control variables, the estimated coefficients of financing constraints are significantly positive, indicating that exporting firms with greater financing constraints have higher labor income shares. Facing tighter financing, enterprises choose labor as a substitution of capital, thus increasing the labor income share. Our estimates are different from the analysis conducted for all Chinese enterprises by Luo and Chen [65] since exporting enterprises can play the labor advantage by participating in GVC production. The coefficient of enterprises' exporting intensity is also significantly positive, showing that being more inclined to export, the enterprise enjoys a larger labor income share, which is similar to the conclusion of Zhou et al. [16]. Enterprise's size presents a negative effect on labor income share. Enterprises with larger sizes have a smaller share of labor income. Finally, the elder enterprises tend to have a larger labor income share since the estimate of the enterprise's age is positive.

5.2. Robustness Check. We have modified some index measurements in the benchmark estimate, which may bring bias to the results. Therefore, this paper employs different alternative indicators and uses different estimation methods for robustness testing. In detail, we take government tax subsidies into consideration when measuring labor income share. The methodology of Koopman et al. is applied to measure the target market's GVC position [60]. We use the median value instead of the average as a reference when defining competition strategy. As for the methods, we adopt Tobit estimation in replace of Logist transformation method and take balanced panel samples excluding the effect of the entry into and exit from the export market on enterprise's labor income share instead of unbalanced ones. Table 3

Table 3: The estimate of regressions using substitute variables and methods.

	Substitute variables			Substitute methods	
	Labor income share	GVC position	Competition strategy	Tobit	Balanced panel
*cs*1	0.241*** (0.0084)	0.222*** (0.0093)	0.209*** (0.0088)	0.207*** (0.0076)	0.226*** (0.0289)
*cs*2	−0.026*** (0.0078)	−0.025*** (0.0087)	−0.030*** (0.0083)	−0.207*** (0.0090)	0.018 (0.0254)
*cs*3	0.272*** (0.0096)	0.255*** (0.0106)	0.242*** (0.0089)	0.381*** (0.0104)	0.242*** (0.0352)
gvc	−0.133*** (0.0517)	−0.163*** (0.0627)	−0.160*** (0.0572)	−0.135*** (0.0383)	−0.261 (0.2058)
fc	0.504*** (0.0160)	0.568*** (0.0177)	0.630*** (0.0177)	0.993*** (0.0118)	0.675*** (0.0641)
es	0.288*** (0.0125)	0.135*** (0.0140)	0.126*** (0.0139)	0.283*** (0.0101)	−0.081* (0.0432)
Size	−0.162*** (0.0058)	−0.218*** (0.0071)	−0.238*** (0.0064)	−0.203*** (0.0024)	−0.131*** (0.0202)
Age	0.012*** (0.0007)	0.006*** (0.0007)	0.012*** (0.0007)	0.014*** (0.0004)	0.012*** (0.0019)
Observations	185400	185400	185,400	185400	8632
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
R^2	0.041	0.025	0.038	—	0.059

shows the same estimate results after taking substitute variables and almost the same results using the Tobit method and balanced panel samples. We can say that the estimate of the effect of enterprises' GVC participation strategy on labor income share is somewhat robust. (The estimated results of balanced panel samples did not show a significant negative effect of the target market's GVC position on labor income share. This may be due to the fact that enterprises have made effective integration of production procedures taking place in different target markets.)

5.3. The Effect of Enterprises' Heterogeneity. The above estimates show the effect from enterprises of the average level. Disparities in factors including target market choice, competitors, exporting product types, ownership, and differences in factor endowments in enterprises' locations may also affect the influence of enterprises' GVC participation strategy on its labor income share. To take these factors into consideration, we need to make a classification of samples to exploit the effect of the enterprise's heterogeneity on the benchmark results.

5.3.1. Heterogeneity in Target Market and Competitors. The estimation results are presented in Table 4, which divides the subsample into six categories of firms according to their target markets and competitors, the main target markets for GVC participation into developed countries and developing countries, and the main competitors into developed, developing, and domestic Chinese firms. All the samples are classified into six subsamples. Results show that the coefficient of low-quality and low-price strategy is positive for all target markets and competitors, indicating the low-quality and low-price strategy adopted by Chinese enterprises in whatever target markets and facing any competitors tend to have a positive effect on their labor income share. This is true in reality. Chinese enterprises usually take advantage of their low labor costs. Using a lot of low-skilled labor will increase the labor income share. There exist significant differences in the effect of the high-quality and low-price strategy on labor income share when enterprises choose different target markets and face variant competitors. When competing with developing countries in developed countries or competing with developed countries in developing countries, the high-quality and low-price strategy tends to have a significant negative effect on exporting enterprises' labor income share, while in other cases, the estimated coefficient is not significant. This may be due to the fact that developed countries are more focused on product quality. When competing with enterprises from developed countries in their hometown, Chinese enterprises will not choose the quality strategy. While facing competitors from developing countries, to some extent we have a quality advantage, and the high-quality and low-price strategy is beneficial for Chinese enterprises' GVC participation. Based on this, choosing a high-quality and low-price strategy when competing with developing countries' competitors in developed countries will lower Chinese enterprises' labor income share. Developing countries are more sensitive to product prices than developed countries. Chinese enterprises have no price advantage competing with competitors from developing countries and from home, so they are not choosing the high-quality and low-price strategy. But in competition with developed countries, the high-quality and low-price strategy is a better choice. In light of this, choosing a high-quality and low-price strategy when competing with developed countries in developing countries significantly lowers Chinese enterprises' labor income share. The low-quality and high-price strategy showed a positive effect on labor income share except when competing with developing competitors in developed countries. The estimation results for the other control variables are broadly consistent with the basic estimation results.

5.3.2. Heterogeneity in Export Product Structures and in Total Factor Productivity. Since export product structure and total factor productivity do affect the labor income share to some extent [67], we have classified samples based on these two influencing factors. From the results shown in Table 5, we can see that the estimate for the low-quality and low-price strategy as well as low-quality and high-price strategy is the same for all subsamples, showing a positive effect on labor income share. However, the high-quality and low-price strategy is playing different roles in subsamples with different product

Table 4: The estimate of subsamples classified by target market and competitors.

Target market	Developed countries			Developing countries		
Competitors	Enterprises from developed countries	Enterprises from developing countries	Chinese enterprises	Enterprises from developed countries	Enterprises from developed countries	Chinese enterprises
*cs*1	0.238*** (0.0169)	0.191*** (0.0506)	0.240*** (0.0207)	0.170*** (0.0240)	0.158* (0.0827)	0.201*** (0.0430)
*cs*2	−0.010 (0.0146)	−0.133*** (0.0497)	−0.006 (0.0210)	−0.046** (0.0204)	−0.040 (0.0823)	−0.014 (0.0464)
*cs*3	0.239*** (0.0208)	0.181*** (0.0522)	0.272*** (0.0218)	0.251*** (0.0298)	0.130 (0.0942)	0.211*** (0.0450)
fc	0.710*** (0.0342)	0.722*** (0.0991)	0.551*** (0.0356)	0.705*** (0.0503)	0.193 (0.1244)	0.531*** (0.0658)
es	0.062** (0.0269)	0.021 (0.0698)	0.216*** (0.0274)	0.120*** (0.0427)	−0.007 (0.1496)	0.116** (0.0492)
Size	−0.229*** (0.0120)	−0.216*** (0.0384)	−0.256*** (0.0144)	−0.261*** (0.0168)	−0.119** (0.0582)	−0.274*** (0.0238)
Age	0.014*** (0.0014)	0.023*** (0.0044)	0.018*** (0.0018)	0.008*** (0.0018)	−0.005 (0.0052)	0.005** (0.0026)
Observations	58562	12372	53880	33160	7235	20191
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.041	0.043	0.040	0.041	0.020	0.039

Table 5: The estimated results of regressions on subsamples classified by product category and total factor productivity.

	Export product category			Total factor productivity		
	Intermediates	Consumer goods	Capital goods	Low	Medium	High
*cs*1	0.214*** (0.0131)	0.234*** (0.0162)	0.185*** (0.0293)	0.160*** (0.0237)	0.105*** (0.0175)	0.090*** (0.0183)
*cs*2	−0.031** (0.0120)	−0.020 (0.0167)	−0.000 (0.0222)	0.021 (0.0305)	−0.015 (0.0168)	−0.028** (0.0128)
*cs*3	0.248*** (0.0161)	0.249*** (0.0170)	0.232*** (0.0344)	0.171*** (0.0233)	0.141*** (0.0204)	0.083*** (0.0322)
gvc	−0.082 (0.0797)	−0.327*** (0.1018)	−0.136 (0.1911)	−0.240** (0.1212)	−0.235** (0.1114)	−0.196* (0.1053)
fc	0.679*** (0.0263)	0.552*** (0.0275)	0.707*** (0.0593)	0.556*** (0.0388)	0.506*** (0.0354)	0.704*** (0.0327)
es	0.077*** (0.0214)	0.170*** (0.0208)	0.075 (0.0483)	0.063** (0.0272)	0.065** (0.0265)	0.118*** (0.0294)
Size	−0.234*** (0.0092)	−0.243*** (0.0108)	−0.206*** (0.0195)	−0.209*** (0.0157)	−0.108*** (0.0135)	−0.195*** (0.0115)
Age	0.011*** (0.0010)	0.014*** (0.0014)	0.011*** (0.0020)	0.014*** (0.0016)	0.015*** (0.0013)	0.016*** (0.0014)
Observations	91108	75339	18864	61800	61800	61800
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.040	0.042	0.033	0.020	0.028	0.035

categories and different productivities. By export product category, the high-quality and low-price strategy has a significant negative effect on the labor income share of enterprises mainly exporting intermediates, the same as benchmark results. But the negative effect of such strategy is not significant for enterprises' major conduct consumer goods and capital goods export, similar to the process of trade liberalization improving the quality of imported intermediates [68]. Trade liberalization has intensified the competition of intermediates export market, and enterprises tend to choose a high-quality and low-price strategy rather than a high-quality and high-price strategy to participate in international competition. Facing international competition of consumer goods, enterprises usually have some market powers and prefer the high-quality and high-price strategy. With developing countries as its target market of capital goods export, China is more likely to adopt te low-quality and low-price strategy. (Results in Table 1 showed that most enterprises exporting consumer goods adopt the high-quality and high-price strategy, while those exporting capital goods prefer the low-quality and low-price strategy.) The target market's GVC position has a significant positive effect on labor income share for enterprises that mainly export consumer goods and has an insignificant positive effect for those that mainly export capital goods and intermediates. This may be attributed to the trade pattern that consumer goods exporting to developed countries are of higher quality, and developing countries are the target markets for capital goods export. Seeing from productivity, the high-quality and low-price strategy has a significant negative effect on labor income share of enterprises with high productivity while it has an insignificant effect for enterprises with medium and low productivity. Productivity is positively correlated with product quality [36], and low and medium productivity firms have lower product quality and may not be in a position to adopt a high-quality and low-price strategy.

Table 6: The estimated results of regressions on subsamples classified by ownership and regions.

	Ownership			Regions		
	State-owned/collective-owned enterprises	Private enterprises	Foreign-funded/Hong Kong, Macau, and Taiwan enterprises	Eastern region	Central region	Western region
*cs*1	0.149*** (0.0392)	0.176*** (0.0129)	0.256*** (0.0150)	0.215*** (0.0096)	0.211*** (0.0444)	0.209*** (0.0526)
*cs*2	−0.031 (0.0364)	−0.028** (0.0124)	−0.014 (0.0136)	−0.026*** (0.0090)	0.034 (0.0413)	−0.106** (0.0471)
*cs*3	0.153*** (0.0454)	0.191*** (0.0145)	0.296*** (0.0173)	0.241*** (0.0109)	0.324*** (0.0518)	0.147** (0.0627)
gvc	−0.136 (0.2330)	−0.074 (0.0782)	−0.227** (0.0946)	−0.166*** (0.0597)	−0.352 (0.2532)	0.341 (0.3076)
fc	0.249*** (0.0791)	0.608*** (0.0242)	0.742*** (0.0296)	0.630*** (0.0185)	0.676*** (0.0771)	0.522*** (0.0967)
es	0.077 (0.0665)	0.144*** (0.0183)	0.092*** (0.0236)	0.130*** (0.0144)	0.059 (0.0661)	0.071 (0.0870)
Size	−0.249*** (0.0345)	−0.245*** (0.0082)	−0.298*** (0.0123)	−0.230*** (0.0067)	−0.305*** (0.0289)	−0.406*** (0.0385)
Age	−0.002 (0.0018)	0.004*** (0.0009)	0.055*** (0.0019)	0.014*** (0.0008)	−0.003 (0.0024)	0.007** (0.0029)
Observations	13828	99451	72121	166211	12390	6799
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.019	0.046	0.048	0.039	0.040	0.049

5.3.3. Heterogeneity of Ownership, Region, and Industry. For enterprises of different ownership and in different regions, the low-quality and low-price strategy and low-quality and high-price strategy have shown the same positive effect on their labor income share, as shown in Table 6. The high-quality and low-price strategy plays a significant negative effect on private enterprises and those in eastern regions while it has an insignificant effect on other subsamples, which has been verified in Li et al. [32]. They have found that private enterprises and eastern region enterprises are more likely to adopt the high-quality and low-price strategy in international competition and pass on the losses from high-quality and low-price competitive strategies to labor. The estimated coefficients of GVC status in the target market show that the higher the GVC status of foreign firms and eastern region firms exporting to countries with higher GVC status, the lower their labor income share, and foreign firms or eastern region firms exporting mainly to developed countries have relatively higher capital-labor ratios of their export products.

Table 7 shows the estimates of regressions on subsamples classified by industries. Results are almost the same as the benchmark regressions for most industries. However, for industries of beverage manufacturing, wood processing industry, printing industry, petroleum processing, and chemical fiber manufacturing industry, the effects of competition strategy on labor income share are not significant.

5.3.4. Endogeneity Problem. Considering the possible omission of explanatory variables affecting labor income share in this paper, and the possible mutual causality between firms' competitive strategies for participating in GVC and labor income share, this can lead to the endogeneity problem. In this paper, a two-way fixed effects model is used in the empirical test to control the endogeneity problem to some extent. Since no effective instrumental variables are found, this paper uses the difference method to examine the dynamic relationship between variables. The difference method can eliminate a part of the endogeneity problem caused by omitted variables that do not change over time and alleviate the endogeneity problem to some extent, thus weakening the influence of endogeneity on the conclusions of this paper. Firstly, relative quality and relative price are regressed after making the difference, and the results are shown in (1) and (2) of Table 8. Then, the differential of competitive strategy is defined according to the differential of relative price and relative quality of firms' exports, and if both relative price and relative quality of exports increase compared with the previous period, it is defined as the differential of high quality and high price; i.e., the differential of high quality and high price indicates that both relative quality and relative price increase compared with the previous period, and the differential of other competitive strategies is redefined similarly. The estimation results of the difference method for competitive strategies are shown in (3)–(7) of Table 8. The signs and significance of the estimated coefficients of relative price and relative quality of firms' exports are consistent with the basic estimation results, and the effects of the competitive strategies constructed using relative price and relative quality on labor income shares are similar to the basic estimation results. The estimated coefficients of the target market GVC position are not significant, indicating that exporters choose to participate in the target market GVC mainly by choosing the horizontal position of GVC position rather than the increment of GVC position. The effect of firms' export intensity on labor income share is also not significantly related to its increment, and the estimated results of other variables are consistent with the basic estimation results.

Table 7: The estimated results of regressions on subsamples classified by industries.

Industry code and name	Low-quality and low-price	High-quality and low-price	Low-quality and high-price	Observations
13 Farm and sideline food processing industry	0.270***	−0.245***	0.154***	8539
14 Food industry	0.164***	−0.196***	0.189**	3741
15 Beverage manufacturing	0.126	−0.109	−0.065	1151
17 Textile industry	0.103***	−0.170***	0.112***	20622
18 Textile clothing, shoes, and hat manufacturing	0.085***	−0.207***	0.071***	16356
19 Leather, fur, and feather manufacturing	0.095***	−0.142***	0.048*	8194
20 Wood processing and wood, bamboo, rattan, palm grass products industry	0.029	−0.091*	0.071	3617
21 Furniture manufacturing	0.205***	−0.255***	0.134***	3745
22 Paper making and paper products industry	0.117*	−0.145**	0.481***	1861
23 Printing and recording media replication	−0.033	0.015	0.077	1257
24 Cultural education and sports goods	0.113***	−0.131***	0.123***	6019
25 Petroleum processing, coking and nuclear fuel processing	0.038	−0.065	0.234	728
26 Chemical raw materials and chemical products manufacturing	0.210***	−0.205***	0.162***	10952
27 Pharmaceutical industry	0.179***	−0.073**	0.141**	3043
28 Chemical fiber manufacturing	0.093	0.015	0.090	670
29 Rubber products	0.109**	−0.069	0.109*	2565
30 Plastic products	0.123***	−0.153***	0.154***	7891
31 Nonmetal mineral products	0.174***	−0.141***	0.135**	9103
32 Black metal smelting and rolling processing industry	0.137**	−0.173**	0.165*	1710
33 Nonferrous metal smelting and rolling processing industry	0.090*	−0.257***	0.207**	1983
34 Metal products	0.125***	−0.164***	0.132***	10923
35 Ordinary machinery manufacturing	0.180***	−0.128***	0.105***	14329
36 Special equipment manufacturing	0.142***	−0.100***	0.134***	7082
37 Transportation equipment manufacturing	0.169***	−0.200***	0.155***	8119
39 Electrical machinery and equipment manufacturing	0.125***	−0.134***	0.113***	12289
40 Communication equipment, computer, and other electronic equipment manufacturing	0.097***	−0.080***	0.145***	8956
41 Instrument and culture, office machinery manufacturing	0.094**	−0.089**	0.166**	3704
42 Handicrafts and other manufacturing	0.147***	−0.177***	0.177***	6238

Table 8: Estimation results of the difference method.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δrp	0.013*** (0.0035)	0.014*** (0.0034)					
Δrq	−0.177*** (0.0070)	−0.180*** (0.0069)					
$\Delta cs0$			−0.116*** (0.0086)				
$\Delta cs1$				0.145*** (0.0093)			0.182*** (0.0106)
$\Delta cs2$					−0.133*** (0.0094)		−0.021* (0.0109)
$\Delta cs3$						0.159*** (0.0107)	0.203*** (0.0121)
Δgvc		0.079 (0.0835)	0.060 (0.0838)	0.068 (0.0838)	0.072 (0.0838)	0.065 (0.0838)	0.080 (0.0835)
Δfc		0.589*** (0.0260)	0.587*** (0.0261)	0.589*** (0.0261)	0.589*** (0.0261)	0.586*** (0.0261)	0.584*** (0.0260)
Δes		0.015 (0.0195)	0.028 (0.0196)	0.027 (0.0196)	0.031 (0.0196)	0.030 (0.0196)	0.019 (0.0195)
ΔSize		−0.286*** (0.0131)	−0.284*** (0.0131)	−0.284*** (0.0131)	−0.284*** (0.0131)	−0.284*** (0.0131)	−0.287*** (0.0131)
Observations	97265	97265	97265	97265	97265	97265	97265
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.011	0.028	0.020	0.021	0.020	0.021	0.027

Table 9: The estimate results of regressions introducing interact terms.

