

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.Doi Number

# Interactive Analysis of Epidemic Situations Based on a Spatiotemporal Information Knowledge Graph of COVID-19

BINGCHUAN JIANG<sup>1</sup>, XIONG YOU<sup>1</sup>, KE LI<sup>1</sup>, TINGTING LI<sup>1</sup>, XIAOJUN ZHOU<sup>1</sup>, LIHENG TAN<sup>1</sup>

<sup>1</sup>Institute of Geospatial Information, Strategic Support Force Information Engineering University, Zhengzhou 450052, China

Corresponding author: XIONG YOU (e-mail: youarexiong@163.com).

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 41801319.

**ABSTRACT** In view of the lack of data association in spatiotemporal information analysis and the lack of spatiotemporal situation analysis in knowledge graphs, this paper combines the semantic web of the geographic knowledge graph with the visual analysis model of spatial information and puts forward the comprehensive utilization of the related technologies of the geographic knowledge graph and big data visual analysis. Then, it realizes the situational analysis of COVID-19 (Coronavirus Disease 2019) and the exploration of patient relationships through interactive collaborative analysis. The main contributions of the paper are as follows. (1) Based on the characteristics of the geographic knowledge graph, a patient entity model and an entity relationship type and knowledge representation method are proposed, and a knowledge graph of the spatiotemporal information of COVID-19 is constructed. (2) To analyse the COVID-19 patients' situations and explore their relationships, an analytical framework is designed. The framework, combining the semantic web of the geographic knowledge graph and the visual analysis model of geographic information, allows one to analyse the semantic web by using the node attribute similarity calculation, key stage mining, community prediction and other methods. (3) An efficient epidemic prevention and anti-epidemic method, which is of referential significance, is proposed. It is based on experiments and the collaborative analysis of the semantic web and spatial information, allowing for real-time situational understanding, the discovery of patients' relationships, the analysis of the spatiotemporal distribution of patients, super spreader mining, key node analysis, and the prevention and control of high-risk groups.

**INDEX TERMS** COVID-19; geographic knowledge graph; spatiotemporal big data; visual analysis

## I. INTRODUCTION

**D**URING the COVID-19 (Coronavirus Disease 2019) outbreak, related methods and spatiotemporal big data have been continuously used to form differentiated prevention and control measures. For example, measures such as COVID-19 patients' maps, the population flow and distribution of potential patients, the spatiotemporal tracking of case trajectories, the allocation of medical resources in epidemic areas and the differentiated control of the epidemic have been used in the areas where they may have social and economic impacts. Applications include the spatial transmission of the epidemic, the dynamic analysis of disease conditions and the allocation of emergency resources. In general, spatiotemporal big data played a significant role in the emergency stage

of the epidemic prevention and control, but they have no significant role in real-time disease monitoring and the spatiotemporal prediction of the epidemic's spread. Because of the limitation of the data model, it is difficult to associate spatiotemporal big data with multifactorial data related to the impact of the epidemic. Single, single domain and single mode data analyses conduct analyses and make predictions through the adjustment, correlation and optimization of an algorithm's components that focus on a single domain. Thus, they lack the comprehensive analysis of multiple domains and multiple modalities. Therefore, it is necessary to consider comprehensively analysing the association of large-scale heterogeneous data and using heterogeneous data fusion to expand the multimodal knowledge support of epidemic in-

fection prediction. This expansion requires integrating multiple data factors such as spatial and temporal distributions, population migration, and patients' relationships for collaborative analysis. This integration must be done to analyse the underlying causes of the outbreaks and accurately discover the spatiotemporal mechanism of the epidemic spread, which will provide scientific and technological knowledge that can be used to prevent similar events from happening again.

Knowledge graphs, a branch of artificial intelligence and the most important way to represent knowledge in the big data era, have been widely used in intelligent search, intelligent Question & Answering (Q & A) [1], [2] and knowledge recommendation [3], [4]. Faced with the epidemic situation, relevant scholars and industries in the field of knowledge graph research have constructed the knowledge graph of medical resources based on medical data, encyclopaedias, news, announcements, etc. The graph can provide technical support such as medical resource tracing, epidemic early warnings, etc. Constructing a patient knowledge graph can realize the functions of super spreader mining, key node discovery, prevention and control measures for high-risk groups, etc.

