Traffic Accident Risk Factor Identification Based on Complex Network

Yang Chen and Yuanchang Deng*
Guangdong Provincial key laboratory of intelligent transportation system, school of intelligent engineering, Sun Yat-Sen university, Guangzhou, 510275, China
Email: dengyc@mail.sysu.edu.cn

Abstract. Identifying the risk characteristics of traffic accidents is of great significance for in-depth understanding and promotion of traffic safety. 85 serious traffic accidents were selected from the road traffic accident liability certification and serious accident reports issued by the traffic police department, and 110 risk points and 218 triggering relationships were abstracted based on network theory analysis to establish a traffic accident risk network. Analyse and compare indicators such as network density and network diameter to reveal the overall nature of the traffic accident risk network. The key risk factors of the risk network are analysed and identified using node indicators such as degree and betweenness. The results show that most traffic accidents are triggered by human factors, among which the driver factor is the main factor. Speeding and improper emergency measures are the two most important factors that need to be controlled. Based on the results of this analysis, this article puts forward management suggestions to improve the level of traffic safety.

Keywords. Accident analysis, risk factors, complex networks, traffic safety.

1. Introduction
According to the “Statistical Annual Report on Road Traffic Accidents of the People's Republic of China”, there were 20,049,049 road traffic accidents across the country in 2017, with 63,772 deaths; direct property losses were 121311.3 million yuan [1]. Once a road traffic accident occurs, it will bring irreparable losses to individuals, families, and society. It is important to investigate the risk factors of traffic accidents. The influencing factors of traffic accidents are usually divided into these four aspects, including human behavior factors [2], vehicle mechanical factors, driving road factors and environmental factors [3].

Case analysis is a commonly used method for studying risks. In addition to case analysis, methods commonly used to study risk include Bayesian model [4-5], neural network model, support vector machine and other methods. Hu Liwei took the raw data processing and analysis of 6,261 road traffic accidents that occurred in the Yunnan Plateau from 2010 to 2015 as an example to analyze the impact of the special geological and meteorological environment in the plateau area on road traffic risks [6]. Song Yinghua, et al. extracted relevant data from 95 investigation reports on major road accidents and used social network analysis (SNA) to build a network of relationships among the causes of major road accidents to gain a deeper understanding of traffic accident risks [7]. Wang Enda, et al. modeled 383 major traffic accidents in my country and conducted a quantitative study on the accident risk caused by the combination of different types of risk factors [8]. With the rise and wide application of artificial intelligence technology, IFOA-GRNN and apriori algorithms have also been used for risk prediction and risk factor analysis [9].
The case analysis method can only analyse a single accident and does not focus on the study of the correlation between risk factors. Therefore, some scholars use the method of accident chain analysis to analyse the accident risk [10-12]. However, the accident chain alone cannot explore the complex interactions between risks, and more accurate models must be used for quantitative analysis. The complex network model provides a theoretical basis for analyzing the relationship between risk factors. Deng et al. identified the interrelationships between 105 risks and 135 risks from 126 coal accidents and established a coal risk network based on complex network theory. The study found that among all the risks, roof collapse, fire, and excessive gas concentration are three. The most valuable risk control target [13]. Jia et al. established a risk network model using rail transit vehicle system components and component connection relationships as nodes and edges and carried out research on key component risks and risk network evaluation [14]. Zhou et al. established a risk network model for 67 typical bridge accidents and 64 tunnel accidents. Research has shown that accidents occur mainly due to management and worker factors, while pure environmental factors have no significant impact on the occurrence of accidents [15].

However, there is no domestic literature report about using complex network models to explore the risk factors of traffic accidents. There is a big difference between road traffic accidents in China and foreign countries. In view of this, the author will collect typical domestic traffic accident cases, and on the basis of a complex network model, explore the risk factors of traffic accidents and the relationship between risk factors, and then provide traffic Propose reasonable suggestions for safety management.

2. Data Collection
In the road traffic accident responsibility certification and serious accident reports issued by the traffic police department, a road traffic accident database was established after analyzing the accident. Selected typical and major cases from the database, such as drunk driving, pedestrians being hit by red lights, transportation of dangerous explosives, cliff falling accidents, etc. The accidents selected in this article are all fatal. The transportation system is a complex system composed of people, vehicles, roads, and the environment. On this basis, this article divides the risk factors of road traffic accidents into driver factors, other factors, vehicle factors, road factors, and weather factors.

