Complexity

The Complexity of Mobility and Human Dynamics in Cities

Lead Guest Editor: Francisco Prieto
Guest Editors: Rosa M. Benito, Xiaowen Dong, and Burçin Bozkaya

The Complexity of Mobility and Human Dynamics in Cities

Complexity

The Complexity of Mobility and Human Dynamics in Cities

Lead Guest Editor: Francisco Prieto
Guest Editors: Rosa M. Benito, Xiaowen Dong, and Burçin Bozkaya

Contents

Hindawi
Complexity
Volume 2022, Article ID 8843000, 20 pages
https://doi.org/10.1155/2022/8843000

WILEY | Hindawi

Research Article

Modelling on Car-Sharing Serial Prediction Based on Machine Learning and Deep Learning

Nihad Brahimi,[1] **Huaping Zhang**,[1] **Lin Dai**,[1] **and Jianzi Zhang**[2]

[1]*School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China*
[2]*Chang'an Automobile Co., Ltd., Chongqing, China*

Correspondence should be addressed to Huaping Zhang; kevinzhang@bit.edu.cn

Received 19 August 2020; Revised 29 November 2021; Accepted 7 December 2021; Published 5 January 2022

Academic Editor: Rosa M. Benito

The car-sharing system is a popular rental model for cars in shared use. It has become particularly attractive due to its flexibility; that is, the car can be rented and returned anywhere within one of the authorized parking slots. The main objective of this research work is to predict the car usage in parking stations and to investigate the factors that help to improve the prediction. Thus, new strategies can be designed to make more cars on the road and fewer in the parking stations. To achieve that, various machine learning models, namely vector autoregression (VAR), support vector regression (SVR), eXtreme gradient boosting (XGBoost), k-nearest neighbors (kNN), and deep learning models specifically long short-time memory (LSTM), gated recurrent unit (GRU), convolutional neural network (CNN), CNN-LSTM, and multilayer perceptron (MLP), were performed on different kinds of features. These features include the past usage levels, Chongqing's environmental conditions, and temporal information. After comparing the obtained results using different metrics, we found that CNN-LSTM outperformed other methods to predict the future car usage. Meanwhile, the model using all the different feature categories results in the most precise prediction than any of the models using one feature category at a time

1. Introduction

Predicting the future is considered as one of the most challenging tasks in applied sciences. Computational and statistical methods are used for deducting dependencies between past and future observed values in order to build effective predictors from historical data. Transport answers to people's desire to participate in different activities in different places [1]. Cars have become a part of the mobility ecosystem owing to the flexibility and freedom that they provide [2]. People are more dependent on cars for both intercity and intracity transit, causing traffic congestion and parking difficulties [3]. "Looking for a parking space creates additional delays and impairs local circulation. In central areas of large cities, cruising may account for more than 10% of the local circulation as drivers can spend 20 minutes looking for a parking spot", said by Dr. Jean-Paul Rodrigue Department of Global Studies and Geography of Hofstra University. Many rental models are emerged to solve these parking problems as one of them is the car-sharing model, which aims to distribute cars within a city for use at a low cost. In this fashion, individuals can exploit all the benefits of a private vehicle without the hassles of lease payments, maintenance, or parking. The program comprises one-way or round-trip, depending on whether the pick-up and the drop-off stations are the same or not [4].

The car-sharing system provides an option to numerous people who opt not to own a vehicle, and they use this system whenever a private vehicle is needed. This system usually fixes the cost on the price per minute that includes a quote of variables such as fuel, price per kilometre, and the share of fixed costs for the operator like maintenance, rebalancing, insurance, and parking [5]. Besides helping in decreasing the level of congestion and managing the lack of parking lots, car-sharing systems have many other advantages such as the reduction of vehicle ownership that leads to efficient use of road and infrastructure, economical savings for the users, and diminution of air and noise pollution [5].

However, this program is facing many issues [6], one of them is the nonsuitable distribution of vehicles within car-sharing systems. As a result, cars tend to be easily accessible in low-demand parking lots in excess whereas an insufficient number of vehicles are available in high-demand parking lots [6]. For car-sharing companies, this problem causes a major financial loss. To improve the car usage rate, the companies employ a variety of techniques that hold great promises in car-sharing predictions.

Over the last few years, machine learning and deep learning have proved their efficiency and got recognition in different fields. Machine learning approaches make use of learning algorithms that make inferences from data to learn new tasks [7] and are widely adopted in a number of massive and complex data-intensive fields such as medicine, astronomy, and biology [8–11]. Deep learning models yield good results in the fields of computer vision and natural language processing [12–15] where it can automatically extract multidimensional features and effectively extract the data patterns for classification or regression [16].

In our work, a multivariate time series approach is presented, and it aims to predict the car usage in the short term and to investigate the factors that help to improve its prediction accuracy. Multiple machine learning models and deep learning models have already fulfilled their promises for multivariate time series prediction approaches and also have proved their ability in extracting meaningful understandings that are hard for humans to analyze and infer [5]. Those models were performed with different features set including the past usage levels, Chongqing's environmental conditions, and the temporal information.

The rest of the paper is organized as follows. Section 2 presents a literature review of current studies on times series models. Section 3 gives a description of the studied problem. A time series analysis is presented in Section 4. Section 5 demonstrates the framework of our approach. Section 6 describes the experimental framework used to evaluate the performance of used models for the multivariate time series approach. Finally, Section 7 concludes the paper and outlines future work directions.

2. Literature Review

Car-sharing has become one of the most popular research subjects in transportation. Many studies have been conducted, but to the best of our knowledge, no work has been done in the scientific literature that compares different machine learning and deep learning models in predicting the future usage of the car-sharing system and in investigating the factors that help improve its prediction accuracy. Interesting works are instead the following:

Studies related to this topic can be categorized into numerous subgroups including the followings [5, 17]: (i) user characteristics: it investigates the ways the users interact with the service [18, 19]; (ii) characterizing the shaping service in charge of the provisions and distribution of the cars around the city [20, 21]; (iii) car demands prediction level [22, 23].

Car demands prediction levels for car-sharing systems can be formulated as a time series prediction problem. Time series uses many approaches, such as the autoregressive integrated moving average (ARIMA) model that focuses on extracting the temporal variation patterns of the traffic flow and uses them for prediction; support vector regression (SVR) model, that captures complex nonlinearities, [20] demonstrated that this approach generally performs better on traffic flow time series; extreme gradient boosting (XGBoost), [21] showed that this model improves the prediction's precision and efficiency. Before starting the calculation, XGBoost sorts the traffic data according to the feature values and also realizes parallel computing on feature enumerations.

Recently, machine learning methods have been challenged by deep learning methods on traffic prediction. Deep learning approaches have a strong ability to express multimodal patterns in data, in order to reduce the overfitting problem and to obtain high prediction accuracy. In addition, as a traffic flow process is complicated in nature, deep learning algorithms can represent traffic features without prior knowledge, which has good performance for traffic flow prediction.

Xu and Lim [22] used an evolutionary neural network to prove the effectiveness of this algorithm and its possible usage as a tool for forecasting the net flow of a car-sharing system in order to offer the vehicle in the shortest time possible with the best accuracy; [23] attempted using the deep belief network (DBN) to define a deep architecture for traffic flow prediction that learns features with limited prior knowledge.

The abovementioned models require the input length to be predefined and static, and they cannot automatically determine the optimal time lags. To remedy these problems, many works have been done such as [24] used a model called long short-term memory recurrent neural network (LSTM RNN) that capture the nonlinearity and randomness of traffic flow more effectively and automatically determine the optimal time lags; [25] presented a novel long short-term memory neural network to predict travel speed using microwave detector data, where the future traffic condition is commonly relevant to the previous events with long time spans; Mo et al. [26] predicted the future trajectory of a surrounding vehicle in congested traffic by using the CNN-LSTM. To the best of our knowledge, no work is found in the literature on car-sharing time series prediction using CNN-LSTM.

Regarding the investigation of factors improving the prediction, [6] conducted the study about the effect of seasonal factors on the bookings of cars in Montreal, and after analysing the results, it was concluded that the usage outcomes scored better in summer season.

With respect to the above works, our approach presents the following highlights:

(i) Comparison of various machine learning and deep learning models to predict the future number of bookings made by car-sharing users using different metrics

(ii) Investigation of the factors that help predicting the car-sharing usage by estimating the relationship between data features and model performances

3. Problem Description

The aim of this study is to predict the number of vehicles that are going to be used in the parking stations at a given moment and to investigate the factors that improve the accuracy of predictions.

The number of vehicle usage at a given time t_x is likely to be correlated with a set of features [6], which are as follows:

(i) The past usage (F^x_{usage}): the history usage is tracked to build prediction models. It comprises of the number of car-sharing transactions based on the data from a car-sharing operator located in Chongqing, China.

(ii) Temporal information (F^x_{time}): the time at which the past usages have been acquired. Since the car demands may vary over time, we partition the time period into segments to capture different temporal trends (e.g., holidays/working days, 1 h timeslots) [6].

(iii) The environmental conditions (F^x_{weather}) at that time: the user transportation habits are usually affected by the weather conditions.

Table 1 summarizes the description of the feature categories of the car rental time series.

3.1. Car Rental Prediction Problem. Hereafter, we formulate the multivariate regression problem addressed in this paper [6]. It consists of predicting the car usage based on the values of features belonging to categories F^x_{usage}, F^x_{time}, and F^x_{weather}. Let T be the historical time period considered for training, t_1, ..., t_{k-1} be the past time points in T, and t_k be the current sampling time. We will denote usage (t_j) $[1 \le j \le k]$ the usage level at time t_j. We will give 1 h timeslots as a prediction horizon [6].

Since the future car usage is related to multiple features, the multivariate regressor R is expressed as follows:

$$\text{usage}(t_{k+1}) = R\left(F^x_{\text{usage}}, F^x_{\text{time}}, F^x_{\text{weather}}, 1 \le x \le k\right), \quad (1)$$

where F^x_{usage} is the usage levels of cars, F^x_{time} is the temporal information at t_x, and F^x_{weather} is the weather conditions in the area at time t_x [6].

3.2. Factors Investigation Problem. Another objective of this research is to determine the factors that help to predict vehicle usage. The features considered in this study are classified into different categories, namely, the past usage, temporal information, and the environmental conditions. We studied numerous machine learning and deep learning models, merging the different feature categories or using them separately one by one, aiming to find features that improve the prediction accuracy of the models.

4. Time Series Analysis

Time series is a sequential collection of recorded observations in consecutive time periods, and they can be univariate or multivariate [27, 28]. We may perform time series analysis with the aim of either predicting future values or understanding the processes driving them [29].

To address the problems stated in the previous section, multiple machine learning and deep learning models were performed.

4.1. Machine Learning Models

4.1.1. Vector Autoregression (VAR). Vector autoregression is a forecasting algorithm used when two or more time series influence each other. It is considered as an autoregressive model because the predictors are not only lags of the series but also past lags of itself [30]. Suppose we measure three different time series variables, denoted by $x_{t,1}$, $x_{t,2}$, $x_{t,3}$. The vector autoregression model of order 1 denoted as VAR(1) is as follows [30]:

$$x_{t,1} = a_1 + \Phi_{11}x_{t-1,1} + \Phi_{12}x_{t-1,2} + \Phi_{13}x_{t-1,3} + w_{t,1}, \quad (2)$$

$$x_{t,2} = a_2 + \Phi_{21}x_{t-1,1} + \Phi_{22}x_{t-1,2} + \Phi_{23}x_{t-1,3} + w_{t,2}, \quad (3)$$

$$x_{t,3} = a_3 + \Phi_{31}x_{t-1,1} + \Phi_{32}x_{t-1,2} + \Phi_{33}x_{t-1,3} + w_{t,3}. \quad (4)$$

The variable a_i is a k-vector of constants serving as the intercept of the model. Φ_{ij} is a time-invariant $(k \times k)$-matrix, and W_{ti} is a k-vector of error terms.

Each variable is a linear function of the lag 1 values for all variables in the set. In general, for a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable [31, 32].

4.1.2. eXtreme Gradient Boosting (XGBoost). XGBoost is an efficient and scalable implementation of gradient boosting framework by Friedman et al. [33] and Friedman et al. [34] The package includes an efficient linear model solver and tree learning algorithm [35]. XGBoost fits the new model to new residuals of the previous prediction and then minimizes the loss while adding the latest prediction [36]. What makes it unique is that it uses "a more regularized model formalization to control overfitting, which gives it better performance"—Tianqi Chen. XGBoost is used for supervised learning problems, where we use the training data x_i to predict a target variable y_i. After choosing the target variable y_i, we need to define the objective function to measure how well the model fits the training data, and it consists of two parts, training loss and regularization term, as follows:

$$\text{obj}(\theta) = L(\theta) + \Omega(\theta), \quad (5)$$

where θ denotes the parameters that we need to learn from data, L is the training loss function, and Ω is the regularization term. A common choice of L is the mean-squared error, which is given by:

$$L(\theta) = \sum_i (y_i - \overline{y_i})^2. \quad (6)$$

And the regularization term controls the complexity of the model, which helps us to avoid overfitting [27].

4.1.3. Support Vector Regression (SVR). The foundations of support vector machines (SVM) have been laid by Vapnik and Chervonenkis, and the methodology is gaining in popularity. The foundations of SVM that deal with classification problems are called support vector classification (SVC) and those of SVM that deal with modelling and prediction are called support vector regression (SVR)) [28].

Most real-world problems cannot be modelled by using linear forms [31]. SVR methodology allows to handle real-world problems. Here are some common kernels used in the SVR modelling [25]:

(1) Linear kernel: $x * y$

(2) Polynomial kernel: $[(x * x_i) + 1]d$

(3) Radial basis function (RBF): exp{-ϒ $|x - x_i|^2$}

4.1.4. K-Nearest Neighbors (kNN). K-nearest neighbors (kNN) is an efficient and intuitive method that has been used extensively for classification in pattern recognition [32]. It is a distance-based classifier, which implies that it implicitly presumes that the smaller the distance between two points is, the more similar they would be [37]. KNN classification algorithm is by far more popular than KNN regression [37]. In the KNN regression model, the derived information from the observed data is applied to forecast the amount of predicted variable in real time [38]. In other words, it estimates the response of a testing point X_t as an average of the responses of the k closest training points, $X_{(1)}, X_{(2)}, \ldots, X_{(k)}$, in the neighborhood of X_t [32]. Let $X = \{X_1, X_2, \ldots, X_M\}$ be a training data-set consisting of M training points, each of which possesses N features [32]. The Euclidean distance is used to calculate how close each training point X_i is to the testing point X_t using

$$d(x_t, x_i) = \sqrt{\sum_{n=1}^{N} (x_{t,n} - x_{i,n})^2}, \quad (7)$$

where N is the number of features, $x_{t,n}$ is the nth feature values of the testing point X_t, and $x_{i,n}$ is the nth feature values of the training point X_i. Some other methods are Manhattan, Minkowski, and Hamming distance methods.

4.2. Deep Learning Models

4.2.1. Long Short-Term Memory (LSTM). Long short-term memory neural network (LSTM NN) was initially introduced by Hochreiter and Schmidhuber (1997) [21]. The primary objectives of LSTM NN are to overcome the vanishing gradients problem of the standard recurrent neural network (RNN) when dealing with long-term dependencies [39]. Its features are especially desirable for traffic prediction in the transportation domain [40]. Figure 1 shows the architecture of long short-term memory cell. The core concept of LSTM is to be composed of recurrently connected memory blocks, each of which contains one or more memory cells, along with three multiplicative "gate" units: One input gate i_t with corresponding weight matrix W_{xi}, W_{hi}, W_{ci}, b_i; one forget gate f_t with corresponding weight matrix W_{xf}, W_{hf}, W_{cf}, b_f; one output gate o_t with corresponding weight matrix W_{xo}, W_{ho}, W_{co}, b_o. All of those gates are set to generate some degrees, by using current input x_i, the state h_{t-1} that the previous step generated, and the current state of this cell c_{t-1}, for the decisions whether to take the inputs, forget the memory stored before, and output the state generated later. Just as these following equations demonstrate [39]:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \quad (8)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \quad (9)$$

$$g_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + W_{cc}c_{t-1} + b_c), \quad (10)$$

$$c_t = i_t g_t + f_t c_{t-1}, \quad (11)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \quad (12)$$

$$h_t = o_t \tanh(c_t). \quad (13)$$

The network controls the flowing information by its sigmoid layer which outputs numbers between zero and one ($S(t) = 1/(1 + e^{-t})$).

4.2.2. Gated Recurrent Unit (GRU). The gated recurrent unit (GRU) architecture contains two gates: an update gate z_t decides how much the unit updates its activation or content and a reset gate r_t allows to forget the previously computed state [41]. The model is defined by the following:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \quad (14)$$

$$r_t = \sigma(W_r x_r + U_r h_{t-1} + b_r), \quad (15)$$

$$H_t = \tanh(W_H x_t + U_H (h_{t-1} r_t) + b_H), \quad (16)$$

$$h_t = z_t H_t + (1 - z_t) h_{t-1}, \quad (17)$$

where h_t represents the output state vector at time t, while H_t is the candidate state obtained with a hyperbolic tangent, x_t represents the input vector at time t, and the parameters of the model are W_z, W_r, W_H (the feed-forward connections), U_z, U_r, U_H (the recurrent weights), and the bias vectors b_z, b_r, b_H [42].

4.2.3. Convolutional Neural Network (CNN). Convolutional neural networks (CNNs) are analogous to traditional artificial neural networks (ANNs) as they are comprised of neurons that self-optimize through learning

Table 1: Categories of features.

Feature category (abbrv.)	Feature description
Usage level (F^x_{usage})	#Rented-cars shows the hourly variation in the number of car rentals in Chongqing between 2017 and 2019
Time (F^x_{time})	Take-date: car-sharing rental date Take-hour: car-sharing rental hour Take-month: month of the year for sharing rental Season: the seasons of the year (winter, spring, summer, autumn)
Weather (F^x_{weather})	Temperature, precipitation, humidity, wind speed, etc.

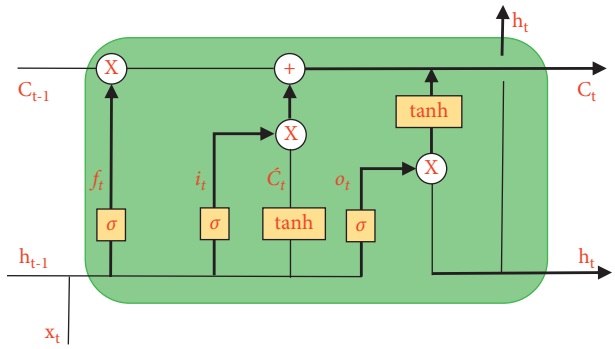


Figure 1: The architecture of long short-term memory cell.

[43]. They were initially developed for computer vision tasks; nevertheless, there have been a few recent studies applying them to time series forecasting tasks. CNNs are comprised of three types of layers including convolutional layers, pooling layers, and fully connected layers as shown in Figure 2.

The convolutional layer will determine the output of neurons connected to local regions of the input through the calculation of the scalar product between their weights and the region connected to the input volume [43]. There are two important techniques used in the convolutional layers to accelerate the training process: local connectivity and weight sharing. The two techniques are implemented using a filter with a specific kernel size which defines the number of nodes that share weights. Their usage decreases significantly the number of learned and stored weights and allows the network to grow deeper with fewer parameters. The pooling layer is usually incorporated between two successive convolutional layers [44].

The main idea of pooling is to reduce the complexity for further layers by down-sampling [45]. Max-pooling is one of the most common types of pooling methods as it performs better [44]. It consists in partitioning the image to subregion rectangles and only returning the maximum value of the inside of that subregion [45].

The fully connected layers are simply, feed-forward neural networks. They form the last few layers in the network [46]. The input to the fully connected layer is the output from the final pooling or convolutional layer, which is flattened and then fed into the fully connected layer in order to perform the same duties found in standard ANNs [43, 46].

4.2.4. CNN-LSTM Model. CNN-LSTM is a hybrid model built by combining CNN with LSTM for improving the accuracy of forecasting [47]. Figure 3 shows the architecture of the CNN-LSTM model. The model comprises of two main components: the first component consists of convolutional and pooling layers in which complicated mathematical operations are performed to filter the input data and extract the useful information. More specifically, the convolutional layers apply convolution operation between the raw input data and the convolution kernels, producing new feature values [48]. The convolution kernel can be considered as a window that contains coefficient values in a matrix form. This window slides all over the input matrix applying convolution operation on each subregion of it. The result of all these operations is a convolved matrix that represents a feature value.

The convolutional layers are usually followed by a nonlinear activation function and then a pooling layer. A pooling layer is a subsampling technique that extracts certain values from the convolved features and produces new matrices (i.e., summarized versions of the convolved features that are produced by the convolutional layer).

The second component exploits the generated features by LSTM which possesses the ability to learn long-term and short-term dependencies through the utilization of feedback connections and dense layers [48].

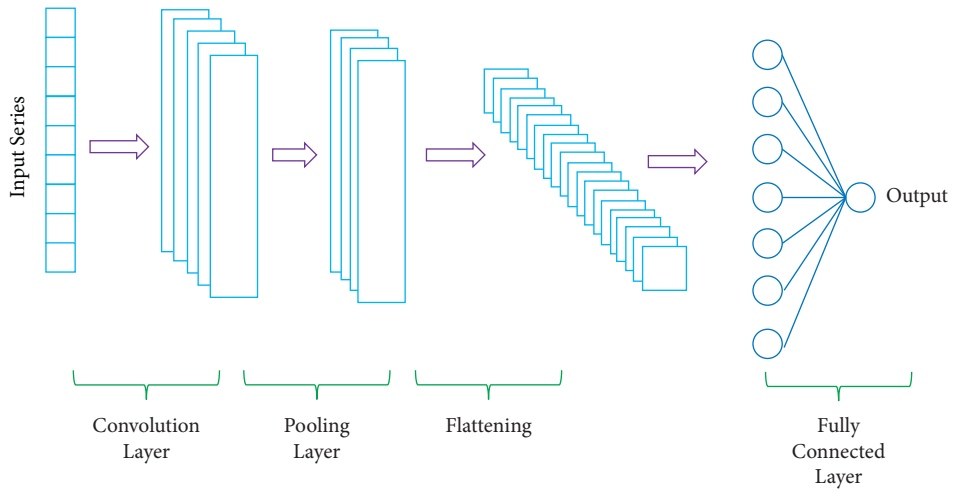


Figure 2: A simple architecture of convolutional neural network.

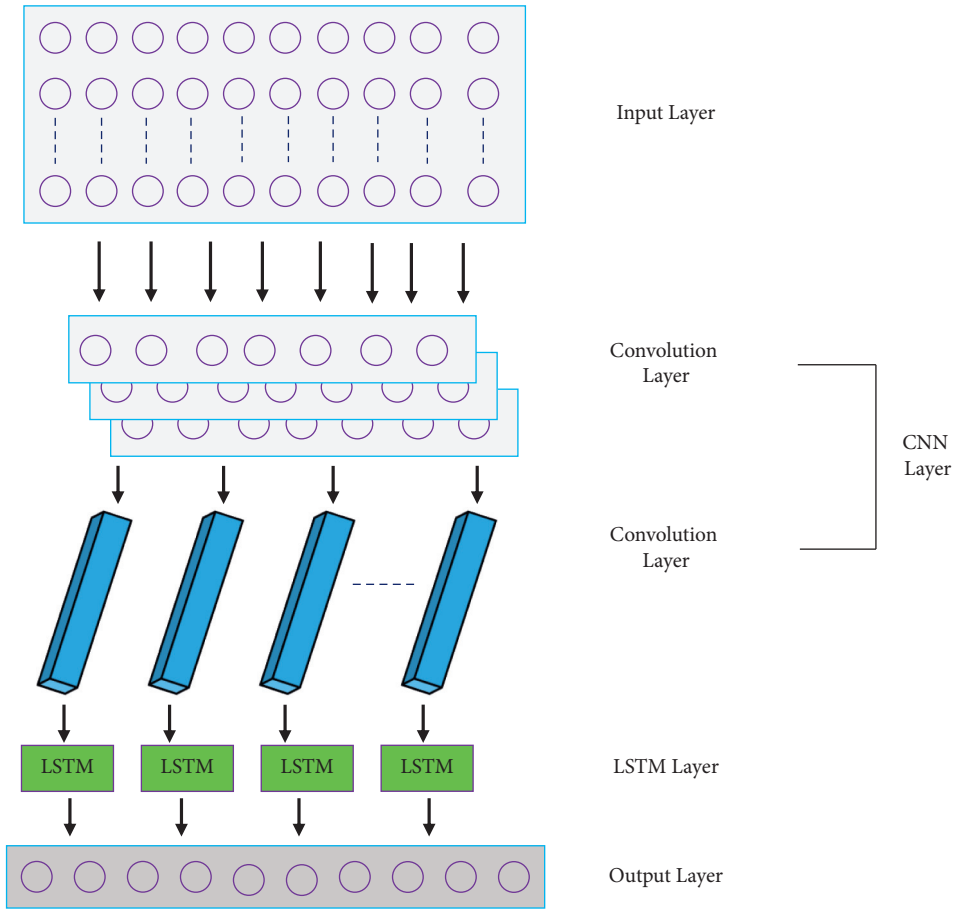


Figure 3: Architecture of CNN-LSTM.

4.2.5. Multilayer Perceptron (MLP). Multilayer perceptrons (MLPs) are deep artificial neural networks and are often applied to supervised learning problems [49]. As we can see from Figure 4, a multilayer perceptron consists of three types of layers: An input layer to receive the signal, an output layer that performs the required tasks to make a decision or prediction about the input, and an arbitrary number of hidden layers that are the true computational engine [49, 50]. In MLP, the data flow in the forward direction from the input to output layer, and the neurons are trained with the backpropagation learning algorithm on a set of input-output pairs. The training involves adjusting the parameters, or the weights and biases, of the model in order to minimize errors [49].

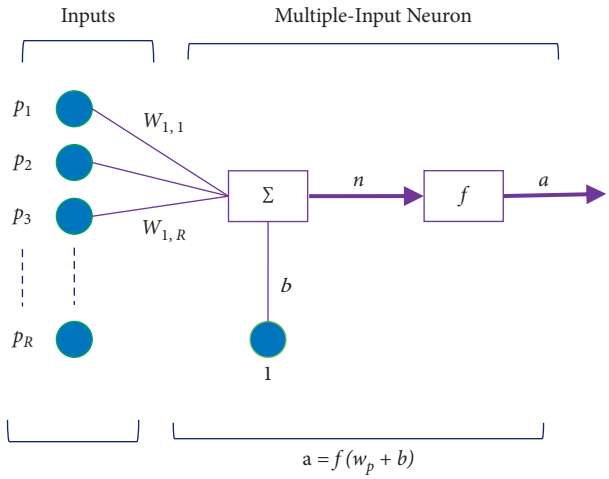


Figure 4: Multilayer perceptron architecture.

4.3. Vanishing Gradients Problem. Artificial neural networks often experience training problems due to vanishing and exploding gradients. The training problem is amplified exponentially especially in deep learning due to its complex artificial neural network architecture [51]. The vanishing gradient is one example of unstable behavior that may be encountered during the training with gradient-based methods (e.g., back propagation) [52]. The neural network's weights receive an update proportional to the partial derivative of the error function with respect to the current weight in each iteration of training. In some cases, the gradient tends to get smaller as we move backward through the hidden layers. This means that neurons in the earlier layers learn much more slowly than neurons in later layers preventing the weight from changing its value [52].

Several approaches exist to reduce this effect in practice, for example, through careful initialization, hidden layer supervision, and batch normalization [53]. In our work, batch normalization has been used, as it was effective in augmenting the performance of the deep neural network.

5. Framework of the Approach

Figure 5 shows the process of car-sharing usage prediction and factors investigation approach based on machine and deep learning models.

5.1. Collecting Chongqing's Car-Sharing Operator Data. Chongqing car-sharing operators' data set contains more than 1 M records of car-sharing usage over 860 parking lots, from January 1st, 2017, 00:00:00 to January 31st, 2019, 23:00:00. The initial records were obtained at different time intervals, and for study purposes, the data are aggregated by hours, days, and weeks, for the whole network.

5.2. Collecting Chongqing's Weather Data. The web crawling technology with selenium was used to extract the hourly Chongqing's weather conditions data from January 1st, 2017, 00:00:00 to January 31st, 2019, 23:00:00 [54].

5.3. Data Preprocessing. A data preprocessing is performed on the data set to improve the performance [55].

5.3.1. Processing Missing Values. Some values were missing in the data set from Chongqing's car-sharing operator. Due to its numerical meaning, we replaced the missing values by the mean of the previous and next hour number of car-sharing usage. This method yields better results compared to the removal of rows. The detailed calculation is shown as follows:

$$C^i_{j,k} = \frac{\left(C^i_{j,k-1} + C^i_{j,k+1}\right)}{2}, \tag{18}$$

where $C^i_{j,k}$ represents station i's missing value on the k^{th} hour of jth day of the year. After handling the missing values of Chongqing's car-rental operator data set, we merged it with Chongqing's weather conditions data set based upon dates and time.

5.3.2. Encoding the Categorical Data. Since the final data set, combining Chongqing's car-sharing operator and weather data, contains some categorical data such as weather condition and season, we converted the categorical data to numerical data using one-hot encoding method. It consists of representing each categorical variable with a binary vector that has one element for each unique label and marking the class label with a 1 and all other elements 0 [56].

5.3.3. Clustering Car-Sharing Parking Stations. To identify and understand the car-rental behaviors across stations and reveal the relationships between the time of a day and usage [57], we organized the parking stations with similar patterns into five distinct classes as follows:

(i) Class A: daily rented cars
(ii) Class B: frequently used cars
(iii) Class C: sometimes used cars
(iv) Class D: occasionally used cars

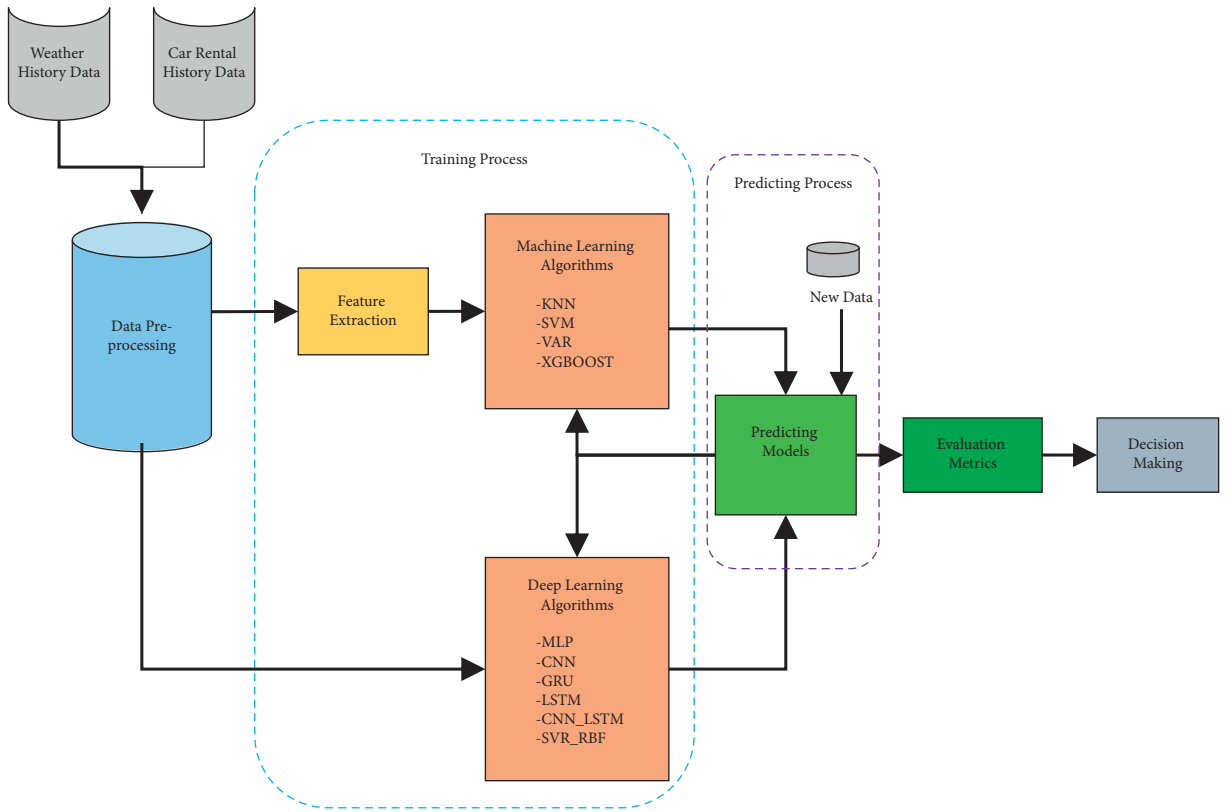


Figure 5: The framework of the proposed approach.

(v) Class E: unlike other parking stations, cars of this class are cars rarely used

where Class A, B, C, D, and E have different parking stations' IDs such as 16, 104, 6, 28, and 25, respectively. To simplify the large set of data and make it understandable, we used the grouped frequency table to cluster the 860 parking stations into five classes [58].

First, the usage frequencies of the parking stations were put in order, and then, the range was calculated as below [59]:

$$\text{range} = \text{largestfrequency value} - \text{smallest frequency value}. \tag{19}$$

Second, an approximate class width was calculated by dividing the range by the number of classes:

$$\text{class width} = \frac{\text{range}}{\text{number of classes}}. \tag{20}$$

The lowest usage frequency represents the first minimum data value.

Third, the next lower class value was calculated by adding the class width to the lowest usage frequency.

$$\text{lower class value} = \text{lowest usage frequency} + \text{class width}. \tag{21}$$

This step was repeated for the other minimum data values until the chosen classes number was created.

Fourth, the upper class limits (that are the highest values possible in the class) were calculated by subtracting 1 from the class width and adding that to the minimum data value.

$$\text{upper class limit} = \text{lowest usage frequency} + (\text{Class width} - 1). \tag{22}$$

Finally, the list of classes is obtained by including in each class the usage frequencies that are greater than the lower class value and smaller than the upper class limit.

5.4. Deseasonalization. Stationarity is an important concept for time series analysis. Some experts believe that neural networks are able to model seasonality directly and that no prior deseasonalization is required, whereas others believe the contrary. The results in [60] show that a prior data processing is required to construct a forecasting model. To test our time series, Augmented Dickey-Fuller Test (ADF Test) was conducted with machine learning and deep learning models to make predictions more accurate [61]. We employed differencing to remove seasonality from the nonstationary time series after the ADF Test [61, 62].

5.5. Scaling. The scaling phase is crucial to move the time series into a reasonable range. In our work, MinMaxScaler was used to scale each feature to a given range.

5.6. Splitting the Data Set. After handling the previously mentioned steps, we prepared our data set properly. We split the data between training and test sets. The training set starts from January 1st, 2017, to December 31st, 2018, and the test set from January 01st, 2019, to January 31st, 2019. Nested cross-validation with an outer loop equal to ten and an inner loop equal to five is used to calculate and to compare each model error. All models had to use the same validation procedure for consistency matters.

6. Experiments

6.1. Data Set. The experiments were performed on the preprocessed Chongqing's car-sharing operator data set combined with Chongqing's weather data set, to extract features that help to predict car usage, and to demonstrate the effectiveness of deep learning, more precisely the CNN-LSTM comparing to other models.

We have implemented the proposed models using a PC with an i7 Intel (R) Core™i7-7500U CPU running at 3.00 GHz and 8 GB RAM with the Windows 10 operating system, under the Python 3.7 development environment. The following packages were installed: TensorFlow 1.14.0; Keras 2.2.4-tf; Pandas 0.23.4; Sklearn 0.21.1; Numpy 1.18.1; Matplotlib 3.1.0; Statsmodels 0.10.1.

The hourly weather observations include time, temperature, humidity, wind speed, pressure, precipitation, and weather conditions. To have a basic idea about what weather conditions are typically associated with the city of Chongqing, we calculated the normalized frequency distributions of every weather condition (Table 2).

From Table 2, fair is the most prevalent meteorological condition in Chongqing, closely followed by fog, light rain, partly cloudy, and cloudy.

From Table 3, July and August are the hottest months of the year with an average temperature of 28°C, and January is the coldest month of the year with the temperature of 7°C.

6.2. Evaluation Metrics. The evaluation metrics are the measure that reflects how close the prediction matches the historical data. They are useful in comparing prediction methods on the same set of data [63].

6.2.1. Mean Absolute Error (MAE). MAE is calculated as the mean of the absolute predicted error values. The MAE is popular as it is easy to both understand and compute.

$$\text{MAE} = \text{mean}\left(\text{absolute}\left(\text{expected}_{\text{value}} - \text{predicted}_{\text{value}}\right)\right). \tag{23}$$

6.2.2. Mean Square Error (MSE). MSE is known for putting more weight on large errors. It is calculated as the average of the squared predicted error values.

$$\text{MSE} = \text{mean}\left(\left(\text{expected}_{\text{value}} - \text{predicted}_{\text{value}}\right)^2\right). \tag{24}$$

6.2.3. Root Square Mean Error (RMSE). The mean-squared error described above is in the squared units of the predictions. It can be transformed back into the original units of the predictions by taking the square root of the mean-squared error score [64]:

$$\text{RMSE} = \text{sqrt}(\text{MSE}). \tag{25}$$

The RMSE is chosen as an evaluation metrics because it penalizes large prediction errors more compared with mean absolute error (MAE).

6.2.4. Mean Absolute Percentage Error (MAPE). Mean absolute percentage error (MAPE) is one of the most widely used measures of forecast accuracy, due to its advantages of scale independence and interpretability [65]. It is calculated using the absolute error in each period divided by the observed values that are evident for that period and consequently averaging those fixed percentages [66]. MAPE indicates how much error in prediction is compared with the real value. The MAPE can be defined by the following formula:

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|, \tag{26}$$

where A_t is the actual value, F_t is the forecast value, and n denotes the number of fitted points.

6.2.5. Root-Mean-Squared Log Error (RMSLE). The root-mean-squared log error (RMSLE) is the RMSE of the log-transformed predicted and target values [54]. RMSLE only considers the relative error between predicted and actual values, and the scale of the error is nullified by the log-transformation [67]. The formula for RMSLE is represented as follows:

$$\text{RMSLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(p_i + 1\right) - \log\left(a_i + 1\right)\right)^2}, \tag{27}$$

where n is the total number of observations in the data set, p_i is the prediction of target, a_i is the actual target for i, and $\log(x)$ is the natural logarithm of x, $\log_e(x)$.

6.3. Experiments and Analysis. Many experiments were conducted on parking stations of different classes to extract which features improved the prediction and to predict the demands accordingly. Before applying the different models, some tests were performed on our time series.

6.3.1. Granger Causality Tests. The Granger causality test is a statistical hypothesis test for determining whether one time series is useful for predicting another [68]. In other words, it

Table 2: Chongqing's weather conditions normalized frequency distributions.

