WILEY | Hindawi

Research Article

Analysis of Traffic Accident Based on Knowledge Graph

Liyan Zhang [iD],[1] **Min Zhang** [iD],[1] **Jiazhen Tang** [iD],[1] **Jian Ma** [iD],[1,2] **Xiaoke Duan** [iD],[1]
Juan Sun [iD],[3] **Xiaofei Hu,**[3] **and Suchuan Xu** [iD][3]

[1]*School of Civil Engineering, Suzhou University of Science and Technology, 1701 Binhe Road, New District, Suzhou 215011, China*
[2]*General Hospital of People's Liberation Army of China Eastern Theater Command, Nanjing, China*
[3]*School of Business, Suzhou University of Science and Technology, 99 Xuefu Road, Huqiu District, Suzhou 215009, China*

Correspondence should be addressed to Jian Ma; 9764634@qq.com

Traffic accident data include multidimensional dynamic and static factors such as "people, vehicles, roads, and environment" at the time of the accident, which is one of the important data sources for improving the traffic safety environment. Based on the case data of traffic accidents and the construction idea of knowledge graph, the knowledge demand, knowledge modeling, knowledge extraction, and knowledge storage of traffic accidents are analyzed in detail. Finally, the traffic accident knowledge graph is constructed. The visualization analysis of accident is carried out from four angles: accident portrait, accident classification, accident statistics, and accident correlation path. The visualized graphic network displayed by the traffic accident knowledge combines human cognition with machine cognition, which improves human's ability to understand massive and complicated data. The theoretical system of constructing traffic accident knowledge graph has certain reference significance in the follow-up research on the analysis of massive traffic accident data.

1. Introduction

The prosperity of the transportation industry not only brings positive benefits to the society but also brings negative impacts that are difficult to reconcile [1]. Among them, the frequent occurrence of road traffic accidents has become one of the important factors inhibiting the steady development of cities. Faced with the severe road traffic safety situation, it is urgent to carry out research on road safety analysis and active prevention and control of road risks. The traffic management departments have accumulated a large amount of accident data in their daily management work. How to discover and reuse the potential value of these data, and dig out the potential laws and inducing factors of accidents, has become a major research hotspot today [2]. In 2017, "the 13th Five-Year Plan for Road Traffic Safety" issued by the State Council pointed out that the comprehensive collection of traffic accident data can promote the improvement of traffic safety big data to provide data basis and theoretical analysis support for the improvement of traffic safety [3]. At present, the domestic traffic control departments have accumulated a large number of original accident data by adopting standardized accident information collection technology, including the specific data of "people, vehicles, roads, and environment" related to the accident. However, the data value has not been fully excavated and remains in the descriptive statistics of the four indicators of the accident. How to excavate the hidden value of traffic accident data to prevent and reduce the occurrence of traffic accidents, and combine the relevant advanced traffic safety technology [4–8] to put it into practical application instead of "empty talk," has become one of the key directions of current research.

There are many ways of traffic accident data mining. At present, there are three main research angles.

(1) Descriptive analysis of accident data based on traditional statistical analysis. The road traffic accident information system of China contains more than 60 items of accident data, which can describe the situation objectively and comprehensively situation when the accident occurs. Simple statistical analysis of the data collected in the system is the basic reference for China to formulate traffic safety management planning [9]. In addition, in 2005, the Shanghai Traffic Police Corps, together with Tongji University and the German Volkswagen Group, carried out the research on road traffic accidents in China and statistically analyzed the national standardized accident data collected. The mechanism of traffic accidents was summarized [10].

(2) Based on big data algorithm, the hidden value information of accident data is mined deeply from the point of view of data association and data collision. The literature [11] adopted the analysis method of data mining technology and multistandard decision-making method to mine the French traffic accident database BAAC, ranked the importance of the mining association rules by ELECTRE method, and selected the association rules with higher ranking as the basis for formulating accident prevention strategies and policies, so as to improve the road safety environment in France. The literature [12–14] used Apriori algorithm or FP growth algorithm to mine association rules among various factors in traffic accidents to provide data decision support for accident early warning and traffic safety management.

(3) Visual analysis of accident data is carried out based on data understanding. The combination of human cognition and machine cognition through the human–computer interface improves the human ability to understand huge and complex data. The literature [15] designed and implemented a multi-view association visualization analysis system based on the spatial semantic enhancement model of traffic accident data, revealing the spatial semantic mode of traffic accidents for users and contributing to in-depth analysis of the causes of traffic accidents.

Knowledge graph (KG) [16] is a new research method of data mining, which represents the mutually independent entities and their relationships in the objective world in the structured form of the graph and forms the basic units of the graph in the form of "entity–relation–entity" triple. The relationship is the link that connects the entity, and both the entity and the relationship can have attributes, thus forming a network structure [17, 18]. In short, through the graph structure pattern of triple [19], KG transforms the knowledge in the objective world into a structure that can be understood and processed by machines and can display intuitive and visual characteristics to be understood and reused by human beings. The literature [20] established a traffic knowledge graph based on multisource and heterogeneous data, which combines the four elements of "people, vehicles, roads, and

environment." At the same time, based on the relationship among three kinds of traffic events (traffic accident, traffic congestion, and traffic feedback), the traffic reason graph was established, and the recognition of text traffic incident and Weibo traffic incident was realized by using knowledge and reason graph. It provided a solution for finding traffic problems and early warning of traffic accidents. The literature [21] (2021) used Word2vec word vector model to extract and classify the keywords (accident features and accident cause attributes) in traffic accident text and generated the knowledge graph of traffic accident domain based on Neo4j and realized the visual analysis of it based on Gephi. Traffic accident data, as the basic data of traffic safety research, can provide multidimensional data support for traffic safety management decision-making. The integration of knowledge related traffic accident based on KG can effectively give full play to the value of data resources and inject knowledge support into traffic safety management decision-making, which has important research value to improve traffic safety.

Based on the structured traffic accident case data, this paper establishes the graph structure and visual traffic accident knowledge graph, that is, the traffic accident knowledge graph that integrates "people, vehicles, roads, and environment." The knowledge is stored by using Neo4j graph database. The multidimensional and multilevel analysis of accidents is realized by using Cyber sentences, including accident portraits, accident classification, accident statistics, and accident association path analysis.

2. The Definition of Knowledge Graph and Its Constituent Elements

Taking the word "knowledge graph" apart analyzes its meaning. First of all, "knowledge," from a philosophical perspective, is the achievements obtained by human beings from various ways of life and production, which come from the objective world, and is the systematic understanding of human beings to summarize, refine, and sublimate all kinds of facts, descriptions, information, and so on. From the "system of knowledge and wisdom of data information" described by Rowley J, that is, the DIKW system [22], as shown in Figure 1, the formation of "knowledge" goes through a process from data to information and then to the transformation of knowledge, which is the cognition of human beings after processing data. The ultimate destination of "knowledge" is the "wisdom" used by human beings, that is, the application of knowledge. Second, the meaning of "graph" is to form a network structure in the form of "graph." In the study of graph theory, in 1878, "graph" was first proposed by Sylvester [23]. A graph is composed of multiple nodes and multiple edges as shown in Figure 2, which shows a simple graph composed of five nodes and six edges. To sum up the above definition, the knowledge graph is to express the knowledge in the form of graph. The nodes of a graph represent concepts or entities, and the edges represent the relationship between nodes. With an objective fact that "Taiwan is a provincial administrative region of China" as an example, the semantic relationship of this fact can be
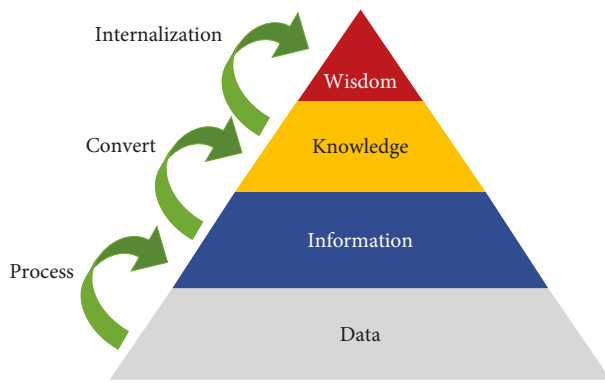
FIGURE 1: Schematic diagram of DIKW system.