Most traffic accidents happen in an instant. To investigate their causes, most traffic accidents can be prevented and avoided in advance. The eventual occurrence of every road traffic accident requires certain key nodes in the process. The retrospective analysis of key nodes is essential to the prevention of traffic accidents. Therefore, we can decompose each traffic accident into several key nodes in the order of occurrence, and then abstract each key node into a risk. These risks can be combined into a chain of accidents linked end to end in chronological order. For example, in a typical rollover accident, the driver drove a vehicle with a potential safety hazard due to weak safety awareness, and the vehicle was overloaded with cargo, and it rolled over while driving on the road, overwhelming the workshop on the roadside, Which eventually caused the workers to crush to death. By abstracting the key nodes in the accident process as risk points, the following accident chain is obtained:

Weak safety awareness→ driving dangerous vehicles→ overloading goods→ overturning→ overwhelming the workshop→ run over workers.

Some risk points will appear in different accident chains at the same time. Using the same risk point, all accident chains can be combined to form a risk trigger network.

3. Network Model
All road traffic risks and the trigger relationship between risks can establish a risk trigger network:

\[ G = (\mathcal{R}_{\text{driver}} + \mathcal{R}_{\text{human}} + \mathcal{R}_{\text{motor}} + \mathcal{R}_{\text{road}} + \mathcal{R}_{\text{worker}} + \mathcal{E}_{\text{trigger}}) \]

where \( \mathcal{R}_{\text{driver}} = \{r_1, r_2, \ldots, r_m\} \) indicates the set of driver risk factors, which is represented by \( \triangle \) in the road traffic risk trigger network;

\( \mathcal{R}_{\text{human}} = \{r_1, r_2, \ldots, r_n\} \) Indicates the set of other people's risk factors, which is represented by \( \square \) in the road traffic risk trigger network;
\[ R_{\text{veh}} = \{ r_1, r_2, \ldots, r_i, \ldots, r_m \} \] indicates the set of vehicle risk factors, represented by ○ in the road traffic risk trigger network;

\[ R_{\text{road}} = \{ r_1, r_2, \ldots, r_i, \ldots, r_r \} \] indicates the set of road risk factors, which is represented by in the road traffic risk trigger network;

\[ R_{\text{weather}} = \{ r_1, r_2, \ldots, r_i, \ldots, r_w \} \] indicates the set of weather risk factors, which is represented by in the road traffic risk trigger network;

\[ E_{\text{trigger}} = \{ (r_i, r_j), j \neq k \} \] indicates the set of trigger relationships among road traffic risk factors.

### 3.1. Network Scale

The network size is defined as the total number of nodes in the network organization. Then the scale of the road traffic risk trigger network is:

\[
\sum_{n} \text{nodes} = |R_{\text{driver}}| + |R_{\text{human}}| + |R_{\text{motor}}| + |R_{\text{road}}| + |R_{\text{weather}}|
\]

### 3.2. Network Diameter

The network diameter is defined as the maximum distance between any two nodes in the network. Then the diameter of the road traffic risk trigger network is:

\[
d_{\text{max}} = \max_{i,j} d_{ij}
\]

### 3.3. Average Path Length

In network organization, the average distance between any two nodes and is defined as the path length of these two nodes. For the road traffic risk trigger network, the average characteristic path length equation is:

\[
l_{\text{avg}} = \frac{\sum_{i,j} d_{ij}}{C_n}
\]

### 3.4. Clustering Coefficient

The clustering coefficient is an index used to indicate the degree of clustering of nodes in a graph. The aggregation coefficient of a single node in the network graph can be expressed as:

\[
C_i = \frac{\sum_{j<k} d_{ijk}}{d_i(d_i-1)}
\]

### 3.5. Degree Centrality

Degree centrality is the most direct measure of node centrality in network analysis. In a one-way network, we can divide the degree centrality into out-degree centrality and in-degree centrality. The scale of the road traffic risk trigger network is represented by \( N \), and the trigger relationship among road traffic risks constitutes the adjacency matrix \( A = [a_{jk}]_{m \times n} \). Then \( C_{\text{out}}(j) \) represents the out-degree centrality of the road traffic accident risk, and \( r_j \) represents the total number of risks that can be directly triggered: \( C_{\text{out}}(j) = \sum_{i} a_{ji} \). \( C_{\text{in}}(j) \) represents the in-degree centrality of the road traffic accident risk \( r_i \), and represents the total number of other risks that can trigger the risk: \( C_{\text{in}}(j) = \sum_{i} a_{ij} \).

