ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Study on Path Optimization and Operational Efficiency Enhancement of Toll Road Network Based on Improved Dijkstra's Algorithm

Mengyao Liu[1,2,*], *Ting Li*[1] *and Hongwei Sun*[2]

[1] School of Transportation Engineering, Chang'an University, Xi'an, Shaanxi, 710018, China
[2] School of Economics and Management University, Shaanxi College of Communications Technology, Xi'an, Shaanxi, 710018, China

SUMMARY: *This paper first introduces the improved Dijkstra's algorithm and graph theory to analyze the effect of network structure construction and the effect of shortest path calculation of highways in toll road network of Guangxi Province; after that, using Data Envelopment Analysis (DEA), a model of toll road operation efficiency is constructed to analyze the trend of change and the characteristics of spatial distribution of the operation efficiency of toll roads in 29 provinces of China from 2020 to 2024. After that, we constructed a toll road operational efficiency model using the data envelopment analysis (DEA), and analyzed the trend and spatial distribution characteristics of toll road operational efficiency from 2020 to 2024 in 29 provinces in China. The results show that: the inter-city and county connectivity of the toll road network in Guangxi is better (the average degree is 14.5014), the inner cities and counties are closely connected (the clustering coefficient is 0.7931), and the proximity of each node of the highway network is larger (0.3191~0.6387), and the nodes have a better connectivity efficiency, especially Nanning City has the smallest average shortest distance between Nanning City and other nodes. When the Euclidean distance between the starting site and the destination site is small, the distribution of the shortest path ratio is more dispersed; with the increase of the Euclidean distance, the distribution of the shortest path ratio is gradually centralized, and the value is taken on both sides of 1.228. There is no obvious trend of growth in the operational efficiency of the toll road in the period of 2020-2024, and it shows a distribution pattern of gradual decrease from east to west. The operational efficiency of toll roads has a strong positive spatial correlation, and the eastern and northwestern regions show significant "high-high" and "low-low" clustering characteristics, respectively.*

KEYWORDS: *Improved Dijkstra's algorithm; Graph theory; DEA method; Shortest path; Operational efficiency*

1 Introduction

The toll roads can be defined as a modern version of highways transportation infrastructures, which help reduce the distance between cities, improve the traffic efficiency, as well as stimulate the development of regions and rural or urban areas. Toll roads are an essential part of building national economies [1-3]. Nowadays, the toll roads in China have become one of the world's largest and densest highway systems, which include numerous transport facilities and junctions [4]. However, amid the vast transportation networks, some challenges have arisen, such as differences in functionality among various types of highways, patterns of

*ASDF_89441@163.com
https://doi.org/10.65102/is2026439 Published: 30 April, 2026

spatiotemporal distribution of traffic flows, and traffic loads of specific road sections. The issues may not only negatively affect the service potential of the toll-road network but also impose additional demands on national transport management, transportation infrastructure development, and engineering constructions. Therefore, studying the characteristics of toll-road operations is highly necessary [5, 6].

Over the past few years, owing to the fast evolution of information technology and big data analysis, there have been major improvements in the process of collecting, processing, and analyzing traffic data [7, 8]. Particularly, the massive amounts of traffic data obtained from facilities like highway checkpoints and traffic control centers contain very valuable data on the functioning of the network of roads. The data include numerous aspects such as traffic volume, speed, lane occupancy, accidents, and driving behavior, which makes it possible to base decision-making about transport much more reliably [9, 10]. Hence, finding an effective way to evaluate the performance of toll road networks using multisource traffic data analysis has become one of the most important issues in transportation studies [11].

Dijkstra's algorithm can be described as one of the classic methods for solving the shortest path problem in graphs. The method involves a process known as a greedy search, whereby the algorithm expands outwards from layers defined by the edges' weights and chooses the node closest to the starting point each time, until the target node is reached [12-14]. In its basis, Dijkstra's algorithm has been used in several applications ranging from route optimization through roads to the analysis of road network efficiency.

Firstly, this paper presents an overview of graph theory and the measures used to describe network structural attributes. The study then highlights some limitations that may occur due to the application of the traditional Dijkstra algorithm in the optimization of routes in the toll road network, which include a wide search range, the need to visit numerous nodes, time-consuming computations, and the tendency of adjacency matrices to increase the storage capacity of the algorithm. Based on the above information, the paper develops an elliptic search algorithm to limit the search range of the shortest route. In this regard, the paper uses the toll road network in Guangxi as a prototype to establish the network structure as a full road network and calculates the shortest route in the road network. Finally, the toll road network operational efficiency of 29 Chinese provinces during the period 2020-2024 is determined by using a toll road network operation efficiency model and DEA.

2 Toll road network path optimization based on improved Dijkstra's algorithm

2.1 Concepts and statistical indicators of graph theory and complex networks

2.1.1 Graph theory and complex networks

(1) Basic concepts of graph theory

Graph theory is a branch of mathematics, the study of which is a generalization of graphs. A concrete network is expressed by an abstract graph $G=(V,E)$, where V is the set of points consisting of all nodes in the network, and E is the set of edges consisting of all nodes connected in the network, and for each edge e in E, there are two points (u,v) in V expressing each other. Denote the number of vertices as $N=|V|$ and the number of edges as $L=|E|$.

(2) Definition of complex network

Complex networks can be represented by the graph theory of the set of graphs, so that the advantages of the operation is intuitive image, and easy to understand. There are two ways to describe the network: in general, for simple networks can be used in the form of data lists, including the number of nodes N, the number of edges between the nodes $l(e)$, the connection between the nodes; for a larger amount of data, the network of more connections between the nodes can be used in the form of matrices, which is conducive to the computer storage and operation. Usually, the network used only to describe the relative position of the nodes and the connection status of the edges is called a topological network. In complex networks, the adjacency matrix A_{ij} can be established to represent the structural information of the topological network, and the matrix element a_{ij} denotes the connection state of node i and node j, if node i and node j are connected to each other, then let $a_{ij} = 0$; otherwise, let $a_{ij} = 1$. The adjacency matrix A_{ij} reflects the most basic information of the network and is the basis for analyzing the network behavior and network characteristics.

2.1.2 Complex network statistical indicators

1) Degree and degree distribution

Degree (k) is a characteristic quantity used to characterize how many connected nodes there are. Denote the degree of vertex i as K_i. In the highway network, the larger the degree value of a provincial city or a route means that the point is more important to a certain extent. An undirected graph consisting of n vertices can be represented in degree using the adjacency matrix, i.e.:

$$K_i = \sum_{j=1}^{n} A_{ij} \tag{1}$$

where n represents the total number of programs, A_{ij} denotes whether node i and node j are connected or not, if articulated then $A_{ij} = 1$ and vice versa, the mean value of the degree in the whole network is given by the expression:

$$\bar{K} = \frac{1}{n} \sum_{i=1}^{n} K_i \tag{2}$$

The degree distribution can be expressed as a function $p(k)$, where $p(k)$ is the probability that the degree of any node is exactly k.

Degree distribution consists mainly of exponential, power-law, and Poisson distributions. Concerning node degree measures, the highest degree is $N-1$, which implies that all pairs of nodes in the network have direct links between them, whereas the lowest degree is 0, meaning that the node remains disconnected from other nodes. The normalized node-degree measure formula is:

$$K_i' = \frac{k_{v_i}}{N-1} \tag{3}$$

Through the empirical examination of the real network, node degree refers to the number of connections a provincial capital has with other provincial level cities. The high degree signifies that the city has more connections with other cities, thus becoming a significant node in the network.

2) Proximity to the center

Closeness Centrality can be described as the inverse of the average shortest path length from node i to all other nodes in the network, reflecting how close a particular node is to the network itself. It is expressed mathematically as:

$$C(v_i) = 1 / \sum_{j=1}^{n} d(v_i, v_j) \tag{4}$$

where $d(v_i, v_j)$ denotes the shortest distance from node v_i to v_j, i.e., the number of the least number of nodes traversed by the path connecting the two nodes, which is interpreted as the number of provincial municipalities that need to be passed through the highway network.

Considering the size of the network, closeness centrality needs to be standardized, and the standardized expression is:

$$C'(v_i) = (N-1) / \sum_{j=1}^{n} d(v_i, v_j) \tag{5}$$

The proximity closeness centrality provides a measure of the density of proximity between cities. The average of this metric within the entire network indicates how efficient the operation of the network is. The higher the index, the more efficient the network operates.

