systems

MDPI

Article

A Local–Transit Percolation and Clustering-Based Method for Highway Segment Importance Ranking

Huizhe Lyu [1,2,3], Yang Li [1,2], Chenxu Liu [1,2], Zhonghao Li [1,3], Lin Xu [1,3], Wei Wang [1] and Jun Chen [1,2,4,5,*]

[1] School of Transportation, Southeast University, No. 2 Southeast University Road, Nanjing 211189, China; lyvhuizhe@seu.edu.cn (H.L.); l_yang@seu.edu.cn (Y.L.); xclcx@seu.edu.cn (C.L.); l_zh@seu.edu.cn (Z.L.); linxu@seu.edu.cn (L.X.); wwang@seu.edu.cn (W.W.)
[2] Key Laboratory of Transport Industry of Comprehensive Transportation Theory (Nanjing Modern Multimodal Transportation Laboratory), Ministry of Transport, No. 56 Baoshan Temple Road, Nanjing 211135, China
[3] National Demonstration Center for Experimental Road and Traffic Engineering Education, Southeast University, No. 2 Southeast University Road, Nanjing 211189, China
[4] Hangzhou International Urbanology Research Center, No. 2318 Yu Hang Tang Lu, Hangzhou 311121, China
[5] Zhejiang Urban Governance Studies Center, No. 2318 Yu Hang Tang Lu, Hangzhou 311121, China
* Correspondence: chenjun@seu.edu.cn

Abstract: The impact of disturbances on a transportation network varies depending on the location and characteristics of the affected highway segments. Given limited resources, it is crucial to prioritize the protection and repair of highway segments based on their importance to maintaining overall network performance during disruptions. This paper proposes a novel method for ranking the importance of highway segments, leveraging a novel local–transit percolation and clustering-based method. Initially, the highway network is constructed by Graph theory, and the k-means clustering method is applied considering each segment's transit and local traffic flows. Subsequently, a local–transit percolation model is constructed to generate an initial ranking of segments based on the size of the second-largest clusters during the percolation phase transition. A secondary ranking is performed by refining the results from the clustering phase. Results of a control experiment show that, compared to baselines, the proposed ranking approach demonstrates a significantly improved ability to sustain network demand and connectivity when high-ranked segments are moved. The model uncertainty analysis was conducted by adding noise to the gantry records, and the experiments demonstrated that the model exhibits robustness under noisy conditions. These findings highlight the effectiveness and superiority of the proposed method.

Keywords: highway; resilience; percolation theory; clustering; segment importance ranking

check for updates

Academic Editors: Katarzyna Antosz, Jose Machado, Erika Ottaviano, Pierluigi Rea, Camelia Claudia Avram and Vijaya Kumar Manupati

Received: 17 November 2024
Revised: 27 December 2024
Accepted: 29 December 2024
Published: 2 January 2025

Citation: Lyu, H.; Li, Y.; Liu, C.; Li, Z.; Xu, L.; Wang, W.; Chen, J. A Local–Transit Percolation and Clustering-Based Method for Highway Segment Importance Ranking. *Systems* **2025**, *13*, 28. https://doi.org/10.3390/systems13010028

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Highways are vital for transporting passengers and cargo both within and between cities. Highways have broad coverage and allow for fast travel, contributing to high traffic efficiency in the transportation network. However, this also makes them vulnerable to disturbing events such as natural disasters, traffic accidents, and road maintenance. These events typically reduce highway capacity or even disrupt the highway's operation. Highways need to maintain connectivity and ensure travel reliability during disruptions.

Transportation resilience refers to the ability to maintain its original state under disruptions. The concept of resilience [1] consists of two abilities: (1) the ability to maintain functionality under disruptions and (2) the ability to recover from the disruptions. Based on the timing of the disruption event's impact, resilience can be categorized into three stages:

prevention, resistance, and recovery. At the micro level, transportation resilience focuses on the ability of transportation facilities to resist damage and recover quickly. It also considers the capability of vehicles and participants to maintain safety and travel efficiency during disturbances. These methods, however, do not apply to large-scale disruptions. At the macro level, disturbances generally cause traffic congestion and flow shifts on localized road segments, which can then lead to large-scale failures in the highway network, known as cascading failures. The impact of each highway segment on the network depends on its location and connections. Given limited resources, it is crucial to prioritize protecting and repairing the segments that are of high importance on the highway network. Therefore, ranking the importance of highway segments is necessary to enhance network resilience.

Segment importance is typically assessed by examining performance changes in the traffic network pre- and post-disturbance. This assessment can be approached through simulation-based methods and complex network-based methods. Simulation-based methods involve traffic flow reassignment pre- and post-disturbance using traffic simulations. It analyzes changes in traffic flow across segments and calculates metrics such as network efficiency and congestion levels. Traffic flow simulation tools include microscopic simulation software like VISSIM, AIMSUN, SUMO, and MITSIMLab, along with macroscopic models and software such as METANET, CTM, and Visum [2]. Zhou and Wang [3] considered segment vulnerability and potential, using the CTM-U model to simulate various traffic events. They identified the segments that most affected network performance under different scenarios. Almotahari et al. [4] considered accessibility, redundancy, and the importance of carrying traffic demand, proposing the Link Criticality Index metric. Kumar et al. [5] defined critical segments based on the impact of removing each segment on network efficiency. Network efficiency is inversely proportional to the average travel time between all origin–destination pairs. This method handles network disconnections but requires multiple traffic assignments. Traffic flow simulation could accurately demonstrate the impact of disturbances on traffic flow. However, when applied to large-scale traffic networks, it involves high modeling complexity and significant computational demands.

Compared to traffic simulation, complex network-based methods represent traffic infrastructure as graphs composed of vertices and edges, offering a more straightforward and computationally efficient analysis approach. Traditional complex network theory often evaluates topological performance under disturbance conditions, including metrics such as degree, centrality, betweenness, and shortest path lengths. However, as for transportation networks, not only the topological performance but also functional performance, such as vehicle speed and traveler demand, need to be considered. Percolation theory addresses these aspects by describing changes in connected clusters as nodes or edges are removed [6]. This leads to a phase transition process in the network. Percolation theory is widely used in assessing network resilience. Li et al. [7] calibrated traffic flow–speed–density functions under rainy conditions using highway loop detector data. By integrating flow–speed–density functions with percolation theory, they identified critical segments that should be prioritized for protection during rainfall. Liu et al. [8] regarded airline integration as a percolation process. They used critical integration time as an indicator of airline reliability. Zhou et al. [9] proposed a method for large-scale network assessment based on percolation theory. This approach is employed to analyze the stochastic dynamics of road networks under the influence of various risk factors. Network connectivity and efficiency were defined to characterize road network topology and functional performance quantitatively. OD traffic refers to the traffic traveling on the roadway network and is classified as local OD traffic and transit OD traffic. Cao et al. [10] identified that the percolation transition in a network can manifest in different forms, including first-order, second-order, or dual-phase transitions. The nature of the transition is contingent upon the strength of the dependencies between

nodes and the density of the dependency links within the network. Kryven et al. [11] explained the mechanism behind multiple discontinuous phase transitions in multilayer networks. The percolation process reflects a situation where multilayer networks might decompose. Such decomposition may extend across all layers of the network, especially when the networks experience uncertain failures. Hayasaka et al. [12] discovered that the sequential addition of edges in a specific order can trigger explosive percolation, wherein the majority of nodes in a network suddenly become interconnected. This phenomenon is particularly evident in networks where connections are not random but instead are determined by specific threshold conditions. Therefore, percolation theory can effectively assess changes in the resilience of highway networks under disturbances. However, there are still some issues with the methods mentioned above:

(1) Highway travel demands are not classified. Highways are crucial for long-distance transport, with transit traffic significantly impacting their performance. Current methods for ranking segment importance fail to consider how changes in transit and local traffic flow can impact the overall highway network.
(2) Existing percolation theory typically relies on a single indicator and lacks research on phase transitions considering multiple indicators. As a result, existing approaches fail to simulate percolation phase transitions based on the relative relationships between transit traffic and local traffic flow.