	(1)	(2)	(3)	(4)	(5)	(6)
*cs*0	−0.106*** (0.0083)				−0.098*** (0.0078)	
*cs*0 × *gvc*	−0.285*** (0.0923)					
*cs*1		0.140*** (0.0071)				0.110*** (0.0089)
*cs*1 × *gvc*		0.056 (0.0677)				
*cs*2			−0.157*** (0.0076)			
*cs*2 × *gvc*			−0.097 (0.0855)			
*cs*3				0.111*** (0.0094)		
*cs*3 × *gvc*				−0.088 (0.1075)		
*cs*0 × *icr*					0.298* (0.1597)	
*cs*1 × *ent*						0.065*** (0.0099)
gvc					−0.196*** (0.0574)	−0.251*** (0.0796)
fc	0.633*** (0.0178)	0.632*** (0.0177)	0.633*** (0.0177)	0.634*** (0.0178)	0.634*** (0.0178)	0.672*** (0.0237)
es	0.142*** (0.0139)	0.135*** (0.0139)	0.139*** (0.0139)	0.145*** (0.0139)	0.143*** (0.0139)	0.150*** (0.0192)
Size	−0.245*** (0.0064)	−0.245*** (0.0064)	−0.244*** (0.0064)	−0.244*** (0.0064)	−0.245*** (0.0064)	−0.251*** (0.0091)
Age	0.012*** (0.0007)	0.012*** (0.0007)	0.012*** (0.0007)	0.012*** (0.0007)	0.012*** (0.0007)	0.007*** (0.0010)
Observations	185400	185400	185400	185400	185400	103251
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.029	0.031	0.032	0.029	0.029	0.034

5.3.5. Further Discussions. In order to test the possible association between exporters' participation in GVC in choosing their target market and their competitive strategy, as well as the effect of the interaction between other variables on labor income share, this paper adds an interaction term to further test. Results are depicted in Table 9. Column 1 has presented the results introducing interactions between the target market's GVC position and high-quality and high-price strategy. The estimate coefficients are all negative, showing the interactions have significantly affected the influence on labor income share. With the increase in the target market's GVC position, the negative effect of the high-quality and high-price strategy on labor income share is enlarged. This may be due to the fact that countries with higher GVC positions paid more attention to product quality and enterprises exporting to these countries preferred the high-quality and high-price strategy, increasing the capital intensity of products while forcing down the labor income share [36]. This is consistent with the conclusions of Flach that enterprises exporting to high-income countries tend to adopt the high-quality and high-price strategy [69]. Columns 2 to 4 showed the estimated results after introducing interactions between low-quality and low-price strategy, high-quality and low-price strategy, and low-quality and high-price strategy and target market's GVC position; results did not show a significant effect on labor income share. Fan et al. found that enterprises exporting high-quality products to countries with higher GVC positions are more likely to import high-quality intermediates [70]. Due to this, we introduced interactions between import intermediates contained in enterprises' export and high-quality and high-price strategy, and estimated coefficients in column 5 are significantly positive, showing that importing intermediates in export production eased the negative effect of the high-quality and high-price strategy on labor income share. This can be explained by the fact that the high quality of exporting goods can be attributed to importing high-quality intermediates and thus lowered the capital intensity of the production procedure. However, we did not find interactions existing between import intermediates in enterprises' export with other competitive strategies. Li concluded that newly entered enterprises are more likely to adopt a price competition strategy [32]. Based on this, we check the effect of interactions between the entry of enterprises and the low-quality and low-price strategy. (We borrowed the three-year judgment rule proposed by Li et al. to define the entry and exit of enterprises [33].) Significant positive estimates in column 6 told us that enterprises' new entry into export markets has enlarged the positive effect of the low-quality and low-price strategy on labor income share.

6. Conclusions

Based on the existing literature, this paper portrays the competitive strategies of firms' participation in GVC and examines the impact of firms' competitive strategies of participation in GVC on labor income share from the perspective of target market selection using data of Chinese manufacturing firms from 2000 to 2007. Conclusions are as follows.

First, with export product quality being the same, with higher exporting prices, enterprises enjoy higher labor income share. With the same exporting price, exporting products of higher quality will lower the enterprise's labor income share. Compared with price competition, quality competition showed a larger effect on labor income share.

Second, both high-quality and high-price competition strategies have significant negative effects on labor income share, while both low-quality and low-price competition strategies have significant positive effects on labor income share. The positive effect of the low-quality and low-price strategy and low-quality and high-price strategy does not vary among heterogeneous enterprises, while labor income share will receive various effects from the high-quality and low-price strategy due to enterprises' heterogeneity. Only when competing with developing countries in developed markets or competing with developed countries in developing markets, the high-quality and high-price strategy will play a significant positive role in affecting labor income share. The high-quality and low-price strategy has laid a positive-negative effect on eastern private enterprises' labor income share.

Third, the higher the GVC status of the target market when firms participate in GVC, the lower the labor income share of exporting firms. But introducing import intermediates will ease the negative effect of the high-quality and high-price strategy. The new market entry by firms amplifies the positive effect of the low-quality, low-price strategy on labor income share.

The 19th National Congress of the Communist Party of China has proposed that we must adhere to the first place of quality and efficiency and keep in line with supply-side reform and push forward the revolution of quality, efficiency, and driving force in economic development to improve total factor productivity. Although in a short period, the low-quality and low-price strategy did ease the decrease in China's labor income share to some extent in GVC production, with the aging of the population, the export competitiveness by virtue of cheap labor will not last. Based on improving products quality and enterprises' benefit, we aim to apply quality competition instead of traditional price competition to shoot more trade benefits in GVC and to assure a certain share of laborers' wage income in the primary distribution of national income. In addition, Chinese exporters should take "One Belt, One Road" as an opportunity to reshape their competitive strategy of participating in global value chains and build a regional value chain division of labor system led by themselves.

Data Availability

The data presented in this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by Project of Philosophy and Social Science Foundation of Guangdong (No. GD20YYJ05), Foundation of Natural Science of Guangdong (No. 2019BT02H594), major Project of National Social Science Foundation of China (No. 21ZDA097), and Project of Humanities and Social Science Research Youth Foundation of Ministry of Education (No. 18YJC630050).

References

[1] D. Li, L. Liu, and H. Wang, "The U curve of labor share in GDP during economic development," *Economic Research Journal*, no. 1, pp. 70–82, 2009.

[2] X. Lu and L. Tian, "Firms' size differentiation and labor income share," *The Journal of World Economy*, no. 9, pp. 27–48, 2020.

[3] Y. Su, C. Chen, and J. Lan, "VAT Reform and labor income share: evidence from Chinese listed companies," *Finance and Trade Economics*, no. 3, pp. 44–61, 2021.

[4] J. Wan and X. Wei, "The impact of minimum wage regulation on the labor income share of enterprises: theoretical analysis and micro-evidence," *Journal of Finance and Economics*, no. 7, pp. 64–78, 2020.

[5] X. Shi, W. Gao, Y. Lu, and M. Li, "Capital market allocation efficiency and labor income share: evidence from the share-trading reform," *Economic Research Journal*, no. 12, pp. 21–37, 2019.

[6] D. Chen and S. Chen, "The relative price of capital labor, the elasticity of substitution and labor income share," *The Journal of World Economy*, no. 12, pp. 73–97, 2018.

[7] W. F. Stolper and P. A. Samuelson, "Protection and real wages," *The Review of Economic Studies*, vol. 9, no. 1, pp. 58–73, 1941.

[8] A. E. Harrison, *Has Globalization Eroded Labor's Share? Some Cross-Country Evidence*, University Library of Munich, Munich, Germany, Article ID 39649, 2005.

[9] D. Acemoglu, "Changes in unemployment and wage inequality: an alternative theory and some evidence," *The American Economic Review*, vol. 89, no. 5, pp. 1259–1278, 1999.

[10] L. Karabarbounis and B. Neiman, "The global decline of the labor share," *Quarterly Journal of Economics*, vol. 129, no. 1, pp. 61–103, 2014.

[11] B. Decreuse and P. Maarek, "Can the HOS model explain changes in labor shares? A tale of trade and wage rigidities," *Economic Systems*, vol. 41, no. 4, pp. 472–491, 2017.

[12] M. W. L. Elsby, B. Hobijn, and A. Şahin, "The decline of the U.S. labor share," *Brookings Papers on Economic Activity*, no. 2, pp. 1–63, 2013.

[13] P. Bai and Z. Yang, "The bargaining power of labor and the proportion of labor income: an analysis of the impact of the financial crisis," *Management World*, no. 5, pp. 78–91, 2019.

[14] S. Xie, Y. Lu, and X. Cai, "The wage bargaining power of workers in an open economic system," *Social Sciences in China*, no. 5, pp. 40–59, 2019.

[15] D. Tang, "Globalization and labor income share: an analysis based on the bargaining strength between labor and capital," *Management World*, no. 8, pp. 23–33, 2011.

[16] Z. Zhou, H. Quan, and T. Chen, "Export, effect of deep labor division and labor income share," *Journal of International Trade*, no. 2, pp. 16–26, 2017.

[17] X. Wu and J. Shao, "Research on the heterogeneity impact of export participation on labor income share of manufacturing enterprises," *Journal of International Trade*, no. 1, pp. 14–27, 2019.

[18] J. Zhang, Z. Chen, and X. Zhou, "Research on the inhibitory effects of exporting on labor's share," *The Journal of Quantitative & Technical Economics*, no. 7, pp. 44–60, 2012.

[19] L. Zhang, J. Li, and X. Xu, "International trade, biased technological progress and factor income distribution," *China Economic Quarterly*, vol. 11, pp. 410–428, 2012.

[20] W. Xiao and M. Zhou, "Change of trade mode and decrease of labor share: empirical evidence from China's industrial sector," *Journal of ZheJiang University (Humanities and Social Sciences)*, vol. 40, no. 5, pp. 154–163, 2010.

[21] Z. Peng and B. Zhang, "The evolution of China's international division of labor income and the decomposition of its

determinants," *China Industrial Economics*, no. 6, pp. 62–80, 2018.
[22] X. Huang, S. Xu, and J. Lu, "Trade liberalisation and labour income share variation: an interpretation of China's deviation from the ss," *The World Economy*, vol. 34, no. 7, pp. 1071–1087, 2011.
[23] P. Böckerman and M. Maliranta, "Globalization, creative destruction, and labor share change: evidence on the determinants and mechanisms from longitudinal plant-level data," *Oxford Economic Papers*, vol. 64, pp. 259–280, 2012.
[24] L. Brandt, J. Van Biesebroeck, and Y. Zhang, "Creative accounting or creative destruction? Firm-Level productivity growth in Chinese manufacturing," *Journal of Development Economics*, vol. 97, no. 2, pp. 339–351, 2012.
[25] M. Yu and Z. Liang, "Trade liberalization and China's labor income share: empirical analysis based on data of manufacturing trading enterprises," *Management World*, no. 6, pp. 54–63, 2014.
[26] R. N. Ahsan and D. Mitra, "Trade liberalization and labor's slice of the pie: evidence from Indian firms," *Journal of Development Economics*, vol. 108, pp. 1–16, 2014.
[27] Z. Liu and J. Zhang, "Forming breakthrough and strategies of captive network in developing countries at global outsourcing system: based on a comparative survey of GVC and NVC," *China Industrial Economics*, no. 5, pp. 39–47, 2007.
[28] G. Gereffi and J. Lee, "Why the world suddenly cares about global supply chains," *Journal of Supply Chain Management*, vol. 48, no. 3, pp. 24–32, 2012.
[29] G. Sui, Z. Sun, and W. Chen, "Global value chain embeddedness and labor income share: based on China's theory and empirical analysis," *Journal of International Trade*, no. 2, pp. 96–112, 2021.
[30] G. Sui, Z. Sun, and W. Chen, "Upstream degree of global value chain, FDI and factor income distribution of manufacturing enterprises," *International Economics and Trade Research*, no. 11, pp. 52–66, 2020.
[31] P. Jin, "Holding to the solid foundation of real economic development," *Qiushi*, no. 7, pp. 24–26, 2012.
[32] K. Li, W. Jiang, and L. Song, "The mystery of changes in the quality of China's exports: a micro-level explanation based on market entry," *Social Sciences in China*, no. 3, pp. 80–103, 2014.
[33] Y. Liu and Y. Ding, "Has China realized quality-driven export growth in ICT industry: an empirical study based on three margins and gravity equation," *China Industrial Economics*, no. 1, pp. 52–64, 2015.
[34] H. Han and C. Yu, "Price competition or quality competition: an empirical study on the competitiveness of China's export products," *Industrial Economic Review*, no. 6, pp. 127–135, 2014.
[35] X. Li, J. Zhou, X. Lu, and J. Hu, "Does the export product quality affect export volume," *Economic Research Journal*, no. 8, pp. 114–129, 2015.
[36] H. Fan and G. Guo, "The relationship between exporting price, exporting quality and productivity: evidence from China," *The Journal of World Economy*, no. 2, pp. 58–85, 2015.
[37] E. Dinopoulos, S. Kalyvitis, and M. Katsimi, "Variable export price elasticity, product quality, and credit constraints: theory and evidence from Greek firms," *Journal of International Economics*, vol. 104, Article ID 102135, 2020.
[38] X. Dai and P. Jin, "Intra-product specialization, institution quality and export sophistication," *Economic Research Journal*, no. 7, pp. 4–17, 2014.
[39] C. Luo and N. Wei, "The price effect of Chian's export: evidence from products of machinery and transport equipment," *Finance & Trade Economics*, no. 8, pp. 104–118, 2015.
[40] F. Li, "Export competition and export quality of Chinese enterprises," *Journal of Dalian University of Technolgy*, vol. 36, no. 3, pp. 86–92, 2015.
[41] D. Sheng and Y. Wang, "The mystery of the low-price exports of China's enterprises: an analysis from the perspective of enterprise's markup ratio," *Management World*, no. 5, pp. 8–23, 2012.
[42] H. Fan, Y. Li, S. Xu, and S. R. Yeaple, "Quality, variable markups, and welfare: a quantitative general equilibrium analysis of export prices," *Journal of International Economics*, vol. 125, Article ID 103327, 2020.
[43] T. Zhong and M. Yu, "External demand, competitive strategy and export behavior of multi-product enterprises," *China Industrial Economics*, no. 10, pp. 119–137, 2020.
[44] A. Maurer and C. Degain, "Globalization and trade flows: what you see is not what you get," *Journal of International Commerce, Economics and Policy (JICEP)*, vol. 3, pp. 1–27, 2012.
[45] X. Gao, J. Huang, and K. Yuan, "Value chain embedded position and export domestic value added rate," *The Journal of Quantitative &Technical Economics*, no. 6, pp. 41–61, 2019.
[46] Z. Zhou and K. Zhu, "The factor ownership structure of foreign-funded enterprises and the attribution of export value-added income," *China Industrial Economics*, no. 1, pp. 118–135, 2020.
[47] S. Zhang, "Does global value chain lead to decrease of labor income share: empirical analysis of China's industrial panel data," *Economic Perspectives*, no. 10, pp. 39–48, 2015.
[48] S. Liu, N. Gu, and X. Chen, "Global value incorporation, factor endowments structure and the proportion of labor income: empirical study based on international data," *The Economist*, no. 3, pp. 96–104, 2016.
[49] P. Bastos and J. Silva, "The quality of a firm's exports: where you export to matters," *Journal of International Economics*, no. 2, pp. 99–111, 2010.
[50] K. Manova and Z. Zhang, "Export prices across firms and destinations," *Quarterly Journal of Economics*, vol. 127, no. 1, pp. 379–436, 2012.
[51] A. K. Khandelwal, P. K. Schott, and S.-J. Wei, "Trade liberalization and embedded institutional reform: evidence from Chinese exporters," *The American Economic Review*, vol. 103, no. 6, pp. 2169–2195, 2013.
[52] B. Shi and W. Shao, "The measurement and decisive factors of Chinese enterprises' exporting product quality," *Management World*, no. 9, pp. 90–106, 2014.
[53] J. Xu, J. Tong, and Q. Mao, "RMB exchange rate, product quality and firms'export activity: an empirical analysis based on Chinese manufacturing enterprises," *Journal of Financial Research*, no. 3, pp. 1–17, 2015.
[54] R. C. Feenstra and J. Romalis, "International prices and endogenous quality," *Quarterly Journal of Economics*, vol. 129, no. 2, pp. 477–527, 2014.
[55] M. Yu and R. Zhang, "Estimating China's manufacturing export quality: pitfalls and remedy," *China Economic Quarterly*, vol. 16, pp. 463–484, 2017.
[56] K. Li and W. Jiang, *How Chinese Manufacturers Compete in Global Markets: Price Competition VS Quality Competition*, Working Paper, 2016.
[57] S. Zhou, Z. Lan, and H. Fu, "Division status of China's manufacturing industry in global value chains: a study based

on koopman's GVC position indices," *Journal of International Trade*, no. 2, pp. 3–12, 2014.
[58] S. Ma, H. Zhang, and X. Wang, "Financing constraints and moving up in the global value chain: theory and evidence from Chinese processing trade enterprises," *Social Sciences in China*, no. 1, pp. 83–107, 2017.
[59] W. Chen, P. Zhao, and J. Fang, "Chains' position, trade revenue and competitiveness in global value chains: an analysis based on trade in value added accounting framework," *International Business Research*, no. 4, pp. 5–18, 2017.
[60] R. Koopman, W. Powers, Z. Wang, and S. Wei, *Give Credit where Credit Is Due: Tracing Value Added in Global Production Chains*, NBER Working Paper, Cambridge, MA, USA, Article ID 16426, 2010.
[61] Z. Wang, S. Wei, and K. Zhu, *Quantifying International Production Sharing at the Bilateral and Sector Levels*, NBER Working Paper, Cambridge, MA, USA, Article ID 19677, 2013.
[62] C. Bai and Z. Qian, "Factor income share in chain: the story behind the statistics," *Economic Research Journal*, no. 3, pp. 27–41, 2009.
[63] A. Antoniades, "Heterogeneous firms, quality, and trade," *Journal of International Economics*, vol. 95, no. 2, pp. 263–273, 2015.
[64] X. Wei, Z. Dong, and Y. Liu, "Political relationship, system environment and labor income share: a study based on the investigation data of China's private enterprises," *Management World*, no. 5, pp. 35–46, 2013.
[65] C. Luo and L. Chen, "Does financial constraints lead to decline in labor income shares: an empirical study based on China's enterprise data supplied by world bank," *Journal of Financial Research*, no. 3, pp. 29–42, 2012.
[66] R. C. Feenstra, Z. Li, and M. Yu, "Exports and credit constraints under incomplete information: theory and evidence from China," *The Review of Economics and Statistics*, vol. 96, no. 4, pp. 729–744, 2014.
[67] X. Zhang and Y. Lu, "Effect of merchandise trade structure changes on labor income share," *The Journal of Quantitative & Technical Economics*, no. 1, pp. 59–76, 2014.
[68] M. Yu and L. Li, "Trade liberalization and quality upgrading of intermediate inputs: evidence from Chinese customs data," *China Economic Quarterly*, vol. 15, pp. 1011–1028, 2016.
[69] L. Flach, "Quality upgrading and price heterogeneity: evidence from Brazilian exporters," *Journal of International Economics*, vol. 102, pp. 282–290, 2016.
[70] H. Fan, Y. A. Li, and S. R. Yeaple, "Trade liberalization, quality, and export prices," *The Review of Economics and Statistics*, vol. 97, no. 5, pp. 1033–1051, 2015.

