International Journal of Geo-Information

MDPI

Article

Modeling Population Mobility Flows: A Hybrid Approach Integrating a Gravity Model and Machine Learning

Jingjing Liu [1,2,3], Lei Xu [3], Le Ma [2] and Nengcheng Chen [1,2,3,*]

[1] Hubei Luojia Laboratory, Wuhan University, Wuhan 430079, China; liujingjing@whu.edu.cn
[2] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[3] National Engineering Research Center for Geographic Information System, School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China
* Correspondence: cnc@whu.edu.cn

Abstract: Population mobility between cities significantly affects traffic congestion, disease spread, and societal well-being. As globalization and urbanization accelerate, understanding the dynamics of population mobility becomes increasingly important. Traditional population migration models reveal the factors influencing migration, while machine learning methods provide effective tools for creating data-driven models to handle the nonlinear relationships between origin and destination characteristics and migration. To deepen the understanding of population mobility issues, this study presents GraviGBM, an expandable population mobility simulation model that combines the gravity model with machine learning, significantly enhancing simulation accuracy. By employing SHAPs (SHapley Additive exPlanations), we interpret the modeling results and explore the relationship between urban characteristics and population migration. Additionally, this study includes a case analysis of COVID-19, extending the model's application during public health emergencies and evaluating the contribution of model variables in this context. The results show that GraviGBM performs exceptionally well in simulating inter-city population migration, with an RMSE of 4.28, far lower than the RMSE of the gravity model (45.32). This research indicates that distance emerged as the primary factor affecting mobility before the pandemic, with economic factors and population also playing significant roles. During the pandemic, distance remained dominant, but the significance of short distances gained importance. Pandemic-related indicators became prominent, while economics, population density, and transportation substantially lost their influence. A city-to-city flow analysis shows that when population sizes are comparable, economic factors prevail, but when economic profiles match, living conditions dictate migration. During the pandemic, residents from hard-hit areas moved to more distant cities, seeking normalcy. This research offers a comprehensive perspective on population mobility, yielding valuable insights for future urban planning, pandemic response, and decision-making processes.

Keywords: urban mobility; COVID-19; geospatial data; data-driven algorithms; lightGBM

check for updates

Citation: Liu, J.; Xu, L.; Ma, L.; Chen, N. Modeling Population Mobility Flows: A Hybrid Approach Integrating a Gravity Model and Machine Learning. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 379. https://doi.org/10.3390/ijgi13110379

Academic Editors: Wolfgang Kainz, Hartwig H. Hochmair, Hao Li and Levente Juhász

Received: 17 June 2024
Revised: 10 October 2024
Accepted: 28 October 2024
Published: 30 October 2024

Copyright: © 2024 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As the region where the majority of the world's population resides, cities exhibit high levels of population and economic activity, influencing the direction of people's mobility, modes of interaction, and the nature of social and entertainment activities [1–5]. People gather in cities, creating a wide range of social and economic interactions while also migrating between cities, which facilitates the dissemination of knowledge and creativity [6–9]. In recent years, research on human mobility has garnered extensive attention and interest [10–12], covering various topics. These topics include the exploration of population mobility patterns within cities [13–15], rural-to-urban migration [3,16], international migration influenced by natural and political-economic factors [17–20], predictions of traffic

and human flow [21–23], and the bidirectional relationship between virus and population mobility in the context of COVID-19 [10,24–26]. All of these domains benefit from theories and methods that model population mobility. Population movement, transportation issues, and virus transmission are fundamentally processes of mobility, where the quantities of interest, such as populations, modes of transportation, and pathogens, change in a spatial distribution. Therefore, a profound understanding and accurate description of spatial mobility characteristics contribute to our understanding of urban resilience, sustainability, and safety, ultimately promoting the continued development of urban spaces [27,28].

Population mobility models elucidate and expound on migration across regions, urban areas, or nations. Researchers use mathematical and statistical methods to understand why people migrate, what influences migration, and its socio-economic impacts on both origin and destination in traditional models [11,29]. Three major theoretical frameworks emerged in the 20th century. The gravity model, introduced by Alexander von Humboldt in the early 19th century, suggests that migration (M) is proportional to the populations (P_1 and P_2) of two regions and inversely proportional to the distance (D) between them, expressed as $M = \frac{k(P_1 \times P_2)}{D^2}$, where k is a scaling constant [30–32]. The radiation model, proposed by Brian J.L. Berry in the 1950s, explains urban population movement, emphasizing the attractiveness of urban centers and inter-city distances [33,34]. The intervening opportunities model, advanced by Erving Goffman in the 1950s, focuses on individual decision-making, suggesting that people may change their migration plans based on encountered opportunities [35].

Traditional population migration models, such as the gravity model, describe the general patterns of population migration, but their representations are constrained by linear assumptions. Advancements in artificial intelligence (AI) have introduced data-driven methods like machine learning and neural networks into population migration research, which can handle complex nonlinear relationships and high-dimensional data. Considering population mobility as a spatial interaction between origin supply and destination demand, these methods offer higher predictive accuracy [36]. Robinson and Dilkina [37] used machine learning to predict migration trends, addressing data imbalance and hyperparameter optimization. Azizi and Yektansani [38] compared eight models for illegal immigration prediction. Carammia et al. [39] combined traditional data with social media for refugee predictions. Giang et al. [40] found backpropagation neural networks most accurate for labor migration. Best et al. [41] used random forests for migration decisions in Bangladesh. Deep learning models, especially LSTM (Long Short-Term Memory) and GNN (Graph Neural Network), have shown strong capabilities in handling time series and spatial data, as demonstrated by Golenvaux [42] and Terroso-Sáenz [43]. Wang and Chen [44] developed a dual gravity model for predicting passenger flows. Golenvaux et al. used long short-term memory networks (LSTM) with Google Trends data to predict international migration, showing that the LSTM approach significantly outperformed standard artificial neural networks and traditional gravity models [42]. Terroso-Sáenz and Muñoz employed graph neural networks (GNN) to account for potential relationships between large geographic areas, validating their approach with nationwide mobility data and current weather conditions in Spain [43]. Their model combined graph convolutional layers and LSTM layers to achieve highly accurate mobility flow predictions. Simini proposed a Deep Gravity model by integrating deep learning with traditional gravity models, which proved to be advantageous in capturing complex mobility patterns and enhancing predictive accuracy [15]. Wang and Chen designed a dual gravity model to describe and predict railway and highway passenger flows, enabling predictions even with incomplete data [44].

LightGBM is an efficient gradient-boosting decision tree method widely used in regression and classification tasks, being particularly suitable for large-scale datasets [45]. Although lightGBM has not yet been applied to modeling inter-city population migration, its advantages in handling nonlinear relationships and capturing feature interactions have been well established [29,46]. In this study, we employ lightGBM to simulate inter-city

population migration, aiming to construct a nonlinear framework that elucidates the complex relationship between urban characteristics and population migration.

AI-driven methods have significant advantages in handling large-scale data, capturing complex nonlinear relationships and improving predictive capabilities, whereas traditional models offer theoretical explanatory power and stability [47]. Integrating multi-source data and various modeling approaches can more comprehensively and accurately reflect the dynamic changes in population migration. We devised a GraviGBM model which builds upon the gravity model and incorporates multi-source geographical data as inputs. This model takes into account geographical variables from five dimensions: land-use areas, public facilities, economic development, population, and epidemic data. We trained a lightGBM model using this comprehensive dataset. The GraviGBM model, once trained, was utilized to estimate population flows between cities in China. Analyzing the decision-making process of the model provides a systematic explanation for the factors driving population mobility. This machine learning integration with the gravity model allows for the direct simulation of population movements between cities without being constrained by coarse grid limitations. This model's structure offers a higher spatial resolution in its simulation capabilities compared to the Deep Gravity model [15]. The GraviGBM model offers enhanced geographical interpretability and is particularly well suited for specific applications in urban planning. Its utility extends to the direct formulation of policies in areas such as transportation, housing, and infrastructure, making it a valuable tool for decision-makers in these domains.

The outbreak of COVID-19 represents a significant global public health event that profoundly impacted people's travel patterns [10,26,48]. Many researchers have dedicated their efforts to investigating how COVID-19 has influenced travel behavior and patterns. This unexpected event offers a unique window for studying the changes in population mobility patterns induced by major public health crises [49–52]. In our model, we incorporate COVID-19-related indicators to extend the application of the model in the context of such major health emergencies, exploring the changes in migration patterns during the pandemic.

To gain deeper insights into the decision-making processes behind travel and the specific reasons for predictions, we employed the SHAP value method for further interpretation of the GraviGBM model. This method allows us to reveal the key variables influencing overall mobility, as well as the geographic characteristics affecting individual flow samples between specific origins and destinations [53–56]. Additionally, the application of SHAPs provides a systematic perspective for understanding the broader impacts of health crises on social and economic systems by elucidating the structural changes in population mobility models before and after the pandemic.

Addressing the systematic modeling of collective population mobility generation and providing a comprehensive explanation of the underlying reasons, this study aims to conduct the following:

(1) Development of the GraviGBM Model: Construct an innovative GraviGBM model to simulate intercity population migration, providing an extensible framework for analyzing population flows. (2) Variable analysis based on SHAP values: Utilize SHAP methods to explore the relationship between urban characteristics and population migration, revealing the key variables influencing migration decisions. (3) Impact factor changes under COVID-19: Extend the GraviGBM model to analyze the effects of the COVID-19 pandemic on population migration factors, assessing changes in the contribution of model variables during public health emergencies. The research outcomes contribute to a deeper understanding of population mobility behavior, assisting governments and urban planners in improving urban planning and resource allocation, optimizing transportation networks, and enhancing urban sustainability. Additionally, they help to identify behavioral changes during health crises and the effectiveness of response strategies. This knowledge can be applied in future health events, natural disasters, emergencies, and crisis situations to ensure the equitable distribution of resources.

2. Data and Methods

The methodological framework of this study can be roughly divided into two parts (Figure 1). The first part involves constructing a model for population flows between Chinese cities using GraviGBM. City geographic features, distances between cities, and epidemic indicators are used as input features for the model. The gravity model and Single-Feature GraviGBM serve as baseline models to evaluate the performance of the GraviGBM model. The second part of the study builds upon the model constructed in the first part and utilizes Shapley Additive Explanations (SHAP) to analyze the factors influencing population flows between cities before and after the COVID-19 pandemic.