Solving these problems requires extensive highway data, which can be obtained from gantry records, including vehicle license plates and passage timestamps. This data helps calculate traffic conditions and demands on highways. Clustering algorithms can classify highway segments based on travel demands by grouping similar data records based on shared characteristics [13–15]. There are various clustering methods in the field of transport, such as connectivity-based clustering, center of mass-based clustering, and density-based clustering algorithms. In traffic applications of clustering methods, Song et al. [16] performed hierarchical clustering on 120,000 road sections in Beijing based on factors like points of interest (POI), achieving successful geographical classification. Kosanin et al. [17] analyzed urban traffic accident data using parameter-free clustering to identify high-risk road sections. Deng et al. [18] investigated dependencies within urban road networks, identifying and clustering failure-dependent clusters to enhance network reliability. Similarly, Vivek et al. [19] developed an unsupervised machine-learning algorithm based on density clustering. They quantified the risk and improved the stability of urban road networks. In addition, Cogonit et al. [20] developed a traffic state taxonomy by clustering similarity metrics to verify the spatial distribution characteristics of urban traffic clusters. Clustering algorithms are widely used in transportation research and applied to traffic pattern recognition, road segmentation, and traffic safety analysis. Based on the origin and destination of travel on highway routes, clustering algorithms classify the demand for travel.

This paper focuses on comparing transit traffic and in-city traffic using highway gantry data. A local–transit percolation model is proposed by combining clustering algorithms, which helps rank the importance of highway network segments at a hierarchical level. Firstly, highway segments are clustered based on the ratio of transit flow to intrastate traffic. This helps identify routes that primarily serve either transit or intrastate traffic. Secondly, based on the clustering results, a percolation phase transition simulation is performed on the motorway network. This simulation helps classify internal travel demand, accurately assessing the impact of disturbances on both transit and internal traffic flow. The implementation of a local–transit percolation phase transition improves the accuracy of ranking road section importance.

The remaining parts of the paper are structured as follows: Section 2 introduces the specific theory of the proposed local–transit percolation and clustering-based method for highway segment importance ranking, Section 3 takes some highways in Hebei Province, China as an example for a case study, Section 4 discusses the effectiveness of the proposed method and Section 5 provides a conclusion and outlook.

2. Methods

2.1. Methodological Framework

This section provides details about the proposed local–transit percolation and clustering-based method for highway segment importance ranking. The methodological framework comprises four main components: the data processing module, the link clustering module, the percolation module, and the critical links ranking module. The framework is illustrated in Figure 1.

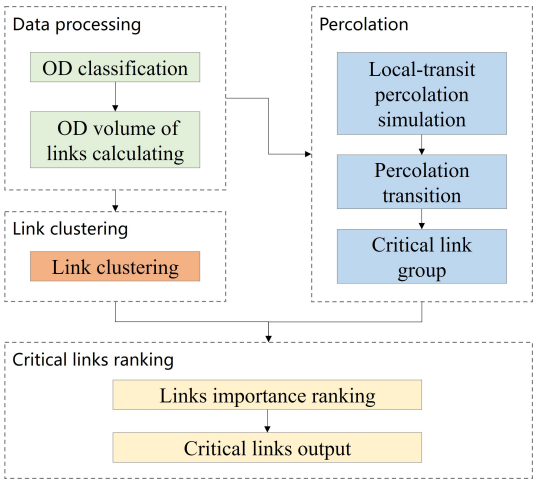


Figure 1. Framework of the proposed local–transit percolation and clustering-based method for highway segment importance ranking.

The data processing module primarily calculates each highway link's transit and local OD traffic flow based on gantry data, which is called OD classification. The link clustering module categorizes links according to their external/internal OD traffic, distinguishing links primarily serving transit versus local traffic. In the percolation simulation module, a local–transit percolation simulation is performed using the transit and local OD traffic flow of each segment. This approach identifies the failing segments during the percolation phase transition, allowing for the initial ranking of critical links within the network. Finally, the critical link ranking module integrates clustering results with percolation simulation outputs to rank critical link importance.

2.2. Data Processing

2.2.1. Data Fields Description

Highway gantry data is utilized in the study. High-speed gantry ETC utilizes RFID technology for automatic identification of passing vehicles, and is able to read the information in the fields related to vehicle travel. It records vehicle details, such as license plate number and vehicle type, as well as trip information like timestamp and location. Each record represents a time trajectory point of a vehicle's trip on the highway. Additionally, geographic location data of gantries are used, including fields such as the city to which the gantry belongs, name, and latitude and longitude coordinates. This data enables the selection of specific gantries within the research area and links them to precise geographic locations. The main data fields are presented in Table 1.

2.2.2. Highway Network Modeling Based on Gantry Data

Traditional highway network modeling methods represent the highway network as an undirected graph $G = (V, E)$, where V represents the set of vertices in the network, typically including highway toll stations and interchanges, and E denotes the edges. This modeling approach requires processing extensive geographical information data, which is time-intensive. Additionally, it does not account for the heterogeneity of traffic flows in different directions on the same segment.

To address this limitation, the paper proposes a highway network modeling method based on gantry data. When all segments V in the highway network G are equipped with gantry detectors r, highway network modeling can be achieved by utilizing vehicle travel trajectories between adjacent gantries. This approach allows for more detailed tracking of traffic flow and travel patterns across the network G. Specifically, for any edge $e_{u,v}$, the travel time $t_{u,v}$ is determined by the average travel time of all vehicles crossing this edge within the time period T. In this study, geographic data processing is not the primary focus. Therefore, this method simplifies the modeling process and improves efficiency without the need for complex geographic information processing, as shown in Algorithm 1. However, in other application scenarios that require geographic information support, this method may need to be combined with additional geographic data processing techniques.

Algorithm 1 Highway Network Modeling Method Based on Gantry Data

Input: Grantry records of vehicles
Output: Graph file G
//Step 1. Initialization
Create an empty directed graph $G = (V, E)$.
//Step 2. Add vertices
for any gantry r **do**
  Add vertex r to graph $G : V = V \bigcup \{r\}$.
end for
//Step 3. Add edges
Group all data by license plate number and sort each group by timestamp in chronological order
for each license plate group **do**
  Extract gantry information from each consecutive record pair (u, v):
  if edge (u, v) does not exist in graph G **then**
    Establish a directed edge (u, v), with direction from u to v. Initialize the edge's average travel time $t_{u,v} = 0$ and the vehicle count $q_{u,v} = 0$;
  end if
  if edge (u, v) does not exist in graph G **then**
    pass;
  end if
  Calculate the travel time for this license plate between the two gantries by $t'_{u,v}$;
  Update the number of vehicles traveling between the two gantries by $q_{u,v} = q_{u,v} + 1$;
  Update the average travel time for vehicles traveling between the two gantries by $t_{u,v} = [(q_{u,v} - 1)t_{u,v} + t'_{u,v}]/q_{u,v}$.
end for
//Step 4. Remove non-adjacent edges
for each edge $e_{u,v}$ in the graph $G = (V, E)$ **do**
  for each edge $e_{v,x}$ originating from vertex v **do**
    if an edge $e_{u,x}$ exists in graph G **then**
      Delete edge $e_{u,x}$;
    else
      Continue;
    end if
  end for
end for
Output the network file G.