3) Intermediary Centrality

Betweenness centrality is also used to measure the importance of nodes in a complex network, with particular emphasis on the hub-transfer function of nodes in path selection throughout the network. The expression of betweenness centrality is:

$$C_B(v_i) = \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{P_{ijk}}{P_{jk}} \tag{6}$$

where P_{jk} is the number of shortest paths between node j to node k; P_{jk} is the number of shortest paths between node j to node k through node i. Considering the scale of nodes in the whole network, the intermediary centrality index is standardized, and the standardized expression is obtained as:

$$C'_B(v_i) = \frac{C_B(v_i)}{C_{N-1}^{2}} \tag{7}$$

A provincial capital having a larger betweenness centrality value implies that it appears more frequently in the shortest path. This would increase the status of the provincial capital as a transfer hub and enhance the control exerted by it within the whole system. Hence, this particular index is used to measure the efficiency of the whole system.

4) Other Indicators

In addition to the basic evaluation indices discussed above, complex network analysis also includes several other commonly used indicators, which describe the structural characteristics of the network from different perspectives..

(1) Average path length

The average path length (denoted by L) is the average value of the distance d_{ij} between node i and node j in the network. The maximum value of the length of the distance between two points in the network is called the diameter of the network and is denoted as D, i.e., $D = \max_{i,j} d_{ij}$. The average path length in a network is the average of the distance lengths between pairs of nodes in the network. i.e.:

$$L = \frac{2}{N(N-1)} \sum_{i \neq j} d_{ij} \tag{8}$$

where N represents the number of nodes in the network.

(2) Clustering coefficient

Clustering coefficient (denoted by c), defined as the probability that any two neighbor nodes of the node are connected to each other, the nodes directly connected to node j in the network are called the neighbor nodes of j, and the number of neighbor nodes is denoted as M_j, and the number of realistically existed edges between the nodes is E_j, and the number of all possible edges is $C_{M_j}^2$, and its The clustering coefficient c_j of node j is:

$$c_j = \frac{E_j}{C_{M_j}^2} = \frac{2E_j}{M_j(M_j - 1)} \tag{9}$$

The average clustering coefficient, $\bar{c}$, can be expressed as:

$$\bar{c} = \frac{1}{N(n)} \sum_j c_j \tag{10}$$

2.2 Improvement of Dijkstra's algorithm

2.2.1 Analysis of Dijkstra's algorithm

The shortest path problem between two nodes in a real city road network is generally solved using the Dijkstra algorithm. As the process goes on, the network nodes can be usually categorized into three kinds, i. e., unlabelled nodes, temporary nodes with labels and permanent nodes with labels.

When implementing Dijkstra’s algorithm in calculating the shortest distance between nodes in urban road networks there will be issues like too large a searching range, searching more nodes and excessive computing time. Consequently, in this paper, we carry out an extensive research on the ellipse search algorithm and rectangle search algorithm due to the specificity of road networks. The searches ranges and the formulae of the regions are discussed and empirical comparisons of both the algorithms are made.

2.2.2 Dijkstra's algorithm for storage structure optimization

(1) Adjacency matrix

Adjacency matrix is a data structure that uses an array of $n \times n$ to store the relationship between nodes and nodes in the graph, using the adjacency matrix of the graph can not only determine the connection relationship between two nodes, but also intuitively get the size of the weights of the edges between the two nodes, if the node v_i and the node v_j are connected, then the v_i rows of v_j in the array of the array is the weight of the edge formed by these two nodes. If graph G is an assignment graph, then the corresponding adjacency matrix is an $n \times n$ matrix with the following expression:

$$\operatorname{arc}[i][j] = \begin{cases} w_{ij}, \text{If } (v_i, v_j) \in E \text{ Or } \langle v_i, v_j \rangle \in E \\ 0, \mathit{If}\ i = j \\ \infty, \mathit{Other} \end{cases} \tag{11}$$

Figures 1 and 2 show a directed graph and its adjacency matrix, respectively. It can be seen that the number of elements of the adjacency matrix is n^2, so it can be obtained that the space complexity of storing the urban road network with the adjacency matrix is $O(n^2)$, that is, it is only related to the number of vertices, but not related to the number of edges or arcs. Therefore, this paper does not use the storage method of adjacency matrix, but the storage method of adjacency table.

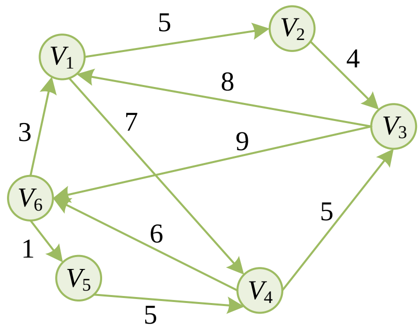


Figure 1: Directed graph

$$\begin{bmatrix} \infty & 5 & \infty & 7 & \infty & \infty \\ \infty & \infty & 4 & \infty & \infty & \infty \\ 8 & \infty & \infty & \infty & \infty & 9 \\ \infty & \infty & 5 & \infty & \infty & 6 \\ \infty & \infty & \infty & 5 & \infty & \infty \\ 3 & \infty & \infty & \infty & 1 & \infty \end{bmatrix}$$

Figure 2: Adjacency matrix

(2) Adjacency table

The adjacency table creates an array $a[v](v = 0,1,2,\ldots,n)$ for each node in the graph G to store the information of the edges connected to that node, i.e., each element of the array represents an edge from node v. The information of each edge consists of three parts, which are the number corresponding to the neighbor w of v, the weight of the edge vw, and the pointer to the next neighbor.

2.2.3 Elliptic Search Algorithm

1) Detailed description of the ellipse search algorithm

In the classical Dijkstra algorithm, the search area is circular. Although the circular search strategy yields a fairly accurate answer, the situation is that the search area is very large and the complexity of the algorithm is also too high, which has a negative impact on real-time performance of the process. The figure illustrating the Dijkstra algorithm with the elliptic search constraint is given below (Figure 3). According to this approach, the first and second nodes can be treated as the foci of the ellipse, namely F1 and F2, and their distance equals 2c. Consequently, the elliptical search strategy will be considered as a series of concentric ellipses created with the help of the beginning and ending nodes.

Let the graph be $ov_1 = 5$ and $v_1 d = 2$, $ov_2 = 2$ and $v_2 d = 8$. By quoting the above assumptions about the traditional algorithm, when using the elliptic algorithm, the size of $ov_2 + v_2 d$ and $ov_1 + v_1 d$ is first compared, and since $ov_1 + v_1 d < ov_2 + v_2 d$, the temporarily labeled node v_1 is noted as labeled, and the above steps are repeated to continue with the the next round of search.

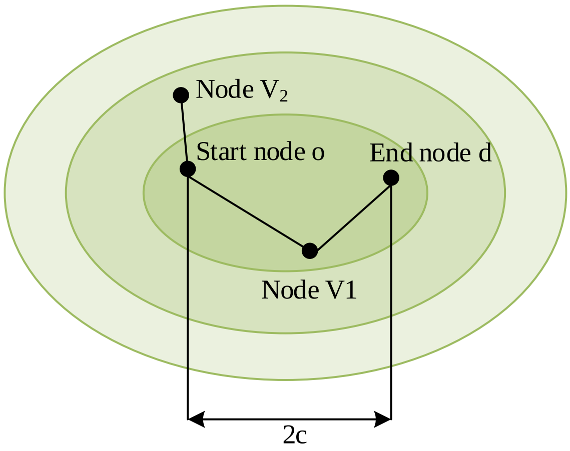


Figure 3: Dijkstra Algorithm for Elliptical Restricted Search Area

2) Determination of the search range of the ellipse

To determine the search range of the elliptic algorithm, it is necessary to find the specific value of the long axis of the ellipse, that is, the value of a. After finding the value of a, the boundary of the elliptic search algorithm can be expressed as an elliptic trajectory satisfying the formula $|ov_i| + |v_i d| = 2a$, and then search for the shortest paths between nodes inside the ellipse. The calculation process of the elliptic trajectory is given below:

Assume that the long axis of the ellipse is $r|od|$, r represents the confidence coefficient, and $|od|$ represents the straight-line distance between two points o and d. $|od|$ is easy to find when the coordinates of the two points are known, so the key to determining the search range of the algorithm lies in finding the appropriate value of the coefficient r, and the computation process for the value of r is given below:

A is a collection of point pairs for finding the shortest path composed of the starting node set O and the terminating node set D:

$$A = O \times D\{(o,d) \mid o \in O \,\&\, d \in D\} \tag{12}$$

A schematic diagram of the values taken by r is shown in Figure 4. Place each of the obtained ratios r_i into the set T, and for the set T, find a parameter r_0 such that the $r_i \le r_0$ confidence level is 95%, and after finding the r_0 that satisfies the condition, it can be brought to the formula $2a = r_0 |od|$ to compute the long axis of the ellipse:

$$r=\frac{P_{od}}{D_{od}}=\frac{P_{od}\,|_1+P_{od}\,|_2+\ldots+P_{od}\,|_k}{kD_{od}\,|_1}$$
$$=\frac{1}{k}\left(\frac{1}{\cos\theta_1}+\frac{1}{\cos\theta_2}+\ldots+\frac{1}{\cos\theta_k}\right)=\frac{1}{k}\sum_1^k\frac{1}{\cos\theta_i} \tag{13}$$

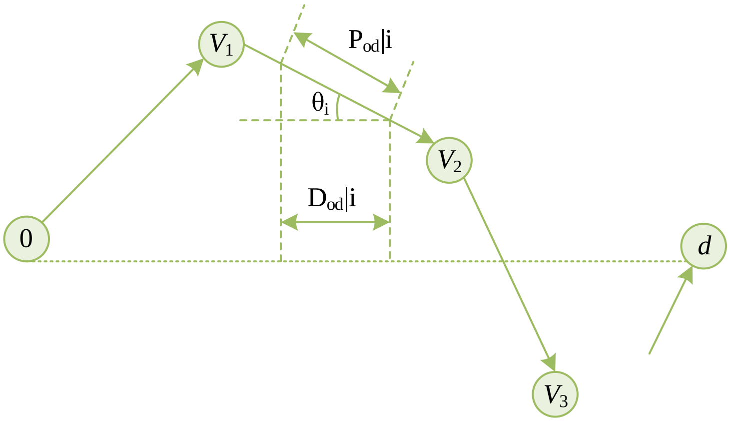


Figure 4: r value Diagram

In the real urban road network, it is known from the a priori knowledge that the angle θ satisfies the relation $\theta\in\left(0,\frac{\pi}{4}\right)$. In order to ensure that the required elliptic region can cover all valid nodes, and to avoid the situation that the search region is not large enough to lead to the situation that the shortest path cannot be found within the search range, this paper takes the maximum value of r, i.e., the value of r when θ_i takes the maximum value of $\frac{\pi}{4}$:

$$r_{\max}=\frac{1}{k}\sum_1^k\frac{1}{\cos\theta_i}=\sqrt{2} \tag{14}$$

If the starting node o is farther away from the terminating node d, then k obtains a larger value. At this time, the value of r is considered to obey the normal distribution, i.e. $N(\mu,\sigma^2/k)$, and the formulas for μ and σ^2 are as follows:

$$\mu=E\left(\frac{1}{\cos\theta}\right)\approx 1.122 \tag{15}$$

$$\sigma^2=D\left(\frac{1}{\cos\theta}\right)\approx 0.01412 \tag{16}$$

From the above analysis, it can be seen that when the value of k is increasing, the value of r is closer to the normal distribution.

3) Elliptic algorithm process

(1) Initialize the algorithm. Input the start node o and the end node d of the shortest path to be sought, and calculate the distance between the two points when the coordinate values of the two points are known;

(2) Determine the search area. Take the node o and the node d entered in the first step as the two foci of the ellipse, determine the ellipse trajectory with the length of $r|od|$, i.e., $\sqrt{2}\,|od|$ as the long axis of the ellipse, and determine the inner part of the ellipse as the searching region of the algorithm.

(3) Determine whether the node is within the search range. Set v_i for the node to be judged, at this time to determine whether v_i to meet the conditions $|ov_i|+|v_i d|<2a$, if satisfied, that the search range of the existence of the node, and retain the node, turn the step 4; if not satisfied, that the search range of the existence of the node is not the existence of the node, then exclude the node.

(4) The node retained in the third step is searched using Dijkstra's algorithm, and the algorithm ends after finding the shortest path od.

3 Effectiveness of improved Dijkstra's algorithm for path optimization of toll road network

This chapter takes Guangxi Expressway as a typical case, builds the topology of its road network, and analyzes the road network by combining the relevant static indexes and dynamic indexes of complex network theory.

3.1 Analysis of the effect of constructing the network structure of the toll road network

3.1.1 Road network construction

(1) Selection of the network modeling method

At present, the techniques widely used in modeling of complicated networks in transportation include Space-R, Space-L, and Space-P. Of these, the method of Space-P, by developing the network of highways, better represents the connections between regions and makes it easier to analyze correlations between cities. Therefore, this research employs Space-P technique to design the topology of Guangxi Highway Network.

(2) Topology construction

In this paper, the Space-P approach is adopted in order to establish a topology model for the expressway network of Guangxi province. The cities and counties where there are toll station points are taken as nodes, and the relationships between the nodes form the edges. Using complex network theory, a graph can be generated with 70 nodes and 456 edges.

3.1.2 Static characterization of the road network

(1) Node degree distribution

The i-th degree of a node in a network graph means the number of edges that link to the given node, whereas the average degree of the network is the average value of the degrees of the nodes in the network. Finding out the degree distribution of the nodes assists in finding out how interconnectivity behavior occurs in the network graph. The degree distribution of the Guangxi Highway Network is depicted in Figure 5 below. The average degree of a node in the case of the Guangxi Highway Network is 14.5014, which means an average of 15 cities/counties are connected to each other in the network.

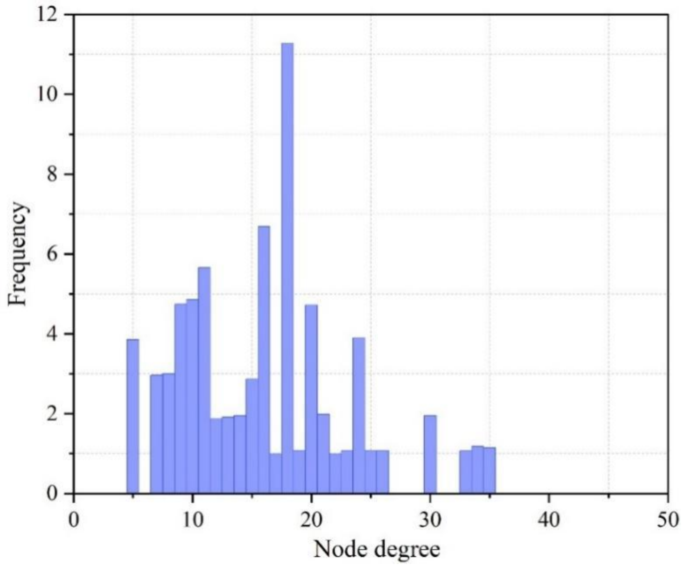


Figure 5: Distribution of Network Degree Values in Guangxi Expressway Network

(2) Clustering coefficient

Clustering coefficient of the node in a complex network is used to measure the local connection of nodes. Higher the value of the clustering coefficient, the higher level of connectivity among cities and counties will be there in the network. Clustering coefficients of highway network nodes in Guangxi have been estimated and the outcomes are presented in Figure 6. The measured value of the coefficient, C = 0.7931, suggests that cities and counties in Guangxi are tightly connected.

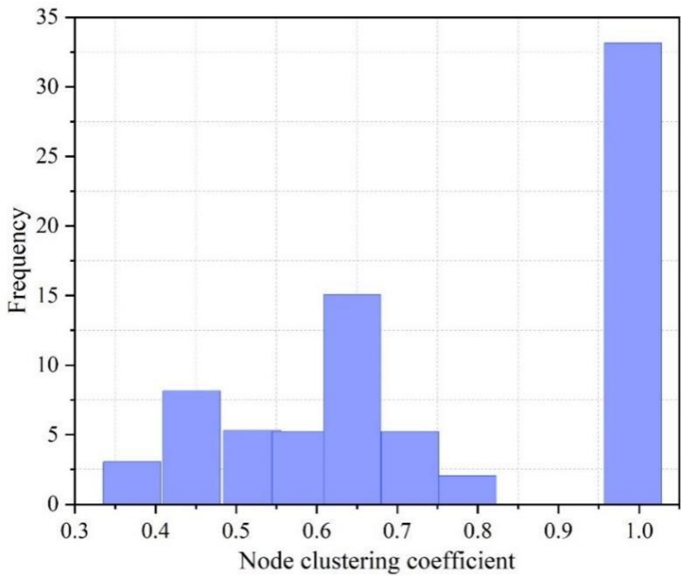


Figure 6: Distribution of node clustering coefficients

3.1.3 Road network centrality analysis

(1) Proximity centrality

The results of the distribution of node proximity centrality are shown in Figure 7. This paper analyzes the distribution of node proximity centrality of road network in more than 60 regions of Guangxi and finds that the proximity value of each node of the highway network of Guangxi Zhuang Autonomous Region ranges from 0.3191 to 0.6387, with a larger proximity and a better connectivity efficiency of the nodes.