### 3.6. Intermediate Centrality

The centrality of network nodes is also called betweenness, which refers to the proportion of the shortest path between all pairs of nodes in the network passing through the node. The greater the
centrality of the node, the more the shortest path through the node, the more obvious its pivotal role in the entire risk network, and the stronger its influence and control. In the road traffic risk trigger network, the median centrality of risk \( r_j \) can be expressed as the number of times the shortest path between all risks in the risk trigger network and other risks traverses. The median centrality \( B_j \) of risk \( r_j \) can be expressed as:

\[
B_j = \sum_{i=1}^{n} \left( \frac{n_i(j)}{n_i} \right) \]

Where, \( n_{ij} \) is the number of shortest paths from risk \( r_i \) to risk \( r_j \); \( n_i(j) \) is the number of shortest paths from risk \( r_i \) to risk \( r_j \) through risk \( r_j \) at the same time.

4. Network Analysis
With the continuous development of complex network theory, many results have been achieved in the statistical indicators of network structure, which is also the basis for statistical description of various topological features. This study uses typical indicators such as network diameter, network density, average path length, clustering coefficient, degree centrality, and intermediate centrality to explore the nature of the road traffic accident risk trigger network.

4.1. The Overall Topology of the Network
In the traffic accident risk network, as shown in figure 1, there are 110 risk points, and the network scale is 110. Under normal circumstances, the more complex the network, the more complex the network structure, and the greater the difficulty of research. However, the types of traffic accidents are limited, and we select representative accident types for research, so the scale of the network will not be too large.

Figure 1. Traffic accident risk network.

In the traffic accident risk network, the network diameter is 10. The larger the diameter of the risk network, the more key nodes that a traffic accident passes through, and the more likely it is to take measures against some of the key nodes in the accident chain to prevent accidents.

In the traffic accident risk network, the network density is 0.018. The higher the density, the closer the relationship between the risk points in the traffic accident risk network, the faster the key nodes are triggered, and the more difficult it is to prevent traffic accidents.

The small world network has the characteristics of high aggregation coefficient and short average path length. The aggregation coefficient of the traffic accident risk trigger network is 0.054, and the
average path length is 3.48, which conforms to the small world characteristics. For small-world networks, by changing a few nodes or edge structures, the nature of the network can be drastically changed. For the traffic accident risk network, finding a way to slightly change the key risk points can significantly improve safety performance and reduce the possibility of traffic accidents. The degree distribution function of the traffic accident risk network is defined in equation (1):

\[ P(k) = 2.496k^{-1.611} \left( R^2 = 0.9101 \right) \]  

(1)

The degree distribution function is a contemplative distribution, as shown in figure 2, indicating that the traffic accident risk network conforms to the nature of the scale-free network. This means that in practical applications, several highly critical risk factors can be identified in the traffic accident risk network, and these critical risk factors can be controlled to effectively improve the safety level.

![Figure 2. Degree distribution of traffic accident risk network.](image)

4.2. The Nature of Network Nodes

Analyzing the nodes of the risk network can explore the deep-seated causes of accidents and provide suggestions for traffic safety management.

4.2.1. Out-of-Degree Centrality. In the traffic risk trigger network, the top 10 risk points are sorted out according to the degree of centrality, as shown in table 1.

| Number | Risk point                              | Classification       | Out-of-degree centrality |
|--------|-----------------------------------------|----------------------|--------------------------|
| 6      | Over-speed driving                       | Driver factors       | 10                       |
| 26     | Driving a dangerous vehicle              | Driver factors       | 9                        |
| 2      | Avoid giving way in time                  | Driver factors       | 6                        |
| 20     | Cargo overload                           | Motor factors        | 6                        |
| 30     | Improper emergency measures              | Driver factors       | 6                        |
| 100    | Drug driving                             | Driver factors       | 5                        |
| 8      | Vehicle collision                        | Motor factors        | 5                        |
| 62     | Inattention                              | Driver factors       | 5                        |
| 3      | Not driving in a prescribed lane         | Driver factors       | 4                        |
| 60     | hit and run                              | Driver factors       | 4                        |
| 23     | Improper driving operation               | Driver factors       | 4                        |
| 55     | Foggy day                                | Weather factors      | 4                        |
| 68     | Drunk driving                            | Driver factors       | 4                        |
In the traffic risk trigger network, the greater the centrality of the risk, the more likely the risk point is to trigger the risk.