Table 1. Example of gantry data.

Field	Example
Trade ID	G2*****01
Toll Gantry ID	G2*****10
Trans Time	2023-07-07 00:00:00
Vehicle Plate	H D****11
City	Xingtai City
Gantry Name	Yunhe West Toll Station—Hebei-Shandong Border 1
LAT	36.883127
LNG	115.702484

2.2.3. Traffic Classification Statistics for Highway Segments

The impact of disturbances on highway segment flows is analyzed. Vehicles are classified as transit or local trips based on their origin and destination. The origin and destination of each trip are recorded by gantries. Local trips occur within the same city. Transit trips span across city boundaries. The classification algorithm is detailed in Algorithm 2. This process provides insights into the primary traffic-bearing functions of each segment. It also forms the basis for ranking segment importance in further analysis.

Algorithm 2 Traffic Classification Algorithm for Highway Segment Flow

Input: Grantry records of vehicles
Output: Categorized traffic volume statistics
//Step 1. Determine OD classification for each vehicle
Group all data by license plate number and sort each group in chronological order:
for each license plate group **do**
  Check the first and last records of the vehicle's trip (i.e., the origin and destination gantries and their respective cities)
  if the origin and destination are within the same city **then**
    Classify the vehicle as an local trip.
  else
    Classify the vehicle as an transit trip.
  end if
end for
//Step 2. Traffic classification calculation for gantry
for each gantry **do**
  Initialize counters for gantry transit traffic and local traffic.
  for each vehicle passing through this gantry **do**
    if the vehicle OD is classified as an local trip **then**
      Increment the gantry transit traffic by 1.
    else
      Increment the gantry local traffic by 1.
    end if
  end for
end for
//Step 3. Traffic classification calculation for segments
for each segment in graph G **do**
  The local and transit flow of a road segment are the averages of the local and transit flows from the gantries at both ends of the segment.
end for

2.3. Segment Clustering Method Based on Traffic OD

Vehicles passing through each highway segment are recorded, and cumulative traffic flow and proportions for the two types of traffic are calculated. A clustering analysis is conducted for the segments with gantry-based trips. The k-means clustering method, a widely used unsupervised learning algorithm, is applied to partition a set of unlabeled data points into k clusters. The core idea is to iteratively assign data points to cluster centers, minimizing the distance between each data point and its assigned centroid. The

objective is to achieve high similarity among data points within clusters, while maximizing the differences between points in different clusters. In the k-means clustering method, the similarity between a data point x and a centroid μ is typically measured using the Euclidean distance, defined as:

$$d(x, \mu) = \sqrt{\sum_{i=1}^{m} (x_i - \mu_i)^2} \tag{1}$$

where $x = (x_1, x_2, \ldots x_m)$ is the feature vector of the data point, $\mu = (\mu_1, \mu_2, \ldots \mu_m)$ is the feature vector of the centroid, and m is the dimensionality of the features.

In each iteration, k-means updates the position of the centroids. For each cluster C_i, the new centroid μ_i is the mean of all data points in that cluster. The formula is:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \tag{2}$$

where μ_i is the centroid of the ith cluster, $|C_i|$ is the number of data points in cluster C_i, and $x \in C_i$ represents data points belonging to cluster C_i. The core objective of k-means clustering is to minimize the sum of squared errors within clusters. By iteratively updating centroids and reassigning data points, the objective function is gradually reduced. The objective function is defined as:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 \tag{3}$$

where k is the number of clusters, C_i is the ith cluster, and $\sum_{x \in C_i} ||x - \mu_i||^2$ represents the squared Euclidean distance between the data point x and the centroid, which is equal to $[d(x, \mu)]^2$. The k-means algorithm is characterized by its simplicity, ease of implementation, and high computational speed, making it well-suited for handling large-scale datasets. In this study, highway segments are classified into two categories based on the OD locations of traffic flows. Since k-means clustering requires the user to specify the number of clusters in advance, the cluster count is set to $k = 2$. This classification of highway segments defines the primary service areas for segment traffic, distinguishing between those primarily serving local trips and those serving transit trips. It provides a rational basis for the hierarchical resilience analysis of highway segments in subsequent research.

2.4. The Local–Transit Percolation and Clustering-Based Method

Percolation theory simulates the process of network degradation by removing edges or vertices and is commonly used to analyze network connectivity. Percolation theory helps analyze the impact on network traffic flow when certain highway segments or nodes fail due to accidents, maintenance, natural disasters, or other causes.

Traditional percolation theory has a single percolation indicator. The size of the giant component (GC) is usually used to measure network connectivity. When the percolation threshold q reaches a certain value q_c , the GC suddenly fragments into multiple smaller clusters. At this point, the size of the second largest cluster (SLC) reaches its maximum, leading to a significant decline in overall network connectivity. This abrupt phenomenon is called the percolation phase transition, and q_c is known as the critical percolation threshold. Edges that fail at the percolation phase transition point q_c significantly impact the connectivity of the graph.

However, percolation theory with one percolation indicator is inadequate for simulating percolation process that consider both local and transit traffic flows. To address

this limitation, we propose a local–transit percolation model, expanding the percolation threshold to incorporate both local and transit flows. Unlike percolation theory with one percolation indicator, the local–transit percolation process has multiple percolation phase transition points. For a highway graph G with n vertices, where edge e_{ij} connects vertex i and vertex j, each edge e_{ij} has local flow q_{ij}^1 and transit flow q_{ij}^2. q_1 and q_2 are set to define the percolation thresholds for vertex removal. In addition, q_1 is the percolation threshold for local traffic flow, and q_2 represents the threshold for transit traffic flow. When $q_{ij}^1 \geq q_1$ and $q_{ij}^2 \geq q_2$, the state of edge e_{ij} is as shown in Equation (4). For each pair (q_1, q_2), segments whose local and transit flow is below the defined threshold will be removed.

$$e_i j = \begin{cases} 1, & \text{if } q_{ij}^1 \geq q_1 \text{ and } q_{ij}^2 \geq q_2 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Referring to the role of the SLC in the percolation phase transition, the local–transit percolation simulation uses the size of the SLC corresponding to each percolation threshold pair (q_1, q_2) as the primary criterion for ranking the importance of edges. A larger SLC indicates a more significant decrease in highway network connectivity under this percolation threshold pair. Edges that fail at this point are more critical to the highway network. Each percolation phase transition point (q_1, q_2) corresponds to a set of failed edges W at this threshold, as shown in Equation (5).

$$\mathbf{W} = \bigcup_I \{e_{ij} | q_{ij}^1 \in (q_{c1}^I - \Delta q_1, q_{c1}^I], q_{ij}^2 \in (q_{c2}^I - \Delta q_2, q_{c2}^I]\} \tag{5}$$

where I the percolation phase transition, q_{ij}^1 indicates the local flow on segment e_{ij}, q_{ij}^2 represents the transit flow on segment e_{ij}, q_{c1}^I is the critical percolation threshold for local flow under I, q_{c2}^I is the critical percolation threshold for transit flow under I, and Δq_1 is the unit interval of the local percolation threshold, which is the smallest unit of change in local travel. Furthermore, Δq_2 is the unit interval of the transit percolation threshold, which is the smallest unit of change in transit travel.