In addition, the core of a knowledge graph is a large-scale semantic web [5]. The advantage of a semantic web is that it can well establish the semantic relationships between entities and store them in a simple formalized format for convenient and fast retrieval and analysis. Knowledge graphs are stored in triples (or quads or quintuples) for the purpose of weight reduction and simplification, which facilitates the identification by computers and reduces the computing requirements [6]. The advantage of the spatiotemporal information analysis model is that it can make full use of spatial location information, visualize the spatial distribution, and discover the time and spatial patterns by using the geographic analysis model. It is difficult to visualize the spatial distribution and spread of an epidemic situation when only using a semantic web. Complicated relationship information such as the patient's social network relationships cannot be described when using only the geographic spatiotemporal model. Therefore, the authors consider the combination of the semantic web and geographic information model to give full play to their advantages to grasp the spatiotemporal distribution situation and relationship links.

In view of the lack of data association in spatiotemporal information analysis and the lack of spatiotemporal situation analysis in knowledge graphs, this paper combines the semantic web of the geographic knowledge graph with the visual analysis model of spatial information and puts forward the comprehensive utilization of related technologies of the geographic knowledge graph and big data visual analysis. Then, it realizes the situational analysis of COVID-19 and the exploration of patient relationships through interactive collaborative analysis.

The main contributions of the paper are as follows.

1) Based on the characteristics of the geographic knowledge graph, a patient entity model and an entity rela-

tionship type and knowledge representation method are proposed, and a knowledge graph of the spatiotemporal information of COVID-19 is constructed.

- 2) To analyse the COVID-19 patients' situations and explore their relationships, an analytical framework is designed. The framework, combining the semantic web of the geographic knowledge graph and the visual analysis model of geographic information, allows one to analyse the semantic web by using the node attribute similarity calculation, key stage mining, community prediction and other methods.
- 3) An efficient epidemic prevention and anti-epidemic method that has referential significance is put forward. Based on experiments and the collaborative analysis of the semantic web and spatial information, it is found that it allows for real-time situational understanding, the discovery of patients' relationships, the analysis of the spatiotemporal distribution of patients, super spreader mining, key node analysis, and the prevention and control of high-risk groups.

## II. RELATED WORK

### A. ANALYSIS OF INFECTIOUS DISEASE PREDICTION

In the medical field, infectious disease prediction analysis has ranged from model predictions to big data-driven research. Increasingly more attention has been paid to the application of artificial intelligence and big data in the prevention and prediction of infectious diseases.

The early dynamic model provided a basic method using qualitative analysis and quantitative analysis to predict infectious diseases [7], [8]. However, due to the small sample data and limited model factors considered, the output results were a qualitative prevention test and management mode reports. With the development and utilization of multi-type sensors, various kinds of epidemic related data with more types and more complete samples have been obtained. Many scholars began to use machine learning and deep learning to achieve infectious disease prediction [9]– [11]. However, the training model is highly dependent on the size and accuracy of the training data set. The setting of the training model parameters is often subjective. The epidemic prediction itself has a large number of independent variables, making it a complicated problem, which could be solved using the combination of a complex network system and a simulation method [12] that could simulate the epidemic spreading process. Complex network and agent simulation [13], [14] that considers individual autonomous behaviour and the interaction between individuals can reflect the impact of a dynamic social network and individual autonomous behaviour on epidemic transmission. The core of this method is simulation modelling, but it does not take into account the attributes of all individuals (such as patients), the relationships between individuals, and the spatial and temporal factors (such as behavioural patterns, geographical distribution, and environmental impacts) that affect individuals.

Spatiotemporal elements are used in the medical research

of infectious diseases in three ways. The first one is using them to analyse the spatial-temporal distribution and epidemic situation of diseases based on a spatial-temporal geographical simulation [15]–[17]. Such methods can provide good spatial and temporal references for the spatial spread of the epidemic, the dynamic analysis of the epidemic, and the allocation of emergency resources, but it is difficult to analyse the underlying causes of the outbreak and the development of the epidemic. The second one is where the elements act as spatiotemporal influencing factors of the epidemic spreading in the process of building a simulation model of medical infectious diseases, which takes into account the geographical space-time factors in the prediction [18]. Such methods consider social and economic indicators, meteorological factors, population density, road traffic, spatial patterns, personnel movement and other factors, which provide a more comprehensive perspective and entry point for the current monitoring of the epidemic situation. However, the number of influencing factors that a single model can consider is limited. In addition, the variable factors such as changes in time and space could be hard to model. The third way is using them in interactive data analysis driven by multi-dimensional data. Using spatial-temporal big data technology, such methods integrate more environmental elements into collaborative analysis. The data-driven big data analysis method provides a new idea for the prediction analysis and mining of infectious diseases, which makes up for the shortcomings of the prediction model that is weak at complex model building [19]–[21].