FIGURE 2: A graph consisting of five nodes and six edges.

expressed as a language that can be understood by machines in the form of "China, provincial administrative region, Taiwan." Among them, "China" and "Taiwan" are two nodes, and the edge is used to indicate that the relationship between the two nodes is a "provincial administrative region."

Some domestic experts and scholars have also made some descriptions on the definition of knowledge graph. The literature [24] defined it as: "The knowledge graph describes concepts, entities and their relationships in the objective world through a structured way, expressing the information of the Internet into a form closer to the human cognitive world. It provides a better ability to organize, manage and understand the massive information on the Internet." The literature [25] defined it as: "Knowledge graph is essentially a knowledge base of semantic network [26], that is, a knowledge base with directed graph structure, in which the nodes of the graph represent entities or concepts, and the edges of the graph represent various semantic relationships between entities or concepts." The unique definition of knowledge graph has not been given in academic circles, but the commonness of these definitions is that entities and relations are the basic elements of knowledge graph.

(1) Entity: it refers to the concrete things that exist in the objective world, corresponding to the ontology in the semantic network. For example, in a traffic accident, the entity can be the name of the person involved in the accident, illegal behavior, license plate, a certain road, rainy day, and so on.

(2) Concept: it is also known as the type of entity, which is an abstract generalization of things that share common characteristics. For example, the concept of "person" is a summary of "the name, age, and illegal behavior of the person involved in the accident," in which "person name and illegal act" are the entity of "person."

(3) Relationship: it refers to a connection between different concepts or entities. For example, there is a "party to the accident" relationship between "accident" and "person."

In addition, attributes can be used to describe certain characteristics of an entity or relationship. For example, the attributes of age, gender, and so on, can be included in the entity of "name."
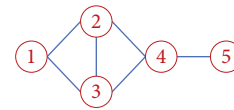
General KG and domain KG are the two major categories of current KG. The former, as its name implies, is a general domain-oriented and common-sense knowledge graph, which mainly serves web search and encyclopedic questions and answers, such as DBpedia [28], Yago [29], and Freebase [30], while the latter is oriented to a specific vertical professional field with professional knowledge, such as financial knowledge graph [31], medical knowledge graph [32], and so on. The traffic accident knowledge graph constructed in this paper is obviously aimed at the domain knowledge graph of the vertical professional field of transportation.

3. Construction of Traffic Accident Knowledge Graph

Traffic accident knowledge graph is a knowledge graph oriented to the professional field of transportation, and its construction process and objectives need to be determined according to the requirements of professional knowledge. Following the construction principle of "beginning with demand and ending from application," this paper divides the life cycle of constructing traffic accident knowledge graph into five stages, that is, knowledge demand, knowledge modeling, knowledge extraction, knowledge storage, and knowledge application as shown in Figure 3. Among them, knowledge modeling, knowledge extraction, and knowledge storage are the core links of constructing knowledge graph. Knowledge modeling means that based on the analysis of data sources, concepts and relational patterns are selected to form a knowledge structure that meets knowledge demand, that is, concept and relational schema design. Knowledge extraction is to extract knowledge elements from multi-source data. Knowledge storage is to store the acquired knowledge in the database for the application of knowledge.

As the data of this study are structured road traffic accident case data, the main problem is how to map the entities, relationships, and attributes contained in a larger number of data into the database completely. Based on the above life cycle diagram, the specific process of constructing the traffic accident knowledge graph is described as follows: first, according to the existing data sources, the knowledge demand for traffic accident analysis is determined. Then, knowledge extraction of entities, relationships, and attributes is carried out from structured data. All elements are written into the Neo4j graph database to complete the construction of the traffic accident knowledge graph. Finally, based on this KG, multidimensional visualization analysis of accident data is realized by the Cypher query statement.

3.1. Knowledge Demand Analysis of Traffic Accidents. The application value of knowledge graph in traffic accidents is mainly reflected in the following aspects:
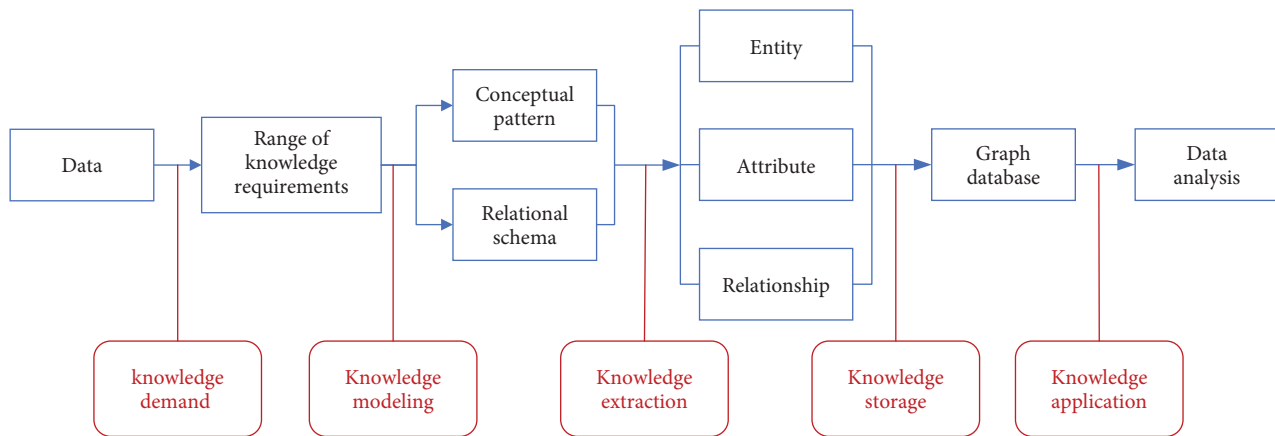
FIGURE 3: Life cycle diagram of traffic accident knowledge graph.

(1) Knowledge graph can play a comprehensive data supporting role in accident analysis. On the basis of the knowledge graph of traffic accidents, multilevel and multidimensional accident analysis can be carried out according to the knowledge semantic relation network, which broadens the ideas of accident analysis and accident prevention.

(2) The accident knowledge network formed by the knowledge graph can present all kinds of accident query results in a visual way. Given an entity, it can search for another entity along its relationship path and finally display the accident query results with a network diagram composed of entities and relationships, which is a visual approach of accident knowledge interpretation.

(3) Knowledge graph can realize the accident risk analysis with human as the core. The accident risk database of the driver is established by integrating the driver's accident record, traffic violation, driving vehicle, age, driving age, and other information into the map. Through the selection of risk characteristics and quantitative construction of accident risk early warning model, accident early warning for drivers, especially long-distance truck drivers, can reduce the occurrence of serious accidents.

At present, the knowledge graph is still in the exploratory research stage in the aspects of traffic accident knowledge integration and accident prevention. With the continuous development of artificial intelligence and big data technology, knowledge graph will play a huge advantage and wide application prospect in accident analysis, accident risk prediction, traffic management decision support, and other aspects.

The data used in this paper are traffic accident case data, which is structured relational data. The data cover the period from January to December 2017, with 9,941 pieces of data. Each row of data represents the information recorded by the traffic police after an accident, including the accident number, the location of the accident (road), the time point of the accident, the type of road, the jurisdiction to which the road belongs, the cause of the accident, the form of the

accident, the identity of the parties involved in the accident, negligence and illegal behavior, the license plate information of the vehicle involved in the accident, and the type of vehicle.

Combining the above advantage analysis and data source structure, the traffic accident knowledge graph constructed in this paper aims to effectively mine the value information contained in the traffic accident case data. A graph database containing traffic accident characteristic factor information and accident result information is established, based on which multidimensional and multilevel and visual accident analysis is realized, such as accident portrait, accident distribution, accident statistics, and so on.