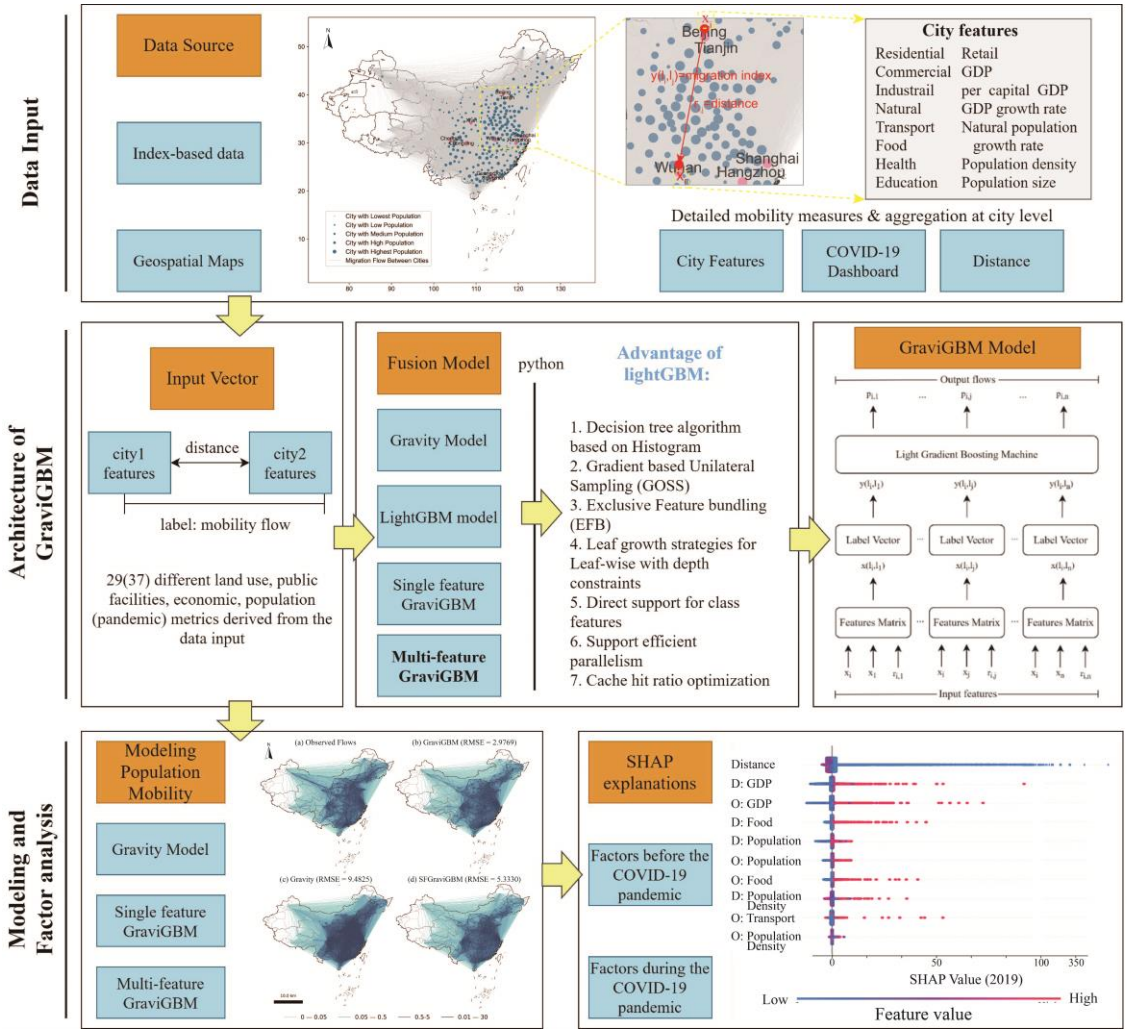


Figure 1. The flowchart of the analytical framework.

2.1. Data Sources

2.1.1. Mobility Flows

The human mobility data utilized in this study is publicly accessible via the Gaode Transportation Big Data Platform [57]. This dataset leverages individuals' location information to capture the migration trajectory of the population, providing valuable insights into urban population dynamics, urban infrastructure, crowd behavior, and other relevant factors. The Gaode mobility estimates are derived from an extensive volume of data, including over 120 billion location-based service (LBS) inquiries daily from more than 1.1 billion mobile devices, and encompassing more than 1.5 billion points of interest (POI).

While the Gaode mobility data does not directly disclose specific population flow figures, it offers two relative indicators: the migration scale index and the in/out migration

index. The migration scale index represents the overall intensity of population movement for a given city and reflects the total volume of in- and out-migration for that city within the dataset. It is a single value calculated daily for each city, offering a comparative view of migration levels across multiple cities. On the other hand, the in/out migration index measures the direction and magnitude of population movement between city pairs, reflecting the net flow of people entering or leaving a city. Depending on whether the focus is on total migration or migration flows between specific city pairs, the migration scale index and the in/out migration index reflect the migration volume at the city level or between city pairs, respectively. Although the specific algorithm used to compute these indexes has not been disclosed by Gaode, the indexes are known to correlate positively with actual migration volumes. The collection window for the in/out migration index is set at 8 h, allowing for the determination of origin–destination flows between pairs of prefectures. This index has been utilized to construct a city relation matrix, wherein the magnitude of the inter-city migration index reflects the intensity of population movement.

This study utilizes the in/out migration indexes from 2019 and 2020, covering all prefecture-level and higher-level cities across China, excluding Hong Kong, Macau, and Taiwan. Prior to data utilization, a meticulous examination was conducted, leading to the exclusion of cities with missing data. The final dataset comprises inter-city mobility information for 295 city pairs.

2.1.2. City Features

This study selected 15 indicators as urban geographic features, derived from OpenStreetMap (OSM) [58] and statistical yearbooks. Eight features (residential, commercial, industrial, natural areas and number of transport, food, health, education, and retail facilities), encompassing the origin and destination characteristics of population migration flows, were computed using OpenStreetMap (OSM) data, while the remaining six features (GDP, per capita GDP, GDP growth rate, natural population growth rate, population density, and population size) were obtained from statistical yearbooks [59]. All geographic features used in the following sections of this study are derived from the sources mentioned above.

2.1.3. COVID-19 Dashboard

The data of daily confirmed incidence comes from the COVID-19 dashboard provided by DXY.cn, which provides near-real-time data collected from government press releases [60]. We selected four key COVID-19 indicators: confirmed counts, suspected counts, death counts, and cured counts. The specific definitions of these indicators are provided in Table 2.

Furthermore, we opted to use the Stringency index from the Oxford COVID-19 Government Response Tracker (OxCGRT) as a policy factor to gauge the severity of travel restrictions or social distancing measures in different regions [61]. This index provides a measure of the government's responsiveness across a range of indicators, encompassing containment and closure policies, which are sometimes referred to as lockdown policies.

To assess the potential multicollinearity among the selected indicators—confirmed counts, suspected counts, death counts, cured counts, and the Stringency index—we conducted a multicollinearity check. The results indicate that both suspected counts and cured counts exhibited multicollinearity issues, while confirmed counts, suspected counts, and the Stringency index passed the multicollinearity check, with Variance Inflation Factors (VIFs) below 5, specifically: confirmed counts (VIF = 1.0139), suspected counts (VIF = 1.0134), and the Stringency index (VIF = 1.0274).

2.2. Research Area

This study focuses on modeling intercity population migration, selecting 295 major prefecture-level cities in mainland China as research cases. The dataset used in the model consists of the characteristics of these 295 cities and the migration indices between each pair of cities. The geographical distribution of the case cities and examples of intercity

migration are shown in Figure 2. Different color dots are used solely to visually distinguish key cities, with no data-related significance attributed to color.

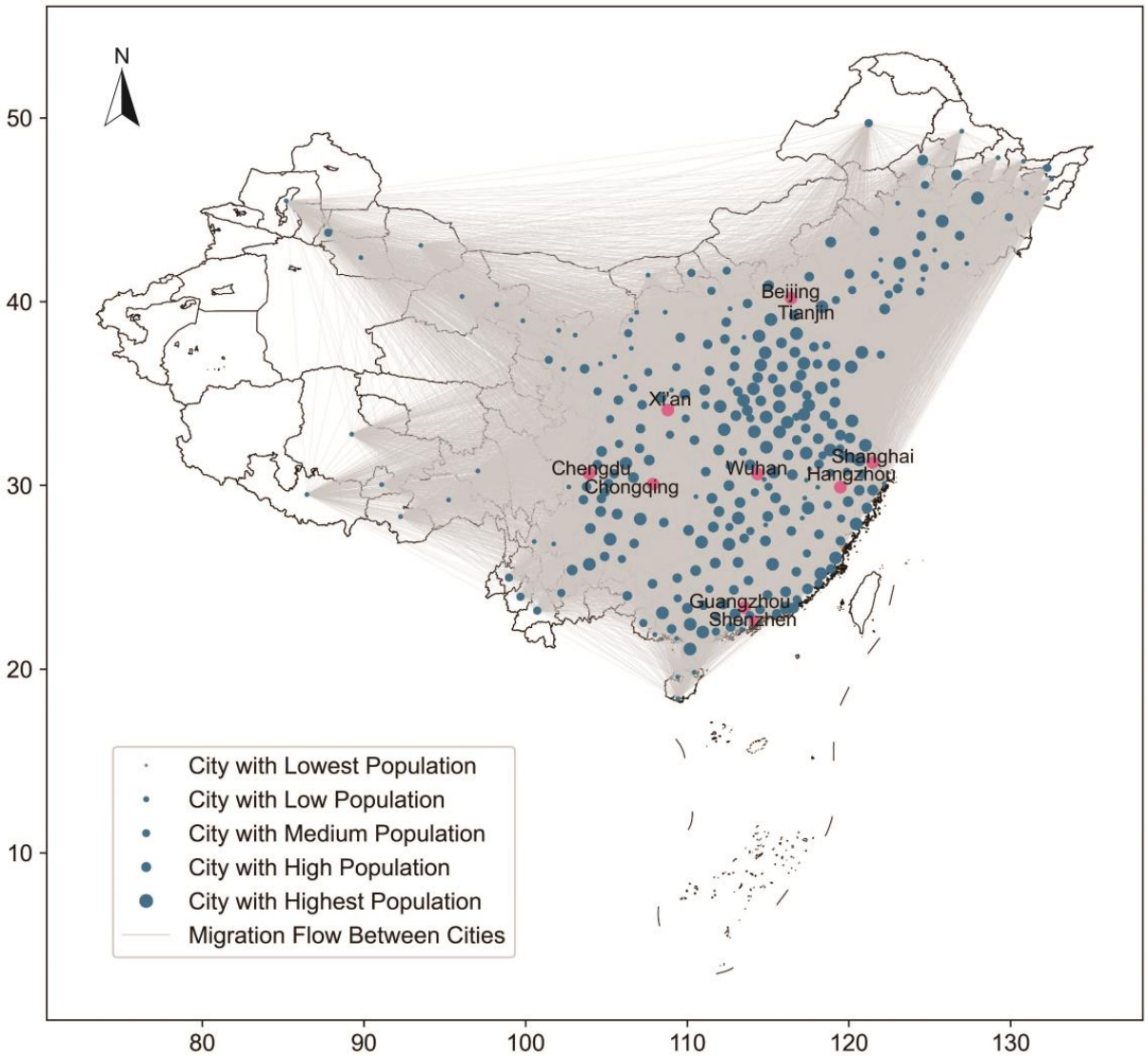


Figure 2. Research area.

2.3. Light Gradient-Boosting Machine

LightGBM is an advanced algorithm developed within the Gradient-Boosting Decision Tree (GBDT) framework, first introduced by Microsoft Research Asia in 2017 [62]. It incorporates several optimizations over traditional GBDT methods, including a histogram-based decision tree algorithm, Gradient-based One-Side Sampling (GOSS), Exclusive Feature Bundling (EFB), and a leaf-wise tree growth strategy with depth constraints [63]. The histogram-based approach discretizes continuous floating-point feature values into K integer values, constructing a histogram that facilitates efficient statistical aggregation during data traversal. Following this process, the optimal split point is determined based on the aggregated statistical metrics [64]. Furthermore, lightGBM's leaf-wise optimization selects the leaf node with the maximum gain for splitting (Figure 3), iteratively enhancing accuracy compared to the conventional level-wise tree growth strategy found in many GBDT algorithms [65]. To mitigate overfitting and prevent excessively deep trees, lightGBM incorporates a maximum depth restriction, ensuring both efficiency and robustness in handling large-scale distributed datasets.