The data-driven spatiotemporal analysis method enhances the collaborative analysis ability of multi-source heterogeneous data, emphasizes the relevance of data, makes better use of the correlation of data, allows for the mutual verification of data cross domains and disciplines, and improves the accuracy and scientificity of epidemic prediction. By fully using time series data, the intermediate process of the epidemic spread can be accurately controlled [22], [23]. In addition, prediction grading on different levels and stages of the epidemic makes it more conducive to determining epidemic prevention and control measures. The introduction of different types of related data results in problems such as how to establish association models between data and how to realize the collaborative analysis of large-scale multi-source heterogeneous data; furthermore, such large scale data samples make the model calculation a problem [24].

## B. GEOGRAPHICAL KNOWLEDGE GRAPH

The knowledge graph is essentially a networked knowledge base that is linked by entities with attributes through relationships. In other words, a knowledge graph is a knowledge base with a directed graph structure, where the nodes of the graph represent entities or concepts, and the edges represent various semantic relationships between entities/concepts.

A geographic knowledge graph is the expansion of a knowledge graph using geography, and it is a structured geo-semantic knowledge base. By formally describing the

concept, entity, and attribute and their relationships in the field of geography, the concept and entity are connected with each other, forming a network knowledge structure [25].

The core of a geographic knowledge graph is to construct a large-scale geographic knowledge semantic web, which is essentially a large-scale directed network graph. The core technology is geographic entity association model building, which allows one to build relationships between large amounts of multi-source heterogeneous epidemic situation data and to achieve the collaborative analysis of multiple sources and multiple elements. With the rapid development of artificial intelligence, big data and other emerging technologies, geographic knowledge graphs have become an effective way to realize the intelligent organization of geographic information, and it has been rapidly recognized by relevant experts in the field of geography. Studies have been done on the aspects of geographic entity extraction [26], [27], topology and orientation relationship extraction [28] and geographic knowledge graph storage [29]–[32].

Related technology research results on the construction of geographic knowledge graphs mainly include the construction framework of the geographic knowledge graph put forward by Lu Feng *et al.* [33]; the analysis and interpretation of the key technology of the construction of the geographic knowledge graph for the intelligent application of a virtual geographic environment that was conducted by Jiang Bingchuan *et al.* [25]; the CrowdGeoKG constructed by Jiaoyan Chen *et al.* [34], who used wikidata and human geography knowledge to strengthen the geographic entities of OpenStreetMap; and the formal representation model of geographic knowledge proposed by Wang *s et al.* [35], which can deal with the spatial, temporal and dynamic characteristics of geographic knowledge. A geographic knowledge graph can provide a good reference for forming the knowledge graph of the spatial-temporal information of COVID-19. The modelling and representation of spatial-temporal features could be realized using a knowledge graph. The knowledge graph of the spatiotemporal information of COVID-19 should include spatial and temporal characteristics, geographic entities, and data types such as space-time object, social network and regional environment data.

A large-scale spatiotemporal knowledge graph needs to combine spatiotemporal data representation models (such as maps, spatiotemporal distributions, etc.) for collaborative analysis. Yuhan Sun [36] proposed a universal geographic knowledge graph indexing framework, Riso-tree, to implement semantic web and geospatial indexing, and constructed a location-aware search and query based on the geographic knowledge of interactive maps [37]. Having combined the characteristics of trajectory data and the definition of a knowledge graph, Wu Xian *et al.* [38] extracted the entities, relationships, and attributes of trajectory data and constructed the trajectory map, which supports basic queries, range queries, nearest neighbour queries, keyword queries and trajectory mode queries. Yunpeng Xiao *et al.* [39] proposed a rumour propagation dynamics model based on

an evolutionary game and anti-rumour information, and the model can effectively describe the propagation of rumours and the dynamic change rule of the influence of anti-rumour information, and further proposed a group behavior model for rumor and anti-rumor [40]. In paper [41], Yunpeng Xiao et al. used user multidimensional attributes and evolutionary games combined with the traditional susceptible-infected-recovered (SIR) epidemic model, which was used to quantify the impacts of external and internal driving factors on group state transitions during hotspot propagation.

In the application of geographic knowledge graphs, geographic knowledge association queries based on knowledge graphs are widely used. Typical examples are Geo-Wiki [42] (a geo-semantics sharing network system) and KIDGS [43] (a geographical knowledge-informed digital gazetteer service). Lin Ye et al. [44], [45] stated that to conduct VGE knowledge engineering, it is necessary to realize the intelligent transformation from “data-information-knowledge-smart” and realize intelligent virtual geographic environment services. Robert Laurini [46] presented a tentative conceptual framework for managing practical geographic knowledge, and the geographic knowledge base (GKB) includes geographic objects, geographic structures, geographic relations, geographic rules, geographic ontology, a gazetteer, physico-mathematical models, and external knowledge. Jiang Bingchuan et al. [47] stated that combining a spatial-temporal data model and semantic web model is one of the main ways to implement an intelligent geographic information service, and they also realized intelligent Q & A and interaction with a virtual geographic environment by using a geographic knowledge graph. Liu Ju et al. [48] realized the intelligent querying of Chinese Guqin masters’ related information based on a knowledge base and a geographic information model. Liu Junnan et al. [49] built an event representation model centred on events that combined spatiotemporal and semantic features, which could organize multi-source heterogeneous data using knowledge graph technologies to represent terrorist event relationships and attribute information.