Weather condition	Frequency	Weather condition	Frequency
Fair	0.42962	Light drizzle	0.00135
Fog	0.13210	T-storm	0.00130
Light rain	0.12954	Partial fog	0.00055
Partly cloudy	0.12487	Heavy T-storm	0.00050
Cloudy	0.10191	Wintry mix	0.00025
Mostly cloudy	0.02330	Fair/windy	0.00020
Haze	0.01707	Thunder in vicinity	0.00015
Mist	0.01285	Heavy T-storm/windy	0.00010
Light rain shower	0.01155	Heavy rain	0.00010
Rain	0.00341	T-storm/windy	0.00010
Patches of fog	0.00236	Shallow fog	0.00005
Light rain with thunders	0.00236	Heavy rain shower/windy	0.00005
Thunder	0.00226	Showers in the vicinity	0.00005
Rain shower	0.00190	Light rain/windy	0.00005

Table 3: Chongqing's monthly temperature records.

Month average	Average temperature	Max/min average
January	7	5/10
February	9	7/12
March	14	11/17
April	19	15/23
May	22	18/25
June	25	21/29
July	28	24/33
August	28	24/33
September	24	20/27
October	18	16/21
November	14	11/16
December	9	7/11

is an approach that analyses the causal relationships between different variables of the time series.

After analysing the results shown in Table 4, we observed that all the given p value were smaller than the significance level (0.05). For example, the value of 0.0003 at (row 4) represents the p value of the Grangers causality test for temperature_x causing number_rented_cars_y (p value (0.0003) < significance level of 0.05).

From the results, we can infer that all the variables are good candidates to help predicting the number of rented cars.

6.3.2. Machine Learning Configuration

(1) eXtreme Gradient Boosting. A grid search was created for XGBoost model in order to locate the most optimal hyperparameters for the data set.

(2) Vector Autoregression. To select the right order of the VAR model, we iteratively fit increasing orders of the VAR model and picked the order that gave a model with the least AIC [29]. In our work, we have chosen the lag 4 model. Before predicting the future values of the target variable with VAR, we used serial correlation of residuals to check if our model is able to explain the patterns in the time series. The obtained scores in Table 5 show that our model is able to capture all the patterns without leftover.

(3) Support Vector Regression. The RBF kernel was chosen in this study for its good performance and advantages in time series prediction problem that has been proved in past researches [69, 70]. The penalty parameters are all tuned for the best using of a grid search method. Predictions are computed with optimal combination of cost and gamma parameters.

(4) K-Nearest Neighbors. The most important hyper-parameters for KNN are as follows: the number of neighbors (K) and the distance metric, and they determine the way in which the nearest neighbors are chosen.

To choose the number of neighbors (K), a grid search was performed. Moreover, the distance metric plays a crucial role in the nearest neighbor algorithm. Most of the papers in the references used the Euclidean distance. Reference [71] did a comparison between the Euclidean and Manhattan distance and found that statistically, no distance metric is significantly better than the other. The Euclidean distance has been selected since it is the most used in time series forecasting with KNN regression works.

6.3.3. Deep Learning Configuration

(1) Long Short-Term Memory and Gated Recurrent Unit Models. In this study, LSTM and GRU models have one neuron in the output layer. Both neural network models were designed with only one hidden layer, a number of 50 epochs were chosen, and the learning rate was set to 0.01. The input and hidden neurons for the GRU model were the same as those for the LSTM model and were set to 41 and 50, respectively [56].

(2) Convolutional Neural Network. CNN is one of the most successful deep learning methods, and its network structures include 1D CNN, 2D CNN, and 3D CNN. 1D CNN is used in this paper, and it can be well applied to time series analysis. The detailed process of the 1D CNN is described as follows:

In our experiment, we used various layers including one convolutional layer with a kernel size of 2 and 64 filters, a max-pooling layer, along with that a rectified linear unit

Table 4: Granger causality test results.

Variables	Number_rented_cars
Day_of_week	0.0004
Take_month	0.0289
Season	0.0001
Temperature	0.0003
Humidity	0.0000
Wind_speed	0.0000
Pressure	0.0000
Precipitation	0.0000
Weather_condition	0.0004
Weekdays_weekend_x	0.0003

Table 5: Check for serial correlation of residuals.

Columns	Value
Number_rented_cars	2.03
Day_of_week_x	2.0
Take_month_x	2.0
Season_x	2.0
Temperature_x	2.03
Humidity_x	2.01
Wind_speed_x	2.0
Pressure_x	2.03
Precipitation_x	2.0
Weather_condition_x	2.01
Weekdays_weekend_x	2.0

(ReLU) activation function is applied in the convolutional and output layers. To minimize the mean-squared error, the gradient descent backpropagation algorithm and Adam optimizer were used; following that, a dropout rate of 0.5 was employed to avoid the overfitting [72], and a number of 70 epochs were used for training the model.

(3) CNN-LSTM. In our implementation, we utilized a version of the CNN-LSTM model that consists of two one-dimensional convolutional layers with a kernel size equal to 5, 32, and 64 filters, respectively, followed by a max-pooling layer, a LSTM layer with 50 units, and a dense layer of 32 neurons [48], In order to avoid overfitting during training, the dropout was adjusted to 0.2. A number of 100 epochs were chosen to train the model.

(4) Multilayer Perceptron Regressor. MLP was applied on our multivariate time series, with ReLU as an activation function to train the regression model. Three hidden layers were used with a number of 50, 35, and 10 hidden neurons, respectively. The training was performed for 50 epochs for MLP with a learning rate of 0.01.

6.4. Prediction Result. For the good organization of the paper, and not being redundant in our explanations, we only discussed the result analysis of the class "A" as other classes exhibit the same behavior and lead to the same conclusion.

To perform the comparison while fitting the models with different features, the results of only CNN-LSTM for deep learning models and XGBoost for machine learning models are described in our analysis. In the same way, each of the applied models had the analysis of the same results.

In our investigation, metrics such as MAE, MSE, RMSE, MAPE, and RMSLE are used in respective order to make the comparison. Note that the smallest errors are shown in bold text in Tables 6–13.

6.4.1. Univariate Time Series

(1) Machine Learning Models. As shown in Table 6, XGBoost reduced the error by (93.14%, 99.14%, 93.36%, 69.17%, 93.76%) against VAR; (93.69%, 99.48%, 94.17%, 83.09%, 94.70%) against SVM; and (93.93%, 99.65%, 94.44%, 84.17%, 95.05%) against KNN.

(2) Deep Learning Models. As it can be seen from Table 7, CNN-LSTM yielded the best results and had smaller evaluation error compared to all other models. CNN-LSTM decreased the evaluation error by (44.19%, 48.37%, 34.44%, 1.03%, 24.14%) against LSTM; (51.91%, 57.97%, 35.15%, 11.71%, 40.21%) against GRU; (61.15%, 58.58%, 79.38%, 37.38%, 38.76%) against CNN; and (67.43%, 95.75%, 69%, 88.24%, 61.37%) against MLP.

6.4.2. The Effect of Weekends Information. We might not forget or underestimate the impact of weekends on the prediction while building prediction models with time series data. They are significant since they can add peculiarity to the outcomes.

(1) Machine Learning Models. From Table 8, it can be noticed that using the XGBoost model with weekend features enhanced the improvement rate of (1.14%, 0.62%, 1.08%, 0.42%, 0.24%) compared to univariate time series.

XGBoost outperformed VAR, SVM, and KNN. It reduced evaluation error at the rate of (93.20%, 99.19%, 93.34%, 69.04%, 93.76%) contrary to VAR; (93.74%, 99.47%, 94.14%, 82.98%, 94.46) contrary to SVM; and (94.00%, 99.65%, 94.37%, 84.04%, 95.04%) contrary to KNN.

(2) Deep Learning Models. The addition of weekend features improved prediction accuracy as it can be observed from Table 9, where adding the weekend feature to the univariate time series of class "A" improved the results with a rate of (2.4%, 4.95%, 2.54%, 4.23%, 0.43%) for CNN-LSTM.

Regarding models' comparison, CNN-LSTM achieved the best results with an improvement rate of (43.75%, 46.52%, 28, 50%, 0.88%, 37.98%) against LSTM; (46.47%, 57.96%, 35.16%, 4.87%, 35.45%) against GRU; (40.38%, 59.34%, 67.17%, 48.70%, 40.56%) against CNN; and (65.92%, 89.22%, 67.79%, 75.46%, 60.88%) against MLP.

6.4.3. The Effect of Weather Information. In addition to rental information in the city of Chongqing, we can leverage weather data to improve prediction at different times of the day.

Table 6: Machine learning for univariate time series evaluation results.

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.07272**	**0.00975**	**0.09868**	**0.30398**	**0.07189**
VAR	1.06025	1.20552	1.4851	0.98598	1.15297
SVM	1.15297	1.86949	1.69395	1.7977	1.3575
KNN	1.19834	2.77452	1.77467	1.91974	1.45087

(a) Class A

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.09125**	**0.02252**	**0.14942**	0.18352	**0.06025**
VAR	0.45087	0.36788	0.60653	**0.14527**	0.09057
SVM	0.72289	1.04231	1.01927	1.14777	0.960432
KNN	0.76353	1.48786	1.18645	1.25337	0.72234

(b) Class B

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.1179**	**0.03489**	**0.18489**	**0.29007**	**0.1179**
VAR	0.43931	0.34352	0.5861	0.50319	0.43931
SVM	0.81499	0.99441	0.99721	1.1251	0.59046
KNN	0.83915	1.86306	1.36483	1.22767	1.30815

(c) Class C

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.03654**	**0.00314**	**0.05391**	**0.04028**	**0.03299**
VAR	0.43459	0.34597	0.58819	0.18361	0.43459
SVM	0.54589	0.69401	0.83355	1.16683	0.54795
KNN	0.61745	1.14995	1.07216	1.18775	0.90985

(d) Class D

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.04051**	**0.00285**	**0.05338**	**0.00683**	**0.04051**
VAR	0.08543	0.08782	0.29143	0.14278	0.08659
SVM	0.14529	0.11864	0.34444	0.53728	0.08543
KNN	0.27331	0.13296	0.56446	1.9495	0.47325

(e) Class E

(1) Machine Learning Models. Table 10 shows that applying XGBoost reduced the MAE by 2.68%; the MSE by 3.18%; the RMSE by 0.25%; the MAPE by 6.84%; and the RMSLE by 1.56% compared to the univariate time series results. Correspondingly, it reduced the MAE by 1.56%; the MSE by 2.58%; the RMSE by 0.33%; the MAPE by 6.98%; and the RMSLE by 1.32% relative to the weekend results.

XGBoost outperformed the other models as it decreased the error values by (93.28%, 99.21%, 93.35%, 76.26%, 93.72%) compared to VAR; (93.49%, 99.41%, 94.15%, 94.20%, 94.05%) compared SVM; and (94.08%, 99.56%, 94.26%, 94.54%, 95.10%) compared to KNN.

(2) Deep Learning Models. From Table 11, our findings show that the CNN-LSTM model was (37.97%, 37.59%, 35.82%, 5.49%, 23.42%) more accurate than LSTM; (39.82%, 57.73%, 34.93%, 2.62%, 38.70%) more accurate than GRU; (40.21%, 58.84%, 58.46%, 43.80%, 40.76%) more accurate than CNN; and (65.25%, 82.77%, 52.96%, 76.67%, 60.67%) more accurate than MLP.

Moreover, it was observed after a thorough examination that integrating weather data as a feature improved the prediction accuracy. When the CNN-LSTM model was applied to class "A" data with weather features, it reduced the MAE by 2.67%; the MSE by 6.04%; the RMSE by 3.03%; the MAPE by 11.10%; and the RMSLE by 5.02% compared to the univariate time series results. Similarly, it also reduced the MAE by 0.28%; the MSE by 1.16%; the RMSE by 0.51%; the MAPE by 7.17%; and the RMSLE by 5.43% compared to the weekend's results.

Table 7: Deep learning for univariate time series evaluation results.

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.04083**	**0.00364**	**0.06035**	**0.08108**	**0.04843**
LSTM	0.07316	0.00705	0.09206	0.08192	0.06384
GRU	0.08491	0.00866	0.09306	0.09151	0.081
CNN	0.10512	0.00879	0.2927	0.14614	0.07909
MLP	0.12537	0.08568	0.19468	0.77838	0.12537

(a) Class A

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.00616**	**0.00714**	**0.0074**	**0.05434**	0.00616
LSTM	0.04043	0.04718	0.04977	0.08423	**0.00404**
GRU	0.04404	0.05247	0.04989	0.18194	0.03956
CNN	0.03939	0.08383	0.09927	0.77724	0.13109
MLP	0.1261	0.05533	0.28953	0.88942	0.1261

(b) Class B

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.02172**	**0.00126**	**0.03547**	0.09764	**0.02864**
LSTM	0.03332	0.00243	0.06397	**0.07606**	0.03059
GRU	0.0338	0.00252	0.04974	0.65142	0.03546
CNN	0.12709	0.00983	0.14989	0.73808	0.06779
MLP	0.13366	0.0851	0.29171	0.91345	0.12709

(c) Class C

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.00913**	0.00852	**0.01545**	0.54338	**0.00927**
LSTM	0.01596	**0.00621**	0.02222	**0.49018**	0.01717
GRU	0.01663	0.08175	0.02151	0.74992	0.01612
CNN	0.12113	0.08763	0.28592	0.53462	0.12113
MLP	0.12187	0.15624	0.29603	0.80945	0.12187

(d) Class D

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.01213**	**0.00198**	**0.04331**	**0.00727**	**0.02345**
LSTM	0.02244	0.00466	0.13236	0.02329	0.02687
GRU	0.03553	0.00475	0.06827	0.33448	0.03836
CNN	0.03746	0.08376	0.06893	0.67174	0.0418
MLP	0.1227	0.08708	0.28941	0.71862	0.1227

(e) Class E

6.4.4. Combined Effect of Historical, Weekends, and Weather Information

(1) Machine Learning Models. The evaluation error of XGBoost was lower as shown in Table 12. It reduced the error by (93.35%, 99.28%, 93.76%, 73.71%, 80.32%) relative to VAR; by (93.57%, 99.47%, 94.47%, 94.21%, 93.81%) relative to SVM; and by (94.12%, 99.59%, 94.62%, 94.54%, 95.16%) relative to KNN.

Our findings show that when all features were used together, XGBoost performed better with rate of (3.96%, 12.92%, 6.65%, 6%, 2.85%). Similarly, it also performed (2.85%, 12.38%, 6.72%, 6.14%, 2.62%) better than when history combined with weekend features were used and (1.31%, 10.6%, 6.41%, 5.48%, 1.31%) better than when history combined with weather features were used.

Table 8: Machine learning for historical data with weekends data evaluation results.

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.07189**	**0.00969**	**0.09876**	**0.30527**	**0.07172**
VAR	1.05771	1.19652	1.482066	0.98598	1.14989
SVM	1.14782	1.83432	1.68569	1.79397	1.2942
KNN	1.19799	2.77371	1.75573	1.91257	1.44645

(a) Class A

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.09057**	**0.02244**	**0.15007**	0.18352	**0.05771**
VAR	0.44645	0.36339	0.60281	**0.13388**	0.09025
SVM	0.72168	1.03896	1.01929	1.14777	0.78527
KNN	0.76351	1.40766	1.18639	1.24379	0.7216

(b) Class B

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.1175**	**0.03476**	**0.18457**	**0.28931**	**0.1175**
VAR	0.43805	0.34159	0.58445	0.50319	0.43805
SVM	0.81479	0.99398	0.99698	1.1251	0.57837
KNN	0.83915	1.86244	1.36463	1.22597	0.93895

(c) Class C

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.03299**	**0.00314**	**0.05544**	**0.03984**	**0.02959**
VAR	0.43452	0.34574	0.588	0.18361	0.43452
SVM	0.54589	0.69409	0.83352	1.16626	0.54589
KNN	0.61696	1.14954	1.07216	1.18884	0.89971

(d) Class D

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.01878**	**0.00222**	**0.04717**	**0.00583**	**0.01878**
VAR	0.08543	0.08481	0.29123	0.14278	0.08527
SVM	0.14422	0.11862	0.34426	0.53462	0.08543
KNN	0.27324	0.13282	0.36444	1.94985	0.37284

(e) Class E

(2) Deep Learning Models. From Table 13, our findings reveal that when CNN-LSTM was fitted with all features together, it performed better with rate of (1.41%, 2.34%, 1.28%, 8.82%, 7.98%) than when only history features were used; likewise, it was (4.04%, 8.24%, 4.28%, 8.17%, 12.60%) better than when history combined with weekend features were used, and (1.68%, 3.47%, 1.79%, 8.33%, 12.97%) better than when history combined with weather features were used.

In comparison to all other models, CNN-LSTM has much lower evaluation error. It decreased the error by (36.80%, 23.39%, 31.18%, 46.46%, 29.66%) against LSTM, by (37.03%, 53.09%, 35.41%, 74.91%, 37.21%) against GRU, by (36.939%, 58.25%, 31.52%, 88.65%, 42.98%) against CNN, by (38.79%, 59.17%, 36.02%, 88.52%, 31.97%) when compared to MLP.

6.5. Comparison of the Results. One of our objectives is to compare the accuracy of various machine learning and deep learning models in predicting the future number of car-sharing transactions. After analysing the results obtained in our study, the following are our findings:

6.5.1. Machine Learning. First, with regard to the results obtained with the machine learning models and after the comparison based on the evaluation metrics, it shows that XGBoost gave the best results, followed by VAR, SVR, and KNN. The XGBoost model had several advantages in model prediction such as complete feature extraction, good fitting effect, and high prediction accuracy.

Second, the SVR prediction series failed to capture random and nonlinear patterns. Hence, it failed to perform well, while XGBoost and VAR forecast series were able to capture random walk patterns.

Third, KNN performed the worse compared to the other machine learning models because of the high number of inputs.

6.5.2. Deep Learning. After comparison of results, we can deduce that CNN-LSTM generated better outcomes followed by LSTM, GRU, CNN, and MLP.

The hybrid CNN-LSTM model yielded better performance on the strength of its capability in supporting very long input sequences that can be read as subsequences by the CNN model and then formed together by the LSTM model.

Besides the CNN-LSTM model, the long short-term memory model achieved good results on account of its ability to learn patterns from sequenced data more effectively.

The key difference between the gated recurrent unit model and the long short-term memory model is that GRU is less complex than LSTM, as it only has two gates (i.e., reset and update) while LSTM has three gates (including

Table 9: Deep learning for historical data with weekends data evaluation results.

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.03985**	0.00346	**0.05882**	**0.07765**	**0.04864**
LSTM	0.07084	0.00647	0.08226	0.07834	0.07843
GRU	0.07445	0.00823	0.09072	0.08235	0.07535
CNN	0.06685	0.00851	0.17917	0.16054	0.08183
MLP	0.11695	**0.0321**	0.18264	0.31643	0.12436
(a) Class A					
	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.00549**	**0.00712**	**0.0062**	**0.02987**	**0.00549**
LSTM	0.03908	0.0343	0.01044	0.03156	0.02043
GRU	0.03774	0.01249	0.04989	0.32077	0.03437
CNN	0.03362	0.03393	0.09543	0.41409	0.07012
MLP	0.11776	0.01509	0.18421	0.74706	0.11776
(b) Class B					
	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.02107**	**0.00126**	**0.03544**	**0.07989**	**0.0274**
LSTM	0.02945	0.00229	0.05796	0.09897	0.02942
GRU	0.03309	0.00249	0.04929	0.56722	0.03524
CNN	0.11495	0.00257	0.13197	0.31063	0.03386
MLP	0.12668	0.03283	0.18118	0.6956	0.11495
(c) Class C					
	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.00841**	0.00543	**0.01444**	0.32857	**0.00848**
LSTM	0.00947	**0.00524**	0.02185	**0.16148**	0.0155
GRU	0.01529	0.0302	0.02081	0.28902	0.00966
CNN	0.10715	0.03437	0.17378	0.67375	0.10715
MLP	0.11859	0.03674	0.18727	0.8023	0.11859
(d) Class D					
	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.00943**	**0.00188**	**0.04145**	**0.00517**	**0.01213**
LSTM	0.02032	0.00464	0.05192	0.01976	0.02659
GRU	0.03531	0.00471	0.06658	0.32383	0.03742
CNN	0.03576	0.03396	0.06862	0.49758	0.03509
MLP	0.11089	0.03635	0.18428	0.70503	0.11089
(e) Class E					

input, output, and forget). By comparing the two models using the different evaluation metrics, it can be concluded that the LSTM model had good memory for longer sequences as compared to the GRU model, and it outperformed in the tasks requiring modelling the long-distance relations.

The CNN produced quite impressive results because of the ability of its convolutional layer in identifying patterns between the time steps. Contrary to the LSTM model, the CNN model is not recurrent, and it can only train the data that are inputted by the model at a particular time step.

Unlike the other models, the multilayers perceptron model performed worse. The model received inputs and didn't treat them as sequence data, which led to temporal dependencies and sequence patterns loss.

6.5.3. Comparison between Machine Learning and Deep Learning Models. It can be inferred from the obtained results that the deep learning models outperformed all the machine learning time series prediction models. From the different results of the different models, we noticed that CNN-LSTM gave the best performance measures and achieved the most accurate prediction results. Furthermore, Figure 6 shows that the dashed line of predicted values almost coincides with the one of real values, which proves that the hybrid CNN-LSTM model generated good results.

Figure 7 shows a comparison between the two best machine learning and deep learning models of Class A, and it can be noticed that CNN-LSTM slightly outperformed the XGBoost model.

Table 10: Machine learning for historical data with weather data evaluation results.

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.07077**	**0.00944**	**0.09843**	**0.10385**	**0.07077**
VAR	1.05296	1.192	1.480543	0.43748	1.12674
SVM	1.08711	1.60693	1.68354	1.79101	1.1896
KNN	1.19562	2.14944	1.7146	1.90031	1.44379
(a) Class A					
	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.09003**	**0.02233**	**0.1498**	0.14143	**0.05296**
VAR	0.44379	0.36095	0.60079	**0.13328**	0.09003
SVM	0.72164	1.03892	1.02093	1.14312	0.69041
KNN	0.76347	1.4106	1.18769	1.23748	0.72168
(b) Class B					
	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.11705**	0.03418	**0.18679**	**0.27058**	**0.11705**
VAR	0.43726	**0.34063**	0.58364	0.41371	0.43726
SVM	0.80968	0.98371	0.99181	1.12363	0.54982
KNN	0.83909	1.86221	1.36451	1.22429	0.83915
(c) Class C					
	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.02945**	**0.00307**	**0.05401**	**0.03076**	**0.02951**
VAR	0.43389	0.34523	0.58756	0.1595	0.43389
SVM	0.54583	0.69474	0.83312	1.16663	0.54594
KNN	0.61329	1.14941	1.07194	1.18616	0.81517
(d) Class D					
	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.01711**	**0.00215**	**0.04635**	**0.00462**	**0.01711**
VAR	0.08469	0.08491	0.29139	0.1393	0.08519
SVM	0.13407	0.11862	0.34422	0.52147	0.08471
KNN	0.2732	0.13283	0.36446	1.94885	0.37281
(e) Class E					

6.5.4. The Computational Time. The computational time of various machine and deep learning models can be found in the following tables:

Table 14 shows that XGBoost has faster computational time while SVM is the more demanding. For deep learning models, Table 15 shows that the computational time of CNN-LSTM is bigger than the LSTM, GRU, CNN, and MLP models.

Machine learning models exhibit faster computational time, while deep learning models take a longer time because of their high number of parameters and their complex mathematical formulas.

Table 11: Deep learning for historical data with weather data evaluation results.

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.03974**	**0.00342**	**0.05852**	**0.07208**	**0.046**
LSTM	0.06407	0.00548	0.09118	0.07627	0.06007
GRU	0.06604	0.00809	0.08993	0.07832	0.07504
CNN	0.06647	0.00831	0.14089	0.13935	0.07765
MLP	0.11436	0.01985	0.12441	0.30888	0.11695

(a) Class A

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.00449**	**0.00505**	**0.00639**	**0.02946**	**0.00449**
LSTM	0.03302	0.01164	0.00767	0.02954	0.03342
GRU	0.00712	0.01249	0.04986	0.10656	0.03386
CNN	0.03309	0.01212	0.0738	0.27765	0.04365
MLP	0.08257	0.01421	0.1101	0.32499	0.08257

(b) Class B

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.02061**	**0.00111**	**0.03333**	**0.06524**	**0.02502**
LSTM	0.0248	0.00206	0.04984	0.07459	0.02668
GRU	0.02939	0.00247	0.04974	0.28738	0.03112
CNN	0.09692	0.00248	0.04984	0.22379	0.034
MLP	0.12534	0.01729	0.1315	0.94313	0.09692

(c) Class C

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.0076**	**0.00354**	**0.0142**	0.20089	**0.00822**
LSTM	0.01172	0.00418	0.02168	0.27274	0.01112
GRU	0.01203	0.01016	0.02078	**0.17646**	0.0118
CNN	0.06987	0.01151	0.10078	0.45685	0.06987
MLP	0.07914	0.02542	0.10539	0.5951	0.07914

(d) Class D

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.00598**	**0.00172**	**0.04102**	**0.00427**	**0.00943**
LSTM	0.01912	0.00461	0.04713	0.01093	0.03641
GRU	0.0316	0.00468	0.06813	0.19444	0.03412
CNN	0.03492	0.01408	0.06839	0.29567	0.03855
MLP	0.09834	0.01498	0.11865	0.62538	0.09834

(e) Class E

Table 12: Machine learning for combined features evaluation results.

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.06984**	**0.00849**	**0.09212**	**0.10335**	**0.06984**
VAR	1.0498	1.17777	1.47572	0.39307	0.3548
SVM	1.08607	1.59434	1.66544	1.78393	1.12886
KNN	1.18772	2.08258	1.7107	1.89279	1.44327

(a) Class A

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.08978**	**0.0223**	**0.14396**	**0.11954**	**0.0498**
VAR	0.44327	0.36056	0.60046	0.13314	0.08978
SVM	0.7216	1.03716	1.01919	1.14185	0.61201
KNN	0.76191	1.41458	1.18536	1.23693	0.72143

(b) Class B

	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.11671**	**0.03406**	**0.18643**	**0.25127**	**0.11671**
VAR	0.43631	0.33916	0.58238	0.38869	0.43631
SVM	0.80446	0.97407	0.98695	1.12311	0.27791
KNN	0.839	1.86212	1.36414	1.22258	0.53914

Table 12: Continued.

	MAE	MSE	RMSE	MAPE	RMSLE
			(c) Class C		
	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.01955**	**0.00291**	**0.05102**	**0.02952**	**0.0295**
VAR	0.43306	0.34416	0.58665	0.15809	0.43306
SVM	0.54572	0.69482	0.83307	1.16535	0.54581
KNN	0.61314	1.14905	1.071	1.18552	0.43926
			(d) Class D		
	MAE	MSE	RMSE	MAPE	RMSLE
XGBoost	**0.01691**	**0.00204**	**0.04636**	**0.00572**	**0.01591**
VAR	0.08465	0.08488	0.29134	0.13918	0.07416
SVM	0.1173	0.11817	0.34376	0.52012	0.0847
KNN	0.27281	0.13251	0.16539	1.94874	0.17384
			(e) Class E		

Table 13: Deep learning for combined features evaluation results.

	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.03918**	**0.00334**	**0.05777**	**0.01527**	0.13296
LSTM	0.06199	0.00436	0.08394	0.02852	**0.04233**
GRU	0.06222	0.00712	0.08944	0.06087	0.06018
CNN	0.06213	0.008	0.08436	0.13451	0.06741
MLP	0.06401	0.00818	0.09029	0.13296	0.07424
			(a) Class A		
	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.00237**	**0.00101**	**0.00315**	**0.01554**	**0.0029**
LSTM	0.0029	0.00151	0.00364	0.02853	0.03108
GRU	0.00365	0.00245	0.04954	0.10156	0.03227
CNN	0.03153	0.00894	0.05426	0.18847	0.00337
MLP	0.06972	0.00401	0.09456	0.18756	0.06972
			(b) Class B		
	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.01766**	**0.00108**	**0.03282**	**0.01725**	**0.02415**
LSTM	0.00252	0.00142	0.04932	0.07215	0.02534
GRU	0.02914	0.00232	0.09914	0.151	0.02541
CNN	0.07412	0.00243	0.05017	0.17203	0.02956
MLP	0.06779	0.01018	0.10765	0.42891	0.07412
			(c) Class C		
	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.00701**	**0.00147**	**0.01351**	**0.14796**	0.00818
LSTM	0.01175	0.0042	0.02069	0.28051	**0.00795**
GRU	0.00783	0.00443	0.02069	0.15548	0.01206
CNN	0.0584	0.00519	0.0783	0.28875	0.0584
MLP	0.04972	0.00613	0.07204	0.59483	0.04972
			(d) Class D		
	MAE	MSE	RMSE	MAPE	RMSLE
CNN-LSTM	**0.0027**	**0.0016**	0.04046	**0.0009**	**0.0027**
LSTM	0.01707	0.00454	**0.03545**	0.0059	0.02613
GRU	0.03159	0.00443	0.06789	0.19402	0.03386
CNN	0.0314	0.00412	0.06813	0.28588	0.03419
MLP	0.04576	0.00488	0.06422	0.426525	0.04576
			(e) Class E		

6.6. Features Investigations. All the used time series prediction models showed that the prediction results were more accurate when we used the different features categories, namely, the past usage levels feature (e.g., number of car-sharing transactions), temporal information features (e.g., season, weekdays/weekend), and environmental condition features (e.g., temperature, humidity, wind speed, pressure, precipitation, weather conditions) together.

It also showed that the environmental condition features dominated other features, and it was followed by the

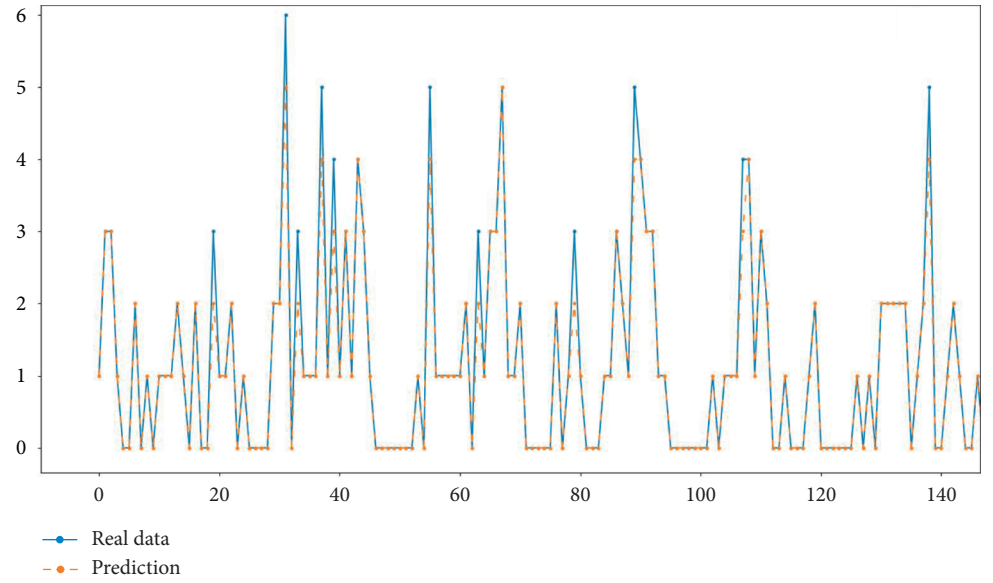


Figure 6: Comparison of the predicted value and the real value for CNN-LSTM.

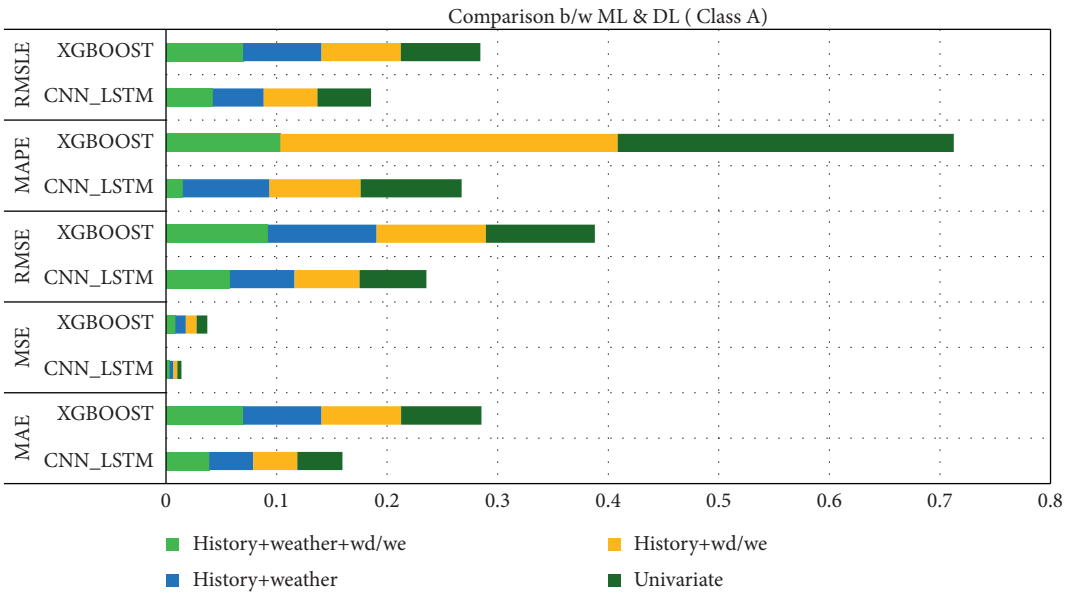


Figure 7: Comparison between best machine learning and deep learning models.

Table 14: Computational time for machine learning models.

Deep learning	Time (s)
XGBoost	10.2203
VAR	34.5798
SVM	65.2421
KNN	20.4125

Table 15: Computational time for deep learning models.

Deep learning	Time (s)
CNN-LSTM	872.2856
LSTM	543.7666
GRU	542.7666
CNN	395.3688
MLP	262.8907

temporal information features. We got the worst results when using only the number of car-sharing transactions based on the data from a Chongqing's car-sharing operator.

7. Conclusions

This research paper, through applying different machine learning and deep learning models to multivariate time series, aims to predict the car usage and to investigate the factors that help to improve the predictions' accuracy.

The evaluation of the different machine learning and deep learning models with MAE, MSE, RMSE, MAPE, and RMSLE reveals that the hybrid model (CNN-LSTM) gives substantially smaller errors as compared to the standalone used models. The experimental results show that the

utilization of the CNN-LSTM model on the number of car-sharing transactions, along with environmental conditions and temporal information features together, yields the highest prediction accuracy. The principal idea of the hybrid model is to efficiently amalgamate the advantages of two deep learning techniques. It exploits the ability of convolutional layers for extracting useful knowledge and learning the internal representation of time series data as well as the effectiveness of long short-term memory (LSTM) layers for remembering events for short and very long time.

Furthermore, through our experimental analysis, we conclude that even though LSTM models constitute an efficient choice for car-sharing time series prediction, their usage along with additional convolutional layers provides a significant boost in enhancing the forecasting performance. Although CNN-LSTM requires high search time due to its sensitivity to various hyperparameters and its high complexity, it shows the highest forecasting accuracy and the best performance.

All the results of the used models confirm that the car-rental usage is more sensitive to environmental conditions than temporal information that means the impact of weather on car-rental transportation deserves more attention at research. However, our work is limited to temporal features. Future studies can extend on adding more features such as the time span of data, spatiotemporal variables, and expand the model to consumers' habits [73–76].

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 study.

Acknowledgments

This research work was funded by the Beijing Municipal Natural Science Foundation (Grant no. 4212026) and National Science Foundation of China (Grant no. 61772075). The authors are thankful to them for their financial support.

References

[1] S. Petit, "World vehicle population rose 4.6% in 2016," *Wards Intelligence*, vol. 17, Jun. 08, 2020, https://wardsintelligence.informa.com/WI058630/World-Vehicle-Population-Rose-46-in-2016.

[2] T. Russell, "Why cars will always Be A main form of transportation," Jun. 08, 2020, https://www.forbes.com/sites/quora/2019/07/30/why-cars-will-always-be-a-main-form-of-transportation/#70d8cab72ab8.

[3] S. Formentin, A. G. Bianchessi, and S. M. Savaresi, "On the prediction of future vehicle locations in free-floating car sharing systems," *2015 IEEE Intelligent Vehicles Symposium (IV)*, vol. 2015, pp. 1006–1011, 2015.

[4] L. Cagliero, S. Chiusano, E. Daraio, and P. Garza, "CarPredictor: forecasting the number of free floating car sharing vehicles within restricted urban area," in *Proceedings of the 2019 IEEE International Congress on Big Data, BigData Congress*, pp. 72–76, Milan, Italy, July 2019.

[5] S. Hosseini, *Data-driven Time Series Demand Forecasting of Car-Sharing Services*, The case study of DriveNow in Munich, Germany, 2018.

[6] A. de Lorimier and A. M. El-Geneidy, "Understanding the factors affecting vehicle usage and availability in carsharing networks: a case study of communauto carsharing system from montréal, Canada," *International Journal of Sustainable Transportation*, vol. 7, no. 1, pp. 35–51, Jun. 2012.

[7] J. Burrell, "How the machine 'thinks': understanding opacity in machine learning algorithms," *Big Data & Society*, vol. 3, no. 1, p. 205395171562251, Jan. 05, 2016.

[8] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP Journal on Applied Signal Processing*, Springer International Publishing, vol. 2016, no. 1, p. 67, Dec. 01, 2016.

[9] G. Battineni, N. Chintalapudi, and F. Amenta, ""Machine learning in medicine: performance calculation of dementia prediction by support vector machines (SVM)," Informatics Med," *Unlocked*, vol. 16, Article ID 100200, Jan. 2019.

[10] E. E. Vos, P. S. Francois Luus, C. J. Finlay, and B. A. Bassett, "A generative machine learning approach to RFI mitigation for radio astronomy," *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, vol. 2019, pp. 1–6, 2019.

[11] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2063–2079, 2018.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Aug. 15, 2020, http://image-net.org/challenges/LSVRC/2015/.

[13] A. Agrawal et al., "VQA: visual question answering," Aug. 15, 2020.

[14] H. Peng, J. Li, Y. Song, and Y. Liu, "Incrementally learning the hierarchical softmax function for neural language models.," Aug. 15, 2020, http://www.aaai.org.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[16] J. Li, H. Peng, L. Liu et al., "Graph CNNs for urban traffic passenger flows prediction," in *Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. SmartWorld/UIC/ATC/ScalCom/CBDCo*, pp. 29–36, Guangzhou, China, October 2018.

[17] C. Morency, M. Trépanier, B. Agard, B. Martin, and J. Quashie, "Car sharing system: what transaction datasets reveal on users' behaviors," in *Proceedings of the IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 284–289, Enshi, China, September 2007.

[18] Y. Liu, C. Liu, X. Lu, M. Teng, H. Zhu, and H. Xiong, "Point-of-Interest demand modeling with human mobility patterns," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 947–955, Halifax NS Canada, August 2017.