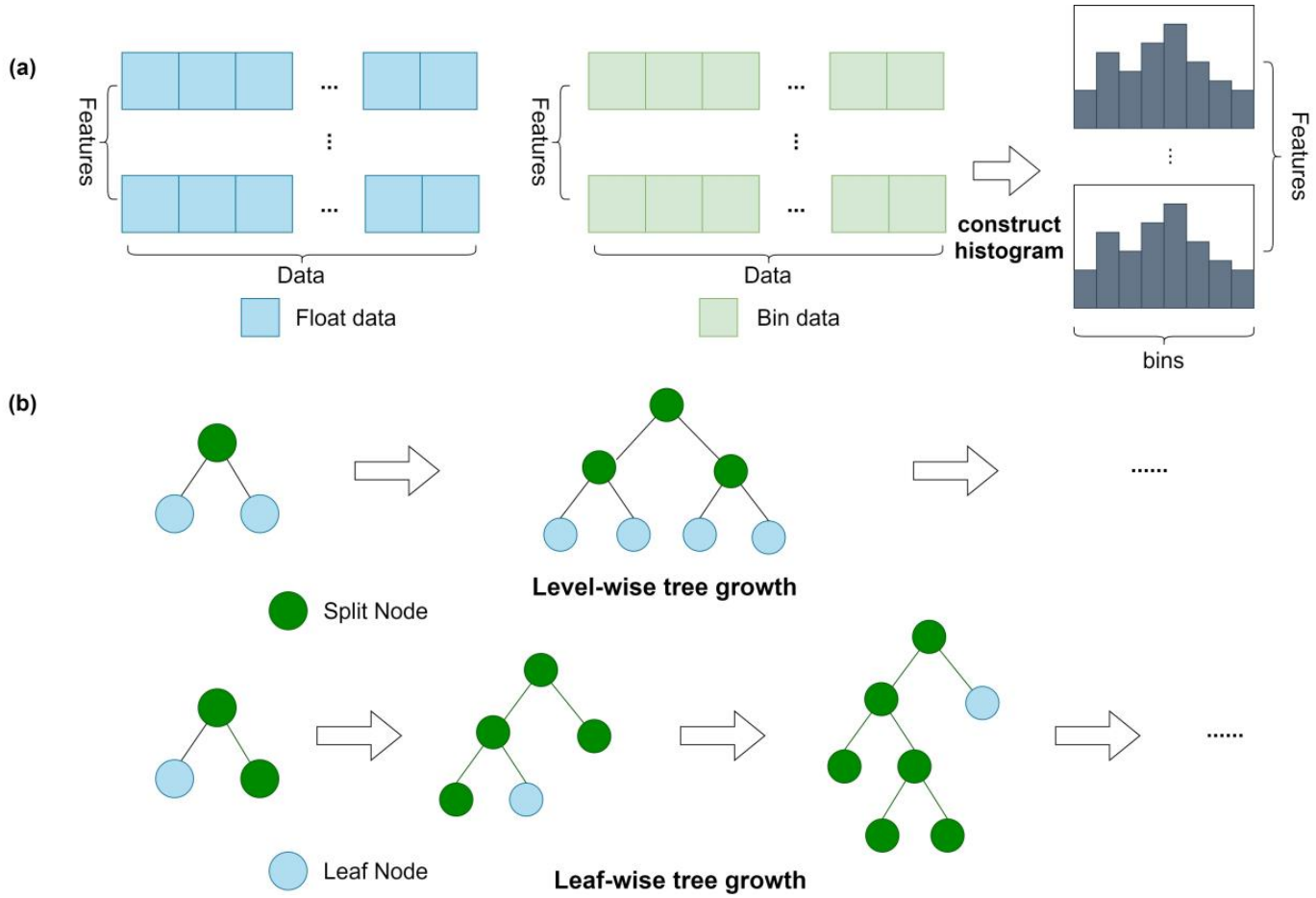


Figure 3. Principles of the lightGBM Algorithm. (**a**) Histogram-based decision tree algorithm. (**b**) Schematic diagram of level-and leaf-wise strategy tree growth.

2.4. Derivation of GraviGBM

GraviGBM is derived from a state-of-the-art model of flow generation known as the gravity model [15]. Building upon the fundamental principles of the gravity model, we employ the Light Gradient-Boosting Machine (lightGBM) model to better capture the nonlinear relationship between population, distance, and migration, as well as to incorporate additional geographic features to naturally extend the gravity model.

The gravity model serves as the theoretical foundation for GraviGBM. In fact, the gravity model is considered to be the state-of-the-art model for generating flows [11,66,67] and is preferred over a null model where flows are randomly distributed across the edges of the mobility network. Assuming a certain type of relationship, these ideas can be extended in a broader context:

$$T_{ij} = Km_i m_j f(r_{ij}) \tag{1}$$

where T_{ij} represents the expected number of trips from location i to location j. K is a constant; the masses m_i and m_j correspond to the number of trips leaving location i or attracted to location j. The function $f(r_{ij})$, referred to as a "friction factor" or "deterrence function", is a decreasing function of the distance r_{ij}.

In the original gravity model and its subsequent improved versions, masses are typically functions of population, but this relationship is not necessarily linear. As the gravity model has evolved, other factors such as economic indicators represented by per capita GDP have also been considered [66,68]. Friction factors can be modeled using a variety of approaches, including power laws, exponential forms, or even combinations of both, as illustrated in Equation (2). However, the optimal expression of this formula

may vary depending on the purpose of the trip, the spatial intensity of travel locations, and the mode of transportation [69]. In different application scenarios, a single distance factor may not be sufficient to quantitatively capture the variations in travel costs between the origin and destination. Consequently, the attributes of the two locations can provide a better description of the direction and magnitude of flow between them.

$$f(r_{ij}) = \alpha r_{ij}^{-\beta} e^{-r_{ij}/r_c} \tag{2}$$

In the context of modeling long-term population flows between cities, it is evident that apart from the distance factor, the natural, economic, and other geographic features of the two cities play a crucial role as variables in the friction coefficient. Therefore, we adopt the principles of the gravity model and interpret the flow problem between the two cities as a regression problem. By inputting the population of the two locations, the distance between them, and other factors that influence population flows, we utilize lightGBM to capture the complex nonlinear relationship between population, friction coefficients, and flow. The model outputs simulated flows between the origin and destination, providing insights into the population movement between the two cities.

2.5. Architecture of GraviGBM

The procedure for estimating the flow $p(l_i, l_j)$ between locations l_i and l_j using the extended gravity model GraviGBM can be divided into three detailed steps (Figure 4). Firstly, we construct input vectors $x(l_i, l_j) = \text{concat}[x_i, x_j, r_{i,j}]$ for $j = 1, \ldots, n$ by concatenating the following input features: x_i, which includes characteristics such as population density and economic indicators of the origin location l_i; x_j, which includes similar features for the destination l_j; and the calculated Euclidean distance $r_{i,j}$ representing the shortest path between the two locations.

Secondly, for each origin l_i, we generate n input vectors $x(l_i, l_j)$, where each vector corresponds to a potential destination within the region of interest. The feature vectors $x(l_i, l_j)$ and their associated labels $y(l_i, l_j)$ are then fed into the lightGBM model.

Finally, the entire dataset is employed in the training phase, utilizing lightGBM to transform the problem into a regression task that establishes the relationship between the features of the origin $x(l_i, l_j)$ and destination cities $y(l_i, l_j)$ and the migration flow. In each boosting iteration, lightGBM sequentially builds decision trees. The construction of each tree involves fitting the residuals based on the predictions of the previous tree using the gradient descent algorithm. Each layer is optimized and enhanced based on the previous layers. The splitting of each node is determined by maximizing information gain or Gini impurity. Through iterative learning across multiple layers, the model's performance and predictive ability are gradually improved, resulting in an optimal model structure. The final model directly outputs the predicted flow $p(l_i, l_j)$ from i to j. This predicted flow, similar to the input labels, possesses a clear direction.

In addition to the city features (Table 1) and epidemic indicators (Table 2), the GraviGBM model incorporates the Euclidean distance between two locations as an additional input feature. The Euclidean distance represents the shortest path between two points in three-dimensional space. Therefore, in the absence of an epidemic, each flow in the GraviGBM model is described by 29 features (14 geographical features for the origin, 14 geographical features for the destination, and the distance between them). During an epidemic, the input features increase to 35, with an additional 3 epidemic indicators (confirmed counts, suspected counts, stringency index) for both the origin and destination.

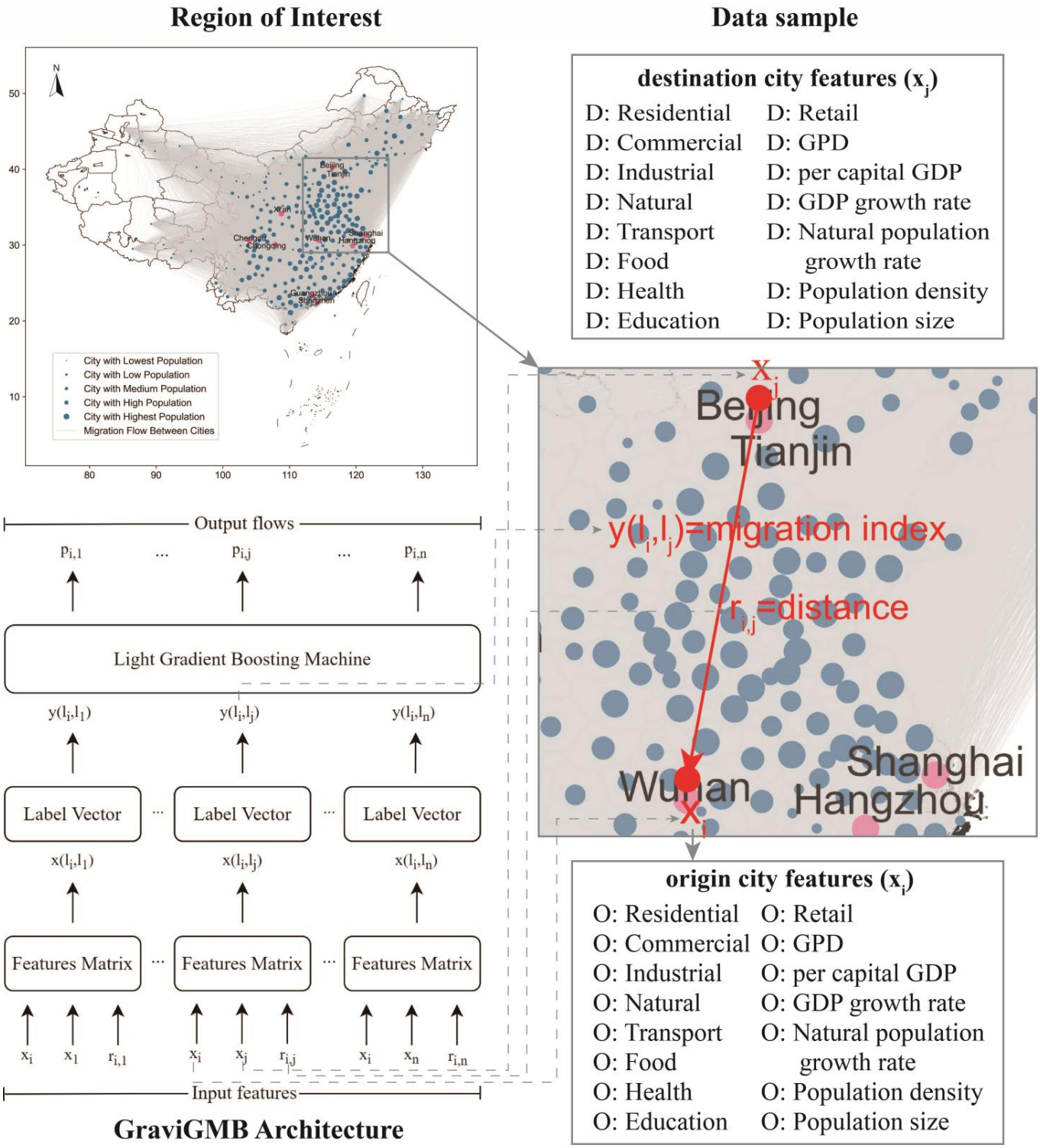


Figure 4. Flow generation and architecture of GraviGBM.