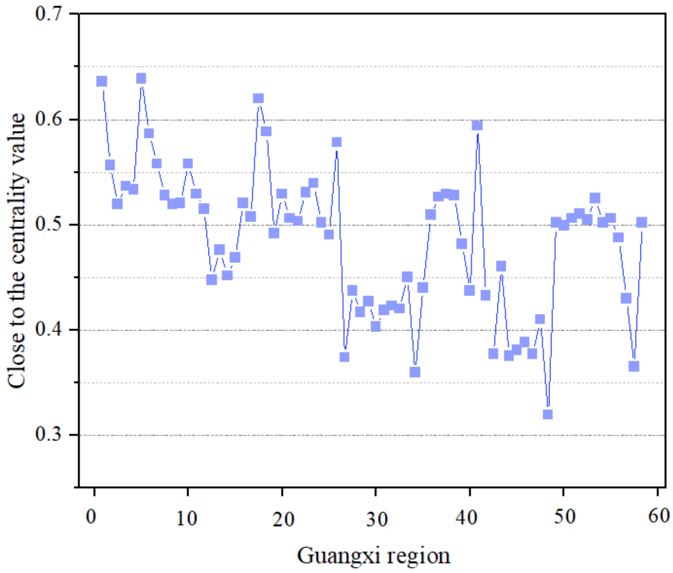


Figure 7: The nodes are distributed in a manner that is close to centrality

(2) Mediated Centrality

The results of the distribution of node intermediary centrality are shown in Figure 8. It can be seen that the distribution of intermediary centrality of the road network is uneven, reflecting the large difference in the importance of the road network connecting the cities. Points with larger values of intermediary centrality are mostly distributed in the south-central part of Guangxi and the surrounding areas of Nanning City, including the cities of Nanning, Baise and Liuzhou, while the intermediary centrality values of the nodes in the northern region of Guangxi are smaller. Among them, the mediated centrality value of Nanning city region, for example, is 0.2286, with a significant transportation hub status.

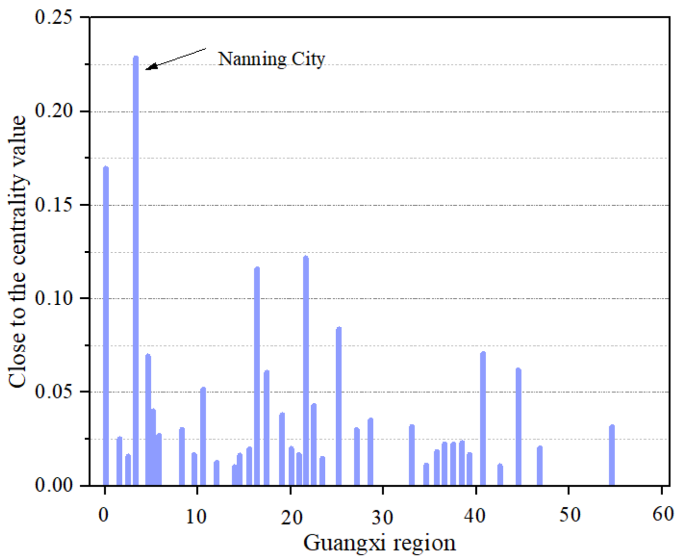


Figure 8: Results of the distribution of node intermediary centrality

3.1.4 Dynamic characterization of the road network

This paper chooses global network efficiency and the size of the largest connected subgraph as indicators of robustness.

(1) Failure analysis of a single node in the network

In this sub-section, one by one node is deleted from the network, keeping the other 60 nodes intact to analyze the effect of each node on the global efficiency of the network. After deleting each node, the global efficiency of the network is calculated again. The variation in global efficiency after deleting one node at a time can be seen in Figure 9. The deletion of the 47th node gives the highest value of global efficiency for the Guangxi highway network, which is 0.5584. On the other hand, the deletions of the 1st and 5th nodes give the lowest value of global efficiency, which are 0.5331 and 0.5276, respectively. And the ability of the traffic transit is poor; by contrast, Nanning, as the capital city of Guangxi, not only has a bypass highway, but also has a high speed road. In contrast, Nanning, as the capital city of Guangxi, not only has a bypass highway, but also is an important node where several expressways converge, and is a traffic transit center in the central and western regions of Guangxi.

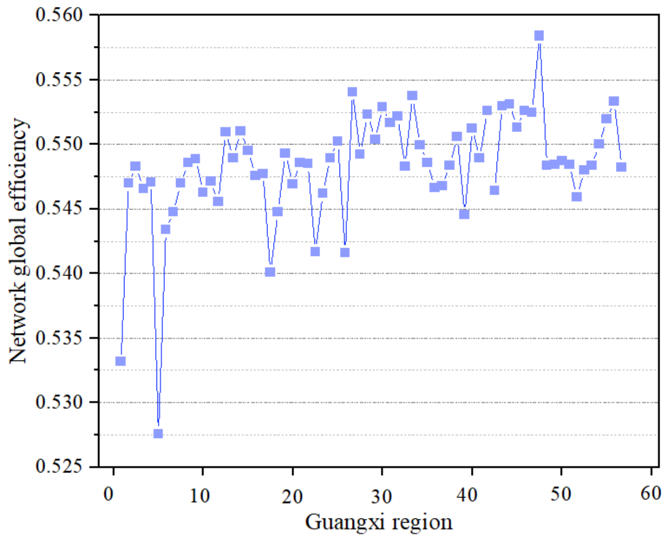


Figure 9: Global efficiency change with a single node failure

(2) Analysis of Random and Deliberate Attacks

In this section, network robustness is measured based on the relative change in the size of the largest connected component when subjected to random and deliberate attacks. For the deliberate attack case study, nodes are ranked according to their importance and then removed accordingly using three different criteria for evaluation, namely node degree, node betweenness centrality, and node efficiency. The results of these experiments in each of the attack types have been summarized in Figure 10. Based on the rate at which the relative change in the size of the largest connected component decreases, the ranking of these attacks is as follows: node-degree-based attack < node-betweenness-centrality-based attack < node-efficiency-based attack. Hence, the attack that is the most effective is the one based on node betweenness centrality because it causes the greatest change in the earliest time. Therefore, the most robust way of ranking nodes is based on betweenness centrality.

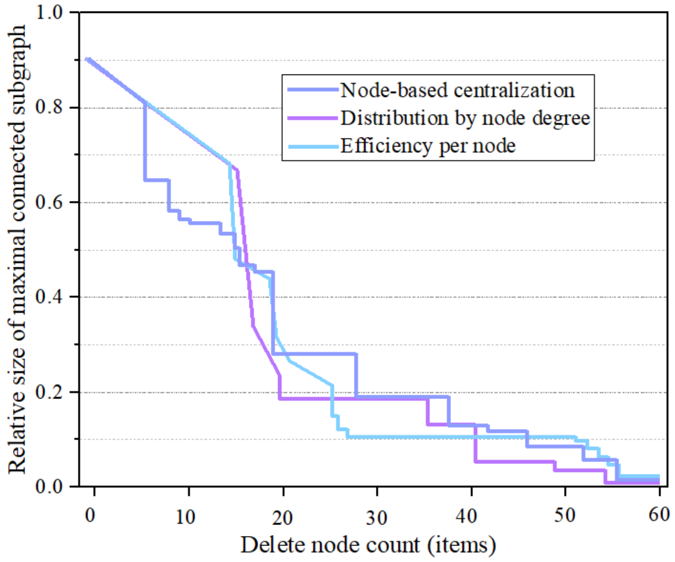


Figure 10: final result of different attack mode

3.2 Analysis of the computational effect of the improved algorithm in the shortest path of highway

3.2.1 Analysis of the statistical pattern of the shortest path ratio

Here, the system created to compute the shortest path can be applied to perform statistical analysis on the shortest-path ratio distribution in the urban road network based on the attained experimental data. The validity of the theoretical assumptions regarding the shortest-path ratio that were discussed earlier should be tested on a specific road network and the appropriate values of the statistical parameters are determined. Altogether, three hundred pairs of origin-destination sites have been chosen at random. With Dijkstra-based shortest-path query system, the following data are collected on every pair: the Euclidean distance |sd|, the number of nodes on the shortest path, n, and the shortest-path ratio, r. Lastly, the shortest-path ratio and the number of nodes on the shortest path are calculated with the help of Dijkstra algorithm.