The geographic knowledge graph is a sound improvement over the knowledge graph, which has been successfully applied in fact-based knowledge Q & A, geographic knowledge association searches and intelligent interaction. In the face of the epidemic, the constructed epidemic Knowledge Q & A system and patient relationship query system have played positive roles in the popularization of epidemic knowledge and in self-protection measures. However, due to the incomplete correlation data and inadequate correlation analysis methods, it is difficult to carry out in-depth analysis, which makes it a problem to closely follow and further predict the epidemic situation.

In summary, the related research on knowledge graphs can direct the construction of epidemic knowledge graphs in these two aspects.

- 1) The construction of an ontology model of geographical knowledge. Geographical ontology research started

earlier, and has formed a relatively complete set of ontology systems and ontology model construction methods, which can provide ontology system guidance for the construction of the COVID-19 patients’ spatiotemporal information knowledge graph; however, it needs to be improved based on the characteristics of the medical field.

- 2) The geographic knowledge extraction and representation model. This model has certain referential significance for the extraction of different geographical knowledge, especially knowledge with time and spatial variables such as spatiotemporal information and geographic events. However, it requires further research regarding the formalized representation of the geographic spatial-temporal model about the epidemic’s spread in order to build a model or representation that presents both factual knowledge and epidemic changes.

### III. CONSTRUCTION OF COVID-19 PATIENTS’ INFORMATION KNOWLEDGE GRAPH

Referring to the basic process of building a geographic knowledge graph, the basic framework for building the spatial-temporal information knowledge graph of COVID-19 patients is shown in Figure 1. The knowledge sources of the spatiotemporal information knowledge graph of patients include the following: social network data, personal relationship data, news data, migration data, epidemic surveillance data, trajectory data, and basic geographic information data. The data sources, formats and extraction methods are shown in Table 1.

#### A. ONTOLOGY CONSTRUCTION

The ontology layer of COVID-19 patients’ knowledge graph is a subset of geographic ontology.

##### 1) DEFINE THE CONCEPT SET

ON\_Concept:  $\{C\}$  is defined as follows:  $\{C_1$ : patients,  $C_2$ : residence,  $C_3$ : hospital,  $C_4$ : vehicle,  $C_5$ : location,  $E$ : event ontology $\}$ .

Here,  $C_1$ : refers to the confirmed COVID-19 patients,  $C_2$ : refers to the patients’ residence,  $C_3$  refers to the hospital where the patients get medical treatment,  $C_4$ : refers to the patients’ means of transport,  $C_5$ : refers to the location of the patients; and  $E$ : event ontology, which mainly includes medical treatment events, confirmed events, travel events, contact events, shopping events, gathering events, home events and fever events.

Define the relationship in the ontology ON\_Relationship as follows (take the patient ontology and other types of ontology as examples):  $\{R\} = \{R(C_1, C_1), R(C_1, C_2), R(C_1, C_3), R(C_1, C_4), R(C_1, C_5), R(C_1, E)\}$  where,

①  $R(C_1, C_1)$  is the relationship between patients;

②  $R(C_1, C_2)$  is the relationship between the patients and the places they live, i.e., “residence”;