Table 1. Descriptions of geographic features in origin and destination cities.

Dimension	Indicator	Description
Land-use areas	Residential Commercial Industrial Natural	Total area (in km^2) for each possible land-use class
Public facilities	Transport	Total count of Points of Interest (POIs) related to each possible transport facility, e.g., bus/train station, bus stop, car parking
	Food	Total count of POIs related to food facilities, e.g., bar, cafe, restaurant
	Health	Total count of POIs related to health facilities, e.g., clinic, hospital, pharmacy
	Education	Total count of POIs related to education facilities, e.g., school, college, kindergarten
	Retail	Total count of POIs related to retail facilities, e.g., supermarket, department store, mall
Economic development	GDP	Total monetary value of all finished goods and services produced within a country's borders over an annual basis
	Per Capita GDP	Average economic output per person
	GDP growth rate	Percentage change in GDP over an annual basis
Population	Natural population growth rate	rate (in percentage) at which a population grows or declines over an annual basis
	Population density	Number of individuals per unit area in a city
	Population size	Total number of individuals residing in a city

Table 2. Indicators for assessing COVID-19.

Dimension	Indicator	Description
Pandemic	Confirmed counts	Total number of individuals officially diagnosed with COVID-19 through testing
	Suspected counts	Total number of individuals showing COVID-19 symptoms or potential exposure under investigation, but not yet confirmed cases
	Stringency index	A numerical scale measuring the strictness of government policies in response to COVID-19, ranging from 0 to 100, where higher values indicate more stringent measures such as lockdowns, travel bans, and social distancing regulations

Considering the similarity in form between GraviGBM and the gravity model, we use the latter as a baseline to evaluate the predictive performance of GraviGBM. Furthermore, we consider a simplified version of the GraviGBM model called the single Feature GraviGBM model (SFGraviGBM) to assess the impact of complex geographical features on the gains provided by the gravity model. SFGraviGBM also employs the lightGBM model, but retains the input features of the gravity model, consisting solely of population size and the distance between cities.

To evaluate the model's performance across different geographical regions, provide insights into its generalization capabilities, and ensure robustness across various scenarios, we implemented spatial cross-validation. We divided the dataset into three regions based on geographic location: the West, Central, and East. However, this division resulted in the loss of some label data due to interregional migration, such as migration from western

cities to central cities. Nevertheless, this division was solely for the purpose of spatial cross-validation. In the final modeling and simulation stage, we used the entire dataset.

The spatial cross-validation process consists of three rounds:

- Round 1:

  Training set: Central + West
  Test set: East

- Round 2:

  Training set: East + West
  Test set: Central

- Round 3:

  Training set: East + Central
  Test set: West

In the final model training phase, the characteristics of city pairs (origin–destination) within the study area are used as input features, while the population migration flow from the origin to the destination, represented by the Gaode migration index, serves as the label for simulation. The root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) are selected as performance evaluation metrics. RMSE measures the difference between the predicted and actual values, and is calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}$$

where y_i represents the actual values, $\widehat{y}_i$ represents the predicted values, and n is the number of observations. The role of RMSE is to quantify the accuracy of the model's predictions, with a lower RMSE indicating better predictive performance.

The advantage of MAE is its lower sensitivity to outliers, providing an additional perspective for evaluating predictive performance. MAE also evaluates prediction accuracy and is computed as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i|$$

R^2 quantifies the model's ability to fit the observed flow, and is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

GraviGBM adopts a non-traditional machine learning approach that focuses on calculating intercity migration flows based on the characteristics of both the origin and destination cities. Unlike conventional supervised learning methods that rely on labeled data representing individual features, our dataset is constructed using city pairs (origin–destination), with each record reflecting the migration flow between cities. This design means that the model's input depends on the features of both city pairs rather than individual city attributes. By leveraging the structure of lightGBM, the model effectively captures the relationship between city pair characteristics and migration flows, thereby uncovering the underlying drivers of intercity migration.

2.6. SHAP

SHAP (Shapley Additive exPlanation) is an explainable machine learning model inspired by the Shapley value that can provide explanations for the output of machine learning models [70]. The Shapley value originates from coalition game theory and was introduced by Shapley in 1953 [71]. Its core idea is to compute the marginal contribution of each feature to the model's output. For each predicted sample, the model generates a prediction value and the assigned value to each feature in that sample is the SHAP value.

This contribution is calculated by weighting and summing all of the possible combinations of feature values. We employ SHAP values to provide both global and local interpretability for GraviGBM. For global interpretation, we utilize collective SHAP values to explain the magnitude of the positive and negative effects of prediction factors (geographic variables, distance, population, epidemic indicators) on predictions. For local interpretation, we utilize single-sample analysis (specific population migration flow) to observe the positive and negative effects of prediction factors in specific predictions and their variations.

$$\phi_j(val) = \sum_{S \subseteq \{x_1, \cdots, x_p\} \setminus \{x_j\}} \frac{|S|!(p - |S| - 1)!}{p!} (\mathrm{val}(S \cup \{x_j\}) - \mathrm{val}(S)) \quad (3)$$

where $S \subseteq \{x_1, \cdots, x_p\} \setminus \{x_j\}$ represents all subsets of the feature set $\{x_1, \cdots, x_p\}$ excluding feature x_j; $|S|$ is the number of features in subset $\boldsymbol{S}$; $\mathrm{val}(S \cup \{x_j\})$ denotes the predicted value when feature x_j is included in subset $\boldsymbol{S}$; $\mathrm{val}(S)$ is the predicted value based only on the features in subset $\boldsymbol{S}$; and p is the total number of features. The notation n! (factorial) indicates the product of all positive integers up to n and is used here to calculate the number of ways to arrange the features. This formula calculates the Shapley value $\phi_j(val)$, which represents the average marginal contribution of feature x_j across all possible subsets $\boldsymbol{S}$. The sum of all SHAP values corresponds to the total difference between the model's actual prediction and the average prediction when no features are included.

We employed SHAP values to provide both global and local interpretability for GraviGBM. In the first scenario, we used collective SHAP values to explain the magnitude of the positive and negative effects of the predictive factors (geographical variables, distance, population, epidemic indicators) on the predictions. In the latter scenario, we utilized single-sample SHAP values to observe the positive and negative effects, as well as their variations, of predictive factors in specific predictions, focusing on a particular population migration flow.

3. Results

3.1. Modeling Population Mobility

We first conducted spatial cross-validation before training on the entire dataset. The dataset was divided into three regions—Western, Eastern, and Central—and three rounds of cross-validation were performed. Overall, GraviGBM performed well in the cross-validation, with results from each round (Table 3) indicating the model's effectiveness and robustness. Specifically, the generalization performance of the Central and Western datasets to the Eastern dataset was notably low, while the inclusion of the Eastern dataset in training significantly enhanced the model's overall generalization ability. The final average results demonstrate that the model has strong generalization capacity across different regions and can effectively adapt to the characteristic differences among these areas.

Table 3. Results of spatial cross-validation.

	RMSE	**MAE**	**R^2**
Round 1	6.8991	0.5701	0.5800
Round 2	1.6189	0.2165	0.8225
Round 3	1.6189	0.1734	0.9075
Mean	3.23	0.32	0.77

After completing the spatial cross-validation, we employed the GraviGBM model framework to train a population migration model. The input features include the characteristics of both the origin and destination cities, as well as the distance between them, while the label is defined as the migration index from the origin to the destination city. The goal of this model is to establish the relationship between urban characteristics and inter-city population migration, revealing the underlying logic of population movement. We utilized the entire dataset, comprising 86,730 records that capture all population migrations among

295 cities in China for the year 2019 to train the model and evaluate its simulation capabilities. This comprehensive dataset allows us to gain a deeper understanding of the intrinsic relationship between urban characteristics and population migration.

We employed the GraviGBM model framework using city features and distances between cities as input features, and the Gaode Mobility Index as the label, to construct a population flow model at the city level in China for the year 2019. This dataset comprised 86,730 records, representing the movement of people among 295 cities in China. To ensure a robust model evaluation, 80% of these data were used for training the model, while the remaining 20% was used for testing.

The gravity model, serving as the theoretical foundation for GraviGBM, is used as a baseline model to evaluate the improvement of the tree structure of lightGBM in capturing the complex nonlinear relationships involved in simulating population flows between cities. The Single Feature GraviGBM model (SFGraviGBM), as a baseline model, investigates the importance of complex geographical features beyond population and distance in constructing the population flow model.

Figure 5 and Table 4 illustrate the performance of the GraviGBM model and the baseline model. From the results, it is evident that GraviGBM achieves an RMSE of 2.98, accurately simulating both high and low values of population flow between cities and providing a comparison of population flow indices between different city pairs. Its performance significantly surpasses that of the gravity model (RMSE = 6.73), which, while able to describe the existence of population flows between cities, struggles to accurately estimate the actual volume of these flows. The performance of SFGraviGBM falls between the two models (RMSE = 5.33).

Table 4. Model performance.

Model	RMSE	MAE	R^2
GraviGBM	2.98	0.29	0.80
SFGraviGBM	5.33	0.45	0.37
Gravity	9.48	1.64	0.20

3.2. Factors Influencing Population Mobility Before the COVID-19 Pandemic

Shapley Additive Explanations (SHAPs) are employed to explain the impact of each input feature on the predictive performance of GraviGBM. The Shapley value of a feature represents its positive or negative effect, as well as its magnitude on the model's predictions. We utilized SHAPs to provide global explanations of intercity population flows in China before the COVID-19 pandemic (Figures 6 and 7). Additionally, we selected four sets of typical flows (Figures 8 and 9) to analyze and compare the diverse and complex factors influencing population movements in different scenarios. The features are arranged vertically on the *y*-axis according to their correlation, with higher correlations placed higher on the axis. The left side of the *y*-axis represents the values of the features (standardization has been applied to factors unrelated to the pandemic). The bar lengths represent the contributions of each feature to the model. The positive and negative contributions of the features, along with the average predicted value E[f(x)] of the model, collectively determine the final predicted value of the flow between a specific origin–destination pair.