A correlation between the shortest-path ratio and the number of nodes in the shortest path is depicted in Fig. 11 shown in the figure below. Additionally, Fig. 12 represents the variation in the distribution of the shortest-path ratio with the variation in the Euclidean distance between the two nodes. Both figures show that at small distances between origins and destinations, the shortest path passes through less nodes and the distribution of the shortest-path ratio has a greater variance. But as the distance increases, the number of nodes that are involved in the shortest path becomes more, and the standard deviation of the shortest-path ratio is reduced. And the value is supposed to be lying on either side of the theoretical value of 1.228. Also, the Jinan road network statistics parameter r0 is 1.3614 (95%) using the 300 shortest path ratio data.

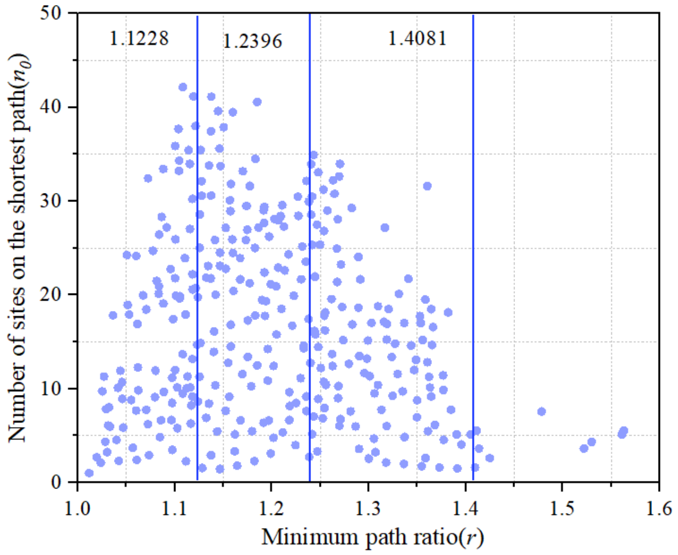


Figure 11: r-n_0 scatter plot in urban road network

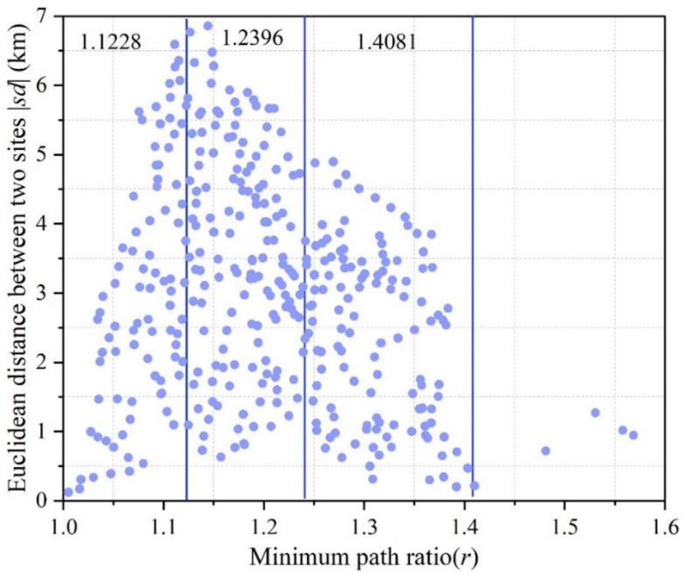


Figure 12: r-|sd| scatter plot in urban road network

3.2.2 Elliptic algorithm vs. Dikstra's algorithm

In this part, a comparative analysis is carried out between the elliptic algorithm and Dijkstra's algorithm through analyzing the number of stations covered and the time consumed to find the shortest path, based on the experimental data obtained from the shortest-path query system. In the urban road network, 100 site pairs are chosen randomly, and the shortest path is calculated by using the Dijkstra algorithm and elliptic algorithm. Figure 13 shows the |sd|−n diagram of the urban road network. The number of stations covered by an algorithm directly represents the scope of the searching process of the algorithm. From the figure above, it can be observed that the greater the distance between the starting point and destination in Euclidean space, the smaller the number of stations traversed by the elliptic algorithm compared with the Dijkstra algorithm, which means that the elliptic algorithm consumes less time. However, for the same Euclidean distance, the number of stations traversed by the elliptic algorithm is highly unstable. The reason is mainly that the urban road network has a more complicated topology, there is an uneven distribution of sites, and the number of stations searched has high fluctuation in case of close proximity between the starting point and destination.

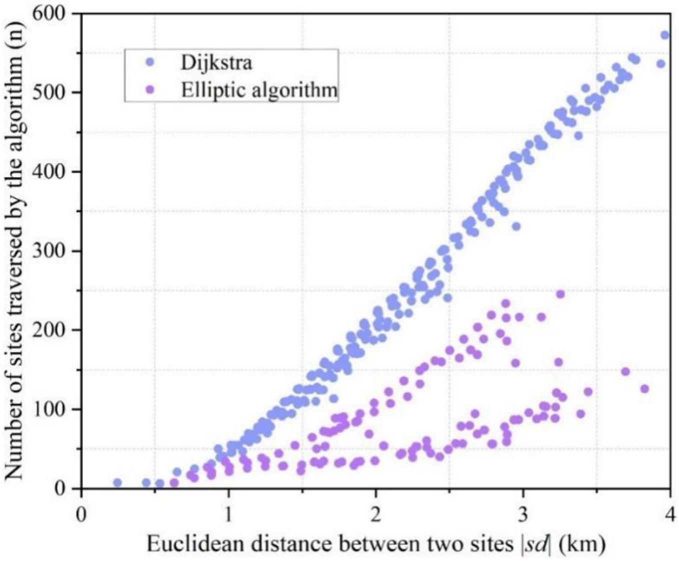


Figure 13: Scatter plot of |sd|-n in urban road network

The graph depicting the shortest path query times with Dijkstra's algorithm and the elliptic algorithm is presented in Figure 14. Such a variation in query time is due to the manner in which the number of stations that the algorithm traverses is distributed, and it is greater when the Euclidean distance between the starting point and endpoint grows. It can be explained by the fact that Dijkstra's algorithm complexity in urban roads is O(n log n) when it comes to time.

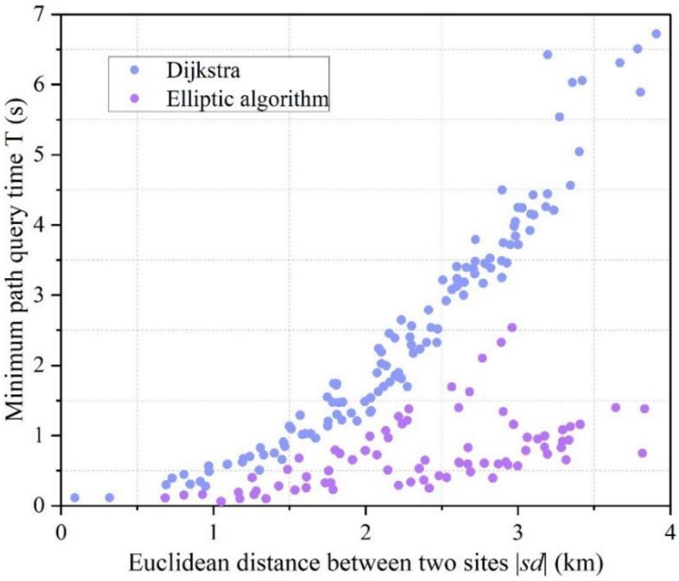


Figure 14: Scatter plot of |sd|-T in urban road network

To improve the understanding of the real-world performance of the elliptic algorithm and Dijkstra's algorithm, the 100 samples are divided into four subgroups according to the Euclidean distance between station pairs at a step of 1.0 km. The average Euclidean distance, average number of stations visited, and average shortest-path query time for each subgroup are calculated. The dependence of average distance and average number of stations and the dependence of average distance and average query time in the city street network are illustrated in Figures 15 and 16, respectively. The number of stations visited by the two algorithms is roughly approximated using a quadratic fit. This analysis reveals the pattern of shortest-path query time depending on the increase of the Euclidean distance between stations and confirms that the elliptic algorithm is superior to Dijkstra's algorithm in terms of time complexity.