Among the risk factors ranked by the degree of centrality, the proportion of driver factors is much higher than vehicle factors and weather factors. The fog factor is the largest out of the weather factors, which means that fog has the highest probability of accidents in severe weather. Driver factors include speeding, driving potentially dangerous vehicles, untimely avoidance, improper emergency measures, and inattention. Among them, speeding has the highest degree of center degree, indicating that compared with other risk points, speeding is most likely to induce traffic accidents and cause serious consequences. Backtracking the accident chain found that the risk points that can be triggered by speeding include sideslip, delayed avoidance, vehicle collisions, rear-end collisions, etc. The risk of this peak is difficult to control due to multipath.

The risk factors that induce accidents often have similarities. For example, both drunk driving and drug driving belong to the driver’s knowledge of breaking the law. The dedication factors such as untimely avoidance, improper emergency measures, inattention, improper driving, etc., all belong to the driver's driving skill level. Among these key risk factors, the driver factor is still the main factor causing the accident. Traffic control departments should take measures to control these key factors that easily trigger risks, reduce the possibility of key nodes triggering more risk points, and cut off the chain of accident propagation.

4.2.2. In-Degree Centrality. In the traffic risk trigger network, sorted by in-degree centrality, the top 10 risk points are sorted out, as shown in table 2.

| Number | Risk point                     | Classification | In-degree centrality |
|--------|--------------------------------|----------------|----------------------|
| 8      | Vehicle collision              | Motor factors  | 15                   |
| 24     | Driver and passenger casualties| Driver factors | 11                   |
| 66     | Rear-end collision             | Motor factors  | 9                    |
| 6      | Over-speed driving             | Driver factors | 8                    |
| 12     | Passenger casualties           | Human factors  | 8                    |
| 2      | Avoid giving way in time       | Driver factors | 7                    |
| 25     | Driver casualties              | Driver factors | 7                    |
| 30     | Improper emergency measures    | Driver factors | 6                    |
| 23     | Improper driving operation     | Driver factors | 5                    |
| 37     | Collision pedestrian           | Motor factors  | 5                    |

In the traffic risk trigger network, the greater the risk in-degree centrality, it indicates that this is the consequence that is easily caused in a traffic accident. Vehicle collisions, driver and passenger injuries and deaths, and rear-end collisions have the largest in-degree centrality, indicating that the most common forms of traffic accidents are facing and facing collisions between vehicles, which in turn cause injuries to the driver and passengers in the car.

Backtracking the accident chain for analysis, it is found that the precursor factors of vehicle collisions include not timely avoidance, not driving in the prescribed lane, unclear traffic signs and so on. Vehicle collisions will induce more risk factors, such as vehicle disintegration, passenger casualties, driver casualties, collisions with pedestrians, and so on.

Remove the dangerous factors that need to be prevented, such as speeding, delayed avoidance, and collisions with pedestrians are also common consequences. This shows that the consequences of most traffic accidents are threats to human safety, and pedestrians are more likely to be injured in traffic accidents.

4.2.3. Middle Centrality. In the traffic risk trigger network, the top 10 risk points are sorted out according to the size of the centrality, as shown in table 3.
Table 3. Top risk factors by the order of the betweenness.

| Number | Risk point                       | Classification  | Middle centrality |
|--------|----------------------------------|-----------------|-------------------|
| 6      | Over-speed driving               | Driver factors  | 571.559           |
| 30     | Improper emergency measures      | Driver factors  | 403.35            |
| 4      | Rollover                         | Motor factors   | 363.826           |
| 2      | Avoid giving way in time         | Driver factors  | 363.752           |
| 62     | Inattention                      | Driver factors  | 325.675           |
| 26     | Driving a dangerous vehicle      | Driver factors  | 280.933           |
| 20     | Cargo overload                   | Driver factors  | 252.691           |
| 60     | Hit and run                      | Driver factors  | 243.512           |
| 8      | Vehicle collision                | Motor factors   | 200.144           |
| 79     | Falling off a cliff              | Motor factors   | 188.745           |
| 37     | Collision pedestrian             | Motor factors   | 171               |
| 82     | Vehicle disintegration           | Motor factors   | 155               |
| 32     | The intersection did not slow down| Driver factors  | 151.914           |

In the traffic risk trigger network, the greater the centrality of the risk point, the more paths through the risk point. In the traffic accident risk network, the middle degree of speeding (No. 6) is the highest, which indicates that speeding is a key risk factor in the traffic accident risk network, which seriously affects traffic safety. Speeding, improper emergency measures, untimely avoidance, and inattention play a pivotal role in the entire network. If measures can be taken to control these factors, the safety performance of the entire network will be improved. Accidents cannot be completely avoided, but the key factors can be changed to reduce the probability of accidents.