[19] C. Boldrini, R. Bruno, and M. Conti, "Characterising demand and usage patterns in a large station-based car sharing system," *2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, vol. 2016, pp. 572–577, 2016.

[20] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6164–6173, Apr. 2009.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] J.-X. Xu and J. S. Lim, "A new evolutionary neural network for forecasting net flow of a car sharing system," in *Proceedings of the 2007 IEEE Congress on Evolutionary Computation*, pp. 1670–1676, Singapore, Sep. 2007.

[23] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.

[24] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proceedings of the 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, pp. 153–158, Chengdu, China, December 2015.

[25] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, May 2015.

[26] X. Mo, Y. Xing, and C. Lv, "Interaction-aware trajectory prediction of connected vehicles using CNN-LSTM networks," *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, vol. 2020, pp. 5057–5062, 2020.

[27] "Introduction to boosted trees — xgboost 1.2.0-SNAPSHOT documentation," Jun. 08, 2020, https://xgboost.readthedocs.io/en/latest/tutorials/model.html.

[28] K. Kavaklioglu, "Modeling and prediction of Turkey's electricity consumption using Support Vector Regression," *Applied Energy*, vol. 88, no. 1, pp. 368–375, Jan. 2011.

[29] "Vector autoregression (VAR) - comprehensive guide with examples in Python - machine learning plus," Jun. 08, 2020, https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/.

[30] "11.2 Vector Autoregressive models VAR(p) models | STAT 510," Jun. 08, 2020, https://online.stat.psu.edu/stat510/lesson/11/11.2.

[31] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 4, pp. 276–281, 2004.

[32] C. Hu, G. Jain, P. Zhang, C. Schmidt, P. Gomadam, and T. Gorka, "Data-driven method based on particle swarm optimization and k-nearest neighbor regression for estimating capacity of lithium-ion battery," *Applied Energy*, vol. 129, pp. 49–55, 2014.

[33] J. Friedman, R. Tibshirani, and T. Hastie, "Additive logistic regression: a statistical view of boosting," *The Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[34] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[35] T. Chen and T. He, "xgboost: eXtreme Gradient Boosting," 2020, https://medium.com/m/global-identity?redirectUrl=https%3A%2F%2Ftowardsdatascience.com%2Fxgboost-extreme-gradient-boosting-how-to-improve-on-regular-gradient-boosting-5c6acf66c70a.

[36] J. Brownlee, "A gentle introduction to XGBoost for applied machine learning," Jun. 08, 2020, https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/.

[37] "Distance metrics and K-nearest neighbor (KNN) | by luigi fiori | medium," Aug. 23, 2021.

[38] F. Modaresi, S. Araghinejad, and K. Ebrahimi, "A comparative assessment of artificial neural network, generalized regression neural network, least-square support vector regression, and K-nearest neighbor regression for monthly streamflow forecasting in linear and nonlinear conditions," *Water Resources Management*, vol. 32, no. 1, pp. 243–258, 2018.

[39] P. Zhou, W. Shi, J. Tian et al., "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany, August 2016.

[40] J. Jian Zheng, C. Cencen Xu, Z. Ziang Zhang, and X. Xiaohua Li, "Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network," in *Proceedings of the 2017 51st Annual Conference on Information Sciences and Systems, CISS 2017*, pp. 1–6, Baltimore, MD, USA, May 2017.

[41] T.-T.-H. Le, J. Kim, and H. Kim, "Classification performance using gated recurrent unit Recurrent Neural Network on energy disaggregation," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 105–110, Jeju, Korea (South), Jul. 2016.

[42] J. Li, X. Wang, Y. Zhao, and Y. Li, "Gated recurrent unit based acoustic modeling with future context,," 2018, https://arxiv.org/abs/1805.07024.

[43] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," pp. 1–11, 2015, http://arxiv.org/abs/1511.08458.

[44] "Convolutional layer - an overview | ScienceDirect topics," Aug. 23, 2021, https://www.sciencedirect.com/topics/engineering/convolutional-layer.

[45] S. Albawi, T. A. M. Mohammed, and S. Alzawi, "Layers of a convolutional neural network," *IEEE*, 2017.

[46] "A comprehensive guide to convolutional neural networks — the ELI5 way | by sumit saha | towards data science," Aug. 23, 2021, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[47] T. Li, M. Hua, and X. Wu, "A hybrid CNN-LSTM model for forecasting particulate matter (PM2.5)," *IEEE Access*, vol. 8, pp. 26933–26940, 2020.

[48] I. E. Livieris, E. Pintelas, and P. Pintelas, "A CNN-LSTM model for gold price time-series forecasting," *Neural Computing & Applications*, vol. 32, no. 23, pp. 17351–17360, 2020.

[49] "Multilayer perceptron - an overview | ScienceDirect topics," Aug. 23, 2021, https://www.sciencedirect.com/topics/computer-science/multilayer-perceptron.

[50] "A beginner's guide to multilayer perceptrons (MLP) | pathmind," Aug. 23, 2021, https://wiki.pathmind.com/multilayer-perceptron.

[51] H. H. Tan and K. H. Lim, "Vanishing gradient mitigation with deep learning neural network optimization," in *Proceedings of the 2019 7th Int. Conf. Smart Comput. Commun. ICSCC 2019*, pp. 1–4, Sarawak, Malaysia, June 2019.

[52] "Vanishing gradient problem - Wikipedia," Aug. 23, 2021, https://en.wikipedia.org/wiki/Vanishing_gradient_problem.

[53] J. Kolbusz, P. Rozycki, and B. M. Wilamowski, "The study of architecture MLP with linear neurons in order to eliminate the "vanishing gradient" problem," *Artificial Intelligence and Soft Computing*, vol. 10245, no. 2013, pp. 97–106, 2017.

[54] *Root Mean Squared Log Error - Machine Learning with Spark - Second Edition [Book]*, https://www.oreilly.com/library/view/machine-learning-with/9781785889936/4317943c-9452-4013-99b9-d267a2820b23.xhtml, Aug. 23, 2021.

[55] X. Ma, R. Cao, and Y. Jin, "Spatiotemporal clustering analysis of bicycle sharing system with data mining approach," *Information*, vol. 10, no. 5, pp. 1–14, 2019.

[56] "3 ways to encode categorical variables for deep learning," Jun. 09, 2020, https://machinelearningmastery.com/how-to-prepare-categorical-data-for-deep-learning-in-python/.

[57] G. Petneházi, "Recurrent neural networks for time series forecasting," *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.

[58] "How to create a grouped frequency table," Jun. 09, 2020, https://sciencing.com/create-grouped-frequency-table-5531910.html.

[59] "Grouped frequency distribution," Jun. 09, 2020, https://www.mathsisfun.com/data/frequency-distribution-grouped.html.

[60] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *European Journal of Operational Research*, vol. 160, no. 2, pp. 501–514, 2005.

[61] M. Theses and B. T. Ojemakinde, "Trace: Tennessee research and creative exchange support vector regression for non-stationary time series," 2006, https://trace.tennessee.edu/utk_gradthes/1756.

[62] S. Arslankaya and V. Öz, "Time series analysis of sales quantity in an automotive company and estimation by artificial neural networks," *Sakarya University Journal of Science*, vol. 22, no. 3, p. 1, 2018.

[63] C. Chen, J. Twycross, and J. M. Garibaldi, "A new accuracy measure based on bounded relative error for time series forecasting," *PLoS One*, vol. 12, no. 3, p. e0174202, Article ID 0174202, 2017.

[64] J. Brownlee, "Time series forecasting performance measures with Python," Jun. 09, 2020, https://machinelearningmastery.com/time-series-forecasting-performance-measures-with-python/.

[65] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *International Journal of Forecasting*, vol. 32, no. 3, pp. 669–679, 2016.

[66] U. Khair, H. Fahmi, S. A. Hakim, and R. Rahim, "Forecasting error calculation with mean absolute deviation and mean absolute percentage error," *Journal of Physics: Conference Series*, vol. 930, no. no. 1, p. 012002, 2017.

[67] O. P. Abimbola, A. R. Mittelstet, T. L. Messer, E. D. Berry, S. L. Bartelt-Hunt, and S. P. Hansen, "Predicting *Escherichia coli* loads in cascading dams with machine learning: an integration of hydrometeorology, animal density and grazing pattern," *The Science of the Total Environment*, vol. 722, p. 137894, 2020.

[68] W. Wei, "Granger causality test - an overview | ScienceDirect topics," Jun. 09, 2020, https://www.sciencedirect.com/topics/social-sciences/granger-causality-test.

[69] Y. Ding, X. Song, and Y. Zen, "Forecasting financial condition of Chinese listed companies based on support vector machine," *Expert Systems with Applications*, vol. 34, no. 4, pp. 3081–3089, 2008.

[70] S. S. Eslamian, S. A. Gohari, M. Biabanaki, and R. Malekian, "Estimation of monthly pan evaporation using artificial neural networks and support vector machines," *Journal of Applied Sciences*, vol. 8, no. 19, pp. 3497–3502, 2008.

[71] F. Martínez, M. P. Frías, M. D. Pérez, and A. J. Rivera, "A methodology for applying k-nearest neighbor to time series forecasting," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 2019–2037, 2019.

[72] I. Koprinska, D. Wu, and Z. Wang, ""Convolutional neural networks for energy time series forecasting," proc. Int. Jt. Conf," *Neural Networks*, vol. 2018, 2018.

[73] J. Kang, K. Hwang, and S. Park, "Finding factors that influence carsharing usage: case study in seoul," *Sustainability*, vol. 8, no. 8, p. 709, Jul. 2016.

[74] J. Firnkorn and M. Müller, "What will be the environmental effects of new free-floating car-sharing systems? The case of car2go in Ulm," *Ecological Economics*, vol. 70, no. 8, pp. 1519–1528, Jun. 2011.

[75] S. de Luca and R. Di Pace, "Modelling users' behaviour in inter-urban carsharing program: a stated preference approach," *Transportation Research Part A: Policy and Practice*, vol. 71, pp. 59–76, Jan. 2015.

[76] J. Yuan, Y. Zheng, and X. Xie, ""Discovering regions of different functions in a city using human mobility and POIs," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax NS Canada, June 2020.

Hindawi
Complexity
Volume 2021, Article ID 1782354, 15 pages
https://doi.org/10.1155/2021/1782354

Review Article

Complex Urban Systems: Challenges and Integrated Solutions for the Sustainability and Resilience of Cities

Riccardo Gallotti,[1] **Pierluigi Sacco**,[1,2] **and Manlio De Domenico**[1]

[1]*Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo, TN, Italy*
[2]*IULM University, Via Carlo Bo. 1, 20143 Milano, Italy*

Correspondence should be addressed to Riccardo Gallotti; rgallotti@gmail.com, Pierluigi Sacco; pierluigi.sacco@iulm.it, and Manlio De Domenico; mdedomenico@fbk.eu

Received 2 July 2020; Revised 9 March 2021; Accepted 4 September 2021; Published 5 October 2021

Academic Editor: Rosa M. Benito

For decades, from design theory to urban planning and management, from social sciences to urban environmental science, cities have been probed and analyzed from the partial perspective of single disciplines. The digital era, with its unprecedented data availability, is allowing for testing old theories and developing new ones, ultimately challenging relatively partial models. Our community has been in the last years providing more and more compelling evidence that cities are complex systems with emergent phenomena characterized by the collective behavior of their citizens who are themselves complex systems. However, more recently, it has also been shown that such multiscale complexity alone is not enough to describe some salient features of urban systems. Multilayer network modeling, accounting for both multiplexity of relationships and interdependencies among the city's subsystems, is indeed providing a novel integrated framework to study urban backbones, their resilience to unexpected perturbations due to internal or external factors, and their human flows. In this paper, we first offer an overview of the transdisciplinary efforts made to cope with the three dimensions of complexity of the city: the complexity of the urban environment, the complexity of human cognition about the city, and the complexity of city planning. In particular, we discuss how the most recent findings, for example, relating the health and wellbeing of communities to urban structure and function, from traffic congestion to distinct types of pollution, can be better understood considering a city as a multiscale and multilayer complex system. The new challenges posed by the postpandemic scenario give to this perspective an unprecedented relevance, with the necessity to address issues of reconstruction of the social fabric, recovery from prolonged psychological, social and economic stress with the ensuing mental health and wellbeing issues, and repurposing of urban organization as a consequence of new emerging practices such as massive remote working. By rethinking cities as large-scale active matter systems far from equilibrium which consume energy, process information, and adapt to the environment, we argue that enhancing social engagement, for example, involving citizens in codesigning the city and its changes in this critical postpandemic phase, can trigger widespread adoption of good practices leading to emergent effects with collective benefits which can be directly measured.

1. Introduction

Cities offer one of the most challenging test beds for any complexity-oriented modeling approach. The reason is simple: they present a multiscale structure integrating a multitude of social and technological subsystems. While being large enough to be amenable to macromodeling, at the same time, they are not large enough to be exclusively approached at that scale, thus raising ambition for detailed microstructural analysis and understanding. For these reasons, the modeling and analysis of cities sit naturally at the mesoscale, at the edge between micro and macro, and offer an ideal environment for the development of "statistical mechanics" of human interaction.

On the basis of these premises, it is paradoxically not surprising that the most authoritative and celebrated account of how cities "work," which has informed a countless number of different approaches and analysis of all sorts, is Jane Jacobs' book "The Death and Life of Great American Cities" [1], which is essentially an autoethnography of the

experience of the city par excellence, New York City, that is, an approach that could not be farther away from conventional scientific standards: subjective observations about a single city through time. Despite its lack of "hard scientific" method, the book has been so influential and was celebrated to become a sort of conceptual map for all scientists aiming at building solid scientific explanations and analyses of the urban dynamics. Such influence stems from its unique capacity of summing up, through the author's gaze, so many different, subtly related aspects of the essence of cities' functioning and living. Not incidentally, one of its major insights among many is that cities thrive only if they are able to maintain their own form of highly idiosyncratic complexity. If they fall for structural oversimplification and loss of diversity, they decay and possibly eventually die. Traditional top-down planning practices have not successfully passed the urban complexity test, due to their inability to credibly address the mutability of social interactions in urban settings through their rigid schemes and their consequent tendency to micromanage environmental complexity rather than enable its generative potential [2]. Moreover, Jacobs' lesson reminds us that the complexity of the city is not only about the manifold aspects of the urban environment and their interrelations but also about the complexity of our own mental representation of the city. Jacobs' insights are also about what planning means for the city and about the subtle balance between self-organization and intelligent design, and here too there is enormous inspiration for readers who are accustomed to think in complexity science terms. There are, therefore, at least three different dimensions of complexity that one should keep in mind thinking of the city: the complexity of the urban environment, the complexity of human cognition about the city, and the complexity of city planning. A comprehensive approach to urban complexity should be able to encompass all three and even more so in the complex postpandemic scenario with which all cities will have to cope in the coming years.

2. Modeling Urban Complexity

Research on the complexity of urban environments has a long tradition, and although its roots are difficult to trace back, a fundamental text is Christopher Alexander's "A City Is Not a Tree" [3]. In this short, insightful essay, later developed into a book [4], Alexander makes use of biological analogies to explore the inherent geometrical properties of urban organization, largely prefiguring the complexity science of the next two decades. As shown in his later book [5], Alexander clearly understands the relationship between the emergent macrostructures of the city and the microlevel of building construction patterns and their compositional space grammar, thus characterizing architectural building rules as subject to adaptive pressures. Alexander then goes further on [6] to identify vernacular architecture as a self-organizing system of space organization which reflects an extremely complex system of socioenvironmental cognition and constitutes an ideal bridge between the complexity science of urban environments and that of their mental representations. Finally, in the monumental 4-book series, "The Nature of Order" [7–10], Alexander arrives at an all-encompassing evolutionary synthesis of human and biological organization structures, where he finally investigates issues such as why certain human settlements have more "life" in them than others. One might think of a city's "liveliness" in terms of an ensemble of emergent structural properties that result from the coevolution of built environments and human interaction, following a logic that closely resembles that of biological design.

It is from these bases that the literatures on shape grammars [11], space syntax [12], and their inevitable confluences [13] take off, building the premises of a computational approach to urban form and function. The marriage with the concurrently upcoming complexity science would not only be inevitable [14, 15] but necessary, with the urban dimension becoming one of the natural testing grounds for fractal [16], agent-based [17], and cellular automata modeling [18] of the emergent order properties of multiscale systems, once again naturally coalescing into a unified theory of urban complexity [19].

Complexity-based approaches to urban issues have since then proliferated to practically every sphere of city life [20, 21]: transportation systems [22], utilities and infrastructure [23, 24], pollution [25], and crime [26], just to limit ourselves to a few examples. These approaches have allowed the development of much deeper insights into the nature and effects of structural interdependencies across urban environments [27], also deriving from the unique tension between the general nonlinear effects typical of all urban environments with local, specific factors and dynamics [28, 29]. The rapidly increasing availability of large databases and the big data revolution in social and urban sciences has further boosted this tendency, leading to a new wave of complexity-oriented urban science which is likely still in its early phase [30], and it is consequently challenged by the need of developing proper analytical methods for the extraction of reliable behavioral information [31, 32]. There is therefore room to expect that the application of complexity science to the modeling, analysis, and understanding of urban systems is a long-term scientific endeavor rather than a transitory phase and that this will have profound effects on many dimensions of science, society, and the economy.

From a complexity science perspective, cities cannot be simply viewed as structures in space but also as functional systems of flows and networks [33]. In 1961, Gilbert used a special class of networks, namely, random geometric graphs, to model the structure of spatially embedded networks and the effects of spatial constraints on the system [34]. The core idea is to consider spatially distributed nodes representing, for instance, geographic areas which are connected to each other if their distance is within some spatial scale used as a reference. This class of models is desirable for studying the structure and the function of complex systems like a city, consisting of areas connected by transportation infrastructures [35].

Nowadays, network modeling and analysis of urban ecosystems is a widely adopted framework to cope with the

complexity of cities and of their societies at different scales [36]. The analysis of the Boston underground transportation system through the lens of global and local efficiency in information flows revealed that its underlying logic of construction is, in fact, a small-world principle [37]. Complex networks have been used for geographical modeling and, by means of combined cellular models of land and behavior, it has been shown that they provide a compelling framework for growth dynamic that is consistent with large-scale regularities, such as fractality or power-law scaling relations [38]. The analysis of the dual graph representation, where roads and junctions are mapped into nodes and edges, respectively, of six urban street networks characterized by different patterns and historical roots revealed their unique connectivity patterns with respect to nongeographic systems [39]. Network science has been used for spatial analysis of the topology of Singapore inferred from human-generated data, by identifying city hubs, centers, and other elements which are essential to characterize urban interactions. Results from longitudinal analysis suggest that Singapore is rapidly developing towards the designed polycentric urban form [40]. An example of network analysis in action is shown in Figure 1.

Models and analytical tools borrowed from or inspired by complexity science are proliferating, providing convincing evidences of their application to real cities [42–46], from human mobility [47–50] and traffic congestion [51–55] to energy consumption [56], air quality [57, 58], climate [59], and health and wellbeing [60–63], as well as accessibility to important facilities like hospitals [64]. The city is seen as a huge complex system which grows and expands [65, 66] and whose spatial organization [67, 68] dynamically experiences a transition from monocentric to polycentric [69, 70].

The relevance of complexity modeling tools to understand urban ecosystems ignited an unprecedented deluge of open and crowdsourced information about the topology of the city and its fundamental constituents [71, 72], as well as about its function, directly related to the behavior of its inhabitants inferred from the data they generate, such as phone call detail records [47, 73–77], transactions [78–83], GPS trajectories [49, 84–88], and geo-tagged social media [89–101].

2.1. The Multilayer Structure of the City. Very recently, it has been suggested that a new level of complexity characterizes cities. In fact, accounting for multiplexity [102, 103] of transportation, that is, multimodality [104] (see Figure 2), and interdependencies [105, 106], that is, structural and functional relationships with other systems (see Figure 3), allows, on the one hand, to gain new insights about the functioning of a city and the complex society it hosts. On the other hand, it allows to better understand its resilience to targeted policies, such as infrastructural interventions, or to random perturbations, such as unexpected failures in transportation or energy networks, as well as catastrophic events.

Multilayer models for transportation infrastructures [104, 107, 108] group together connections from the same transportation mode, assign them to a layer, and sometimes couple layers with each other, the last one depending on whether information about time or economic cost to move across layers is available or not [109].

Different perspectives can be adopted. For instance, one might assign different submodes (e.g., distinct lines of the tube) to different layers [110, 111] or group them together within the same layer encoding a unique means of transportation [104]. The two approaches, applied to the backbone of 9 different cities in Europe, from small towns to megacities, highlighted different vulnerabilities and provided a framework for testing improvements [112]. The analysis of topological pathways across layers revealed, for instance, that London's public transportation is designed to minimize redundancies [113]. The London underground network exhibits patterns that are not observed in other systems, like social networks. The calculation of clustering, that is, the tendency to triadic closure, revealed mechanisms to avoid redundant connections, with 3-mode triangles more frequent than 2-mode triangles which, in turn, are more frequent than single-mode triangles [114]. The analysis of the interdependence between the street network and the subways of London and New York City unraveled similar mechanisms, with the underground network acting as a decentralizing force which pushes congestion towards the end of its lines. It has been found that uneven spatial distributions of accessibility might emerge if the speed of subways is increased without a systemic view of the city [115].

While transportation networks with their multiplexity and interdependency are fundamental to enhance our understanding of the city, it is similarly of utmost importance to include other systems, for example, sociotechnical and ecological ones, which are in turn shaped by urban activities, in the integrated picture required to maintain some control on city sustainability and resilience.

2.2. The Multilayer Dynamics of the City. Tightly related to the modeling of the urban backbone is the analysis of its flows. In fact, the city is a complex system consisting of geographic areas which integrate local flows of goods and people within the overall urban ecosystem. It is not surprising that modeling and understanding both individual and collective human mobility patterns [47, 49, 75, 116] play an important role in our understanding of the city: the identification and quantification of functional patterns, for example, daily mobility motifs [117], can be used to drive the development of transportation systems and enhance urban infrastructures [118].

The abundance of available urban data is already being used to better understand one of the most important urban problems: traffic congestion. Only recently the traditional assumption that people follow the minimum cost path [121] has been challenged [122], showing that routing mechanisms accounting for the complexity of the city have

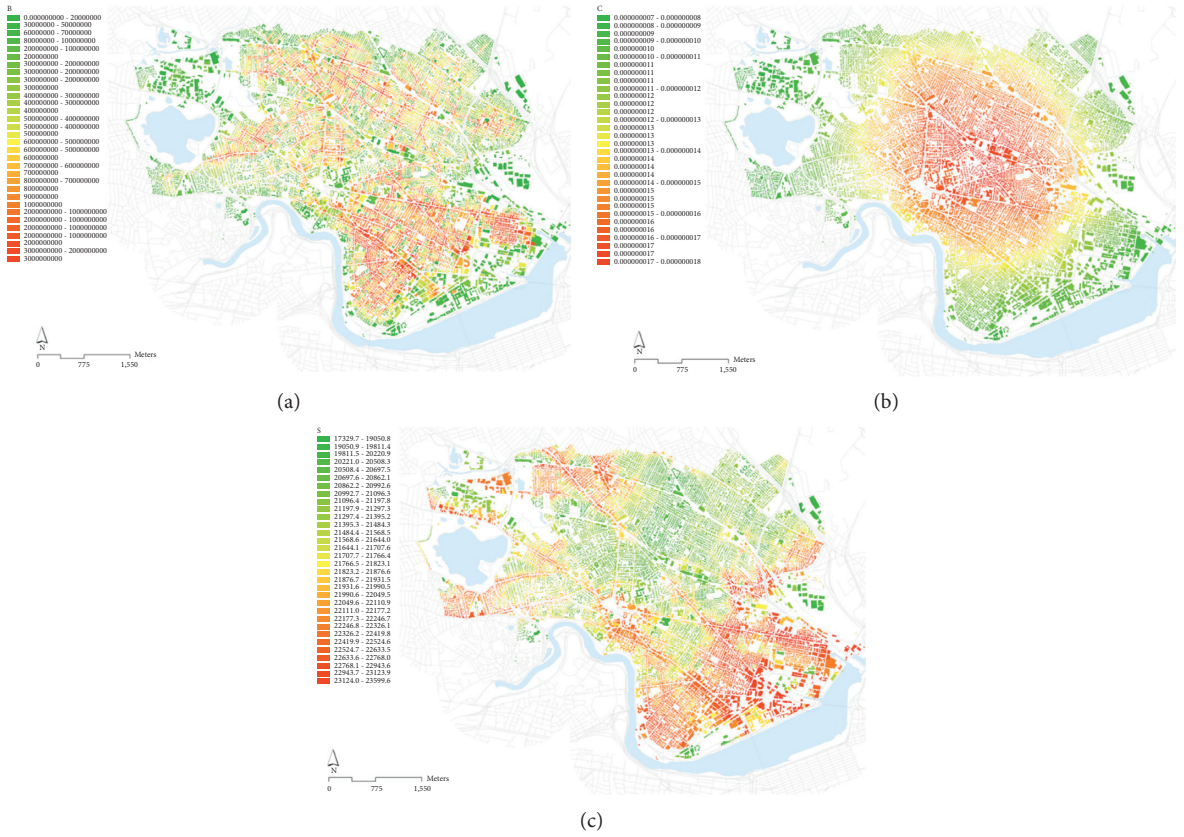


Figure 1: Analysis of an urban network. The road network of Cambridge & Somerville (MA) as seen through the lens of network science. Nodes represent buildings and color encodes different descriptors quantifying the relevance of each node with respect to different criteria: (a) betweenness, (b) closeness, and (c) straightness centrality, respectively. Figure reproduced with permission from [41], all rights reserved.

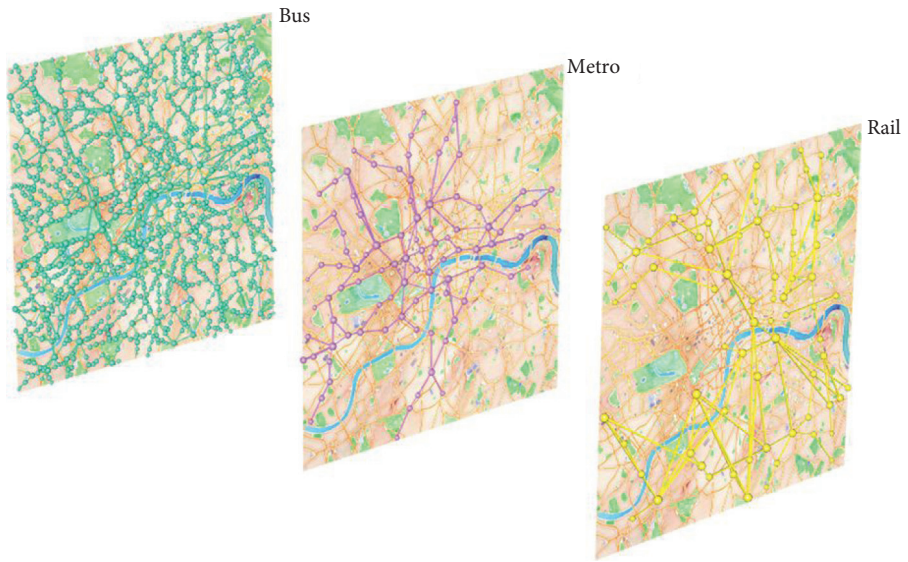


Figure 2: Multiplex urban transportation networks. Multimodal transportation network of London, consisting of three layers encoding different types of infrastructural connections: bus, tube, and rail. Figure reproduced with permission from [107], all rights reserved.

huge potential in mitigating traffic [53], especially in hypothetical smart cities [123]. Stochastic theories of urban growth, validated on US and OECD empirical data, show that congestion shapes cities, revealing intriguing relationships between mobility patterns and scaling laws, such as the dependence on population size of the total number of miles driven daily, the total length of the road network, the total traffic delay, the total consumption of

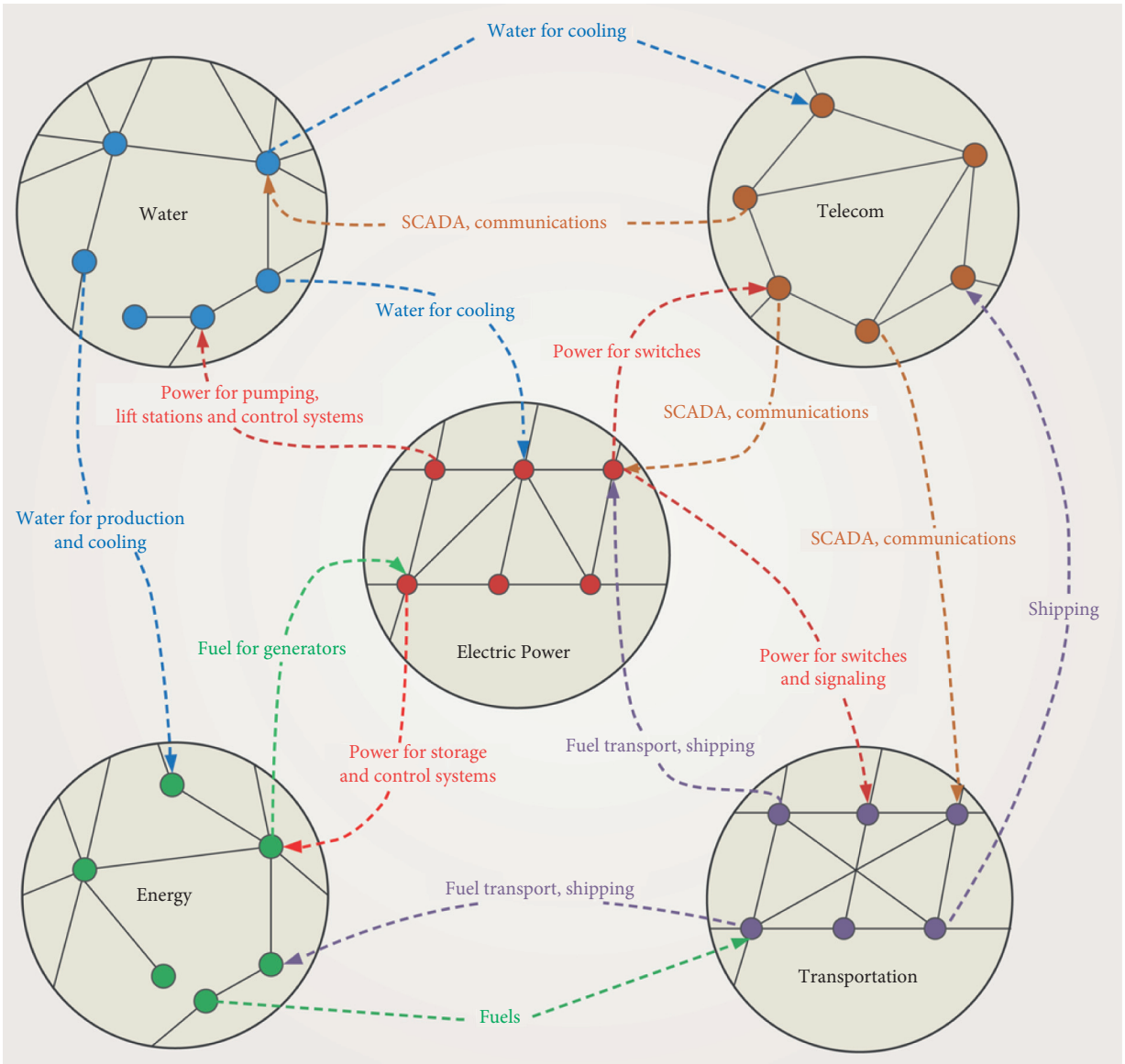


Figure 3: Interdependent urban networks. Illustration of complex relationships among different subsystems and infrastructures [119]. Connectivity patterns include feedback and feedforward paths, as well as branching topologies, whose structure and dynamics are better understood within the framework of multilayer systems [102, 103, 105, 106]. Figure reproduced with permission from [120], all rights reserved.

gasoline, the quantity of CO_2 emitted, and the relation between area and population of cities [70]. Remarkably, those results highlight the unsustainability of cities whose transportation infrastructure mostly relies upon traffic-sensitive modes, despite polycentrism [70]. The feedback between the urban backbone and routing systems is responsible for the emergence of congestion which temporarily degrades the functionality of the city, reminiscent of the slower-is-faster effect [124] and of the Braess's paradox [125], further supporting the hypothesis that a complexity science perspective, at the crossroads of multiple disciplines, is required. Despite its emergent nature, multilayer modeling [126] and analysis [127, 128] are enhancing our ability to anticipate congestion phenomena, at least from a theoretical perspective.

As for structure, also urban and interurban dynamics can be understood in terms of interdependent processes. This is the case, for instance, of human movements and epidemics spreading, with integrated models proposed for large-scale mobility [131] between cities, based on different sources of human-generated information, from commuting flows [132, 133] to mobile phone data [134–136] and geo-referenced social media [137]. Similarly, mobile phone data [138] and credit transactions [78] have been used to model and predict human mobility within a city, showing that predictive performances improve when information about social patterns is accounted for [139–141]. Social multiplexity inferred from the intersection of popular online platforms revealed that users connected on both platforms tend to have more similar neighborhoods, as well as more

similar social and spatial properties on both platforms with respect to users connected on just one layer [142]. Recently, human flows with recurrent mobility patterns within the city of Medellín in Colombia have been stratified by socioeconomic classes characterizing the city, to unveil the geographic location of patches triggering the epidemic state at the critical point of the process. Remarkably, those patches depend on the social mixing between classes and mobility [129] (see Figure 4). This result is extremely relevant for policy and decision-making, which cannot overlook the knowledge of the conditions under which such types of critical regimes are expected [130].

The analysis of critical properties provides the methodological baseline for understanding and quantifying the resilience of the city. The multiplex structure of London's public transportation network enhances its robustness to random failures of single stations and of entire routes. Analysis of the empirical distribution of check-ins and checks-out shows that passengers travel along fastest paths in a network affected by real disruptions, offering a basis for data-driven policies to enhance the navigability of the city [109]. The spatial constraints play a crucial role for critical properties: in contrast to other systems which are not embedded in space, interdependent urban networks are rather sensitive to failures, and abrupt collapse can be driven by any small fraction of interdependent nodes [143].

Multilayer modeling is therefore fundamental for better understanding city resilience and its complex, interdependent structures and dynamics. Urban ecosystems are, in fact, the result of growing networks which are interconnected and coevolving, inducing strong correlations across layers that can alter their response to social, economic, and environmental processes [144].

3. An Integrated View of Complex Urban Systems

3.1. The Cognitive Challenges of Urban Complexity. As already remarked, urban environments are defying not only from the viewpoint of the modeling of urban environments and of their structural interdependencies but also in terms of the challenges they pose to human cognitive systems [145]. For them to be useful, we need to integrate the insights deriving from the analysis of urban dynamics into mental models that enable us to represent, enact, and assess our strategies of navigation, utilization, and governance of the urban space. Such a feature of modeling creates a fascinating parallel with literary fiction [146] and contributes to explaining why narratives have always traditionally been a key tool to conceptualize, represent, and communicate the complexity of urban environments [147]. This is true not only for urban designers and policy-makers but also for local communities and citizens alike. This is why urban cognition is a key frontier of knowledge transmission and application of our findings from research on city complexity.

It has indeed long been recognized that humans have intrinsic cognitive limits in processing information [148]. Given the increasing complexity of highly urbanized cities, such limits have now become a challenge in the interaction of citizens with the urban environment [149]. The concept linking individual experiences to the increasingly complex circumstances of urban life is that of cognitive load [150]. A person's brain can be overloaded when there are too many inputs to reckon with or when sequences of inputs hit so fast that a new input is still being processed when the new one arrives. Under overload, individuals adapt their behavior by changing priorities and recurring to simplifying choice heuristics.

Redefining traditional tools such as maps to deal with the complexity of urban environments is a key step in this process. Recent studies on visual search strategies [151–153] have demonstrated that the time needed to find a route in a transportation network grows with the complexity of its map, with a transition in search strategies from following metro lines to the scattering of eye focus all over the map [152]. A similar transition from directional to isotropic random search has been observed for visual search of hidden objects when one increases the number of distractors [151]. The ability to manage complex "mental maps" is thus limited, and only extensive training on spatial navigation can push this limit with morphological changes in the hippocampus [154]. Therefore, transportation network structures may be too confusing, requiring to wade through too much detail to figure out whether the service is useful [155]. To measure the cognitive load associated with the visual search of a route in transportation networks, an information perspective has been proposed [111] to quantify the difficulty to navigate them. Using a measure of "information search" associated with a trip that goes from one route to another [156], it has been possible to characterize our difficulty to navigate in a public transit map to identify and measure the cognitive limit. To overcome such limits, a "fractal" approach to cartography is in a sense an obvious move, but what is less obvious is the "just in time" adaptive attitude that it commands upon planners, interest groups, and citizens [157].

Transferring these kind of ideas from specialists to nonspecialists is difficult and calls for a profound revision of individual mindsets, not to speak of mental models, and especially so if such knowledge has to become useful and applicable in specific problem-solving contexts [158]. This is an especially burning issue in view of the deliberation processes that support public decision-making in modern democracies. The appeal to oversimplified, and outright incorrect, solutions to complex urban challenges may be especially tempting for politicians and policy-makers when citizens are basically unable to grasp the subtleties of urban policy dilemmas and their implications for their own interests [159]. This can only be balanced by integrating such knowledge into the practical experience and local capability building processes of citizens [160]. But this "pedagogical" change cannot happen without a substantial redefinition of the professional culture of planners and urban experts of all sorts [161].

