Article

Identifying Traffic Congestion Patterns of Urban Road Network Based on Traffic Performance Index

Jinrui Zang 1, Pengpeng Jiao 1,*, Sining Liu 2, Xi Zhang 3, Guohua Song 4 and Lei Yu 5

1 Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture, Beijing 100044, China
2 Beijingxi Railway Station, China Railway Being Group Co., Ltd., Beijing 100844, China
3 Beijing Key Laboratory of Urban Traffic Operation Simulation and Decision Support, Beijing Transport Institute, Beijing 100073, China
4 Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China
5 College of Science and Technology, Texas Southern University, Houston, TX 77004, USA
* Correspondence: jiaopengpeng@bucea.edu.cn; Tel.: +86-186-1192-0632

Abstract: Urban congestion has become a global problem with urbanization and motorization. The analysis of time-varying traffic congestion patterns is necessary to formulate effective management strategies. The existing studies have focused on traffic flow patterns developed by the volume, speed and density of road sections in a limited district, while the long-time analysis of congestion patterns of the macro road network at the city level is inadequate. This paper aims to recognize traffic congestion patterns of the urban road network based on the traffic performance index (TPI) of 699 days in 2018, 2019 and 2021 in Beijing. The self-organizing maps (SOM) method improved by an automatic clustering number determination algorithm is proposed to cluster congestion patterns based on time-varying TPI. The traffic congestion of the macro road network is clustered into Mondays, Fridays, ordinary weekdays, congested weekdays, weekdays of winter and summer vacation, Saturdays, Sundays and festivals patterns. Patterns of Mondays and congested weekdays have a prominent morning peak, while patterns of Fridays, ordinary weekdays, and weekdays of winter and summer vacation have a prominent evening peak. Saturdays, Sundays and festivals are less congested than weekday patterns. It is verified that the SOM method proposed in this paper clusters traffic congestion into more detailed and accurate patterns, and it is applicable to TPI clustering in different years. The degree of congestion in 2021 increases by 7.15% in peak hours and decreases by 7.50% in off-peak hours compared with that in 2019 due to COVID-19. This method is helpful for traffic management in terms of making decisions according to different congestion patterns.

Keywords: congestion patterns; traffic performance index; clustering methods; traffic conditions

1. Introduction

With the increase in motor vehicles, traffic congestion is becoming increasingly serious, which has caused problems such as a reduction in traffic efficiency and the increase in the incidence of traffic accidents. Scientific and quantitative evaluation of traffic congestion is the fundamental work for easing traffic congestion. To quantitatively evaluate the degree of urban congestion, most cities have established traffic monitoring systems. TPI is one of the crucial indicators of intelligent transportation systems and can effectively evaluate the traffic status on the road network in Beijing. Beijing had 21.89 million residents and 6.22 million motor vehicles in 2021. Beijing is the most congested city during peak hours. The congested sections during peak hours are mainly concentrated in the ring road and connecting road. TPI is widely used because of its important role in traffic management, such as assessment of transportation demand management policies and traffic congestion alleviation tasks.
Travelers tend to take into account travel mode and travel time more than travel path in the multimodal transportation system. The traffic congestion condition in the destination area or even the entire road network is important reference information for travelers. Affected by different travel purposes on weekdays, weekends and holidays, daily traffic congestion presents different characteristics. Dates with the same congestion change characteristics are defined as having the same traffic congestion pattern in this paper. Due to the travel regularity of residents, traffic congestion can be divided into typical limited patterns. The traffic congestion pattern can be identified by analyzing the time-varying rule of traffic congestion in the road network between different days. The traffic flow status in the urban road network is described efficiently with limited patterns. However, traffic flow patterns are mostly developed based on the flow, speed, density and other indicators of road sections on specific road type in the existing studies. Dynamic traffic flow and static road networks are not considered comprehensively. The overall traffic congestion situation of the complex macro road network cannot be accurately evaluated.

In order to comprehensively evaluate the traffic congestion of the macro road network and effectively describe the traffic flow status of the road network with typical limited patterns, the traffic congestion patterns of the macro road network based on TPI are studied in this paper. First, a clustering method based on the self-organizing maps (SOM) neural network is proposed. The method of automatically determining the clustering numbers is constructed to improve the disadvantage that SOM requires the manual determination of clustering number. Second, clustering indexes are developed such as characteristic indexes of traffic congestion, TPIs weighted by the time-varying coefficient, and time-varying TPIs. Third, the evaluation method of the clustering effect is proposed. Finally, differences of congestion patterns in different years are analyzed as the case study. The results of this paper will provide travel suggestions for travelers and help traffic managers to predict and prevent serious congestion dates and take targeted congestion control measures for different congestion situations.

2. Literature Review

Congestion has become a serious problem in China. Evaluation indexes of traffic conditions have been developed, such as the Travel Time Index (TTI) [1], the Roadway Congestion Index (RCI) [2], and TomTom’s Worldwide Congestion Index (CI) [3]. These indexes provided effective methods for characterizing traffic congestion patterns and their temporal variations at a macroscopic level, in which TPIs are widely used. One of the most important applications of TPI is to identify the characteristics of traffic congestion. Sun et al. [4] classified the traffic performance characteristics of urban trunk roads and identified frequent traffic jams based on TPI data. Sun et al. [5] identified traffic patterns of main trunk roads using TPI data. Wei et al. [6] explored the spatiotemporal pattern of traffic congestion performance in large cities by using TPI data, which varies substantially between cities and different days of the week. Wei et al. [7] proposed a framework to identify turn-level bottlenecks in large-scale road networks based on TPI data. TPI is also an important index for traffic management and planning. Cheng et al. [8] constructed a mixed-integer second-order cone programming model for electric vehicle charging stations considering traffic congestion based on TPI data. Yan et al. [9] presented a spatial-temporal electric vehicle charging load profile simulation method considering weather and traffic based on TPI data. The calculation and prediction methods of TPI have been further improved. Yu et al. [10] proposed a method to calculate the optimal TPI based on the macroscopic fundamental diagram (MFD). Cheng et al. [11] proposed a combination prediction model of the traffic congestion index on weekdays based on LightGBM-GRU, considering seasonal factors and weekly periodicity. Yi et al. [12] proposed a deep learning method using LSTM-RNN models for predicting TPI in highways. Above all, TPI plays a critical role in traffic congestion evaluation, traffic planning and management. The changing characteristics and patterns of TPIs need to be further studied to facilitate better traffic decision-making.
Existing studies have indicated that when classified properly into different categories, the traffic congestion patterns of different dates in each category have similar characteristics, which provides the possibility of predicting the future traffic conditions according to the categorized characteristics. Methods have been studied on the division of traffic congestion patterns, based on different data sources and research methods in existing studies, as shown in Table 1.

Table 1. Literature on the division of traffic congestion patterns.