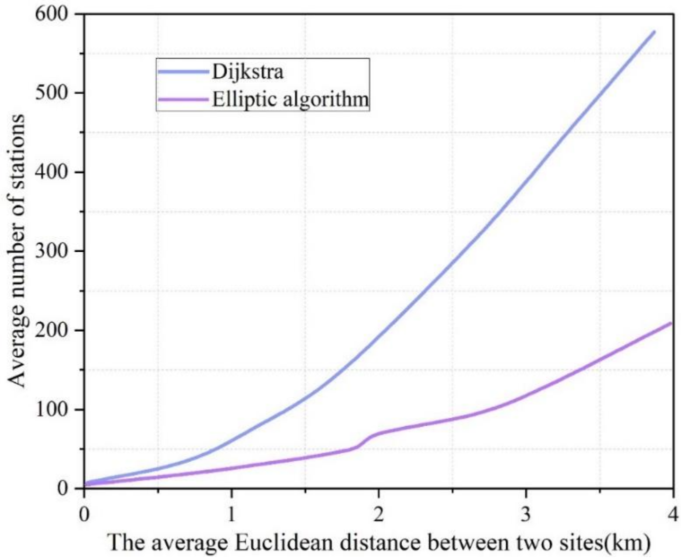


Figure 15: Relationship between average distance and average number of stations

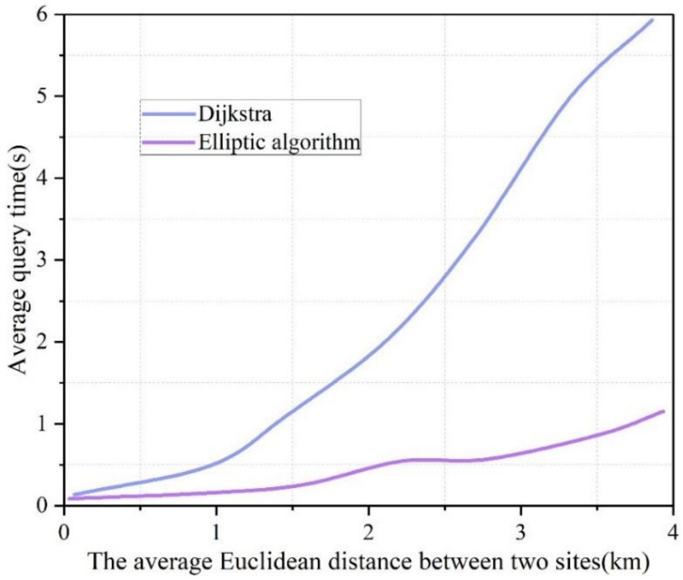


Figure 16: Relationship between average distance and average query time

4 Analysis of the effect of operational efficiency improvement of toll road network based on DEA

In order to evaluate the operational efficiency of the toll roads and the spatial pattern of their operational efficiency, this paper uses the DEA method to analyze the operational efficiency of toll roads in 29 provinces from 2020 to 2024. On this basis, this paper will further explore the characteristics of time and space differences through Theil Index and Moran's I Index.

4.1 DEA-based cost efficiency model for toll road management and maintenance

4.1.1 Basic DEA theory

Data Envelopment Analysis (DEA) is an efficiency evaluation method. The earliest CCR model is built on the concept of single-input, single-output efficiency and provides a solution framework for evaluating multi-objective problems. Assuming that a production process can be viewed as an activity in which each decision-making unit (DMU) uses a certain quantity and type of production factors to generate a certain quantity of output, let there be a set of $DMU_j\left(j=1,2,\cdots,n\right)$; each DMU has m kinds of inputs $x_i(i=1,2,\cdots,m)$, input weights $v_i(i=1,2,\cdots,m)$; each DMU has q kinds of outputs $y_r(r=1,2,\cdots,q)$, output weights $u_r(r=1,2,\cdots,q)$; and the current measurement is DMU_k. The linear programming model from the input perspective can be expressed as:

$$\begin{cases} \max \dfrac{\sum_{r=1}^{q} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}} \\ s.t.\ \dfrac{\sum_{r=1}^{q} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \\ v \geq 0; u \geq 0 \end{cases} \tag{17}$$

It should be noted that the CCR model assumes constant returns to scale. Nevertheless, real-life processes are not always operating at an optimal scale. This means that the technical efficiency calculated using the CCR model includes scale efficiency and, thus, is often called comprehensive efficiency (TE). Based on these assumptions and after many years of further study, scientists have proposed a series of DEA models. If one assumes variable returns to scale, then the BCC model will be constructed. The efficiency calculated by the BCC model does not depend on scale efficiency and is called pure technical efficiency (PTE). BCC model is based on the CCR dyadic model with the addition of constraints $\sum_{j=1}^{n}\lambda_j=1\,(\lambda\geq 0)$, and λ denotes the linear combination coefficients of DMUs.

4.1.2 Technical routes

This paper evaluates the management and maintenance efficiency of highways in different regions, which is actually a study on whether the inputs of operation and management costs and maintenance costs can be effectively matched with the toll revenues. Firstly, we collect and review the statistical bulletin of toll roads issued by the provincial and municipal transportation authorities, and organize the data related to expressways; secondly, we select single-kilometer toll, operation and management cost, and maintenance cost as the outputs and inputs, and input the relevant parameters by using the DEAP2.1 software, and compute the results of the CCR and BCC models under the input perspective, respectively; Lastly, we assess the efficiency gains achieved by the relevant provinces and municipalities, and then recommend approaches for improving the non-DEA effective decision making units based on Data Envelopment Analysis (DEA). After that, we analyze and make further recommendations regarding non-DEA effective decision-making units and provide the technical steps of the DEA based cost efficiency model for managing toll roads, as depicted in Figure 17 below.

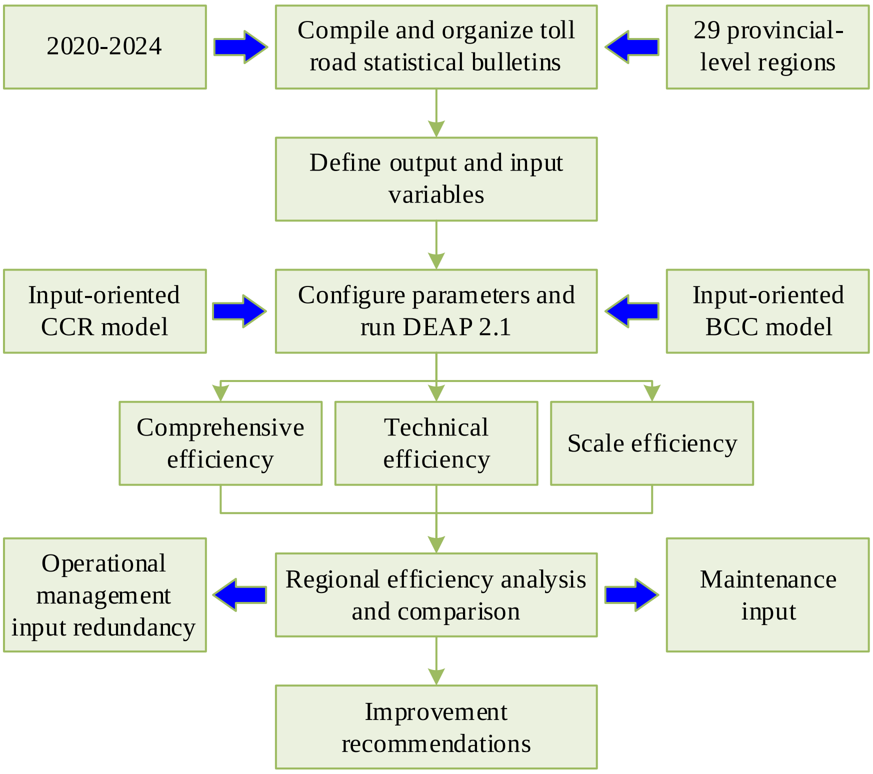


Figure 17: Technical Roadmap of Model Based on DEA

4.2 Measurement of Operational Efficiency of Toll Roads

The purpose of this segment is to determine the efficiency performance of toll roads within 29 provinces located in the eastern, central, and western parts of China from 2020 to 2024.

4.2.1 Trends in operational efficiency and characteristics of spatial distribution

Based on the DEA model and toll road input-output data selected for analysis in this study, operational efficiency of toll roads in 29 provinces in China during the period of 2020–2024 is assessed, and the results are shown in Table 1. Generally, the mean value of operational efficiency of China's toll roads in regional levels is 0.7100, implying that China's regional toll road industry shows relatively high efficiency. The conclusion can be considered to agree with previous researches on operational efficiency of toll road industries.