The ultimate sense of the challenges posed by applying a complexity science perspective to urban systems is that a city is a complex system emerging from the collective behavior of individuals who are themselves complex systems. This

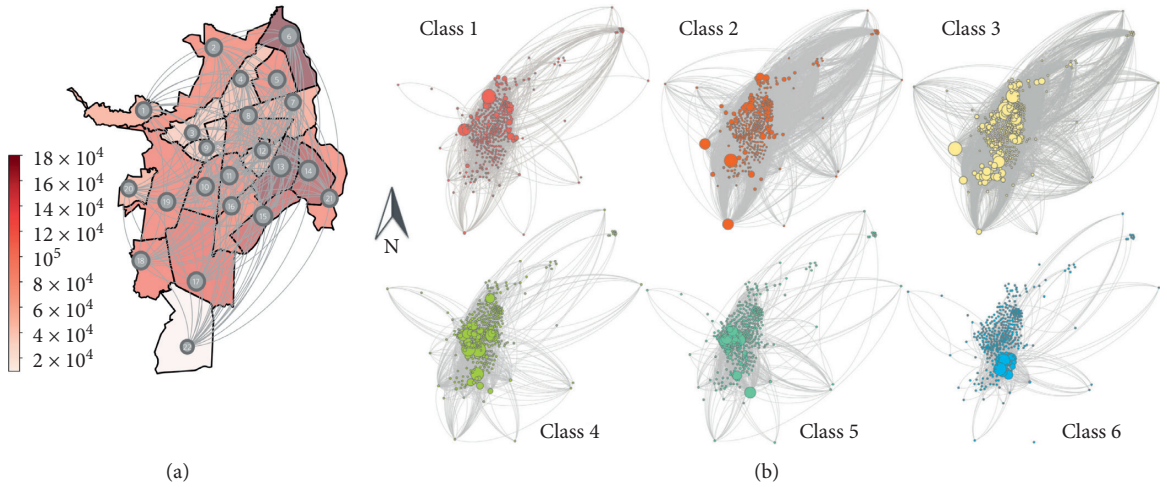


Figure 4: Multilayer urban dynamics. (a) Empirical mobility flows among different geographic areas in the city of Cali (Colombia). Each area is modeled as a metapopulation, that is, a node with population dynamics. (b) The city of Medellín (Colombia) stratified with respect to socioeconomic classes. Each class allows to build a layer of a multilayer model of the city, where each metapopulation network encodes human flows of a specific class. Figure reproduced with permission from [129] and [130], all rights reserved.

multiscale complexity feeds back in terms of cognitive constructs into the minds of the very individuals that make it exist [162]. Multiscale complexity is common to all large-scale social phenomena, from markets to States to online communities. However, the fact that the city is the socio-spatial context that mediates most human experiences makes it especially sensitive and to some extent fundamental to most other forms of social phenomena. Not incidentally, as noted by Portugali [162], Prigogine in his Nobel lecture indicated the city as the natural metaphor of social complexity. In a sense, then, the city's capacity of functioning as a complex system also depends on how the mental representations of the city, elaborated by its inhabitants, influence their own choices and behaviors and how this in turn reflects into certain spatial and behavioral patterns at the macro-scale. It is therefore fundamental to think of the city in terms of systems ecology and to develop new ecologically informed models of sociospatial cognition that mediate between the subjective and objective levels of urban experience. This is, for instance, the intuition behind the notion of cognitive affordances as the basis of a complexity-driven urban epistemology: the individual and social capacity to use the city is the result of a contextual negotiation built on previous experience but also sensitive to local variations of the urban codes of meaning [163]. As previous experience with pushing buttons in various different circumstances can guide our choice when we face a button to be possibly pushed in a certain new, unfamiliar circumstance, individuals and communities are getting increasingly familiar with maintaining a style of open-ended adaptive learning [164]. This cognitive fluidity becomes a necessity for citizens facing a constantly shifting and evolving urban environment that can, often unpredictably, challenge their assumptions, play with their perceptions, and put pressure upon their consolidated behavioral repertoires [165]. An affordances-based approach to urban epistemology can also be usefully tested on the basis of experimental trials with animal cognition, such as in the case of rats [166]. Thinking in terms of affordances creates a natural setting for closing the attitude-action gap that is at the root of the failure of effective collective action in urban environments [167], by repurposing them as "playable" spaces, both individually and collectively [168]. It is by regaining a shared sense of collective intentionality in the urban space that urban self-organization can be both better understood and more effectively governed [169]. But this increased need of a collective awareness calls for a substantial upgrading of our ambition in enabling participative, inclusive practices in the urban space. It is therefore necessary that these goals become a relevant item in the urban policy agenda.

3.2. Planning and Complexity: Adaptive Urban Policy-Making. From the point of view of planning theory, the main takeaway of a couple of decades' experience of dealing with cities from a complexity-focused perspective is the necessity of a radical redefinition of planning practices as an interplay between institutionally driven design and urban self-organization principles [170]. The stakes are high. Planning practices are mainly justified to secure fair and inclusive access to urban resources. At the same time, cities may look very different when seen from different vantage points corresponding to different levels of benefit and privilege [171]. The consequent danger is that the management of urban complexity may also function as a convenient smokescreen to decline political responsibility and accommodate certain vested interests at the expense of others [172], irrespective of the declared intentions. However, to the contrary, it is also possible that seriously accounting for the self-organization dimension of urban processes into planning practices may function as an enabler of community initiative and active participation [173]. In essence, taking advantage of the lesson of complexity in planning means creating the context for decentralized,

collaborative action at least as much as prescribing centralized, top-down forms of planning [174], thus laying the premises for a reflexive approach to planning [175].

The real issue from the planning perspective is therefore how to empower citizens and local communities to play an active role in this cocreative process and to be able to conceptualize, promote, and assess collective action in the pursuit of common interests [176] and in the exercise of their right to the city [177]. One possible solution is rediscovering rituals of collective pleasure in public spaces as a foundation for a common social intentionality [178], in the spirit of a repurposed collective action aimed at social change along the lines of Gramscian thinking, consequently turning the urban environment into a "playable" public domain [179]. Many collective practices of public art in the public domain are exploring possibilities and breaking new ground in this direction, by engaging local communities to take an active, propositional attitude towards urban complexity in terms of shared agency and not of passive delegation to planners and high-level stakeholders [180]. In this perspective, we look with great interest at gamification-based approaches insofar as they are intended not as domesticating formats for passive engagement and manipulative conditioning but rather as a smart deployment of collective resources and talent [181, 182] and of self-empowering behavioral change [183], which appeals to expressive rather than to instrumental motives [184]. In this sense, gamification has shown promise as a surprisingly practical and flexible tool to pursue complex technical and sociopolitical collective goals [185], due to two main characteristics: its intrinsic narrative potential and appeal, which as already remarked may effectively function in conveying detailed contextual information in complex urban environments [186], and its endogenous scoring metrics that, while providing direct feedback as to the efficacy of certain individual or collective actions, also build motivation and engagement [187].

In this regard, the pandemic crisis and the consequent necessity to rely upon contact tracing apps to mitigate the spreading of the contagion can be regarded as a missed opportunity. In countries whose local culture is characterized by a strong civic sense and on the emphasis on collective responsibility and duty as well as by a mature stage of the digital transition, such as many Far-Eastern societies [188, 189], contract tracing apps have been massively adopted and have functioned well despite the inevitable concerns on privacy [190], keeping the level of contagion and human losses remarkably low [191] if compared to that of more individualistic societies [192] where contract tracing has been implemented relatively late and dedicated apps have been adopted by a minority [192]. Even in a situation of high personal risk, reliance on responsibility and fear as incentives to app adoption has not proven effective in mobilizing citizens to behave prosocially, despite a generalized declared willingness to download contact tracing apps across most countries [193]. On the other hand, the apps themselves were merely designed, following a purely functional logic, as carriers of information and control devices in a moment in which people felt a sudden, heavy burden from the almost complete disruption of their previous social life. Maybe considering social incentives to communication and exchange that were put particularly under stress during the pandemic crisis could have been a key to a stronger motivation to adopt the app by choosing suitable, motivationally salient cues. For instance, people could be shown how the probability of lifting up social restrictions would be affected by the rate of adoption of the app and consequently how that probability could change for every extra thousand people adopting it, inviting them to contribute by adopting the app themselves and convincing their acquaintances to do the same. Or, alternatively, in terms of appeal to responsibility, people could receive a constantly updated estimate of the probability to be infectious given their history of social contacts, so as to discourage risky behavior through appeal to regret [194].

4. Outlooks

Most applications of complexity science to urban issues tend to focus, for understandable reasons, on specific dimensions of the urban environment. Clearly, studying the structure and the dynamical evolution of urban utilities networks, transportation systems, or resource flows is already more than enough to challenge our modeling abilities. However, if we want to fully acknowledge the implications of the previous analysis and in particular the necessity of integrating the three dimensions of modeling, cognition, and governance, we need to look at urban environments not from a sector-specific but from a system-wide perspective that cuts across different sectors and dimensions of the urban fabric. Such an integrated framework to understand cities through data-powered tools sits at the edge between many different disciplines such as statistical physics, social sciences, economics, digital health and wellbeing, and engineering, to name only a few ones. The state-of-the-art applications reviewed above provide us with a promising benchmark by characterizing urban complexity in terms of multiplex networks and possibly point to the multiplex city as a computationally appropriate and conceptually scalable [195] representation of urban complexity [196], which will allow the development and deployment of new urban governance strategies, as well as the redesign of old ones [54].

Urban environments clearly raise specific health issues, which are becoming of increasing relevance once we are reminded that, according to recent estimations, up to 70% of the world's population will be living in urban areas by 2050 [197], and even if in the postpandemic scenario this trend could be subject to changes in the medium-long term, cities will inevitably keep on playing a pivotal role in future economies and societies [198]. This extraordinary scale of urbanization is affecting both the environmental impact of human activities and the nature and scope of human health and wellbeing issues by posing new adaptive challenges [199]. The Ottawa global health milestones agreement Charter puts the concept of the promotion of healthy practices and lifestyles at the center of a socially and financially sustainable approach to public health [200], where, as formalized by the salutogenesis [215] paradigm [201], the

focus shifts from the causes and reinforcing conditions of diseases to those that favor and preserve health. The crucial step that enables individuals to successfully adapt to the stressful conditions of the urban environments of socioeconomically advanced societies is to enhance their capacity to cope with such critical factors by improving resilience at all social scales [202]. The urban dimension of salutogenesis clearly becomes all the more central in the postpandemic scenario and will likely leave a deep trace in future urban policies [203].

Although these principles are today widely recognized, finding viable approaches to their implementation in urban environments is not easy task. However, a clever use of the incentive systems connected to digital participation may provide, as discussed above, an innovative platform to motivate people to pursue healthy habits and lifestyles while promoting other public interest goals at the same time. To address this issue, it is crucial to devise data-driven solutions for the promotion of innovative practices of urban health and wellbeing to promote a salutogenic approach to urban space through proactive access to, and use of, the varied mix of wellbeing-enhancing assets in the public domain. Such promotion may be accomplished by means of gamified participatory practices which deeply engage people in the active pursuit of integrated salutogenic goals [204, 205].

Strategies to encourage citizens to pursue health promotion goals by actively engaging with their urban environment include the following:

(i) Fostering inclusiveness and relationship building instead of social stratification and segregation [206]

(ii) Promoting cultural participation by arousing curiosity about, and involvement with, urban cultural heritage and poorly known landmarks [207]

(iii) Creating new opportunities to adopt healthy habits, including increased mobility, while reducing depletion of natural capital [208, 209]

To understand to what extent an integrated approach that takes culture as a driver of urban change may become a powerful basis for a systemic view of urban functioning and change which invites citizens to be proactive in the pursuit of their own quality of life and wellbeing, two aspects of special importance are sociability and cultural experience. Sociability is an issue of increasing relevance in the public health agenda, as loneliness is now recognized as a serious public health problem [210]. Cultural experience is increasingly connected to health and wellbeing issues by a rapidly growing literature [211, 212]. The eudaimonic [216] approach to wellbeing provides a conceptual framework encompassing all three dimensions within the overarching salutogenic framework, as it postulates that a harmonic integration of different spheres of human existence best fits human sociopsychological development [213] and promotes health more effectively than narrower approaches [214]. In an eudaimonic perspective, the urban environment becomes an extremely rich and potentially stimulating playground on multiple dimensions: as a natural context for urban mobility, as an elective space of sociability, and as a theater of cultural experiences. Breathing new life into the postpandemic urban fabric will be not only a problem of technical efficiency and of provision of adequate services tailored to the new social demands and needs but also an exercise of collective sense-making. In the forthcoming scenario, it will therefore be crucial to directly involve citizens in collective problem-solving processes to codesign the postpandemic city, both to build upon collective intelligence and to rebuild social cohesion after a long, critical period of isolation and social alienation and as a powerful driver of widespread adoption of prosocial practices, attitudes, and new habits without which the cities of the future might become less and less livable. This is in particular true in view of the strong call that urban life in postpandemic cities will make for widespread citizens' responsibility and attention to safety concerns and public health norms compliance. Taking this forced start of a new urban cycle as an opportunity to bring urban sociability at the core of urban policies through the active contribution of citizens may lead to new, emergent forms of social cooperation whose collective benefits may be directly measured and, consequently, may further reinforce individual and social motivations to pursue a more sustainable, cohesive route to urban development. It is unlikely that the new challenges will be effectively tackled through business-as-usual policy approaches [2], and there will be the need to refresh what looked like the "default" conviction only one year ago, namely, that cities are the undisputed, trend-setting centers of gravity of social, economic, and cultural life. With the increasing diffusion of remote smart working, the choice to live in a city will be less instrumental and will have to be motivated by the intrinsic richness of opportunity of the urban environment as to the promotion of socially rewarding encounters and mind-opening, stimulating experiences. In the postpandemic scenario, sociability habits and patterns in the urban space will have to be rebuilt almost from scratch, and new demands for social connection and meaningful interaction will inevitably arise. Likewise, the postpandemic city will have to rethink and repurpose its public space as a theater of collective expression and inspiration. We need to be ready to tackle these challenges through a clear vision of the multilayered complexity of urban dynamics, with a strong focus on the livability and liveliness of urban environments. However important, these dimensions of sense-making have been marginally dealt with by complexity models and analyses. The time has come to integrate them into a full-fledged approach to urban systems.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors equally contributed to this work.

References

[1] J. JacobsVintage, New York, NY, USA, 1992.

[2] J. C. Scott, *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*, Yale University Press, New Haven, USA, 2020.

[3] C. Alexanderp. 1965, 1964.

[4] C. Alexander, *A City Is Not a Tree*, Sustasis Press/Off the Common Books, Amsterdam, Holland, 2017.

[5] C. Alexander, *A Pattern Language: Towns, Buildings, Construction*, Oxford University Press, Oxford, UK, 1977.

[6] C. Alexander, *The Timeless Way of Building*, Vol. 1, Oxford University Press, New York, NY, USA, 1979.

[7] C. Alexander, "The phenomenon of life: book one the nature of order: an essay on the art of building and the nature of the universe," in *A Vision of a Living World: The Nature of Order*, Book 3, Center for Environmental Structure, New York, NY, USA, 2002.

[8] C. Alexander, *The Process of Creating Life*, Vol. 10, Center for Environmental Structure, Berkeley, CA, 2002.

[9] C. Alexander, *A Vision of a Living World: The Nature of Order*, Book 3, Center for Environmental Structure, New York, NY, USA, 2004.

[10] C. Alexander, *The Luminous Ground: An Essay on the Art of Building and the Nature of the Universe*, Taylor & Francis, Milton Park, UK, 2004.

[11] G. Stiny, "Introduction to shape and shape grammars," *Environment and Planning B: Planning and Design*, vol. 7, no. 3, pp. 343–351, 1980.

[12] K. Karimi, "A configurational approach to analytical urban design: 'Space syntax' methodology," *Urban Design International*, vol. 17, no. 4, pp. 297–318, 2012.

[13] T. V. Heitor, J. P. Duarte, and R. M. Pinto, "Combing grammars and space syntax: formulating, generating and evaluating designs," *International Journal of Architectural Computing*, vol. 2, no. 4, pp. 491–515, 2004.

[14] P. M. Allen and M. Sanglier, "Urban evolution, self-organization, and decisionmaking," *Environment & Planning A: Economy and Space*, vol. 13, no. 2, pp. 167–183, 1981.

[15] D. S. Dendrinos and H. Mullally, *Urban Evolution: Studies in the Mathematical Ecology of Cities*, Oxford University Press, Oxford, USA, 1985.

[16] M. Batty and P. A. Longley, *Fractal Cities: A Geometry of Form and Function*, Academic Press, Cambridge, Massachusetts, USA, 1994.

[17] I. Benenson, "Multi-agent simulations of residential dynamics in the city," *Computers, Environment and Urban Systems*, vol. 22, no. 1, pp. 25–42, 1998.

[18] 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.

[19] M. Batty and P. M. Torrens, "Modelling and prediction in a complex world," *Futures*, vol. 37, no. 7, pp. 745–766, 2005.

[20] M. Barthelemy, *The Structure and Dynamics of Cities*, Cambridge University Press, Cambridge, USA, 2016.

[21] M. Barthelemy, "The statistical physics of cities," *Nature Reviews Physics*, vol. 1, no. 6, pp. 406–415, 2019.

[22] H. Lu and Y. Shi, "Complexity of public transport networks," *Tsinghua Science & Technology*, vol. 12, no. 2, pp. 204–213, 2007.

[23] M. Zarghami and S. Akbariyeh, "System dynamics modeling for complex urban water systems: application to the city of Tabriz, Iran," *Resources, Conservation and Recycling*, vol. 60, pp. 99–106, 2012.

[24] Y. Zhang, Z. Yang, B. D. Fath, and S. Li, "Ecological network analysis of an urban energy metabolic system: model development, and a case study of four Chinese cities," *Ecological Modelling*, vol. 221, no. 16, pp. 1865–1879, 2010.

[25] S. Kai, L. Chun-Qiong, and L. Si-Chuan, "Self-Organized Criticality: emergent complex behavior in PM10 pollution," *International Journal of Atmospheric Sciences*, vol. 2013, Article ID 419694, 7 pages, 2013.

[26] M. Fonoberova, V. A. Fonoberov, I. Mezic, J. Mezic, and P. J. Brantingham, "Crime and violence in urban settings," *The Journal of Artificial Societies and Social Simulation*, vol. 15, p. 2, 2012.

[27] K. M. Faust, D. M. Abraham, and D. DeLaurentis, "Coupled human and water infrastructure systems sector interdependencies: framework evaluating the impact of cities experiencing urban decline," *Journal of Water Resources Planning and Management*, vol. 143, no. 8, Article ID 04017043, 2017.

[28] L. M. A. Bettencourt, J. Lobo, D. Strumsky, and G. B. West, "Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities," *PLoS One*, vol. 5, no. 11, Article ID e13541, 2010.

[29] J. Johnson, A. Nowak, P. Ormerod, B. Rosewell, and Y.-C. Zhang, *Non-Equilibrium Social Science and Policy: Introduction and Essays on New and Changing Paradigms in Socio-Economic Thinking*, Springer, Berlin, Germany, 2017.

[30] B. Lepri, F. Antonelli, F. Pianesi, and A. Pentland, "Making big data work: smart, sustainable, and safe cities," *EPJ Data Science*, vol. 4, p. 1, 2015.

[31] M. Lenormand, T. Louail, M. Barthelemy, and J. Ramasco, *Spatial Accuracy 2016*, 2016.

[32] R. Gallotti, R. Louf, J.-M. Luck, and M. Barthelemy, "Tracking random walks," *Journal of The Royal Society Interface*, vol. 15, no. 139, 2018.

[33] M. Batty, *The New Science of Cities*, MIT Press, Cambridge, USA, 2013.

[34] E. N. Gilbert, "Random plane networks," *Journal of the Society for Industrial and Applied Mathematics*, vol. 9, no. 4, pp. 533–543, 1961.

[35] R. Gallotti, G. Bertagnolli, and M. De Domenico, *EPJ Data Science*, vol. 10, p. 1, 2021.

[36] M. Barthélemy, "Spatial networks," *Physics Reports*, vol. 499, no. 1-3, pp. 1–101, 2011.

[37] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, 2001.

[38] C. Andersson, K. Frenken, and A. Hellervik, "A complex network approach to urban growth," *Environment & Planning A: Economy and Space*, vol. 38, no. 10, pp. 1941–1964, 2006.

[39] S. Porta, P. Crucitti, and V. Latora, "The network analysis of urban streets: a dual approach," *Physica A: Statistical Mechanics and its Applications*, vol. 369, no. 2, pp. 853–866, 2006.

[40] 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.

[41] A. Sevtsuk and M. Mekonnen, "Urban network analysis. A new toolbox for ArcGIS," *Revue Internationale de Géomatique*, vol. 22, no. 2, pp. 287–305, 2012.

[42] Y.-H. Tsai, "Quantifying urban form: compactness versus 'sprawl'," *Urban Studies*, vol. 42, no. 1, pp. 141–161, 2005.
[43] M. Guérois and D. Pumain, "Built-up encroachment and the urban field: a comparison of forty European cities," *Environment & Planning A: Economy and Space*, vol. 40, no. 9, pp. 2186–2203, 2008.
[44] N. Schwarz, "Urban form revisited-selecting indicators for characterising European cities," *Landscape and Urban Planning*, vol. 96, no. 1, pp. 29–47, 2010.
[45] C. K. Gately, L. R. Hutyra, and I. Sue Wing, "Cities, traffic, and CO2: a multidecadal assessment of trends, drivers, and scaling relationships," *Proceedings of the National Academy of Sciences*, vol. 112, no. 16, pp. 4999–5004, 2015.
[46] R. Ewing and S. Hamidi, "Compactness versus sprawl," *Journal of Planning Literature*, vol. 30, no. 4, pp. 413–432, 2015.
[47] C. Song, T. Koren, P. Wang, and A.-L. Barabási, "Modelling the scaling properties of human mobility," *Nature Physics*, vol. 6, no. 10, pp. 818–823, 2010.
[48] T. Louail, M. Lenormand, M. Picornell et al., "Uncovering the spatial structure of mobility networks," *Nature Communications*, vol. 6, no. 1, 2015.
[49] R. Gallotti, A. Bazzani, S. Rambaldi, and M. Barthelemy, "A stochastic model of randomly accelerated walkers for human mobility," *Nature Communications*, vol. 7, 2016.
[50] H. Barbosa, M. Barthelemy, G. Ghoshal et al., "Human mobility: models and applications," *Physics Reports*, vol. 734, pp. 1–74, 2018.
[51] D. Helbing, "Traffic and related self-driven many-particle systems," *Reviews of Modern Physics*, vol. 73, no. 4, pp. 1067–1141, 2001.
[52] D. Li, B. Fu, Y. Wang et al., "Percolation transition in dynamical traffic network with evolving critical bottlenecks," *Proceedings of the National Academy of Sciences*, vol. 112, no. 3, pp. 669–672, 2015.
[53] S. Çolak, A. Lima, and M. C. González, "Understanding congested travel in urban areas," *Nature Communications*, vol. 7, 2016.
[54] A. Solé-Ribalta, S. Gómez, and A. Arenas, "Decongestion of urban areas with hotspot pricing," *Networks and Spatial Economics*, vol. 18, no. 1, pp. 33–50, 2018.
[55] J. Depersin and M. Barthelemy, "From global scaling to the dynamics of individual cities," *Proceedings of the National Academy of Sciences*, vol. 115, no. 10, pp. 2317–2322, 2018.
[56] F. Le Néchet, "Cybergeo," *European Journal of Geography*, 2012.
[57] B. Stone, "Urban sprawl and air quality in large US cities," *Journal of Environmental Management*, vol. 86, no. 4, pp. 688–698, 2008.
[58] E. Uherek, T. Halenka, J. Borken-Kleefeld et al., "Transport impacts on atmosphere and climate: land transport," *Atmospheric Environment*, vol. 44, no. 37, pp. 4772–4816, 2010.
[59] A. Martilli, "An idealized study of city structure, urban climate, energy consumption, and air quality," *Urban Climate*, vol. 10, pp. 430–446, 2014.
[60] R. Ewing, G. Meakins, S. Hamidi, and A. C. Nelson, "Relationship between urban sprawl and physical activity, obesity, and morbidity-update and refinement," *Health & Place*, vol. 26, pp. 118–126, 2014.
[61] D. E. Newby, P. M. Mannucci, G. S. Tell et al., "Expert position paper on air pollution and cardiovascular disease," *European Heart Journal*, vol. 36, no. 2, pp. 83–93, 2014.
[62] M. B. Rice, P. L. Ljungman, E. H. Wilker et al., "Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the framingham heart study," *American Journal of Respiratory and Critical Care Medicine*, vol. 191, no. 6, pp. 656–664, 2015.
[63] W. Li, K. S. Dorans, E. H. Wilker et al., "Residential proximity to major roadways, fine particulate matter, and hepatic steatosis," *American Journal of Epidemiology*, vol. 186, no. 7, pp. 857–865, 2017.
[64] J. Nicholl, J. West, S. Goodacre, and J. Turner, "The relationship between distance to hospital and patient mortality in emergencies: an observational study," *Emergency Medicine Journal*, vol. 24, no. 9, pp. 665–668, 2007.
[65] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, "Growth, innovation, scaling, and the pace of life in cities," *Proceedings of the National Academy of Sciences*, vol. 104, no. 17, pp. 7301–7306, 2007.
[66] L. M. A. Bettencourt, "The origins of scaling in cities," *Science*, vol. 340, no. 6139, pp. 1438–1441, 2013.
[67] A. Bertaud, 2004.
[68] V. Volpati and M. Barthelemy, "The spatial organization of the population density in cities," 2018, http://arxiv.org/abs/1804.00855.
[69] R. Louf and M. Barthelemy, "Modeling the polycentric transition of cities," *Physical Review Letters*, vol. 111, no. 19, 2013.
[70] R. Louf and M. Barthelemy, "How congestion shapes cities: from mobility patterns to scaling," *Scientific Reports*, vol. 4, no. 1, 2014.
[71] M. Haklay and P. Weber, "OpenStreetMap: user-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.
[72] P. Neis, D. Zielstra, and A. Zipf, "The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011," *Future Internet*, vol. 4, no. 1, pp. 1–21, 2011.
[73] C. Ratti, D. Frenchman, R. M. Pulselli, and S. Williams, "Mobile landscapes: using location data from cell phones for urban analysis," *Environment and Planning B: Planning and Design*, vol. 33, no. 5, pp. 727–748, 2006.
[74] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
[75] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
[76] V. Soto and E. Frías-martínez, "Robust land use characterization of urban landscapes using cell phone data," 2011.
[77] T. Louail, M. Lenormand, O. G. C. Ros et al., "From mobile phone data to the spatial structure of cities," *Scientific Reports*, vol. 4, 2014.
[78] S. Hasan, C. M. Schneider, S. V. Ukkusuri, and M. C. González, "Spatiotemporal patterns of urban human mobility," *Journal of Statistical Physics*, vol. 151, no. 1-2, pp. 304–318, 2012.
[79] Ľ. Jánošíková, J. Slavík, and M. Koháni, "Estimation of a route choice model for urban public transport using smart card data," *Transportation Planning and Technology*, vol. 37, p. 638, 2014.
[80] S. Sobolevsky, I. Bojic, A. Belyi et al. in *Proceedings of the 2015 IEEE International Congress on Big Data*, June 2015.
[81] M. Lenormand, T. Louail, O. G. Cantú-Ros et al., "Influence of sociodemographic characteristics on human mobility," *Scientific Reports*, vol. 5, no. 1, 2015.
[82] M. K. El Mahrsi, E. Come, L. Oukhellou, and M. Verleysen, "Clustering smart card data for urban mobility analysis,"

IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 3, pp. 712–728, 2017.

[83] R. Di Clemente, M. Luengo-Oroz, M. Travizano, S. Xu, B. Vaitla, and M. C. González, "Sequences of purchases in credit card data reveal lifestyles in urban populations," *Nature Communications*, vol. 9, no. 1, 2018.

[84] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on GPS data," in *Proceedings of the 10th International Conference on Ubiquitous Computing-UbiComp'08.*, pp. 312–321, ACM, Seoul, Korea, September 2008.

[85] A. Bazzani, B. Giorgini, S. Rambaldi, R. Gallotti, and L. Giovannini, "Statistical laws in urban mobility from microscopic GPS data in the area of florence," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2010, no. 05, 2010.

[86] R. Gallotti, A. Bazzani, and S. Rambaldi, "Towards a statistical physics of human mobility," *International Journal of Modern Physics C*, vol. 23, 2012.

[87] R. Gallotti, A. Bazzani, and S. Rambaldi, "Understanding the variability of daily travel-time expenditures using GPS trajectory data," *EPJ Data Science*, vol. 4, no. 1, p. 18, 2015.

[88] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, and A.-L. Barabási, "Returners and explorers dichotomy in human mobility," *Nature Communications*, vol. 6, no. 1, 2015.

[89] F. Kling and A. Pozdnoukhov, "When a city tells a story," in *Proceedings of the 20th International Conference on Advances in Geographic Information Systems-SIGSPATIAL12*, 2012.

[90] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A tale of many cities: universal patterns in human urban mobility," *PLoS One*, vol. 7, no. 5, Article ID e37027, 2012.

[91] T. H. Silva, P. O. S. V. de Melo, J. M. Almeida, J. Salles, and A. A. F. Loureiro, "A comparison of foursquare and instagram to the study of city dynamics and urban social behavior," in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing-UrbComp 13*, August 2013.

[92] A. Noulas, C. Mascolo, and E. Frias-Martinez, "Exploiting foursquare and cellular data to infer user activity in urban environments," in *Proceedings of the 2013 IEEE 14th International Conference on Mobile Data Management*, June 2013.

[93] R. Feick and C. Robertson, "A multi-scale approach to exploring urban places in geotagged photographs," *Computers, Environment and Urban Systems*, vol. 53, pp. 96–109, 2015.

[94] T. Shelton, A. Poorthuis, and M. Zook, "Social media and the city: rethinking urban socio-spatial inequality using user-generated geographic information," *Landscape and Urban Planning*, vol. 142, pp. 198–211, 2015.

[95] E. Chaniotakis and C. Antoniou, "Use of geotagged social media in urban settings: empirical evidence on its potential from twitter," in *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*, September 2015.

[96] Y. Hu, S. Gao, K. Janowicz, B. Yu, W. Li, and S. Prasad, "Extracting and understanding urban areas of interest using geotagged photos," *Computers, Environment and Urban Systems*, vol. 54, pp. 240–254, 2015.

[97] G. Lansley and P. A. Longley, "The geography of Twitter topics in London," *Computers, Environment and Urban Systems*, vol. 58, pp. 85–96, 2016.

[98] S. Cvetojevic, L. Juhasz, and H. Hochmair, "Positional accuracy of twitter and instagram images in urban environments," *GI_Forum*, vol. 4, no. 1, pp. 191–203, 2016.

[99] J. D. Boy and J. Uitermark, "How to study the city on instagram," *PLoS One*, vol. 11, no. 6, Article ID e0158161, 2016.

[100] P. Giridhar, S. Wang, T. Abdelzaher, R. Ganti, L. Kaplan, and J. George, "On localizing urban events with instagram," in *Proceedings of the IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, May 2017.

[101] T. R. Meyer, D. Balagué, M. Camacho-Collados et al., "A year in Madrid as described through the analysis of geotagged twitter data," *Environment and Planning B: Urban Analytics and City Science*, vol. 46, no. 9, pp. 1724–1740, 2018.

[102] M. De Domenico, A. Solé-Ribalta, E. Cozzo et al., "Mathematical formulation of multilayer networks," *Physical Review X*, vol. 3, no. 2, Article ID 041022, 2013.

[103] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of complex networks*, vol. 2, no. 3, pp. 203–271, 2014.

[104] R. Gallotti and M. Barthelemy, "Anatomy and efficiency of urban multimodal mobility," *Scientific Reports*, vol. 4, no. 1, 2014.

[105] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.

[106] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, "Networks formed from interdependent networks," *Nature Physics*, vol. 8, no. 1, pp. 40–48, 2012.

[107] R. Gallotti and M. Barthelemy, "The multilayer temporal network of public transport in Great Britain," *Scientific Data*, vol. 2, no. 1, 2015.

[108] R. G. Morris and M. Barthelemy, "Transport on coupled spatial networks," *Physical Review Letters*, vol. 109, no. 12, 2012.

[109] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, "Navigability of interconnected networks under random failures," *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8351–8356, 2014.

[110] L. Alessandretti, M. Karsai, and L. Gauvin, "User-based representation of time-resolved multimodal public transportation networks," *Royal Society Open Science*, vol. 3, no. 7, 2016.

[111] R. Gallotti, M. A. Porter, and M. Barthelemy, "Lost in transportation: information measures and cognitive limits in multilayer navigation," *Science Advances*, vol. 2, no. 2, Article ID e1500445, 2016.

[112] A. Aleta, S. Meloni, and Y. Moreno, "A Multilayer perspective for the analysis of urban transportation systems," *Scientific Reports*, vol. 7, no. 1, p. 44359, 2017.

[113] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," *Nature Communications*, vol. 6, no. 1, p. 6864, 2015.

[114] E. Cozzo, M. Kivelä, M. D. Domenico et al., "Structure of triadic relations in multiplex networks," *New Journal of Physics*, vol. 17, no. 7, Article ID 073029, 2015.

[115] E. Strano, S. Shai, S. Dobson, and M. Barthelemy, "Multiplex networks in metropolitan areas: generic features and local effects," *Journal of The Royal Society Interface*, vol. 12, no. 111, 2015.

[116] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[117] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González, "Unravelling daily human mobility motifs," *Journal of The Royal Society Interface*, vol. 10, no. 84, 2013.
[118] C. G. Prato, "Route choice modeling: past, present and future research directions," *Journal of Choice Modelling*, vol. 2, no. 1, pp. 65–100, 2009.
[119] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies," *IEEE Control Systems Magazine*, vol. 21, no. 11, 2001.
[120] J. Gao, D. Li, and S. Havlin, "From a single network to a network of networks," *National Science Review*, vol. 1, no. 3, pp. 346–356, 2014.
[121] J. G. Wardrop and J. I. Whitehead, "Correspondence. Some theoretical aspects of road traffic research," *Proceedings-Institution of Civil Engineers*, vol. 1, no. 5, pp. 767-768, 1952.
[122] A. Lima, R. Stanojevic, D. Papagiannaki, P. Rodriguez, and M. C. González, "Understanding individual routing behaviour," *Journal of the Royal Society Interface*, vol. 13, no. 116, 2016.
[123] M. De Domenico, A. Lima, M. C. González, and A. Arenas, "Personalized routing for multitudes in smart cities," *EPJ Data Science*, vol. 4, no. 1, p. 1, 2015.
[124] C. Gershenson and D. Helbing, "When slower is faster," *Complexity*, vol. 21, no. 2, pp. 9–15, 2015.
[125] D. Braess, "Über ein Paradoxon aus der Verkehrsplanung," *Unternehmensforschung Operations Research - Recherche Opérationnelle*, vol. 12, no. 1, pp. 258–268, 1968.
[126] A. Solé-Ribalta, S. Gómez, and A. Arenas, "Congestion induced by the structure of multiplex networks," *Physical Review Letters*, vol. 116, no. 10, 2016.
[127] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas, "Ranking in interconnected multilayer networks reveals versatile nodes," *Nature Communications*, vol. 6, no. 1, p. 6868, 2015.
[128] S. Manfredi, E. Di Tucci, and V. Latora, "Mobility and congestion in dynamical multilayer networks with finite storage capacity," *Physical Review Letters*, vol. 120, 2018.
[129] D. Soriano-Paños, L. Lotero, A. Arenas, and J. Gómez-Gardeñes, "Spreading processes in multiplex metapopulations containing different mobility networks," *Physical Review X*, vol. 8, 2018.
[130] J. Gómez-Gardeñes, D. Soriano-Paños, and A. Arenas, "Critical regimes driven by recurrent mobility patterns of reaction-diffusion processes in networks," *Nature Physics*, vol. 14, no. 4, pp. 391–395, 2018.
[131] V. Belik, T. Geisel, and D. Brockmann, "Natural human mobility patterns and spatial spread of infectious diseases," *Physical Review X*, vol. 1, 2011.
[132] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21484–21489, 2009.
[133] D. Brockmann and D. Helbing, "The hidden geometry of complex, network-driven contagion phenomena," *Science*, vol. 342, no. 6164, pp. 1337–1342, 2013.
[134] A. Lima, M. De Domenico, V. Pejovic, and M. Musolesi, *Scientific Reports*, vol. 5, 2015.
[135] J. T. Matamalas, M. De Domenico, and A. Arenas, "Assessing reliable human mobility patterns from higher order memory in mobile communications," *Journal of The Royal Society Interface*, vol. 13, no. 121, 2016.
[136] P. Bosetti, P. Poletti, M. Stella, B. Lepri, S. Merler, and M. De Domenico, "Reducing measles risk in turkey through social integration of syrian refugees," 2019, http://arxiv.org/abs/1901.04214.
[137] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, "Geo-located Twitter as proxy for global mobility patterns," *Cartography and Geographic Information Science*, vol. 41, no. 3, pp. 260–271, 2014.
[138] X.-Y. Yan, C. Zhao, Y. Fan, Z. Di, and W.-X. Wang, "Universal predictability of mobility patterns in cities," *Journal of The Royal Society Interface*, vol. 11, no. 100, 2014.
[139] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011.
[140] M. De Domenico, A. Lima, and M. Musolesi, "Interdependence and predictability of human mobility and social interactions," *Pervasive and Mobile Computing*, vol. 9, no. 6, pp. 798–807, 2013.
[141] J. L. Toole, C. Herrera-Yaqüe, C. M. Schneider, and M. C. González, "Coupling human mobility and social ties," *Journal of the Royal Society Interface*, vol. 12, no. 105, 2015.
[142] D. Hristova, A. Noulas, C. Brown, M. Musolesi, and C. Mascolo, "A multilayer approach to multiplexity and link prediction in online geo-social networks," *EPJ Data Science*, vol. 5, no. 1, p. 24, 2016.
[143] A. Bashan, Y. Berezin, S. V. Buldyrev, and S. Havlin, "The extreme vulnerability of interdependent spatially embedded networks," *Nature Physics*, vol. 9, no. 10, pp. 667–672, 2013.
[144] J. Y. Kim and K.-I. Goh, "Coevolution and correlated multiplexity in multiplex networks," *Physical Review Letters*, vol. 111, no. 5, 2013.
[145] J. Portugali, *Complexity, Cognition and the City*, Springer Science & Business Media, Berlin, Germany, 2011.
[146] R. Frigg, "Models and fiction," *Synthese*, vol. 172, no. 2, pp. 251–268, 2010.
[147] V. Mukhija, "Learning from invisible cities: the interplay and dialogue of order and disorder," *Environment & Planning A: Economy and Space*, vol. 47, no. 4, pp. 801–815, 2015.
[148] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.
[149] S. Milgram, "The experience of living in cities," *Science*, vol. 167, no. 3924, pp. 1461–1468, 1970.
[150] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," *Educational Psychologist*, vol. 38, no. 1, pp. 63–71, 2003.
[151] H. F. Credidio, E. N. Teixeira, S. D. S. Reis, A. A. Moreira, and J. S. Andrade, "Statistical patterns of visual search for hidden objects," *Scientific Reports*, vol. 2, no. 1, p. 920, 2012.
[152] M. Burch, K. Kurzhals, and D. Weiskopf, "Visual task solution strategies. in public transport maps ET4S@ GIScience," pp. 32–36, 2014.
[153] M. Burch, K. Kurzhals, M. Raschke, T. Blascheck, and D. Weiskopf, "A task-based view on the visual analysis of eye-tracking data," *Eye Tracking and Visualization, Mathematics and Visualization*, vol. 3, 2017.
[154] E. A. Maguire, D. G. Gadian, I. S. Johnsrude et al., "Navigation-related structural change in the hippocampi of taxi drivers," *Proceedings of the National Academy of Sciences*, vol. 97, no. 8, pp. 4398–4403, 2000.