<table>
<thead>
<tr>
<th>Literature</th>
<th>Data</th>
<th>Method</th>
<th>Traffic Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wen et al. [13]</td>
<td>TPI data</td>
<td>Hierarchical clustering method</td>
<td>Normal Weekdays, Key Congested Weekdays, and Most Congested Weekdays</td>
</tr>
<tr>
<td>Calafate et al. [14]</td>
<td>Traffic flows</td>
<td>Clustering heuristic</td>
<td>Monday, Tuesday and Wednesday, Thursday, Friday, Saturday, and Sunday.</td>
</tr>
<tr>
<td>Sun et al. [15]</td>
<td>TPI data</td>
<td>Hierarchical clustering algorithm</td>
<td>Workdays, Latter half of vacation (October 4th–8th), and Weekends and the beginning of vacation (October 1st–3rd)</td>
</tr>
<tr>
<td>Nguyen et al. [16]</td>
<td>Speed</td>
<td>Clustering analysis</td>
<td>Disturbance, high frequency of disturbances, bottleneck congestion, small scaled homogeneous congested traffic, homogeneous and disturbances clusters</td>
</tr>
<tr>
<td>Toshniwal et al. [17]</td>
<td>Traffic volumes</td>
<td>CSTAT model</td>
<td>Saturdays, Sundays, public holidays, rainy days and weekdays</td>
</tr>
<tr>
<td>Loo and Huang [18]</td>
<td>Speed and CI</td>
<td>Natural Cities Model</td>
<td>Weekday, Saturday, and Sunday</td>
</tr>
<tr>
<td>Behara et al. [19]</td>
<td>OD matrices</td>
<td>DBSCAN-based framework</td>
<td>Saturdays, Sundays, public holidays, regular weekdays, and weekday school holidays</td>
</tr>
<tr>
<td>Li et al. [20]</td>
<td>Traffic congestion index</td>
<td>Weighted K-means clustering</td>
<td>H (holiday) pattern, MTW (Monday, Tuesday, and Wednesday) pattern, TF(Thursday and Friday) pattern</td>
</tr>
</tbody>
</table>

As shown in Table 1, classical clustering algorithms have been used to identify traffic congestion patterns in the existing studies, such as the hierarchical, k-means, and DBSCAN clustering method. The neural network algorithm is widely used because of its good performance. The SOM neural network clustering method is used in this paper to analyze traffic congestion patterns on the road network.

As a practical method of classifying patterns, the clustering method has been studied for decades [21,22]. Many studies divide traffic flow patterns based on clustering methods. Li et al. [23] established a regional congestion evaluation model by a clustering method based on congestion characteristics, which were divided into multiple mode, double peak mode and evening peak prolonged mode, and the prewarning criteria, including red alert, orange alert and yellow alert, were determined. Cheng et al. [24] developed a comprehensive classification indicator system including the ample degree of road network, traffic flow, speed and occupancy, which divided the traffic state into a fluent state, basic fluent state, slight congestion, moderate congestion, and severe congestion based on the improved FCM clustering method. Fu et al. [25] divided traffic statuses into serious congestion, moderate congestion, mild congestion, and no congestion based on the queuing time index (QTI), and the thresholds for different levels were obtained using the cluster analysis method to describe the traffic status of a signalized intersection. Neves et al. [26] proposed a method to divide traffic patterns of road mobility using biclustering, which
combines speed limits, vehicle passage frequencies, and the spatial extent of congested road segments. Zhang et al. [27] proposed a fuzzy weighted c-means classification method based on volume, speed and occupancy, which divided the traffic state into free flow, synchronous flow and congested traffic flow. The link-based traffic flow, speed and occupancy indicators within a certain area have been mainly used in the existing studies, while few have been conducted on traffic congestion patterns for an entire traffic network.

Above all, this paper intends to identify traffic congestion patterns based on TPI data of macro road network using SOM clustering method.

3. Methodology

The difference in travel purpose on different days leads to different traffic congestion patterns. To recognize the typical patterns of congestion change with time, three clustering methods of TPI curves are constructed in this paper. First, the self-organizing map (SOM) neural network clustering algorithm was used to cluster TPIs in 2018 due to its good clustering performance in the traffic flow field. A method for automatically determining the optimal number of clusters is constructed to improve the disadvantage that SOM requires manual determination of the number of clusters. Second, three clustering systems are developed, including characteristic indexes of traffic congestion, TPIs weighted by time-varying coefficients, and time-varying TPI. Third, evaluation indexes of the clustering effect are proposed to select the optimal clustering method, including consistency within categories, differences between categories and coverage ratio. Finally, the congestion patterns of 2019 and 2021 are constructed as a case study. Differences in congestion patterns in different years are analyzed. The general methodology of this research is shown in Figure 1.

![Figure 1. Methodological design.](image)

4. Data Preparation

The TPI is developed to evaluate the traffic condition of the urban road network, which quantifies the traveler’s perception of traffic congestion. To quantitatively evaluate the traffic congestion, the proportion of severe congestion mileage for the entire road network is converted to TPI (a number ranging from zero to ten). To help travelers intuitively understand road congestion through the TPI, the TPI is divided into five levels to represent different congestion conditions, namely, [0, 2), [2, 4), [4, 6), [6, 8) and [8, 10], corresponding to congestion levels of smooth traffic, no congestion, slight congestion, moderate congestion and severe congestion. The calculation method of the TPI is as follows, as shown in Figure 2.
Proportion of severe congestion (PSC) mileage for the entire road network

Corresponding relationship between PSC and TPI in Table 2

Figure 2. Development flow of TPI.

Step 1: Calculate the congestion level of each road based on speed in different road types, as shown in Table 2.

Table 2. Interval division of road type speed, the proportion of the congestion mileage and TPI corresponding traffic congestion level.

<table>
<thead>
<tr>
<th>Traffic Congestion Level</th>
<th>Smooth</th>
<th>Basically Smooth</th>
<th>Slight Congestion</th>
<th>Medium Congestion</th>
<th>Severe Congestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Express road speed</td>
<td>(65, +∞)</td>
<td>(50, 65]</td>
<td>(35, 50]</td>
<td>(20, 35]</td>
<td>(∞, 20]</td>
</tr>
<tr>
<td>Main road speed</td>
<td>(40, +∞)</td>
<td>(30, 40]</td>
<td>(20, 30]</td>
<td>(15, 20]</td>
<td>(∞, 15]</td>
</tr>
<tr>
<td>Secondary road speed</td>
<td>(35, +∞)</td>
<td>(20, 35]</td>
<td>(15, 25]</td>
<td>(10, 15]</td>
<td>(∞, 10]</td>
</tr>
<tr>
<td>Proportion of severe</td>
<td>[0%, 4%]</td>
<td>[5%, 8%]</td>
<td>[11%, 14%]</td>
<td>[11%, 14%]</td>
<td>[14%, +∞]</td>
</tr>
<tr>
<td>congestion mileage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPI</td>
<td>[0, 2]</td>
<td>[2, 4]</td>
<td>[4, 6]</td>
<td>[6, 8]</td>
<td>[8, 10]</td>
</tr>
</tbody>
</table>

The road types include express roads, main roads, and secondary roads. The speed comes from the floating car data (FCD) system in Beijing.