From a regional perspective, the operational efficiency values of toll roads in eastern, central, and western regions are 0.8770, 0.7080, and 0.5586 respectively. As a result, the operational efficiency value in the eastern region is relatively higher than in other two regions. There is an evident reason for such phenomenon. This is because eastern coastal region is characterized by high traffic volumes on toll roads due to better economic foundation and geographical location. Contrary to that, central and western regions have slower economic growth rate, less population, and remoteness geographically, and therefore low traffic volume on toll roads.

In terms of inter-provincial perspective, significant differences still exist in the operational efficiency of toll roads in various provinces. The top five provinces based on toll road operational efficiency include Shanghai, Anhui, Guangdong, Zhejiang, and Fujian. Furthermore, the implementation of regional policies such as Yangtze River Delta Urban Agglomeration, Urban Cluster of Middle Reaches of Yangtze River, and Pearl River Delta Urban Agglomeration has broken the existing trend of efficiency declining from east to west and initiated convergence in the distribution pattern of efficiency among provinces of east, middle, and west regions.

Table 1: Operation efficiency of toll roads in 2020-2024

Area	Province	2020	2021	2022	2023	2024	Mean
The east area	Beijing	0.5449	0.9302	0.8209	0.9456	0.9188	0.8321
	Tianjin	0.9256	0.8032	0.8335	0.8563	0.8402	0.8518
	Hebei	0.7265	0.5990	0.8068	0.8670	1.0166	0.8032
	Liaoning	0.6521	0.9811	0.6814	0.7899	0.8621	0.7933
	Shanghai	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Jiangsu	1.0000	1.0000	0.7276	1.0000	0.8751	0.9205
	Zhejiang	0.8952	1.0000	0.9027	1.0000	0.9657	0.9527
	Fujian	0.8291	0.9721	0.9982	0.9014	1.0036	0.9409
	Shandong	0.6927	0.8238	0.6071	0.7596	0.6288	0.7024
	Guangdong	0.9204	0.9446	1.0000	1.0000	0.9996	0.9729
	Mean	0.8187	0.9054	0.8378	0.9120	0.9111	0.8770
The middle area	Shanxi	0.6197	0.6409	0.7099	0.7464	0.6477	0.6729
	Jilin	0.8124	0.6924	0.8675	0.5768	0.6244	0.7147
	Heilongjiang	0.5757	0.7059	0.5961	0.7100	0.9735	0.7122
	Anhui	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Jiangxi	0.5010	0.8373	0.6261	0.8792	0.7160	0.7119
	Henan	0.6450	0.6772	0.5353	0.6714	0.7659	0.6590
	Hubei	0.4773	0.8882	0.7446	0.5426	0.5405	0.6386
	Hunan	0.3085	0.6503	0.5306	0.9153	0.3666	0.5543
	Mean	0.6175	0.7615	0.7013	0.7552	0.7043	0.7080
The west area	Inner Mongolia	0.7204	0.6701	0.4532	0.5356	0.6802	0.6119
	Guangxi	0.7524	0.8403	0.8561	0.6799	0.7804	0.7818
	Chongqing	0.6143	0.5879	0.5420	0.4736	0.7570	0.5950
	Sichuan	0.5796	0.4681	0.3425	0.3164	0.5092	0.4432
	Guizhou	0.4095	0.5686	1.0000	0.5060	0.4830	0.5934
	Yunnan	0.4576	0.4159	0.6541	0.5710	0.3871	0.4971
	Shanxi	0.7787	0.8579	0.8178	0.7034	0.7756	0.7867
	Gannan	0.5614	0.3678	0.4372	0.4430	0.6538	0.4926
	Qinghai	0.4271	0.4541	0.3970	0.4107	0.4119	0.4202
	Ningxia	0.4904	0.5292	0.4512	0.5101	0.5403	0.5042
	Xinjiang	0.3653	0.4157	0.5009	0.3151	0.4929	0.4180
	Mean	0.5597	0.5614	0.5865	0.4968	0.5883	0.5586
National average		0.6650	0.7359	0.7052	0.7122	0.7319	0.7100

The time-varying trend of the average value of toll road operational efficiency is shown in Figure 18. The operating efficiency of toll roads in China, eastern, central and western regions shows obvious fluctuating trends between 2020 and 2024, for example, between 2020 and 2021, the operating efficiency of highways in the three major regions is obviously in an upward trend, a phenomenon that may be related to the resumption of production in various regions at the end of the Xinguang epidemic in 2021. And between 2022 and 2023, the changes in each region appear different trends, this phenomenon may be caused by the changes in the mileage of toll roads.

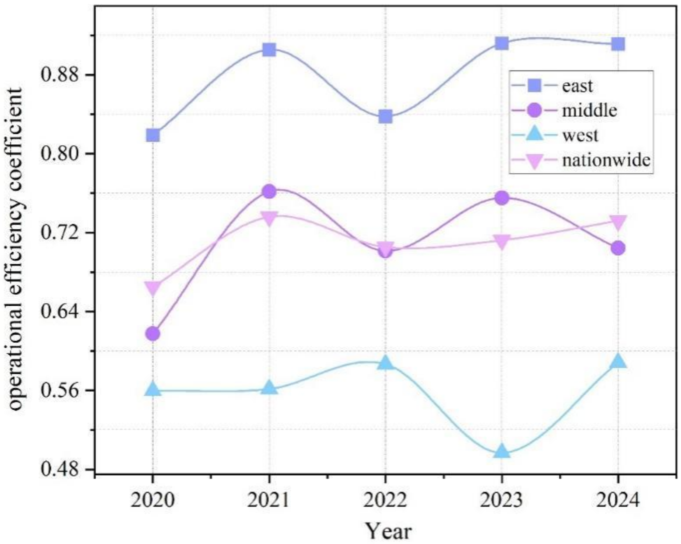


Figure 18: Time-varying Trend of Average Operating Efficiency of Toll Roads

4.2.2 Regional relevance of operational efficiency

For investigating the spatial correlation of toll road operational efficiency, exploratory spatial data analysis was utilized to calculate the global Moran’s I index of operational efficiency of China’s toll roads between 2020 and 2024. As shown in Table 2 below, the values of all Morans’I are positive and statistically significant at the 1% significance level, indicating a statistically significant positive spatial correlation in toll road operational efficiency. In other words, there exist clear spatial clusters, which mean that the operational efficiency of toll roads in one province can affect the operation efficiency of neighboring provinces’ toll roads, because of which toll roads in adjacent regions can complement each other due to improvement in highway transportation and factors mobility, resulting in a strong spatial agglomeration phenomenon. Transport facilities create agglomeration and diffusion effects on economic development of local as well as nearby regions, thus promoting economic activities and mobility of factors in nearby regions. With an increase in traffic flow, operational efficiency of toll roads improves.

Table 2: Moran's I test for the operational efficiency of China's toll roads

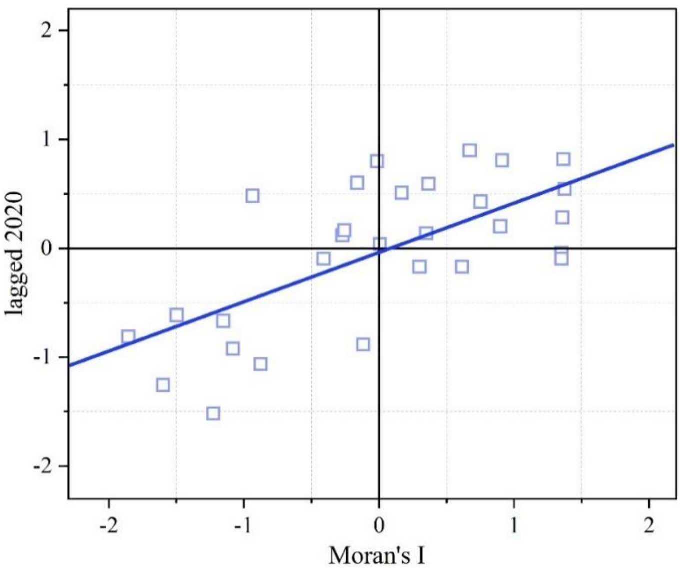
Year	Moran’s I	*Z*	*P*
2020	0.4759	3.9321	0.0000
2021	0.6370	5.1263	0.0000
2022	0.5912	4.8847	0.0000
2023	0.5231	4.2754	0.0000
2024	0.5669	4.5903	0.0000

Figure 3 reveals the local spatial clustering for operational efficiency of toll roads in China, while Figure 19 shows the local Moran's I index for operational efficiency of toll roads, where panel (a) and panel (b) are for the year 2020 and 2024, respectively. It can be noticed from these results that most of the provinces fall under the first quadrant, which means high–high agglomeration, while a few other provinces fall under the third quadrant, depicting high-low agglomeration. High-high agglomeration is seen in those provinces that are mostly situated in the east and central parts of China, such as Shanghai, Jiangsu, Zhejiang, Fujian, and Guangdong, while high-low agglomeration occurs in Beijing, Liaoning, and Hebei.