3.2. Knowledge Modeling Design of Traffic Accident. Knowledge modeling design includes conceptual pattern and relational pattern, which is for the abstract mapping of the concepts and relationships of real things. A concept is an abstract description of a certain type of entity in the objective world. Entities of the same type may have different attributes. Relationship is to explain the existence of some kind of link between entities, which is diverse.

According to the demand of accident knowledge, this paper initially forms the core concept of traffic accident and then combines professional knowledge to expand the relationship between entities, so as to lay a good foundation for knowledge extraction [33]. The core concepts in the field of traffic accidents established in this paper are shown in Table 1. The field relationship of traffic accidents is shown in Table 2.

In order to distinguish concepts, relations, and attributes, this paper adopts different distinguishing marks to express them. The concept labels are in English and uppercase. The relationship labels are in English and the content words are in uppercase. The attribute labels are in English, and the content words are in lowercase, such as concept tag: ACCIDENT, PEOPLE, VEHICLE, ENVIRONMENT, etc.; relationship tag: Located_in, Jurisdiction_over, Weather, etc.; attribute tag: accident_time, person_age, person_gender, etc.

TABLE 1: Core concepts in the field of traffic accidents.

Name of concept	Coding of concepts	Optional attributes of entities in a concept
Accident itself	ACCIDENT	Time of occurrence (year, month, day, hour, minute, second), number of deaths, number of injuries, direct economic losses, etc.
People involved	PEOPLE	Age, sex, negligent act, illegal act, etc.
Vehicles involved	VEHICLE	Model, color, brand, etc.
Traffic environment	ENVIRONMENT	Weather, lighting, etc.
Road	ROAD	The level of the road, the length of the road, condition of road surface, etc.
Traffic management department	DEPARTMENT	Qualitative description
Cause of accident	CAUSE	Qualitative description
Nature of accident	NATURE	Qualitative description
Determination of accident liability	JUDGMENT	Qualitative description

TABLE 2: Traffic accident area relation table.

Name of relationship	Coding of relationships	The entity connected in front	A rear-facing entity
Location of accident	Located_in	The ID of the accident	Name of road
Jurisdiction of road	Jurisdiction_over	Name of department	Name of road
Weather of accident	Weather	The ID of the accident	Sunshine, rain, fog, etc.
Nature of accident	Character	The ID of the accident	Collision moving vehicles, scraping pedestrians, fixed objects, etc.
Cause of accident	Due_to	The ID of the accident	Violation of traffic signals, failure to yield according to regulations, etc.
Party to the accident	Person	The ID of the accident	The unique identification ID of the person involved
The mode of transportation of the person involved	Transportation	The ID of the person involved	Vehicle
Liability of the person involved in the accident	Responsibility	Party to the accident	Voluntary consultation, primary responsibility, secondary responsibility, coresponsibility, etc.

According to the concept and relationship description in the field of traffic accidents, the conceptual and relational pattern structure diagram of traffic accident knowledge graph finally formed is shown in Figure 4.

3.3. Knowledge Extraction of Traffic Accident. Entity is an objective thing in the real world, a concrete example in the concept layer, and a key knowledge element in the knowledge graph. Entity extraction is a process of recognition, that is, entities with specific meaning are identified in some way. The entity scope in the field of traffic accidents studied in this paper is mainly the concepts listed in the conceptual pattern design.

In general, entities and relationships have strong personalized features, and they are constantly increasing and updating, so knowledge extraction needs to be carried out according to the given corpus and features. Knowledge extraction technology refers to the technology of obtaining entities and relationships from multisource and heterogeneous data, which is not only the basis of constructing knowledge graph but also the derivation of big data technology. With the explosive growth of data, how to obtain useful knowledge from the data is the current technical difficulty. According to the different data source structure, it can be divided into knowledge extraction of structured data, semistructured data and unstructured data. The corresponding extraction method is shown in Figure 5.

The accident data based on this paper are a structured traffic accident data table, which is mapped into RDF triple by direct mapping method, so as to realize the knowledge extraction of entities and relations. According to the concepts and relationship patterns designed above, a total of 17182 entities and 51992 relationships are extracted. The extraction results of different types of entities and relationships are shown in Table 3.

3.4. Storage and Visualization of Traffic Accident Knowledge Graph Based on Neo4j. Neo4j is a graph database based on Java, which is used to store graph structure data of entities and relationships. Two entities and relations form a knowledge unit, and the relationship is used to connect two nodes, which is directional. Both entities and relationships can have attributes, which have names and various values. Tags are used to distinguish between different types of entities and relationships.

CREATE statement, LOADCSV, and Neo4j import are currently the main methods to import triple data into Neo4j in batches [33]. The running speed, advantages and
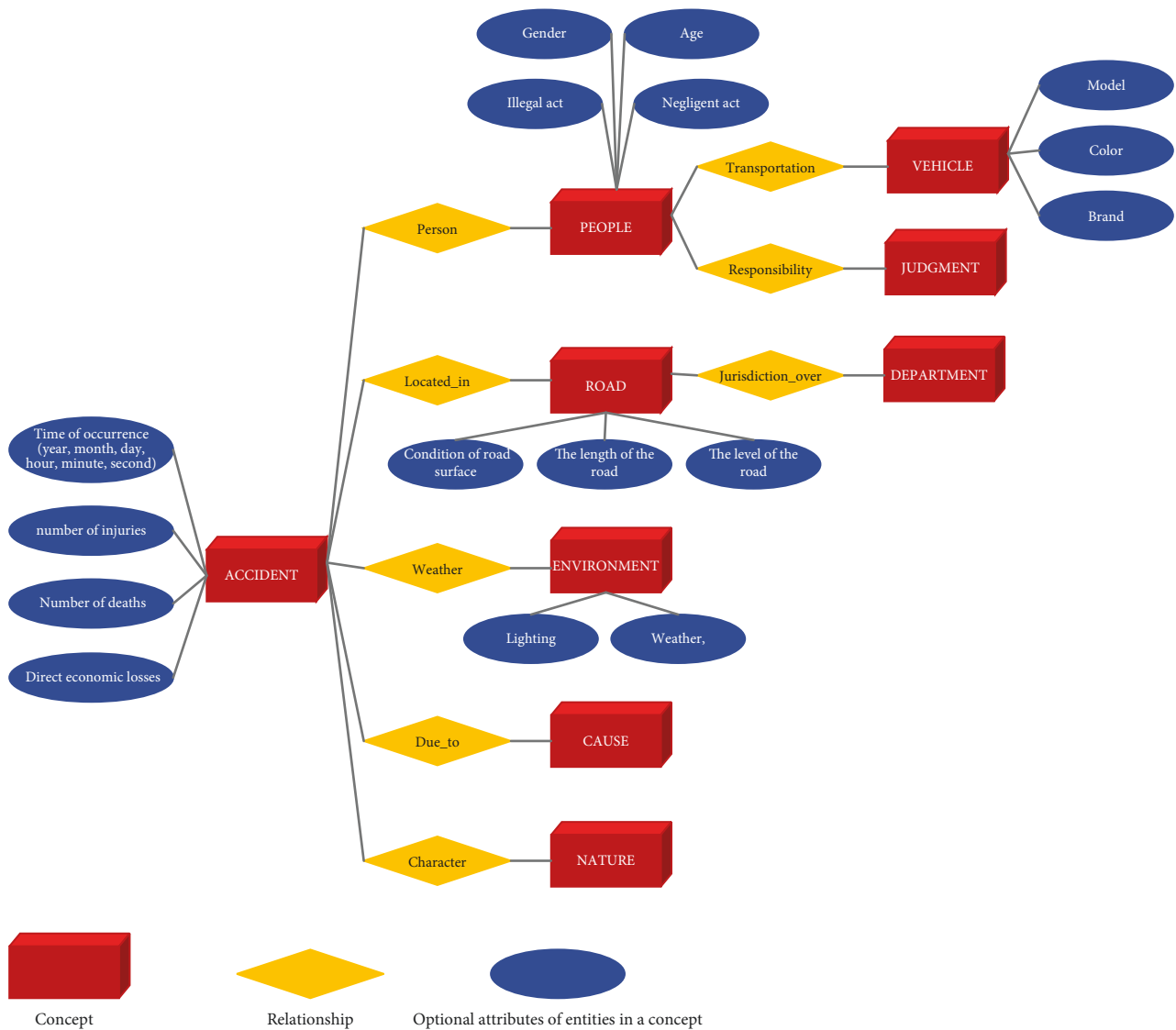
Figure 4: Conceptual and relational schema structure diagram of traffic accident knowledge graph.
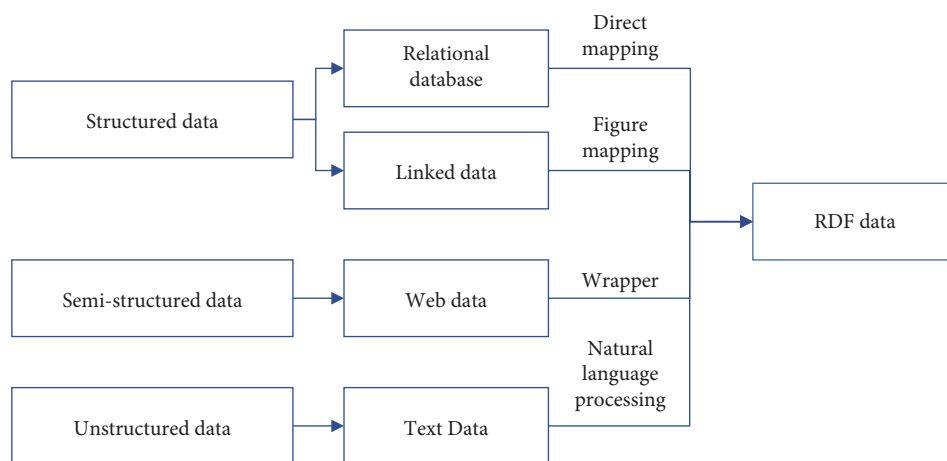