[155] J. Walker, *Human Transit: How Clearer Thinking about Public Transit Can Enrich Our Communities and Our Lives*, Island Press, Washington, DC, USA, 2012.
[156] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, "Networks and cities: an information perspective," *Physical Review Letters*, vol. 94, 2005.
[157] S. Encarnação, M. Gaudiano, F. C. Santos, J. A. Tenedório, and J. M. Pacheco, "Fractal cartography of urban areas," *Scientific Reports*, vol. 2, no. 1, p. 527, 2012.
[158] M. J. Jacobson, "Problem solving, cognition, and complex systems: differences between experts and novices," *Complexity*, vol. 6, no. 3, pp. 41–49, 2001.
[159] O. Karaman, "Urban neoliberalism with islamic characteristics," *Urban Studies*, vol. 50, no. 16, pp. 3412–3427, 2013.
[160] A. Walker, "Introduction," *The Asia Pacific Journal of Anthropology*, vol. 2, no. 2, pp. 1–20, 2001.
[161] N. J. Habraken, "Cultivating complexity: the need for a shift in cognition," *Complexity, Cognition, Urban Planning and Design*, pp. 55–74, Springer, Berlin, Germany, 2016.
[162] J. Portugali, "What makes cities complex?," *Complexity, Cognition, Urban Planning and Design*, pp. 3–19, Springer, Berlin, Germany, 2016.
[163] L. Marcus, M. Giusti, and S. Barthel, "Cognitive affordances in sustainable urbanism: contributions of space syntax and spatial cognition," *Journal of Urban Design*, vol. 21, no. 4, pp. 439–452, 2016.
[164] K. Samuelsson, M. Giusti, G. D. Peterson, A. Legeby, S. A. Brandt, and S. Barthel, "Impact of environment on people's everyday experiences in Stockholm," *Landscape and Urban Planning*, vol. 171, pp. 7–17, 2018.
[165] W. S. Rauws, M. Cook, and T. Van Dijk, "How to make development plans suitable for volatile contexts," *Planning Practice & Research*, vol. 29, no. 2, pp. 133–151, 2014.
[166] O. Yaski, J. Portugali, and D. Eilam, "City rats: insight from rat spatial behavior into human cognition in urban environments," *Animal Cognition*, vol. 14, no. 5, pp. 655–663, 2011.
[167] R. O. Kaaronen, "Affording sustainability: adopting a theory of affordances as a guiding heuristic for environmental policy," *Frontiers in Psychology*, vol. 8, p. 1974, 2017.
[168] C. Rawlinson and M. Guaralda, "Out of control," in *Proceedings of the 8th International Design and Emotion Conference*, pp. 1–12, University of the Arts London, Central Saint Martins College of Art & Design, New Zealand, 2012.
[169] M. Hasanov and J. Beaumont, "The value of collective intentionality for understanding urban self-organization," *Urban Research & Practice*, vol. 9, no. 3, pp. 231–249, 2016.
[170] S. Zhang, G. De Roo, and T. Van Dijk, "Urban land changes as the interaction between self-organization and institutions," *Planning Practice & Research*, vol. 30, no. 2, pp. 160–178, 2015.
[171] "Landscape and urban planning," *International Journal of Landscape Science, Planning and Design*, vol. 139, p. 63, 2015.
[172] S. Nooteboom, "Impact assessment procedures for sustainable development: a complexity theory perspective," *Environmental Impact Assessment Review*, vol. 27, no. 7, pp. 645–665, 2007.
[173] B. Boonstra and L. Boelens, "Self-organization in urban development: towards a new perspective on spatial planning," *Urban Research & Practice*, vol. 4, no. 2, pp. 99–122, 2011.
[174] S. Moroni and S. Cozzolino, "Action and the city. Emergence, complexity, planning," *Cities*, vol. 90, pp. 42–51, 2019.
[175] E. Lissandrello and J. Grin, "Reflexive planning as design and work: lessons from the port of amsterdam," *Planning Theory & Practice*, vol. 12, no. 2, pp. 223–248, 2011.
[176] G. Ferilli, P. L. Sacco, and G. Tavano Blessi, "Beyond the rhetoric of participation: new challenges and prospects for inclusive urban regeneration," *City, Culture and Society*, vol. 7, no. 2, pp. 95–100, 2016.
[177] P. Marcuse, "From critical urban theory to the right to the city," *City*, vol. 13, no. 2-3, pp. 185–197, 2009.
[178] D. Sommer and P. L. Sacco, "Optimism of the will," *Sustainability*, vol. 11, no. 3, p. 688, 2019.
[179] R. Van Melik, I. Van Aalst, and J. Van Weesep, "Fear and fantasy in the public domain: the development of secured and themed urban space," *Journal of Urban Design*, vol. 12, no. 1, pp. 25–42, 2007.
[180] P. L. Sacco, S. Ghirardi, M. Tartari, and M. Trimarchi, "Two versions of heterotopia: the role of art practices in participative urban renewal processes," *Cities*, vol. 89, pp. 199–208, 2019.
[181] O. Devisch, A. Poplin, and S. Sofronie, "The gamification of civic participation: two experiments in improving the skills of citizens to reflect collectively on spatial issues," *Journal of Urban Technology*, vol. 23, no. 2, pp. 81–102, 2016.
[182] S.-K. Thiel and P. Fröhlich, "Gamification as motivation to engage in location-based public participation?" *Progress in Location-Based Services 2016*, pp. 399–421, Springer, Berlin, Germany, 2017.
[183] R. J. Lin, S. Ramakrishnan, H. Chang, S. Spraragen, and X. Zhu, "Designing a web-based behavior motivation tool for healthcare compliance," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 23, no. 1, pp. 58–67, 2013.
[184] A. L. Hillman, "Expressive behavior in economics and politics," *European Journal of Political Economy*, vol. 26, no. 4, pp. 403–418, 2010.
[185] I. S. Mayer, "The gaming of policy and the politics of gaming: a review," *Simulation & Gaming*, vol. 40, no. 6, pp. 825–862, 2009.
[186] K. Heljakka and P. Ihamäki, "Out of the classroom, into the park: gamified art experiences and learning through play in urban playscapes," in *Proceedings of the 16th Annual Hawaii International Conference on Art & Humanities*, pp. 9–11, Honolulu, Hawaii, January 2018.
[187] J. Banfield and B. Wilkerson, "Increasing student intrinsic motivation and self-efficacy through gamification pedagogy," *Contemporary Issues in Education Research*, vol. 7, no. 4, pp. 291–298, 2014.
[188] H.-Y. Cheng, S.-W. Jian, D.-P. Liu, T.-C. Ng, W.-T. Huang, and H.-H. Lin, "Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset," *JAMA Internal Medicine*, vol. 180, no. 9, p. 1156, 2020.
[189] Y. J. Park, Y. J. Choe, O. Park et al., "Contact tracing during coronavirus disease outbreak, South Korea, 2020," *Emerging Infectious Diseases*, vol. 26, no. 10, pp. 2465–2468, 2020.
[190] S. Park, G. J. Choi, and H. Ko, "Information technology-based tracing strategy in response to COVID-19 in South Korea-privacy controversies," *Jama*, vol. 323, no. 21, p. 2129, 2020.
[191] R. Steinbrook, "Contact tracing, testing, and control of COVID-19-learning from taiwan," *JAMA Internal Medicine*, vol. 180, no. 9, p. 1163, 2020.
[192] F. Rowe, O. Ngwenyama, and J.-L. Richet, "Contact-tracing apps and alienation in the age of COVID-19," *European*

Journal of Information Systems, vol. 29, no. 5, pp. 545–562, 2020.

[193] S. Altmann, L. Milsom, H. Zillessen et al., "Acceptability of app-based contact tracing for COVID-19: cross-country survey study," *JMIR mHealth and uHealth*, vol. 8, no. 8, Article ID e19857, 2020.

[194] O. Plonsky and I. Erev, "To get people to adopt tracing applications, minimize the probability they will regret it," 2020.

[195] V. V. Makarov, A. E. Hramov, D. V. Kirsanov et al., "Interplay between geo-population factors and hierarchy of cities in multilayer urban networks," *Scientific Reports*, vol. 7, no. 1, 2017.

[196] U. Rossi, "The multiplex city," *European Urban and Regional Studies*, vol. 11, no. 2, pp. 156–169, 2004.

[197] S. Angel, A. M. Blei, D. L. Civco, and J. Parent, *Atlas of Urban Expansion*, Lincoln Institute of Land Policy Cambridge, Massachusetts, MA, USA, 2012.

[198] R. Florida, A. Rodriguez-Pose, and M. Storper, "Cities in a post-COVID world," in *Papers in Evolutionary Economic Geography (PEEG) 2041*, Utrecht University, Department of Human Geography and Spatial Planning, Utrecht, Netherlands, 2020.

[199] R. B. Patel and T. F. Burke, "Urbanization-an emerging humanitarian disaster," *New England Journal of Medicine*, vol. 361, no. 8, pp. 741–743, 2009.

[200] L. Potvin and C. M. Jones, "Twenty-five years after the ottawa charter: the critical role of health promotion for public health," *Canadian Journal of Public Health*, vol. 102, no. 4, pp. 244–248, 2011.

[201] M. Eriksson and B. Lindström, "A salutogenic interpretation of the Ottawa Charter," *Health Promotion International*, vol. 23, no. 2, pp. 190–199, 2008.

[202] M. Castleden, M. McKee, V. Murray, and G. Leonardi, "Resilience thinking in health protection," *Journal of Public Health*, vol. 33, no. 3, pp. 369–377, 2011.

[203] K. Y. Lai, C. Webster, S. Kumari, and C. Sarkar, "The nature of cities and the Covid-19 pandemic," *Current Opinion in Environmental Sustainability*, vol. 46, pp. 27–31, 2020.

[204] M. D. Lee, "Gamification and the psychology of game design in transforming mental health care," *Journal of the American Psychiatric Nurses Association*, vol. 22, no. 2, pp. 134–136, 2016.

[205] A. Tonkin and L. Whiting, *Play and Playfulness for Public Health and Wellbeing*, Taylor & Francis Group, Milton Park, UK, The Science of Public Health, 1st edition, 2019.

[206] B. E. Kok, K. A. Coffey, M. A. Cohn et al., "How positive emotions build physical health," *Psychological Science*, vol. 24, no. 7, pp. 1123–1132, 2013.

[207] E. Grossi, G. Tavano Blessi, and P. L. Sacco, "Magic moments: determinants of stress relief and subjective wellbeing from visiting a cultural heritage site," *Culture Medicine and Psychiatry*, vol. 43, no. 1, pp. 4–24, 2019.

[208] L. D. Frank, M. A. Andresen, and T. L. Schmid, "Obesity relationships with community design, physical activity, and time spent in cars," *American Journal of Preventive Medicine*, vol. 27, no. 2, pp. 87–96, 2004.

[209] H. W. Kohl, C. L. Craig, E. V. Lambert et al., "The pandemic of physical inactivity: global action for public health," *The Lancet*, vol. 380, no. 9838, pp. 294–305, 2012.

[210] K. Gerst-Emerson and J. Jayawardhana, "Loneliness as a public health issue: the impact of loneliness on health care utilization among older adults," *American Journal of Public Health*, vol. 105, no. 5, pp. 1013–1019, 2015.

[211] P. L. Sacco, *Economia Della Cultura*, vol. 27, p. 165, 2017.

[212] D. Fancourt and S. Finn, 2019.

[213] C. D. Ryff and B. H. Singer, "Know thyself and become what you are: a eudaimonic approach to psychological well-being," *Journal of Happiness Studies*, vol. 9, no. 1, pp. 13–39, 2008.

[214] P. Miquelon and R. J. Vallerand, "Goal motives, well-being, and physical health: an integrative model," *Canadian Psychology/Psychologie canadienne*, vol. 49, no. 3, pp. 241–249, 2008.

[215] The concept of salutogenesis broadens the scope of public health policies from mainly addressing diseases and pathogenic factors to promoting healthy environments, individual and collective attitudes, and lifestyles, and more generally all kinds of factors that contribute to the maintenance of a healthy condition.

[216] The eudaimonic approach to well-being is based upon a balanced joint pursuit of physical, psychological, social and developmental dimensions of human existence.

Hindawi
Complexity
Volume 2021, Article ID 8632086, 14 pages
https://doi.org/10.1155/2021/8632086

Research Article

Mapping Population Dynamics at Local Scales Using Spatial Networks

José Balsa-Barreiro,[1,2] **Alfredo J. Morales**,[1] **and Rubén C. Lois-González**[1,2]

[1]*Human Dynamics, MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2]*ANTE Research Group, Galician Studies and Development Institute (IDEGA), Santiago de Compostela 15782, Spain*

Correspondence should be addressed to José Balsa-Barreiro; jobalbar@gmail.com

Received 14 July 2020; Revised 1 October 2020; Accepted 25 March 2021; Published 31 May 2021

Academic Editor: Ãtila Bueno

Nowadays, around half of the global population lives in urban areas. This rate is expected to increase up to two-thirds by the year 2050. Most studies analyze urban dynamics in wide geographic ranges, focusing mainly on cities. According to them, the global population is spatially distributed (and polarized) in two extremes: large urban agglomerations and rural deserts. However, this remark is excessively general and imprecise. For this reason, it remains essential to analyze these dynamics at other spatial scales. A close-up look in thinly populated regions shows how urban dynamics are also noticeable. In this paper, we analyze spatiotemporal patterns of population distribution in a predominantly rural area by applying a local-scale approach. These patterns are represented by using spatial networks with nodes representing the human settlements and links showing hierarchies between nodes. This case study is conducted in a small municipality located in northwestern Spain. It is a predominantly rural area with a very particular spatial pattern of population distribution.

1. Introduction

The global population is increasingly concentrated in urban areas. According to United Nations (UN), more than 4 billion people live nowadays in urban areas, which represent 55.2 percent of the global population [1]. The most eye-catching aspect related to urbanization is the continuous growth in recent decades. In 1960, only 33.6 percent of the global population was living in cities, 21.6 percent less than today. This same trend towards urbanization will continue in the upcoming years. Recent UN projections forecast that the rate of urban population is expected to be more than two-thirds by 2050.

The spatial pattern of the global population is increasingly polarized. Large megalopolises with millions of inhabitants coexist with immense empty spaces. These spatial inequalities are apparent from different perspectives. Territorially, around 50 percent of the global population is currently concentrated in just around one percent of the planet's surface [2]. Over time, developing countries are under a more strong urbanization process compared to developed countries in the past. It favors the emergence of *overurbanization processes* in these countries showing rates of urban population considerably larger than expected for their levels of economic development and wealth concentration [3, 4].

In the past, people have mostly lived in very low-density rural settings. In 2007, the rate of urban population at a global scale exceeded 50 percent by showing how the major relevance of urban dynamics is relatively recent. It is particularly notorious in the last two centuries with the emergence of the industrial revolution. Over time, the urbanization process experienced by western countries was relatively slow in line with the emergence of industrial activities and wealth concentration in cities. In consequence, people from rural areas migrated to cities where the most of labor opportunities were concentrated. Nowadays, developing countries located in East Asia and Africa experience a very rapid increase in urbanization rates [5]. Conversely, the generalized lack of future job opportunities in rural areas explains the massive migration to cities. Thus, the population flows that developed countries experienced in the

last two centuries are being replicated nowadays in developing countries, but more rapidly. It explains the exponential growth rates experienced in large cities located in developing countries without adequate infrastructures to support their urban growth. The emergence of *pseudourbanization* or *false-urbanization* processes [6, 7] helps to understand the majority of negative dynamics related to these areas in terms of poverty, marginality, social deprivation, and increasing violence rates. At a global scale, traces drawn by the centers of gravity related to relevant sociodemographic indicators show this trend. In Balsa-Barreiro et al. [8], it is observed how the global wealth is moving towards the global East, while traces related to the increase of population and urbanization rates are shifting to the global South.

Cities concentrate people, goods, means of production, and services. These offer great benefits due to the proximity between labor opportunities and potential workforce, which allow them to reduce transportation costs and to reach a more efficient use of resources, among other benefits. Urban regions favor a larger and more flexible labor market, where companies find a vast reservoir of the workforce and where workers can find a great number of employment opportunities. The proximity between both agents within urban areas allows to increase labor productivity and to boost the potential exchange of knowledge and ideas [9]. For these reasons, population and wealth are likely to grow at once within cities [10], which helps to understand why the richest countries are *urban economies* [11]. However, the relationship between urban population and wealth is not strictly linear. Dobbs and Remes [12] analyzed the quantitative weight of the 2,600 largest global cities. These cities concentrated 38 percent of the global population, but 72 percent of global GDP. This study evidences the emergence of enormous inequalities between cities and regions on a global scale [13].

The attractiveness of cities and rapid urbanization step up conflicts related to aspects such as gentrification, social segregation and polarization, pollution, and mobility. An example related to their environmental impact is as follows: cities account nowadays for more than 70 percent of global greenhouse-gas emissions and city dwellers generate more than 2 billion tons of waste per year, a rate that is expected to increase to 3.40 billion tons by 2050 [14]. In this sense, experts warn about the emergence of the so-called *urban diseconomies* which refer to a bunch of negative externalities derived from constrained mobility, the poor accessibility between districts in terms of travel-times, the predatory living costs, and/or the excessive employment competitiveness in cities, among other factors [15–17].

Positive and negative externalities related to urbanization are distributed across the territory in an unbalanced way. Although the vast majority of the world is nonurban, it becomes more dependent on cities than ever. Cities are always dependent on external resources in terms of land, water, energy, and food, among others. These demands increase the pressure on agrarian and forestry lands surrounding cities, overloading natural landscapes and leading to forced changes in land uses. A well-established example is shown in the repeated emergence of intentional wildfires in areas surrounding many south European cities [18, 19].

In response to this, experts and policy authorities must offer solutions according to the principles of sustainable development, territorial convergence, and social cohesion. Thus, all people should have access to the same labor opportunities, irrespective of their place of origin. Strongly encouraged by this statement, the *New Urban Agenda* adopted by the UN at the Habitat III conference [20] was exclusively focused on the power of cities as driving forces for sustainable development at a global scale.

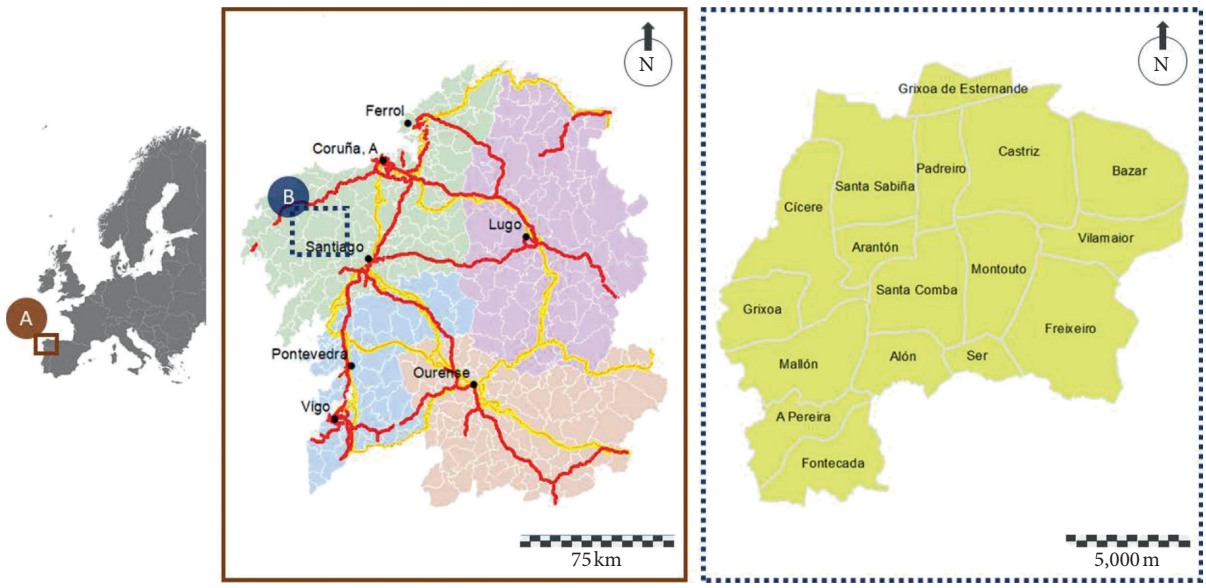
On this basis, a comprehensive perspective of the entire territory beyond the cities is still lacking. One essential issue refers to the multiscalability of urbanization processes, an aspect that is addressed in this paper. Our objective here is to identify urban patterns in territories that are not properly cities by using a local-based approach. For this purpose, we analyze the spatial variability in population flows in a small Spanish municipality in a dataset covering an extensive time period, starting in the late 19th century.

The remainder of this paper is organized as follows. In Section 2, we contextualize the study area presenting its main geographical aspects. In Section 3, we explain the methodology used for mapping spatial networks. In Section 4, we conduct a multiscale analysis of the population change in our study area since the late 19th century. Results are shown in charts and spatial networks depending on the spatial scale and data aggregation. These results are discussed subsequently in Section 5. Finally, we close with the conclusions of the study where the most relevant aspects are summarized.

2. Study Area

The study area corresponds to a small municipality located in Galicia, in northwestern Spain (Figure 1). The municipality of Santa Comba has 9,635 inhabitants distributed in 203 square kilometers [21]. Its population density is 47.5 inhabitants per square kilometer, which is around half compared to its region (92 inhabitants). Its main town, named identically as the municipality, is the most populated settlement counting 28.3 percent of the total population (2,731 inhabitants in 2015). According to Goerlich et al. [22, 23], it would be an eminently rural municipality, without a single human settlement exceeding 5,000 inhabitants. This accords with the methodology employed by Eurostat, which considers eminently rural municipalities with a population density lower than 100 inhabitants per square kilometer [24]. On the one hand, according to Zoido and Arroyo [25], this municipality is at the intermediate interval between rural and urban, with a population lower than 10,000 inhabitants.

The region of Galicia, where the study area is located, presents a particular spatial pattern of population distribution. Traditionally, this spatial pattern was characterized by high fragmentation and dissemination across the entire territory. This region concentrates nearly 50 percent of all the singular population entities located in Spain. In comparison, this rate is about 10 times larger than its demographic weight at a national level [21]. The municipality of Santa Comba is



(a) (b)

Figure 1: Location of the study area. The municipality of Santa Comba is located in the northwestern sector of the region of Galicia, Spain. In the box (*A*), the black-labeled nodes correspond to the most populated cities, whereas the indicated lines show the most important road infrastructures. High-capacity roads are displayed in red color and railways in yellow. In box (*B*), the distribution in parishes within the study area is shown. (a) Zoom *A*. (b) Zoom *B*.

located next to the so-called *Atlantic Axis*, the area where the major and most thriving cities within the whole region are located. However, despite its geographical closeness, this municipality decreases its population over a long period showing a similar behavior like the surrounding area located on the west side: the so-called *Costa da Morte* [26, 27].

Like the entire region, the internal distribution in parishes within our particular study area presents a great significance to fully understand its territorial structure nowadays (Figure 1). The *parish* is a territorial figure with an ecclesiastic origin and without any official administrative competence. However, it constitutes an essential figure for understanding social and power relationships between people, especially in rural areas. This figure is the key to understand the topology and evolution of the spatial pattern of population distribution within this region.

Within our particular study area, a total of 175 human settlements distributed across 17 parishes were registered by official population censuses published since the late 19th century. Some of these settlements were eventual because they emerged and/or faded away sporadically in just some of the censuses.

Population data were collected for the period from 1888 to 2015. These data were extracted from the *Nomenclator de Población*, an official census published by the *Spanish Statistical Office* [28]. Albeit with some exceptions, this dataset is published every ten years.

3. Methodology

In this paper, we analyze the internal population flows in a rural area covering an extensive time period, starting from the late 19th century. Spatial networks are implemented for mapping these flows. These spatial networks consist of nodes and links. The nodes represent human settlements and these are located according to their spatial coordinates. The links connect the nodes according to a two-folded hierarchical criterion: territorial dependence and total population.

The implementation of these spatial networks is shown graphically in Figure 2. The most populated human settlement of the entire municipality determines the main node, which is shown in Figure 2(a). This node is the central hub for a first-order network, where this main node connects with the most populated settlement in each parish. In total, 17 nodes and 16 links compose this first-order network, which is shown in Figure 2(b). A second-order network is implemented for each parish. This network connects the most populated settlement in each parish with all the other settlements within this same parish. Thus, a bunch of 17 individual minor networks is implemented. These second-order networks are shown in Figure 2(c). Finally, the entire spatial network results from merging the first-order network and all the independent second-order networks (Figure 2(d)). All the human settlements officially reported since the late 19th century are included in this comprehensive spatial network.

The topology of the spatial network shown in Figure 2 is adapted to one particular year, that is, in this case, 2015. This topology varies depending on the population data in each census year. Thus, the most populated settlement in the whole municipality determines the topology of the first-order network. The most populated settlement in each parish not only determines the topology of the different second-order networks but also defines partially the topology of the first-order network at its edges. Similarly, the number of nodes depends on the total number of settlements officially reported in one particular census year.

Figure 2: Implementation of the spatial network according to 2015 census data. (a) The most populated settlement for the whole study area is represented by a red node. This node is the central hub of the whole network. (b) A first-order network is established by linking this central hub with the most populated settlement in each parish. (c) A bunch of second-order networks is established by linking the most populated settlement in each parish with all the settlements within this same parish. (d) The final network results from merging the first-order and second-order spatial networks.

The historical population flows among settlements within our study area are shown in Section 4. There we compare population data for the whole network in two different years. The size of nodes represents the population in the most recent year. The hierarchy of links is based on the population data in the most recent year. The color of nodes refers to relative variations in the number of inhabitants between both years. We use a simple color legend where only three colors depict quantitative variations in the number of inhabitants for each node. Three additional colors are used for mapping the emergence or demise of any settlement in a concrete census year. In case of settlements not officially reported for one of the census years, these are represented, but not connected to the network. Finally, the color of nodes designates the color of links. In the case of a link connecting two differently colored nodes, it will show a color gradation between edges. To gain a detailed view, we present graphically this methodology in Figure 3.

4. Results

Our study area counted 9,635 inhabitants in 2015, 563 inhabitants more than in 1888. However, this growth was not constant over time. Data show two oppositional trends separated by a clear turning point in 1960 when this municipality peaked at 13,951 inhabitants (Figure 4). Since then, this municipality had lost nearly one-third of its population. The reason behind this is the collapse of traditional agrarian societies, which forced a large number of people to emigrate. Before the 1960s, most of these people migrated to South

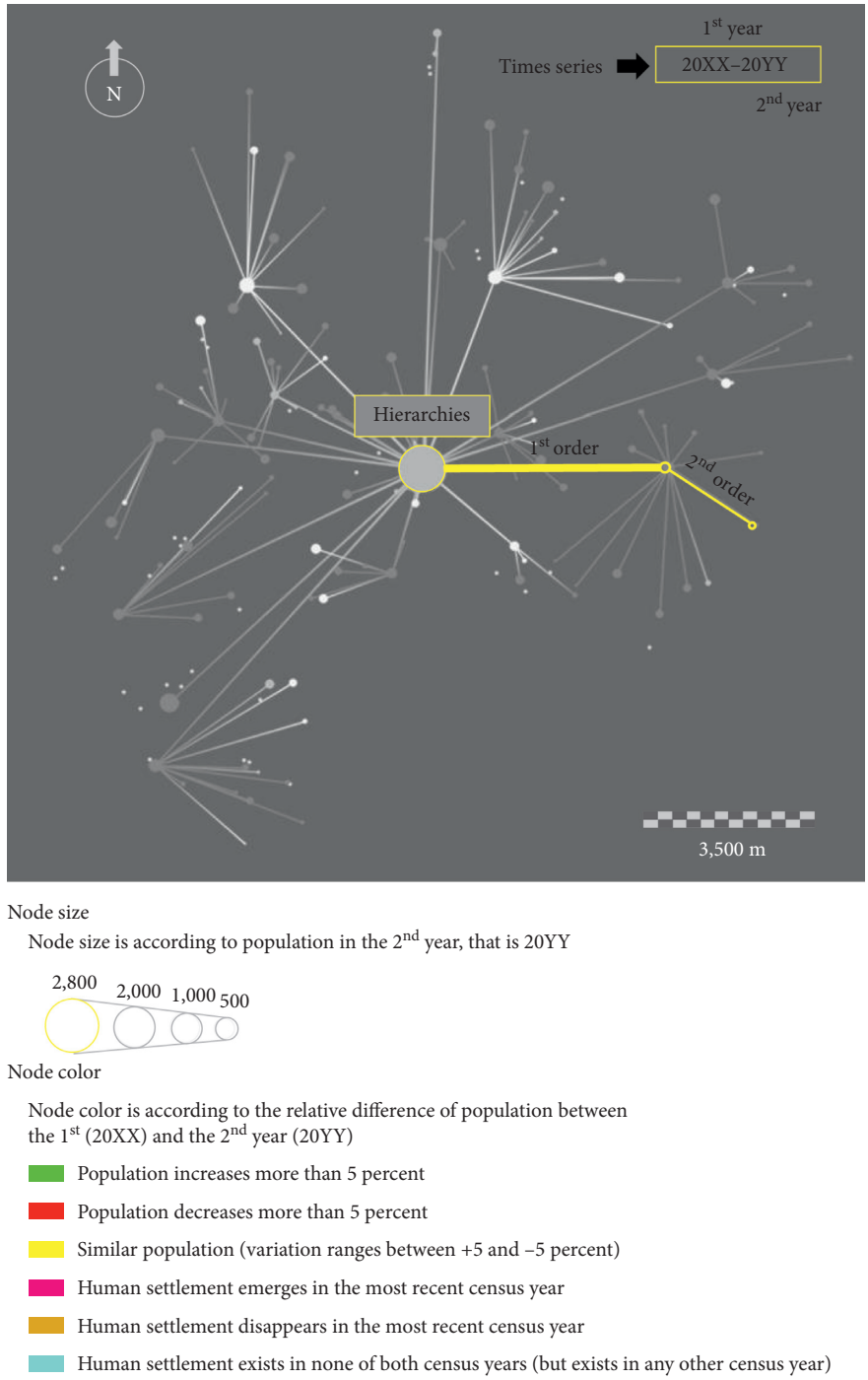


Figure 3: Methodology for mapping population dynamics using spatial networks. This methodology is applied in Section 4.

American countries and later on there was a trend towards Central European countries. In the last decades, the majority of new migrants decided to move to the most important cities nearby [29].

A more comprehensive analysis of data allows us to observe important changes in the spatial pattern of population distribution. Figure 5 gives a synthetic view of an acute internal redistribution of the population at different scales over time. The size of boxes represents the relative population for all the parishes and settlements. Each minor box represents a single settlement. The aggregation of minor boxes with the same color refers to those which are part of the same parish. Three different years were considered: 1888, 1960, and 2015. 1880 and 2015 correspond to the first and the last census years, respectively. The 1960 census is relevant because it is the peak and the turning point in the whole time series. The complete dataset for each year is organized hierarchically in these treemaps. Boxes sharing color are spatially distributed according to their areas. The minor boxes are located close to the bottom-right corner and the

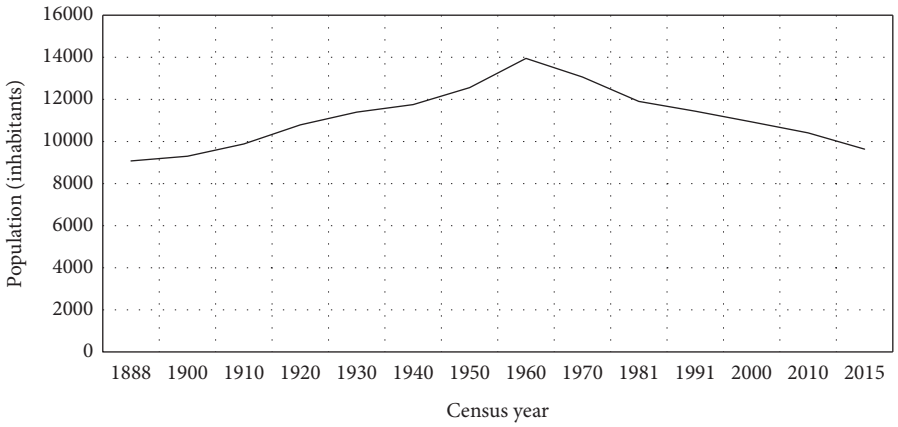
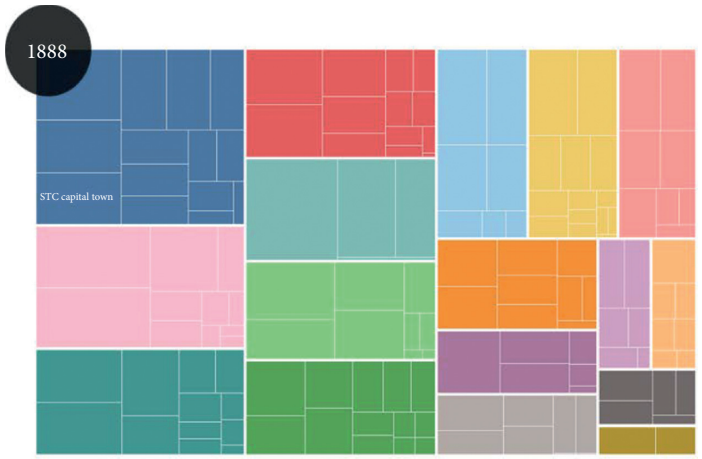
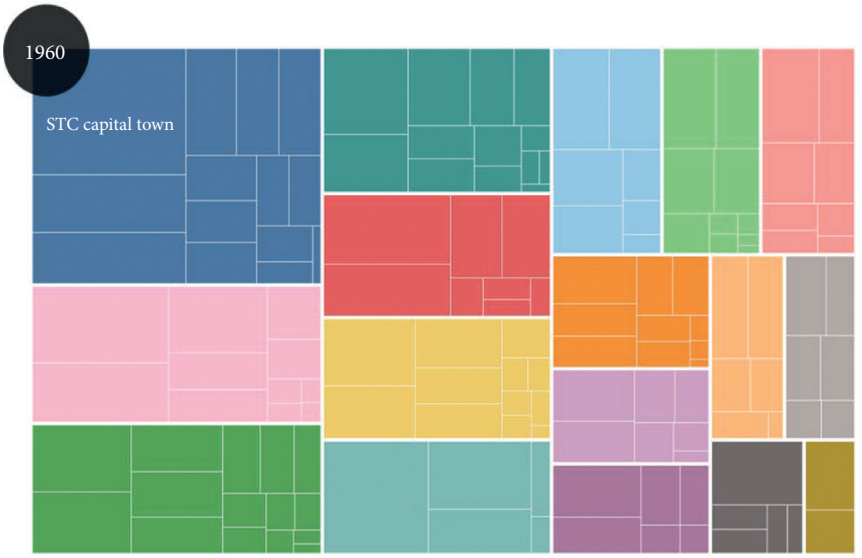


Figure 4: Population in the study area between 1888 and 2015.

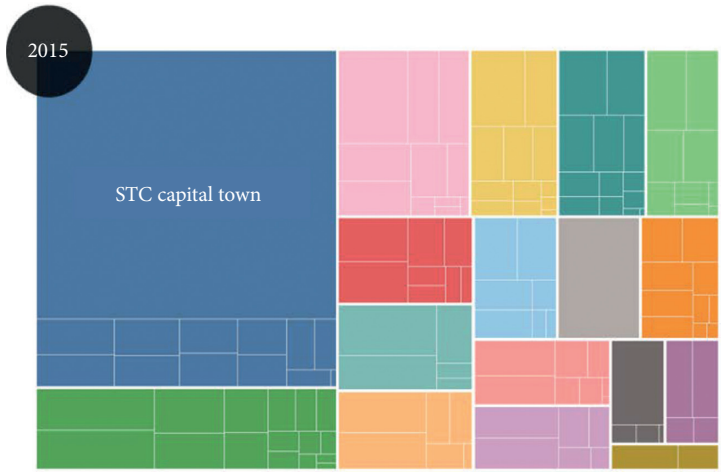


(a)



(b)

Figure 5: Continued.



(c)

Figure 5: Cross-census comparison between (a) 1888, (b) 1960, and (c) 2015. The size of the complete chart for each year represents the total population in the whole municipality. Each settlement corresponds to a minor box, whose size depends on its relative population. The boxes with the same color are part of the same parish.

major ones close to the top-left corner. The color legend for the same parish is kept for all the years.

The most evident difference is the utmost importance of the capital town, whose box is labeled. According to data, it is the most populated node since 1940. In 1888, it was only the ninth most populated settlement in this municipality, with only 1.7 percent of the total population. In 2015, its population was 28.3 percent, a relative weight 16 times larger. In Figure 6, we represent the relative growth over time of the capital town in comparison to the rest of the municipality before and after 1960 and the time when the maximum population is reached.

As we can observe, both lines tend to diverge after 1960. This means that the population is increasingly concentrated in one single node: the capital town. Although this node counted in 2015 with only 2,731 inhabitants, it can be equated to a child-size city. Thus, this capital town agglomerates a relevant number of urban activities that are proper central places such as business activities, commercial stores, public endowments, and office services, among others. Meanwhile, the rest of the settlements endures a growing crisis in terms of population.

Therefore, despite this study area being predominantly rural and thinly populated, we observe an acute internal redistribution of population. In this way, it demonstrates how urban dynamics result to be scale-independent, emerging even in thinly populated regions. In our study area, this is more evident after 1960, when the capital town only represented 4.7 percent of the total population. Half a century later, its population represents over six times more (28.3 percent), showing average growth rates of around 4 percent in relative terms in the last few decades (Figure 7).

A more comprehensive analysis for better understanding the actual population dynamics within this municipality is using spatial networks and fine-grained data over the whole time series. The following figures show a cross-census comparison for all the settlements officially reported. This is done using spatial networks, where nodes represent human settlements and links show hierarchies among nodes. The methodology for data mapping was previously introduced in Section 3.

Figure 8 shows the population dynamics for the periods running from 1888 to 1960, and from 1960 to 2015. In the left figure, the spatial network presents a very balanced configuration related to their nodes. Many nodes presented a similar size and the green color was predominant. This means that the population was evenly distributed across the territory, in a face of population growth numbers. In the right figure, the main node adopts an extreme significance in terms of size and color. Thus, most of the network show negative population dynamics, which is the exact opposite of what is happening with the main node.

These same population dynamics are represented in Figure 9 for the periods in between successive censuses, which is around ten years for most of the cases. Among other important aspects, we can observe a clear and progressive variation of the predominant colors over time, shifting from mostly green to red tones in the most recent decades. Furthermore, some very significant modifications in the topology of the whole network are shown. The clearest change emerges after 1940 when the current capital town became the most populated node, which is located in the center of the whole network.