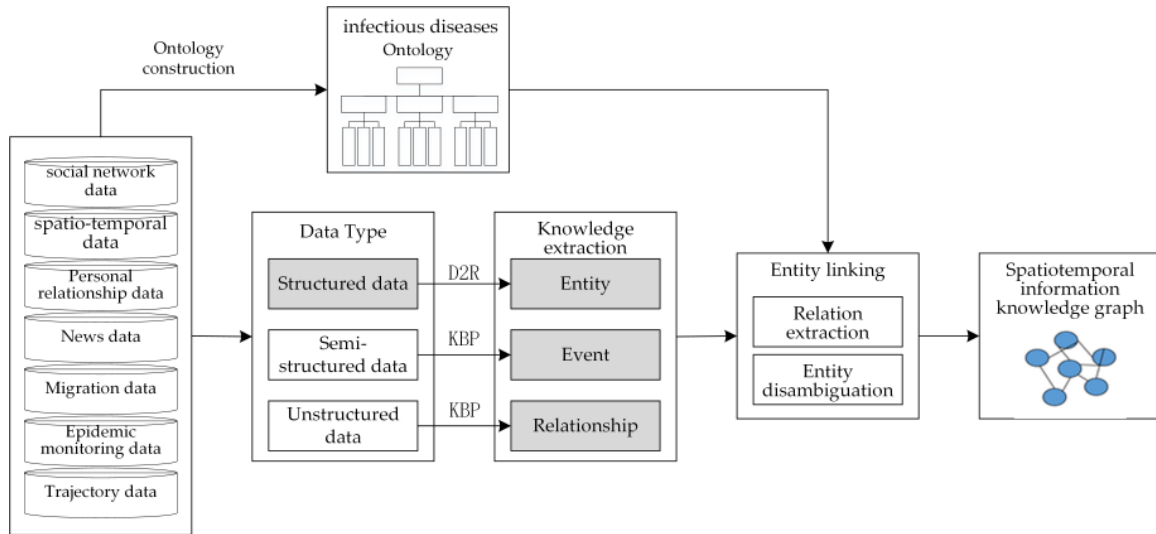


FIGURE 1. The basic framework for building the spatial-temporal information knowledge graph of COVID-19 patients.

TABLE 1. SOURCES OF THE SPATIOTEMPORAL INFORMATION OF COVID-19 PATIENTS

Data category	Data Sources	Data Format	Extraction method	Knowledge type
Social network data	WeChat and Weibo data	Unstructured text	Wrapper extraction	Character entities, social relationships, geographic and hidden information
Personal relationship data	Household registration, WeChat, and Weibo	Structured household registration database, unstructured text	D2R, wrapper extraction	Character entities and character relationships
News data	News website	Unstructured text	Wrapper extraction	News
Migration data	Railway, civil aviation, and highway administrations	Structured database	Wrapper extraction	Tracking data and vehicle entities
Outbreak Monitoring data	National and provincial health committees	Unstructured text	Wrapper extraction	Patient entities and patient relationships
Trajectory data	Mobile positioning and surveys	Structured database	Wrapper extraction	Character entities and, displacement events
Basic geographic information data	Maps, Images, and Geographic Names	Structured database	Wrapper extraction	Geographical Name Entities and Geographical Knowledge

③ $R(C_1, C_3)$  is the relationship between patients and the hospitals where they get medical treatment, i.e., “hospital in”;

④ $R(C_1, C_4)$  is the relationship between patients and the transportation, i.e., ”take”;

⑤ $R(C_1, C_5)$  is the relationship between the patients and their locations, i.e., “location”; and

⑥ $R(C_1, E)$  is the relationship between patients and events, i.e., “has-event”.

## 2) ENTITY SEMANTIC MODEL AND RELATIONSHIP DEFINITION

The entity conceptual model is shown in Table 2.

The knowledge graph of COVID-19 patients has clear characteristic spatiotemporal information, mainly including the following, as shown in Figure 2.

①Physical position characteristics. All entities  $e$  have basic location characteristics, including  $f_e = \{lng, lat, D\}$ , where lng represents longitude, lat represents latitude, and  $D$  describes the scope of the entity and mainly refers to the

TABLE 2. ENTITY CONCEPTUAL MODEL

Patient entity
+Patient entity identification code:Entity_ID
+Patient entity name:Name
+Patient entity classification:Classification
+Patient entity relationship:Relationship
+Patient entity properties:
+Patient entity location:
+Patient events:

administrative radius of the area where the patient is located.

②Triple related to the location. In the triple  $(h, r, t)$ , at least one of the two entities,  $h$  and  $t$ , contains a location feature. At present, the entities with location features in the graph include residences, departures, destinations, hospitals for medical treatment, locations and so on.

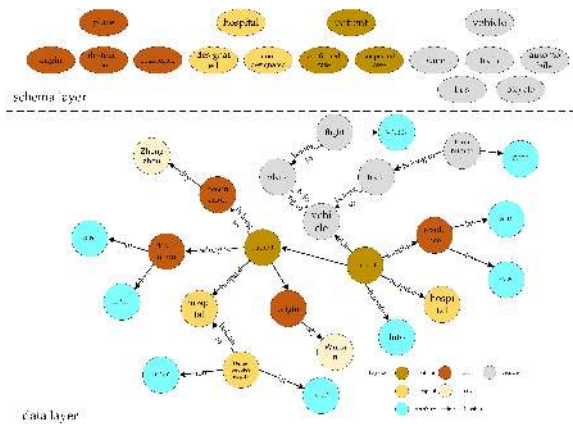


FIGURE 2. The COVID-19 patients' entities and relationships.