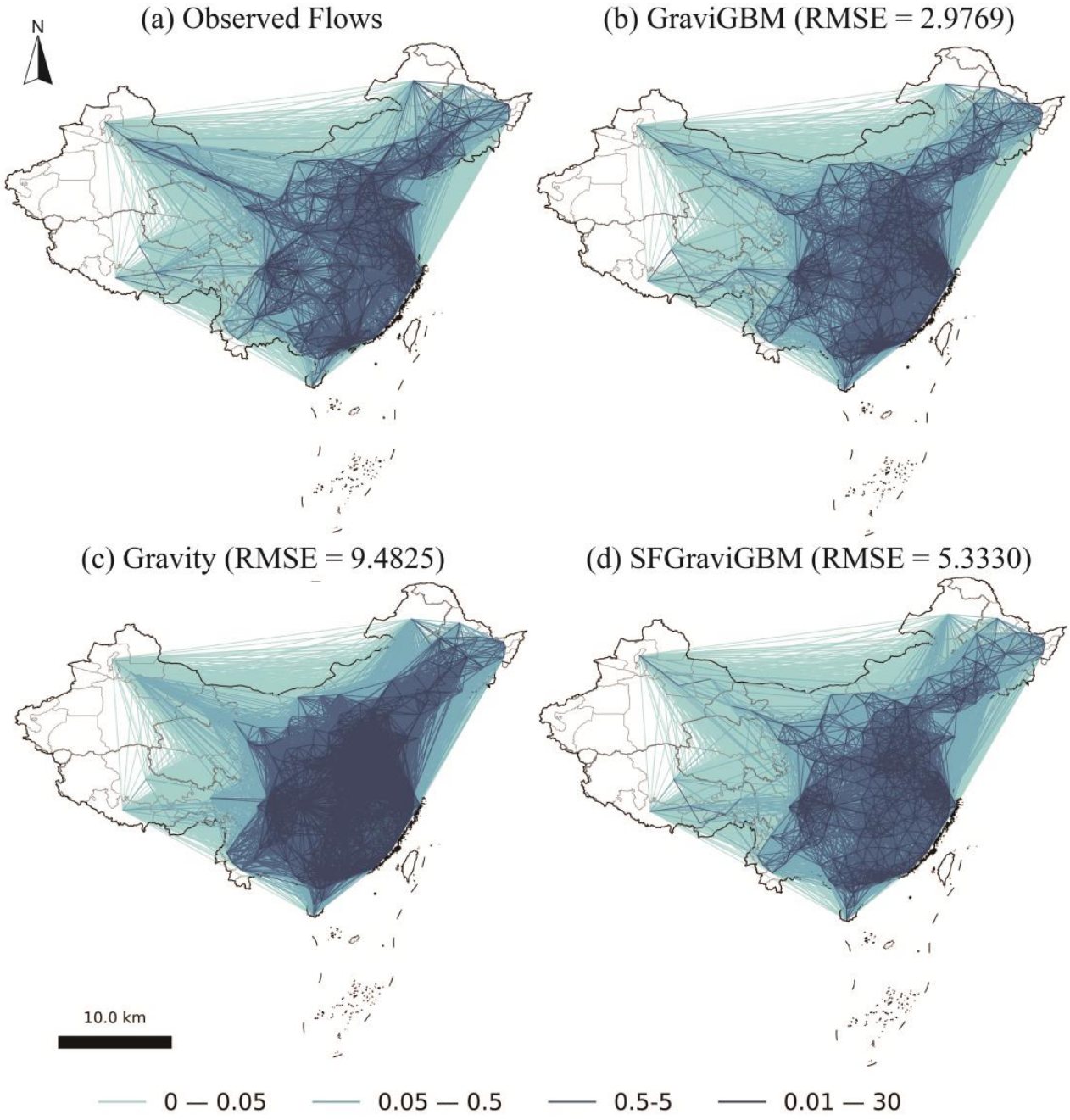


Figure 5. Real flows versus generated flows. (**a**) Visualization of the mobility network describing the observed flows. (**b**) The flows generated using GraviGBM. (**c**) The flows generated using the gravity model. (**d**) The flow generated using Single Feature GraviGBM. The gradient from yellow to red indicates the population mobility index increasing from 0 to 728. The root mean squared error (RMSE) reflects the accuracy of the estimated values.

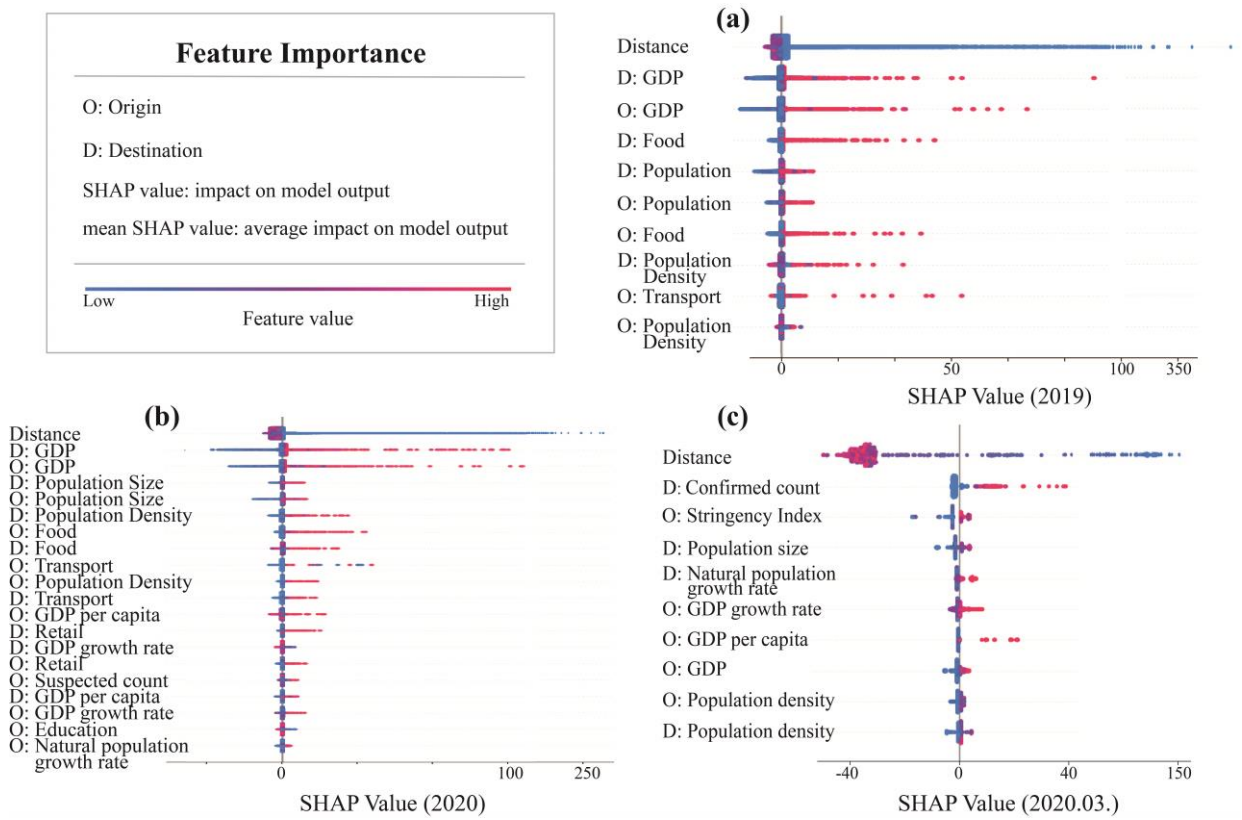


Figure 6. Explanation of GraviGBM before and during the COVID-19 pandemic. The features are displayed along the vertical axis, arranged in descending order of relevance, with the most significant ones at the top and the least significant ones at the bottom. Each data point represents a source-destination pair, where blue dots indicate pairs with lower feature values and red dots represent pairs with higher feature values. (**a**) Distribution of Shapely values for top ten features in GraviGBM for 2019. (**b**) Distribution of Shapely values for top twenty features in GraviGBM for 2020. (**c**) Distribution of Shapely values for top ten features in GraviGBM for March 2020.

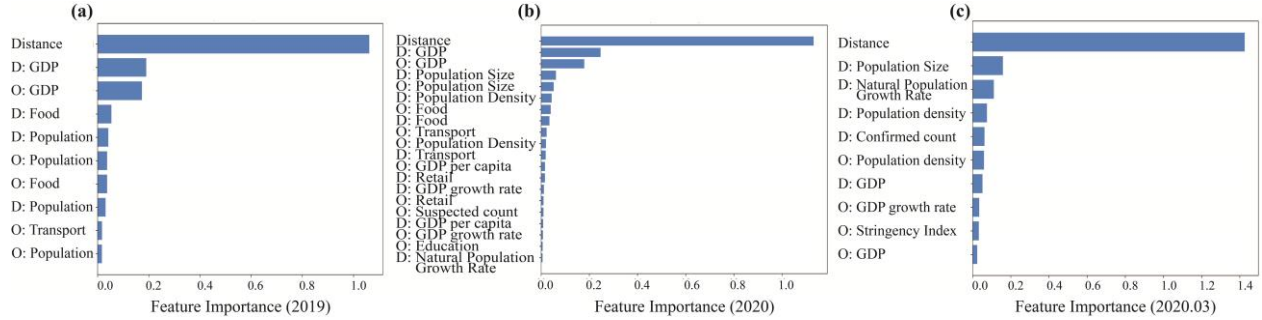


Figure 7. Feature importance bar chart for GraviGBM with SHAP. (**a**) Feature importance for top ten features in GraviGBM for 2019. (**b**) Feature importance for top twenty features in GraviGBM for 2020. (**c**) Feature importance for top ten features in GraviGBM for March 2020.

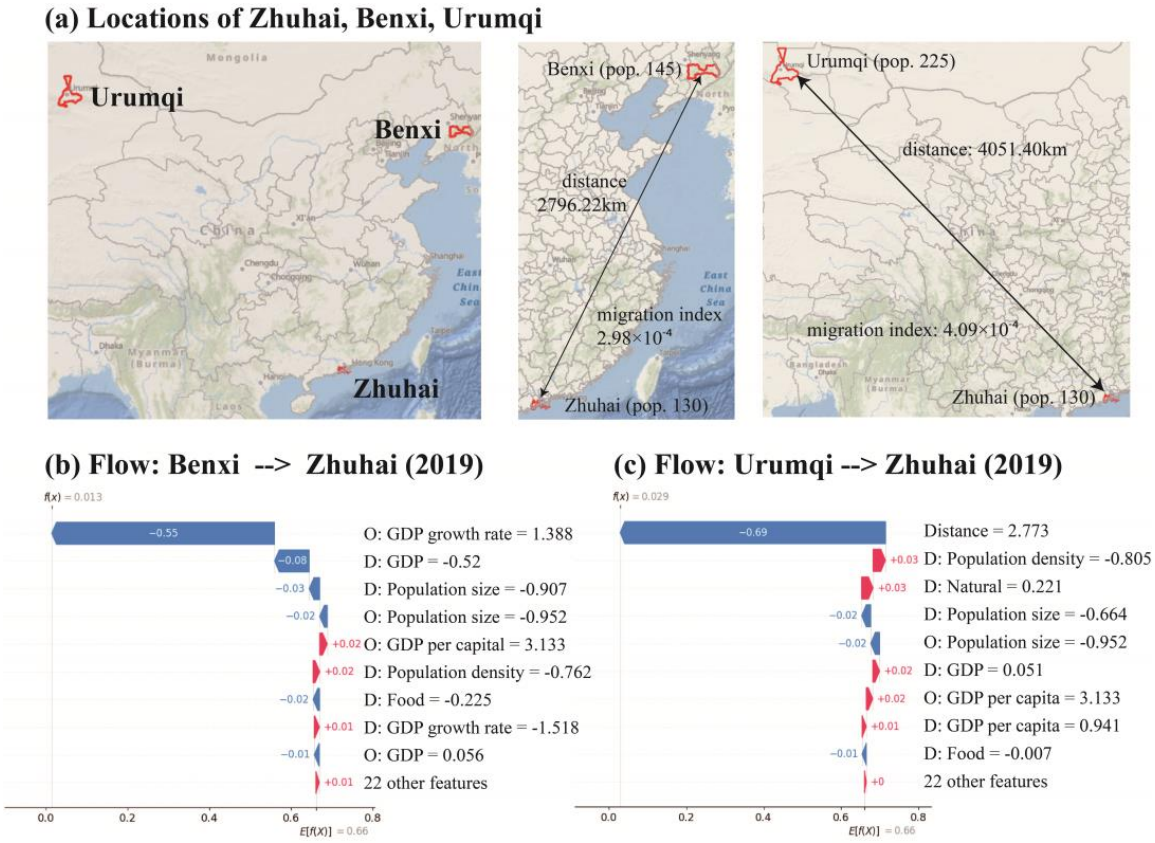


Figure 8. Explanation of generated flows before the COVID-19 pandemic. (**a**) Geographic positions, population, and distance between Benxi, Urumqi, and Zhuhai. Benxi (1.45 million) and Zhuhai (1.3 million) have the similar population. Urumqi (CNY 341.3256 million) and Zhuhai (CNY 344.423 billion) have the similar GDP values. (**b**,**c**) SHAP values for the flow from Benxi to Zhuhai and from Urumqi to Anhui during 2019.