Figure 5: Schematic diagram of knowledge extraction method for multisource heterogeneous data.

TABLE 3: Extraction results of different types of entities and relationships.

Category	The encoding of the entity or relationship type in the concept	The number of entities or relationships	Total
Entity in concept	ACCIDENT	5083	
	PEOPLE	9152	
	VEHICLE	68	
	ROAD	2453	
	ENVIRONMENT	4	17182
	DEPARTMENT	8	
	CAUSE	34	
	NATURE	13	
	JUDGMENT	7	
Relationship	Located_in	5083	
	Weather	5083	
	Person	9911	
	Due_to	5092	
	Character	5083	51992
	Jurisdiction_over	2604	
	Transportation	9572	
	Responsibility	9564	

TABLE 4: Comparison of methods for importing data in batches into Neo4j.

Methods	Running speed	Advantage	Disadvantage	Scope of application
CREATE	Slow (1 k nodes/s)	The statement is simple, easy to use and understand and can realize the real-time insertion of new data	Slow speed	1 – 1 w nodes
LOAD CSV	General (5 k nodes/s)	In addition to the advantages described in the CREATE, you can load not only local CSV, but also remote CSV	The data must be converted to CSV	1 w – 10 w nodes
Neo4j import	Fast (tens of thousands of nodes/s)	The data import speed is fast and the running time is short	Neo4j must be disabled during the import process; you cannot reimport new data using import in an established database	Over ten million nodes

disadvantages, and of application of the three data batch import methods are shown in Table 4.

This paper mainly uses the method of Neo4j import is mainly used to realize the batch import of data, that is, the py2neo module package in Python is used to realize the rapid import of CSV files and the creation of nodes and relationships. Finally, the traffic accident domain knowledge, mainly involving accident cases, is formed as shown in Figure 6.

4. Case Analysis of Traffic Accidents Based on Knowledge Graph

The knowledge application of domain knowledge graph should be based on the needs of this field. This paper constructs a knowledge network around the knowledge elements of "people, vehicles, roads, environment, and accident results," which makes the accident analysis results visual. Through the analysis of traffic accidents, the summary of the occurrence law of accidents and the multidimensional classification and statistics of accidents are realized, which provides data support for improving the level of traffic safety management.

Cypher is a declarative query language in Neo4j, which can query and update data efficiently. The commonly used statements include MATCH (used to match the pattern of the graph), WHERE (used with MATCH to add constraints to the pattern of the graph), and RETURN (to determine the type of results returned, which can be entities and relationships of the graph structure, or tables). The traffic accident analysis process based on knowledge graph is shown in Figure 7.

4.1. Accident Portrait.
Enterprise portraits, customer portraits, product portraits, and so on are one of the major applications of knowledge graph. This kind of portrait is that the knowledge graph fuses multisource data to make a more comprehensive description of the characteristics of entity objects and presents them in a visual way. This paper takes a single road traffic accident as the center and comprehensively describes the accident situation through its associated entities, relationships, and attributes, which constitutes an "accident portrait."

The traffic accident portrait based on the knowledge graph can describe the location, time, cause of the accident, basic information of the accident parties, negligence and illegal behavior, and the final handling of the accident, namely, the determination of responsibility, and so on. Taking an accident portrait as an example is as shown in Figure 8. At 1 : 25 on January 1, 2017, an intervehicle
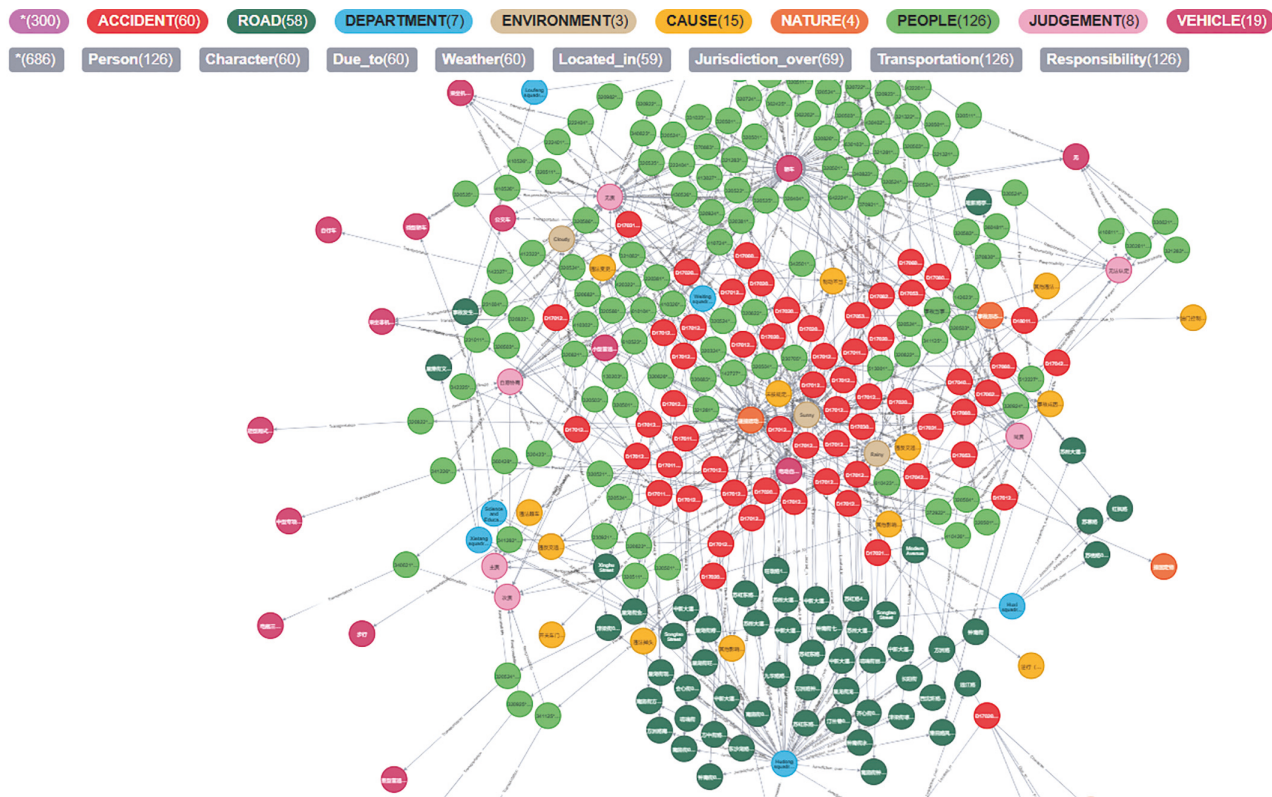
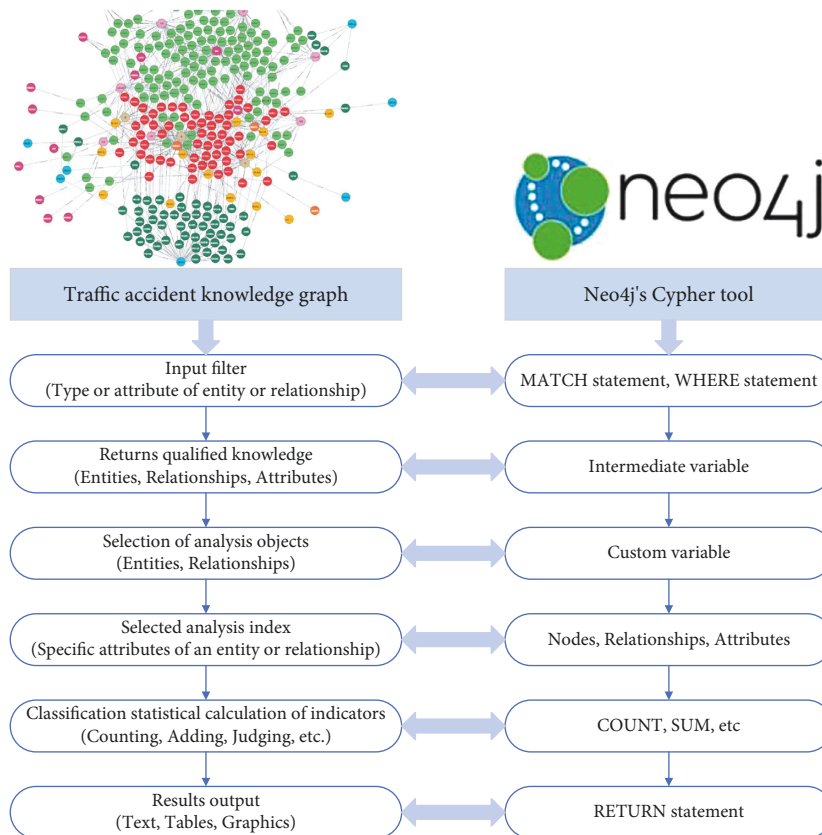FIGURE 6: Visualized map of traffic accident knowledge.