5. Discussion

Geographic science analyzes the relationships among natural and social systems by considering all their interdependences over the territory. Although many of these relationships were already addressed in the past, their spatial dynamics seem to be very diverse and complex. For this reason, it is strongly recommended to implement a multiscale approach for checking the spatial behavior of these dynamics at different spatial scales. The *analytical geography* proposes the

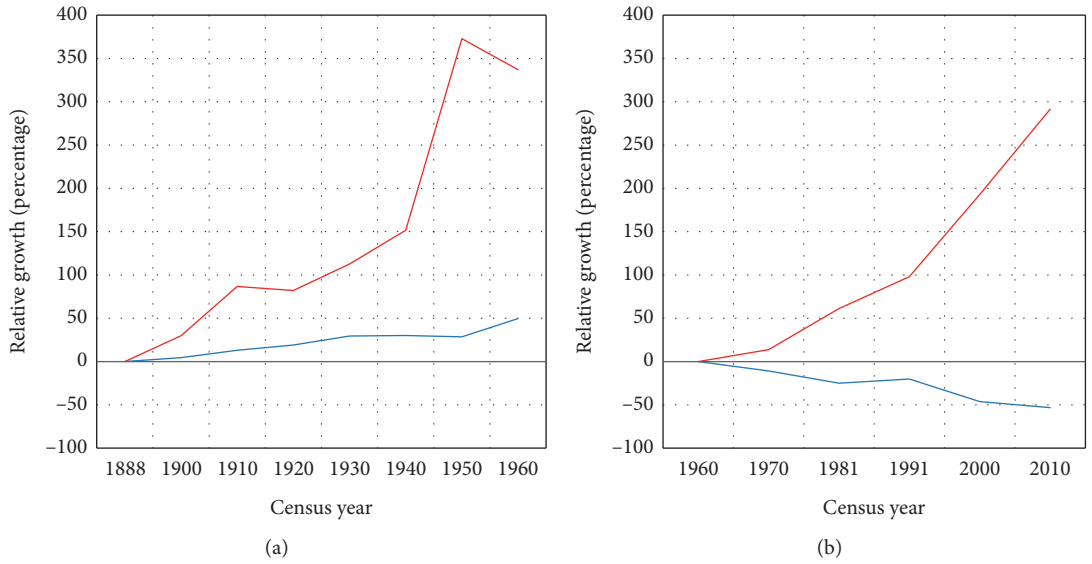


Figure 6: Relative growth of the capital town (red line) in comparison to the rest of the municipality excluding the capital town (blue line) before and after 1960.

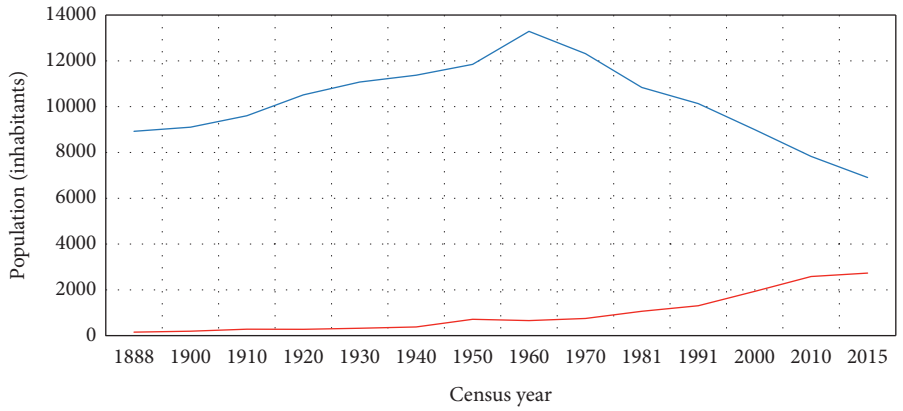


Figure 7: Population in the capital town (red line) and the whole municipality excluding the capital town (blue line) between 1888 and 2015.

use of cartographic models based on networks, flows, and graphs to measure this complexity across the territory.

In this paper, we show how dynamics only related to urban spaces emerge also in thinly populated regions. For years, dynamics related to urbanization were only analyzed in large cities and metropolises. However, these dynamics and related mechanisms do not only emerge in populated cities, but these are also evident in predominantly rural regions. Our study area is a good sample. In population terms, this area is facing a very regressive trend for the last half-century, losing nearly a third of the population. However, it has experienced a very acute process of internal redistribution of the population, where people have moved from the vast majority of rural settlements to the capital town. According to the spatial networks, the settlements located at the edges have reduced their importance, while the centrally located node is taking up more weight.

In broad terms, the segregation between urban and rural spaces is increasingly reinforced over time. Spatial networks show how this process is progressive, being the smallest nodes the first ones to disappear. It seems to be a factor more relevant than the geographical distance of nodes concerning the most populated town, which only counts with 2,731 inhabitants. However, this town has the power to attract people from the whole network, behaving like a *child-size city* at a local scale.

Depopulation in rural areas is more than a purely quantitative process. Beyond losing population, rural areas present increasingly weaker structures from a qualitative perspective. The young people of working ages are precisely the most expelled group, while elderly people are multiplying. This explains why rural areas experience a progressive weakening of their demographic structure and, in consequence, the population decline tends to be more severe over time.

The use of spatial networks for mapping population dynamics is very adequate for simplifying their complexity. This mapping strategy presents some advantages compared to choropleth maps, thanks to the structure based on nodes

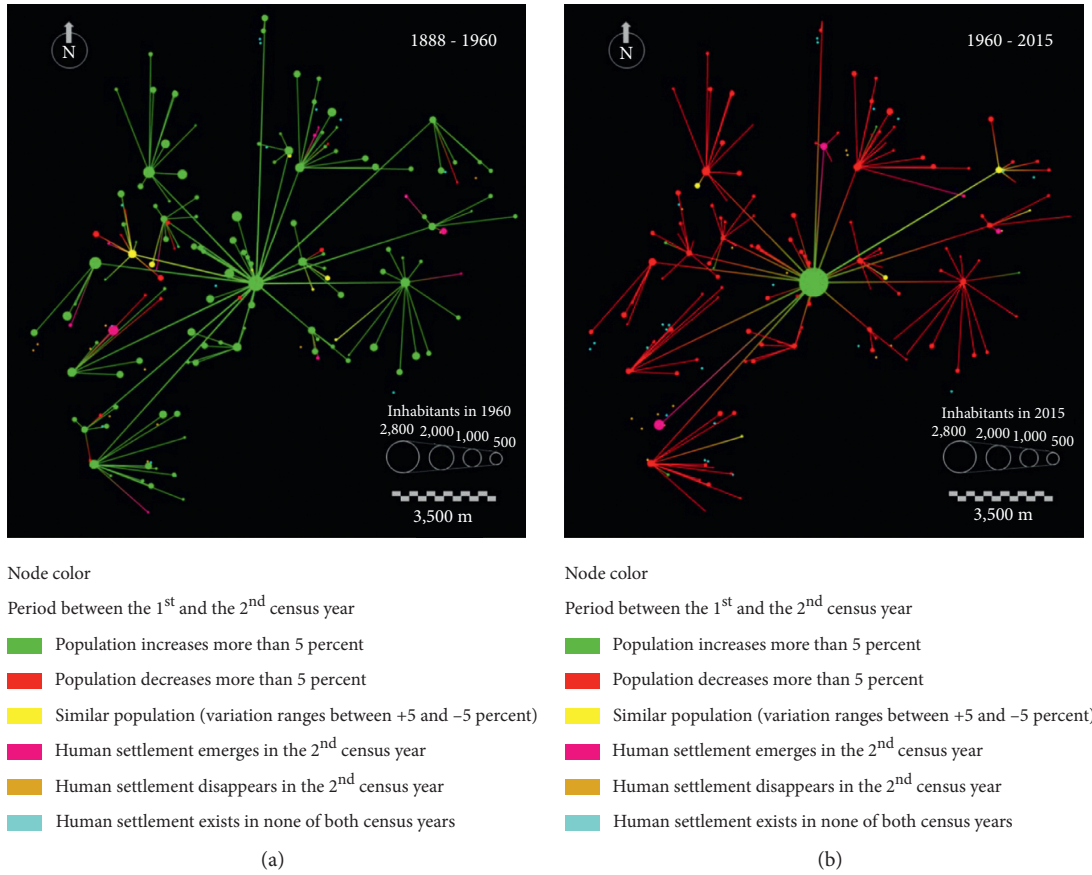
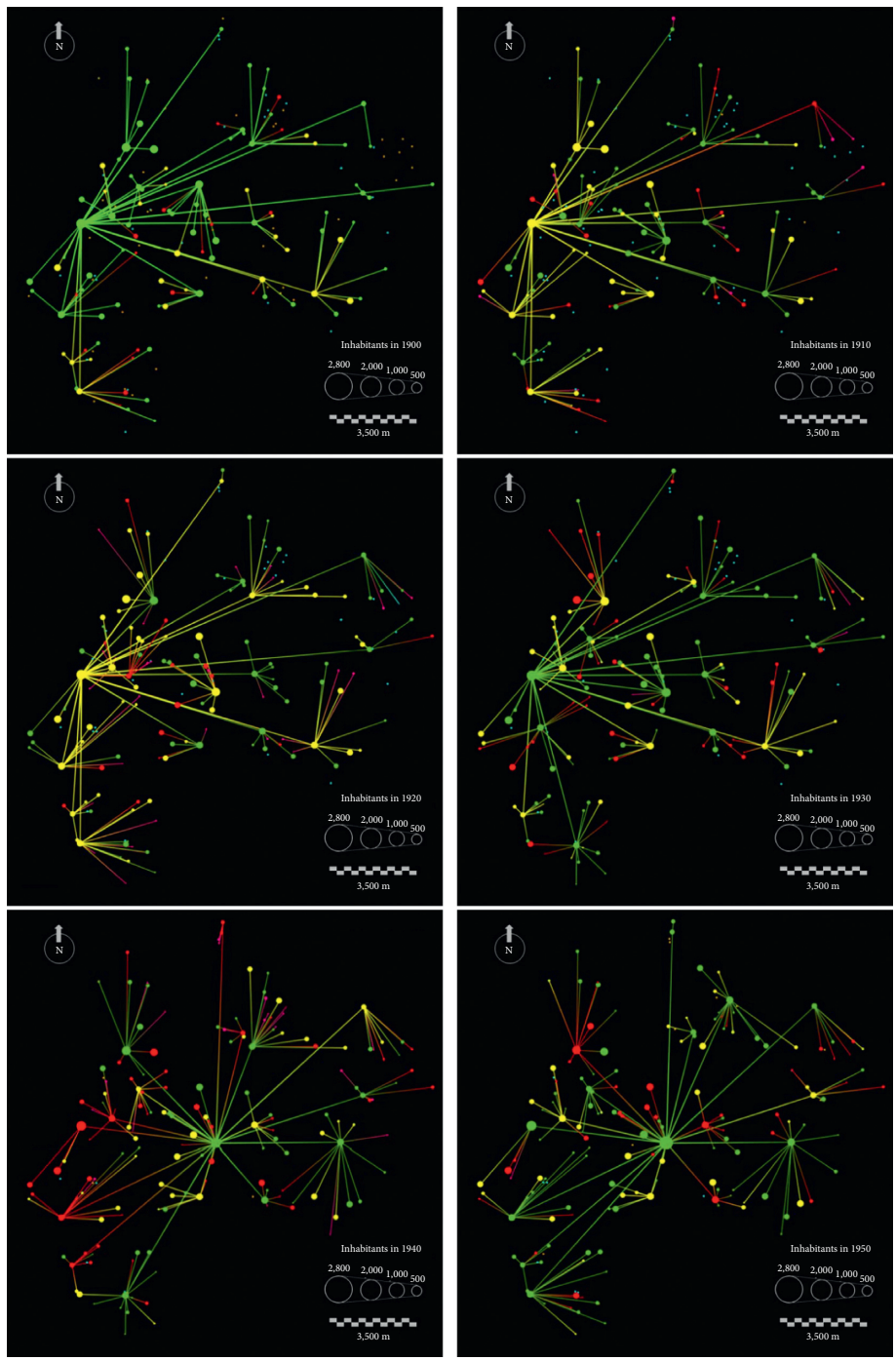


Figure 8: Population change from 1888 to 1960 (left figure) and from 1960 to 2015 (right figure). Node size refers to the total population in the most recent year. Node color shows the relative variation between the first and the second year in each figure.

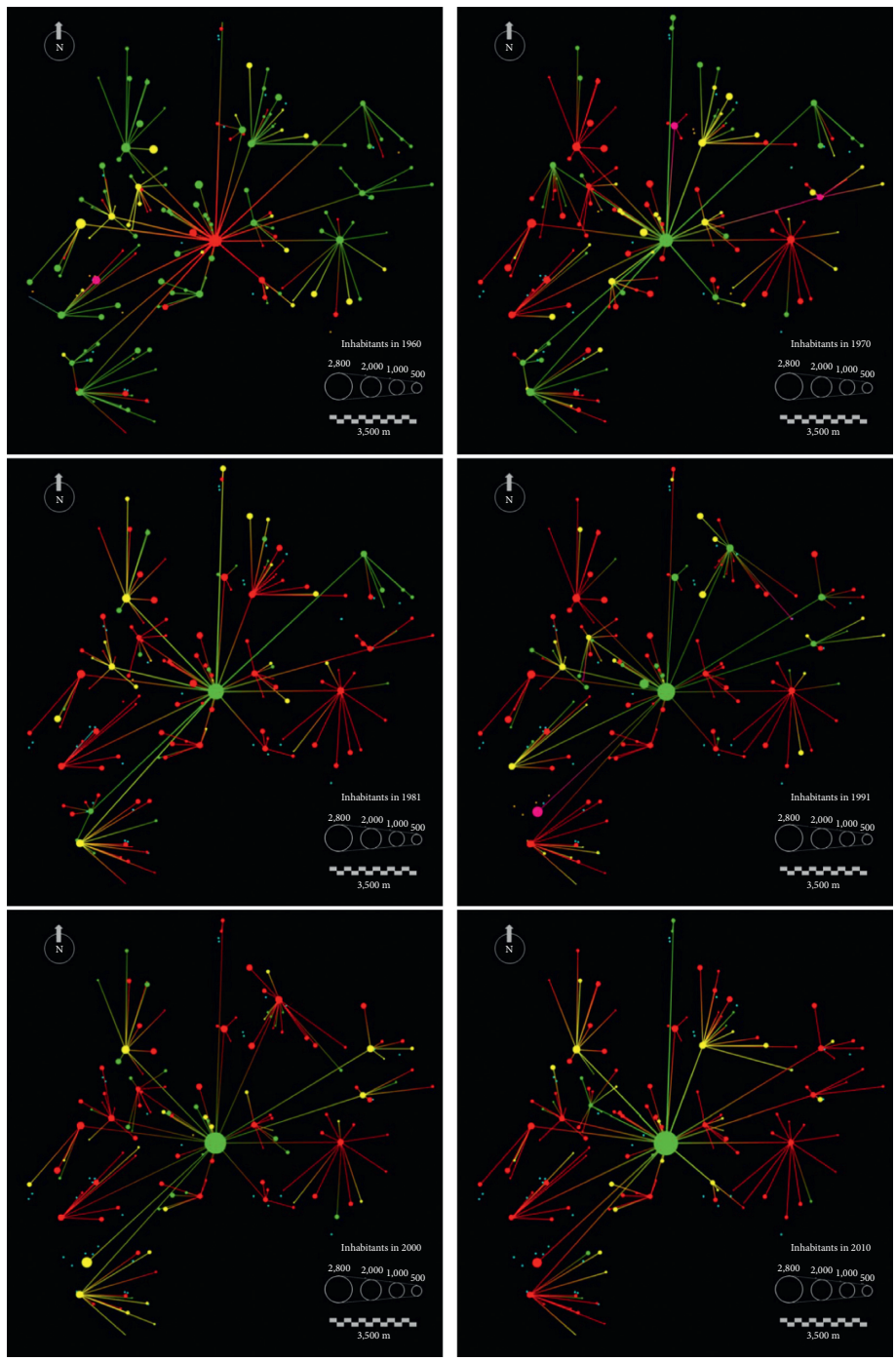
and links. We can represent not only fine-grained population data by nodes but also the changing hierarchies between nodes by links. It is important to note that links do not represent actual population flows in real-time, but these can be a proxy indicator. We also must emphasize that the spatial networks are not a closed system. Migration abroad was a very important factor for understanding the population dynamics in this municipality. Many people migrated to Brazil until 1960 and Central Europe afterward. Each of these migration flows presented its particularities. The migration to Brazil was mainly permanent with a great number of migrants that never returned. The migration to Central European countries was mainly temporary with most of these migrants returning after some years. However, most of them did not return to their hometowns but instead they settled down in the capital town where many of them opened new businesses. Their investments contributed decisively to the rapid growth of the capital town [29, 30].

We must carefully consider some limitations in this study. First, our study area belongs to a region with specific particularities from a demographic perspective. On the one hand, the spatial pattern of population distribution diverges substantially from the rest of Spain. In 2015, an average municipality in Spain counted with 5,716 inhabitants distributed in 7.6 settlements, whereas in Galicia it counted with 8,709 inhabitants distributed in 96.3 settlements. Therefore, an average settlement in Spain had approximately 752 inhabitants, eight times more population than one located in the region of Galicia [28]. The most outstanding aspect behind this spatial pattern is the fragmentation and dissemination of population across the region. Another relevant aspect refers to major changes in the topology in some of the spatial networks shown. In this study, the most significant change is visible after 1940 with the shift from an agrarian society to the prevailing one, dominated by an urban economy that is mostly concentrated in the capital town. Traditional economies in this region were based on small farms for family survival. The familiar structure counted with many children who were expected to contribute like major labor forces. Spatially, this demographic model was traditionally based on a balance between population and resources. In the event of any temporary imbalance, migration abroad turned into the main alternative for many families [31–33]. The transition to a new model was evident after the 1960s with the end of subsistence farming, the attraction of industrial cities, and the subsequent development of service-based economies. In our study area, the number of dwellers in the main node was exponentially increased since 1960 (+313.8 percent). The capital town is nowadays the most important trading center for an extensive region beyond the administrative borders of this municipality. Also, the major changes in our spatial network, after 1940, show how the *parish*, a territorial figure with great relevance



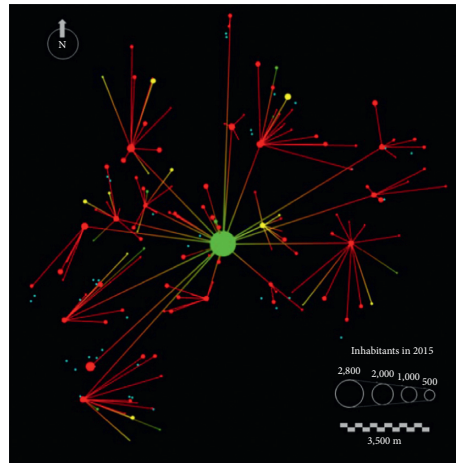
(a)

Figure 9: Continued.



(b)

Figure 9: Continued.



Node color

Period between the 1st and the 2nd census year

- Population increases more than 5 percent
- Population decreases more than 5 percent
- Similar population (variation ranges between +5 and −5 percent)
- Human settlement emerges in the 2nd census year
- Human settlement disappears in the 2nd census year
- Human settlement exists in none of both census years

(c)

Figure 9: Population change for all the intercensal periods from 1888 to 2015. The period represented in each figure is shown in the upper right-hand corner. Node size refers to the total population in the most recent year. Node color shows the relative variation between the first and the second year in each figure. (a) 1888–1900, (b) 1900–1910, (c) 1910–1920, (d) 1920–1930, (e) 1930–1940, (f) 1940–1950, (g) 1950–1960, (h) 1960–1970, (i) 1970–1981, (j) 1981–1991, (k) 1991–2000, (l) 2000–2010, and (m) 2010–2015.

in the past, had lost most of its influence due to the secularization process and the decline of social power by the Church.

Another relevant aspect is the spatial relationship between population and resources nowadays. Population decline results in losing competitiveness and job positions, and vice versa. This creates a downward spiral for adopting unsustainable strategies for territorial management [26, 34]. The ever-increasing concentration of people and wealth is leading to a distorted vision of the territory based on purely urban perspectives. It would explain why rural areas are increasingly under pressure, which is observed in forced changes of land uses in rural areas surrounding cities. According to the science of complex networks, those systems that concentrate an excessive number of interdependencies in a small group of nodes are more vulnerable by presenting more probabilities to fail [35, 36].

In response, policies for sustainable development must be encouraged to achieving a real convergence between urban and rural territories. This is particularly important in the context of the COVID-19 pandemic, which might lead to major changes in the spatial pattern of population distribution. In this sense, Kotkin [37] anticipates the end of the so-called *megacities* and the coming age of dispersion, with more people willing to live in less densely populated regions.

We have adopted a local-based approach in this study to verify the multiscalability of population dynamics. It is now crucial to adopt the right interventions to counterbalance the negative externalities. Small case studies like the one presented might be field trials for the actual implementation of policies and interventions. In this way, this can be an efficient way to assess the results and to optimize the costs.

In short, we have identified the presence of urban dynamics in areas rarely addressed in the mainstream literature on urban studies. Somehow, this paper lays the groundwork for future research aiming to deepen the understanding of urban dynamics at any spatial scale. Our findings can be extrapolated to other different regions across the world and these may contribute to adopting the right policies for promoting resilience in territorial management. Our findings are valuable for experts such as urban planners, policymakers, and other competent authorities with expertise in territorial management and spatial planning.

6. Conclusions

The objective of this paper is to demonstrate the multiscalar nature of certain urban dynamics. For this, we analyze population dynamics in a thinly populated region in decline for a long time series. Spatial networks are implemented to compare population data between two different years. The results show how only one of 175 nodes increases continuously its size, exactly the opposite of what occurs in the rest of the nodes. The process of concentration of population and

the segregation between rural and urban spaces is visually displayed.

A recurrent downward spiral of depopulation and economic decline threatens the sustainability of the majority of rural regions. In response, policy-makers and experts must implement the right policies to adopt integrated territorial management without overemphasizing urban territories.

Data Availability

Population data used in this research are available at the *Spanish Statistical Office* website for the most recent decades. Historical data are only available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank Heike M. Grünsteudel for her contribution to review this paper.

References

[1] Department of Economic and Social Affairs, *World Urbanization Prospects: 2018 Revision*, Department of Economic and Social Affairs, UN Population Dynamics, New York, NY, USA, 2018.

[2] T. Misra, "Half the world's people live on 1 percent of its land," 2016, https://www.theatlantic.com/international/archive/2016/01/world-population-distribution/422871/.

[3] M. Timberlake and J. Kentor, "Economic dependence, overurbanization, and economic growth: a study of less developed countries," *The Sociological Quarterly*, vol. 24, no. 4, pp. 489–507, 1983.

[4] P. Graves and R. Sexton, "Overurbanization and its relation to economic growth for less developed countries," *Economy Forum*, vol. 8, no. 1, pp. 95–100, 1979.

[5] D. H. Potts, "Whatever happened to Africa's rapid urbanization?" *World Economics*, vol. 13, no. 2, pp. 17–29, 2012.

[6] Y. Liu, Z. Li, and J. Jin, "Pseudo-urbanization or real urbanization? Urban China's emergence of administrative regions and its effects: a case study of Zhongshan city, Guangdong province," *China Review*, vol. 14, no. 1, pp. 37–59, 2014.

[7] S. Batishcheva, "World urbanization prospects and the problem of its infrastructural provision," *Economic Analysis*, vol. 46, no. 1-2, pp. 72–81, 2013.

[8] J. Balsa-Barreiro, Y. Li, A. Morales, and A. S. Pentland, "Globalization and the shifting centers of gravity of world's human dynamics: implications for sustainability," *Journal of Cleaner Production*, vol. 239, Article ID 117923, 2019.

[9] A. Pentland, *Social Physics: How Social Networks Can Make Us Smarter*, Penguin Books, New York, NY, USA, 2014.

[10] I. S. Gill and C.-C. Goh, "Scale economies and cities," *The World Bank Research Observer*, vol. 25, no. 2, pp. 235–262, 2010.

[11] S. A. Frick and A. Rodríguez-Pose, "Change in urban concentration and economic growth," *World Development*, vol. 105, pp. 156–170, 2018a.

[12] R. Dobbs and J. Remes, "Trends. The shifting urban economic landscape: what does it mean for cities?" in *Proceedings of the World Bank's Sixth Urban Research and Knowledge Symposium* Washington, DC, USA, 2012.

[13] M. Chen, H. Zhang, W. Liu, and W. Zhang, "The global pattern of urbanization and economic growth: evidence from the last three decades," *PLoS One*, vol. 9, no. 8, Article ID e103799, 2014.

[14] World Bank, "Solid waste management," 2019, https://www.worldbank.org/en/topic/urbandevelopment/brief/solid-waste-management.

[15] S. A. Frick and A. Rodríguez-Pose, "Urban concentration and economic growth VOX CEPR Policy Portal," 2018, https://voxeu.org/article/urban-concentration-and-economic-growth.

[16] O. Serpell, "Wrong about urbanization? How emerging factors could shift people away from cities," 2018, https://kleinmanenergy.upenn.edu/policy-digests/wrong-about-urbanization.

[17] R. Florida, "The myth that urbanization means prosperity," 2017, https://www.theatlantic.com/business/archive/2017/12/developing-world-city-size-urbanization/548468/.

[18] J. Balsa Barreiro and T. Hermosilla, "Socio-geographic analysis of the causes of the 2006's wildfires in Galicia (Spain)," *Forest Systems*, vol. 22, no. 3, pp. 497–509, 2013.

[19] A. Badía and M. Pallarés, "Spatial distribution of ignitions in Mediterranean periurban and rural areas: the case of Catalonia," *International Journal of Wildland Fire*, vol. 15, no. 2, pp. 187–196, 2006.

[20] Habitat III conference (2016): "The United Nations Conference on Housing and Sustainable Urban Development". Quito (Ecuador), http://habitat3.org.

[21] Spanish Statistical Office (INE), https://www.ine.es, 2019.

[22] F. J. Goerlich and I. Cantarino, "Estimaciones de la población rural y urbana a nivel municipal," *Estadística Española*, vol. 57, no. 186, pp. 5–28, 2015.

[23] E. Reig, F. J. Goerlich and, and I. Cantarino, *Delimitación de áreas rurales y urbanas a nivel local. Demografía, coberturas de suelo y accesibilidad*, Fundación BBVA, Bilbao, Spain, 2016.

[24] L. Dijkstra and H. Poelman, "A harmonised definition of cities and rural areas: the new degree of urbanisation," European Commission, Brussels, Belgique, 2014.

[25] F. Zoido and A. Arroyo, "La población de España," in *Tendencias demográficas durante el siglo XX en España*, Arroyo, Ed., pp. 17–75, Spanish Statistical Office, Madrid, Spain, 2003.

[26] J. Balsa-Barreiro, "Unsustainability of territorial models from a demographic point of view: the case of A Costa da Morte (Galicia, Spain)," *Papeles de Población*, vol. 19, no. 78, pp. 167–206, 2014.

[27] J. Balsa-Barreiro and S. Landsperger, "A Costa da Morte (Galicia, España): un modelo demográfico antagónico al español. Análisis de su evolución demográfica en el siglo XXI," *Journal of Iberian and Latin American Research*, vol. 21, no. 1, pp. 63–86, 2015.

[28] Spanish Statistical Office (INE), "Nomenclátor: población del Padrón Continuo por Unidad Poblacional," 2015, https://www.ine.es/nomen2/index.do.

[29] J. Balsa Barreiro, "Evolución del urbanismo y de la disposición territorial del poblamiento en un municipio coruñés de interior: caso de Santa Comba (Galicia, España)," *Revista de Urbanismo*, vol. 24, pp. 76–118, 2011.

[30] J. C. Sánchez, "Bases para el análisis geohistórico del poblamiento rural tradicional en Galicia," *Boletín de la Asociación de Geógrafos Españoles*, vol. 62, pp. 75–99, 2013.

[31] X. Martínez-Filgueira, D. Peón, and E. López-Iglesias, "Intra-rural divides and regional planning: an analysis of a traditional emigration region (Galicia, Spain)," *European Planning Studies*, vol. 25, no. 7, pp. 1237–1255, 2017.

[32] S. Calvo-Iglesias, U. Fra-Paleo, and R. A. Diaz-Varela, "Changes in farming system and population as drivers of land cover and landscape dynamics: the case of enclosed and semi-openfield systems in Northern Galicia (Spain)," *Landscape and Urban Planning*, vol. 90, no. 3-4, pp. 168–177, 2009.

[33] A. Bouhier, "La Galice, essaigéographiqued'analysed'interpretation d'un vieuxcomplexeagraire," Ph.D. dissertation, University of Poitiers, La Roche-Sur-Yon, France, 1979.

[34] C. Jones, "The end of economic growth? Unintended consequences of a declining population," *National Bureau of Economic Research*, vol. 26651, 2020.

[35] J. Balsa-Barreiro, A. Vie, A. Morales, and M. Cebrian, "Deglobalization in a hyper-connected world," *Palgrave Communications*, vol. 6, p. 28, 2020.

[36] A. Vie and A. Morales, "How connected is too connected? Impact of network topology on systemic risk and collapse of complex economic systems," *Computational Economics*, vol. 15, 2020.

[37] J. Kotkin, "The coming age of dispersion," 2020, https://quillette.com/2020/03/25/the-coming-age-of-dispersion/.

Hindawi
Complexity
Volume 2021, Article ID 1283012, 19 pages
https://doi.org/10.1155/2021/1283012

Research Article

Network Structure of Intercity Trips by Chinese Residents under Different Travel Modes: A Case Study of the Spring Festival Travel Rush

Rong Zhang, Jinghu Pan, and Jianbo Lai

College of Geography and Environmental Science, Northwest Normal University, Lanzhou 730070, China

Correspondence should be addressed to Jinghu Pan; panjh_nwnu@nwnu.edu.cn

Received 23 June 2020; Revised 11 March 2021; Accepted 21 March 2021; Published 1 April 2021

Academic Editor: Burçin Bozkaya

With the advent of big data, the use of network data to characterize travel has gradually become a trend. Tencent Migration big data can fully, dynamically, immediately, and visually record the trajectories of population migrations with location-based service technology. Here, the daily population flow data of 346 cities during the Spring Festival travel rush in China were combined with different travel modes to measure the spatial structure and spatial patterns of an intercity trip network of Chinese residents. These data were then used for a comprehensive depiction of the complex relationships between the population flows of cities. The results showed that there were obvious differences in the characteristics of urban networks from the perspective of different modes of travel. The intercity flow of aviation trips showed a core-periphery structure with national hub cities as the core distribution. Trips by train showed a core-periphery structure with cities along the national railway artery as the core. This gradually decreased toward hinterland cities. Moreover, the intercity flow of highway trips indicated a spatial pattern of strong local aggregation that matched the population scale.

1. Introduction

By highlighting social and economic factors, networks of population flow can be developed. For example, social network analysis is an important interdisciplinary research method in recent years from the perspective of the relationships between the behaviors of subjects. The migration and flow of populations are regarded as activities with production factors allocated in space. These factors promote the reagglomeration and diffusion of social and economic factors [1]. Travel has a clear origin and destination, which, together with population flows during certain periods, constitute a population trip network. A population trip network relies on an urban network, with cities as the network nodes. The direction and intensity of population flows represent the relationship between the nodes. The emergence of a floating population not only changes the spatial distribution of the population but also affects the development of the regional economy [2]. It plays an important role in accelerating the development of the urban and rural economy, and it promotes urbanization, upgrades the industrial structure, and optimizes the regional allocation of labor resources [3]. The 2016 Report of China's Floating Population Development estimated that China's floating population reached 247 million in 2015, accounting for 18% of the total population. In addition, the direction of Chinese population flow has changed in recent years, with the labor force, especially migrant workers, returning from eastern coastal cities to central and western cities, and labor- and resource-intensive industries transferring to central and western regions [4].

The phenomenon of population flow around the world cannot be ignored in the development of today's era. The flow of population between cities is represented by a particular type of network structure [5]. Wei et al. [6] proposed that a population flow network is a typical directed-weighted geographic network. Most existing research on population flow in China is concentrated around the Spring Festival [2, 7]. However, some scholars have analyzed the characteristics of population flow on National

Day and the subsequent Mid-Autumn Festival [1]. When selecting data, most researchers have traditionally used provincial census data and sample survey data [8–10]. However, with the rapid advancement of globalization and information technology, static data used in traditional research have been unable to meet the requirements for spatial analysis. It is more difficult to explore the increasingly complex relationships between cities using such data [11, 12]. A census cannot accurately grasp the scale of daily flow between cities and the characteristics of inflow and outflow, nor can it obtain the routes and direction of population migration within a relatively continuous time interval. This is because it only focuses on a specific time period, or analyzes the macro-patterns of population flow under laws of long-term evolution. Thus, it is impossible to analyze the increasingly complex interactions between cities from the perspective of flow space [13, 14].

As a direct representation of intercity interconnections and interactions, population migration has been a hot issue for geographers [15]. Current research focuses on long-distance intercity trip networks represented by aviation [16], daily intercity commuting networks represented by high-speed railways [17], and the characteristics and travel modes of intercity trip networks based on time-dependent microblog "check-in" data and network attention data [18]. The traffic routes that residents rely on for intercity trips are important tools for urban internal and external connections, and they reflect the ability and degree of urban connections and communication [19]. To some extent, traffic routes can be regarded as the basic support and key link of an urban network spatial structure. With the development of the social economy, the "spatiotemporal compression" effect caused by high speed travel modes, such as high speed rail and aviation, greatly improves the mobility of residents. Intercity trips are networked, integrated, dynamic, and personalized, and they are time- and mode-dependent. An intercity trip network with different time scales and modes can reflect complex geospatial connections.

Under the effect of globalization, urban space evolves from local space to flow space. The connections between cities form a hierarchical network; the horizontal connection between nodes is the point of emphasis in such a network model. In recent years, research on urban networks has emerged as a new paradigm of geography. Geographers in this field seek to identify the importance and characteristic attributes of urban nodes and analyze the hierarchical structure and relevance of urban networks [20]. Some geographers believe that global cities are the centrality for commanding and controlling global capital. They used a series of network analysis indicators, such as the degree centrality, betweenness centrality, and closeness centrality, to reveal the importance of network nodes [20]. In studies of global city networks based on the index of degree centrality, the centrality and power are usually equal. Boschken [21] and Alderson and Beckfield [22] described the positions of global cities in a network. However, "global cities are influential only insofar as they can influence hinterland cities [23]. Cook et al. [24] pointed out that when focusing on economic exchange, the dominant exchanges (with stronger power) are easier to control and influence exchange behavior, compared to those with a large number of potential opportunities (with higher centrality). Then, Zachary proposed the concepts of recursive centrality and recursive power in 2011 [25], and renamed them, respectively, alter-based centrality and alter-based power [26]. In addition, he thought that, in world city networks, the agglomeration of resource elements such as labor, capital, and information into world cities and the outward diffusion of resource elements from world cities are all a performance of centrality. Centrality is the unification of resource aggregation and diffusion. Power represents the influence and dominance of a city in the process of resource circulation. The power of a city is determined by the location of the network and the role it plays. Neal [25] drew two hypothetical world city network structure diagrams. He thought that, in comparison, the central node of the larger network is "central, but lack of power," and the central node of the smaller network is "power, but relatively lack of center." Through recursion, Neal explained that city centrality and power not only depend on the scale of network connections but also on the capital capacity (such as economic, cultural, and other similar representations) of its own and related branches [27]. Therefore, we selected these two indicators for a comprehensive evaluation of the status of a city in the network.

The objectives of this study were as follows: (1) to evaluate the intercity trip characteristics and intensity of three trip networks during the Spring Festival travel rush using population flow data, to simulate the path of the population flow process in 346 cities in China; (2) to explore spatial differences between the three trip networks, and reveal the complex structural characteristics of the intercity trip network; (3) to analyze the spatial structure of Chinese urban networks and measure the urban network hierarchy and aggregation spatial patterns under different travel modes; and (4) to explore the differences of urban network characteristics under different travel modes and discuss the rationality and necessity of spatiotemporal big data in the study of daily population flow.

2. Related Work

A network is composed of abstract nodes and edges of connecting nodes, and it abstractly describes the complex and intertwined objective world [28]. Cities are not isolated in regional space. They have a complex interaction with each other, thus forming an urban network with a specific spatial structure and functional organization [29]. Through the exchange of materials, information, finance, and population flow among cities, the interaction and complementary advantages among cities and regions are strengthened, forming complex networks of different scales and levels [30]. This kind of network takes the city as the center and the element of circulation as the medium. It forms a spatial structure of nodes, axes, and domains in a certain area. Under the background of the acceleration of globalization, the urban system of Western developed countries gradually presents a

trend of transformation from a hierarchical-scale model to a network model [31]. This prompted scholars to change their perspective and reexamine the new spatial structure of cities. The focus of their attention also changed from the urban-hierarchical scale, spatial form, and functional evolution to the structure and relationships of urban networks [2, 6, 32]. Castells [33] proposed the concept of a "space of flows." Flows, networks, and network nodes are summarized as the basic elements of "flow space." Among them, information flow, population flow, and capital flow are regarded as the flow elements, and companies, enterprises, cities, and countries are regarded as the network nodes. Different flow elements produce different network nodes. The different attributes of network nodes affect the movement of flow elements, as well as the patterns of the entire network. The transformation of an urban hierarchical system and functional division system is jointly driven by the "flow space" and the traditional "local space." The emergence of "flow space" makes the traditional centralized place model; based on the theory of hierarchical scale, it gradually turns to an open, flow-based, and polycentric network model [34]. Xu et al. [2] used network analysis to reveal the relationships between migration data and urban development.

With the development of Internet technology and intelligent terminals, the collection and analysis of resident trip data have diversified. The era of big data makes it possible to obtain resident movement patterns through massive spatial and temporal trajectories of individual granularity. Mobile computing devices with geo-positioning can be used to track individual spatial movements and record spatial and temporal data over a long time with high precision (bus swipe-card records, social network "check-in" data, the movement of taxis, etc.) [35, 36]. In China, many Internet companies provide location-based services to users, such as Baidu, Tencent, and Sina. The data in this study were obtained from the Tencent Migration big data platform (https://heat.qq.com/qianxi.php), and were downloaded using Python. These data were mainly obtained by third-party users using the positioning data provided by the Tencent location service. It covers most users' complete long-distance and short-distance trip behavior by taking the day as the statistical unit. Data on children, the elderly, and those who do not use location-based services are not available. To a certain extent, this avoids the underestimation and virtual increase of data caused by short-distance and long-distance trips. At the same time, the continuous development of geographic information technology facilitates data collection, storage, analysis, and visualization, and provides greater support for the study of the structural characteristics of trip networks [37, 38].