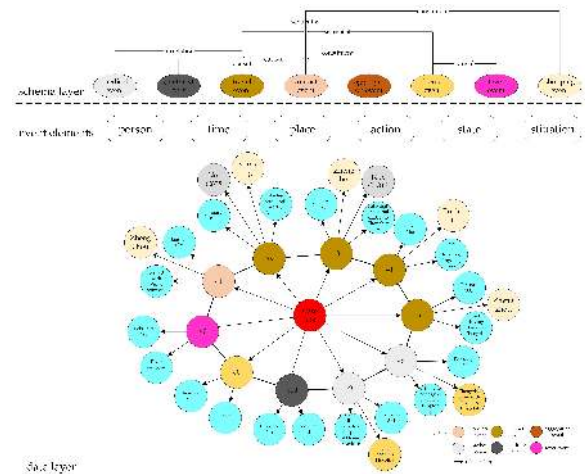


FIGURE 3. The schema of events and relationships.

### 3) EVENT SEMANTIC MODEL AND RELATIONSHIP DEFINITION

An event includes 6 elements: who, when, where, action, state and what. An event, expressed as  $e$ , could be defined as a sextuple as follows:

$$e = (E, T, P, A, S, W)$$

The letters  $E, T, P, A, S$  and  $W$  represent the subject (including entities such as people and medical organizations), time, place, actions (such as trigger words for actions), state set and situations (event descriptions), respectively. For example, in "patient 23 returned to Taikang County by taking private car from Wuhan on January 7", the event subject is "patients 23", the time is "January 7", the position is "Wuhan", "Taikang county", the actions are "take, return", the state set is "return", and the event description is "take private car to return".

The relationships between events include constituent relationships, sequential relationships, causal relationships and correlation relationships. The relationship between events can be linked through time, place and participants.

For example: "Patient XXX, female, 25 years old, now living in the Evergrande Oasis of Minghu Office of the Economic Development Zone, contacted the returnees from Wuhan in Zhengzhou on January 18. She took the G658 high-speed rail (carriage 11) to Xinxiang on January 19 and went back to Zhengzhou on January 21 by high-speed rail G1813 (carriage 4). Then, she drove to Ruzhou, PingDingshan on the same day and drove back to Zhengzhou on January 30 at last. Since then, she had been stayed at home until fever symptoms occurred on February 1 when she drove to Zhengzhou Seventh People's hospital. On February 3, she drove to the people's Hospital of Henan Province for treatment and was diagnosed on February 6"

As shown in Figure 3, there is a causal relationship between the contact event  $e_1$  "contact with Wuhan returnees on January 18" and the fever event  $e_7$  "fever symptoms on February 1". The migration event  $e_2$  "take G658 high-speed rail (carriage 11) to Xinxiang on January 19" and the

migration event  $e_3$  "take high-speed rail G1813 (carriage 4) to Zhengzhou on January 21" are sequential. The fever event  $e_7$  "fever symptoms on February 1" and the event  $e_8$  "drive to Zhengzhou Seventh People's Hospital on February 1" have a causal relationship.

### 4) SEMANTIC WEB REPRESENTATION

The spatiotemporal information knowledge graph of COVID-19 patients expresses the concept, entity, attribute and event of the COVID-19 subfield with resource the description framework (RDF) as a triple, namely, " $\langle s$  (subject),  $P$  (predicate),  $O$  (object)  $\rangle$ ". Then, we establish the relationships among its elements, entities and events. It is represented by the directed graph of "point— edge", as shown in Figure 4.

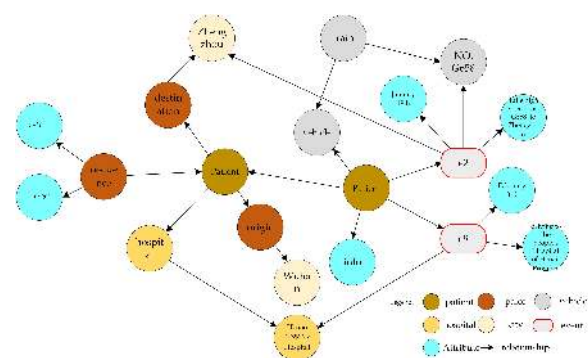


FIGURE 4. The representation of the COVID-19 spatiotemporal information knowledge graph.