Through the local–transit percolation approach detailed above, we can rank the obtained set **W** based on the size of the SLC corresponding to each critical percolation threshold pair (q_1, q_2). Next, the importance of highway segments within the edge set **W**, under the same SLC, are ranked. Based on the clustering results of local and transit travel flows across all segments e_{ij}, the segments are categorized into those primarily serving local travel and those primarily serving transit travel. Consequently, the clustering results divide the percolation threshold surface into two regions: the local travel percolation zone Z_1 and the transit-priority percolation zone Z_2. If the percolation threshold pair (q_1, q_2) falls within Z_1, this indicates that the segments under this threshold are mainly dedicated to local travel flow. In this case, the magnitude of local flow q_{ij}^1 serves as the secondary ranking criterion for sorting the edges in set **W**. Similarly, when the percolation threshold pair (q_1, q_2) is within Z_2, the magnitude of transit flow q_{ij}^2 is used as the secondary ranking criterion for the same set.

In summary, the algorithm for importance ranking of road segments is as shown in Algorithm 3. This algorithm enables automatic importance ranking of highway segments based on gantry records, which can assist management in organizing road segments more efficiently. First, the edges in W are filtered according to the critical thresholds of local and transit flows; then, the importance weights of each edge at the percolation point for network connectivity are computed, and the edges are ranked accordingly. The set **W** is then divided into two priority intervals: Z_1 for local travel and Z_2 for transit flow. When the percolation threshold pair (q_1, q_2) falls within Z_1, q_{ij}^1 is used for secondary sorting; when it falls within

Z_2, q_{ij}^2 is used. The final output is the set of edges **W** ordered according to both primary and secondary criteria, identifying critical road sections for different flow conditions.

Algorithm 3 The Highway Segment Importance Ranking Algorithm by Local–Transit Percolation and Clustering-Based Method

Input: W(set of edge), q_{ij}^1(the magnitude of local flow), q_{ij}^2(the magnitude of transit flow), (q_1, q_2) (the percolation threshold pair)
Output: sorted **W**
// Step 1. Classify edges in **W** into two categories
Category 1: Edges primarily serving local traffic;
Category 2: Edges primarily serving transit traffic;
// Step 2. Determine the region for the threshold pair (q_1, q_2)
if (q_1, q_2) is in Z_1 (local-priority percolation zone) **then**
    Set secondary sorting criterion to local traffic flow size q_{ij}^1;
else if (q_1, q_2) is in Z_2 (transit-priority percolation zone) **then**
    Set secondary sorting criterion to transit traffic flow size q_{ij}^2;
end if
// Step 3. Initialize an empty list sorted **W** to store the sorted edges
for edge e_{ij} in **W** **do**
    Calculate the importance score for e_{ij} based on the secondary sorting criterion;
    Insert e_{ij} into sorted **W** in descending order of importance score;
    return sorted **W**.
end for

3. Case Study

3.1. Data Set

The data set consists of travel records from highway gantries in Hebei Province. It covers the period from 1 July to 7 July 2023, with an average of approximately 20 million records per day. The scope of this research includes four cities in Hebei Province: Shijiazhuang, Hengshui, Xingtai, and Handan, covering 437 gantries. Data preprocessing and cleaning are conducted. Each vehicle has a unique TRADE ID and license plate number. Using the TRADE ID and license plate, the spatiotemporal locations of vehicles are extracted. For the time field, the travel time records of the same vehicle are detected, and the time outliers that deviate significantly from normal driving characteristics are identified and eliminated as time error records. Finally, outliers and null values are cleaned. After data filtering, the average daily travel data in the case area is approximately five million records.

Based on the highway network modeling method in Section 2.2.2, a directed graph network of the study area is established, as shown in Figure 2.

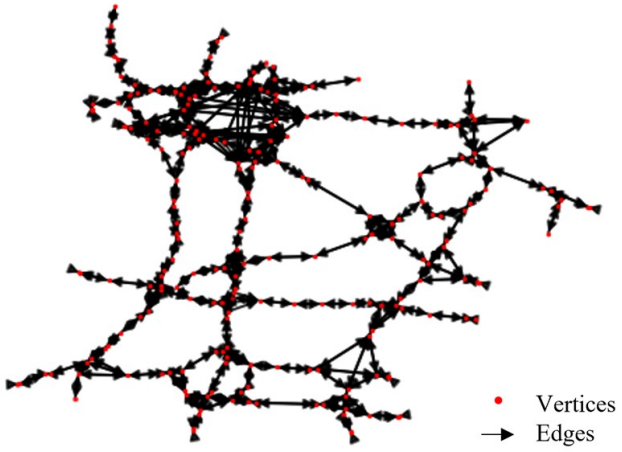


Figure 2. Directed graph of the study area.

In addition, some global topological metrics of the network, including the number of nodes, edges, and the average shortest path length, are shown in Table 2. Table 3 provides a detailed statistical analysis of the clustering coefficient, in-degree, and out-degree of the nodes. The average clustering coefficient of the study network is 0.004, and the average degree is 1.482, indicating that the network is relatively sparse. The majority of nodes are disconnected neighborhoods (clustering coefficient = 0), though a small subset of nodes may belong to locally dense regions (maximal clustering coefficient = 0.167). The distributions of in-degree and out-degree are relatively uniform, with low standard deviations, suggesting the absence of prominent hub nodes. The degree distribution is concentrated, with most nodes having degrees in (1,2).

Table 2. Vertices, Edges, and Average Shortest Path of Network.

Metric	Value
Average Shortest Path	175
Number of Edges	649
Number of Vertices	438

Table 3. Cluster Coefficient and Node Degree of Network.

	Clustering Coefficient	In-Degree	Out-Degree
mean	0.004	1.482	1.482
std	0.020	1.157	1.143
min	0.000	0.000	0.000
mean of top 25%	0.000	1.000	1.000
mean of top 50%	0.000	1.000	1.000
mean of top 75%	0.000	2.000	2.000
max	0.167	9.000	9.000

3.2. Classification Statistics of Highway Segment Flow

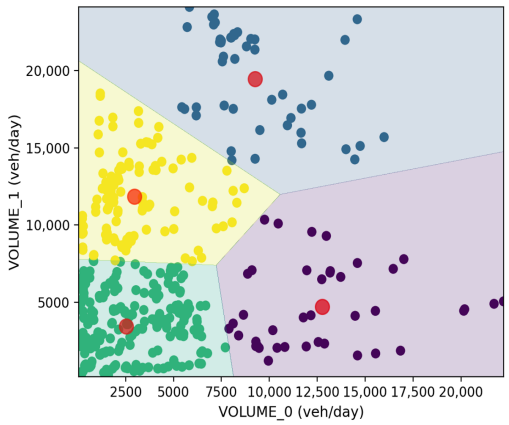
Using the method described in Section 2.2.3, highway segment flow based on 7-day gantry flow data was classified. The daily average was calculated to obtain the segment flow classification statistics for the study area. Figure 3 presents an example of the results, where $VOLUME_0$ denotes local travel volume, and $VOLUME_1$ represents transit travel volume.