Step 2: Calculate the percentage of vehicle kilometers of travel (VKT) of different types to the total VKT of all roads. VKT is the accumulation of mileage of all vehicles on the road, which is calculated by Equation (1).

\[
VKT_i = \sum_{i=1}^{N_i} VKT_{R_i} = \sum_{i=1}^{N_i} Q_{R_i} \times L_{R_i}
\]  

(1)

where, \(i\) is the road type, including express road, main road and secondary road. \(VKT_i\) is the VKT (pcu*km) of road type \(i\), \(N_i\) is the number of road links in road type \(i\), \(VKT_i\) is the VKT of road \(R_i\), \(Q_{R_i}\) is the number of vehicles passing road \(R_i\), and \(L_{R_i}\) is the length of road \(R_i\). The percentage of \(VKT_i\) to the total VKT of all roads is calculated by Equation (2).

\[
P_{VKT_i} = \frac{VKT_i}{\sum_1^N VKT_i}
\]  

(2)

Step 3: Calculate the proportion of severe congestion mileage \(P_{SCR_i}\) for road type \(i\), as shown in Equation (3).

\[
P_{SCR_i} = \frac{L_{SCR_i}}{\sum_1^{M_i} L_{R_i}}
\]  

(3)

where \(L_{SCR_i}\) is the length of the road under the severe congestion of road type \(i\), \(M_i\) is the number of roads under severe congestion.

Step 4: Calculate the proportion of severe congestion mileage \(P_{VKTi}\) for the entire road network, as shown in Equation (4).

\[
P_{VKTi} = \sum_1^3 P_{SCRi} \times P_{VKTi}
\]  

(4)
Step 5: Convert $P_{VKTi}$ into TPI, as shown in Table 2.

Since travelers tend to have lower tolerance about severe congestion, which has a greater delay impact on the travel time, the proportion of the length of severely congested roads is used to determine the TPI in Beijing. Since roads with smooth traffic may have an offset effect on the level of congestion of the entire road network, using the proportion of severely congested roads can alleviate that effect and provide a more accurate evaluation of the congestion level. Notably, the development process or the specific calculation of the TPI is not the focus of this paper.

Based on the historical and real-time FCD, the TPI curve is formed by using the time as the horizontal axis and TPI as the vertical axis. TPI curves of a week and a day of National Day in 2021 are illustrated in Figure 3 as an example. There are significant differences such as the peak value and time of the morning and evening peak among the curves for different days. There are different patterns of congestion on different days.

![Figure 3. Samples of TPI curves.](image)

The TPI data at 15-min intervals of 365 days from January to December 2018 are used to develop clustering models. The TPI data of 212 days from January to September 2019 and 122 days from June to October 2021 are used as a case study to verify the adaptability of the model in different years and analyze the impact of COVID-19 on traffic congestion patterns. The data structure of the TPI of the entire network in Beijing includes date, time, week, holiday, weather and TPI, as shown in Table 3.

<table>
<thead>
<tr>
<th>Date (mm/dd/yyyy)</th>
<th>Time</th>
<th>Week</th>
<th>Holiday</th>
<th>Weather</th>
<th>TPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/4/2021</td>
<td>7:00</td>
<td>Monday</td>
<td>National Day</td>
<td>Light rain</td>
<td>1.46</td>
</tr>
<tr>
<td>10/4/2021</td>
<td>7:30</td>
<td>Monday</td>
<td>National Day</td>
<td>Light rain</td>
<td>3.29</td>
</tr>
<tr>
<td>10/22/2021</td>
<td>7:00</td>
<td>Friday</td>
<td>Weekday</td>
<td>Fine</td>
<td>1.76</td>
</tr>
<tr>
<td>10/22/2021</td>
<td>7:15</td>
<td>Friday</td>
<td>Weekday</td>
<td>Fine</td>
<td>8.02</td>
</tr>
</tbody>
</table>

5. Traffic Congestion Patterns Clustering Method
5.1. Clustering Algorithms
5.1.1. SOM Neural Network Clustering Algorithm

The SOM neural network was proposed by Professor Kohonen [28] and has been widely applied in different research fields [29–32]. The algorithm assumes that there are topological structures or orders in the input object, which can realize dimension reduction mapping from the input space to the output plane. The SOM network includes an input layer and output layer. The input layer corresponds to a high-dimensional input vector, and the output layer is composed of a series of ordered nodes organized on a two-dimensional grid. The input and output nodes are connected by weight vectors. In the learning
process, SOM finds the output layer unit with the shortest distance to update it. The weights of adjacent regions are updated so that the output nodes maintain the topological characteristics of the input vector. It is proven that it has a good clustering effect in the field of traffic flow. SOM clustering requires the following steps [32]:

Step 1: Initialize the network and assign the initial value to each node weight of the output layer.
Step 2: Randomly select the input vector from the input samples to find the weight vector with the minimum distance from the input vector.
Step 3: Define the winning unit, and adjust the weight in the adjacent area of the winning unit to make it close to the input vector.
Step 4: Provide new samples and conduct training.
Step 5: Shrink the neighborhood radius, reduce the learning rate, repeat until it is less than the allowable value, and output the clustering results.

5.1.2. Automatic Clustering Number Determination Algorithm

The number of clusters needs to be input in the SOM clustering process. A method to automatically determine the optimal number of clusters is proposed to improve the SOM algorithm. Based on the characteristics of the TPI curves in Beijing, the silhouette measure was adopted to determine the number of categories. The silhouette coefficients of different clustering numbers of TPI curves are automatically calculated based on the SOM algorithm. The optimal clustering number is determined when the profile coefficient reaches the maximum.

The silhouette measure $Sil(a)$ is a number ranging within $[-1, 1]$, as shown in Equation (5):

$$Sil(a) = \frac{a_a - b_a}{\max\{a_a, b_a\}} = \frac{\min_{K' \neq K(a)} \{\text{mean}_{A_c \in K'} |A_aA_c|\}}{\text{mean}_{A_b \in K(a)}|A_aA_b|}$$

where $a_a$ is the minimum value of the mean distance between points $A_a$ and $A_c$, $K'$ is a different category from $K(a)$, $|A_aA_c|$ is the distance between points $A_a$ (in $K(a)$) and $A_c$ (in $K'$), $b_a$ is the mean distance between points $A_a$ and $A_b$, $A_b \in K(a)$, $K(a)$ is the category of point $A_a$, and $|A_aA_b|$ is the distance between points $A_a$ and $A_b$.

A higher value of $Sil(a)$ theoretically means a better clustering result. For each clustering result of category number $C$, the mean silhouette value $P_C$ of all the $M_2$ points needs to be calculated by Equation (6):

$$P_C = \frac{1}{M_2} \sum_{a=1}^{M_2} Sil(a)$$

when $P_C$ reaches the maximum value, $C$ is determined to be the optimal number of clustering categories.

5.2. Clustering Indexes

Clustering indexes are developed based on the characteristics of traffic congestion, the value of TPIs and the weight of the time variation coefficient.

5.2.1. Key Characteristic Indexes of Traffic Congestion

Characteristic indexes are designed based on traffic congestion characteristics at different times of the day, such as the central tendency and dispersion degree of TPIs at peak and trough hours. The characteristic indexes of congestion are shown in Table 4.
Table 4. Characteristic indexes of traffic congestion.