## IV. ANALYSIS METHOD OF EPIDEMIC SITUATION BASED ON SPATIOTEMPORAL INFORMATION KNOWLEDGE GRAPH OF COVID-19 PATIENTS

### A. ANALYSIS TASKS

To realize the monitoring and analysis of epidemic situations by combining the patient semantic relation network model

and spatiotemporal information visual analysis model, the following analysis tasks need to be done:

- 1) Epidemic situation analysis: analyse the regional distribution of patients to understand the current situation,
- 2) Patients' relationship analysis: analyse patients who are infected by gathering them for super spreader mining,
- 3) Analysis of the spatial and temporal patterns of the patients: analyse patients' activity areas to set up early warning and protection areas, and
- 4) Early warning and prediction of high-risk groups: find the possible cross-infection patients based on the relationships between patients and other entities.

### B. ANALYSIS FRAMEWORK

As shown in Figure 5, the basic analytical framework for the comprehensive geographic knowledge graph is based on the knowledge graph semantic web model and the temporal and spatial information visual analysis model. Almost all nodes in the spatiotemporal information knowledge graph have spatiotemporal position information, and all types of nodes can be mapped to a unified spatiotemporal framework. This framework can take full advantage of the characteristics of the semantic web map analysis and geographical spatiotemporal analysis for comprehensive analysis and application. The graph analysis based on the knowledge graph semantic model includes functions such as relationship analysis, network graph distribution, link prediction, and graph visualization. The analysis based on spatiotemporal data models includes functions such as geographical spatiotemporal distribution, spatiotemporal trajectory analysis, character and event analysis, and spatiotemporal situation analysis.

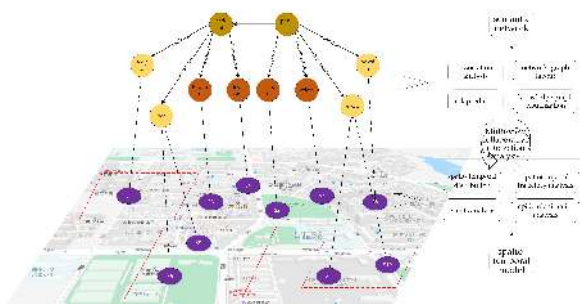


FIGURE 5. Collaborative analysis of the semantic web and spatiotemporal model.

### C. NODE CLUSTERING AND SIMILARITY ANALYSIS OF COVID-19 PATIENTS' SEMANTIC WEB

#### 1) SIMILARITY CALCULATION OF PATIENTS' NODES BASED ON NODE ATTRIBUTES

The patients' spatiotemporal information knowledge graph is highly dynamic. The entities in the network and the relationships between them would evolve over time and space. New nodes might show up while the existing ones might disappear, and currently disconnected nodes might be connected. Link prediction can be used to analyse the potential

relationships between the confirmed patients and to report on an epidemic situation that requires an early warning.

Node similarity calculation methods are divided into attribute-based methods and link-based methods. The COVID-19 patients' knowledge graph semantic web is a directed graph with relatively clear node attributes where the attribute-based method should be applied. It calculates the similarity by comparing the attribute values of the nodes. First, the attribute vectors are constructed for each node, and each attribute is regarded as a dimension of the multi-dimensional space. For example,  $v_p$  and  $v_q$  are two nodes in the network and both of them have  $m$  attributes, which could be expressed as follows:

$$v_p = (e_{p1}, e_{p2}, \dots, e_{pm})$$

$$v_q = (e_{q1}, e_{q2}, \dots, e_{qm})$$

The nodes are mapped to the low dimensional space, and then the node similarity is calculated using the Euclidean distance formula and cosine similarity formula of vector space.

$$\text{dist}(v_p, v_q) = \sqrt{\sum_{i=1}^m (v_{pi} - v_{qi})^2} \quad (1)$$

$$\text{sim}(v_p, v_q) = \frac{\vec{v}_p \cdot \vec{v}_q}{\|v_p\| \cdot \|v_q\|} \quad (2)$$

#### 2) SUPER SPREADER MINING BASED ON BETWEENNESS CENTRALITY ALGORITHM

The betweenness centrality, which is the probability that a node is located on the shortest path connecting any two nodes, is closely related to the length of the path. The COVID-19 patients' knowledge graph semantic web has rich knowledge of node attributes and relationships (edges), which can realize super spreader mining using the betweenness centrality algorithm.

The betweenness centrality of a node [50] refers to the ratio of the number of shortest paths passing through the node to the number of shortest paths between any two nodes in the network. Its value indicates the network's control ability over information transmission. The larger the value is, the greater the number of shortest paths between any two nodes in the network passing through the node, and the more important the node.