From the global model analysis of population flows between Chinese cities in 2019, it is observed that, consistent with the classical gravity model, the most significant factor influencing population flows is distance, with its importance far exceeding other factors. This negative effect implies that cities that are closer in distance are more likely to experience population migration. However, the GraviGBM model reveals that GDP, both in the source and destination, plays a crucial role in population migration between cities, with an importance that is even three times higher than that of population size (Figure 6a). The aligned SHAP value and feature value indicate that the economic activity level of both the source and destination cities contributes to the occurrence of population flows, with a greater emphasis on the destination city's GDP. The population variable in GraviGBM exhibited a positive effect, with nearly equal importance for both the origin and destination populations. This observation aligns with the assumptions of classical models. Further observations of the SHAP values for population density reveal a positive effect for the source city's population density, while for the destination city, it generally exhibits a negative effect, with a lower global importance compared to the source city. This indicates that in the populous China, using only differences in population size to predict intercity flows is not appropriate, and population density may be a more suitable input feature within the population dimension. Interestingly, we found that the number of food facilities at the destination is of even greater importance than population size, making it the fourth most important feature.

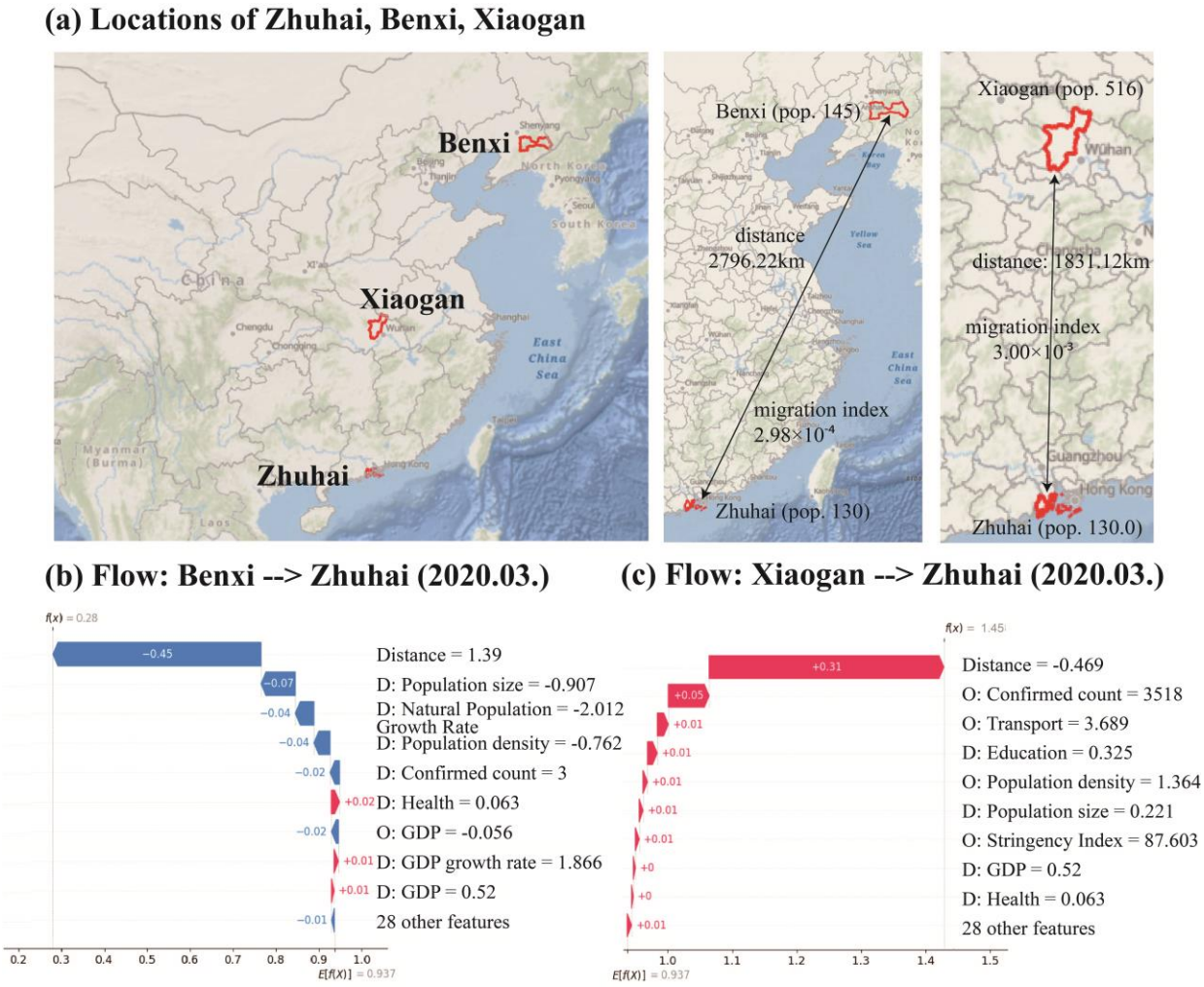


Figure 9. Explanation of generated flows after the COVID-19 pandemic. (**a**) Geographic position, shape, population, and distance between Benxi, Xiaogan, and Zhuhai. (**b**,**c**) SHAP values for the flow from Benxi to Zhuhai and from Xiaogan to Zhuhai during March 2020.

Distance is often seen as a barrier to migration. Cities that are closer tend to have lower migration costs (such as time and money), making it easier to attract population inflows when there are no significant differences in living, working, and social conditions. Moreover, GDP reflects a city's economic vitality and job opportunities. Cities with strong economic development can typically offer more job positions and a higher standard of living, attracting more people. Population size is also a key factor in determining a city's attractiveness. A larger population often implies richer resources and opportunities, thereby attracting more migration. Correspondingly, migration can decrease the population of the origin city, creating a positive feedback loop. Regarding another population-related factor—population density—higher density can indicate resource scarcity, prompting people to seek more attractive destinations. Conversely, lower population density in destination cities may reflect a more relaxed living environment, suitable for incoming populations. This dynamic suggests that relying solely on population size to predict migration is insufficient. The number of food facilities represents the convenience of urban living. More food facilities may indicate a richer array of lifestyle choices and a higher quality of life, especially for young people and families, which can be an important factor in their decision to migrate.

We also selected typical city pairs to study the contribution of input features to the predicted flow values for a single flow (a specific origin–destination pair) and compared it with the global model. For instance, Benxi and Zhuhai, with similar population sizes of

1.45 million and 1.3 million, respectively, are predicted to generate minimal population flows according to the calculations of the gravity model formula. However, in reality, there are frequent population flows between these two cities. Our proposed GraviGBM model can accurately simulate such data, addressing the limitation of the gravity model in capturing flows between cities with similar population sizes. In our specific analysis of the 2019 flow from Benxi to Zhuhai (Figure 8b), we found that the GDP growth rate of the source city is the most significant contributing factor to this flow, with a SHAP value of −0.55, far exceeding other factors. The SHAP value of the destination city's GDP (−0.08) also surpasses the population size of the origin city (−0.03) and the population size of the destination city (−0.02). Among the top nine important features, five are related to the economy (origin city's GDP growth rate, origin city's GDP, origin city's GDP per capita, destination city's GDP growth rate, destination city's GDP). This result demonstrates that when population sizes are similar, economic factors becomes the dominant factor driving population flows between two cities. Interestingly, in contrast to the analysis of global factors' contribution, distance plays a negligible role in the flow from Benxi to Zhuhai.

To further observe the changes in factors influencing population flows when economic factors are no longer the dominant factor alongside distance, we selected the population flow from Urumqi to Zhuhai. The GDP of these two cities in 2019 was CNY 341 million and CNY 344 million, respectively, which serves to reduce the impact of economic factors, allowing for a clearer comparison with the population flow from Benxi to Zhuhai. The results reveal that distance is the dominant factor in this flow, with a SHAP value of −0.69. The destination's population density and natural land-use are the two most important factors, apart from distance. The significance of natural land-use as one of the most important factors in this flow may be related to the specific characteristics of Zhuhai as a city. Zhuhai ranked first in the 2019 list of China's Most Livable Cities list [72], which suggests that factors related to living environment are prioritized as primary considerations for migration between cities with similar GDP levels.

These phenomena indicate that the level of economic development in cities directly influences people's migration decisions. In cities with similar population sizes, economic opportunities become the primary attraction for migrants. Well-developed cities offer richer resources and better living conditions, making people more willing to relocate in pursuit of a higher quality of life. When the economic conditions of the destination city significantly surpass those of the origin city, migrants may overlook the costs associated with distance. For example, despite the considerable distance between Benxi and Zhuhai, the allure of economic factors outweighs the time and money required for migration, a pattern observable in many pairs of cities with similar economic conditions. Conversely, when the economic conditions of both the origin and destination cities are similar, the influence of distance becomes more pronounced, emerging as the main driving force behind migration. This suggests that people are more likely to consider geographical factors when selecting their migration destination, especially in scenarios where economic opportunities are relatively balanced. Additionally, the livability of a city also affects people's willingness to migrate. In contexts of similar economic conditions, the superiority of a city's environment and quality of life can significantly enhance its attractiveness, making the focus on living quality an increasingly important consideration in migration decisions.

3.3. Factors Influencing Population Mobility During the COVID-19 Pandemic

To compare the factors influencing population mobility before and after the COVID-19 pandemic, city-specific characteristics, inter-city distances, and flow data from 2020 and March 2020 were used to construct inter-city population mobility models. Additionally, four pandemic-related statistical indicators, namely confirmed counts, suspected counts, and the stringency index, were incorporated into the models to better simulate population mobility patterns under the influence of the pandemic. Considering the significant reduction in population flow during the pandemic, the actual flow was multiplied by 1000 times during modeling to enhance the significance of these features.

From the results, it can be observed that the top ten factors influencing the overall model in 2020 remain the same as in 2019, and the direction of their impact also stays consistent (Figure 6b). Pandemic-related influencing factors had a negligible impact on the population mobility model for the entire year of 2020. Given that the concentrated outbreak of COVID-19 in China occurred in early 2020, we selected March 2020 to construct a monthly scale population mobility model using GraviGBM to further analyze changes in factors affecting population mobility during the pandemic. The results show that distance remained the most important factor influencing population mobility in March 2020, but the impact of shorter distances on mobility increased significantly. GDP was no longer the primary influencing factor; instead, the population size and natural growth rate of the destination became more influential. The population size of the destination continued to have a positive effect, as it was before the pandemic. However, after the outbreak of the pandemic, the destination's population natural growth rate exhibited a prominent positive effect. This could be because higher population growth rates are often associated with better healthcare systems and public health conditions, and during the pandemic, people tended to choose such places to ensure their own and their families' safety and health. Additionally, pandemic-related factors such as the destination's confirmed count, as well as the origin's stringency index, had a positive effect on population mobility. Economic factors, population density, transportation, and other non-pandemic-related influential factors no longer played a significant role under pandemic conditions.