<table>
<thead>
<tr>
<th>Index</th>
<th>Meaning in the TPI Curve</th>
<th>Index</th>
<th>Meaning in the TPI Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Maximum TPI of 0:00–24:00</td>
<td>X15</td>
<td>Average TPI of 10:00–12:00</td>
</tr>
<tr>
<td>X2</td>
<td>Minimum TPI of 0:00–24:00</td>
<td>X16</td>
<td>TPI variance of 10:00–12:00</td>
</tr>
<tr>
<td>X3</td>
<td>Ratio of maximum TPI to minimum TPI of 0:00–24:00</td>
<td>X17</td>
<td>Average TPI of 14:00–16:00</td>
</tr>
<tr>
<td>X4</td>
<td>Average TPI of 7:00–9:00</td>
<td>X18</td>
<td>TPI variance of 14:00–16:00</td>
</tr>
<tr>
<td>X5</td>
<td>TPI variance of 7:00–9:00</td>
<td>X19</td>
<td>Total TPI of 0:00–24:00</td>
</tr>
<tr>
<td>X6</td>
<td>Average TPI of 17:00–19:00</td>
<td>X20</td>
<td>Average TPI of 0:00–24:00</td>
</tr>
<tr>
<td>X7</td>
<td>TPI variance of 17:00–19:00</td>
<td>X21</td>
<td>TPI variance of 0:00–24:00</td>
</tr>
<tr>
<td>X8</td>
<td>Ratio of average TPI of 7:00–9:00 to 17:00–19:00</td>
<td>X22</td>
<td>Maximum TPI of 7:00–9:00</td>
</tr>
<tr>
<td>X9</td>
<td>Average TPI of 0:00–6:00</td>
<td>X23</td>
<td>Minimum TPI of 7:00–9:00</td>
</tr>
<tr>
<td>X10</td>
<td>TPI variance of 0:00–6:00</td>
<td>X24</td>
<td>Ratio of maximum TPI to minimum TPI of 7:00–9:00</td>
</tr>
<tr>
<td>X11</td>
<td>Average TPI of 20:00–24:00</td>
<td>X25</td>
<td>Maximum TPI of 17:00–19:00</td>
</tr>
<tr>
<td>X12</td>
<td>TPI variance of 20:00–24:00</td>
<td>X26</td>
<td>Minimum TPI of 17:00–19:00</td>
</tr>
<tr>
<td>X13</td>
<td>Average TPI of 12:00–14:00</td>
<td>X27</td>
<td>Ratio of maximum TPI to minimum TPI of 17:00–19:00</td>
</tr>
<tr>
<td>X14</td>
<td>TPI variance of 12:00–14:00</td>
<td>X28</td>
<td>Ratio of maximum TPI of 7:00–9:00 to 17:00–19:00</td>
</tr>
</tbody>
</table>

The Pearson correlation coefficients are calculated to reduce the correlation among characteristic indexes of traffic congestion. The redundant indexes with Pearson correlation above 0.8 are eliminated, and only one key index that can reflect the congestion characteristics is retained. Indexes of X3, X4, X5, X6, X7, X11, X13, X14, and X16 are selected as key indexes, which reflect characteristics of traffic congestion at morning peak, evening peak, night, trough hours and off-peak.

5.2.2. Time-Varying TPI

The time-varying TPI directly reflect the characteristics of congestion at different moments. TPIs of 15-min granularity in 24 h are selected as clustering indexes.

5.2.3. TPI Weighted by Time-Varying Coefficient

The weight is assigned to the clustering index to represent the different unit lengths for different dimensions. The axes are lengthened or shortened according to the weights. Therefore, the weight assignment significantly impacts the clustering result.

The coefficient of variation of the TPI at each time point was used as the weight in this paper. The coefficient of variation is a parameter used to measure the variation degree of the sample data. For the TPI at a specific time point with a higher variation degree, the difference in its congestion pattern is also greater; therefore, it should be assigned a higher weight, and vice versa. The coefficient of variation is shown in Equation (7):

\[
CV = \frac{\sigma_{X_t}}{E_{X_t}} = \sqrt{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_{1i} - E_{X_1})^2 / \left( \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i} \right)^2}
\]  

where \(E_{X_t}\) is the mean value of the TPI.

5.3. Evaluation Indexes of the Clustering Results

Indexes are used to evaluate clustering results, including consistency within categories, differences between categories, coverage ratio and silhouette.

5.3.1. Consistency within Categories

A strong consistency is requested within a category. The coefficient of variation of the TPI is used to measure the consistency. The calculation method of the coefficient of variation within-category is shown in Equation (8). Let \(CV_t\) be the coefficient of variation within a category; then,

\[
CV_t = \frac{1}{n_t} \sum_{i=1}^{n_t} CV_t^i = \frac{1}{n_t} \sum_{i=1}^{n_t} \left( \frac{1}{\sum_{j=1}^{n_t} CV_t^j} \right)
\]
where \( n \) is the number of categories, \( CV_k I \) is the mean value of the coefficient of variation for all time points in a day, \( l \) is the number of time intervals in a day, and \( CV_k i \) is the coefficient of variation of the TPI at time \( i \) in category \( k \).

The degree of consistency rises when the value of the \( CV_I \) decreases.

5.3.2. Differences between Categories

Greater differences among clustering categories generally indicate a better clustering result. Therefore, the coefficient of variation of the mean values of all categories is proposed as the index to measure the differences between categories, as shown in Equation (9). Let \( CV_{II} \) be the coefficient of variation between categories; then,

\[
CV_{II} = \frac{1}{l} \sum_{i=1}^{l} \left( \frac{E_{II,i}}{\sigma_{II,i}} \right) \tag{9}
\]

where \( l \) is the number of time intervals in a day, and \( CV_{II,i} \) is the coefficient of variation of \( TPI_k^j \) of all categories, as shown in Equation (10).

\[
TPI_k^j = \frac{1}{m} \sum_{j=1}^{m} TPI_{k,j,i} \tag{10}
\]

where \( TPI_k^j \) is the mean TPI of time \( i \) in category \( k \), \( TPI_{k,j,i} \) is the TPI of time \( i \) in day \( j \) in category \( k \), and \( m \) is the number of days in a category.

The differences increase when the \( CV_{II} \) rises.

5.3.3. Coverage Ratio

In terms of clustering the daily traffic patterns, it is ideal to classify all the dates into limited categories. However, days might be classified into incorrect categories because of the existence of inclement weather, incidents, important events, and an imperfect clustering method. The proportion of days correctly classified is defined as the coverage ratio, the average Pearson correlation coefficient of which with other variables of the same category is greater than 0.9, as shown in Equation (11).

\[
CR = \frac{D_1}{D} \times 100\% \tag{11}
\]

where \( D_1 \) is the number of correctly classified days and \( D \) is the number of all days.

The result of the clustering improves when the value of the coverage ratio becomes greater.

5.4. Discussion of Results

The SOM neural network clustering algorithm improved by the method of automatically determining the optimal clustering number are used to cluster the TPIs of 2018 under three clustering indexes. TPIs in 2018 are clustered into eight categories. The morning peak of weekend and festival patterns appears at 10:00–12:00, which is later than weekday patterns at 7:00–9:00. The evening peak of weekend and festival patterns appears at 16:00–18:00, which is earlier than weekday patterns at 17:00–19:00. The optimal clustering method is determined by comparing different clustering evaluation indexes.