Hindawi
Complexity
Volume 2021, Article ID 4580917, 12 pages
https://doi.org/10.1155/2021/4580917

Research Article

Study on the Temporal and Spatial Differentiation of Provincial Tourism Efficiency in Eastern China and Influencing Factors

Huadi Wang[1] **Shaogui Xu,**[2] **Qiuyi Xie,**[3] **Jianying Fan,**[4] **and Nianxing Zhou**[3]

[1]*School of Education Science, Nanjing Normal University, Nanjing, Jiangsu 210046, China*
[2]*School of Management, Jinan University, Guangzhou, Guangdong 510630, China*
[3]*School of Geography, Nanjing Normal University, Nanjing, Jiangsu 210023, China*
[4]*School of Tourism and Culture, Nanning Normal University, Nanning, Guangxi 530100, China*

Correspondence should be addressed to Nianxing Zhou; zhounianxing@263.net

Received 22 June 2021; Revised 10 August 2021; Accepted 9 September 2021; Published 4 October 2021

Academic Editor: Jun Li

Tourism efficiency can be used to effectively measure the utilization of regional tourism resources and the state of tourism economic development. Based on the super efficiency DEA model, Malmquist index, and spatial econometric model, this article measures the static and dynamic tourism efficiency of 11 provinces and cities in eastern China for 2010 to 2019. In combining ArcGIS 10.0 and MATLAB 2016b software, this article studies the temporal and spatial differentiation of tourism efficiency in eastern provinces and cities and influencing factors. The results show that (1) the overall tourism efficiency of eastern provinces is at a high level and relatively stable, but the regional distribution is quite varied, and problems of spatial imbalance are prominent; (2) the overall tourism efficiency of eastern provinces is increasing, and the change index of technical efficiency contributes the most, followed by the change index of scale efficiency; and (3) industrial status, traffic conditions, tourism resource endowment, and the labour force are the main factors affecting the temporal and spatial differentiation of tourism efficiency in eastern provinces and cities, while the level of economic development and information technology have no significant impact.

1. Introduction

At present, China's economy is transitioning from a stage of rapid growth to a stage of high-quality development. Tourism plays an important role in optimizing the economic structure, promoting the transformation of growth drivers, and meeting people's spiritual and cultural needs to meet their desire for a better life. It is also crucial for China under the current new normal to improve tourism efficiency and take the path of sustainable development. As an effective index of resource utilization capability and effect based on input-output [1], efficiency can effectively measure the comprehensive impact of tourism on a region to make a corresponding evaluation of the tourism development of different provinces and cities. It is of great significance to promote the development of regional tourism and increase tourism economic income to systematically evaluate the tourism efficiency of various provinces and cities through time and space.

Domestic and foreign research on tourism efficiency has been carried out from many dimensions, and the research results are relatively comprehensive. In terms of the object of research, researchers have examined the tourism efficiency of different levels of tourism destinations, different types of tourism activities, and tourism departments with different functions [2–5]. In terms of research content, researchers have conducted beneficial explorations of the efficiency of tourism poverty alleviation [6], tourism resource conversion [7], and tourism ecological development [8, 9]. In addition to conducting research on the efficiency of tourism hotels, tourism transportation, tourism destinations, and travel agencies from a micro perspective [10], domestic scholars have also carried out research at the macro and meso levels in examining urban agglomerations, provinces, cities, and counties as well as at the micro level in studying scenic areas and forest parks [11–16]. The scope, depth, methodologies, and intersectionality of research on tourism efficiency are also expanding [10]. Despite this work, research on the eastern China economic belt, a leading area of tourism economic development, is still relatively limited. Taking 11 provinces and cities in eastern China as its basic research unit, this study constructs an index system of tourism efficiency measurement using data related to tourism and the economy for 2010 to 2019. The static super efficiency DEA model and dynamic Malmquist index are used to measure the level and development trends of tourism efficiency in China's eastern provinces and cities. Meanwhile, with the help of a spatial econometric model, the authors analyse the temporal and spatial differentiation characteristics and influencing factors of regional tourism efficiency from the perspective of space and visualization and regulate and predict differences in tourism efficiency. This work is expected to promote the high-quality development of the tourism economy in China's eastern region and guide the tourism economic development of China's central and western provinces, which started later and lags behind that of China's eastern provinces, to facilitate the development of China's tourism economy.

2. Study Area and Data

2.1. Study Area. To formulate long-term national economic development plans, land development and renovation plans, and regional layout and regional economic development policies and services, China is divided into three major economic zones: the eastern, central, and western regions. The eastern economic zone is located in the eastern region of Asia, along the eastern border of the Eurasian continent and along the west coast of the Pacific Ocean, covering a total area of approximately 950,000 square kilometres. The region includes three cities, Beijing, Tianjin, and Shanghai, and eight provinces, Hebei, Liaoning, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan, with a total of 11 provincial administrative units. Due to the influence of its geographic location, development foundations, market conditions, and human capital, the region has formed the most developed socialist economy in China. All of the major metropolises of mainland China, such as Beijing, Shanghai, Shenzhen, and Guangzhou, are located in eastern China. Therefore, it is appropriate and significant to use the DEA-Malmquist index combined with the exploratory spatial data analysis method to conduct an in-depth study of tourism efficiency in this area.

2.2. Data Source and Index Selection

2.2.1. Data Source. To ensure the study's authenticity, integrity, and continuity, all data used are from the China Statistical Yearbook (2010–2019) and China Tourism Statistical Yearbook (2010–2019) as well as from the statistical yearbooks of eastern provinces and cities and the Statistical Bulletin of National Economic and Social Development (2010–2019). Due to missing data for some years for some provinces and cities, interpolation is adopted to supplement the data to ensure the rationality of calculations and meet the processing requirements of this study.

2.2.2. Index Selection

(1) Setting of Input-Output Elements of Tourism Efficiency. The index of tourism efficiency evaluation often adopts the production input factors of economics, which are generally defined as land and natural resources, labour, capital, and so on [17]. Among these, the impact of land on tourism development is relatively limited, and the number of starred hotels and A-level scenic areas receiving capital investment also reflects the scale of land investment to a certain extent, so it is not taken as an input variable [18]. The labour force injects vitality into regional tourism development, representing an important source of potential for its development. However, due to the lack of data on the ideal input index of tourism practitioners, on the basis of increasing the number of employees, the selection of employees in the tertiary industry as labour input can be regarded as a means of index selection in line with the characteristics of the tourism industry. Capital factors mainly include various tourism service facilities and enterprises that promote the development of the tourism industry. We use the number of A-level scenic areas, star hotels, and travel agencies that reflect the value of tourism resource endowment and the reception and service capacity of the tourism industry as input variables of capital [19]. Tourism output factors mainly include economic effects, social effects, and ecological effects. To reduce the impacts of other external environmental variables, such as travel convenience, travel preferences, and city reputation, we use domestic tourism income and the number of domestic tourists as output indicators to reflect economic and social effects. Based on the change in the statistical standard of the total area of afforestation, the area of artificial afforestation for a given year is used as the index of ecological effect output (Table 1).

(2) Setting of Influencing Factors of Tourism Efficiency. As tourism efficiency is affected by macro- and micro-level aspects, the following six factors are used based on the existing literature [20–22]. ① The economic development level (ECON): the level of economic development directly affects tourism project cooperation between provinces and urban areas and then affects the development of tourism efficiency. Improving infrastructure, introducing advanced technology, and innovating tourism products require a certain degree of capital investment. Here, the variable is expressed by the per capita GDP of provinces and cities. ② Industrial status (STAT): the status of the tourism industry can reflect the government's attention to regional tourism development to a certain extent and reflect the agglomeration capacities of regional tourism factors. The stronger this capacity is, the broader the flow of regional tourism production factors becomes. We use the proportion of domestic

tourism income of GDP to express the state of the tourism industry. ③ Tourism resource endowment (RESO): tourism resource endowment affects the flow and allocation of tourism investment elements between provinces and cities as well as tourism flow and tourism investment, thus affecting the division of labour and cooperation in the tourism industry between provinces and cities. The evaluation rules of the national standard classification and of the quality of scenic areas (GB/T17775-2003) are adopted. The minimum score for scenic areas specified by Rule 2 is selected as the score for the resource endowment evaluation of scenic areas at corresponding levels. The resource endowment evaluation value of the decision-making unit is obtained by summing the scores for all scenic areas above grade 3A of the unit [23]. ④ Traffic conditions (TRAN): tourism resources in the eastern provinces are distributed unevenly across space, and traffic conditions greatly affect the development of tourism. As highways represent the most important form of provincial and municipal transportation, the highway mileage (km) of each province and city is used to represent traffic conditions. ⑤ The labour force (HR): the tourism industry is a labour-intensive industry. Labour indirectly affects the development efficiency and performance of the tourism economy. The number of employees in the tertiary industry is used to represent the labour force. ⑥ The information development level (INFO): the informatization development level affects the spatial agglomeration and expansion of the tourism economy and can be expressed by the total level of post and telecommunications business. However, due to the lack of statistical calibre and data for certain years, the total amount of postal business is used to express this value.

3. Methods

3.1. Super Efficiency DEA Model. The DEA (Data Envelopment Analysis) model was first proposed by Charnes, Cooper, and Rhodes in 1978. DEA is a new field that integrates operational research, mathematical economics, and management science [24] and that uses mathematical programming models, including linear programming, multiobjective programming, generalized optimization with a tapered structure, semi-infinite programming, and random programming, to evaluate the relative effectiveness of multiple inputs and/or output "departments" or "units" (DMU) [25]. DEA is a commonly used means to measure economic efficiency and can effectively measure tourism efficiency. However, due to the particularities of the tourism industry, the corresponding input variables are relatively controllable, while the output variables are not. We use the input-oriented model of the DEA method and the method proposed by Andersen and Petersen [26]. The super efficiency model of the improved CCR designed to further distinguish the effectiveness of effective DMU is used to evaluate and rank the research objects. Assuming n role units, input variable X_{ij} ($I = 1, 2, \ldots, m$) in m and output variable Y_{ij} ($r = 1, 2, \ldots, p$) in P, the super efficiency CCR-DEA calculation model is written as follows:

$$\min\left[\theta - \varepsilon\left(\sum_{i=1}^{m} S_i^- + \sum_{r=1}^{p} S_r^+\right)\right],$$
$$\text{s.t.}\begin{cases} \sum_{j=1,j\neq k}^{n} \lambda_j x_{ij} + S_i^- = 0x_0, & i = 1,2,\ldots,m, \\ \sum_{j=1,j\neq k}^{n} \lambda_j y_{ij} - S_r^+ = y_0, & r = 1,2,\ldots,p, \\ \lambda_j \geq 0, & j = 1,2,\ldots,n, \\ S_i \geq 0, \\ S_r \geq 0. \end{cases} \tag{1}$$

In the formula, X_0 and y_0 are input and output vectors of selected decision unit DUM_0, respectively; λ is the combination ratio of n decision units in an effective θ combination newly constructed relative to DUM_0; S_i^- and S_r^+ are relaxation variables of input and output, respectively; and θ is the comprehensive efficiency value of DMU. When $\theta < 1$, DMU is non-DEA effective. When $\theta \geq 1$, DMU is DEA effective; the greater θ is, the higher the level of efficiency is [27].

3.2. Malmquist Index. As the change in tourism efficiency involves a dynamic process, tourism efficiency will change with time. The relative efficiency calculated by the static DEA model cannot be compared with the optimal efficiency frontier constructed by it; that is, the value cannot be directly compared with tourism efficiency over the years. Therefore, the Malmquist index is introduced to conduct an intertemporal analysis of tourism efficiency and a dynamic evaluation of tourism efficiency [28]. The Malmquist index of total factor productivity was first proposed by Malmquist (1953) and later improved by Fare R (1994). The Malmquist Index (MI) is expressed as follows [29]:

$$\text{MI}_{t+1} = \left[\frac{D^{t+1}\left(x^{t+1}, y^{t+1}\right)}{D^{t+1}\left(x^t, y^t\right)} \times \frac{D^t\left(x^{t+1}, y^{t+1}\right)}{D^t\left(x^t, y^t\right)}\right]^{1/2}. \tag{2}$$

In the formula, $D_t(x^{t+1}, y^{t+1})$ represents the distance between the DMU in period $t+1$ and the production front surface in period t; that is, all DMUs in period t are used to construct the production front surface to measure the efficiency of a DMU in period $t+1$. $\text{MI}_{t+1} \in (0, +\infty)$. If $\text{MI}_{t+1} > 1$, then compared with that of period t, the tourism efficiency of period $t+1$ is higher. If $\text{MI}_{t+1} < 1$, then compared with that of period t, the tourism efficiency of period $t+1$ is lower. If $\text{MI}_{t+1} = 1$, relative to period t, there is no change in tourism efficiency in period $t+1$.

3.3. Spatial Weight Matrix Setting

3.3.1. Spatial Autocorrelation Test. Before setting the spatial weight matrix, the spatial data should be analysed to determine the nonrandomness or spatial autocorrelation of distribution. This approach involves spatial analysis with a recognition function. By identifying and distinguishing the characteristics of geographic data and then observing the atypical location, we find the corresponding spatial correlation, clustering, or hot spot pattern; connect the geographic space to the data space; and determine the spatial correlation of economic activities. We use Moran's I index to explore the spatial correlation of regional comprehensive tourism efficiency. The specific formula used is as follows:

$$\text{Moran's } I = \frac{n}{S_0} \times \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} \left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2},$$
$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}. \tag{3}$$

In the formula, n is the province, X_i is the observed value of province i, and W_{ij} represents the spatial weight matrix. The Moran's I index value ranges from −1 to 1. According to Moran's I index value, spatial distribution in the study area can be determined. At a given significance level, a Moran's I of >0 indicates a positive correlation, that is, the units in the study area are spatially clustered; if Moran's I is < 0, this indicates a negative correlation, that is, each unit in the study area is distributed discretely across space; and if Moran's $I = 0$, the space is uncorrelated, that is, each unit in the study area presents an irregular random distribution.

3.3.2. Spatial Weight Matrix. The spatial weight matrix reflects the interdependence of individuals in space. The value is an important basis for further spatial data analysis. In this article, the geographic adjacency weight matrix is used as the spatial weight matrix, and the form is set as follows:

$$W_{ij} = \begin{cases} 1, & \text{If } i \text{ is adjacent to } j \text{ and } i \neq j, \\ 0, & \text{If } i \text{ is not adjacent to } j \text{ or } i = j. \end{cases} \tag{4}$$

That is, when two provinces are adjacent, the value is 1; otherwise, it is 0. At the same time, for our robustness test, we use the geographic distance weight matrix for our spatial econometric analysis. The form of the geographic distance spatial weight matrix is set as follows:

$$W_{ij} = \begin{cases} 1/d_{ij}^2, & i \neq j, \\ 0, & i = j. \end{cases} \tag{5}$$

In the formula, d_{ij}^2 represents the surface distance of latitude and longitude between two cities, and $i \neq j$. When $i = j$, W_{ij} is 0, and the geographic attenuation index is 2.

3.4. Spatial Measurement Model Setting. Spatial panel models can effectively solve the problems of spatially explained variable autocorrelation and measurement error and mainly include the spatial lag model (SLM), spatial error model (SEM), and spatial Dobbin model (SDM) [30]. Considering the possible spatial spillover effect of tourism efficiency and influencing factors in provinces and cities in eastern China, based on the basic spatial panel model, the model is constructed as follows:

$$y_{it} = \beta_0 + \beta_1 \text{ENCO}_{it} + \beta_2 \text{STAT}_{it} + \beta_3 \text{RESO}_{it} + \beta_4 \text{TRAN}_{it} + \beta_5 \text{HR}_{it} + \beta_6 \text{INFO}_{it} + \mu_i + \varepsilon_{it}. \tag{6}$$

In the formula, i is the province, t is the year, and y_{it} is the comprehensive tourism efficiency level (calculated by the super efficiency DEA model). ENCO_{it} is the level of economic development. STAT_{it} is the industrial status (measured by the proportion of domestic tourism revenue of GDP); RESO_{it} is the endowment of tourism resources; TRAN_{it} denotes traffic conditions. HR_{it} is the labour force, and INFO_{it} is the information level. μ_i is the individual fixed effect, and ε_{it} is the error term.

In addition, in view of the possible omission of location factors and other variables in the setting of the econometric model, these unobservable missing variables may also have an impact on the comprehensive efficiency of tourism and lead to spatial dependence, so it is necessary to include the spatial effect in the econometric analysis. The specific spatial measurement model is set as follows:

$$y_{it} = \alpha + \rho \sum_{j=1, j\neq i}^{N} W_{ij} y_{it} + \beta X_{it} + \sum_{j=1, j\neq i}^{N} W_{ij} X_{it} \theta + \mu_i + \nu_t + \varepsilon_{it},$$
$$\varepsilon_{it} = \varphi \sum_{j=1, j\neq i}^{N} W_{ij} \varepsilon_{it} + \phi_{it}. \tag{7}$$

In the formula, ε_{it} is the error term; μ_i and ν_t represent the unobservable regional effect and time effect, respectively; ρ and φ are the spatial lag coefficient and spatial error coefficient, respectively; W_{ij} is the spatial weight matrix, and X is the independent variable vector including the economic development level, industrial status, tourism resource endowment, traffic conditions, labour force, and the information level. Formula (7) is a general nested model of the spatial interaction effect. In the empirical analysis, according to ρ, φ and θ, the spatial econometric model is also different depending on whether the value is 0.