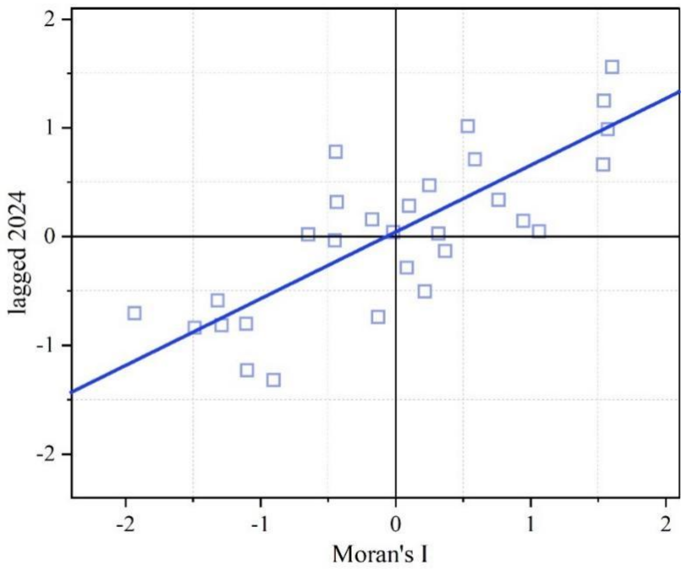
The Moran's I value for toll-road operational efficiency ranges between 0.4759 and 0.5669 from 2020 to 2024, implying an increasing agglomeration effect on toll-road operational efficiency. Simultaneously, the area of "high-high" agglomeration moves from Jiangsu Province and Zhejiang Province to Jiangsu Province, Zhejiang Province, Shanghai Municipality, Shandong Province, and Anhui Province. The joint increase in "high-high" agglomeration areas and decrease in "high-low" agglomeration provinces show that toll-road operational efficiency in a specific province depends on the efficiency of adjacent provinces. Toll-road operational efficiency presents a high level of spatial correlation due to economic and logistics development.

Table 3: China toll road operation efficiency local spatial agglomeration

The rallying point	2020 year	2024 year
High-High	Beijing, Shanghai, Jiangsu, Jiangxi, Anhui, Henan, Hubei, Chongqing, Guizhou, Fujian	Shanghai, Jiangsu, Zhejiang, Anhui, Shandong, Henan, Hubei, Hunan, Guangdong, Chongqing, Fujian
High-Low	Tianjin, Guangdong, Yunnan, Shaanxi	Tianjin, Hebei, Sichuan, Shaanxi
Low to low	Shanxi, Neimenggu, Liaoning, Gulin, Heilongjiang, Sichuan, Gansu, Qinghai, Ningxia, Xinjiang	Inner Mongolia, Liaoning, Gulin, Heilongjiang, Gansu, Qinghai, Ningxia, Xinjiang
Low to High	Hebei, Jiangxi, Shandong, Hunan, Guangxi	Beijing, Shanxi, Jiangxi, Guangxi, Guizhou, Yunnan



(a) 2020



(b) 2024

Figure 19: Local Moran's I Index of Toll Road Operation Efficiency

5 Conclusion

The present paper examines the network construction effect and the shortest path optimization effects of the Guangxi toll highway network with the help of an enhanced Dijkstra algorithm. Based on this, it also assesses the effectiveness of the toll highway network in 29 provinces of China in terms of its operational efficiency with the help of the Data Envelopment Analysis (DEA) methodology in the year 2020-2024.

The results of the computation of the enhanced Dijkstra algorithm indicate that, as shown in the created composite network of the Guangxi highway system, nodes with great betweenness centrality are mostly found in southern central Guangxi and surrounding Nanning, such as Nanning, Baise, and Liuzhou, while nodes located in the northern part of Guangxi have significantly lower betweenness centrality values. In addition, the attack mode on the basis of node betweenness centrality is the first to lead to a failure in highway-network connectivity and finally leads to the ranking of node importance according to betweenness centrality, which is the most effective attack mode.

The general performance of the regional toll road industry in China is comparatively good with a mean of 0.7100. The eastern and central regions have an average value of 0.8770 and 0.7080 respectively, which are higher than the national average, while the average value of the western region is 0.5586 indicating a relatively low level of efficiency. Furthermore, the operating efficiency of toll roads in one province is also highly dependent on the efficiency of toll roads in adjacent provinces and the spatial pattern of China toll roads demonstrates a slow increase in the quantity of high-high agglomerations areas and a steady decrease in the amount of high-low agglomerations provinces.

It should be noted that the shortest paths studied in this paper are all based on the length of the road as the weights of the paths, with the accelerating pace of life, people's concept of time will be enhanced, and they hope to get the shortest paths with time as the weights. Therefore, this paper suggests that future research can be carried out in the area of time-dependent shortest path algorithms.

About the Author

Mengyao Liu (1991-), female, a native of Harbin, Heilongjiang Province, Ph.D. Candidate in Transportation Planning and Management, School of Transportation Engineering, Chang'an University; Lecturer, School of Economics and Management, Shaanxi College of Communications. Her research focuses on Toll Roads, Educational Management. Ting Li (1977-), female, a native of Shangluo, Shaanxi Province, Assistant Research Fellow of Chang'an University. Her research focuses on Urban Transportation Planning.

Hongwei Sun (1993-), male, native of Zhangjiakou, Hebei Province, China. Research interests: computer vision, pattern recognition.

References

[1] Hensher, D. A. (2018). Toll roads–a view after 25 years. Transport Reviews, 38(1), 1-5.

[2] Rohman, M. A., Doloi, H., & Heywood, C. A. (2017). Success criteria of toll road projects from a community societal perspective. Built Environment Project and Asset Management, 7(1), 32-44.

[3] Rahayu, L., & Kipuw, D. M. (2020). The correlation between toll road development and the improvement of local economy (case study: the Soroja Toll road). International journal of sustainable transportation technology, 3(1), 26-36.

[4] Xu, Y., Fan, J., & Xu, H. (2021). Study on the operation efficiency of toll roads in China from the perspective of scale economy. Journal of Advanced Transportation, 2021(1), 8830521.

[5] Akram, M., Shareef, A., & Al-Kenani, A. N. (2024). Pythagorean fuzzy incidence graphs with application in one-way toll road network. Granular Computing, 9(2), 39.

[6] Rahaviana, K. P., & Widyaningsih, N. (2024). The Effect of Service Quality and Toll Price Rates on Toll Road User Satisfaction (Case Study on Kelapa Gading–Pulo Gebang Toll Road). Jurnal Syntax Transformation, 5(04), 673-687.

[7] Shengdong, M., Zhengxian, X., & Yixiang, T. (2019). Intelligent traffic control system based on cloud computing and big data mining. IEEE Transactions on Industrial Informatics, 15(12), 6583-6592.

[8] Al-Sakran, H. O. (2015). Intelligent traffic information system based on integration of Internet of Things and Agent technology. International Journal of Advanced Computer Science and Applications (IJACSA), 6(2), 37-43.

[9] Chang, H. H., & Yoon, B. J. (2018). High-Speed Data-Driven Methodology for Real-Time Traffic Flow Predictions: Practical Applications of ITS. Journal of Advanced Transportation, 2018(1), 5728042.

[10] Park, S. H., Kim, S. M., & Ha, Y. G. (2016). Highway traffic accident prediction using VDS big data analysis. The Journal of Supercomputing, 72(7), 2815-2831.

[11] Ghaffariyan, M. R., Stampfer, K., Sessions, J., Durston, T., Kuehmaier, M., & Kanzian,

C. H. (2010). Road network optimization using heuristic and linear programming. Journal of Forest Science, 56(3), 137-145.

[12] Bing, H., & Lai, L. (2022). Improvement and application of Dijkstra algorithms. Acad. J. Comput. Inf. Sci, 5(5), 97-102.

[13] Alshammrei, S., Boubaker, S., & Kolsi, L. (2022). Improved Dijkstra algorithm for mobile robot path planning and obstacle avoidance. Comput. Mater. Contin, 72(3), 5939-5954.

[14] Dhulkefl, E., Durdu, A., & Terzioğlu, H. (2020). Dijkstra algorithm using UAV path planning. Konya Journal of Engineering Sciences, 8, 92-105.