5.4.1. Clustering Results of Key Characteristic Indexes of Traffic Congestion

The TPI curves and average trend curves of patterns based on key characteristic indexes of traffic congestion are shown in Figure 4.
5.4. Discussion of Results

The SOM neural network clustering algorithm improved by the method of automatically determining the optimal clustering number are used to cluster the TPIs of 2018 under three clustering indexes. TPIs in 2018 are clustered into eight categories. The morning peak of weekend and festival patterns appears at 10:00–12:00, which is later than weekday patterns at 7:00–9:00. The evening peak of weekend and festival patterns appears at 16:00–18:00, which is earlier than weekday patterns at 17:00–19:00. The optimal clustering method is determined by comparing different clustering evaluation indexes.

5.4.1. Clustering Results of Key Characteristic Indexes of Traffic Congestion

The TPI curves and average trend curves of patterns based on key characteristic indexes of traffic congestion are shown in Figure 4. As shown in Figure 4, TPI curves are clustered into eight categories based on key characteristic indexes of traffic congestion, including weekends, weekends and festivals, weekdays with prominent evening peaks, weekdays with prominent morning peaks, Fridays, Mondays, ordinary weekdays, and weekdays of winter and summer vacation patterns. The time-varying trend of TPI curves in each pattern is relatively consistent. The characteristics of weekdays are well described, while the characteristics of the patterns of weekends, weekends and festivals are not well extracted and distinguished. The absolute values of TPIs in some patterns differ greatly, such as pattern 1, pattern 2, pattern 3, and pattern 8. There are multiple curves in pattern 3 classified unreasonably.

The statistical characteristics of each pattern are shown in Table 5.
The following conclusions can be obtained from Table 5:

1. Pattern 1 and pattern 2 are holiday patterns. Pattern 1 is more congested than pattern 2. There are 75 samples in pattern 1, 73 of which are weekends. The maximum TPI of the evening peak is 36.74% higher than that of the morning peak in pattern 1. Pattern 2 has 39 samples, 22 of which are festivals, including the Spring Festival, International Labour Day, Dragon Boat Festival and National Day.

2. Pattern 3, pattern 5, pattern 7 and pattern 8 are weekday patterns with prominent evening peaks, in which the maximum TPI of the evening peak is 33.27%, 10.70%, 11.67%, and 11.34% higher than that of the morning peak, respectively. Pattern 3 has the morning peak with the lowest congestion due to the presence of unreasonably classified samples. Pattern 5 has 44 working day samples, 29 of which are Fridays. The congestion degree of the morning and evening peaks is the highest in pattern 5. Pattern 8 has 37 working day samples of winter and summer vacation, which has a lower degree of congestion due to the lack of students commuting to school. Pattern 7 has 47 working day samples, which is defined as the ordinary weekday pattern.

3. Pattern 4 and pattern 6 are weekday patterns with prominent morning peaks, in which the maximum TPI of the morning peak is 10.09% and 17.29% higher than that of the evening peak, respectively. Pattern 4 has 61 weekday samples without Fridays. Pattern 6 has 33 weekday samples, 25 of which are Mondays. The morning peak congestion is the highest in pattern 6.

5.4.2. Clustering Results of the Time-Varying TPI

TPI curves and average trend curves of patterns based on the time-varying TPI are shown in Figure 5.

As shown in Figure 5, TPI curves are clustered into eight categories based on time-varying TPI, including festivals, Saturdays, Sundays, weekdays of winter and summer vacation, Mondays, Fridays, ordinary weekdays, and congested weekdays. There are three patterns for holidays and five patterns for weekdays. The time-varying trend of TPI curves in each pattern based on time-varying TPI is more consistent based on key characteristic indexes of traffic congestion. The weekday and holiday curves are all divided into detailed patterns, the characteristics of which are well described.

The statistical characteristics of each pattern are shown in Table 6.
5.4.2. Clustering Results of the Time-Varying TPI curves and average trend curves of patterns based on the time-varying TPI are shown in Figure 5.

As shown in Figure 5, TPI curves are clustered into eight categories based on the time-varying TPI, including festivals, Saturdays, Sundays, weekdays of winter and summer vacation, Mondays, Fridays, ordinary weekdays, and congested weekdays. There are three patterns for holidays and five patterns for weekdays. The time-varying trend of TPI curves in each pattern based on time-varying TPI is more consistent based on key characteristic indexes of traffic congestion. The weekday and holiday curves are all divided into detailed patterns, the characteristics of which are well described. The statistical characteristics of each pattern are shown in Table 6.

The following conclusions can be obtained from Table 6:

1. Pattern 1, pattern 2 and pattern 3 are holiday patterns. There are 12 festival samples in pattern 1, which has no evening peak. Pattern 2 has 38 samples, 26 of which are Saturdays. The maximum TPI of the morning and evening peaks are similar in pattern 2. There are 67 weekend and festival samples in pattern 3, 43 of which are Sundays. The maximum TPI of the evening peak is 56.92% higher than that of the morning peak in pattern 3.

Table 6. Characteristics of patterns based on the time-varying TPI.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Weekdays</th>
<th>Weekends</th>
<th>Festivals</th>
<th>Max TPI of Morning Peak</th>
<th>Max TPI of Evening Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>7</td>
<td>5</td>
<td>11</td>
<td>3.98</td>
<td>~</td>
</tr>
<tr>
<td>P2</td>
<td>8</td>
<td>30, 26 Saturdays</td>
<td>10</td>
<td>4.84</td>
<td>5.1</td>
</tr>
<tr>
<td>P3</td>
<td>5</td>
<td>62, 43 Sundays</td>
<td>8</td>
<td>3.25</td>
<td>5.10</td>
</tr>
<tr>
<td>P4</td>
<td>36</td>
<td>1, 1 working day</td>
<td>0</td>
<td>5.91</td>
<td>6.91</td>
</tr>
<tr>
<td>P5</td>
<td>33, 27 Mondays</td>
<td>0</td>
<td>0</td>
<td>7.55</td>
<td>6.3</td>
</tr>
<tr>
<td>P6</td>
<td>43, 34 Fridays</td>
<td>4, 3 working days</td>
<td>0</td>
<td>6.81</td>
<td>7.8</td>
</tr>
<tr>
<td>P7</td>
<td>55</td>
<td>2, 2 working day</td>
<td>0</td>
<td>6.43</td>
<td>6.98</td>
</tr>
<tr>
<td>P8</td>
<td>74</td>
<td>0</td>
<td>0</td>
<td>7.19</td>
<td>6.66</td>
</tr>
</tbody>
</table>
2. Pattern 4, pattern 6, and pattern 7 are weekday patterns with prominent evening peaks, in which the maximum TPI of the evening peak is 16.92%, 14.54%, and 8.55% higher than that of the morning peak, respectively. Pattern 6 has 47 working day samples, 34 of which are Fridays. The congestion degree of the evening peak is the highest in pattern 6. The samples of pattern 4 are weekdays in winter and summer vacation, which has a lower degree of congestion due to the lack of students commuting to school. Pattern 7 has 57 working day samples, which is defined as the ordinary weekday pattern.