	TOLL_GANTRY_ID	TIME_STAMP	VOLUME_0	VOLUME_1
0	S007513003001510010	2023-07-01	643	793
1	S007513003001620010	2023-07-01	696	882
2	S007513003001310010	2023-07-01	592	772
3	S007513003001420010	2023-07-01	613	841
4	G000413002001420010	2023-07-01	8014	22,148
...	...	...	...	...
432	G002013001001220010	2023-07-01	3378	11,807
433	G002013001001420010	2023-07-01	3681	11,178
434	G002013001001310010	2023-07-01	3758	11,320
435	G002013001001510010	2023-07-01	5835	12,432
436	G002013001001620010	2023-07-01	5953	14,352

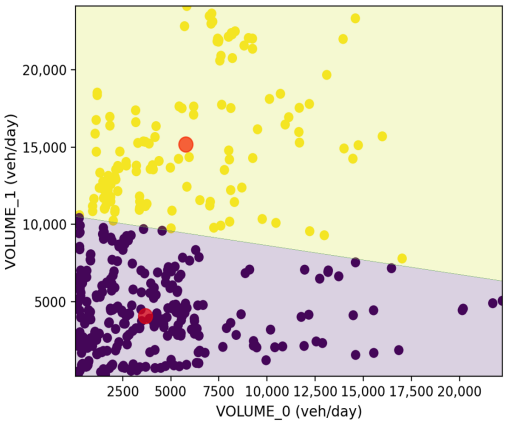
Figure 3. Gantry flow statistics in study area.

3.3. Clustering and Classification of Highway Gantries

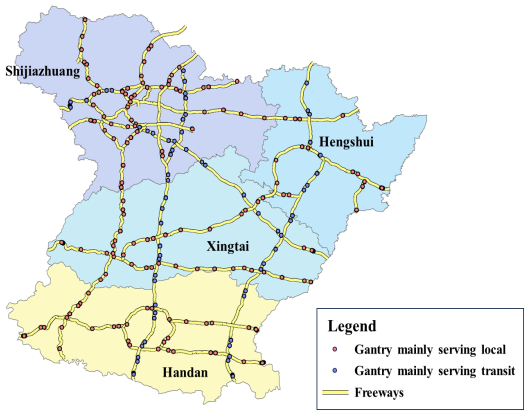
Based on the gantry records, k-means clustering on the gantries was performed according to the method in Section 2.3, using the local and transit traffic flow of each segment as input. According to the classification of gantry traffic and traffic size characteristics, the number of clusters $k = 4$ is set first, and the results are shown in Figure 4a. As can be seen from the figure, the road segments in the upper left corner of the yellow and the lower right corner of the purple area mainly serve transit and local trips, respectively. However, the blue and green areas of the road segments are not obvious in the service classification, which can be classified as mainly carrying high-traffic segments and low-traffic segments. In order to further distinguish the main service traffic type of each road segment, the number of clusters is set $k = 2$. The clustering results are shown in Figure 4b, and the gantry locations are visualized in Figure 4c. In this case, the purple shaded area shows the roadways that mainly serves local trips, and yellow shaded area shows the roadways that mainly serves transit trips.



(**a**) Clustering Results when $k = 4$



(**b**) Clustering Results when $k = 2$



(**c**) Gantry Locations

Figure 4. Gantry Clustering Results.

In Figure 4b, the cluster centroids of local travel transit travel flow and are (3669, 4051) and (5768, 15,195), respectively. The boundary line between the two clusters in the figure is:

$$VOLUME_1 = -0.2VOLUME_0 + 10520 \quad (6)$$

The results show that on gantry segments primarily serving local travel, half of the traffic flow is still transit travel. This is due to the characteristic of highways primarily serving long-distance travel.

3.4. Percolation Simulation and Segments Importance Ranking

Based on the volumes of transit flow and local travel flow on highway network segments, a local–transit percolation simulation of the highway network was conducted. The unit percolation threshold intervals were set as $\Delta q_1 = \Delta q_2 = 1/100$. The changes in the size of the giant component under different (q_1, q_2) values are shown in Figure 5.

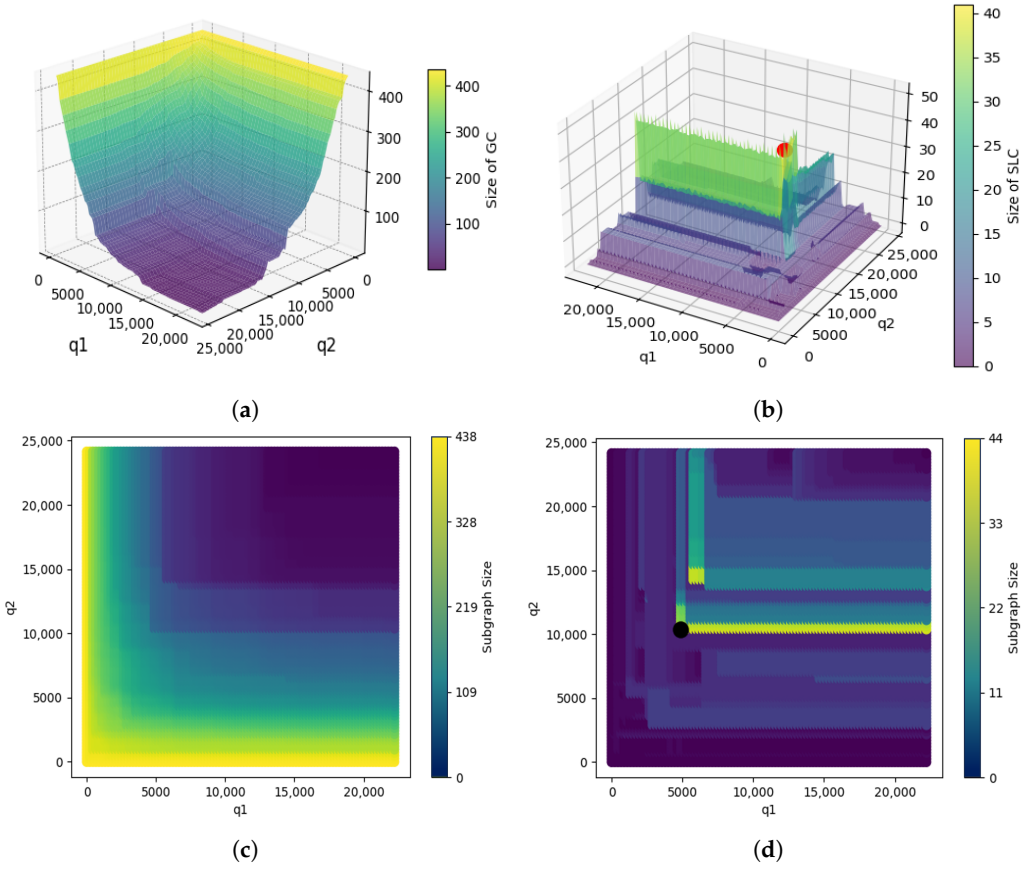


Figure 5. Variations in GC and SLC in Local–Transit Percolation Simulation. (**a**) 3D Plot of GC Size Variation. (**b**) 3D Plot of SLC Size Variation. (**c**) Plane Plot of GC Size Variation. (**d**) Plane Plot of SLC Size Variation.

As Figure 5 shows, as q_1 and q_2 increase, the size of the GC decreases while the size of the SLC increases. When the size of the SLC reaches its maximum, the giant component size decreases rapidly, indicating a percolation phase transition. The results show that when the critical percolation thresholds q_{c1} and q_{c2} are 4889 and 10,379, respectively, the size of the SLC in the network reaches its maximum of 44.