Additionally, we chose the single flow from Benxi to Zhuhai as a representative city to study the contribution of features to predicting the flow values and compared these results with those from Section 3.2, both before and after the pandemic. We found (Figure 9b) that, during the outbreak of the pandemic, distance replaced the origin's GDP growth rate as the most important contributing feature to this flow. The absolute SHAP value for distance far exceeded other features, and it had a negative effect on the predicted values. Economic factors represented by GDP and the GDP growth rate diminished during the pandemic outbreak. Population-related features continued to make important contributions after the pandemic outbreak, and their direction remained the same as before the pandemic. The destination's confirmed count had a negative effect on this flow. The number of confirmed and dead cases at the destination increased people's concerns about infection risk, affecting their willingness and decision to migrate. We also selected Xiaogan City in Hubei Province, which experienced severe outbreaks in March, as the origin and studied the feature decision process for the flow to Zhuhai. The results show (Figure 9c) that the main influencing features had a positive effect on this sample, meaning that the actual flow of this sample was higher than the average flow of the samples during the same period. In contrast to the feature analysis results for Benxi to Zhuhai, distance had a positive impact in this sample. This indicates that during the outbreak, populations in severely affected areas tended to migrate to cities that were farther away. The important pandemic-related indicators for the origin included confirmed count and Stringency index, both of which had a positive effect. In the contemporaneous samples from Benxi to Zhuhai, the important factors were related to the destination's pandemic situation. The phenomenon may signify people's willingness to leave areas with strict lockdown measures for areas with more normalized management.

4. Discussion

The GraviGBM model, built on the classical gravity model, enhances the accuracy of mobility flow predictions by introducing a nonlinear structure and incorporating multiple geographic features. It also offers an expandable framework to include additional factors such as socioeconomic information and pandemic severity. This integration of diverse factors significantly improves population flow modeling by leveraging advanced AI techniques to handle complex multi-source data.

The SHAP method interprets the GraviGBM model, explaining the contributions of geographic variables. We found that population size and distance have effects similar to the classical gravity model, with GDP playing a significant role in intercity migration. During

the COVID-19 outbreak, distance and population size remained crucial, while pandemic-related indicators gained importance, highlighting concerns about health and safety.

The GraviGBM model effectively simulates population flows between cities with similar populations, indicating that migration is influenced by a complex set of factors beyond just distance and population size. Prior to COVID-19, the economy was a decisive factor; after the outbreak, the severity of the pandemic influenced migration decisions, emphasizing the impact of health risks and lockdown measures. This finding underscores the importance of integrating both health and socioeconomic data.

Our research provides a data-driven model for simulating population flows and elucidates the roles of geographic variables. Future research should consider utilizing complex structures like deep learning and examining model transferability to areas with limited data. Multi-scale and geospatial transfer applications are potential directions. Data-driven models leveraging AI technology will offer new perspectives for population migration modeling and will enhance the understanding of migration dynamics.

In our efforts to enhance the linear structure of the gravity model, we explored neural network methods, including linear neural networks, as well as Convolutional Neural Networks (CNN) and Graph Convolutional Networks (GCN). Unfortunately, due to the extreme imbalance in the dataset (with mobility index distribution spanning from 0.000278 to 29.766, and 75% of the data concentrated below 0.00256), neural network approaches tended to overfit on low values, leading to significant discrepancies between simulated results and actual values. Besides lightGBM, we also experimented with other machine learning methods such as random forests, support vector regression, linear regression, and decision tree regression. Among these, lightGBM exhibited the most promising performance, outperforming other methods in mitigating the challenges posed by the highly imbalanced nature of the dataset. The exploration of diverse AI techniques in handling imbalanced geospatial data significantly contributes to geospatial analysis and enhances model robustness.

5. Conclusions

This study introduces GraviGBM, a novel and extensible mobility flow generation model, and employs SHAPs to comprehensively analyze the determinants of population flows, especially within the context of the COVID-19 pandemic. The research outcomes indicate that GraviGBM significantly outperforms traditional models, achieving a lower RMSE that underscores its superior simulation capabilities, enabling a more accurate representation of the relationship between the characteristics of origin and destination cities and population migration. This enhanced simulation capability facilitates a deeper understanding and analysis of migration trends. Prior to the pandemic, distance played a pivotal role in population mobility, but during the crisis, short distances gained prominence as people sought to adapt to changing circumstances. Pandemic-related indicators became crucial determinants of population flows, underscoring the profound impact of public health crises on mobility patterns. Notably, this study sheds light on the intricate interplay between economic and living conditions in city-to-city migration decisions, offering valuable insights for urban planning and decision-making. Moreover, this research highlights the resilience and adaptability of populations during times of crisis, as individuals sought safer and more stable environments. These findings hold significant implications for urban planning, pandemic response strategies, and informed decision-making processes, providing a multifaceted perspective on population mobility in a rapidly changing world.

Author Contributions: Conceptualization, Jingjing Liu; Validation, Jingjing Liu and Le Ma; Data curation, Jingjing Liu; Visualization, Jingjing Liu; Methodology, Jingjing Liu; Writing—original draft, Jingjing Liu; Writing—review and editing, Lei Xu; Formal analysis, Le Ma; Supervision, Nengcheng Chen; Funding acquisition, Lei Xu and Nengcheng Chen. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Special Fund of Hubei Luojia Laboratory (Grant No. 220100034), National Key R&D Program (Grant No. 2018YFB2100500), National Key Research and Development Program for Young Scientist (Grant No. 2021YFF0704400), Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (Grant No. 162301212687).

Data Availability Statement: The original data presented in the study are openly available in Gaode Transportation Big Data Platform (https://trp.autonavi.com/migrate/page.do, accessed on 7 July 2023), OpenStreetMap (https://planet.osm.org/, accessed on 7 July 2023), China Economic and Social Big Data Research Platform (https://data.cnki.net/, accessed on 7 July 2023), DXY-COVID-19-Data 2024 (https://github.com/BlankerL/DXY-COVID-19-Data/, accessed on 7 July 2023), and Oxford COVID-19 Government Response Tracker (https://github.com/OxCGRT, accessed on 7 July 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bettencourt, L.M.A.; Lobo, J.; Helbing, D.; Kühnert, C.; West, G.B. Growth, Innovation, Scaling, and the Pace of Life in Cities. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7301–7306. [CrossRef] [PubMed]
2. Byrne, D. Class and Ethnicity in Complex Cities—The Cases of Leicester and Bradford. *Environ. Plan. A* **1998**, *30*, 703–720. [CrossRef]
3. Sîrbu, A.; Andrienko, G.; Andrienko, N.; Boldrini, C.; Conti, M.; Giannotti, F.; Guidotti, R.; Bertoli, S.; Kim, J.; Muntean, C.I.; et al. Human Migration: The Big Data Perspective. *Int. J. Data Sci. Anal.* **2021**, *11*, 341–360. [CrossRef]
4. Huang, S.; Gan, Y.; Chen, N.; Wang, C.; Zhang, X.; Li, C.; Horton, D.E. Urbanization Enhances Channel and Surface Runoff: A Quantitative Analysis Using Both Physical and Empirical Models over the Yangtze River Basin. *J. Hydrol.* **2024**, *635*, 131194. [CrossRef]
5. Ji, R.; Wang, C.; Cui, A.; Jia, M.; Liao, S.; Wang, W.; Chen, N. Assessing Terrestrial Water Storage Dynamics and Multiple Factors Driving Forces in China from 2005 to 2020. *J. Environ. Manag.* **2024**, *370*, 122464. [CrossRef]
6. Chen, Z.; Gong, Z.; Yang, S.; Ma, Q.; Kan, C. Impact of Extreme Weather Events on Urban Human Flow: A Perspective from Location-Based Service Data. *Comput. Environ. Urban Syst.* **2020**, *83*, 101520. [CrossRef]
7. Lissoni, F. International Migration and Innovation Diffusion: An Eclectic Survey. *Reg. Stud.* **2018**, *52*, 702–714. [CrossRef]
8. Huang, S.; Gan, Y.; Zhang, X.; Chen, N.; Wang, C.; Gu, X.; Ma, J.; Niyogi, D. Urbanization Amplified Asymmetrical Changes of Rainfall and Exacerbated Drought: Analysis Over Five Urban Agglomerations in the Yangtze River Basin, China. *Earth's Future* **2023**, *11*, e2022EF003117. [CrossRef]
9. Ma, L.; Xiang, L.; Wang, C.; Chen, N.; Wang, W. Spatiotemporal Evolution of Urban Carbon Balance and Its Response to New-Type Urbanization: A Case of the Middle Reaches of the Yangtze River Urban Agglomerations, China. *J. Clean. Prod.* **2022**, *380*, 135122. [CrossRef]
10. Li, K.; Li, C.; Xiang, Y.; He, F.; He, S.; Chen, J.; Fang, Y.; Sun, Y. Policy and Newly Confirmed Cases Universally Shape the Human Mobility during COVID-19. *Natl. Sci. Open* **2022**, *1*, 20220003. [CrossRef]
11. Barbosa, H.; Barthelemy, M.; Ghoshal, G.; James, C.R.; Lenormand, M.; Louail, T.; Menezes, R.; Ramasco, J.J.; Simini, F.; Tomasini, M. Human Mobility: Models and Applications. *Phys. Rep.* **2018**, *734*, 1–74. [CrossRef]
12. Pappalardo, L.; Manley, E.; Sekara, V.; Alessandretti, L. Future Directions in Human Mobility Science. *Nat. Comput. Sci.* **2023**, *3*, 588–600. [CrossRef] [PubMed]
13. González, M.C.; Hidalgo, C.A.; Barabási, A.-L. Understanding Individual Human Mobility Patterns. *Nature* **2008**, *453*, 779–782. [CrossRef]
14. Badr, H.S.; Du, H.; Marshall, M.; Dong, E.; Squire, M.M.; Gardner, L.M. Association between Mobility Patterns and COVID-19 Transmission in the USA: A Mathematical Modelling Study. *Lancet Infect. Dis.* **2020**, *20*, 1247–1254. [CrossRef] [PubMed]
15. Simini, F.; Barlacchi, G.; Luca, M.; Pappalardo, L. A Deep Gravity Model for Mobility Flows Generation. *Nat. Commun.* **2021**, *12*, 6576. [CrossRef] [PubMed]
16. Prieto Curiel, R.; Pappalardo, L.; Gabrielli, L.; Bishop, S.R. Gravity and Scaling Laws of City to City Migration. *PLoS ONE* **2018**, *13*, e0199892. [CrossRef]
17. Gray, C.L.; Mueller, V. Natural Disasters and Population Mobility in Bangladesh. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6000–6005. [CrossRef]
18. Myers, N. Environmental Refugees: A Growing Phenomenon of the 21st Century. *Philos. Trans. R. Soc. B Biol. Sci.* **2002**, *357*, 609–613. [CrossRef]
19. Reuveny, R. Climate Change-Induced Migration and Violent Conflict. *Political Geogr.* **2007**, *26*, 656–673. [CrossRef]
20. Salah, A.A.; Pentland, A.; Lepri, B.; Letouze, E.; Vinck, P.; de Montjoye, Y.-A.; Dong, X.; Dagdelen, O. Data for Refugees: The D4R Challenge on Mobility of Syrian Refugees in Turkey. *arXiv* **2018**, arXiv:1807.00523.
21. Ebrahimpour, Z.; Wan, W.; Cervantes, O.; Luo, T.; Ullah, H. Comparison of Main Approaches for Extracting Behavior Features from Crowd Flow Analysis. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 440. [CrossRef]
22. Xie, P.; Li, T.; Liu, J.; Du, S.; Yang, X.; Zhang, J. Urban Flow Prediction from Spatiotemporal Data Using Machine Learning: A Survey. *Inf. Fusion* **2020**, *59*, 1–12. [CrossRef]