3. Pattern 5 and pattern 8 are weekday patterns with prominent morning peaks, in which the maximum TPI of the morning peak is 19.84% and 7.96% higher than that of the evening peak, respectively. Pattern 5 has 33 weekday samples, 27 of which are Mondays. The maximum TPI of the morning peak of pattern 5 is the highest in all patterns. Pattern 8 has 74 weekday samples, which is more congested than pattern 7, which is defined as the congested weekday pattern.

5.4.3. Clustering Results of TPI Weighted by Time-Varying Coefficient

TPI curves and average trend curves of patterns based on TPI weighted by the time-varying coefficient are shown in Figure 6.

Figure 6. TPI curves and average trend curves of patterns based on TPI weighted by the time-varying coefficient.
As shown in Figure 6, TPI curves are clustered into eight categories based on the time-varying TPI, including weekends, Saturdays and festivals, vacation weekdays with prominent evening peak, weekdays with prominent morning peak, Fridays, Mondays, ordinary weekdays, and weekdays of winter and summer vacation. There are two patterns for holidays and six patterns for weekdays. This index overemphasizes the characteristics of morning and evening peak traffic flow. The pattern of vacation weekdays with prominent evening peak is identified, the consistency of curves in which is low. The absolute values of TPIs in some patterns differ greatly, such as pattern 2, pattern 3, and pattern 8.

The statistical characteristics of each pattern are shown in Table 7.

**Table 7.** Characteristics of patterns based on TPI weighted by time-varying coefficient.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Weekdays</th>
<th>Weekends</th>
<th>Festivals</th>
<th>Max TPI of Morning Peak</th>
<th>Max TPI of Evening Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>5</td>
<td>68</td>
<td>9</td>
<td>3.42</td>
<td>5.18</td>
</tr>
<tr>
<td>P2</td>
<td>14</td>
<td>29</td>
<td>20</td>
<td>4.58</td>
<td>4.19</td>
</tr>
<tr>
<td>P3</td>
<td>15</td>
<td>2, 1 working day</td>
<td>8</td>
<td>5.43</td>
<td>6.95</td>
</tr>
<tr>
<td>P4</td>
<td>57</td>
<td>0</td>
<td>0</td>
<td>7.06</td>
<td>6.56</td>
</tr>
<tr>
<td>P5</td>
<td>37, 30 Fridays</td>
<td>3, 3 working days</td>
<td>0</td>
<td>6.89</td>
<td>7.89</td>
</tr>
<tr>
<td>P6</td>
<td>47, 34 Mondays</td>
<td>0</td>
<td>0</td>
<td>7.47</td>
<td>6.37</td>
</tr>
<tr>
<td>P7</td>
<td>52</td>
<td>1, 1 working day</td>
<td>0</td>
<td>6.89</td>
<td>7.34</td>
</tr>
<tr>
<td>P8</td>
<td>34</td>
<td>1, 1 working day</td>
<td>0</td>
<td>5.86</td>
<td>6.44</td>
</tr>
</tbody>
</table>

The following conclusions can be obtained from Table 7:

1. Pattern 1 and pattern 2 are holiday patterns. There are 73 weekend and festival samples in pattern 1, 68 of which are weekends. The maximum TPI of the evening peak is 51.46% higher than that of the morning peak in pattern 1. There are 43 weekend and festival samples in pattern 2. The maximum TPI of the morning peak is 9.31% higher than that of the evening peak in pattern 2.

2. Pattern 3, pattern 5, pattern 7, and pattern 8 are weekday patterns with prominent evening peaks, in which the maximum TPI of the evening peak is 27.99%, 14.51%, 6.53%, and 9.90% higher than that of the morning peak, respectively. Pattern 5 has 40 working day samples, 30 of which are Fridays. The congestion degree of the evening peak is the highest in pattern 5. Pattern 8 has 35 working day samples of winter and summer vacation, which has a lower degree of congestion due to the lack of students commuting to school. Pattern 3 has 17 working day samples of winter and summer vacation, in which the degree of congestion in the evening peak is obviously higher than that in the morning peak. Pattern 7, the ordinary weekday pattern, has 53 working day samples.

3. Pattern 4 and pattern 6 are weekday patterns with prominent morning peaks, in which the maximum TPI of the morning peak is 7.62% and 17.27% higher than that of the evening peak, respectively. Pattern 4 has 57 weekday samples, in which the degree of congestion in the morning peak is higher than that of the evening peak. Pattern 6 has 47 weekday samples, 34 of which are Mondays. The morning peak congestion is the highest in pattern 6.

5.4.4. Evaluation of Clustering Results of Different Indexes

The evaluation indexes are used to evaluate the clustering results of different indexes, including consistency within categories, differences between categories, coverage ratio and silhouette, as shown in Table 8.

**Table 8.** Evaluation indexes values of the clustering results.

<table>
<thead>
<tr>
<th>Index</th>
<th>CV$_{I}$</th>
<th>CV$_{II}$</th>
<th>CR</th>
<th>Silhouette</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key characteristic indexes of traffic congestion</td>
<td>0.18</td>
<td>0.64</td>
<td>76.29%</td>
<td>0.92</td>
</tr>
<tr>
<td>Time-varying TPI</td>
<td>0.19</td>
<td>0.70</td>
<td>90.73%</td>
<td>0.94</td>
</tr>
<tr>
<td>TPI index weighted by time-varying coefficient</td>
<td>0.20</td>
<td>0.69</td>
<td>87.40%</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Ideal clustering results are expected to have higher $CV_{II}$, $CR$, and silhouette and lower $CV_I$. The following conclusions can be obtained from Table 8:

1. The clustering result of the time-varying TPI has the highest values of $CV_{II}$, $CR$, silhouette and lower $CV_I$, which is consistent with practical experience and is beneficial to practical application. The traffic congestion patterns constructed by this method effectively describe 90.73% of the traffic flow on urban roads.

2. The clustering result of key characteristic indexes of traffic congestion has the lowest $CV_I$, $CV_{II}$, $CR$, and silhouette, which has a low proportion of reasonably classified samples.

3. The clustering result of the TPI index weighted by the time-varying coefficient has the highest values of $CV_I$, and silhouette and higher $CV_{II}$ and $CR$, which overemphasizes the influence of morning and evening peaks by the time-varying coefficient. Above all, TPIs are most effectively clustered based on the SOM clustering method improved by automatically determining the algorithm of the optimal clustering number, using the time-varying TPI.