4. Results and Analysis

4.1. Comprehensive Tourism Efficiency Analysis. Based on the data measurement based on the DEA model, the comprehensive tourism efficiency of 11 eastern provinces and cities from 2010 to 2019 is calculated with the help of DEA-Solver software. According to the results (Table 2), the overall tourism efficiency of the eastern provinces and cities is higher than that shown in Table 1 for the past decade. Across provinces and cities, the tourism efficiency of each region varies. Hainan Province has the lowest comprehensive tourism efficiency (0.588), whereas other

Table 1: Descriptive statistics of input-output indices and original data for eastern provinces and cities for 2010 to 2019.

Indices	Indicator	Unit	Definition	Maximum	Minimum	Mean	Standard deviation
Input indicators	Number of employees in the tertiary industry	Ten thousand people	The potential and vitality of tourism	3378.02	165.87	1255.18	682.31
	Number of A-level scenic areas	Home	The value of tourism resource endowment	1292	52	317.45	259.96
	Number of star-rated hotels	Home	The reception capacity of the tourism industry	1175	71	456.18	295.97
	Number of travel agencies	Home	The service capacity of the tourism industry	3281	301	1420.66	736.35
Output indicators	Domestic tourism revenue	One hundred million yuan	The overall economic level of the region	13902.21	235.61	4713.31	3121.61
	Number of domestic tourists	Thousands of people	The level of social activity	93288.00	2521.03	35863.14	20974.24
	Current afforestation area	Thousands of hectares	The ecological environment of the tourism area	371.77	0.71	72.86	88.70

provinces and cities have comprehensive tourism efficiency levels of higher than 0.8, and most provinces and cities have higher levels of tourism efficiency.

In terms of time, the overall tourism efficiency of eastern provinces and cities from 2010 to 2019 maintains a high level, does not change much, shows no significant continuous increase or decrease, or first increases and then decreases or vice versa. The highest efficiency value for 10 years is 1.307, and the lowest value is 1.097. In 2009, the government defined tourism as a strategic pillar industry. The Opinions of the State Council on Accelerating the Development of Tourism in 2009 reported that all efforts should be made to promote investment in and the development of the tourism industry. Several Opinions on Promoting the Reform and Development of Tourism in 2014 called for strengthening and creating room for tourism development. The support and promotion of numerous policies have greatly promoted the development of tourism and the improvement and stability of tourism efficiency.

In terms of regional distribution, there are great differences in comprehensive tourism efficiency among provinces and cities, and the degree of polarization is serious. The average value of comprehensive efficiency over 10 years for Hebei Province is the highest (2.050), and the value for Hainan Province is the lowest (0.588). Among the 11 provinces, the average efficiency levels for Hebei, Tianjin, and Shanghai are higher than 1.5, ranking in the first echelon. The average efficiency levels of Liaoning, Fujian, and Guangdong are higher than 1 but lower than 1.5, ranking in the second echelon. The average efficiency levels of Beijing, Jiangsu, Zhejiang, Shandong, and Hainan are less than 1, and the average efficiency level of Hainan is less than 0.5, ranking in the third echelon. In summary, most provinces and cities in eastern China show high degrees of tourism efficiency and excellent development results. At the same time, the results show that the annual distribution of tourism efficiency is relatively stable and that the regional distribution is unbalanced.

4.2. Dynamic Tourism Efficiency Analysis. Using DEAP2.1 software, the Malmquist index of total factor productivity and its changes in 11 provinces and cities of eastern China from 2010 to 2019 were calculated and decomposed. Among the values measured, the total factor productivity index is measured as the comprehensive efficiency change index × technological progress, and the efficiency change index is also measured as the pure technical efficiency change index × the scale efficiency change index × the technological progress efficiency change index. In other words, the total factor productivity index can be decomposed into the comprehensive efficiency change index and the technological progress efficiency change index, and the comprehensive efficiency change index can be further decomposed into the pure technical efficiency change index and the scale efficiency change index. Through the decomposition of the total factor productivity index, dynamic tourism efficiency can be more clearly analysed.

On the time dimension (Table 3), the total factor productivity of eastern provinces and cities from 2010 to 2019 was higher than 1, indicating that total factor productivity was still on the rise. In the nine periods of tourism efficiency evaluation, the change index of comprehensive efficiency for five periods is higher than 0.95 but less than 1, the change index of comprehensive efficiency for four periods is higher than 1, the change across nine periods show no special signs of regularity, and the average change index of comprehensive efficiency is 0.999. In terms of technological progress efficiency, except for the two periods of 2011–2012 and 2014–2015, the efficiency change index of the other seven periods is higher than 1, and the average value of the technical efficiency change index is 1.065. From a comparative point of view, the change index of technological progress efficiency is the most important driving force of tourism efficiency, while the change index of comprehensive efficiency restricts the improvement of tourism efficiency to a certain extent. After decomposing the comprehensive efficiency change index into pure technical efficiency and the scale efficiency change index, it is found that the pure technical efficiency change index and scale efficiency change

Table 2: Values and ranking of comprehensive tourism efficiency levels in eastern provinces and cities for 2010 to 2019.

DMU	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean	Ranking
Beijing	0.797	0.814	0.833	0.866	0.920	0.910	0.871	0.838	0.796	0.741	0.839	9
Tianjin	1.846	1.818	1.807	2.099	1.446	2.075	1.963	1.871	1.916	1.871	1.871	2
Hebei	1.303	0.912	1.496	1.822	2.250	2.649	2.091	2.763	2.315	2.898	2.050	1
Liaoning	1.370	1.223	1.326	1.281	1.242	1.142	1.197	1.260	1.203	1.227	1.247	4
Shanghai	1.859	1.702	1.918	1.826	1.223	1.942	1.737	1.673	1.381	1.240	1.650	3
Jiangsu	0.734	0.734	0.769	0.823	0.909	0.972	0.998	0.988	0.975	0.954	0.886	7
Zhejiang	0.723	0.794	0.828	0.851	0.890	1.056	0.997	0.960	0.911	0.882	0.889	6
Fujian	0.722	2.557	1.165	1.201	0.703	0.850	0.801	0.943	0.861	0.885	1.069	5
Shandong	1.052	0.825	0.819	0.885	0.794	0.895	0.804	0.826	0.796	0.776	0.847	8
Guangdong	0.899	0.894	1.036	1.174	1.341	1.308	1.222	1.092	0.982	0.744	1.069	5
Hainan	0.765	0.507	0.652	0.460	0.439	0.577	0.621	0.619	0.616	0.626	0.588	10
Mean	1.097	1.162	1.150	1.208	1.105	1.307	1.209	1.258	1.159	1.168	1.182	

Table 3: Overall dynamic tourism efficiency changes of eastern provinces and cities divided by time.

Time	Comprehensive efficiency (effech)	Efficiency of technological progress (techch)	Pure technical efficiency (pech)	Scale efficiency (sech)	Total factor productivity (tfpch)
2010–2011	0.977	1.250	1.005	0.972	1.220
2011–2012	1.052	0.974	1.004	1.048	1.025
2012–2013	0.988	1.049	0.998	0.989	1.036
2013–2014	1.010	1.025	1.000	1.010	1.035
2014–2015	1.032	0.988	0.993	1.039	1.019
2015–2016	0.990	1.094	1.002	0.987	1.082
2016–2017	1.009	1.049	0.995	1.015	1.058
2017–2018	0.976	1.092	0.987	0.989	1.066
2018–2019	0.966	1.090	0.993	0.972	1.053
Mean	0.999	1.065	0.997	1.002	1.065

index are higher than 1 in five periods and less than 1 in four periods. The average efficiency change indexes of the two are 0.977 and 1.002, respectively. This reveals problems with the operation, management, and technologies of the tourism industry over the last ten years. While improving tourism efficiency relying on industrial scale is more effective, this cannot achieve outstanding results. Therefore, the change index of technical efficiency is the main factor that affects the improvement of tourism efficiency in eastern provinces and cities. In descending order, the indexes affect the tourism efficiency total factor productivity index as follows: the technical efficiency change index, scale efficiency change index, and pure technical efficiency change index.

For provinces and cities (Table 4), the average TFP of 11 eastern provinces and cities from 2010 to 2019 is 1.062, and the TFP index of each province and city is higher than 1. Among the values, the average growth rate of the comprehensive efficiency change index is −0.1%, the average growth rate of the technological progress efficiency change index is 6.2%, the average growth rate of the pure technical efficiency change index is −0.3%, and the average growth rate of the scale efficiency change index is 0.2%. Therefore, the overall high level of tourism efficiency found in eastern provinces is mainly due to improvements to the efficiency of technological progress, indicating that the utilization of tourism resources and the management level of the tourism industry in this region have been effectively improved. Further subdivision shows that the improvement of the total factor productivity indexes of Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, and Fujian is jointly driven by comprehensive efficiency and technological progress efficiency. These seven provinces and cities should further improve the management level of the tourism industry by maintaining the current level of tourism efficiency. The comprehensive efficiency change indexes of Beijing, Guangdong, Shandong, and Hainan are less than 1, so the improvement of their total factor production efficiency is mainly driven by the efficiency of technological progress. To increase comprehensive efficiency, these four provinces and cities should formulate measures to improve their own resource utilization levels and support talent according to their low pure technical efficiency or scale efficiency.

4.3. Temporal and Spatial Differentiation and Influencing Factor Analysis

4.3.1. Temporal and Spatial Differentiation. Based on the analysis of the index of tourism development efficiency and total factor productivity for 11 eastern provinces and cities, it is necessary to judge whether there are some correlations based on the spatial relations between provinces and cities and whether these correlations will change with time. According to Formula (3), the panel Moran's I index value of tourism comprehensive efficiency is −0.2486, which is significant at the significance level of 5%, indicating that after controlling the explanatory variables, tourism comprehensive efficiency shows a significant spatial correlation, and

Table 4: Average values of dynamic tourism efficiency for 11 provinces and cities in eastern China from 2010 to 2019.

Province or city	Comprehensive efficiency (effech)	Efficiency of technological progress (techch)	Pure technical efficiency (pech)	Scale efficiency (sech)	Total factor productivity (tfpch)
Beijing	0.992	1.071	0.970	1.022	1.062
Tianjin	1.000	1.094	1.000	1.000	1.094
Hebei	1.000	1.123	1.000	1.000	1.123
Liaoning	1.000	1.062	1.000	1.000	1.062
Shanghai	1.000	1.013	1.000	1.000	1.013
Jiangsu	1.030	1.060	1.000	1.030	1.091
Zhejiang	1.022	1.053	1.000	1.022	1.076
Fujian	1.023	1.010	1.002	1.021	1.034
Shandong	0.972	1.094	1.000	0.972	1.064
Guangdong	0.979	1.060	1.000	0.979	1.038
Hainan	0.978	1.052	1.000	0.978	1.028
Mean	0.999	1.062	0.997	1.002	1.062

tourism comprehensive efficiency is distributed discretely among provinces. To further explore the spatial distribution characteristics of comprehensive tourism efficiency in eastern provinces and cities, the comprehensive tourism efficiency of eastern provinces and cities is spatially clustered and visualized in 2010, 2015, and 2019 (Figure 1).

4.3.2. Analysis of Influencing Factors. According to Moran's *I* index, we find a significant spatial correlation for provincial comprehensive tourism efficiency in eastern China. Therefore, the spatial econometric model is used to analyse the influencing factors of comprehensive tourism and to analyse the direct, indirect, and total spatial spillover effects of each influencing factor to clarify the action mechanism of the influencing factors. According to Elhorst's research [31], the use of the spatial lag model (SAR), spatial error model (SEM), or spatial Doberman model (SDM) is judged through calculation. The specific test results are shown in Table 5.

Since the spatial lag model with spatial fixed effects is more suitable for estimating the econometric equation of formula (7), we estimate the spatial lag model with spatiotemporal dual fixed effects. To facilitate comparison, MATLAB 2016b and its spatial measurement software package are used to estimate and test the relevant models of six factors influencing tourism efficiency in 11 provinces and cities in eastern China, and the results of the panel fixed effect model are displayed in Table 6. According to Table 6, the p value of the measurement results of the provincial spatial SAR model for eastern China is significantly different from 0, indicating significant spatial effects in all regions.

From the model measurement estimation results, the coefficients of various variables of the SAR model have both significant positive and negative effects and insignificant effects. Among them, the impact of economic development and information levels on tourism efficiency is not significant. Tourism resource endowment and the labour force have significant negative effects on tourism efficiency. The impact of industrial status and traffic conditions on tourism efficiency is positive. Specifically, we find the following:

(1) The level of economic development, as an important indicator of the scale and speed of regional economic development, shows high levels of productivity and talent attraction. Although this measure can have a certain impact on tourism efficiency in all areas of tourism investment and tourism demand, the more developed the economy is, the better outcomes are, because tourism and other industries also compete for policies and economic resources. Regarding regional tourism efficiency, while vigorously developing the regional economy, paying attention to the investment and policy preferences of the tourism industry is key to improve tourism efficiency, especially in economically developed areas of China's eastern region.

(2) The level of informatization is of certain significance to the development and innovation of regional tourism. The renewal and use of advanced technology can help the tourism industry reduce its dependence on human factors and have an impact on the comprehensive management of tourism facilities and tourists' participation in tourism activities. In terms of improving tourism efficiency, the variable failed to pass the significance test. To improve innovation in the tourism industry and promote the sustainable development of tourism, the information level should play a more important role in tourism development and further affect the tourism efficiency of eastern provinces and cities.

(3) Tourism resource endowment is an important basis for regional tourism development. Tourist attractions are critical in helping provinces and cities to develop tourism. The quality and quantity of tourism attractions are of great significance to the development of tourism. The negative effect of tourism resource endowment on tourism efficiency coincides with resource curse theory. Rich tourism resources do not necessarily equate to high tourism efficiency [32]. In addition to their insufficient utilization of tourism resources, different provinces and cities lack cooperation in tourism resource development, the interregional division of labour and cooperation is unreasonable, and the development of tourism resources and the noncooperative game of tourism products will also affect tourism efficiency.

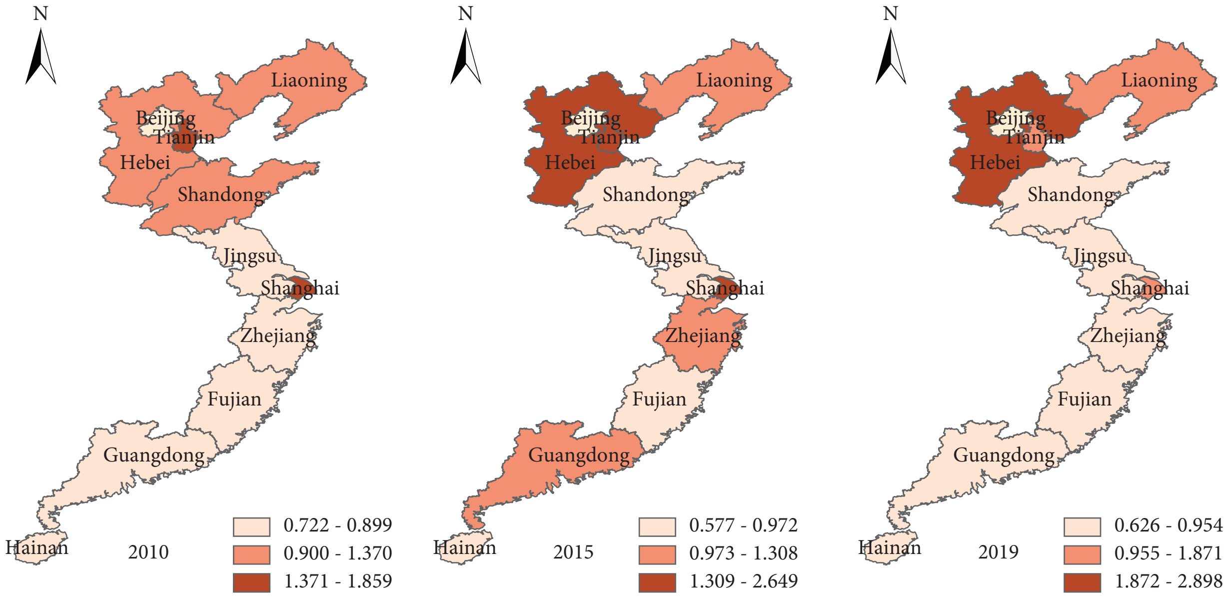


Figure 1: Spatial Distribution map of comprehensive Tourism efficiency (part).

(4) The labour force indirectly affects the development efficiency and performance of the tourism economy. As a labour-intensive industry, the tourism industry has certain requirements for the labour force. However, this does not mean that the more tourism practitioners there are, the higher the level of tourism efficiency will be, in turn meeting the needs of the regional tourism industry. At the same time, too many low-skilled employees may also be an important reason for the negative correlation found between the labour force and tourism efficiency. Therefore, on the one hand, the eastern provinces and cities should formulate differentiated incentive policies for tourism practitioners according to different levels of economic development; on the other hand, these entities should also absorb an appropriate amount of labour according to the needs of the industry.

(5) Industrial status is mainly reflected in attention to industry and the optimization of the industrial structure. Relevant policies can lay a foundation and good external conditions for the development of the tourism industry, and the optimization and upgrading of industrial structure can promote the healthy development of the tourism industry and improve tourism efficiency. Industrial status has a significantly positive impact on tourism efficiency. To improve tourism efficiency, the tourism industry must be a focus and its industrial structure must be optimized.

(6) Traffic conditions support the development of tourism and directly affect the sustainable development of the sector in various provinces and cities. The development of transportation changes the breadth and depth of the spatial economic ties of the tourism industry within provinces and cities to a certain extent and affects tourists' access to areas, and the resulting psychological and spatial distance jointly act on tourists' tourism motivations and demands and can attract more investment from tourism investors. To create more room for tourism development, more development space can be obtained. Therefore, traffic conditions have an important positive impact on the tourism efficiency of all eastern provinces and cities. The continuous improvement of traffic conditions will be critical to promoting the development of regional tourism and should be given sufficient attention.

To more thoroughly and accurately interpret the influencing and transmission effects of the explanatory variables on the explained variables, we continue to analyse direct, indirect, and total spatial spillover effects using a spatial SAR model. Here, LeSage and Pace's method [33] is used to further estimate the spatial spillover effect in the spatial lag model (Table 7). The direct effect reflects the impact of explanatory variables such as the level of regional economic development on comprehensive tourism efficiency. The indirect effect represents the spatial spillover effect of explanatory variables such as the effect of the level of regional economic development on the comprehensive tourism efficiency of surrounding provinces. The overall spatial spillover effect is the sum of the direct spatial spillover effect and indirect spatial spillover effect. Among these effects, the indirect spatial spillover effect does not pass the significance test, so we mainly analyse the direct spatial spillover effect and total spatial spillover effect.