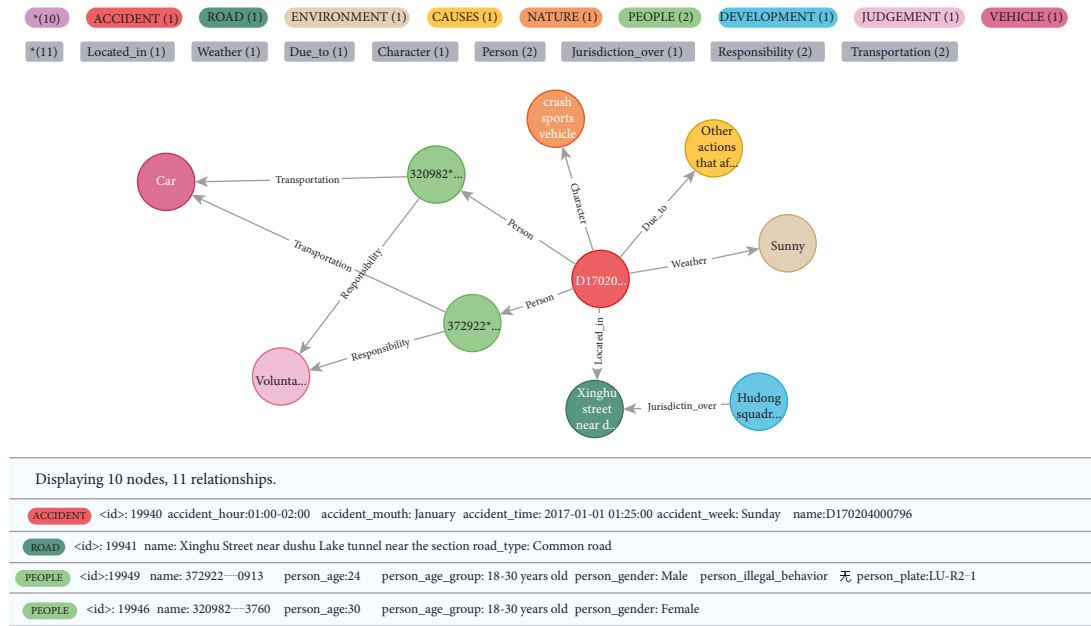FIGURE 7: Traffic accident analysis process based on knowledge graph.

FIGURE 8: An accident portrait based on the traffic accident knowledge graph.

collision accident occurred near Xinghu Street near Dushuhu Tunnel due to other safety actions. The accident was handled by the Hudong Squadron, and both parties involved in the accident voluntarily negotiated the compensation for the accident. From the property column of the person responsible for the accident, we can see whether the person has violated traffic laws and regulations and whether there are other negligent acts that affect traffic safety.

4.2. Accident Classification. All accident cases are carried on the accident classification inquiry according to a certain classification standard. This classification standard can be called accident characteristic dimension, including accident location, accident form, accident cause, and so on. According to the relationship between all kinds of entities associated with the feature dimension, the query statement conforming to Cypher is constructed.

In the traffic accident knowledge graph, in order to realize the accident classification query according to the accident location, namely, the name of road, each name in the concept "ROAD" is regarded as an entity. There is a relationship "Located_in" between "ACCIDENT" and "ROAD." The returned result is all the relationship paths between the two concepts. As a result, the Cypher query sentence is as follows:

"MATCH $p =$ (n: "ACCIDENT")- [r: Located_in] - > (m: "ROAD") RETURN p"

The returned results that visually show the distribution of the accident location are as shown in Figure 9. As can be seen from the picture, there are many accidents on Donghuan Road. Therefore, it is necessary to further investigate the hidden dangers of this road.

4.3. Accident Statistics. Accident statistical indicators usually have statistical indicators and classification indicators. Statistical indicators usually refer to four indicators commonly used in traffic accident data statistics (the number of accidents, the number of deaths, the number of injuries, and direct economic losses). The classification indicators include the location of the accident, the time of the accident, and so on. Users can select indicators as required to construct Cypher query statements. The returned result of "RTURN" is defined by using the "COUNT" and "SUM" commands alone or in combination. If the result users need to return is a table, they can use the "ORDERBY" command to sort the query results.

4.3.1. The Location Analysis of Accident-Prone Places. The number of accidents is selected as the statistical index, the road as the classification index. The name of road is selected as an entity. There is a relationship "Located_in" between the concept "ACCIDENT" and "ROAD." The results need to be returned as "the name of road" and "the type of road," and the number of accidents is counted and sorted in descending order. From this, the Cypher query sentence can be constructed as follows:

"MATCH (n: "ACCIDENT")-[r: Located_in] -> (m: "ROAD") RETURN m.name AS Road, *m*. road_type AS Road_type, COUNT (*) AS Number ORDER BY Number DESC."