Traffic congestion pattern recognition method of this paper is compared with Li et al. [20]. Li et al. [20] proposed a K-means clustering method improved by the coefficient of variation weight to identify traffic congestion patterns. MTW pattern (Monday, Tuesday, and Wednesday), TF pattern (Thursday and Friday) and H pattern (working days within 3 days before and after holidays) are developed based on TPIs from January to December 2017 of six districts in Beijing [20]. The silhouette coefficient of Haidian, Fengtai, Shijingshan, Chaoyang, Dongcheng, and Xicheng are respectively 0.26, 0.35, 0.16, 0.34, 0.30, and 0.29, and the average silhouette coefficient of all congestion patterns is 0.28 [20]. This paper proposed an SOM clustering method improved by the algorithm of automatically determining the clustering numbers. More detailed congestion patterns are developed such as festivals, Saturdays, Sundays, weekdays of winter and summer vacation, Mondays, Fridays, ordinary weekdays, and congested weekdays in this paper. The silhouette coefficient of congestion patterns in this paper is 0.94, which is higher than the reference of Li et al. [20]. The closer the value of the silhouette coefficient is to 1, the better the clustering results. The clustering accuracy of traffic congestion patterns is improved by the SOM clustering method proposed in this paper.

6. Case Study

The TPIs in 2019 and 2021 are clustered as a case study based on the time-varying TPI to verify the rationality of the clustering method. The clustering results in different years are compared to study the impact of COVID-19 on congestion patterns.

6.1. Clustering Results of TPIs in 2019

The TPIs of 212 days from January to September 2019 are clustered into eight categories based on the time-varying TPI, including weekends with prominent evening peaks, Saturdays and festivals, Sundays, weekdays of winter and summer vacation, Mondays, Fridays, ordinary weekdays, and weekdays with prominent evening peaks. The TPI curves and average trend curve of each pattern are shown in Figure 7.
Figure 7. TPI curves and average trend curves of different patterns in 2019.

As shown in Figure 7, there are three patterns for holidays and five patterns for weekdays. Pattern 1, pattern 2 and pattern 3 are weekends and festivals, and the other patterns are weekdays. The characteristics of each pattern are shown in Table 9.


<table>
<thead>
<tr>
<th>Pattern</th>
<th>Weekdays</th>
<th>Weekends</th>
<th>Festivals</th>
<th>Max TPI of Morning Peak</th>
<th>Max TPI of Evening Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>2</td>
<td>22</td>
<td>2</td>
<td>3.24</td>
<td>4.79</td>
</tr>
<tr>
<td>P2</td>
<td>7</td>
<td>18</td>
<td>8</td>
<td>4.46</td>
<td>4.27</td>
</tr>
<tr>
<td>P3</td>
<td>0</td>
<td>17, 13 Sundays</td>
<td>3</td>
<td>3.26</td>
<td>5.43</td>
</tr>
<tr>
<td>P4</td>
<td>28</td>
<td>1, 1 working day</td>
<td>0</td>
<td>5.73</td>
<td>6.60</td>
</tr>
<tr>
<td>P5</td>
<td>25, 16 Mondays</td>
<td>0</td>
<td>0</td>
<td>7.65</td>
<td>6.53</td>
</tr>
<tr>
<td>P6</td>
<td>19, 14 Fridays</td>
<td>0</td>
<td>0</td>
<td>6.89</td>
<td>8.00</td>
</tr>
<tr>
<td>P7</td>
<td>50</td>
<td>1, 1 working day</td>
<td>0</td>
<td>7.15</td>
<td>7.25</td>
</tr>
<tr>
<td>P8</td>
<td>9</td>
<td>1, 1 working day</td>
<td>0</td>
<td>4.39</td>
<td>6.01</td>
</tr>
</tbody>
</table>

The following conclusions can be obtained from Table 9:

1. Pattern 1, pattern 2, and pattern 3 are holiday patterns. There are 24 samples in pattern 1, 22 of which are weekends. The maximum TPI of the evening peak is 47.84% higher than that of the morning peak in pattern 1. There are 25 weekend and festival samples in pattern 2. The maximum TPI of the morning peak is 4.45% higher than that of the evening peak in pattern 2. There are 17 weekend samples in pattern 3, 13 of which are Sundays. The maximum TPI of the evening peak is 66.56% higher than that of the morning peak in pattern 3.

2. Pattern 4, pattern 6, pattern 7, and pattern 8 are weekday patterns with prominent evening peaks, in which the maximum TPI of the evening peak is 15.18%, 16.11%, 1.40% and 36.90% higher than that of the morning peak, respectively. Pattern 6 has
19 working day samples, 14 of which are Fridays. The congestion degree of the evening peak is the highest in pattern 6. Pattern 4 has 39 working day samples of winter and summer vacation, which has a lower degree of congestion due to the lack of students commuting to school. Pattern 7 has 50 weekday samples, in which the degree of congestion in the morning peak is close to that in the evening peak. Pattern 8 has 10 working day samples, in which the degree of congestion in the evening peak is obviously higher than that in the morning peak.

3. Pattern 5 is a weekday pattern with a prominent morning peak, in which the maximum TPI of the morning peak is 17.15% higher than that of the evening peak. Pattern 5 has 25 weekday samples, 16 of which are Mondays. The morning peak congestion is the highest in pattern 5.

6.2. Clustering Results of TPIs in 2021

The TPIs of 122 days from June to October 2021 are clustered into eight categories based on the time-varying TPI, including Saturdays and festivals, Sundays and festivals, weekdays with prominent evening peak, weekdays of winter and summer vacation, weekdays with prominent off-peak, Fridays, ordinary weekdays, and congested weekdays. The TPI curves and average trend curve of each pattern are shown in Figure 8.

![Figure 8: TPI curves and average trend curves of different patterns in 2021.](image)

As shown in Figure 8, there are two patterns for holidays and six patterns for weekdays. Pattern 1 and pattern 2 are weekends and festivals, and the other patterns are weekdays. The morning peak of weekend and festival patterns appears at 10:00–12:00, which is later than weekday patterns at 7:00–9:00. The evening peak of weekend and festival
The following conclusions can be obtained from Table 10:

1. Pattern 1 and pattern 2 are holiday patterns. There are 11 weekend and festival samples in pattern 1, 9 of which are Saturdays. The maximum TPI of the evening peak is 10.43% higher than that of the morning peak in pattern 1. There are 24 weekend and festival samples in pattern 2, 13 of which are Saturdays. The maximum TPI of the morning peak is 50.14% higher than that of the evening peak in pattern 2.

2. Pattern 3, pattern 4, pattern 5, and pattern 6 are weekday patterns with prominent evening peaks, and the maximum TPI of the evening peak is 23.37%, 9.40%, 12.74%, and 9.93%, respectively, higher than that of the morning peak. The difference in congestion degree between morning and evening peaks is the most obvious in pattern 3. Pattern 5 has 11 weekday samples, 10 of which are Fridays. The traffic flows at the morning peak and evening peak are moderate congestion and severe congestion, respectively, in Pattern 5. Pattern 4 has 22 weekday samples of winter and summer vacation, which has a lower degree of congestion due to the lack of students commuting to school. The level of traffic congestion during the off-peak hours is higher than that in other patterns.

3. Pattern 7 and pattern 8 have small differences in the traffic congestion degree between the morning and evening peaks, and the maximum TPI of the morning peak is 0.68% and 0.76%, respectively, higher than that of the evening peak. The traffic flows at the morning peak and evening peak are all moderate congestion in pattern 7 and pattern 8. Pattern 8 is a more congested weekday pattern than pattern 7.