23. Zhao, J.; Wang, Y.; Li, F. Modeling Population Migration Using Feedforward Neural Networks: A Case Study in China. *Appl. Intell.* **2021**, *51*, 567–578.
24. Jia, J.S.; Lu, X.; Yuan, Y.; Xu, G.; Jia, J.; Christakis, N.A. Population Flow Drives Spatio-Temporal Distribution of COVID-19 in China. *Nature* **2020**, *582*, 389–394. [CrossRef] [PubMed]
25. Zhang, M.; Wang, S.; Hu, T.; Fu, X.; Wang, X.; Hu, Y.; Halloran, B.; Li, Z.; Cui, Y.; Liu, H.; et al. Human Mobility and COVID-19 Transmission: A Systematic Review and Future Directions. *Ann. GIS* **2022**, *28*, 501–514. [CrossRef]
26. Yu, L.; Zhao, P.; Tang, J.; Pang, L. Changes in Tourist Mobility after COVID-19 Outbreaks. *Ann. Tour. Res.* **2023**, *98*, 103522. [CrossRef]
27. Kroll, C.; Warchold, A.; Pradhan, P. Sustainable Development Goals (SDGs): Are We Successful in Turning Trade-Offs into Synergies? *Palgrave Commun.* **2019**, *5*, 140. [CrossRef]
28. Le Blanc, D. Towards Integration at Last? The Sustainable Development Goals as a Network of Targets. *Sustain. Dev.* **2015**, *23*, 176–187. [CrossRef]
29. Simini, F.; Barlacchi, G.; Kessler, F.B.; Pellungrini, R.; Pappalardo, L. Human Mobility from Theory to Practice: Data, Models and Applications. In Proceedings of the WWW '19: Companion Proceedings of the 2019 World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 1311–1312. [CrossRef]
30. Anderson, J. The Gravity Model. *Annu. Rev. Econ.* **2010**, *3*, 133–160. [CrossRef]
31. Lewer, J.J.; Van Den Berg, H. A Gravity Model of Immigration. *Econ. Lett.* **2008**, *99*, 164–167. [CrossRef]
32. Westerlund, J.; Wilhelmsson, F. Estimating the Gravity Model without Gravity Using Panel Data. *Appl. Econ.* **2011**, *43*, 641–649. [CrossRef]
33. Hong, I.; Jung, W.-S.; Jo, H.-H. Gravity Model Explained by the Radiation Model on a Population Landscape. *PLoS ONE* **2019**, *14*, e0218028. [CrossRef] [PubMed]
34. Kang, C.; Liu, Y.; Guo, D.; Qin, K. A Generalized Radiation Model for Human Mobility: Spatial Scale, Searching Direction and Trip Constraint. *PLoS ONE* **2015**, *10*, e0143500. [CrossRef] [PubMed]
35. Wilmot, C. *Modeling Hurricane Evacuation Traffic: Testing the Gravity and Intervening Opportunity Models as Models of Destination Choice in Hurricane Evacuation [Technical Summary]*; Louisiana Transportation Research Center: Baton Rouge, LA, USA, 2008.
36. Rodrigue, J.-P. *The Geography of Transport Systems*, 5th ed.; Routledge: London, UK, 2020; ISBN 978-0-429-34632-3.
37. Robinson, C.; Dilkina, B. A Machine Learning Approach to Modeling Human Migration. In Proceedings of the COMPASS '18: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, San Jose, CA, USA, 20–22 June 2018.
38. Azizi, S.; Yektansani, K. Artificial Intelligence and Predicting Illegal Immigration to the USA. *Int. Migr.* **2020**, *58*, 183–193. [CrossRef]
39. Carammia, M.; Iacus, S.M.; Wilkin, T. Forecasting Asylum-Related Migration Flows with Machine Learning and Data at Scale. *Sci. Rep.* **2022**, *12*, 1457. [CrossRef]
40. Giang, N.H.; Nguyen, T.-T.; Tay, C.C.; Phuong, L.A.; Dang, T.-T. Towards Predictive Vietnamese Human Resource Migration by Machine Learning: A Case Study in Northeast Asian Countries. *Axioms* **2022**, *11*, 151. [CrossRef]
41. Best, K.; Gilligan, J.; Baroud, H.; Carrico, A.; Donato, K.; Mallick, B. Applying Machine Learning to Social Datasets: A Study of Migration in Southwestern Bangladesh Using Random Forests. *Reg. Environ. Change* **2022**, *22*, 52. [CrossRef]
42. Golenvaux, N.; Alvarez, P.G.; Kiossou, H.S.; Schaus, P. An LSTM Approach to Forecast Migration Using Google Trends. *arXiv* **2020**, arXiv:2005.09902.
43. Terroso-Sáenz, F.; Muñoz, A. Nation-Wide Human Mobility Prediction Based on Graph Neural Networks. *Appl. Intell.* **2022**, *52*, 4144–4160. [CrossRef]
44. Wang, Z.; Chen, Y. Exploring Spatial Patterns of Interurban Passenger Flows Using Dual Gravity Models. *Entropy* **2022**, *24*, 1792. [CrossRef]
45. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149–3157.
46. Simpson, L.; Finney, N. Spatial Patterns of Internal Migration: Evidence for Ethnic Groups in Britain. *Popul. Space Place* **2009**, *15*, 37–56. [CrossRef]
47. Wang, C.; Ma, L.; Zhang, Y.; Chen, N.; Wang, W. Spatiotemporal Dynamics of Wetlands and Their Driving Factors Based on PLS-SEM: A Case Study in Wuhan. *Sci. Total Environ.* **2022**, *806*, 151310. [CrossRef] [PubMed]
48. Kraemer, M.U.G.; Yang, C.-H.; Gutierrez, B.; Wu, C.-H.; Klein, B.; Pigott, D.M.; Open COVID-19 Data Working Group; Du Plessis, L.; Faria, N.R.; Li, R.; et al. The Effect of Human Mobility and Control Measures on the COVID-19 Epidemic in China. *Science* **2020**, *368*, 493–497. [CrossRef] [PubMed]
49. Zhang, C.; Zhang, K.; Yuan, Q.; Tao, F.; Zhang, L.; Hanratty, T.; Han, J. ReAct: Online Multimodal Embedding for Recency-Aware Spatiotemporal Activity Modeling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Tokyo, Japan, 7–11 August 2017; ACM: New York, NY, USA, 2017; pp. 245–254.
50. Ruktanonchai, N.W.; Floyd, J.R.; Lai, S.; Ruktanonchai, C.W.; Sadilek, A.; Rente-Lourenco, P.; Ben, X.; Carioli, A.; Gwinn, J.; Steele, J.E.; et al. Assessing the Impact of Coordinated COVID-19 Exit Strategies across Europe. *Science* **2020**, *369*, 1465–1470. [CrossRef]
51. Ma, S.; Li, S.; Zhang, J. Spatial and Deep Learning Analyses of Urban Recovery from the Impacts of COVID-19. *Sci. Rep.* **2023**, *13*, 2447. [CrossRef]

52. Pepe, E.; Bajardi, P.; Gauvin, L.; Privitera, F.; Lake, B.; Cattuto, C.; Tizzoni, M. COVID-19 Outbreak Response, a Dataset to Assess Mobility Changes in Italy Following National Lockdown. *Sci. Data* **2020**, *7*, 230. [CrossRef]
53. Lemos, B.M.; Nascimento, F.A.C.D.; Silva, M.A.V.D. Influence of the Spatial Distribution of Jobs in Intervening Opportunities Models. *Transp. Res. Rec. J. Transp. Res. Board* **2023**, *2677*, 1441–1454. [CrossRef]
54. Van den Broeck, G.; Lykov, A.; Schleich, M.; Suciu, D. On the Tractability of SHAP Explanations. *J. Artif. Intell. Res.* **2022**, *74*, 851–886. [CrossRef]
55. Bowen, D.; Ungar, L. Generalized SHAP: Generating Multiple Types of Explanations in Machine Learning. *arXiv* **2020**, arXiv:2006.0715.
56. Rathi, S. Generating Counterfactual and Contrastive Explanations Using SHAP. *arXiv* **2019**, arXiv:1906.09293.
57. Gaode Autonavi Gaode Migration Data 2019. Available online: https://report.amap.com/migrate/page.do (accessed on 7 July 2023).
58. OpenStreetMap Contributors OpenStreetMap Data 2019. Available online: https://planet.osm.org/ (accessed on 7 July 2023).
59. China Economic and Social Big Data Research Platform. Available online: https://data.cnki.net/ (accessed on 4 July 2023).
60. DXY-COVID-19-Data 2024. Available online: https://github.com/BlankerL/DXY-COVID-19-Data/ (accessed on 7 July 2023).
61. Hale, T.; Angrist, N.; Goldszmidt, R.; Kira, B.; Petherick, A.; Phillips, T.; Webster, S.; Cameron-Blake, E.; Hallas, L.; Majumdar, S.; et al. A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* **2021**, *5*, 529–538. [CrossRef] [PubMed]
62. Zhou, Y.; He, J. A Randomized Approach for Crowdsourcing in the Presence of Multiple Views. In Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 685–694.
63. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.
64. Zeng, H.; Yu, X.; Zhang, J. Urban Village Demolition, Migrant Workers' Rental Costs and Housing Choices: Evidence from Hangzhou, China. *Cities* **2019**, *94*, 70–79. [CrossRef]
65. Liang, W.; Luo, S.; Zhao, G.; Wu, H. Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms. *Mathematics* **2020**, *8*, 765. [CrossRef]
66. Lenormand, M.; Bassolas, A.; Ramasco, J.J. Systematic Comparison of Trip Distribution Laws and Models. *J. Transp. Geogr.* **2016**, *51*, 158–169. [CrossRef]
67. Zipf, G. The P1 P2/D Hypothesis: On the Intercity Movement of Persons. In *American Sociological Review*; American Sociological Association: Washington, DC, USA, 1946.
68. Lenormand, M.; Huet, S.; Gargiulo, F.; Deffuant, G. A Universal Model of Commuting Networks. *PLoS ONE* **2012**, *7*, e45985. [CrossRef]
69. Barthélemy, M. Spatial Networks. *Phys. Rep.* **2011**, *499*, 1–101. [CrossRef]
70. GitHub—Shap/Shap: A Game Theoretic Approach to Explain the Output of Any Machine Learning Model. Available online: https://github.com/shap/shap (accessed on 13 June 2024).
71. Štrumbelj, E.; Kononenko, I. Explaining Prediction Models and Individual Predictions with Feature Contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647–665. [CrossRef]
72. Available online: https://top.askci.com/news/20190719/0940461149998.shtml (accessed on 7 July 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.