The direct spatial spillover effect is also known as the intraregional spillover effect. Table 7 shows that the spillover effects of industrial status, traffic conditions, tourism resource endowment, and the labour force on tourism efficiency pass the significance test. The first two factors have positive effects, and the latter two have negative effects,

Table 5: Spatial metrological test results.

Content evaluation methods	Inspection methods	Geographic adjacency matrix	
		Statistics	Probability
SAR model and SEM model test	LM-lag test	6.1879	0.013
	R-LM-lag test	2.2481	0.134
	LM-err test	4.1682	0.041
	R-LM-err test	0.2283	0.633
Fixed effect test for spatial Dubin model	SFE-LR	118.5147	0.0000
	TFE-LR	15.1073	0.1282
Hausman test of SDM model	Hausman test	58.2209	0.0000
Simplified test of SDM model	Wald-lag test	1.3323	0.9698
	LR-lag test	1.4864	0.9604
	Wald-err test	2.2124	0.8992
	LR-err test	2.4679	0.8720

Table 6: Spatial metering estimation results.

Variable	FE model	SAR model
$ENCO_{it}$	−0.189	−0.185
	(0.411)	(−0.465)
$STAT_{it}$	4.835 ***	4.804 ***
	(1.243)	(3.989)
$RESO_{it}$	−0.625 **	−0.623**
	(0.268)	(−2.400)
$TRAN_{it}$	1.048*	1.052*
	(0.607)	(1.789)
HR_{it}	−0.648*	−1.887 *
	(0.354)	(−1.328)
$INFO_{it}$	0.136	0.136
	(0.102)	(1.374)
p		−0.031
		(−0.275)
Log-likelihood		5.2193
Observations	110	110
Number of province	11	11
R-squared	0.299	0.8100

Note: *, **, and *** indicate significance at the levels of 10%, 5% and 1%, respectively. Statistics are in parentheses.

presenting values of 4.862 ($p < 0.01$), 1.076 ($p < 0.1$), −0.641 ($p < 0.05$), and −0.656 ($p < 0.1$), respectively. The more the tourism industry is valued by the eastern provinces and cities, the more easily the development of the tourism industry will produce economies of scale and form a "magnetic field effect," such as talent, funds, and technology, which can not only optimize the industrial structure but also further improve scientific and technological content and technology and innovation spillovers of the tourism industry. Transportation is not only an important facet of tourism infrastructure but also an indispensable prerequisite for promoting the development of tourism. Transportation conditions greatly affect consumers' choices of tourism destinations, which will affect the development of tourism resources and of the whole tourism industry. Good traffic conditions can enable tourists to allocate more time and energy to "having fun" rather than to "travel." In other words, developed transportation networks can enhance the scale effect of tourism activities, help enhance the overall profits of tourism in specific regions, and improve the marginal output of tourism reception resources. It is worth noting that the effects of economic development and information levels do not pass the significance test. The empirical results of this article deviate from the traditional view that economic foundations and innovation ability must contribute to the improvement of industrial efficiency and further confirm that the growth of tourism efficiency in eastern provinces and cities in the studied period and region is not mainly derived from regional economic foundations or driven by innovation. The possible reasons for this are as follows: First, the level of economic development is not necessarily related to regional investment in tourism or the construction of tourism development infrastructure; second, the government-led characteristics of China's tourism development are evident. Compared with those invested in the primary and secondary industries and in other high-tech industries, fewer scientific research funds are invested in the tourism industry, which is considered a typical service industry. In the absence of external constraints, the scientific and technological innovation of tourism enterprises is limited, the application of informatization in tourism is not sufficient, and the role of informatization is relatively

Table 7: Estimation results of direct, indirect, and total effects.

Effect	$ENCO_{it}$	$STAT_{it}$	$RESO_{it}$	$TRAN_{it}$	HR_{it}	$INFO_{it}$
Direct effect	−0.180	4.862 ***	−0.641**	1.076*	−0.656*	0.139
	(−0.482)	(4.006)	(−2.460)	(1.877)	(−1.961)	(1.494)
Indirect effect	0.001	−0.089	0.015	−0.025	0.015	−0.003
	(0.010)	(−0.155)	(0.196)	(−0.181)	(0.181)	(−0.152)
Total effect	−0.179	4.772 ***	−0.626**	1.051*	−0.641*	0.136
	(−0.482)	(3.559)	(−2.408)	(1.837)	(−1.910)	(1.470)

Note: *, **, and *** indicate significance at the levels of 10%, 5%, and 1%, respectively. t statistics are in parentheses.

Table 8: Estimation results of the robustness test.

Effect	$ENCO_{it}$	$STAT_{it}$	$RESO_{t}$	$TRAN_{it}$	HR_{it}	$INFO_{it}$
Direct effect	−0.189	4.817 ***	−0.623**	1.080*	−0.662*	0.138
	(−0.472)	(3.981)	(−2.337)	(1.840)	(−1.956)	(1.426)
Indirect effect	0.003	−0.101	0.011	−0.021	0.015	−0.003
	(0.055)	(−0.194)	(0.153)	(−0.159)	(0.194)	(−0.138)
Total effect	−0.187	4.717 ***	−0.611**	1.059*	−0.647*	0.136
	(−0.475)	(3.626)	(−2.227)	(1.782)	(−1.911)	(1.402)

Note: *, **, and *** indicate significance at the levels of 10%, 5%, and 1%, respectively. Statistics are in parentheses.

limited. The effect of tourism resource endowment and the labour force on tourism efficiency is negative. The possible reasons for this are as follows: the higher the resource endowment is, the more difficult it is to make full use of these resources, or the richness of tourism resources renders the problem of resource homogeneity more prominent and competition for tourism products more intense. The latter occurs because tourism is still a labour-intensive industry, and the low threshold of tourism employment leads to a large influx of personnel into the tourism industry. Excessive labour input is not conducive to the improvement of tourism efficiency. The quality of labour personnel is also an important aspect that has a different impact on tourism efficiency. The total spatial spillover effect is similar to the direct spatial spillover effect. Industrial status (4.772, $p < 0.01$) and traffic conditions (1.051, $p < 0.1$) have a positive effect, tourism resource endowment (−0.626, $p < 0.05$) and the labour force (−0.641, $p < 0.1$) have a negative effect, and the level of economic development and informatization do not pass the significance test.

To eliminate possible endogeneity problems in our data, the geographic distance spatial weight matrix is used to replace the geographical adjacency matrix to test the robustness of the factors that influence tourism comprehensive efficiency and its spatial spillover effect (Table 8). The results show that the benchmark regression results remain unchanged, indicating that the robustness test has the ideal effect.

5. Conclusion and Discussion

5.1. Conclusion. Using the static super efficiency DEA model, dynamic Malmquist index, and spatial econometric model, this article systematically analysed the panel data of 11 eastern provinces for 2010 to 2019 and measured levels of and differences in tourism efficiency in the time and space dimensions. This article further explored the state and influencing factors of the temporal and spatial differentiation of tourism efficiency in these 11 provinces and cities. Our specific conclusions are as follows:

(1) From the static data, the average level of tourism efficiency in eastern provinces for the last ten years is higher than 1, and it has been at a relatively high level overall. However, in terms of tourism efficiency in specific time periods, there are still great differences in the spatial distribution. The leading two cities in terms of average efficiency, with values of above 1.5, are Hebei and Tianjin, while Hainan has the lowest average efficiency level of only 0.588. At the same time, the efficiency values for different years and different provinces and cities have been affected by multiple factors and have changed sporadically, resulting in no obvious trends in the interior region.

(2) In terms of dynamic index changes, the Malmquist index of the overall tourism efficiency of eastern provinces and cities is 1.062 for the past 10 years and is on the rise. Among the contributions of each index to the tourism efficiency total factor productivity index, the change index of technical efficiency is the highest, the change index of scale efficiency is the second highest, and the change index of pure technical efficiency is relatively low. Among provinces and cities, the improvements of the total factor productivity index in Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, and Fujian are jointly driven by comprehensive efficiency and technological progress efficiency, while improvements of the total factor productivity index in Beijing, Guangdong, Shandong, and Hainan are mainly driven by technological progress efficiency.

(3) From the spatial autocorrelation test and calculation obtained from a spatial econometric model, the tourism efficiency of eastern provinces and cities

shows obvious signs of temporal and spatial differentiation and spatial correlation for the past decade. In terms of influencing factors, industrial status and traffic conditions have a significantly positive effect on tourism efficiency; tourism resource endowment and the labour force have a negative effect, economic development and information levels do not pass the significance test, and changes in provincial and municipal tourism efficiency are jointly affected and driven by a variety of factors.

In summary, the overall tourism efficiency of the 11 eastern provinces and cities studied is relatively high, and the annual average efficiency level is relatively stable. However, compared with that in the subdivided regions, tourism efficiency across the studied provinces and cities varies greatly. However, overall tourism efficiency is still increasing and can be further improved in terms of technical, scale, and pure technical efficiency, especially in provinces and cities with low efficiency values. Eastern provinces and cities show obvious patterns of temporal and spatial differentiation in terms of tourism efficiency and spatial correlation. Industrial status, traffic conditions, tourism resource endowment, and the labour force are the main factors driving this differentiation.

5.2. Discussion. Since the institution of the country's reform and opening up policy, China's tourism has developed rapidly. Policy support, capital investment, and scientific planning have in turn undergone more significant progress. The healthy development of tourism and change in tourism efficiency have been affected by multiple factors, including the development status of provinces and cities, industrial development maturity, capital investment, the allocation of resources, and scientific levels [34, 35]. As an important facet of tourism development quality, tourism efficiency is of important practical significance for expanding the scale of tourism development, improving the comprehensive competitiveness of tourism, and realizing the sustainable development of tourism. Examining the development and evolution of tourism efficiency can help tourism in China develop in a more effective manner. Based on the above discussion and analysis of tourism efficiency and its decomposition efficiency in China's eastern provinces from 2010 to 2019, the following discussion considers efficiency levels, efficiency changes, and temporal and spatial differentiation and corresponding influencing factors.

(1) For eastern provinces and cities, technical and scale efficiency influence each other and more significantly affect tourism efficiency. Therefore, existing technical conditions should be utilized to increase scale efficiency such that scale and technical efficiency can be effectively combined to inject improve tourism efficiency in China's eastern provinces and cities.

(2) There are significant regional differences in tourism efficiency across eastern provinces and cities. This unbalanced pattern is closely related to the industrial status of tourism across provinces and cities, regional traffic conditions, the development and utilization of tourism resources, and the absorption of the labour force. Therefore, we should pay attention to changing tourism's industrial structure and improving the utilization efficiency of resource elements, constantly improving regional traffic conditions and the industrial environment, enriching the workforce, and realizing a more balanced and consistent development of tourism efficiency in China's eastern provinces and cities.

This study presents some shortcomings, which are as follows. Due to the complexity and comprehensiveness of the tourism industry, the selection of input indicators must be further improved. The optimization and reorganization of the spatial network of tourism efficiency is controlled by multiple factors, but the discussion of the impacts of various factors on tourism efficiency is still slightly insufficient. In view of a lack of tourism statistical data, the use of alternative index data may lead to biased results, which must be addressed in follow-up studies.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (No. 41671140).

References

[1] M. J. Farrell, "The measurement of productive efficiency," *Journal of the Royal Statistical Society: Series A*, vol. 120, no. 3, pp. 253–278, 1957.

[2] B. Zhou, Y. Xu, and S. K. Lee, "Tourism development and regional production efficiency: evidence from southwestern China," *Tourism Economics*, vol. 25, no. 5, pp. 800–818, 2019.

[3] S. Chaabouni, "China's regional tourism efficiency: a two-stage double bootstrap data envelopment analysis," *Journal of Destination Marketing and Management*, vol. 11, no. 3, pp. 183–191, 2019.

[4] C. D. Köksal and A. A. Aksu, "Efficiency evaluation of A-group travel agencies with data envelopment analysis (DEA): a case study in the Antalya region, Turkey," *Tourism Management*, vol. 28, no. 3, pp. 830–834, 2007.

[5] O. Mhlanga, "Factors impacting airline efficiency in Southern Africa: a data envelopment analysis," *GeoJournal*, vol. 84, no. 3, pp. 759–770, 2019.

[6] K. Wang, H. Lin, C. Gan, and C. Deng, "Spatio-temporal coupling relationship between tourism poverty alleviation efficiency and economic development level in contiguous destitute areas: a case study of Wuling mountain," *Economic Geography*, vol. 40, no. 2, pp. 200–208, 2020.

[7] S. Wang, Y. Chu, A. Guo, and Y. Guo, "Study on high quality development of tourism economy in 18 key provinces along

the one belt and one road initiative: based on the measurement of tourism resource conversion efficiency," *Scientia Geographica Sinica*, vol. 40, pp. 1505–1512, 2020.

[8] Z. Yao, T. Chen, S. Yin, and X. Li, "Regional tourism eco-efficiency Model and an empirical research of Hainan province," *Scientia Geographica Sinica*, vol. 36, pp. 417–423, 2016.

[9] Y. Fang, Z. Huang, D. Li, and F. Wang, "The measurement of Chinese provincial tourism developing efficiency and its spatio-temporal evolution," *Economic Geography*, vol. 35, no. 8, pp. 189–195, 2015.

[10] X. Ma, "Progress and trend of tourism efficiency research home and abroad," *Human Geography*, vol. 27, no. 3, pp. 11–17, 2012.

[11] Z. Wang and Q. Liu, "The evolution and influencing factors of spatial network structure of China's provincial tourism efficiency," *Scientia Geographica Sinica*, vol. 41, pp. 397–406, 2021.

[12] K. Wang, Z. Huang, Y. Tao, and Y. Fang, "Study on spatial characteristics and spillover effects of urban tourism efficiency: a case of Yangtze river delta," *Economic Geography*, vol. 33, no. 4, pp. 161–167, 2013.

[13] P. Shen, N. Zhou, Y. Zhang, K. Wang, and Z. Li, "Spatio-temporal evolution and influencing factors of tourism industry efficiency in Jiangsu province based on Dea-Malmquist index secondary decomposition model," *Resources and Environment in the Yangtze Basin*, vol. 27, pp. 53–62, 2018.

[14] Q.-l. Pan, Y.-q. Song, L. Chen, and Z.-k. Pan, "The spatial pattern and influencing factors of county-scale tourism efficiency in Shaanxi province," *Journal of Natural Resources*, vol. 36, no. 4, pp. 866–878, 2021.

[15] F. Cao, Z. Huang, M. Xu, and K. Wang, "Spatial-temporal pattern and influencing factors of tourism efficiency and the decomposition efficiency in Chinese scenic areas: based on the Bootstrap-DEA method," *Geographical Research*, vol. 34, pp. 2395–2408, 2015.

[16] Z. Ding, Q. Lai, X. Huang, J. Zhu, and P. Lin, "Study on the spatial differences and convergences of the tourism efficiency in the Forest parks of China," *Forestry Economics*, vol. 38, no. 11, pp. 41–48, 2016.

[17] A. Paul, D. William, and H. Gao, *Economics*, Economic Science Press, Beijing, China, 1991.

[18] H. Deng and L. Lu, "The urban tourism efficiencies of cities in Anhui province based on DEA model," *Journal of Natural Resources*, vol. 29, pp. 313–323, 2014.

[19] J. Wei, J. Hu, and L. Zhu, "Spatial-temporal evolution and influence mechanism of tourism development efficiency in Hubei and Anhui provinces," *Economic Geography*, vol. 38, no. 8, pp. 187–195, 2018.

[20] F. Cao, Z. Huang, and F. Yu, "The spatial evolution of travel efficiency of China's national scenic areas and its driving mechanism," *Geographical Research*, vol. 33, no. 6, pp. 1151–1166, 2014.

[21] F. Martínez-Roget and X. A. Rodríguez, "Occupancy level and productivity in rural tourism establishments: the case of Galicia, Spain," *Tourism Economics*, vol. 12, no. 2, pp. 279–289, 2006.

[22] M. Jiang, N. Zhou, and Y. Liang, "The difference of regional tourism spatial structure in Jiangsu province based on aggregation and fractal," *Resources and Environment in the Yangtze Basin*, vol. 21, no. Z2, pp. 81–88, 2012.

[23] K. Wang, Z. Huang, F. Cao, F. Yu, and F. Tang, "Spatial pattern evolution of urban tourism performance and its influence factors in Pan-Yangtze river delta," *Journal of Natural Resources*, vol. 31, pp. 1149–1163, 2016.

[24] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[25] Q. Wei, *Data Envelopment Analysis*, Science Press, Beijing, China, 2004.

[26] P. Andersen and N. C. Petersen, "A procedure for ranking efficient units in data envelopment analysis," *Management Science*, no. 39, pp. 1261–1265, 1993.

[27] J. Hua and C. Ye, "Urban land use efficiency and spatio-temporal differentiation of Guangdong province based on DEA," *Research of Soil and Water Conservation*, vol. 25, pp. 283–288, 2018.

[28] S. Xu, S. Fang, Y. Gan, and X. Qin, "Measurement and spatio-temporal evolution of tourism poverty alleviation efficiency in Guangxi border ethnic areas," *World Geographical Studies*, vol. 30, no. 2, pp. 367–377, 201.

[29] R. Fare, S. Grosskopf, and M. Norris, "Productivity growth, technical progress, and efficiency change in industrialized Countries," *The American Economic Review*, vol. 84, no. 1, pp. 66–83, 1994.

[30] L. Anselin, "Spatial externalities, spatial multipliers, and spatial econometrics," *International Regional Science Review*, vol. 26, no. 2, pp. 153–166, 2003.

[31] J. P. Elhorst, "Dynamic spatial panels: models, methods and inferences," *Spatial Econometrics*, Springer, Berlin, Heidelberg, Germany, 2014.

[32] R. M. Auty, *Sustaining Development in Mineral Economies: The Resource Curse Thesis*, Routledge, London, UK, 1993.

[33] J. LeSage and R. K. Pace, *Introduction to Spatial Econometrics*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2009.

[34] T. Yu, B. Zuo, X. Song, and Y. Wu, "The influence of tourism development on regional economic efficiency: evidence from 283 prefecture-level cities in China," *Geographical Research*, vol. 39, no. 6, pp. 1357–1369, 2020.

[35] Y. Fang, M. Xie, D. Bi, and Y. Xiao, "An exploratory analysis on the spatial correlation characteristics and influential factors of inbound tourism in China: a geographical weighted regression perspective," *Tourism Science*, vol. 28, no. 3, pp. 22–35, 2014.