Hit-and-run is the escape from the scene after the accident, which reflects that the driver's legal literacy and safety awareness need to be improved. The superimposition of factors such as drunk driving and drug driving with hit-and-run accidents often results in secondary accidents and amplifies the serious consequences of accidents. Traffic control departments must take strict precautions.

An intersection is a place where pedestrians and vehicles meet, and it is also a dangerous section with a high frequency of accidents. Only by slowing down can the driver have time to deal with the complicated traffic environment.

5. Conclusion
(1) The calculation of the aggregation coefficient and network density shows that the traffic accident risk network is a small-world network.
(2) The statistical analysis of the nodes of the traffic accident risk network shows that the degree distribution obeys the power law distribution, and the traffic accident risk network has a scale-free feature.
(3) Using degree centrality and intermediate centrality as the measurement, identify the key risk points in the traffic accident risk network, and find that the human factor is dominant, and the driver factor is particularly important.
(4) In response to the identified key risk factors such as speeding and improper emergency measures, propose prevention suggestions and remedial measures for traffic accidents from the four aspects of people, vehicles, roads, and the environment to improve the level of traffic safety.

References
[1] Khorasani Z D 2019 Road traffic injuries measures in the Eastern Mediterranean Region: findings from the Global Status Report on Road Safety – 2015 Journal of Injury and Violence Research 11 149-158.
[2] Moradi A, Nazari S S H and Rahmani K 2019 Sleepiness and the risk of road traffic accidents: A systematic review and meta-analysis of previous studies *Transportation Research Part F: Psychology and Behaviour* **65** 620-629.

[3] Tsubota T, Fernando C, Yoshii T and Shirayanagi H 2018 Effect of road pavement types and ages on traffic accident risks *Transportation Research Procedia* **34** 211-218.

[4] Yu R J, Mohamed A A and Mohamed A 2013 Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors *Accident Analysis and Prevention* **50** 371-376.

[5] Cai M, Cheng G Z, Zhang S L, Wang B and Zhou K 2019 Risk assessment of highway roadside accidents based on bayesian network *Journal of Transportation Systems Engineering and Information Technology* **02** 202-208.

[6] Hu L W, Xue G, Li L Y, Wang M and She T Y 2018 Analysis of coupling of highway traffic risks in geological and meteorological environment of plateau regions *China Journal of Highway and Transport* **01** 110-119.

[7] Song Y H, Tian L and Huo F Z 2019 Establishment of risk index system for major road traffic accidents based on SNA *Journal of Safety Science and Technology* **06** 151-156.

[8] Wang E D and Shi J 2017 Quantification and analysis of traffic crash risk under the superposition of multiple factors based on the mutual information theory *Journal of Transportation Engineering* **06** 1-5+12.

[9] Xie X C, Fu G, Xue Y J Y and Jiang J 2018 Risk prediction and factors risk analysis based on IFOA-GRNN and apriori algorithms: Application of artificial intelligence in accident prevention *Process Safety and Environmental Protection* **122** 169-184.

[10] Wang T, Liu Y M, Gu X P and Qin X H 2018 Vulnerable lines identification of power grid based on cascading fault space-time graph *Proceedings of the CSEE* **20** p 5962-5972+6176.

[11] Zhang Y and Ma J L 2018 Fiber network communication transmission vulnerability assessment method based on accident chain *Laser Journal* **12** 138-142.

[12] Wang J, Zhang M G, Huang X and Chen J R 2019 Scenario analysis of road transportation accidents of inflammable and explosive hazardous chemicals *China Safety Science Journal* **05** 171-177.

[13] Deng Y L, Song L L, Zhou Z P and Gerard O 2017 An approach for understanding and promoting coal mine safety by exploring coal mine risk network *Complexity*.

[14] Jia W Z, Wang Y H, Su H M, Liu Y and Lin S 2019 Risk evaluation of rail vehicles based on risk network *Journal of Transportation Systems Engineering and Information Technology* **04** 143-148.

[15] Zhou Z P, Javier I and Li Q M 2014 Using network theory to explore the complexity of subway construction accident network (SCAN) for promoting safety management *Safety Science* **64** 127-136.