According to the importance ranking method described in Section 2.4, primary ranking of the segments was performed based on the size of the SLC under each percolation threshold. Then, based on the clustering results in Figure 4, secondary ranking of the segments was performed under the same second largest cluster, obtaining the importance ranking results of highway segments, as shown in Figure 6. In addition, as shown in Figure 7, the percolation threshold pairs are sorted by the size of SLC and the number of failed segments of W corresponding to that SLC is counted. Furthermore, the horizontal coordinate is the rank of segments by SLC size. It can be observed that when the size of the SLC is larger, a greater number of edges are removed during the corresponding percolation phase transition, which will accelerate the collapse of the network.

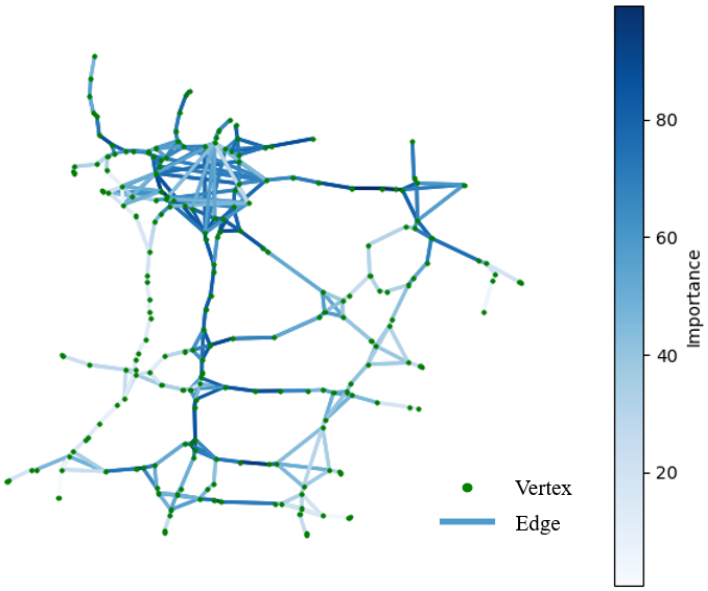


Figure 6. Results of Highway Importance Ranking.

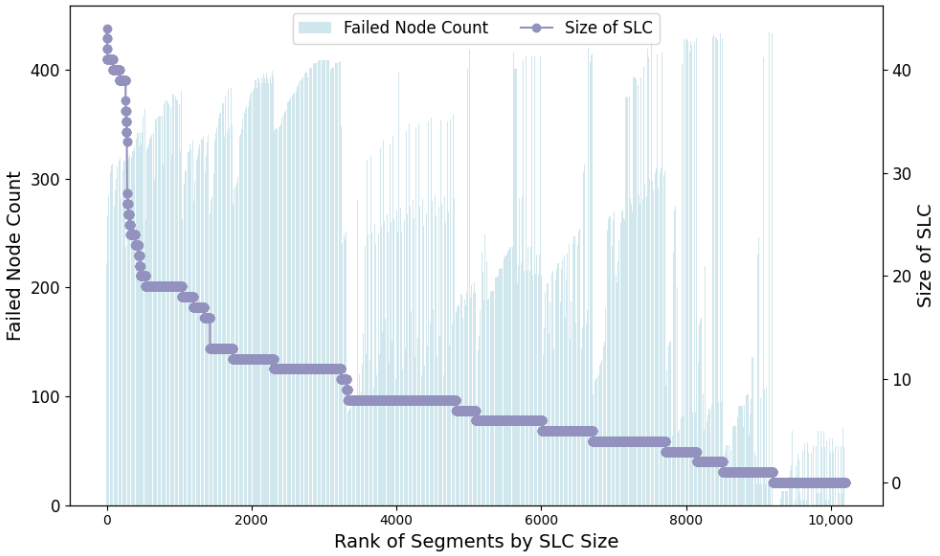


Figure 7. Number of Failure Edges and Size of the SLC.

4. Result and Discussion

4.1. Results

This paper proposes a local–transit percolation and clustering-based method to rank highway segment importance. To evaluate the effectiveness of the proposed method, a control experiment was conducted. Common methods for importance ranking, such as those based on edge betweenness centrality and weighted shortest paths, were used as baselines for comparison. The weighted shortest paths focus on the absolute frequency of edges used in the shortest paths. Furthermore, edge betweenness centrality calculates the proportion of shortest paths between all pairs of nodes that pass through that particular edge. Importance rankings were generated separately for the proposed method and the baselines, and all the segments were sequentially removed. The overlap of the top 100 edges ranked by the three strategies were calculated, as shown in Figure 8. A large overlap suggests that the three strategies produced similar results in terms of ranking the importance of edges. There are 22 edges that are ranked highly across all three strategies; thus, these edges are crucial for the connectivity and functionality of the network.

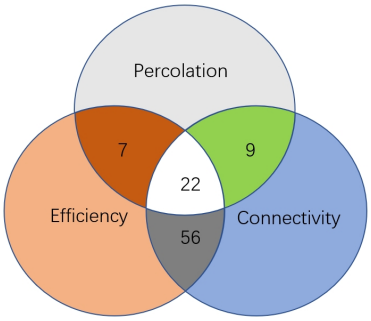


Figure 8. Overlap of the Top 100 Edges Ranked by the Three Strategies.

A comparative analysis was performed to evaluate the effects of three path removal strategies, utilizing five key indicators: local traffic, transit traffic, demand impedance ratio, the size of the giant component, and the sum of the inverse travel times. The size of the GC is commonly employed to assess the overall connectivity of a network, while the sum of the inverse travel times quantifies the connectivity between nodes by aggregating the inverse travel times for all origin–destination pairs. The demand impedance ratio is defined as the ratio of travel demand to travel impedance for each OD pair, which is shown in Equation (7). It evaluates the level of service and efficiency of the network.

$$E = \frac{1}{n_w} \sum_{w \in W} \frac{d_w}{\lambda_w} \tag{7}$$

where n_w is the number of OD pairs in the network set **W**, d_w is the travel demand of OD pair **W**, λ_w is the travel impedance of OD pair **W**.

As edges were progressively removed, the changes in local flow and transit flow for each method were as shown in Figure 9a,b, respectively. Figure 9a shows that the local flow decreased the most under our proposed segment importance ranking method compared to the baselines. This suggests that the method is more sensitive to disruptions in key segments that have higher importance, leading to a more substantial impact on localized traffic. Compared to local flow in Figure 9a, transit flow in Figure 9b decreases faster under our proposed method. This can be attributed to the difference in path selectivity between the two types of flows. Local flow typically has multiple alternative paths available between vertices, which provides resilience to disruptions caused by the removal of individual segments. In contrast, transit flow often relies on a smaller set of critical paths to maintain connectivity across the network. As a result, disruptions to these critical segments have a more evident impact on transit flow, leading to its faster reduction compared to local flow.

Furthermore, the impacts of these removals were analyzed using three indicators: the size of the giant component (GC), the network demand impedance ratio, and the reciprocal sum of travel time. The size of the GC reflects the overall connectivity of the network. The network demand impedance ratio, calculated as the total flow divided by the travel time for all OD pairs and then averaged, represents the network's efficiency under disruptions. As shown in Figure 9c, when the top 50 segments ranked by importance are sequentially removed, the size of the GC decreases the most under the proposed method compared to the baselines. A smaller GC size indicates weaker network connectivity, suggesting that the removal of segments identified by our method has a more substantial impact on the highway network's connectivity. Additionally, the network demand impedance ratio decreases the most across all removed segments, as illustrated in Figure 9d. Furthermore, Figure 9e illustrates the curve of the standard deviation of the demand impedance ratio as segments are removed. A smaller standard deviation indicates fewer highway segments with a high demand impedance ratio. The figure demonstrates that our method results in a lower standard deviation as the top 300 highway segments are removed, suggesting

that our approach efficiently identifies segments with a high demand impedance ratio. In addition, as shown in Figure 9f, when the top 100 segments ranked by importance are sequentially removed, the reciprocal sum of travel time decreases rapidly under the proposed method compared to the baseline methods, indicating that these segments have a significant impact on travel time. Overall, these results highlight the superior performance of the proposed method in capturing the true importance of segments within the network structure compared to the baselines.