Hindawi
Complexity
Volume 2021, Article ID 3994004, 12 pages
https://doi.org/10.1155/2021/3994004

Research Article

Comparative Studies on Trade and Value-Added Trade along the "Belt and Road": A Network Analysis

Guangjun Sui,[1] Jialing Zou ,[1] Shuang Wu ,[2] and Danling Tang[3,4]

[1]*Guangdong Institute for International Strategies, Guangdong University of Foreign Studies, Guangzhou 510420, China*
[2]*Division of Humanities and Social Sciences, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai 519087, China*
[3]*Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China*
[4]*Guangdong Key Laboratory of Ocean Remote Sensing, State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 511458, China*

Correspondence should be addressed to Shuang Wu; sharon9081@life.hkbu.edu.hk

Received 27 May 2021; Revised 21 July 2021; Accepted 7 August 2021; Published 14 August 2021

Academic Editor: Fabio Saracco

With the deepening of regional industrial transfer, traditional trade data cannot fully explain the real level of trade in a region. This paper aims to reveal the changing pattern of real trade along the "Belt and Road" by establishing a value-added trade (VAT) network of the "Belt and Road" countries and comparing it with the trade network. Applying a network method, we analyze and compare the structures, characteristics, evolutions, and underlying dynamisms of both networks. With a thorough interpretation and visualization of the network density, network centrality, trade communities, and influencing factors of both networks in the three time sections of 2005, 2010, and 2015, we come to the following findings: (1) the connectivity of both networks has been greatly enhanced, reflecting a more integrated regional economy. (2) The center-external structure of both networks has been further strengthened and the polarization of the VAT network is more obvious than that of the trade network. (3) The organizations of trade communities and VAT communities are largely different. Specifically, China rapidly increased its core position in the VAT network and incorporated more and more countries into its VAT community. (4) The underlying factors have similar impacts on both networks. While the growth of regional economic size and free trade agreements will enhance both trade and VAT, the economic gaps and population differences among the "Belt and Road" countries will prohibit regional economic ties. Based on these findings, we propose suggestions on further regional economic cooperation by taking advantage of China's core position to promote regional VAT, construct broad trade channels, and enhance trade governance.

1. Introduction

The "Belt and Road initiative" (BRI) is China's greatest international economic envisagement, aiming at stimulating economic development in a vast region covering countries in Asia, Europe, and Africa with the theme of infrastructure development, policy dialogue, unimpeded trade, financial support, and people-to-people exchange. Since it was proposed in 2013, trade ties among the BRI countries have become increasingly closer, facilitating a tighter intraregional trade network. This also deepened the transfer of industries and the division of labor within the BRI region. With the deepening of regional industrial transfer and production cooperation, traditional trade data cannot adequately account for the true level of trade because the export products of one country may contain the import of raw and processed materials from another country. Compared with the trade data, VAT data can counteract on this effect and better reflect the true level of trade [1]. Our paper aims to reveal the changing pattern of the true level of trade of the BRI region by comparing the VAT network with the trade network. Through analyzing the structures, characteristics, evolutions, and underlying dynamisms of both trade network and VAT network of the BRI region, we try to fully interpret the development and change of trade among the BRI countries and propose pertinent suggestions for future regional trade cooperation and production connection.

Since the global and regional trade networks maintain distinctive hallmarks of networks, scholars have taken network methods to analyze the trade networks and explore the intricate relations of trade among countries and regions [2]. For example, Garlaschelli and Loffredo [3] and De Benedictis and Tajoli [4] took network analysis to delineate the global trade network structure. De Benedictis and Tajoli [4] pointed out that the evolution of trade network has great impact on trade relations and trade policies. Using network methods, scholars also analyzed the topology, community structure, and evolution of world trade web [5–7]. Taking a further step, Jiang et al. [8] analyzed the evolution of global trade communities and proposed directions for China's geopolitical strategies. Network methods also played an important role in analyzing product trade networks, especially the global trade networks of energy and mineral products [9, 10], agricultural products [11], waste products [12, 13], and other important products [14, 15]. These studies are not only of great value in comprehending the characteristics of product trade networks but also help to understand the organizations and patterns of global production networks and global geopolitical relations.

The proposition of BRI accelerates research on trade networks in the BRI region. Taking the overall BRI trade network as the research object, scholars analyze the evolutionary characteristics of the network structure, changes in trade communities, and carry out an in-depth description of the BRI trade network in recent years [16, 17]. On this basis, Chong et al. [18] discussed the forming factors and influence of proximity, culture, and institutions on the structure of the BRI trade network. Chen et al. [19] further analyzed the cultural trade network of the BRI countries and its evolution. Incorporating the BRI trade network into the global trade network, Song et al. [20, 21] compared the topological structure of the BRI trade network with that of the global trade network and sorted out the relationship between the trade communities in the BRI trade network and those in the global trade network. Other scholars paid attention to the structures and influencing factors of the trade network of industries and subdivided products among the BRI countries, including the high-end manufacturing [22], agriculture [23], and natural gas [24].

Due to the deepening of international industrial division of labor, the traditional international trade data cannot fully reflect the true level of trade. In order to measure the evolution of the network structure of the real trade between regions, research on trade networks based on value added has recently begun to emerge. Ferrarini [25] first measured and visualized the global vertical VAT network and found that the central position of China's automobile and electronics industry was rising. Cerina et al. [26] and Zhu et al. [27] used the network method to analyze the VAT in the global trade network and described the specific form of the global value chain. By calculating the foreign value-added part of each country's export trade, Amador and Cabral [28] built a global nonweighted VAT network and depicted its structural evolutionary characteristics. They found that large trading countries play a key role in the VAT network. In addition, based on the network analysis method, Sun et al. [29] depicted the topological structure of the global manufacturing VAT network and further discussed the factors affecting its structural evolution. Focusing on the manufacturing VAT network of the BRI countries, scholars found a rapid growth trend after 2003, largely influenced by the regional trade agreement relationship and the increase in economic aggregates [30].

Taken together, there is a growing body of literature on trade among the BRI countries based on network analysis. These studies have portrayed the trade network structure of the BRI region as well as its evolutionary characteristics and influencing factors. Compared with the traditional trade, VAT can better reflect the true level of trade [31]. Research based on VAT network analysis has also shown a rising trend. However, few studies have considered the VAT of the BRI region, let alone to describe its evolutionary characteristics.

To fill in this gap, we try to establish the BRI VAT network and analyze its structure, characteristics, evolution, and underlying dynamism with a comparison of the BRI trade network. In the following parts, we first introduce the method and data we employed for calculating the VAT among the BRI countries and establishing the BRI VAT network as weighted networks. We then compare the BRI trade network with the BRI VAT network by the three descriptors of network density, network centrality, and trade communities, combined with a further exploration of their evolutionary trends. In the next part, we build up a model analyzing the factors that influence the two networks to understand their underlying dynamisms and differences. Based on the above analysis, we finally come to the conclusion and propose pertinent policy suggestions for strengthening the trade links among the BRI countries. By analyzing the evolution of trade and VAT network of the BRI countries, this paper will help understand the significance of the BRI, especially from the perspective how BRI strengthened economic linkages.

2. Method and Data

2.1. Research Area. The BRI insists an ethos of inclusiveness, which does not restrict to a limited geographical space [32]. Yet, for the convenience of research, scholars usually have their own definitions of the spatial scope of the BRI [16, 33]. Given the availability of international input-output table data and based on previous related research, the BRI countries in this article refer to China and the following 63 countries, as shown in Table 1 and Figure 1.

2.2. Data Source. For the network construction, we use two main sorts of data, the export trade data among the BRI countries and the intercountry input–output tables of the BRI region. The export trade data between countries in each year comes from the International Trade Center (https://www.trademap.org/). In terms of the intercountry

Table 1: The BRI countries indicated in this paper.

Subregion	Country
Central Asia	Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan, Turkmenistan
Mongolia and Russia	Mongolia, Russia
Southeast Asia	Vietnam, Laos, Cambodia, Thailand, Malaysia, Singapore, Indonesia, Brunei, Philippines, Myanmar
South Asia	India, Pakistan, Bangladesh, Afghanistan, Nepal, Bhutan, Sri Lanka, Maldives
Central and Eastern Europe	Poland, Czech Republic, Slovakia, Hungary, Slovenia, Croatia, Romania, Bulgaria, Serbia, Montenegro, Kingdom of Macedonia, Bosnia and Herzegovina, Albania, Estonia, Lithuania, Latvia, Ukraine, Belarus, Moldova
West Asia and Middle East	Turkey, Iran, Syria, Iraq, UAE, Saudi Arabia, Qatar, Bahrain, Kuwait, Lebanon, Oman, Yemen, Jordan, Israel, Armenia, Georgia, Azerbaijan, Egypt

Note. The BRI region usually contains 66 countries. Because the dataset we referred to, the Eora26 intercountry input-output table, does not include East Timor and Palestine. We excluded East Timor and Palestine from the 66 countries and got the research area with 64 countries.

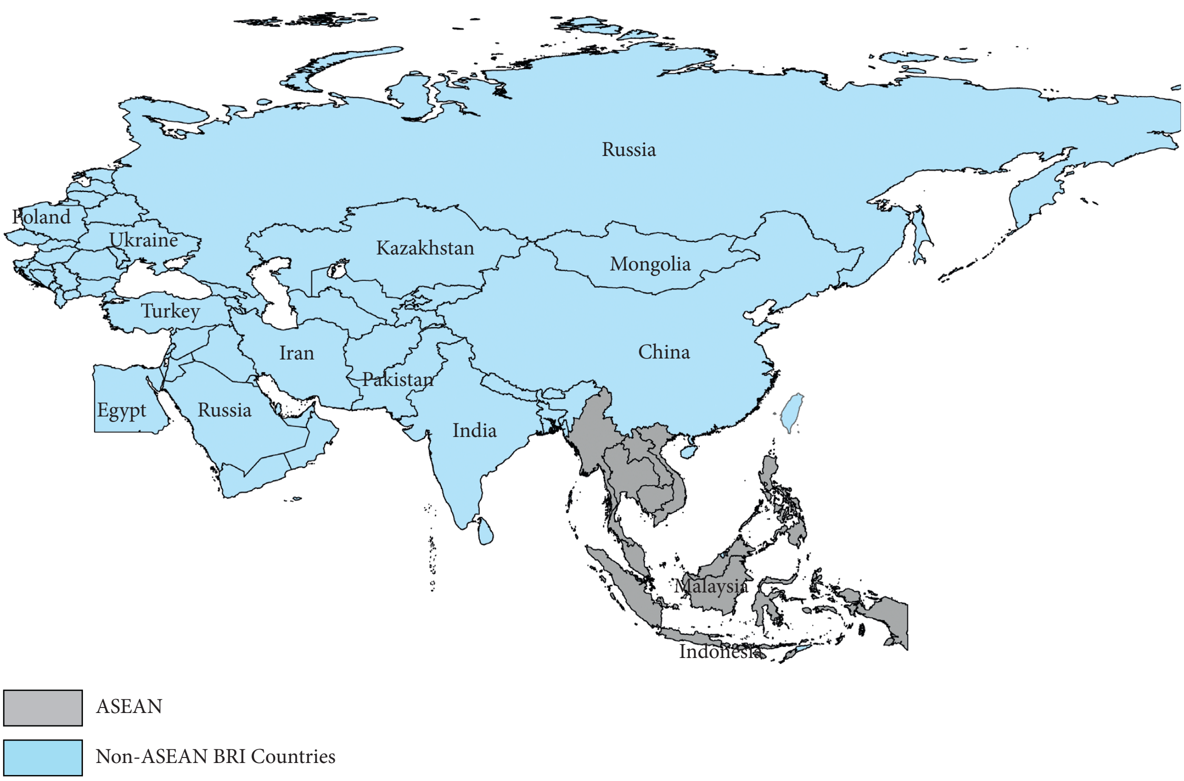


Figure 1: Map of the BRI countries in this paper.

input–output tables, there are a variety of databases compiled by many institutions, including the WIOD, GTAP, Eora26, Exiobase, and others. While most of the databases contain a small number of countries and regions, which is far from enough to cover the BRI countries, the Eora26 database compiled by the Lenzen team from the University of Sydney stands out as a relatively comprehensive one (https://worldmrio.com/eora26/). It covers 189 countries and regions around the world with a time span from 1990 to 2015. In addition, according to Qiao and others' [34] suggestion on data usage, we turn to adopt Eora26's intercountry input-output database. We extract the data of the 64 BRI countries in the years 2005, 2010, and 2015.

2.3. Methods

2.3.1. International Flow Accounting of VAT. International export trade products often contain inputs of imported raw materials from other countries. These raw material importing countries will also share export profits. To measure the true level of trade, it is necessary to calculate the actual value added in the exported products, i.e., the value-added exports. The multiregional input-output model is commonly used to calculate the value-added exports between countries. In the multiregional input-output model, the Leontief model is as follows:

$$X = (I - A)^{-1} f = Bf. \tag{1}$$

Assume that there are m countries and each country is a single sector. By introducing the diagonal matrix of value-added rate $\widehat{V}$, we can obtain

$$V = \widehat{V}X = \widehat{V}(I - A)^{-1}f = \widehat{V}Bf, \tag{2}$$

where $V = \begin{bmatrix} V^{11} & \cdots & V^{1m} \\ \vdots & \ddots & \vdots \\ V^{m1} & \cdots & V^{mm} \end{bmatrix}$, $\widehat{V} = \begin{bmatrix} v^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & v^m \end{bmatrix}$, $A = \begin{bmatrix} A^{11} & \cdots & A^{1m} \\ \vdots & \ddots & \vdots \\ A^{m1} & \cdots & A^{mm} \end{bmatrix}$, $B = (I - A)^{-1} = \begin{bmatrix} c^{11} & \cdots & c^{1m} \\ \vdots & \ddots & \vdots \\ c^{m1} & \cdots & c^{mm} \end{bmatrix}$, $f = \begin{bmatrix} f^{11} & \cdots & f^{1m} \\ \vdots & \ddots & \vdots \\ f^{m1} & \cdots & f^{mm} \end{bmatrix}$, X denotes the total output matrix driven by exports between countries, I denotes the identity matrix with all diagonals being 1, A denotes the direct consumption coefficient matrix, B denotes the Leontief inverse matrix, f denotes the export trade volume between countries, $\widehat{V}$ denotes the diagonal matrix of the value-added rate of each country, and V denotes the value-added export matrix between countries. By calculating the V matrix, we can obtain the specific flow of VAT between two countries.

2.3.2. Network Construction and Characteristic Analysis. (1) *Network Construction*. This paper selects data of the three time sections of 2005, 2010, and 2015 to construct the trade network and VAT network. In the trade network, each country is regarded as a node and the export of one country to another country is used as the weight of the edge from that node to another node. Similarly, in the VAT network, the value-added export of a country to another country is used as the weight of the edge from that node to another node.

(2) *Network Density*. Based on the premise that the trade network and VAT network are weighted networks, we first normalize the trade flows. The normalized trade flow can be represented as W_{ij}, which ranges between 0 and 1. The network density can be calculated as follows: $D = (\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij})/(N(N-1))$, $i \neq j$, for $\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}$ ranges from 0 to $N(N-1)$. Therefore, the density of weighted network will range between 0 and 1.

(3) *Community Detections*. Since there are different trade communities in the BRI trade network, we also identify the trade communities in both of the trade network and the VAT network. Comparative analysis of the patterns and characteristics of the trade communities will help to understand the characteristics of both networks more thoroughly. Community detection is a common method to identify communities in the network [16, 35]. In this paper, we employ the Q function method proposed by Newman [36] to detect trade communities. When Q reaches the maximum, we can obtain the best community division. The Q function is introduced as follows:

$$Q = \frac{1}{2M}\sum_{ij}\left[\left(a_{ij} - \frac{k_i k_j}{2M}\right)\delta(c_i, c_j)\right], \tag{3}$$

where

$$\delta(c_i, c_j) = \begin{cases} 1, & (c_i = c_j), \\ 0, & (c_i \neq c_j). \end{cases} \tag{4}$$

In formulas (3) and (4), c_i denotes the community to which node i in the network belongs, a_{ij} denotes the weight of node i to node j, i.e., the trade flow from country i to country j, k_i denotes the sum of weights of the edge from node i, and M denotes the sum of weights of all edges in the network.

(4) *Eigenvector Centrality*. Centrality measures the importance of the node in the network. For the trade network and VAT trade network, nearly all nodes are connected, but edges are weighted. Hence, the closeness centrality and betweenness centrality cannot well represent the importance of nodes. To solve this problem, in this paper, we use the eigenvector centrality index to measure the importance of nodes in the trade network and VAT network. The functional form of the eigenvector centrality is as follows:

$$EC(i) = x_i = c\sum_{j=1}^{n} a_{ij}x_j, \tag{5}$$

where x_i is the importance of node v_i, node v_j is the neighbor of v_i, and c is a proportional constant. Denoted by $x = [x_1, x_2, x_3, \cdots, x_n]^T$, when it reaches the steady state after several iterations, it can be written in the following matrix form:

$$x = cAx. \tag{6}$$

This means that x is the eigenvector corresponding to the eigenvalue c^{-1} of matrix A. The basic method to calculate the vector x is to give the initial value $x(0)$ and then use the following iterative algorithm:

$$x(t) = cAx(t-1), \quad t = 1, 2, \ldots, \tag{7}$$

until $x'(t) = x'(t-1)$. In the iterative process of each step, if x is divided by the principal eigenvalue λ corresponding to the adjacency matrix A, from this equation, we can obtain a nonzero solution, that is, $x = \lambda^{-1}Ax$. Thus, the constant $c = \lambda^{-1}$.

3. Results

3.1. Comparative Analysis of Network Structure. We use the Gephi0.92 Software to visualize the constructed trade network and VAT network. Since there are too many edges in the network to make clear the visualization, we select the top 1% edges of the trade flows in each network to simplify the visualization. The visualized structures of both networks in the three time sections of 2005, 2010, and 2015 are shown in Figure 2. We can see that the density of both of the trade network and the VAT network increased at each stage from 2005 to 2015, and the weight of the edge became significantly larger. Combined with the network density calculation in Table 2, it fully shows the trade ties among the BRI countries have become closer.