Displaying the returned results in tabular form is as shown in Figure 10. It can be seen that the top 10 roads with more accidents include Xinghu Street, Donghuan Road, Modern Avenue, Jinji Lake Avenue, Fengting Avenue, Zhongxin Avenue East, Zhongyuan Road, Songtao Street, Loujiang Avenue, and Weisheng Road.
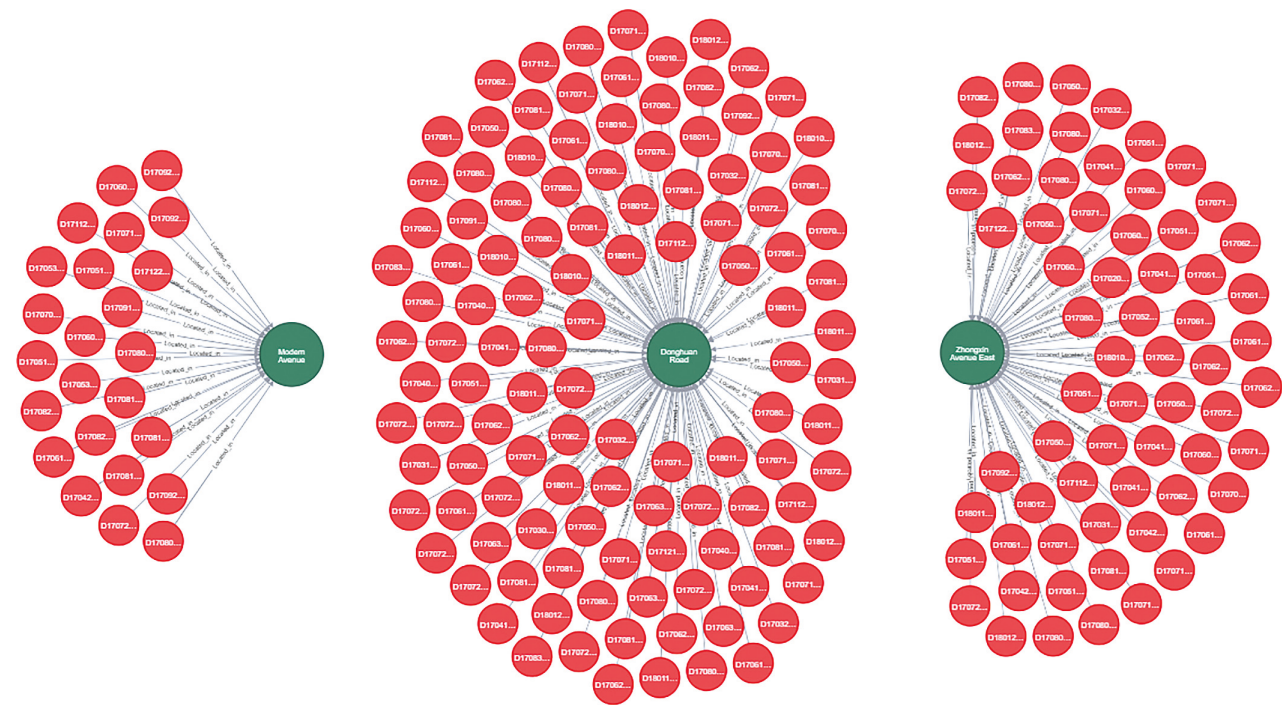
FIGURE 9: Visual display of accident classification by location.

	ROAD	Road_type	Number
1	"Xinghu Street"	"Common road"	160
2	"Donghuan Road"	"Common road"	116
3	"Modem Avenue"	"Common road"	102
4	"Jinji Lake Avenue"	"Common road"	93
5	"Futing Avenue"	"Common road"	66
6	"Zhongxin Avenue East"	"Common road"	65
7	"Zhongyuan Road"	"Common road"	57
8	"Songtao Street"	"Common road"	53
9	"Loujiang Avenue"	"Common road"	51
10	"Weisheng Road"	"Common road"	47

FIGURE 10: Number of accidents per road—top 10 roads.

	Month	Number
1	"March"	914
2	"May"	810
3	"April'	784
4	"June"	652
5	"July"	338
6	"September"	325
7	"November"	308
8	"December"	231
9	"February"	212
10	"October"	197
11	"January"	167
12	"August"	145

(a)

	Week	Number
1	"Friday"	838
2	"Thursday"	802
3	"Wednesday'	745
4	"Monday"	739
5	"Tuesday"	739
6	"Saturday"	636
7	"Sunday"	584

(b)

	Hour	Number
1	"08:00~09:00"	619
2	"09:00~10:00"	395
3	"18:00~19:00"	382
4	"17:00~18:00"	381
5	"07:00-08:00"	375
6	"19:00~20:00"	295
7	"16:00~17:00"	288
8	"15:00~16:00"	268
9	"20:00~21:00"	244
10	"21:00~22:00"	229
11	"10:00~11:00"	228
12	"13:00~14:00"	223
13	"12:00~13:00"	220
14	"14:00~15:00"	217
15	"11:00~12:00"	215
16	"22:00~23:00"	140
17	"23:00~24:00"	90
18	"06:00~07:00"	84
19	"00:00~01:00"	55
20	"02.00~03:00"	34
21	"01:00~02:00"	32
22	"05:00~06:00"	31
23	"04:00~05:00"	21
24	"03:00~04:00"	17

(c)

FIGURE 11: Query results of time factor in accident-prone areas, (a) statistics by month, (b) statistics by week, and (c) statistics by hour.

	Gender	Age	Population
1	"Male"	"18-30 years old"	2389
2	"Male"	"31 '40 years old"	2158
3	"Male"	"41*50 years old'	1159
4	"Female"	"31'40 years old'	888
5	"Female"	"18-30 years old"	799
6	"Male"	"51-60 years old'	636
7	"Female"	"41-50 years old"	417
8	"Female"	"51-60 years old"	269
9	"Unknown"	"Unknow"	229
10	"Male"	"61-70 years old"	226
11	"Female"	"61-70 years old"	153
12	"Male"	"Under 17 years old"	66
13	"Female"	"Under 17 years old"	55
14	"Male"	"Over 71 years old'	46
15	"Female"	"Over 71 years old"	22

FIGURE 12: Gender and age distribution of accident parties.

4.3.2. Analysis of Time Elements of Accident-Prone Places. The time elements of accident-prone places can be year, month, and day, or they can be subdivided into hours, minutes, and seconds. Since the graph constructed in this paper stores the time element as the attribute of the accident, the query of it needs to construct the Cypher statement according to the characteristics of the attribute. This section

selects the month, week, and time of the accident for analysis.