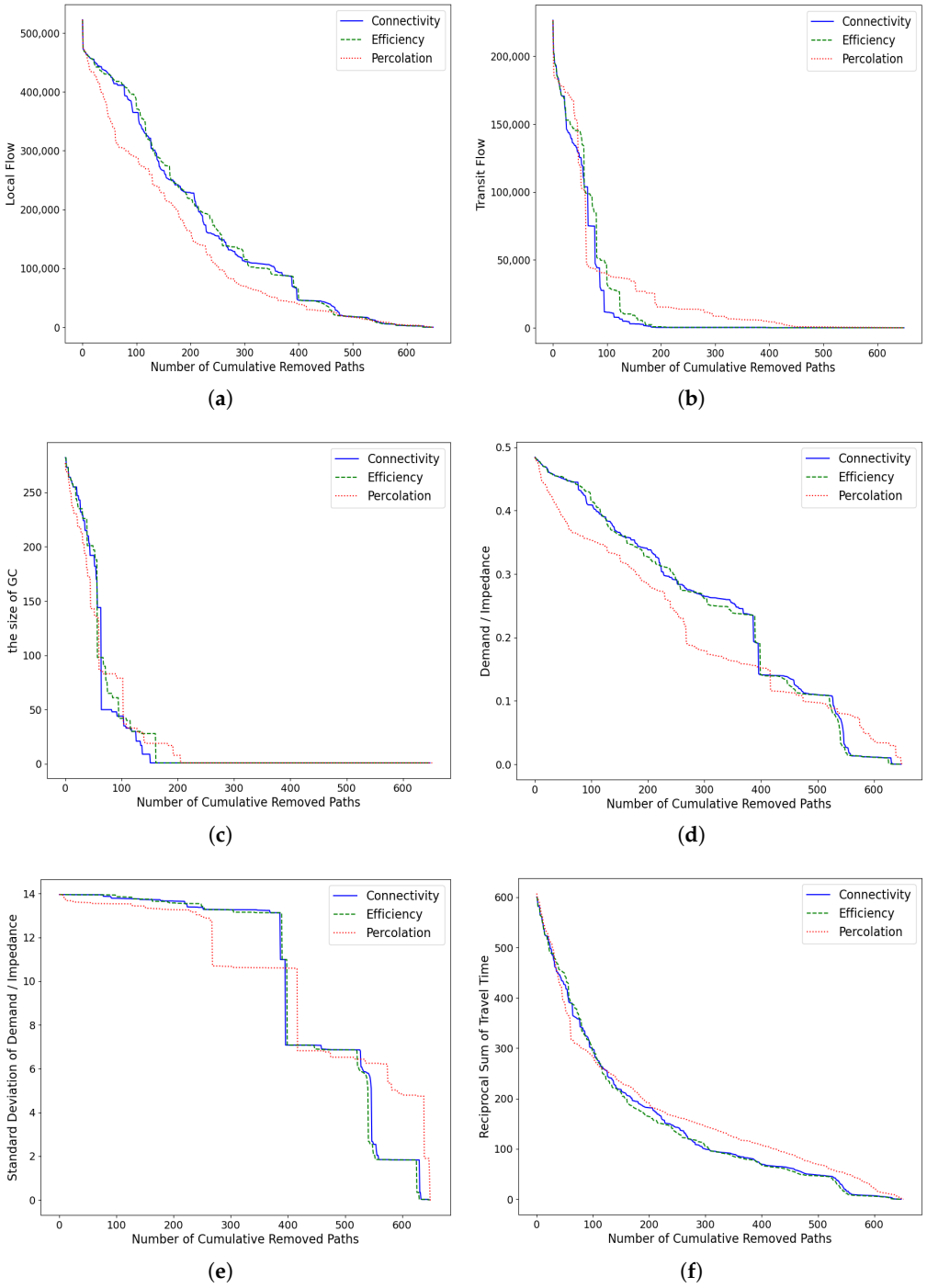


Figure 9. Results of the Control Experiment. (**a**) Variation in Local Traffic Flow. (**b**) Variation in Transit Traffic Flow. (**c**) Variation in the Size of the GC. (**d**) Variation in the Network Demand Impedance ratio. (**e**) Variation in the Standard Deviation of the Demand Impedance Ratio. (**f**) Variation in Reciprocal Sum of Travel Time.

4.2. Discussion

In this study, we proposed a method for ranking the importance of road segments based on highway gantry data. The results demonstrate that our method is effective and advanced in ranking road segment importance. However, it is important to note that the method heavily relies on the accuracy and completeness of the highway gantry records, which may be influenced by various factors, introducing uncertainty into the model.

The quality of gantry monitoring data can be affected by several factors, including the accuracy of the equipment, weather conditions, and fluctuations in traffic flow. These factors may introduce noise or errors into the data, affecting the stability and accuracy of the model. To investigate the impact of this uncertainty, we designed experiments in which Gaussian noise was added to the transit and local flow data of the road segments. The formula for adding noise is as Equation (8).

$$q_i^{noisy} = q_i^{original} + \mathcal{N}(0, \sigma^2_{q_i^{original}} \cdot \alpha)) \tag{8}$$

where $\mathcal{N}(0, \sigma^2)$ represents Gaussian noise with a mean of 0 and variance σ^2. The noise factor α was set to 0, 0.15, 0.30, and 0.45. We then ranked the importance of road segments using our proposed method, as well as two common baseline methods: edge betweenness centrality and weighted shortest paths. Subsequently, we removed road segments in order of their importance and compared the changes in the size of the GC in the network, as shown in Figure 10.

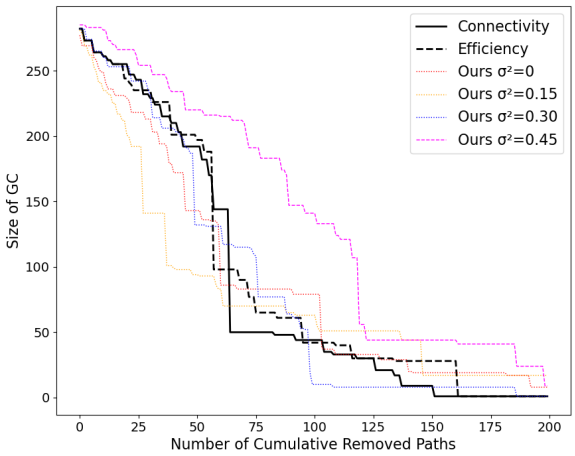


Figure 10. Results of Highway Importance Ranking under Uncertainty.

Figure 10 shows that for the top 50 important road segments, when the noise factor $\alpha < 0.30$, our method consistently leads to the fastest decline in the GC. This suggests that despite the presence of noise in the gantry data, our method retains a high level of robustness. In particular, when the noise level is relatively low, our method outperforms the baseline methods, demonstrating its strong adaptability and stability in complex networks.