Recently, many scholars have used population trip data regarding aviation [39, 40], railways [41], and high speed rail [42] to summarize the spatial connections of national passenger traffic networks. Some scholars have conducted comparative studies of high-speed railways and airline networks in China and revealed laws of regional spatial organization [27]. These studies have enriched the research of different scales of urban network systems to a certain extent. However, they focused on spatial connections and the interaction between the cities from the perspective of a single type of traffic flow. In terms of intercity trip networks, researchers mainly focus on the use of long-distance travel surveys and commuting survey data to directly reveal the characteristics of intercity trip networks. Limtanakool et al. [43] revealed differences in behavior and network structure heterogeneity under different types of travel based on survey data regarding long distance interregional travel in Europe. De Montis et al. [44] analyzed the structural characteristics of intercity travel (commuting) in Italy using a complex network method. Based on the data regarding American aviation flow, Neal [45] systematically discussed the network characteristics of different types of aviation flow (viz., business flow and tourism flow) and different seasons of aviation flow (viz., summer and winter). High precision spatiotemporal information regarding population flow from location-based services provides sufficient and accurate measured flow data for studying travel and population flow. Li et al. [32] used Baidu migration data to analyze the population flow characteristics during the Spring Festival in China. Wei et al. [46] analyzed the characteristics of a Chinese urban network in the transitional period. These studies used big data to analyze individual and group behavior, and to reflect the spatial behavior, spatial cognition, and connection mode behind it. Such data can be used to reflect the decision-making of individuals and groups with regard to their spatiotemporal behavior, and is becoming a hot research frontier for studying travel and population flow.

3. Methods and Data Sources

We used the method of complex network analysis to identify and analyze the structural characteristics of the Chinese residents in an intercity trip network by using indicators such as the dominant flow, alter-based centrality, alter-based power, clustering coefficient, and network cluster structure analysis.

3.1. Dominant Flow. The dominant flow analysis method was first proposed by Nystuen and Dacey [47] in 1961. It is a mature method that simplifies the analysis of an urban network [37]. Briefly, the method involves judging the position of a city in an urban system based on the main factor flow from one city to another, including the maximum dominant flow, the second dominant flow, and other higher dominant flows [48]. In this study, we used dominant flow analysis to recognize the status of a city at the macroscale. The selection of the factor flow was the population flow intensity under different travel modes during the Spring Festival travel rush [20].

3.2. Alter-Based Centrality and Alter-Based Power. Tencent population migration data provide the direction and intensity of migration in various cities within a day under different modes of travel. Based on it, we constructed a bidirectional matrix $L = (L_{ij})$ to characterize the population

flow over the course of one day, where L_{ij} is the population flow intensity from city i to city j. There are 346×346 directional weighted asymmetric matrices.

$$L = \begin{array}{c} \\ i_1 \\ i_2 \\ . \\ . \\ . \\ i_{(n-1)} \\ i_n \end{array} \begin{array}{c} \begin{array}{ccccccc} j_1 & j_2 & . & . & . & j_{(n-1)} & j_n \end{array} \\ \begin{bmatrix} 0 & L_{12} & . & . & . & L_{1(n-1)} & L_{1n} \\ L_{21} & 0 & . & . & . & L_{2(n-1)} & L_{2n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ L_{1(n-1)} & L_{2(n-1)} & . & . & . & 0 & L_{n(n-1)} \\ L_{1n} & L_{2n} & . & . & . & L_{n(n-1)} & 0 \end{bmatrix} \end{array} \quad (1)$$

The population flow weight R_{ij} between city i and city j is calculated, and is regarded as the index for calculating the alter-based centrality and alter-based power:

$$R_{ij} = \frac{L_{ij}^T + L_{ij}}{2}. \quad (2)$$

When measuring the urban alter-based centrality (alter-based power), it is necessary to consider the impact of related cities on the measured city. For the data, there are a number of population flows between certain cities and satellite cities. For example, Langfang accounts for the majority of the population flow associated with Beijing. As a national economic and cultural center, Beijing has a relatively high degree of centrality, which leads to an indirect increase on the centrality of Langfang. Adding the dependency parameter "d_{ij}" can correct the phenomenon that the centrality result is difficult to describe. The status of the urban network and the centrality and power results of the entire network city tend to converge. In comparison, the status of a city with a balanced resource relationship will increase, while the status of a satellite city will decline significantly. This is more in line with the actual development of the city [20]. The formula is as follows:

$$d_{ij} = \frac{R_{ij}}{\mathrm{WDC}_i}, \quad (3)$$

where WDC_i is the weighted degree centrality of city i. Based on the correction, the formulas for the alter-based centrality and alter-based power are as follows:

$$\begin{aligned} \mathrm{AC}_i &= \sum_j \left(1 - d_{ij}\right) \times R_{ij} \times \mathrm{DC}_j, \\ \mathrm{AP}_i &= \sum_j \left(1 - d_{ij}\right) \times \frac{R_{ij}}{\mathrm{DC}_j}. \end{aligned} \quad (4)$$

3.3. Clustering Coefficient. The clustering coefficient is used to describe the interconnection level of nodes [28]. When certain nodes are closely connected, they can form a network cluster. We calculated the clustering coefficient as follows:

$$C_i = \frac{2B_i}{m_i\left(m_i - 1\right)}, \quad (5)$$

where C_i is the clustering coefficient, and B_i is the number of paths between the node and the neighboring nodes of m_i.

3.4. Data Sources. The data in this study were obtained from the Tencent Migration big data platform (https://heat.qq.com/qianxi.php). Tencent is one of the largest Internet-integrated service providers and one of the Internet enterprises with the largest number of service users in China. It mainly involves social networking, communication, games, and other aspects. The representative services are WeChat, QQ, and some Tencent software. In 2018, WeChat monthly active users reached 1.15 billion, and QQ monthly active users reached 650 million. With the ubiquity of smart devices, the number of Tencent users is increasing in China, so these data can cover most regions and users. This is also a new research direction in the study of population mobility.

Tencent location-service data is a value-added service whereby Tencent obtains the location information of mobile terminal users through radio communication networks or external positioning based on the location-based service. Users are provided with corresponding services under the support of a geographic information system platform. Today, mobile phones are widely used and have a positioning function. Thus, the user's trajectory can be recorded. In this way, a large amount of personal mobile information can be collected. Tencent location data were obtained on the premise of protecting user privacy. These data were updated every 24 hours, and covered most of the trip routes. Of course, the data also have the disadvantage that unfinished trips over 24 hours were disassembled. These data are mainly displayed by Excel tables, including the starting place, destination, and the number of people traveling by air, train, and car. In order to study the spatial structure of intercity trip networks in China, we selected 346 cities above the prefecture level as the research objects. We used data from 346 cities regarding trips taken by air, train, and road to represent the spatial correlation intensity of Chinese cities from the perspective of aviation, railways, and highways. Other geographic data were obtained from the National Geomatics Center of China. Figure 1 shows the location of major cities in China. Figure 2 shows the population of prefecture-level cities in China.

The Spring Festival travel rush is a unique social and economic phenomenon in China during the transition period. It mainly refers to the high pressure of various traffic modes caused by nationwide, large-scale population migration. The Chinese Spring Festival travel rush is not only a protracted, large-scale population flow, but also has a variety of purposes (e.g., college students and migrant workers return home to visit parents and relatives, and to sightsee). According to an official Chinese report, the floating population was nearly 3 billion during the Spring Festival travel rush in 2018, accounting for one-third of the world's population. Considering the representativeness of the research period and comparisons with similar research, we chose 40-day intercity trip data of Chinese residents during the Spring Festival travel rush as the basis for building the network. The analysis period was from February 1 to March

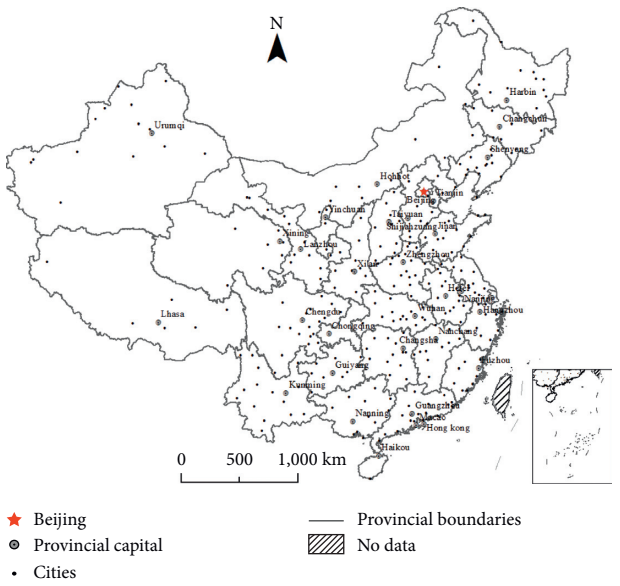


Figure 1: Locations of major cities in China.

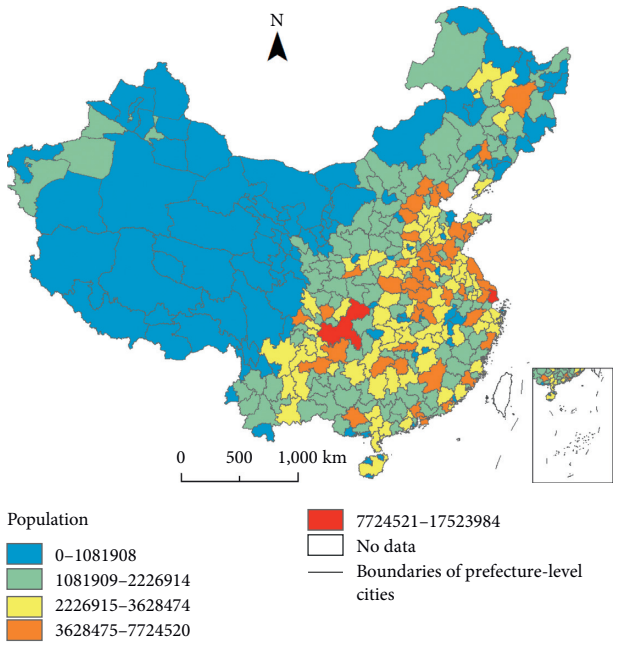


Figure 2: Population of prefecture-level cities in China.

12, 2018. Behind a large number of resident trips is the imbalance between the population and regional economic development. In this study, based on the daily data of population flow during the Spring Festival travel rush in 2018 obtained by Tencent location services, we established the trip relationship network of residents under different travel modes. We analyzed the spatial structure differences of Chinese urban networks under different trip modes, and provided a new perspective for the study of urban network structures. By studying the spatial structure of the intercity trip network of Chinese residents through different travel modes, the multiple spatial characteristics of population migration, resident trips, and urban networks could be revealed from different perspectives. This makes up for the one-sided conclusions drawn from the single-trip mode in the existing research and enriches the regional cognition of the spatial relationship between cities in China.

4. Results

4.1. Intercity Trip Pattern of Residents. We used ArcGIS to visualize the population flow routes and intensity under different travel modes. Figure 1 shows how the natural breakpoint classification method was adopted according to railway classification standards. In order to more clearly show the characteristics of the level of population flow, the first-level lines were not shown in the figure. As a whole, the population flow routes under the three travel modes all showed a pattern of sparseness in the west and density in the east. High population flow routes were concentrated in the east side of the "Huhuanyong Line." From the perspective of the China as a whole, there were different degrees of spatial differentiation and spatial dependence in the spatial connection intensity and spatial connection mode of travel in prefecture-level cities, and their hierarchical characteristics interact:

(1) From the perspective of aviation flow (Figure 3(a)), resident travel formed a "diamond" structure with Beijing, Shanghai, Guangzhou-Shenzhen, and Chengdu-Chongqing as the core. The eastern interregional interweaving phenomenon was significant. There are many long-distance routes, and most of them are connected with cities with a developed economic foundation, such as Shanghai-Chongqing, Chongqing-Beijing, Shanghai-Beijing, and Shenzhen-Shanghai. The population flow of these routes is more than 1.89 million people. In addition, there are routes associated with Xi'an, Guiyang, Wuhan, and Nanjing carrying 9.5 million to 1.89 million people, expanding around the "diamond" structure. Southwest, northwest, and northeast of China have fewer aviation routes. Most of them associate between provincial capitals and economically developed cities in the east, and the number of passengers is fewer than 950,000. The three northeastern provinces are closely connected with Beijing. The five provinces of Inner Mongolia, Xinjiang, Tibet, Qinghai, and Gansu have few connecting routes with each other. In the aviation travel network, people are most likely to choose aviation travel when traveling distances between 1,000 and 1,500 km (Figure 4). It has the characteristics of long distance and short time. The aviation flow clearly reflects the spatial connection and the core-periphery combination at the national scale, which plays a significant role in reflecting the national urban system structure at the macro-scale.

(2) From the perspective of railway flow (Figure 3(b)), in addition to the more obvious "diamond structure," the "three horizontal and one vertical"-shaped skeleton inside the diamond was also

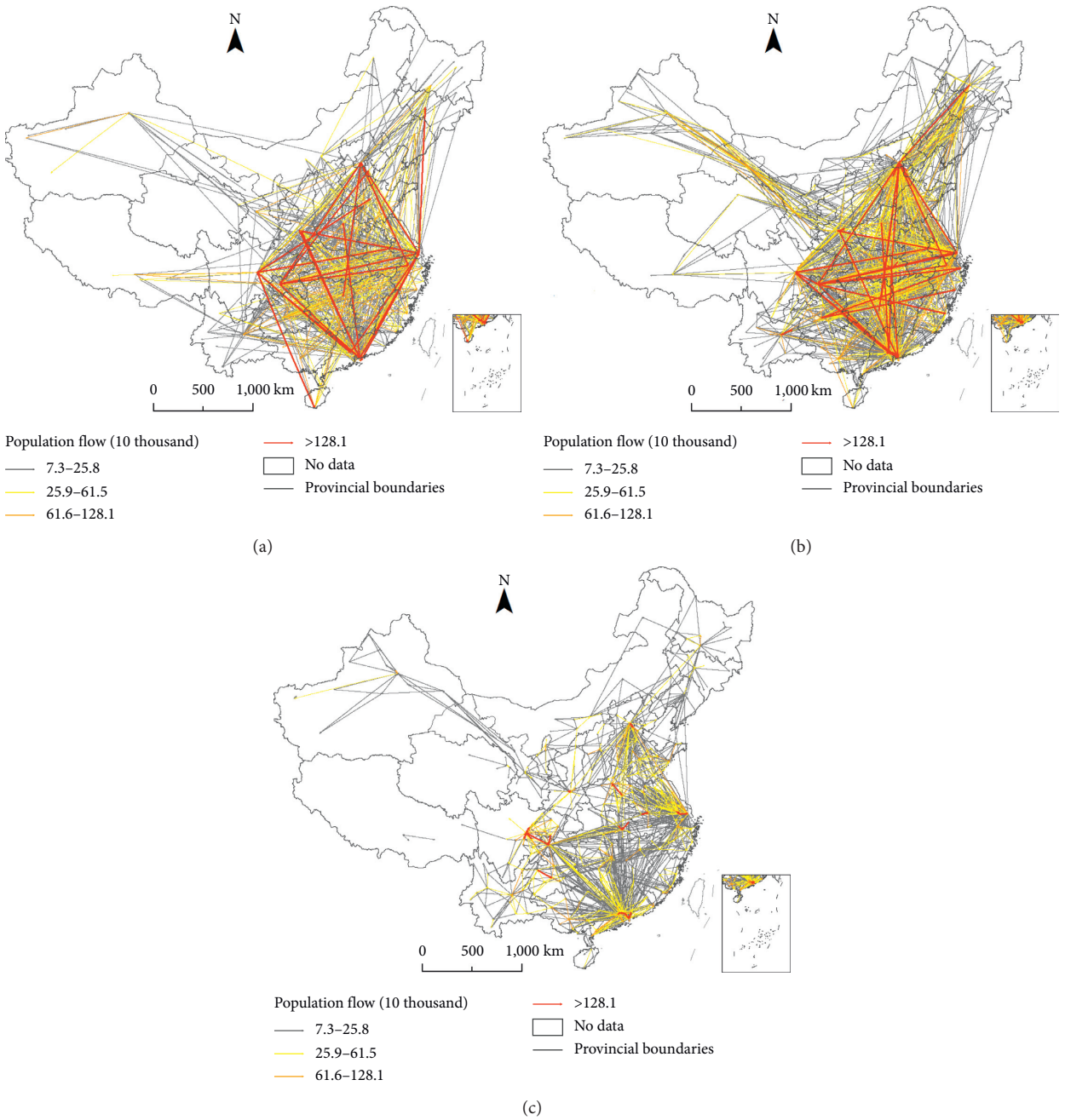


FIGURE 3: Routes and intensity of population flow under different travel modes: (a) aviation; (b) rail; (c) highways. The color of the route represents the intensity of population flow, and the change from gray to red represents an increase in the intensity.

relatively clear. The routes at all levels increased, covering a wide range and large density. Among the three modes of travel, the number of trips carried by rail was the largest, accounting for 49% of the total number of residents traveling in China. When the trip distance was less than 1,500 km, residents chose to travel by rail. Rail travel included long-distance routes like Kashgar to Shanghai, medium- and long-distance routes centered on Beijing, Chengdu, Guiyang, and Zhengzhou, and even short distance routes of only 8 km between Zhuhai and Macao, which is the most flexible way for residents to travel. The number of railway routes in northwestern and northeastern China is significantly more than that in aviation. Railway flow mainly reflects the travel patterns of residents in the hinterland cities along the national railway arteries, such as Beijing-Kowloon, Lanzhou-Xinjiang, Beijing-Shanghai, Beijing-Guangzhou, Lanzhou-Lianyungang, Beijing-Harbin, and Beijing-Baotou. The traffic infrastructure plays a significant role in guiding the spatial connections of travel, especially the "along-the-way effect" of areas covered by high speed rail.

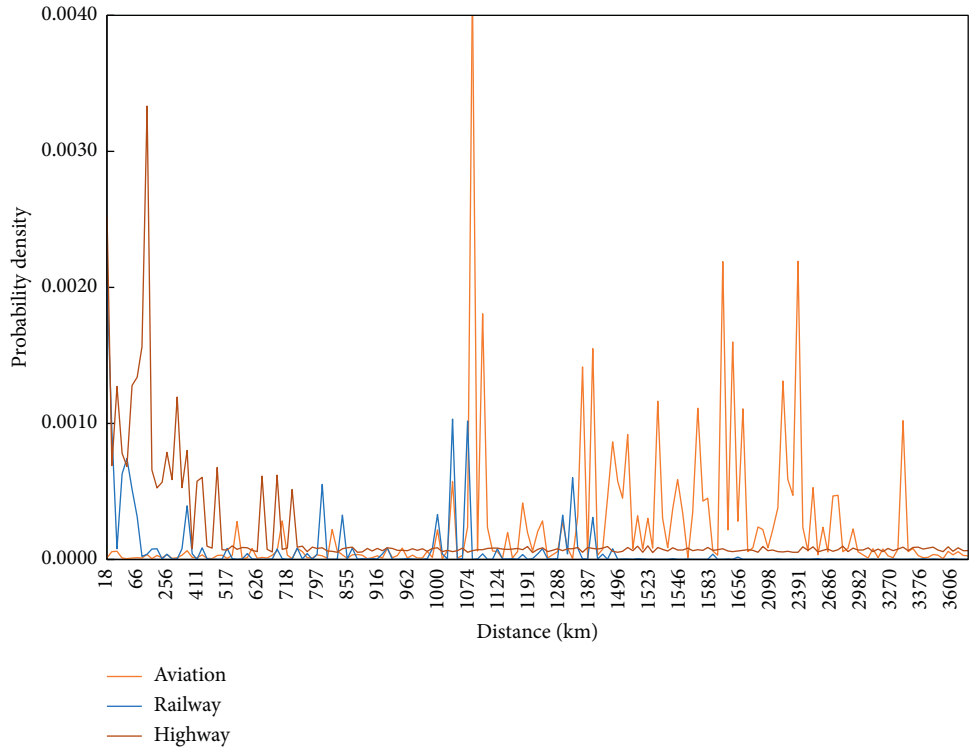


Figure 4: Probability distribution of the distance under different travel modes.

(3) From the perspective of highway flow (Figure 3(c)), the distributions of routes and people showed a pattern of outward divergence from the regional central cities. Because the highway is constrained by speed and facilities, the most connected cities are close to cities. When the distance was less than 350 km, the highway was the preferred mode of travel. Among them, the number of routes carrying 7.32–25.78 million people was the largest, reaching 1642, while there were only 32 routes with more than 1.812 million people. In comparison, the western route was longer, and the eastern route was shorter. Highway flows mainly reflected the spatial dependence and core-peripheral combination of travel on the inner levels of urban agglomerations. This clearly showed the development of each urban agglomeration. The Beijing-Tianjin-Hebei urban agglomeration, Yangtze River Delta, Pearl River Delta, Chengdu-Chongqing urban agglomeration, and the Central Plains urban agglomerations had hierarchical relationships. The hierarchical relationships of the urban agglomerations of Harbin-Changchun, Shandong Peninsula, Guanzhong Plain, Wuhan, and the Beibu Gulf had already appeared, while the remaining urban agglomerations were weak.

In order to compare the population distribution scale of different travel modes in each city, we drew a scatter plot. In the scatter plot, the points represent cities, the x-axis and y-axis represent highways and railways, respectively, and the color represents aviation (Figure 5). The results showed the following characteristics of resident travel: from the perspective of distribution scale, Shanghai, Beijing, Guangzhou, Chongqing, Shenzhen, and Chengdu were all in the forefront in the three travel modes. These six cities are national trip distribution centers. Under the three travel modes, the cities with a low value of the distribution scale accounted for the majority, while the number of cities in medium- and high-value areas was small. From the perspective of trip routes (Table 1), during the Spring Festival travel rush, the total number of passengers carried by aviation was 395.906 million people, and the top 10 routes carried 10% of these passengers. The cities connected by these routes were exactly at the four apexes of the "diamond"-shaped structure. Trains carried the largest number of passengers during the Spring Festival travel rush, up to 1151.46 million people. Among the top 10 routes, except Wuhan, which is the central point of the internal cross-skeleton crossing the "diamond" shape, the "diamond"-shaped structure was still in the center of the route, but the top 10 routes only accounted for 2.6% of the total. During the Spring Festival travel rush, 802.237 million people chose highways as their travel mode. The top 10 routes exactly represented the close connections between the core cities and sub-core cities in the three urban agglomerations of Beijing-Tianjin-Hebei (Beijing and Langfang), the Yangtze River Delta (Shanghai and Suzhou), and the Pearl River Delta (Shenzhen, Dongguan, Huizhou, Guangzhou, and Foshan). The two-way routes of Xi'an and

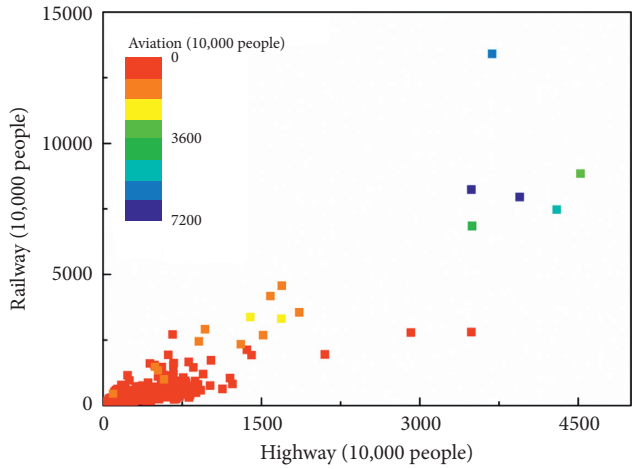


Figure 5: Population distribution scale under different travel modes.

Table 1: Top 10 population flow routes and populations under different travel modes.

Travel mode	Top 10 flow routes and populations (10,000 people)	Proportion
Aviation	From Shanghai to Chongqing (561.97), from Chongqing to Shanghai (561.79), from Chongqing to Beijing (531.66), from Beijing to Chongqing (495.89), from Shenzhen to Chengdu (391.39), from Chengdu to Shenzhen (353.64), from Shanghai to Beijing (335.05), from Beijing to Shanghai (330.75), from Shenzhen to Shanghai (326.07), and from Guangzhou to Shanghai (286.53).	A total of 41.7476 million people, accounting for 10.54% of the total carrying capacity.
Railway	From Shanghai to Chongqing (398.96), from Chongqing to Shanghai (384.79), from Foshan to Guangzhou (309.32), from Guangzhou to Foshan (307.67), from Chongqing to Beijing (296.40), from Changsha to Beijing (296.39), from Beijing to Chongqing (260.11), from Wuhan to Beijing (246.53), from Beijing to Shanghai (244.27), and from Chengdu to Nanjing (238.58).	A total of 29.8301 million people, accounting for 2.60% of the total carrying capacity.
Highway	From Shenzhen to Dongguan (377.45), from Dongguan to Shenzhen (374.35), from Shanghai to Suzhou (270.93), from Suzhou to Shanghai (263.99), from Foshan to Guangzhou (256.25), from Guangzhou to Foshan (250.17), from Xianyang to Xi'an (203.18), from Beijing to Langfang (202.05), from Xi'an to Xianyang (201.19), and from Shenzhen to Huizhou (188.14).	A total of 25.8772 million people, accounting for 3.23% of the total carrying capacity.

Xianyang of the Guanzhong urban agglomeration in the west entered the top 10, while the population flow between Chengdu and Chongqing of the largest urban agglomeration in the west did not enter the top 10 in China under the three travel modes. This deserves further attention.

Figure 6 shows the distribution scale and type of population flow under different travel modes (where the natural breakpoint classification method is adopted according to the railway classification standard). The distribution scale of a city is the sum of the inflow and outflow population, which represents the capacity of a city to receive tourists. From the distribution scale, it can be seen that there were four and six cities in the highest distribution level (>45.674 million people) in aviation and railway, respectively. There was no city in the first level in the highway mode. In the second level (17.187–45.674 million people), the numbers of cities in aviation, railway, and highway were 4, 16, and 10, respectively. In the third level (5.544–17.187 million people), there were 21 cities under the aviation mode. Most were distributed in the eastern region, and most of them were provincial capitals. There were 81 cities on the railway, mainly distributed along the railway route. There were 85 cities on the highway, concentrated in the eastern region and distributed in blocks. In contrast, the spatial differentiation of the intercity population based on aviation trips was the most obvious. It was characterized by high polarization and discrete point embedding, reflecting a core-periphery structure with national hub cities as the core distribution. The pattern of intercity population flow based on railway trips basically showed a core-periphery structure that used the cities along the national railway aorta as the core, and gradually decreased to the hinterland cities. The intercity population flow based on highway travel was a spatial pattern with strong local aggregation that matched the

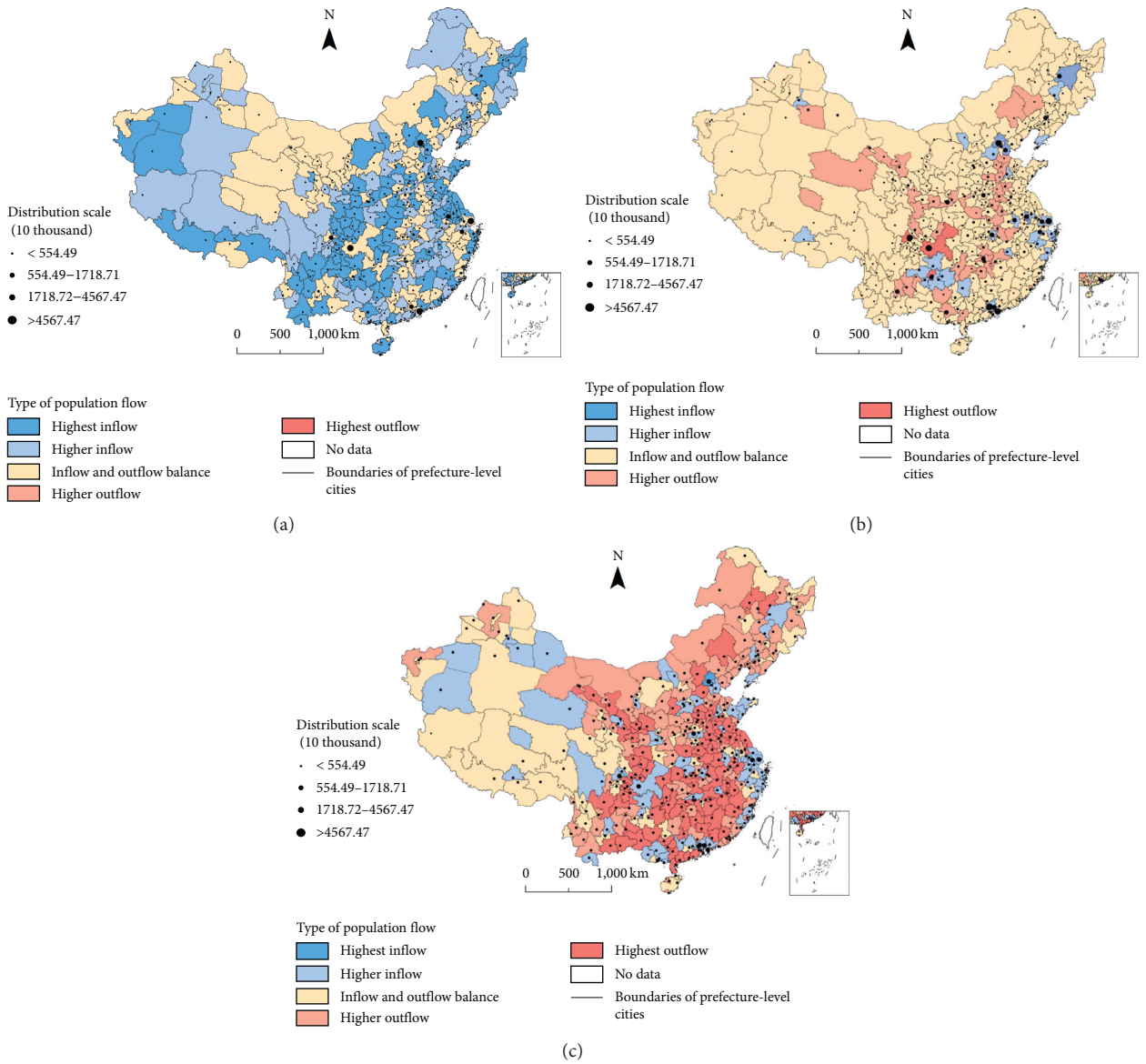


FIGURE 6: Distribution scale and type of population flow under different travel modes: (a) aviation; (b) rail; and (c) highway. Map spots of different colors show the different types of population flow, and the points represent the distribution scale. The larger the point, the stronger the distribution capacity of the cities.

population scale. It formed three strong aggregation regions in space: Beijing-Tianjin-Hebei-Yangtze River Delta-Central Plains, the Sichuan Basin, and the Pearl River Delta.

The type of population flow was obtained by subtracting the inflow population from the outflow population of the city. The gap between the outflow population and the inflow population was displayed, so as to judge the urban population flow direction. Different travel modes presented different types of urban inflow and outflow (Figure 6). (1) In terms of aviation travel (Figure 6(a)), there were 118 cities with the highest population inflow, among which cities with a net inflow of more than 300,000 include Nanjing, Tianjin, Jinan, Nanchang, Wuhan, Hong Kong, Changchun, and other major regional center cities, as well as Sanya, Bijie, Guilin, Lijiang, Nyingchi, and other tourist cities. The cities with the highest population outflow were Shenzhen, Chongqing, Chengdu, Guiyang, Guangzhou, Hangzhou, etc. Except for Xianyang, the cities with a net outflow population of 300,000 or more were all provincial capitals and above. On the whole, the number of cities with net population inflow was large, while the number of cities with net population outflow was relatively small. (2) In terms of railway travel (Figure 6(b)), the highest population inflow cities were only Beijing and Nanjing. The higher population outflow cities, higher population inflow cities, cities with equal inflow and outflow, and highest population outflow cities were cross-distributed. Among them, the highest population outflow cities were Chongqing, Guiyang, Kunming, Shenzhen, etc. The higher population outflow cities and the higher population inflow cities had the same number. Of the 16 cities

with a net outflow population of more than 600,000, nine cities were located in the western region. Of the 23 cities with a net inflow population of more than 600,000, only Urumqi and Qiannan Prefecture were western cities. (3) In terms of highway travel (Figure 6(c)), cities with equal inflow and outflow were mostly distributed in the northwest and northeast regions, and the number of outflow cities was more than that of inflow cities. The cities with the largest population outflow (over 400,000) were Zhoukou, Shaoyang, Shangrao, Ganzhou, Yichun, Yulin, Handan, Huaihua, and Zhaotong, all of which were major labor-export cities in central and western China. Chongqing had a higher population inflow under the highway travel mode. From February 1 to February 14, the number of people entering Chongqing every day was twice that at other times. On February 11 alone, the population of Chongqing reached 761,800 by highway, while on March 1, only 250,000 people entered Chongqing, and the number of inflow routes was shortened from 163 to 102. However, in terms of comprehensive aviation and railway, Chongqing's output of passengers during the Spring Festival transportation was far more than that by highway. Indeed, Chongqing is generally a highly populated outflow city. The scale of urban inflow and outflow was different under different travel modes. This was mainly determined by the city grade and location. The inflow and outflow of high-grade cities such as Beijing and Shanghai were not significantly different, while those of Qingyang and Wuwei were mostly larger than the inflow.

Figure 7 shows the intercity trip network structure under the maximum dominant flow. From the point of view of dominant flow routes, the maximum dominant flow routes of aviation, railway, and highways presented a spatial distribution pattern from long to short, and from complex to simple. The rhombus structure network framework supported by the cross was prominent in the dominant aviation flow (Figure 7(a)). It undertook the function of long-distance transportation, was little influenced by the constraints of geographical space, and had the network characteristics of a "hyperplane." Moreover, the maximum dominant flow of aviation was directly associated with first-tier cities in the different regions of China, and had the function of building the main network framework at a national and even global level, as the main form of framework construction of China's urban association pattern. The maximum dominant flow of railway mostly presented a typical pole axis spatial system pattern (Figure 7(b)). The railway mainly undertook the function of medium- and long-distance transportation. At the macro-level, it served as an important axis belt for a national development strategy. It directly associated and drove the element flow within large regions, and promoted the development of cities along the line to form an economic axis belt. On the other hand, it provided a support axis belt that connected central cities to the core framework. The maximum dominant flow of highways was a star-shaped divergent radiation pattern (Figure 7(c)), which effectively filled in for the whole traffic skeleton and the supporting axis belt. The highway mainly undertook the function of short-distance transportation. Due to the influence of regional divisions caused by the administrative forces of different provinces, highway trips were mostly limited to administrative units within provinces or nearby provinces and regions. The "localization" feature was significant, reflecting the internal connections of the relatively complete regional system. The urban network coverage areas covered by the three maximum dominant flows of aviation, railway, and highways overlapped, but each had its own emphasis. They respectively depicted the network characteristics of the national, regional, and provincial spatial scales, forming interregional interdependence, indispensable element association, and a spatial relationship.

From the number of cities associated with the maximum dominant flow route, we could see that Beijing and Shanghai were associated with 69 and 70 cities, respectively, under the maximum dominant flow of aviation (Figure 7(a)), occupying a dominant position. Shenzhen, Chengdu, and Chongqing were respectively associated with 36, 36, and 27 cities with the maximum dominant flow, which are in the second level. Guangzhou, Urumqi, Wuhan, Kunming, Harbin, Xi'an, and Changsha had more than five related cities. In the maximum dominant flow of the railway (Figure 7(b)), Beijing and Chengdu were all associated with 21 cities of the maximum dominant flow, which was far less than the number of cities associated with the first level of the aviation maximum dominant flow. There were 11 cities, such as Wuhan, Urumqi, Zhengzhou, and Guangzhou, with more than 10 related cities. Chengdu ranked first in the maximum dominant flow of highways (Figure 7(c)), which was related to 15 cities, followed by Zhengzhou, Guangzhou, Xi'an, and Changsha. Among the top 30 cities, except Foshan and Dongguan, were all provincial capitals and above. In the maximum dominant flow, the more dominant flow routes connected by the city, the higher the dependence of that city.

4.2. Structural Characteristics of the Urban Trip Network

4.2.1. Alter-Based Centrality and Alter-Based Power. The calculation results regarding the alter-based centrality (AC) and alter-based power (AP) (Table 2) showed the following. Beijing, Shanghai, Chongqing, Guangzhou, Shenzhen, and Chengdu ranked in the top 10 in all three travel modes. As far as the aviation mode was concerned, Chongqing had the strongest centrality. Shanghai had the strongest alter-based power, which proved that Shanghai had the largest dominance and influence in the entire aviation passenger transportation network. As far as the railway mode was concerned, Beijing had the greatest alter-based centrality and alter-based power. In the railway passenger network, Beijing had an absolute advantage as a national administrative center. It not only connected many routes and had a wide range but also had a strong influence on the routes it connected. As far as the highway mode was concerned, Shenzhen had the highest alter-based centrality. Its highway transportation network was very dense, and migrant workers from the south often traveled in Shenzhen. In addition, Shenzhen was directly connected to Shanghai, Chengdu, Beijing, and other cities through a number of highway routes. It can be said that the distribution capacity

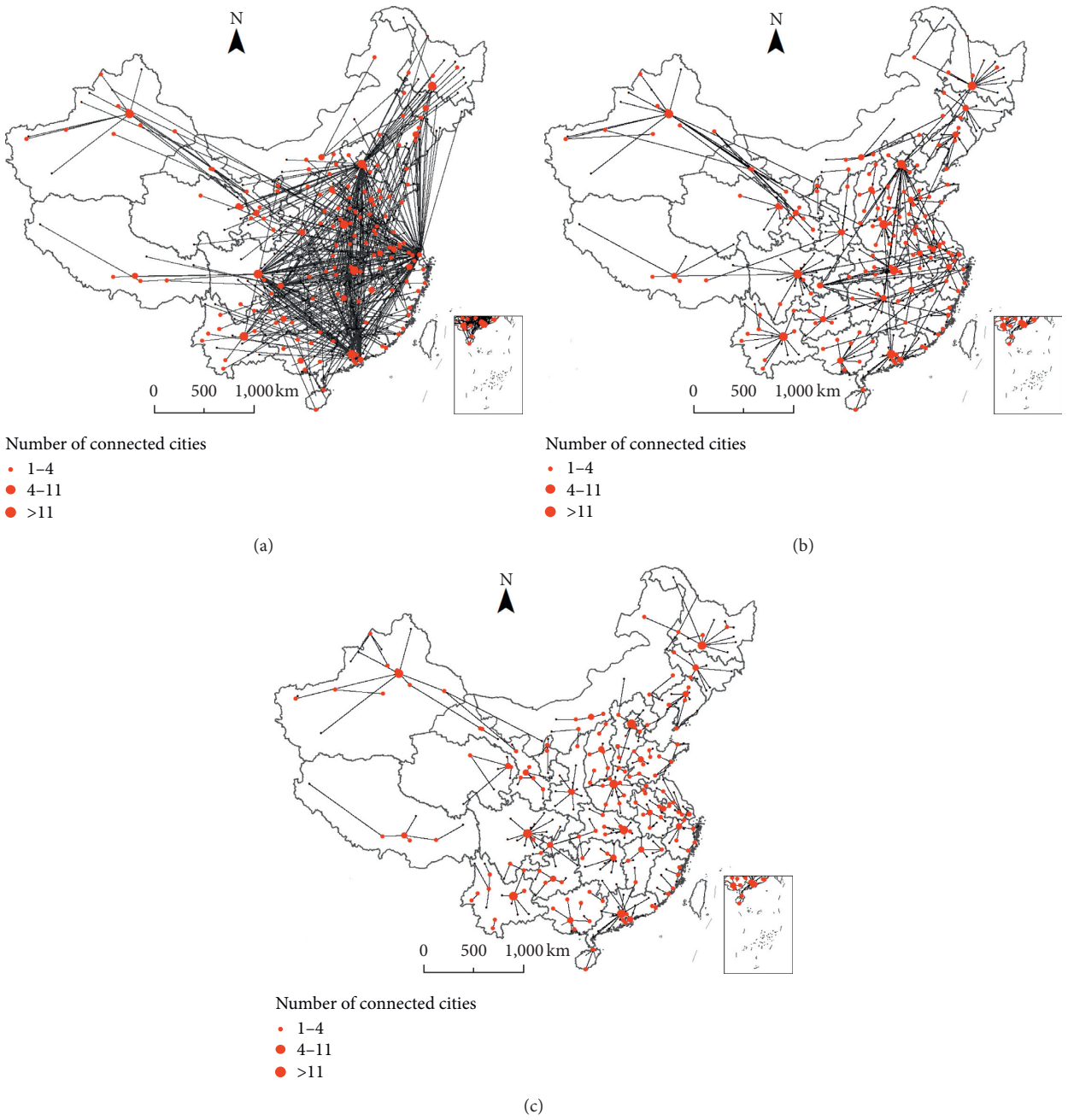


Figure 7: Characteristics of urban-trip association based on the maximum dominant flow under different travel modes: (a) aviation; (b) railway; (c) highway.