The number of accidents is sorted in descending order according to the attributes of each entity in the concept "ACCIDENT" including month, week, and time. The results to be returned are "month," "week," "hour," and "the number of accidents." The number of accidents is counted
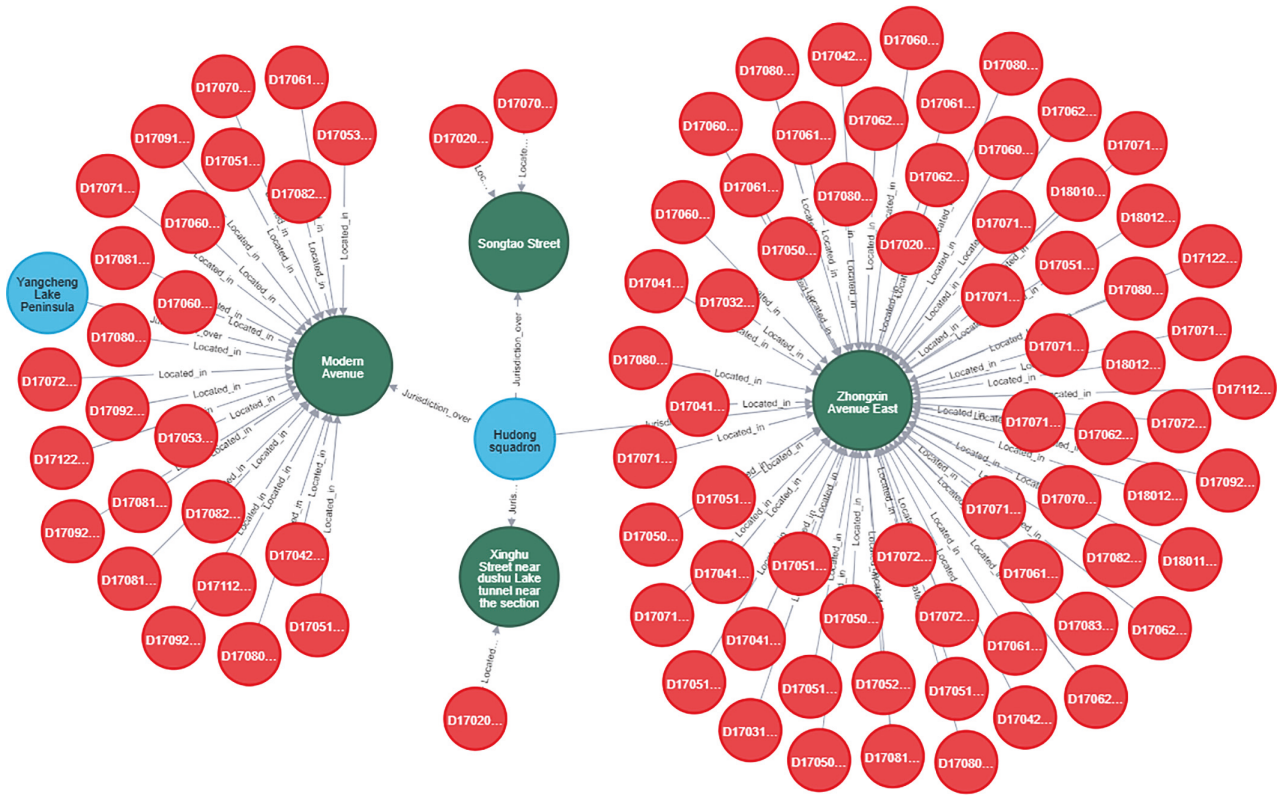
FIGURE 13: Visual display of accident association path.

and sorted in descending order. As a result, the Cypher query sentences can be constructed as follows:

> "MATCH (n: "ACCIDENT) RETURN n. accident_mouth AS Month, COUNT (*) AS Number ORDER BY Number DESC"

> "MATCH (n: "ACCIDENT") RETURN n.accident_week AS Week, COUNT (*) AS Number ORDER BY Number DESC"

> "MATCH (n: "ACCIDENT") RETURN n.accident_hour AS Hour, COUNT (*) AS Number ORDER BY Number DESC"

The returned results are displayed in the form of table pairs and shown in Figure 11.

From the monthly statistics of traffic accidents (Figure 11(a)), March, May, and April were the top three months with more than 750 accidents per month and an average of more than 25 accidents per day. This is mainly due to the fact that March to May is the month of rapid economic development after the Spring festival. Economic development can be inseparable from the development of the transportation industry, while excessive traffic flow caused accidents. The lowest number of accidents occurred in August, which is mainly due to the extreme heat and small road traffic flow.

From the weekly statistics of traffic accidents (Figure 11(b)), the number of accidents on weekdays (Monday to Friday) was basically maintained at more than 700 accidents, while the number of traffic accidents on Sundays is the smallest. This is mainly due to the large commuting traffic on weekdays.

From the time statistics of traffic accidents (Figure 11(c)), the occurrence of traffic accidents is mainly concentrated in the morning and evening peak, which corresponds to the time period from 7 : 00 to 9 : 00 and 17:00 to 19:00. This is mainly due to the rush hour in the morning and evening; drivers and pedestrians who were in a hurry tend to ignore the safety of traffic travel and coupled with the increase of traffic flow, passenger flow, and the conflict between them, which can easily lead to accidents. In addition, the morning and evening peak coincides with the alternating time of day and night, and the switching between street lights and natural light at night can easily lead to deviation in the driver's perspective and line of sight, and then misjudge some traffic conditions. Wrong driving behavior decisions are also the main cause of accidents.

4.3.3. Analysis of the Characteristics of Accident Parties.
The characteristics of accident parties mainly refer to gender and age. According to the goal of accident analysis, the Cypher sentence can be constructed as follows:

> "MATCH (n: "PEOPLE") RETURN n. person_gender AS Gender, n. person_age_group AS Age, COUNT (*) AS Population ORDER BY Population DESC"

The results returned are as shown in Figure 12. Except for people with unknown gender records, male accident victims aged 18 to 50 are the most accident-prone, which was

more than twice as many as female accident victims between 18 and 50 years old. Although male drivers made up a high proportion of all drivers, their greater risk of accidents was also related to their aggressive driving behavior.

4.4. Analysis of Accident Association Paths. There may be one or more intermediate entities among different entities in the knowledge graph, in which an association path is formed by the relationship. Through the analysis of the associated path of accidents, it can find that there are some related accident groups. By focusing on the analysis of intermediate entities, problems existing in traffic management can be found.

The length of the associated path can be represented by the number of intermediate entities, as shown in the following formula:

$$L = N + 1. \tag{1}$$

In the formula, L represents the length of the associated path and N represents the number of intermediate entities, $N \geq 0$.

If the length of an associated path is 2, there is an entity in the middle. The concept of the entity at the beginning and end of the path is "ACCIDENT" and "DEPARTMENT," respectively, and the concept of intermediate entity is "ROAD." From this, the Cypher statement can be constructed as follows:

"MATCH $p = $ (n1: "ACCIDENT")—[r1]—(n2: "ROAD")—[r2]—(n3: "DEPARTMENT") RETURN p"

The results returned are as shown in Figure 13. It can be clearly seen from the picture that the number of accidents associated with the road "Zhongxin Avenue East" is the largest, and the "Hudong Squadron" of the area to which the road belongs should conduct a special investigation of its traffic safety risks.

5. Conclusion and Prospect

To study the analysis method for massive traffic accident data, based on the knowledge graph, this paper constructs a traffic accident knowledge graph, which integrates the four elements of "people, vehicle, road, and environment." The knowledge hidden in a large number of structured traffic accident case data stored by the traffic management department is effectively acquired and reused by using the knowledge graph, and the multidimensional and multilevel analysis of traffic accident data is realized. The visual mesh graph is used to directly show the relationship between all kinds of traffic accident knowledge. The research results obtained not only are helpful for researchers to understand the characteristics of traffic accidents and the relationship between causative factors in a more intuitive way but also can provide direct and effective knowledge support and decision-making basis for traffic management departments to implement reasonable traffic management measures. It is helpful to the prevention of traffic accidents and the overall improvement of traffic safety environment, so it has

important application value. In addition, the method system adopted in constructing traffic accident knowledge graph enriches the theory of traffic data mining in theory and has a certain theoretical research significance.

In the follow-up research, the following points are worth paying attention to:

(1) The knowledge scope of traffic accidents is wide. This paper mainly takes the structured traffic accident case data of specific regions provided by the traffic management department as the knowledge source and constructs the knowledge graph centering on "people, vehicles, roads, and environment." However, due to the single source of knowledge, the constructed KG has certain limitations and does not have universal applicability. In addition to structured data, the carriers of traffic accident-related knowledge exist in a large number of unstructured text records, web pages, pictures and videos. Knowledge extraction from multisource heterogeneous data is the focus of research in the next stage.

(2) The traffic knowledge graph of superior quality can provide comprehensive and reliable knowledge support for various decision-making needs, such as travel decision, safety management decision, and so on. Its application scenarios and values are considerable. This paper only studies one branch of traffic field, namely traffic accident, and only studies the application value of knowledge graph from the angle of accident analysis. The traffic branch studied by knowledge graph and the application value of knowledge support need to be further expanded and excavated.

(3) Due to the data source whose no latitude and longitude coordinates, only the accident frequency calculated by traffic accident knowledge graph is used to determine accident-prone roads, which has certain limitations. In the follow-up, the geographical coordinates corresponding to the accident location will be integrated into the knowledge graph, and the graph algorithm [34–36] will be used to further improve the accuracy of accident-prone road determination.

Data Availability

Data are available on request to the corresponding author.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Liyan Zhang, Min Zhang, and Jiazhen Tang contributed equally to this work and should be considered cofirst authors.