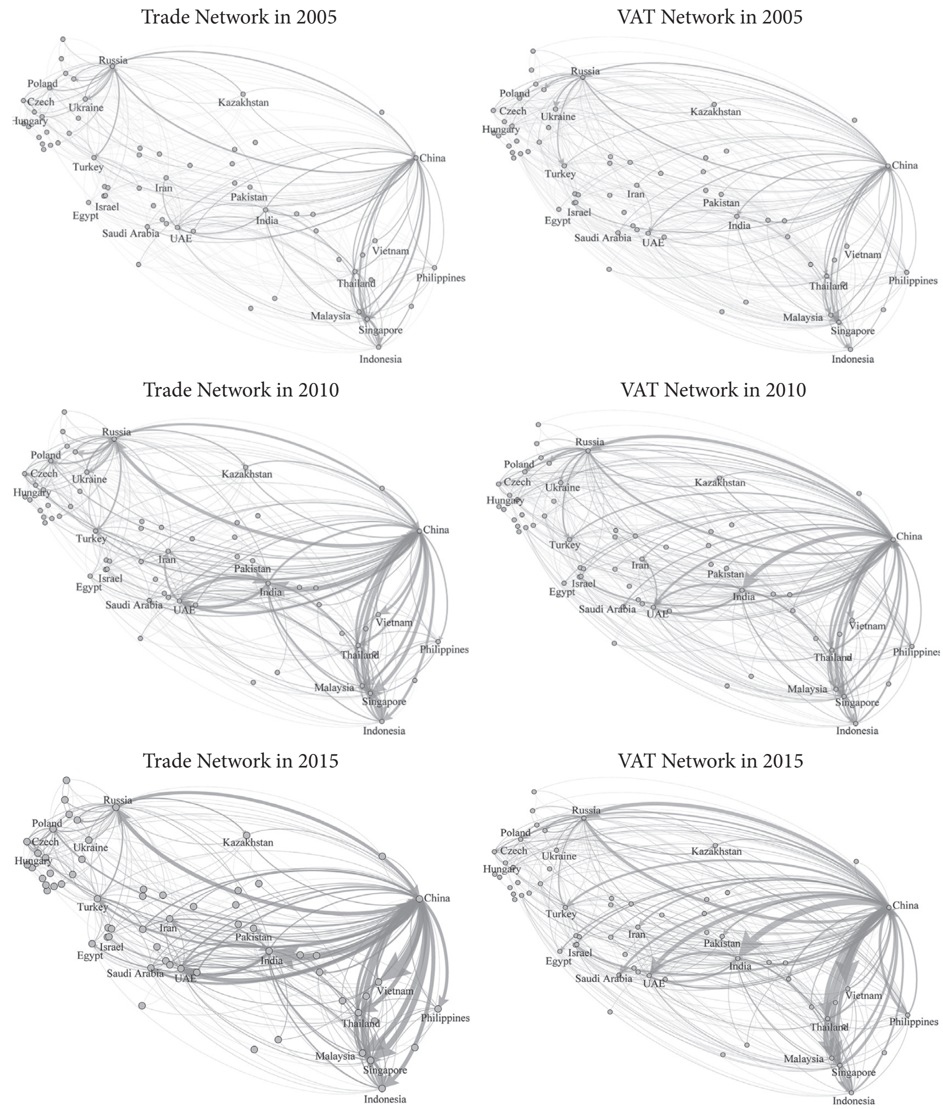


Figure 2: Network structure of the trade network and VAT network of the BRI countries.

Table 2: Network density of the trade network and the VAT network of the BRI countries.

Network density	2005	2010	2015
Trade network	0.184	0.385	0.474
VAT network	0.160	0.337	0.427

From a subregional perspective, we can tell the trade density of Southeast Asian countries is much greater than that of the other regions at various time sections. This is because the ASEAN countries maintain relatively larger economy, and they have signed free trade agreements. Moreover, the manufacturing industry of the ASEAN countries is more advanced, establishing a more closely connected production network and a denser VAT network. In addition, the trade network and VAT network of South Asia with India as the core are developing rapidly. The internal trade links of Central and Eastern Europe with Russia as the core is also becoming closer. From 2005 to 2015, the trade and VAT between China and the countries along the BRI have increased significantly, especially between China and countries in Southeast Asia.

Comparing the trade network with the VAT network, we find that the density of the VAT network is lower than that of the trade network. This is mainly because the value-added rate of export products is usually less than 1. The direction and volume of trade flows are also different. In the trade network, the two-way trade flows between China and the major BRI trading countries such as Southeast Asian countries, Russia, and India are obvious. In the VAT network, however, China's VAT outflow is more obvious. This shows that among the BRI countries, China is in a relatively high value-added position in the industrial chain, resulting in a greater VAT outflow than the inflow.

3.2. Comparative Analysis of Network Centrality

3.2.1. Evolution of Network Centrality. Network centrality is an important indicator to measure the status of nodes in the network. We calculate the centrality of the characteristic vector of both networks in 2005, 2010, and 2015 to show the status of trading countries in the BRI region. By selecting 10 major countries, we then analyze their ranking changes in the trade network and the VAT network. The result is shown in Figure 3.

Figure 3 shows that although China lagged behind Singapore in terms of trade network centrality and ranked second in 2005, it has always occupied the first position in the VAT network all through the years. Then, in 2010 and 2015, China's trade network centrality has also risen to the first place, which fully demonstrates China's core position in the BRI trade network and VAT network. Singapore, however, showed a downward trend of centrality in both trade network and VAT network from 2005 to 2015 and became the third in the rank of VAT network centrality. This is mainly due to the fact that most of Singapore's trade is re-export trade with relatively low value-added rate [37], which makes its participation index in the VAT network lower than that in the trade network.

From 2005 to 2015, countries with significant increases in the network centrality of both trade networks and VAT networks are India, Vietnam, and the United Arab Emirates (UAE). Among them, India's centrality in the VAT network rose from the sixth in 2005 to the third in 2010 and 2015. Vietnam shows the fastest rise in the rank, with its centrality rank in the VAT network growing rapidly from the twelfth in 2005 to the fourth in 2015. This shows that, in addition to rapid economic development, Vietnam also accelerates its integration into the regional production networks and regional value chains. The UAE, as an important country in West Asia, has a slight increase in position in the VAT network, from the eighth in 2005 to the sixth in 2015.

During the same period, the centrality of some countries in the networks has declined. Among them, Indonesia and Russia decreased the most. Indonesia's centrality in the trade network has dropped from the fourth to the eighth, and its centrality in the VAT network dropped from the sixth to the eighth. Because of the slow growth of domestic economic development, Russia's centrality in the value-added network slumped from the third in 2005 to the ninth place in 2015.

3.2.2. Distribution of Network Centrality. The distribution of network centrality reflects the polarization of the network cores, i.e., the variation in trading capacities of the BRI countries. Figure 4 indexes the evolution of centrality distribution of the trade network and the VAT network. It shows that the centrality distribution of either the trade network or VAT network in each specified year largely conforms to the power-law distribution; that is, the centrality of a few head countries are very large, and most other countries are at the edge of the network with low centrality. In detail, we can see that, in the trade network, from 2005 to 2010, although the relative central position of some countries has changed, the overall distribution of network centrality has not changed much. From 2010 to 2015, as China became the regional trading core country, the polarization has become distinctive with the large centrality gap between China and the other BRI countries, indicating China's expansion of its comparative advantage in the trade network. In the VAT network, the polarization of centrality distribution from 2005 to 2015 is more obvious. The centrality gap among the head countries began to expand as well. By comparing the trade network with the VAT network, we can tell a larger gap in centrality distribution of the VAT network than that of the trade network. China's central position in the VAT network is greater than that in the trade network.

Using Gephi 0.92, we further analyze the characteristics and evolution of the trade communities of trade network and VAT network of the BRI countries. The visualized result is shown in Figure 5.

In terms of the trade network, the division of trade communities mainly draws on geographical proximity to trade countries with strong centrality. In other words, the trade communities are shaped as certain large trade countries with neighboring small trade countries. Particularly, in the year of 2005, China and its surrounding Southeast Asian countries formed a main trade community. Russia led a trade community with the former Soviet Union countries. India formed a trade group with South Asia and some West Asian countries. Hungary, Turkey, and other Central and Eastern European countries constituted another trading community. Compared with 2005, the 2010 landscape of trade communities changed mainly because the trade community of China and Southeast Asia has divided into two separate communities, but the two trade communities of India and Middle East countries as well as South Asian countries have merged into one community. By the year 2015, China and ASEAN countries once again formed into one group, and the other trade communities have not changed much.

In terms of the VAT network, the community division also shows the characteristic of geographical proximity to large VAT countries. In 2005, most Southeast Asian countries formed a VAT community centering Singapore, while China formed another one with India and some West Asian countries. Russia still led a community with the former Soviet Union countries in the VAT network. In 2010, the VAT community of China and India has split up, with China turning to group with Southeast Asian countries and India forming a community with South Asian countries. Russia, together with some Central and Eastern European countries, formed a new VAT community. By 2015, the VAT community with China as the core has further expanded to include Russia, Mongolia, Central Asian countries, and Southeast Asian countries. India and South Asian countries remained as a separate VAT community.

Comparing the trade communities with the VAT communities, we find the overall patterns of the trade

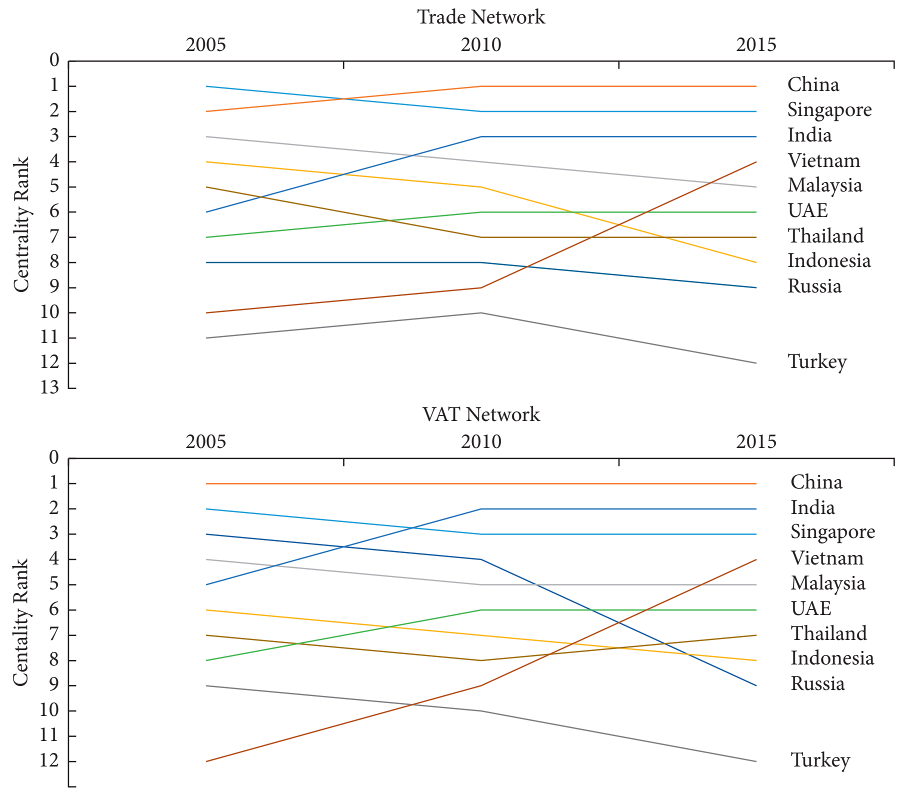


FIGURE 3: Evolution of network centrality of the trade network and the VAT network of major BRI countries.

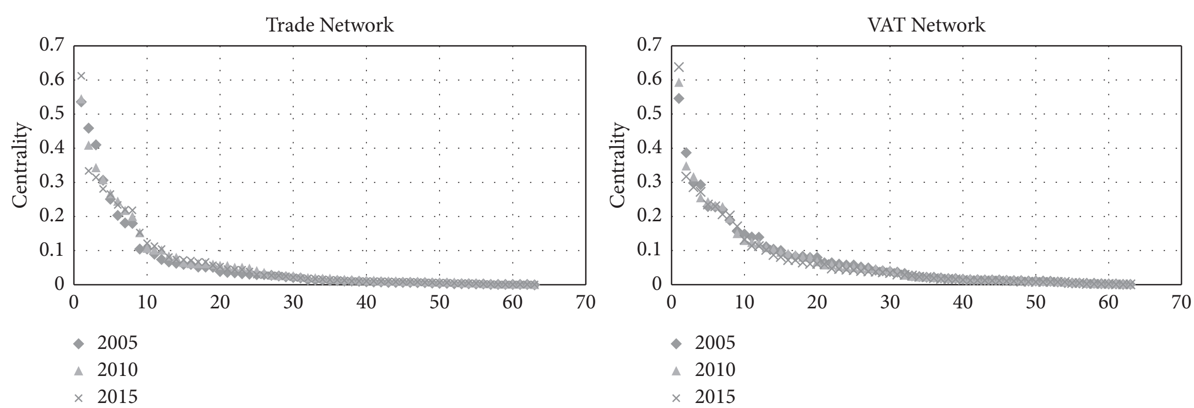


FIGURE 4: Distribution of network centrality of the trade network and the VAT network.

communities and the VAT communities represented differently all through the years. The overall pattern of the trade communities remains stable, while the overall patterns of the VAT communities changed a lot from 2005 to 2015. Specifically, in 2005, while China and Vietnam formed a trade community with other Southeast Asian countries, they were excluded from the VAT community of the Southeast Asian countries and were integrated into a separate one with countries in South Asia, West Asia, and Middle East. As West Asian and Middle East countries mainly relied on oil exports, we speculate China and Vietnam maintained quite amount of resource-based production. For Russia, it was involved in the same trade community with East European countries but formed a different VAT community without East European countries, indicating a weaker production cooperation than trade connection between Russia and the East Europe.

Then in 2010, China and Vietnam established stronger production connection with other Southeast Asian countries by forming an integrated VAT community but cut back trade connections with Singapore, Malaysia, Indonesia, and the others. Russia still maintained strong trade connection but weak production connection with the East Europe. Clearly, East European countries inclined to production cooperation with Central European countries. Such a VAT community composed by East European countries and Central European countries continued to the year 2015.

In 2015, the main difference between trade communities and VAT communities took place in Russia, the former Soviet Union countries, and Central Asian countries. These

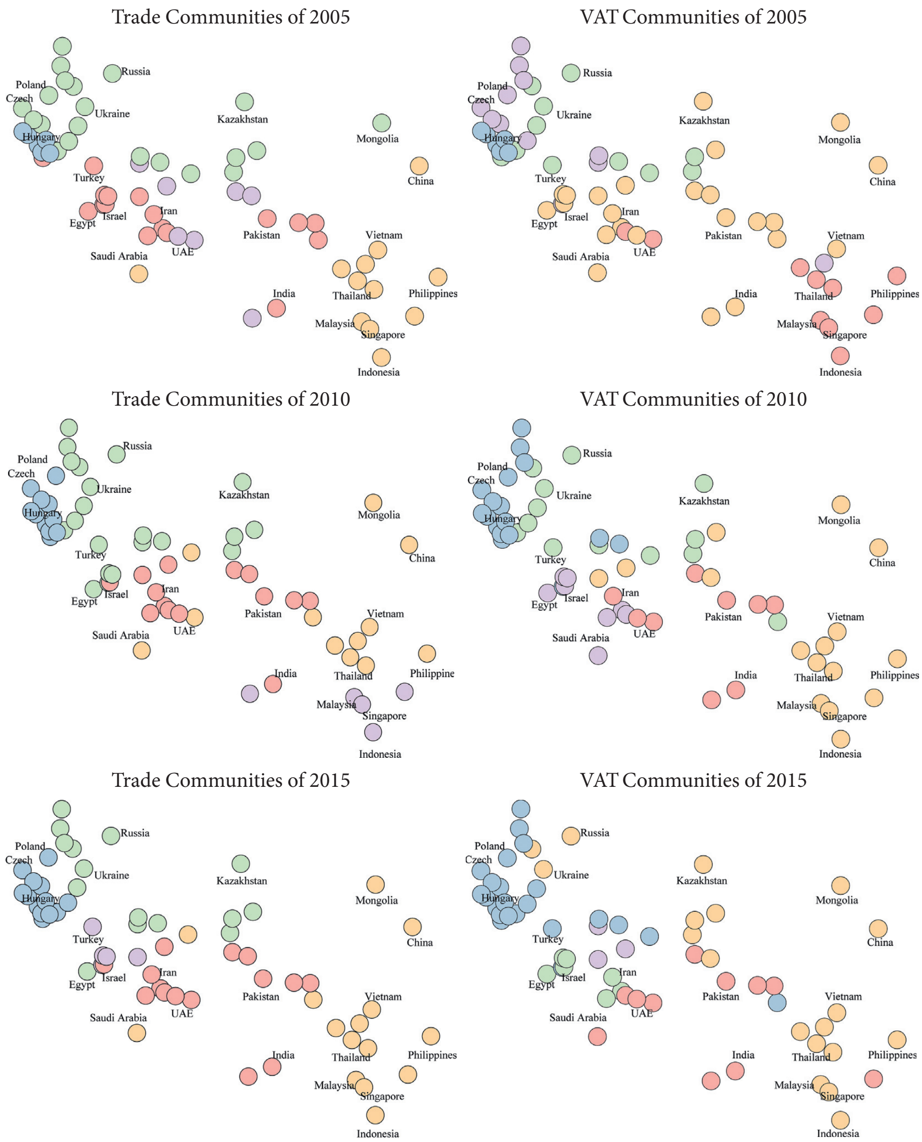


FIGURE 5: Comparison of trade communities and value-added trade communities of the BRI countries. *Note.* The spots indicate the geographical positions of national capitals. Spots in the same color belong to the same community.

countries formed an independent trade community but were separated into two different VAT communities: While Russia and Central Asian countries were involved in the same VAT community with China and Southeast Asian countries, the former Soviet Union countries joined the one with East and Central European countries. This result resonates with previous findings of Russia's decline in centrality and an increasingly obvious centrality polarization of the VAT network with China as the core all through the time period.

4. Factors Influencing Network Development

4.1. Econometric Model. Studies on international trade often include factors in aspects of economy, population, trade relations, and spatial distance to conduct empirical verifications of their impacts on trade network. Based on these studies, we try to investigate the factors that influence the development of BRI trade network and valued-added trade network from the following aspects. First, geographical distance, i.e., the spatial proximity between the trading countries, is an important factor, as we assume the closer the distance between countries, the easier the trade will happen. Economic distance also matters [38, 39]. In this paper, we take the difference in the GDP scale between the two countries as their economic distance. An excessive economic distance often indicates the two countries have large differences in resource endowments and production structure, which is not conducive to bilateral trade. Considering that only the economic distance cannot fully represent the impact of economy, we include the sum of the GDP of the two countries as a factor of the overall economic scale. Population size is another factor we take to reflect the market size of a country, which represents the demand for trade [39]. Moreover, we surmise countries that are bound to certain trade agreements will be more likely to trade with each other. As the BRI countries maintain different and independent trade agreements, such as The China-ASEAN Free Trade Area and the Russia-led EEU regional trade agreement [18], we also take the various trade agreements into the model. In addition, we understand common religion [40, 41] and language [42, 43] help reduce trade costs and facilitate trade. Taken together, we set up the model as follows:

$$\ln(\mathrm{NW}) = f\left[\ln\left(\mathrm{GDP}_{\mathrm{diff}}\right),\ \ln\left(\mathrm{GDP}_{\mathrm{sum}}\right),\ \ln\left(\mathrm{POP}_{\mathrm{diff}}\right),\ \ln\left(\mathrm{geo}_{\mathrm{dist}}\right), \mathrm{FTA}, \mathrm{relig}, \mathrm{lang}_{\mathrm{off}}\right], \tag{8}$$

where NW indicates the network in Quadratic Assignment Procedure (QAP) regression, $\mathrm{NW}_{\mathrm{trade}}$ indicates the trade network, and $\mathrm{NW}_{\mathrm{vat}}$ indicates the VAT network. They are the explained variables in the regression model.