6.3. Comparison of Clustering Results in Different Years

The change rules of traffic congestion in different years have certain similarities and differences. There are some common traffic congestion patterns between different years, such as weekdays of winter and summer vacation, ordinary weekdays, Fridays, and weekends and festivals patterns. Traffic congestion patterns in different years also have different characteristics. For example, a congested weekday pattern exists in 2018 and 2021, but not in 2019; a Monday pattern exists in 2018 and 2019, but not in 2021. To analyze the applicability of the model constructed in this paper in different years and the impact of COVID-19 on congestion, the average trend curves of different patterns in 2018, 2019 and 2021 are compared in Figure 9.
The following conclusions can be drawn from Table 11:

1. Congestion during peak hours in 2019 shows an increasing trend compared with that in 2018. The maximum TPI of peak hours of different patterns in 2019 is higher by 3.94% on average than that in 2018, except for weekdays of winter and summer vacation and Saturdays and festivals. The average TPI between 10:00 and 12:00 in 2019 was lower by 13.68% on average than that in 2018.

2. The congestion of peak hours in 2021 shows an increasing trend compared with that in 2018 and 2019. The maximum TPI of peak hours of different patterns in 2021 increases by 8.35% and 7.15% on average compared with that of 2018 and 2019, respectively. The average TPI between 10:00 and 12:00 of weekday patterns in 2021 is lower by 13.10% and 7.50% on average compared with that of 2018 and 2019.

3. The degree of congestion in the morning and evening peak increases in 2021 after COVID-19. The travel modes of some commuters are shifted from public transport to cars, considering health factors. Thus, the degree of road traffic congestion in the morning and evening peak commuting hours in 2021 is higher than that in 2019. The degree of congestion in off-peak hours decreases in 2021, indicating that travel activities in off-peak hours have decreased due to COVID-19.

Table 11. Average TPI of morning peak, evening peak, and off-peak hours of different patterns in 2018, 2019, and 2021.

<table>
<thead>
<tr>
<th>Contrast Content</th>
<th>Average TPI of Morning Peak</th>
<th>Average TPI of Evening Peak</th>
<th>Average TPI of Off-Peak Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average TPI in Q3</td>
<td>5.36</td>
<td>5.56</td>
<td>5.66</td>
</tr>
<tr>
<td>Weekday of winter and summer vacation</td>
<td>5.91</td>
<td>5.73</td>
<td>6.49</td>
</tr>
<tr>
<td>Ordinary weekday</td>
<td>6.43</td>
<td>7.15</td>
<td>7.44</td>
</tr>
<tr>
<td>Congested weekday</td>
<td>7.19</td>
<td>7.95</td>
<td>6.66</td>
</tr>
<tr>
<td>Friday</td>
<td>6.81</td>
<td>6.89</td>
<td>7.45</td>
</tr>
<tr>
<td>Monday</td>
<td>7.55</td>
<td>7.66</td>
<td>7.15</td>
</tr>
<tr>
<td>Saturday and Festival</td>
<td>4.84</td>
<td>4.46</td>
<td>5.40</td>
</tr>
<tr>
<td>Sunday and Festival</td>
<td>3.25</td>
<td>3.26</td>
<td>3.55</td>
</tr>
</tbody>
</table>
7. Conclusions and Recommendations

Traffic congestion in large cities has become one of the most significant sustainability problems. It is necessary to study the traffic congestion patterns on the road network. The traffic congestion pattern recognition method is proposed in this paper based on the large amount of the TPI data. First, the SOM clustering method improved by the algorithm of automatically determining the optimal number of clusters is proposed. Second, cluster indexes and evaluation indexes are constructed. Third, the clustering results in different years are compared to analyze the impact of COVID-19 on congestion patterns as a case study. This paper provides a reference for other cities to develop the localized traffic congestion patterns. The main findings of this paper are summarized as follows:

(1) TPIs are most effectively clustered based on the SOM clustering method improved by the automatically determining algorithm of the optimal clustering number, using the time-varying TPI. The traffic congestion patterns developed by this method describe 90.73% of the traffic flow on urban roads. Compared with the K-means method used in the existing research of Li et al. [20], the SOM clustering method has higher accuracy in the identification of traffic congestion patterns and effectively improves the silhouette coefficient. The SOM clustering method clusters TPI curves into more detailed traffic congestion patterns, which can describe the characteristics of traffic flow well.

(2) TPIs on the road network can be generally clustered into weekend and festival patterns (Saturdays and festivals, Sundays and festivals) and different weekday patterns (Mondays, Fridays, ordinary weekdays, congested weekday, and weekdays of winter and summer vacation). The level of congestion on weekends and festivals is lower than that on weekdays. The morning peak of weekend and festival patterns appears at 10:00–12:00, which is later than weekday patterns at 7:00–9:00. The evening peak of weekend and festival patterns appears at 16:00–18:00, which is earlier than weekday patterns at 17:00–19:00.

(3) Monday and congested weekday patterns are weekday patterns with prominent morning peaks, and the maximum TPI of the morning peak is 18.50% and 4.36%, respectively, higher than that of the evening peak. Fridays, ordinary weekdays, and weekdays of winter and summer vacation patterns are weekday patterns with prominent evening peaks, and the maximum TPI of the evening peak is 13.53%, 3.09%, and 13.83%, respectively, higher than that of the morning peak.

(4) The degree of congestion in 2021 increases by 7.15% in peak hours and decreases by 7.50% in off-peak hours compared to in 2019. The proportion of travel modes in peak hours has shifted from public transport to cars, and the amount of travel activities in off-peak hours has decreased due to COVID-19.

The traffic congestion patterns in Beijing are studied based on TPIs in this paper. The application effect of the method proposed in this paper in other cities needs to be further studied, and the congestion patterns in different cities should be analyzed in the future.

Author Contributions: This work is the result of collaboration among J.Z., P.J., S.L., X.Z., G.S. and L.Y. All authors have equally contributed, reviewed, and improved the manuscript. All authors have revised the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Fundamental Research Funds for the Universities of Beijing (No. X21063), the Natural Science Foundation of China (NSFC) (No. 52172301, 51678045) and the Major Program of the National Social Science Foundation of China (No. 21ZDA029).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Some or all data and models that support the findings of this study are available from the corresponding author upon reasonable request.
Acknowledgments: This research is supported by the Fundamental Research Funds for the Universities of Beijing (X21063), the Natural Science Foundation of China (NSFC) (No. 52172301 and 51678045) and Major Program of the National Social Science Foundation of China (21ZDA029).

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

References
5. Sun, Q.; Chu, G.; Li, Q.; Zhang, Y. Identification of recurrent congestion in main trunk road based on grid and analysis on influencing factors. J. Adv. Transp. 2022, 4729719. [CrossRef]


29. Pd, A.; Ldg, B.; Rm, A. Smoothed self-organizing map for robust clustering—Sciencedirect. *Inf. Sci.* 2020, 512, 381–401. [CrossRef]


31. Amiri, V.; Nakagawa, K. Using a linear discriminant analysis (lda)-based nomenclature system and self-organizing maps (som) for spatiotemporal assessment of groundwater quality in a coastal aquifer. *J. Hydrol.* 2021, 603, 127082. [CrossRef]


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