Acknowledgments

References

[1] E. Suryani, R. A. Hendrawan, P. F. E. Adipraja, and R. Indraswari, "System dynamics simulation model for urban transportation planning: a case study," *International Journal of Simulation Modelling*, vol. 19, no. 1, pp. 5–16, 2020.

[2] Y. H. Chun and M. K. Cho, "An empirical study of intelligent security analysis methods utilizing big data," *Journal of Logistics, Informatics and Service Science*, vol. 9, no. 1, pp. 26–35, 2022.

[3] Z.-L. Dang, "A brief discussion on the relevant requirements of the 13th five-year plan of road traffic safety in the production of road transportation safety [J]," *Architectural Engineering Technology and Design*, no. 32, p. 2021, 2018.

[4] L.-Y. Zhang, M. Zhang, J. Ma, and J. Ge, "A data-driven control strategy for urban express ramp," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–12, Article ID 6540972, 2021.

[5] L.-Y. Zhang, X.-K. Duan, J. Ma, M. Zhang, Y. Wen, and Y. Wang, "Mechanism of road capacity under different penetration scenarios of autonomous vehicles," *International Journal of Simulation Modelling*, vol. 21, no. 1, pp. 172–183, 2022.

[6] S. Christalin Nelson, M. Tapan Kumar, and G. L. Prakash, "A novel optimized LSTM networks for traffic prediction in VANET," *Journal of System and Management Sciences*, vol. 12, no. 1, pp. 461–479, 2022.

[7] A. Riedel, J. Gerlach, M. Dietsch et al., "A deep learning-based worker assistance system for error prevention: case study in a real-world manual assembly," *Advances in Production Engineering & Management*, vol. 16, no. 4, pp. 393–404, 2021.

[8] Lu Zhang, Y. Yan, W. Xu, J. Sun, and Y. Zhang, "Carbon emission calculation and influencing factor Analysis based on industrial big data in the "double carbon" era," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–12, Article ID 2815940, 2022.

[9] H.-Y. Xu and G. Cai, "Design of road traffic accident handling system[J]," *China Transportation Information Industry*, vol. 1, 2005.

[10] H.-Y. Wang and Y. Wang, "Multidisciplinary and pluralistic research on road traffic accidents[C]," in *Proceedings of the China International Conference on Automotive Safety Technology and Annual Meeting of Automotive Safety Technology of the Chinese Society of Automotive Engineering*, Chinese Society of Automotive Engineering, Changchun China, July 2007.

[11] F. Mazouri, M. C. Abounaima, and K. Zenkouar, "Data mining combined to the multicriteria decision analysis for the improvement of road safety: case of France[J]," *Journal of Big Data*, vol. 6, no. 1, 2019.

[12] D. H. Yang, "Traffic accident association analysis based on data mining technology[J]," *Digital Technology and applications*, no. 10, p. 230, 2017.

[13] Y. Wan, *Research on Vehicle Accident Early Warning Method Based on Urban Traffic Accident data[D]*, Wuhan University of Technology, China, 2020.

[14] F. Su and Q. Yuan, "Analysis of association rules of urban highway traffic accidents[J]," *Journal of Wuhan University of Technology*, vol. 42, no. 04, p. 313, 2020.

[15] Y.-T. Luo, L. Liu, X.-Y. Liu, C.-S. Yin, J.-G. Chen, and W.-J. Xie, "Visual analysis of urban traffic accident data under spatial semantic enhancement[J]," *Chinese Journal of Image and Graphics*, vol. 24, no. 12, pp. 2279–2290, 2019.

[16] A. Singhal, *Official Google Blog: Introducing the Knowledge Graph: Things, Not strings[EB/OL]*, Chennai, Tamil Nadu, 2012.

[17] Y.-H. Xiao, *Knowledge Graph: Concepts and Techniques[M]*, Publishing House of Electronics Industry, Beijing, 2020.

[18] H.-Q. Huang, J. Yu, X. Liao, K. Murphy, and V. Tresp, "Review on knowledge graphs[J]," *Computer system applications*, vol. 28, no. 6, p. 1, 2019.

[19] N. Shadbolt and N. Gibbins, "Resource description framework (RDF)[J]," *Encyclopedia of Gis*, pp. 6–19, 2009.

[20] X.-R. Sun, Y. Meng, and W.-L. Wang, "Weibo traffic incident recognition based on knowledge graph and target detection," *Data Analysis and knowledge Discovery*, vol. 4, no. 12, Article ID 136147, 2020.

[21] Y.-H. Cheng g, J.-Q. Zhang, J.-C. Li, and A. Zhang, "Visual mining and analysis method of accident text data in transportation industry[J]," *Computer Engineering and applications*, vol. 57, no. 21, Article ID 116122, 2021.

[22] J. Rowley, "The wisdom hierarchy: representations of the DIKW hierarchy[J]," *Journal of Information Science*, 2007.

[23] J. J. Sylvester, "On an application of the new atomic theory to the graphical representation of the invariants and cbq, with three appendices," *American Journal of Mathematics*, vol. 1, no. 1, p. 64, 1878.

[24] J.-Z. Li and L. Hou, "A review of research on knowledge graph [J]," *Journal of Shanxi University (Natural Science Edition)*, vol. 40, no. 03, p. 454, 2017.

[25] G.-L. Qi, H. Gao, and T.-X. Wu, "Research progress of knowledge graph[J]," *Intelligence Engineering*, vol. 3, no. 01, pp. 4–25, 2017.

[26] S. Y. Xu and B. Raahemi, "A semantic-based service discovery framework for collaborative environments," *International Journal of Simulation Modelling*, vol. 15, no. 1, pp. 83–96, 2016.

[27] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, and Z. Ives, "DBpedia: a nucleus for a web of open data," in *Proceedings of the 6th International The semantic Web and 2nd Asian Conference on Asian Semantic Web Conference (ISWC'07/ASWC'07)*, Busan, Korea, November 2007.

[28] F. M. Suchanek, G. Kasneci, and G. Weikum, "YAGO: a core of semantic knowledge unifying wordnet and wikipedia," in *Proceedings of the International Conference on World Wide Web*, NY City, June 2007.

[29] B. K. Freebase, "A collaboratively created graph database for structuring human knowledge[J]," *Proc. SIGMOD*, vol. 08, 2008.

[30] Y. Hu Yang, H.-F. Yan, and K. Chen, "Joint entity and relationship extraction algorithm for financial knowledge graph [J]," *Journal of Chongqing University of Technology (Natural Science)*, vol. 34, no. 05, p. 139, 2020.

[31] K.-Q. Yuan, Y. Deng, D.-Y. Chen, B. Zhang, and K. Lei, "Construction technology and research progress of medical knowledge graph[J]," *Computer Application Research*, vol. 35, no. 07, pp. 1929–1936, 2018.

[32] H. S. Moon, H. M. Lee, and S. T. Chung, "Design system of driving scenario by applying static and dynamic stimuli in VR driving simulator based on text network analysis," *Journal of Logistics, Informatics and Service Science*, vol. 9, no. 1, pp. 213–233, 2022.

[33] Y.-C. Zou, *Construction and Application of Traffic Knowledge Graph Based on Multi-Source Data fusion[D]*, Dalian University of Technology, China Liaoning, Dalian, 2020.

[34] R. Lu, Z.-P. Cai, and S. Zhao, "A survey of knowledge reasoning based on KG," *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 5, Article ID 052058, 2019.

[35] W. Xu, H. Y. Sun, A. L. Awaga, Y. Yan, and Y. J. Cui, "Optimization approaches for solving production scheduling problem: a brief overview and a case study for hybrid flow shop using genetic algorithms," *Advances in Production Engineering & Management*, vol. 17, no. 1, pp. 45–56, 2022.

[36] Y. J. Wang, N. D. Wang, S. M. Cheng et al., "Optimization of disassembly line balancing using an improved multi-objective Genetic Algorithm," *Advances in Production Engineering & Management*, vol. 16, no. 2, pp. 240–252, 2021.