However, it is worth noting that when the noise level is high ($\alpha > 0.30$), the performance of our methods is significantly affected, indicating that under extreme data uncertainty, the performance of current methods may not meet expectations. Therefore, future research could further optimize the model to increase its tolerance to data noise or incorporate data cleaning techniques to minimize the impact of noise on the results.

5. Conclusions

Highways are important transportation modes for medium- and long-distance travel, accommodating both transit and local travel demands. They are susceptible to disturbances such as natural disasters, which can affect network efficiency. After such disturbances, limited resources should be preferentially allocated to segments that have a greater impact on network travel and are of higher importance. Therefore, based on highway gantry records, this paper proposes an importance ranking method for highway segments using local–transit percolation and clustering-based method. The main work and conclusions are as follows:

1. Highway network modeling based on gantry data: A highway network model was constructed by graph theory utilizing vehicle travel trajectories recorded by gantries. This approach overcomes the shortcomings of traditional methods in processing geographic information data and improves network modeling efficiency.
2. Classification of segment travel demand: The travel demands on each segment were statistically analyzed. Based on the transit and local travel flows carried by the segments, the k-means algorithm was used to cluster the segments, identifying the main travel demand categories served by each segment.
3. Local–transit percolation simulation: A local–transit percolation theory was proposed and simulations based on the classification results of segment travel flows were performed. A percolation phase transition occurs when the size of the SLC suddenly increases. The size of the SLC corresponding to each percolation threshold was used for primary ranking of the segments. Based on the clustering results, secondary ranking was performed for segments under the same SLC, resulting in the final importance ranking.
4. Case study and uncertainty analysis: A case study was conducted on the highway network within four prefecture-level cities in Hebei Province, China, obtaining importance ranking results. A control experiment was designed to compare the performance changes of the network under different segment removal strategies. Segments with high importance were gradually removed using our method and compared to traditional methods based on edge betweenness centrality and weighted shortest paths. Results show that our method has a greater impact on the overall travel demand fulfillment and network connectivity after the failure of highly important segments. In addition, uncertainty analysis was conducted by adding Gaussian noise to the segment flow data. Results show that when noise factor $\alpha < 0.30$, removing the top 50 ranked highway segments by our method can still make the highway network collapse faster than the baseline method. The control experiment and uncertainty analysis validate the advancement and robustness of the proposed importance ranking method.

However, this paper still has limitations. The study area is divided into transit and local vehicles based on administrative boundaries, without considering the zoning formed by traffic flow characteristics. In addition, there are still many events that could impact on the highway functional performance apart from the vehicle travel types, such as traffic accidents, extreme weather, and road maintenance. In the future, we will focus on segment importance based on travel characteristics and optimize the analysis considering specific traffic events.

Author Contributions: Conceptualization, H.L., Y.L. and J.C.; methodology, H.L. and Y.L.; software, C.L.; validation, Z.L. and L.X.; formal analysis, C.L.; investigation, Z.L.; resources, H.L.; data curation, H.L.; writing—original draft preparation, H.L. and Y.L.; writing—review and editing, H.L. and W.W.; visualization, Y.L.; supervision, J.C.; project administration, J.C.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China grant number U20A20330 and 52131203.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available on request.

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

References

1. Taylor, M.A.; Sekhar, S.V.; D'Este, G.M. Application of accessibility based methods for vulnerability analysis of strategic road networks. *Networks Spat. Econ.* **2006**, *6*, 267–291. [CrossRef]
2. Kim, Y.; Tak, H.Y.; Kim, S.; Yeo, H. A hybrid approach of traffic simulation and machine learning techniques for enhancing real-time traffic prediction. *Transp. Res. Part Emerg. Technol.* **2024**, *160*, 104490. [CrossRef]
3. Zhou, Y.; Wang, J. Critical Link Analysis for Urban Transportation Systems. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 402–415. [CrossRef]
4. Almotahari, A.; Yazici, A. A computationally efficient metric for identification of critical links in large transportation networks. *Reliab. Eng. Syst. Saf.* **2021**, *209*, 107458. [CrossRef]
5. Kumar, A.; Haque, K.; Mishra, S.; Golias, M.M. Multi-criteria based approach to identify critical links in a transportation network. *Case Stud. Transp. Policy* **2019**, *7*, 519–530. [CrossRef]
6. Chang, K.H.; Sheu, J.B.; Chen, Y.J.; Chang, C.H.; Liu, C.H. Practice-based post-disaster road network connectivity analysis using a data-driven percolation theory-based method. *Transp. Res. Part E Logist. Transp. Rev.* **2023**, *171*, 103022. [CrossRef]
7. Li, Y.; Wu, J.; Xiao, Y.; Hu, H.; Wang, W.; Chen, J. Resilience analysis of highway network under rainfall using a data-driven percolation theory-based method. *Phys. A Stat. Mech. Its Appl.* **2024**, *638*, 129639. [CrossRef]
8. Liu, S.; Li, Z.; Zhong, J.; Li, D. Percolation transition in temporal airport network. *Chin. J. Aeronaut.* **2020**, *33*, 219–226. [CrossRef]
9. Zhou, Y.; Li, S.; Wang, J. Modeling stochastic behavior of road networks with disruptions using percolation theory. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 13230–13241. [CrossRef]
10. Cao, Y.Y.; Liu, R.R.; Jia, C.X.; Wang, B.H. Percolation in multilayer complex networks with connectivity and interdependency topological structures. *Commun. Nonlinear Sci. Numer. Simul.* **2021**, *92*, 105492. [CrossRef]
11. Kryven, I.; Bianconi, G. Enhancing the robustness of a multiplex network leads to multiple discontinuous percolation transitions. *Phys. Rev. E* **2019**, *100*, 020301. [CrossRef] [PubMed]
12. Hayasaka, S. Explosive percolation in thresholded networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *451*, 1–9. [CrossRef]
13. Song, Z.; Cai, J.; Yang, Q. Taxi travel distance clustering method based on exponential fitting and k-means using data from the US and China. *Systems* **2024**, *12*, 282. [CrossRef]
14. Wang, W.; Xia, F.; Nie, H.; Chen, Z.; Gong, Z.; Kong, X.; Wei, W. Vehicle trajectory clustering based on dynamic representation learning of internet of vehicles. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 3567–3576. [CrossRef]
15. Lv, Z.; Qiao, L.; Cai, K.; Wang, Q. Big data analysis technology for electric vehicle networks in smart cities. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 1807–1816. [CrossRef]
16. Song, Q.; Hu, J.; Zhang, R.; Zhang, Z. An urban topological map generation method for traffic flow prediction based on road segment clustering with floating vehicle trajectory dataset. In Proceedings of the 19th Cota International Conference of Transportation Professio (CICTP 2019), Nanjing, China, 6–8 July 2019; pp. 2330–2338.
17. Košanin, I.; Gnjatović, M.; Maček, N.; Joksimović, D. A clustering-based approach to detecting critical traffic road segments in urban areas. *Axioms* **2023**, *12*, 509. [CrossRef]
18. Deng, Y.; Liu, S.; Zhou, D. Dependency cluster analysis of urban road network based on percolation. *Transp. Res. Part C Emerg. Technol.* **2023**, *154*, 104264. [CrossRef]
19. Vivek, S.; Conner, H. Urban road network vulnerability and resilience to large-scale attacks. *Saf. Sci.* **2022**, *147*, 105575. [CrossRef]
20. Cogoni, M.; Busonera, G. Stability of traffic breakup patterns in urban networks. *Phys. Rev. E* **2021**, *104*, L012301. [CrossRef] [PubMed]

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.