For the explanatory variables, $\mathrm{GDP}_{\mathrm{diff}}$ denotes the absolute value of the GDP difference and $\mathrm{GDP}_{\mathrm{sum}}$ denotes the sum of GDP. The unit of these two variables is current US dollar. $\mathrm{POP}_{\mathrm{diff}}$ denotes the absolute value of the population difference. Its unit is person. $\mathrm{geo}_{\mathrm{dist}}$ denotes the matrix of geographic distance between the capital cities of the two countries. Its unit is kilometer. FTA denotes the matrix of a free trade agreement between the two countries (1 means there is a free trade agreement and 0 means there is not). relig denotes whether there is a common religion between the two countries. $\mathrm{lang}_{\mathrm{off}}$ denotes whether there is a common official language between the two countries.

4.2. Regression Results. Considering the high correlation of the network data, we draw on the QAP method for regression to avoid the influence from the multicollinearity of variables. The QAP regression method does not make strict requirements on the independence between variables. The purpose of QAP regression is to explore the regression relationship between multiple explanatory variable matrices and one explained variable matrix and to evaluate the regression results based on the regression coefficient and the regression fitted value R^2. Referring to the studies of Liu et al. [44] and Wang et al. [30], who have applied the QAP method in analyzing the BRI manufacturing network, we go further to investigate the factors that influence the trade network and the VAT network of the BRI countries. Due to the availability of data, we only perform the regression analysis for the 2015 data. The regression result is shown in Table 3.

For the trade network, the adjusted goodness of fit, i.e., the adjusted R^2, of the three models are 0.461, 0.460, and 0.455, respectively. This means the variables included in the model can explain 46.1%, 46.0%, and 45.5% of the corresponding network structure changes. The sample size of each model is 3906, indicating the model includes a 63×63 matrix composed of 63 countries (regions) along the BRI. The diagonal elements of the matrix are 0; that is, there is no self-loop in the network.

From the perspective of explanatory variables, the scale of GDP has obvious positive relationship with the trade links. Its coefficient is 4.092 in Model 1, significant at the significance level of 1%. The regression coefficient of economic distance is −0.929, significant at the significance level of 1%, which shows that the economic distance has a significant hindering effect on trade links. The smaller the difference in economic size between countries, the more likely it is to generate stronger trade ties. The regression coefficient of population difference is −0.206, significant at the significance level of 10%, which also shows that the difference in population size has a relatively negative impact on trade links. The regression coefficients of geographic distance in the three models are −1.431, −1.474, and −1.318, respectively, all significant at the 1% significance level. We can tell the greater the spatial distance between the two countries, the weaker the trade links become. The regression coefficient of the common religion is −1.634, and it is significant at the 5% level of significance. Instead of promoting trade flows, the common religion exerts hindering effect on trade connections. The coefficients of FTA in the three models are 3.534, 3.565, and 3.852, and they are all significant at the significance level of 1%, which fully demonstrates that the signing of FTA between the two countries has a very positive effect on promoting trade ties between the two countries. Although the regression coefficient of the common official language is positive, it is not significant at the 10% significance level. It means the common official language has no significant impact on the trade links among the BRI countries.

The adjusted R^2 of the three regression models for the VAT network is 0.369, 0.345, and 0.332, respectively, which means the variables could explain 36.9%, 34.5%, and 33.2% of the VAT network structural changes. From the perspective of regression coefficient, we could see the coefficient of the GDP scale is 2.167, with the significance level of 1%. It shows that the larger the economic size of the two countries are, the closer the VAT ties between the two countries will be. However, the coefficient of economic distance is significantly negative, which adds the meaning that the GDP gap between the two countries should be small to promote the VAT. The coefficient of the population difference is also significantly negative, indicating a closer VAT link between countries with similar population sizes than those with large population difference. The regression coefficient of the common religion is still significantly negative. That is, having a common religion has a certain hindering effect on the VAT network. We speculate that the BRI countries that have common religions may maintain similar industrial structures or at the same level of economic development, which reduces the need of VAT. The free trade agreements still have a very significant positive impact on promoting

Table 3: QAP regression result of the trade network and the VAT network in 2015.

Variables	NW_{trade}			NW_{vat}		
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
Intercept	−55.33***	−53.28***	−54.69***	−24.55***	−21.08**	−22.10***
$\ln(POP_{diff})$	−0.206*			−0.349***		
$\ln(GDP_{sum})$	4.092***	3.914***	3.914***	2.167***	1.865***	1.839***
$\ln(GDP_{diff})$	−0.929***	−0.949***	−0.955***	−0.481***	−0.515***	−0.506***
$\ln(geo_{dist})$	−1.431***	−1.474***	−1.318***	0.062	−0.011	0.139
relig	−1.634**	−1.660**		−1.182**	−1.226**	
FTA	3.534***	3.565***	3.852***	0.803***	0.856***	0.980***
$lang_{off}$	0.750	0.765		−0.321	−0.294	
R^2	0.461	0.460	0.455	0.370	0.346	0.333
Adjusted R^2	0.461	0.459	0.455	0.369	0.345	0.332
Number of observations	3906	3906	3906	3906	3906	3906

Note. ***, **, and *, respectively, indicate significance at 1%, 5%, and 10%.

VAT links. Yet the influence of the geographic distance and the common official language are not significant in the VAT model.

Comparing the regression results of the trade network and the VAT network, we find factors of economic size, economic distance, population difference, common religion, FTA, and common language have similar impacts on both of the trade network and the VAT network. On the one hand, enlarging economic size of the trading countries and facilitating FTA will positively promote the trade ties and VAT ties. On the other hand, economic distance, population difference, and common religion have hindering effects on both networks. Geographical distance has different impacts on the trade network and VAT network. While the geographic distance significantly prevents the trade links between countries, its hindering effect on VAT links is not significant.

5. Conclusions and Policy Suggestions

5.1. Conclusions. In this paper, we first calculate the VAT flows among the BRI countries from 2005 to 2015 to construct the BRI trade network and VAT network. Applying the network analysis method, we further analyze and visualize the structures, characteristics and evolutions of both trade network and VAT network. Then, we investigate into the factors that influence the development of both networks using the QAP method. By comparing the network density, network centrality, trade communities, and the influencing factors of the two networks through the past decade, we draw the following conclusions:

(1) The density of both trade network and VAT network of the BRI countries has been risen. As the density of regional trade network increased rapidly, the BRI countries have been communicating with each other more frequently. Meanwhile, the density of VAT network has also grown largely, reflecting a more closely related regional production network. In a sense, we could tell a pattern of the BRI regional economic integration is taking shape.

(2) The center-external structure of the BRI region has been strengthened. The network centrality of the head countries in the BRI trade network is exponentially larger than that of the other countries. Such centrality gap has been widening continuingly through the years. This reflects a distinct center-external structure of the BRI trade network led by the regional core country of China and several subregional core countries, like Singapore, India, UAE, and Russia. Comparing with the trade network, the center-external structure of regional VAT network is more obvious. With an increasingly growing centrality gap with the other countries, China occupies an absolute core position in the value-added trade network and leads the regional production network.

(3) The patterns of the trade communities and the VAT communities were mainly based on geographical proximity to regional and subregional core countries. They looked largely different, indicating a diverse regional trade and production network. While the economic ties within trade communities have not changed too much from 2005 to 2015, the VAT communities have been gradually drawn together in the shape of several large agglomerations, i.e., the community of China with surrounding countries, the community of the Europe, the South Asian countries, and Middle East countries, and the community of other countries. It is noteworthy that, in 2015, the VAT community with China as the core grew to be larger than any other VAT communities and larger than the trade community with China as the core. It shows that China maintains greater influence in leading the regional VAT network.

(4) The influencing factors have similar impacts on the trade network and the VAT network. Growing economic scale and FTA signing have obvious positive effects on the strengthening trade links and VAT links; economic distance and population difference have a significant inhibitory effect. The main difference is the effect of geographical distance. While it has obvious hindering effect on the trade network, it does not necessarily influence the VAT network.

5.2. Policy Implications. Unimpeded trade is one of the "five links" promoted by the BRI and the top priority of the BRI construction. Strengthening trade links among the BRI countries will facilitate regional economic development and promote the building of a community with a shared future for humankind. As the core country with the highest centrality in both the trade network and VAT network of the BRI region, China can play a more important role in promoting the trade ties among the BRI countries and leading regional socioeconomic development. Based on this research, we try to propose suggestions for a tighter, more deepening and more adequate regional economic cooperation pattern from the perspective of China:

First, China should further accelerate the transformation and upgrading of domestic industries and further enhance its position in the VAT network in the region. At the same time, China should promote the diffusion of industries to countries along the BRI and help promote their further integration into the global production network and global value chain to achieve common economic development.

Second, it is possible for China to advance the construction of broad trade channels within the BRI region. The broad trade channels will help reduce the trade cost between countries and promote intraregional trade links. They will also help regional industrial transfer and industrial connections and further enhance China's core position in the BRI trade network and VAT network.

Finally, China should help speed up the standardization of trade rules of the BRI region by promoting the signing of trade agreements among the BRI countries for reducing trade barriers. Meanwhile, with its central position in both the trade network and the VAT network, it is necessary for China to promote regional economic governance and facilitate intraregional trade links and production integration.

Data Availability

All data in this paper are available upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by Foundation of Natural Science of Guangdong (No. 2019BT02H594), Major Project of National Social Science Foundation of China (No. 21ZDA097), and Guangdong Ordinary Universities Youth Innovative Talents Project (No. 2020WQNCX016).

References

[1] R. C. Johnson and G. Noguera, "Accounting for intermediates: production sharing and trade in value added," *Journal of International Economics*, vol. 86, no. 2, pp. 224–236, 2012.

[2] C. Prell, L. Sun, K. Feng, and W. T. Myroniuk, "Inequalities in global trade: a cross-country comparison of trade network position, economic wealth, pollution and mortality," *PLoS One*, vol. 10, Article ID e144453, 2015.

[3] D. Garlaschelli and M. I. Loffredo, "Structure and evolution of the world trade network," *Physica A: Statistical Mechanics and Its Applications*, vol. 355, no. 1, pp. 138–144, 2005.

[4] L. De Benedictis and L. Tajoli, "The world trade network," *The World Economy*, vol. 34, no. 8, pp. 1417–1454, 2011.

[5] G. Fagiolo, J. Reyes, and S. Schiavo, "On the topological properties of the world trade web: a weighted network analysis," *Physica A: Statistical Mechanics and Its Applications*, vol. 387, no. 15, pp. 3868–3873, 2008.

[6] I. Tzekina, K. Danthi, and D. N. Rockmore, "Evolution of community structure in the world trade web," *The European Physical Journal B*, vol. 63, no. 4, pp. 541–545, 2008.

[7] G. Fagiolo, J. Reyes, and S. Schiavo, "The evolution of the world trade web: a weighted-network analysis," *Journal of Evolutionary Economics*, vol. 20, no. 4, pp. 479–514, 2010.

[8] X. Jiang, Y. Yang, and S. Wang, "Spatial and temporal patterns of evolution of global trade networks during 1985–2015 and its enlightenment to China's geostrategy," *Geographical Research*, vol. 37, pp. 495–511, 2018.

[9] Y. Yang, J. P. H. Poon, Y. Liu, and S. Bagchi-Sen, "Small and flat worlds: a complex network analysis of international trade in crude oil," *Energy*, vol. 93, pp. 534–543, 2015.

[10] X. Hao, H. An, X. Sun, and W. Zhong, "The import competition relationship and intensity in the international iron ore trade: from network perspective," *Resources Policy*, vol. 57, pp. 45–54, 2018.

[11] S. Torreggiani, G. Mangioni, M. J. Puma, and G. Fagiolo, "Identifying the community structure of the food-trade international multi-network," *Environmental Research Letters*, vol. 13, p. 54026, 2018.

[12] N. E. K. Petridis and E. Stiakakis, "Global e-waste trade network analysis," *Resources, Conservation and Recycling*, vol. 158, Article ID 104742, 2020.

[13] G. Dong, T. Qing, L. Tian, R. Du, and J. Li, "Optimization of crude oil trade structure: a complex network analysis," *Complexity*, vol. 2021, Article ID 3480546, 11 pages, 2021.

[14] M. H. Schütz and N. Palan, "Restructuring of the international clothing and textile trade network: the role of Italy and Portugal," *Journal of Economic Structure*, vol. 5, pp. 1–29, 2016.

[15] T. Kito and K. Ueda, "The implications of automobile parts supply network structures: a complex network approach," *CIRP Annals*, vol. 63, no. 1, pp. 393–396, 2014.

[16] Z. Liu, T. Wang, J. W. Sonn, and W. Chen, "The structure and evolution of trade relations between countries along the Belt and Road," *Journal of Geographical Sciences*, vol. 28, no. 9, pp. 1233–1248, 2018.

[17] J. Zou and W. Liu, "Trade network of China and countries along"Belt and Road Initiative"areas from 2001 to 2013," *Scientica Geographica Sinica*, vol. 36, pp. 1629–1636, 2016.

[18] Z. Chong, C. Qin, and S. Pan, "The evolution of the Belt and Road trade network and its determinant factors," *Emerging Markets Finance and Trade*, vol. 55, no. 14, pp. 3166–3177, 2019.

[19] Q. Chen, J. Cheng, and Z. Wu, "Evolution of the cultural trade network in "the belt and Road" region: implication for global cultural sustainability," *Sustainability*, vol. 11, no. 10, p. 2744, 2019.

[20] Z. Song, S. Che, and Y. Yang, "The trade network of the Belt and Road Initiative and its topological relationship to the global trade network," *Journal of Geographical Sciences*, vol. 28, no. 9, pp. 1249–1262, 2018.

[21] Z. Song, S. Che, and Y. Yang, "Topological relationship between trade network in the Belt and Road Initiative area and

global trade network," *Progress in Geography*, vol. 36, pp. 1340–1348, 2017.
[22] H. Xu, T. Sun, and L. Cheng, "Trade patterns and influence factors of high-end manufacturing on "One Belt and One Road"—a study based on the exponential random graph models," *Finance & Trade Economics*, vol. 36, pp. 74–88, 2015.
[23] M. Zhan, "Research on the competitiveness and complementarities of agricultural product trade of countries along "the belt and Road": based on social network analysis method," *Issues in Agricultural Economy*, vol. 2, pp. 103–114, 2018.
[24] C. Zhang, J. Fu, and Z. Pu, "A study of the petroleum trade network of countries along "The Belt and Road Initiative"," *Journal of Cleaner Production* vol. 222, pp. 593–605, 2019.
[25] B. Ferrarini, "Vertical trade maps," *Asian Economic Journal*, vol. 27, no. 2, pp. 105–123, 2013.
[26] F. Cerina, Z Zhu, A. Chessa, and M. Riccaboni, "World input-output network," *PLoS One*, vol. 10, Article ID e0134025, 2015.
[27] Z. Zhu, M. Puliga, F. Cerina, A. Chessa, and M. Riccaboni, "Global value trees," *PLoS One*, vol. 10, Article ID e0126699, 2015.
[28] J. Amador and S. Cabral, "Networks of value-added trade," *The World Economy*, vol. 40, no. 7, pp. 1291–1313, 2017.
[29] T. Sun, H. Xiao, B. Meng, and H. Xu, "Topology and evolutionary pattern of manufacturing gvc network and its influencing factors: a study based on wwz method and social networks," *Management Revue*, vol. 30, pp. 49–60, 2018.
[30] B. Wang, N. Chen, and G. Lin, "Network structure of value-added trade in manufacturing of "Belt and Rood" countries and its influencing factors," *Journal of International Trade*, vol. 3, pp. 85–100, 2019.
[31] F. de Soyres, F. de Soyres, and A. Gaillard, "Value added and productivity linkages across countries," *FRB International Finance Discussion Paper, No. 1266*, Board of Governors of the Federal Reserve System, Washington, DC, USA, 2019.
[32] W. Liu, "Scientific understanding of the Belt and Road Initiative of China and related research themes," *Progress in Geography*, vol. 34, pp. 538–544, 2015.
[33] J. Zou, C. Liu, G. Yin, and Z. Tang, "Spatial patterns and economic effects of China's trade with countries along the Belt and Road," *Progress in Geography*, vol. 34, pp. 598–605, 2015.
[34] X. Qiao, Z. Li, and N. Xiang, "Tracing the embodied carbon in intermediate goods trade and a parallel comparison among multi-regional databases," *Finance & Trade Economics*, vol. 39, pp. 84–100, 2018.
[35] M. T. Schaub, J.-C. Delvenne, M. Rosvall, and R. Lambiotte, "The many facets of community detection in complex networks," *Applied Network Science*, vol. 2, no. 1, p. 4, 2017.
[36] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 26113, 2004.
[37] K. M. Choy, "Trade cycles in a re-export economy: the case of Singapore," *International Economic Journal*, vol. 26, no. 2, pp. 189–201, 2012.
[38] F. Kimura and H.-H. Lee, "The gravity equation in international trade in services," *Review of World Economics*, vol. 142, no. 1, pp. 92–121, 2006.
[39] X. Yao, B. Wang, and Y. Pu, "The network structure characteristics and influence factors of the service intermediate input about the countries along "the Belt and Road"," *World Economy Study* vol. 1, pp. 122–133+136, 2018.
[40] H. L. F. De Groot, G.-J. Linders, P. Rietveld, and U. Subramanian, "The institutional determinants of bilateral trade patterns," *Kyklos*, vol. 57, no. 1, pp. 103–123, 2004.
[41] J. J. Lewer and H. Van den Berg, "Religion and international trade: does the sharing of a religious culture facilitate the formation of trade networks?" *The American Journal of Economics and Sociology*, vol. 66, no. 4, pp. 765–794, 2007.
[42] J. Melitz, "Language and foreign trade," *European Economic Review*, vol. 52, no. 4, pp. 667–699, 2008.
[43] J. Lohmann, "Do language barriers affect trade?" *Economics Letters*, vol. 110, no. 2, pp. 159–162, 2011.
[44] H. Liu, Y. Zhang, and Y. Sun, "The spatial network structure of China's regional development and its determinants," *Economic Review*, vol. 5, pp. 59–69, 2015.