Table 2: Top 10 and bottom 10 cities of alter-based centrality (AC) and alter-based power (AP) under different trip modes.

Rank order	Aviation			Railway			Highway		
	Cities	AC	AP	Cities	AC	AP	Cities	AC	AP
1	Chongqing	1	0.808	Beijing	1	1	Shenzhen	1	0.932
2	Shanghai	0.938	1	Chongqing	0.941	0.481	Guangzhou	0.977	0.951
3	Beijing	0.656	0.778	Shanghai	0.801	0.559	Dongguan	0.96	0.644
4	Shenzhen	0.63	0.716	Guangzhou	0.72	0.643	Chongqing	0.812	0.891
5	Guangzhou	0.464	0.545	Shenzhen	0.645	0.564	Shanghai	0.774	0.733
6	Chengdu	0.443	0.741	Chengdu	0.633	0.564	Suzhou	0.717	0.68
7	Hangzhou	0.27	0.261	Wuhan	0.486	0.286	Beijing	0.697	0.856

Table 2: Continued.

Rank order	Cities	Aviation		Cities	Railway		Cities	Highway	
		AC	AP		AC	AP		AC	AP
8	Xi'an	0.241	0.118	Xi'an	0.441	0.273	Chengdu	0.597	1
9	Nanjing	0.203	0.163	Zhengzhou	0.329	0.261	Foshan	0.566	0.39
10	Xianyang	0.185	0.085	Hangzhou	0.324	0.209	Huizhou	0.419	0.155
...	...	...	...	...	...	...	...	...	...
337	Huaibei	0	0	Daxinganling	0.001	0.002	Guoluo	0.002	0.006
338	Chenzhou	0	0	Kezhou	0.001	0.004	Shannan	0.002	0.004
339	Yushuzhou	0	0	Nujiang	0.001	0.001	Yushuzhou	0.002	0.005
340	Yunfu	0	0	Hetian	0.001	0.002	Aletai	0.002	0.011
341	Shaoguan	0	0	Huangnanzhou	0.001	0.001	Qionghai	0.002	0.017
342	Hezhou	0	0	Shennongjia	0.001	0.001	Kezhou	0.002	0.013
343	Bozhou	0	0	Changdu	0.001	0	Daxinganling	0.001	0.002
344	Shennongjia	0	0	Aletai	0	0.001	Naqu	0.001	0.003
345	Guoluo	0	0	Guoluo	0	0.001	Rikaze	0.001	0.005
346	Wuzhou	0	0	Ali	0	0	Ali	0	0.002

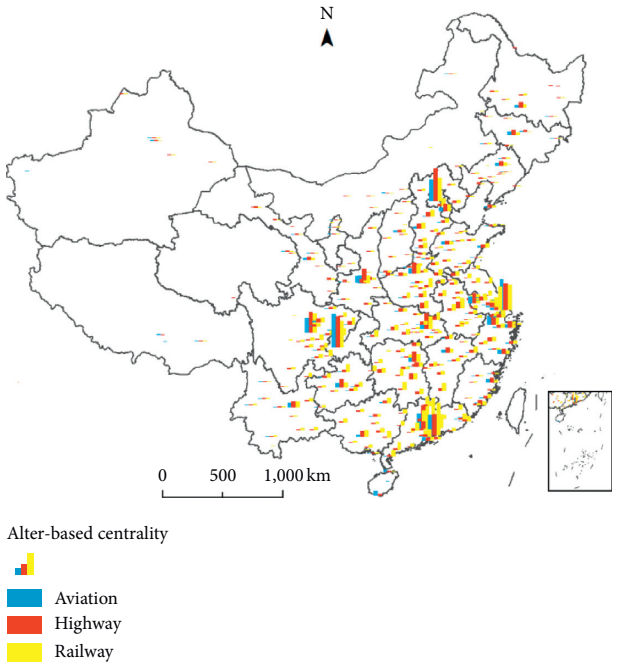


Figure 8: Classification of alter-based centrality under different travel modes.

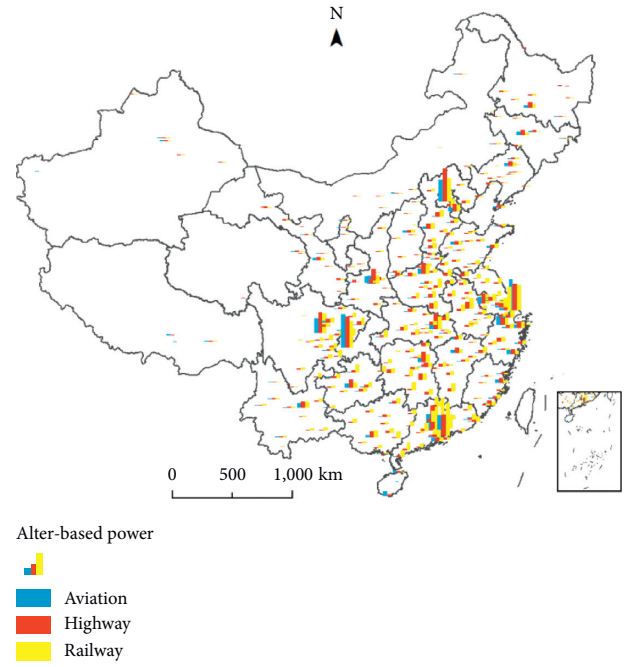


Figure 9: Classification of alter-based power under different travel modes.

was very strong. In terms of alter-based power, Chengdu occupied first place. In the road network, Chengdu was in the middle of the network. The city served as an important carrier of communication between the east and the west, and occupied an important position in the highway network.

We used ArcGIS software to visualize the data (Figure 8). Under different travel modes, city hierarchies are presented by AC values as a pyramid. That is, the AC value was high and the number of cities was small. In aviation mode, the hierarchical structure was the most clear: the number of high-level cities was the smallest and they were scattered, e.g., Shanghai, Chongqing, and Shenzhen. As a high end way of traveling, aviation has a high demand on the site cost. Only cities with strong collection and distribution capacities could support them, and the number was small. In highway mode, the number of high-level cities increased significantly. The highway service distance was short, the number of people traveling on highways was small, and most were concentrated in the surrounding cities. Therefore, the embodiment of the distribution capacity was mostly connected with the surrounding cities, and the gap between the cities was small. In railway mode, the number of high-level cities in terms of their AC values fell between those of highways and aviation. Railways had a wide range of services, they connected many cities, and the hierarchical structure of the AC was between aviation and highways.

Figure 9 shows the AP map under different travel modes of the Spring Festival travel. In aviation, six cities (viz., Shanghai, Chongqing, Beijing, Chengdu, Shenzhen, and Guangzhou) were at the highest level. They had strong

control over the connected routes and surrounding cities. On railways, the lowest-level cities were in Tibet, Qinghai, Inner Mongolia, and Hainan. Other provinces appear with high AP-value cities. On highways, Suzhou, Zhengzhou, Foshan, Xi'an, Dongguan, Kunming, Nanjing, Nanning, and Hangzhou were added to the highest level (0.287–1). As a whole, the distribution of cities with high alter-based power was relatively balanced. Unlike cities with a high alter-based centrality, Yinchuan, Lanzhou, Urumqi, Xining, and other cities were the hub nodes for population flow from west to east, and the AP value was significantly higher than the AC value. These results confirmed that the intermediary and distribution functions of some cities in the process of population flow could be better reflected by alter-based power.

When analyzing urban networks, the alter-based centrality and alter-based power jointly determined the status of a city. Generally speaking, there was a positive correlation between the city's AC and AP values. That is to say, a city with a high AC value had a high AP value, and a city with a strong ability to gather and spread resources had a stronger ability to control those resources. Of course, there were differences in some cities. Neal, by measuring the world urban information network, divided quintessential cities with high centrality and high power, hub cities with high centrality and low power, and gateway cities with low centrality and high power [22]. Thus, we could better identify the status and attributes of cities [39]. The natural breakpoint classification method was used to classify the alter-based centrality and alter-based power of the aviation network (Table 3), and the city type was divided based on the matching relationship between cities at different levels. The ratio of the number of cities at the first level, second level, third level, and other levels was 5 : 8:24 : 309, and the ratio of the number of quintessential, hub, gateway, and edge cities was 23 : 7:7 : 309.

Using the same method to divide the urban hierarchy structure of the railway network, the ratio of the number of cities at the first level, second level, third level, and other levels was 6 : 16 : 63 : 261. The ratio of the number of quintessential, hub, gateway, and edge cities was 33 : 37 : 15 : 261. Compared with aviation travel, the number of other-level edge cities decreased significantly, and the number of quintessential cities increased.

From the perspective of gateway cities, most second-level cities were provincial capital cities with relatively good geographic locations. There were only eight gateway cities at the third level. Based on the statistics of the urban hierarchical structure of the highway trip network, we found that the ratio of the number of cities at the first level, second level, third level, and other levels was 8 : 49 : 118 : 171. The ratio of the number of quintessential, hub, gateway, and edge cities was 73 : 80 : 22 : 171. The number of quintessential cities and hub cities increased significantly. The number of third-level quintessential cities was the largest, and the number of gateway cities was less. In view of the three types of trip networks, the same city has different hierarchical structures in different trip networks. For example, Tianjin is classified as a quintessential city at the third level in the aviation network; in the railway network, it is a gateway city at the second level; and in the highway network, it is a hub city at the second level. As a whole, the number of gateway cities changed less. With the trip mode from aviation to railway to highway, the number of edge cities decreased and the number of third-level cities increased.

4.2.2. Cluster Structure of Urban Trip Network. The network cluster structure refers to a complex system in which clusters are interconnected with other clusters according to certain rules. A network with a clustered structure features a high average clustering coefficient. That is, the higher the average clustering coefficient, the more obvious the clustered structure [3]. In this study, we used the fast-unfolding clustering algorithm to divide the city types under different travel modes during the China Spring Festival rush, so as to obtain small groups with close internal connections and relatively few regional connections. The principle of fast-unfolding clustering is to repeatedly divide through iterative operations, so that the overall modularity of the divided network continues to increase until the network structure no longer changes [49]. Cluster analysis in Gephi software needs to ensure that the value of the modularity reaches the maximum. The greater the modularity value, the better the cluster structure division effect [50]. The statistical table for the urban cluster structure under different travel modes during the Spring Festival rush was obtained through calculation (Table 4). The number of clusters divided under different travel modes was different. The number of urban clusters was the least in aviation mode, at only seven. The difference in the number of cities included in each cluster structure was also the largest. For example, cluster 0 contained the largest number of cities, reaching 113, while cluster 4 and cluster 6 contained only two cities. There were eight clusters divided by railway mode, and the gap in the number of cities included in each cluster also narrowed. The number of clusters was the most under highway mode, with 10 clusters, and the number of cities in each cluster was the most balanced.

A spatial visualization of the cluster structure under different travel modes is shown in Figure 10. The cluster structure of different travel modes was obviously different, among which the average clustering coefficient value was represented as aviation < railway < highway. The cluster structure was the most significant in highway mode, and the aviation cluster structure was relatively weak. Under aviation, it was characterized by fewer categories, discontinuous clusters, a large spatial span, fewer geographical constraints, and a jump distribution. The cluster structure under railway mode had an aggregation state and high spatial connectivity, but there were also discontinuous jump clusters, such as cluster 2 (Jiangsu, Anhui, Zhejiang, Jiangxi, Chongqing, and Guizhou). The cluster structure under highway mode appeared as a block distribution. This was similar to the cluster structure of the railway in general, although the coincidence degree between the cluster boundary and the provincial boundary was higher. In conclusion, due to the limitation of geographical spatial effect, adjacent cities are more likely to form a cluster. Some clusters from the

Table 3: City hierarchy in the aviation network.

Level	Types	Alter-based centrality	Alter-based power	Cities included (and cities number)
Nationwide level	Quintessential world cities	0.464–1.000	0.261–1.000	Chongqing, Shanghai, Beijing, Shenzhen, Guangzhou (5)
Regional level	Quintessential world cities	0.203–0.463	0.118–0.260	Chengdu, Hangzhou, Nanjing (3)
	Hub world cities	0.203–0.463	0–0.117	Xi'an (1)
	Gateway world cities	0–0.202	0.118–0.260	Wuhan, Zhengzhou, Kunming, Guiyang (4)
Local level	Quintessential world cities	0.056–0.202	0.034–0.117	Xianyang, Tianjin, Changsha, Wenzhou, Jinan, Lanzhou, etc. (15)
	Hub world cities	0.056–0.202	0–0.033	Changchun, Hongkong, Zhuhai, Hefei, Shijiazhuang, Taiyuan (6)
	Gateway world cities	0–0.055	0.034–0.117	Urumqi, Qingdao, Hohhot (3)
Other level	Edge cities	0–0.055	0–0.033	Dalian, Foshan, Yinchuan, Guilin, Yangzhou, Ili, etc. (309)

Table 4: Statistical table of urban cluster structure and number of cities under different trip modes.

Aviation		Railway		Highway	
Cluster type	Number of cities	Cluster type	Number of cities	Cluster type	Number of cities
0	113	0	77	0	56
1	44	1	84	1	62
2	98	2	63	2	62
3	72	3	18	3	28
4	2	4	8	4	29
5	15	5	11	5	16
6	2	6	53	6	18
		7	32	7	27
				8	31
				9	17

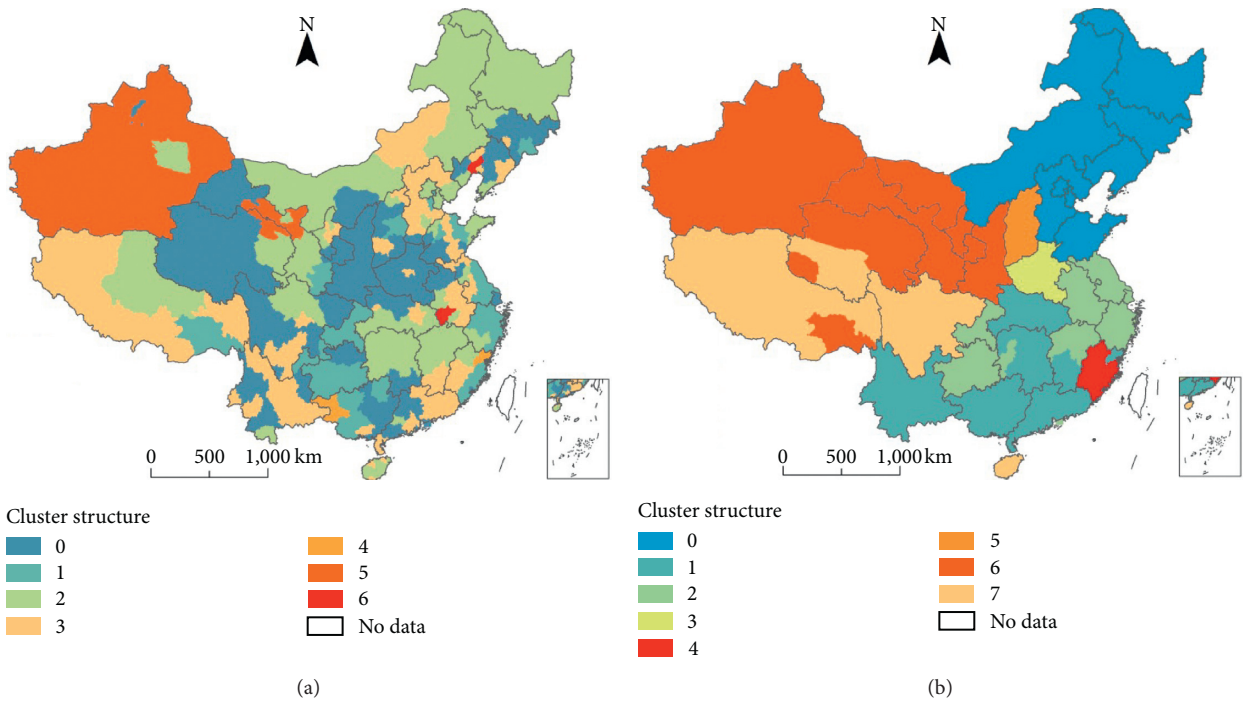
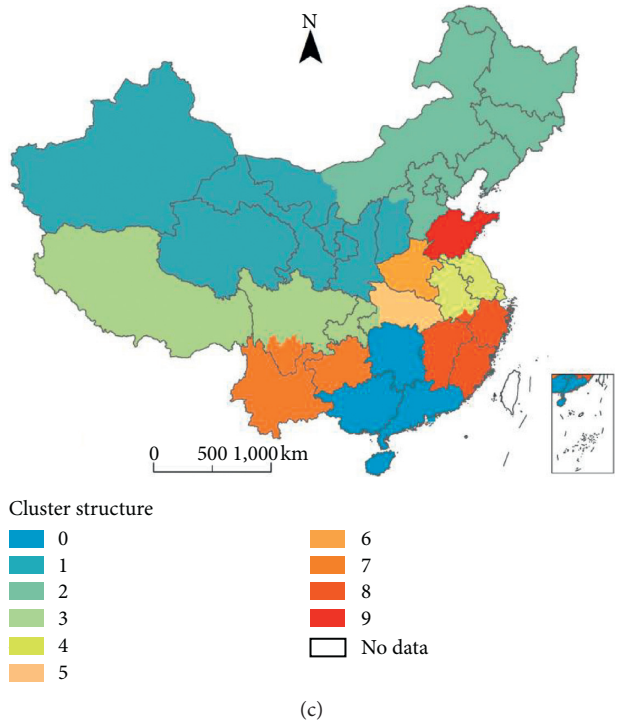


Figure 10: Continued.



(c)

Figure 10: City network cluster structure in China under different travel modes: (a) aviation; (b) railway; and (c) highway.

highway network and aviation network were identical to the railway network, so they showed a high degree of overlap (red), in which the number of overlapping cities reached 16. Eighty-five cities belong to the same cluster under any two traffic modes, shown in blue, and the remaining cities belong to different clusters under the three traffic modes (Figure 11). Although similarities could be observed in the networks representing three fundamentally different types of interactions, the modular structure was inconsistent, which indicated that there were strong differences in people's choice of travel mode [51].

5. Discussion

In the process of calculating the alter-based centrality and alter-based power, we found that when the degree centrality was the same, with the increase of the weighted degree centrality, the status of city also improved [20]. For example, Beijing, Shanghai, Guangzhou, and other cities were all associated with high-intensity population flow, and their alter-based centrality and alter-based power and ranks were pushed up. The network status of Anqing, Qingdao, Nanchang, and other regional central cities improved due to their high correlation and uniform flow distribution. When the dependency parameter was added to equation (3), the centrality with path dependence in a city tended to converge [20]. For example, in the highway trip network, 71.9% of the population flow in Xianyang was associated with Xi'an, which caused its ranking to decrease by 60 relative to the weighted degree centrality. The alter-based centrality and alter-based power also showed a certain degree of reverse nature. Under the same association weight, cities with more association opportunities were more likely to be connected, but cities with path dependence were easier to control. For example, Lanzhou associates mostly with Tianshui, Xining, Zhongwei, and some cities with low centrality. These cities lack communication opportunities and depended highly on Lanzhou. Thus, the status of Lanzhou's western gateway was promoted. Gateway cities were often the hubs of regional network resource allocation, and have high power over neighboring cities, to some extent. This compressed the opportunity and possibility of residents traveling in the region, showing a monopoly of regional resource circulation. This proved that a large number of small-scale cities (network resources can only be exchanged through gateway cities) had the disadvantages of lacking a path and path dependence.

Chinese people have a strong and unique homeland complex. In this sense, because of the Spring Festival in China, there will be a Spring Festival travel rush in China. The main driving factor of population migration during the Spring Festival travel rush is the imbalance of regional economic development. After China's reform and opening up, the eastern region was given priority of development [32, 52]. Driven by economic benefits, a large amount of rural surplus labor began to move to the east and other economically developed areas. In the study, we found that the population flow routes were sparse in the west and dense in the east, and the high population flow routes concentrated on the east side of the "Huhuanyong Line," which is consistent with the conclusions of Wei et al. [46]. The economic development of eastern China is relatively fast, with frequent exchanges between cities, and priority given to transportation infrastructure. The economic development of

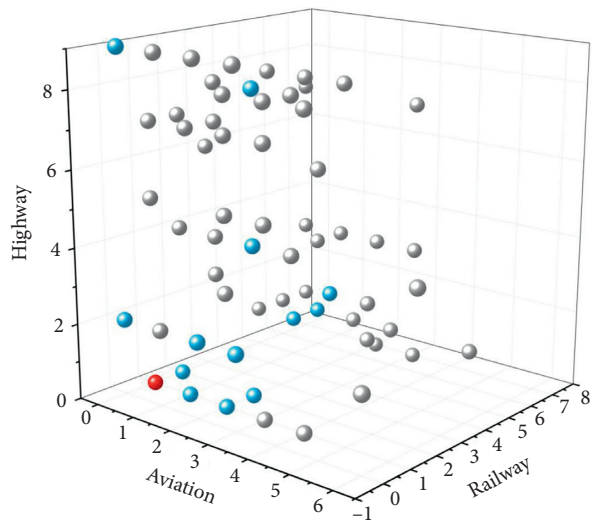


Figure 11: Similarity of cluster structure in the travel modes of aviation, railway, and highway. The cell colors indicate the type of overlap between the three cluster structures. Gray, blue, and red, respectively, indicate that the three cluster structures completely overlap, that two overlap, and that none overlap.

northwestern China is backward and construction is difficult. In view of this, the government should support and improve northwestern transportation routes, improve the convenience of travel, and reasonably plan and arrange the routes. At the same time, it is also necessary to adapt to local conditions and promote economic construction in the west to reduce population outflow. In the analysis of alter-based centrality and alter-based power, we know that although Chengdu and Chongqing are in the western region, they are ranked highly. Chongqing, in particular, is a province with a large population output, which is consistent with the research results of Lai and Pan [3], This phenomenon makes the development of the urban network in this region appear "faulty." The development of surrounding cities is limited, and resources and development opportunities are concentrated in Chengdu and Chongqing because of the lack of a pivot point. In order to alleviate this phenomenon, it is necessary to spread the functions of the cities (Chengdu and Chongqing) to the surrounding region [46]. Population migration brings not only economic development and regional exchanges but also disease transmission, employment, and traffic pressure. Therefore, the government should regularly detect population flows, predict their trends in the next period, and make plans for the occurrence and prevention of emergencies in advance [1]. In addition, population flow can reflect the development of urban agglomerations. That is, the closer the connections between cities, the better the development of urban agglomerations. In future work, we can reduce the attention on individual cities, and instead consider using population flow data to measure the development of urban agglomerations and the internal and external relations of urban agglomerations.

This study compared and analyzed the structural characteristics of intercity networks under multiple travel modes. We improved the unilateral and attribute-oriented defects caused by existing research relying on a single type of travel mode. Tencent Migration data broke through the lag effect brought by traditional statistical data and provided a new data source for large-scale research on intercity travel by the Chinese. Compared with other data, Tencent had more users and higher accuracy. However, due to the generation and acquisition of the data and the protection of personal privacy, it was impossible to obtain social attributes such as the occupation, sex, age, and travel purposes of travelers. It was impossible to further explore the willingness and group effect of the population. In addition, some travel paths may be disassembled, and the origin and destination cannot be studied as network nodes. The status of a transit city was overemphasized, the trip characteristics of users cannot be identified completely, and the spatial characteristics of multiple trips by residents cannot be identified. This will inevitably cause errors in the research results. As a representative period of holidays, the Spring Festival is special. However, we cannot ignore the impact of population flow in other periods on Chinese economic development. Pan and Lai [1] used the same data to study the time series characteristics and scale of population flow on National Day and the subsequent Mid-Autumn Festival. It was concluded that the performance of population flow in this period is quite different from that in the Spring Festival travel rush. In future work, it will be necessary to comprehensively analyze the characteristics of Chinese population flow in combination with multiple time periods, and to clarify the differences between other time periods and the Spring Festival travel rush. In addition, on the basis of continuous time series data, we will excavate the evolution process and formation mechanism of the intercity trip network of Chinese residents. In this study, we discussed the structure of the population trip network based on the macro-characteristics of the floating population under different travel modes, and the spatial pattern of the urban network structure. However, there was no fusion comparison of the multisource data. With the rapid development of geographic big data, it is necessary to use intelligent technology to compare the differences in population flow under multisource data, and to explore more useful knowledge. This study provided a new perspective for the perception of human social activities, that is, big data, and points out a new research idea for global scientific researchers to study human social activity.

6. Conclusions

In this study, we used data from population migration trajectories between cities provided by the Tencent Migration big data platform. Spatial analysis was used to explore the characteristics of an intercity trip network of the Chinese under different travel modes during the Spring Festival travel rush. We determined the hierarchy of this urban network and measured the agglomerated spatial patterns under different travel modes with the help of two indicators: the alter-based centrality and alter-based power. The main conclusions were as follows:

(1) Among the three travel modes, the fewest routes were connected by aviation (8451), followed by highways (10,222) and railway (13,746). When a trip involved traveling less than 350 km, travelers tended to use highways. When trips covered distances less than 1,500 km, railways were the most popular mode of travel, and when the distance was 1,000–1,500 km, aviation was the most popular. In terms of the number of passengers, railways transport the most people, followed by highways and aviation. Among the maximum dominant flow, Beijing and Shanghai played the most important role in controlling domestic aviation trip connections, followed by Chengdu and Chongqing. For railway trips, Beijing and Chengdu occupied the absolute advantage, whereas highways were mostly used for connections between provincial administrative centers and the surrounding cities.

(2) According to the indicators of alter-based centrality and alter-based power, cities could be divided into four types: quintessential cities, hub cities, gateway cities, and edge cities. In the aviation city network, the ratio of the number of cities among these four types was 23 : 7 : 7 : 309, with the largest number of other-level edge cities. In the railway network, the ratio of the number of cities among these four types was 33 : 37 : 15 : 261. The number of other-level edge cities decreased significantly, and the number of quintessential cities increased significantly. The number of quintessential cities at the local level was the largest under the highway mode of travel. The number of hub cities was 80, with a similar number of gateway cities.

(3) During the Spring Festival travel rush, the number of urban clusters under different travel modes was different. Aviation, railway, and highway trips are clustered into 7, 8, and 10 urban clusters in sequence. In the space display, there were fewer categories, discontinuous clusters, fewer geographical constraints, and jump distributions in the aviation travel mode. The cluster structure in the railway travel mode had an aggregation state and high spatial connectivity. The cluster structure under the highway travel mode appeared as a block distribution, and the coincidence degree between the cluster boundary and the provincial boundary was higher.

(4) There were differences in the characteristics of the urban network from the perspective of the different modes of travel. Intercity population flow based on aviation trips was highly polarized with discrete point embedding, reflecting a core-periphery structure with nationwide-level hub cities as the core distribution. The intercity population flow based on railway trips had a core-periphery structure that took cities along the national railway artery as the core and gradually decreased to the hinterland cities. The intercity population flow based on highway trips had a spatial pattern of strong local aggregation that matched with the population scale.

Data Availability

Tencent's population migration data are obtained from the location big data released by Tencent Company, which comes from https://heat.qq.com/qianxi.php.

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 (Grant no. 41661025).

References

[1] 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.

[2] J. Xu, A. Li, D. Li et al., "Difference of urban development in China from the perspective of passenger transport around Spring Festival," *Applied Geography*, vol. 87, pp. 85–96, 2017.

[3] 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. 4, Article ID 04020018, 2020.

[4] W. Zhang, Z. Chong, X. Li, and G. Nie, "Spatial patterns and determinant factors of population flow networks in China: analysis on Tencent location big data," *Cities*, vol. 99, Article ID 102640, 2020.

[5] D. N. Garrett, R. Alasdair, and J. L. Rosenbloom, "An economic geography of the United States: from commutes to megaregions," *PLoS One*, vol. 11, no. 11, Article ID e0166083, 2016.

[6] Y. Wei, W. Song, C. Xiu, and Z. Zhao, "The rich-club phenomenon of China's population flow network during the country's spring festival," *Applied Geography*, vol. 96, pp. 77–85, 2018.

[7] Y. Long, "Redefining Chinese city system with emerging new data," *Applied Geography*, vol. 75, pp. 36–48, 2016.

[8] J. Shen, "Explaining interregional migration changes in China, 1985–2000, using a decomposition approach," *Regional Studies*, vol. 49, no. 7, pp. 1176–1192, 2013.

[9] J. Qi, Z. Wang, Y. Wang, and D. Li, "Visualization and analysis on the spatial-temporal patterns of flow direction of interprovincial migration in China based on origin-destination matrix," *Procedia Environmental Sciences*, vol. 26, pp. 115–118, 2015.

[10] S. Dong, Y. Pu, and Y. Wang, "A research on complex network of Chinese interprovincial migration based on the fifth population census," in *Proceedings of the 2013 21st International Conference on Geoinformatics*, pp. 1–4, Kaifeng, China, June 2013.

[11] Y. Wang, L. Dong, Y. Liu, Z. Huang, and Y. Liu, "Migration patterns in China extracted from mobile positioning data," *Habitat International*, vol. 86, pp. 71–80, 2019.

[12] Y. Wang, F. Wang, Y. Zhang, and Y. Liu, "Delineating urbanization "source-sink" regions in China: evidence from mobile app data," *Cities*, vol. 86, pp. 167–177, 2019.

[13] T. Lv, X. F. Piao, W. Y. Xie, and S. B. Huang, "Study of the attack-resistance of national economy based on data mining analysis of the population flow social network," in *Proceedings of the 2011 International Conference on E-Business and E-Government (ICEE)*, Shanghai, China, May 2011.

[14] X. Liu, Z. Neal, and B. Derudder, "Featured graphic. City networks in the United States: a comparison of four models," *Environment and Planning A: Economy and Space*, vol. 44, no. 2, pp. 255-256, 2012.

[15] X. Li, S. Huang, J. Chen, and Q. Chen, "Analysis of the driving factors of US domestic population mobility," *Physica A: Statistical Mechanics and its Applications*, vol. 539, Article ID 122984, 2020.

[16] Z. Xu and R. Harriss, "Exploring the structure of the U.S. intercity passenger air transportation network: a weighted complex network approach," *GeoJournal*, vol. 73, no. 2, pp. 87–102, 2008.

[17] F. Zhen, Y. Cao, X. Qin, and B. Wang, "Delineation of an urban agglomeration boundary based on Sina Weibo microblog "check-in" data: a case study of the Yangtze River Delta," *Cities*, vol. 60, pp. 180–191, 2017.

[18] H. Roberts, J. Sadler, and L. Chapman, "Using Twitter to investigate seasonal variation in physical activity in urban green space," *Geo: Geography and Environment*, vol. 4, no. 2, Article ID e00041, 2017.

[19] J. Li, H. V. Zuylen, C. H. Li, and S. F. Lu, "Monitoring travel times in an urban network using video, GPS and Bluetooth," *Procedia Social and Behavioral Sciences*, vol. 20, pp. 630–637, 2011.

[20] Z. Y. Zhao, Y. Wei, S. J. Wang, and R. Q. Pang, "Measurement of directed alternative centricity and power of directed weighted urban network: a case of population flow network of China during "Chunyun" Period," *Geographical Research*, vol. 36, no. 4, pp. 647–660, 2017, in Chinese.

[21] H. L. Boschken, "A multiple-perspectives construct of the American global city," *Urban Studies*, vol. 45, no. 1, pp. 3–28, 2008.

[22] A. S. Alderson and J. Beckfield, "Power and position in the world city system," *American Journal of Sociology*, vol. 109, no. 4, pp. 811–851, 2004.

[23] B. Derudder, M. Timberlake, and F. Witlox, "Introduction: mapping changes in urban systems," *Urban Studies*, vol. 47, no. 9, pp. 1835–1841, 2010.

[24] K. S. Cook, R. M. Emerson, M. R. Gillmore, and T. Yamagishi, "The distribution of power in exchange networks: theory and experimental results," *American Journal of Sociology*, vol. 89, no. 2, pp. 275–305, 1983.

[25] Z. Neal, "Differentiating centrality and power in the world city network," *Urban Studies*, vol. 48, no. 13, pp. 2733–2748, 2011.

[26] Z. Neal, "Does world city network research need eigenvectors?" *Urban Studies*, vol. 50, no. 8, pp. 1648–1659, 2013.

[27] H. Yang, F. Dobruszkes, J. Wang, M. Dijst, and P. Witte, "Comparing China's urban systems in high-speed railway and airline networks," *Journal of Transport Geography*, vol. 68, pp. 233–244, 2018.

[28] B. J. L. Berry, "Cities as systems within systems of cities," *Papers of the Regional Science Association*, vol. 13, no. 1, pp. 146–163, 1964.

[29] P.-A. Laharotte, R. Billot, E. Come, L. Oukhellou, A. Nantes, and N.-E. El Faouzi, "Spatiotemporal analysis of bluetooth data: application to a large urban network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1439–1448, 2015.

[30] B. Derudder and Z. Neal, "Uncovering links between urban studies and network science," *Networks and Spatial Economics*, vol. 18, no. 3, pp. 441–446, 2019.

[31] Z. P. Neal, "From central places to network bases: a transition in the US Urban hierarchy, 1900–2000," *City & Community*, vol. 10, no. 1, pp. 49–75, 2010.

[32] J. Li, Q. Ye, X. Deng, Y. Liu, and Y. Liu, "Spatial-temporal analysis on spring festival travel rush in China based on multisource big data," *Sustainability*, vol. 8, no. 11, p. 1184, 2016.

[33] M. Castells, "Globalisation, networking, urbanisation: reflections on the spatial dynamics of the information age," *Urban Studies*, vol. 47, no. 13, pp. 2737–2745, 2010.

[34] W. Cheng, C. L. Xiu, W. Q. Ke, Z. Y. Yu, and Y. Wei, "Hierarchical structures of China's city network from the perspective of multiple traffic flows," *Geographical Research*, vol. 34, no. 11, pp. 2073–2083, 2015, in Chinese.

[35] Y. Lu and Y. Liu, "Pervasive location acquisition technologies: opportunities and challenges for geospatial studies," *Computers, Environment and Urban Systems*, vol. 36, no. 2, pp. 105–108, 2012.

[36] H. Yang, G. Burghouwt, J. Wang, T. Boonekamp, and M. Dijst, "The implications of high-speed railways on air passenger flows in China," *Applied Geography*, vol. 97, pp. 1–9, 2018.

[37] S.-L. Shaw and H. Yu, "A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical-virtual space," *Journal of Transport Geography*, vol. 17, no. 2, pp. 141–149, 2009.

[38] Y. Z. Liu, W. T. Zhang, X. M. Cui, G. D. Zhang, and G. X. Wang, "City pipe network intelligent service based on GIS and Internet of Things," in *Proceedings of the 2014 7th International Conference on Intelligent Computation Technology and Automation*, Changsha, China, October 2014.

[39] J. Wang and F. Jin, "China's air passenger transport: an analysis of recent trends," *Eurasian Geography and Economics*, vol. 48, no. 4, pp. 469–480, 2007.

[40] J. Li, X. Wen, M. Wu, F. Liu, and S. Li, "Identification of key nodes and vital edges in aviation network based on minimum connected dominating set," *Physica A: Statistical Mechanics and Its Applications*, vol. 541, Article ID 123340, 2020.

[41] W. Cao, X. Feng, and H. Zhang, "The structural and spatial properties of the high-speed railway network in China: a complex network perspective," *Journal of Rail Transport Planning & Management*, vol. 9, pp. 46–56, 2019.

[42] J. Wang, D. Du, and J. Huang, "Inter-city connections in China: high-speed train vs. inter-city coach," *Journal of Transport Geography*, vol. 82, Article ID 102619, 2020.

[43] N. Limtanakool, M. Dijst, and T. Schwanen, "A theoretical framework and methodology for characterising national urban systems on the basis of flows of people: empirical evidence for France and Germany," *Urban Studies*, vol. 44, no. 11, pp. 2123–2145, 2007.

[44] A. De Montis, S. Caschili, and A. Chessa, "Time evolution of complex networks: commuting systems in insular Italy," *Journal of Geographical Systems*, vol. 13, no. 1, pp. 49–65, 2010.

[45] Z. Neal, "The devil is in the details: differences in air traffic networks by scale, species, and season," *Social Networks*, vol. 38, pp. 63–73, 2014.

[46] Y. Wei, C. L. Xiu, Z. M. Liu, and W. Chen, ""Spatial pattern of city network in transitional China based on the population flows in "Chunyun" period," *Scientia Geographica Sinica*, vol. 36, no. 11, pp. 1654–1660, 2016, in Chinese.
[47] J. D. Nystuen and M. F. Dacey, "A graph theory interpretation of nodal regions," *Papers in Regional Science*, vol. 7, no. 1, pp. 29–42, 2005.
[48] J. O. Wheeler and R. L. Mitchelson, "Information flows among major metropolitan areas in the United States," *Annals of the Association of American Geographers*, vol. 79, no. 4, pp. 523–543, 1989.
[49] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, Article ID P10008, 2008.
[50] S. A. Tabrizi, A. Shakery, M. Asadpour, M. Abbasi, and M. A. Tavallaie, "Personalized PageRank clustering: a graph clustering algorithm based on random walks," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 22, pp. 5772–5785, 2013.
[51] L. Hedayatifar, R. A. Rigg, Y. Bar-Yam, and A. J. Morales, "US social fragmentation at multiple scales," *Journal of The Royal Society Interface*, vol. 16, no. 159, Article ID 20190509, 2019.
[52] J. Shen, "Increasing internal migration in China from 1985 to 2005: institutional versus economic drivers," *Habitat International*, vol. 39, pp. 1–7, 2013.