Suppose that the number of shortest paths between node pair  $j$  and  $k$  is  $g_{j,k}$ , the number of shortest paths that pass through node  $i$  is  $g_{i,j}(i)$ , and  $C_{j,k}(i)$  represents the betweenness centrality of  $i$ , i.e., the probability that  $i$  is on the shortest path between  $j$  and  $k$ . The calculation formula for the betweenness centrality is as follows [51]:

$$\text{BC}(i) = \sum_{j \neq i \neq k \in v} \frac{g_{j,k}(i)}{g_{j,k}} \quad (3)$$

$g_{j,k}(i)$  is the number of shortest paths between  $j$  and  $k$  that  $i$  passes.  $g_{j,k}$  is the number of the shortest paths between  $j$  and  $k$ .

### 3) SEMANTIC WEB NODE CLUSTERING BASED ON THE LOUVAIN ALGORITHM

COVID-19 patients' spatiotemporal information knowledge graph is essentially a large-scale directed graph. The cluster analysis of network nodes is similar to the community detection of a social network. A community is a set of nodes that are similar to each other and different from other nodes in the network. The community detection algorithm helps to mine some nodes at the centre or edge of the community. For example, it can be used for the analysis of infected patients caused by a family gathering.

The Louvain algorithm [52] is a community detection algorithm based on Modularity, which is efficient, provides good results, and can detect hierarchical community structures. Its optimization goal is to maximize the Modularity of the community network.

The Modularity function is used to measure the quality of the results of the community detection algorithm. It can describe the compactness of detected communities [53]. Its function is defined as follows:

$$Q = \frac{1}{2m} \sum_{i,j} [A_{i,j} - \frac{k_i k_j}{2m}] \delta(c_i, c_j) \quad (4)$$

where  $m$  is the number of edges in the network, and  $A$  is the adjacency matrix. If  $c_i$  equals  $c_j$ , then  $\delta(c_i, c_j) = 1$ ; otherwise, it would be 0.

The formula can be further transformed into the following:

$$Q = \frac{1}{2m} \sum_c [\sum \text{in} - \frac{(\sum \text{tot})^2}{2m}] \quad (5)$$

$\sum \text{in}$  represents the sum of the weights of the edges in community  $c$ , and  $\sum \text{tot}$  represents the sum of the weights of the edges connected to nodes in community  $c$ .

The Louvain algorithm consists of the following steps.

Step 1: Traverse all the nodes  $i$  in the network, assign a single node  $i$  to the community of each neighbouring node, calculate the changes in the Modularity measure  $Q$  represented by  $\Delta Q$ , and add it to the community of the neighbouring node with the maximum change in Modularity measure. Repeat this step until all nodes are no longer changed, and the generated small community is the input of the second step.

Step 2: Compress the graph. All nodes in the same community are compressed into a new node. The weights of the edges between the nodes in the community are converted into the weights of the ring of the new node, and the edge weights between the communities are converted into the edge weights between the new nodes.

Step 3: Iterate these two steps until the algorithm is stable.

TABLE 3. PSEUDOCODE OF LOUVAIN ALGORITHM

Define directed node network graph structure as $G(V, E)$
1) Initialize each point in $G$ as a community, calculate the module degree value at this time and store it in $Q_1, Q_3 = Q_1$ ;
2) $Q_2 = Q_1$ ;
3) for $i = 1$ to $n$ /* $n$ is the number of vertices in the network graph */
4) if $d(v_i) \neq 1$ /* $d(x)$ calculate the degree of vertex $x$ */
5) add points to the community to which they are connected;
6) else
7) remove the vertex $v_i$ from the original community;
8) add $v_i$ to the community that makes $\Delta Q$ the largest;
9) end if
10) end for
11) calculate the modularity value at this time and store it in $Q_1$ ;
12) if $Q_1 > Q_2$ , go to step 2;
13) end if;
14) merge the communities into one super point;
15) store the points contained in different hyper points into the corresponding collection array communities;
16) return communities

When node  $i$  is assigned to the community of neighbouring node  $j$ ,  $\Delta Q$  can be calculated by formula 6

$$\Delta Q = \left[ \frac{\sum \text{in} + k_{i,in}}{2m} - \left( \frac{\sum \text{tot} + k_i}{2m} \right)^2 \right] - \left[ \frac{\sum \text{in}}{2m} - \left( \frac{\sum \text{tot}}{2m} \right)^2 - \left( \frac{k_i}{2m} \right)^2 \right] \quad (6)$$

The pseudocode of the Louvain algorithm is shown in Table 3. The time complexity of the Louvain community discovery algorithm is  $O(n \log n)$ , where  $n$  is the number of nodes in the network. As the data scale increases, the time complexity of the algorithm increases by levels of  $n \log n$ .

## V. EXPERIMENT

### A. EXPERIMENT DATA

The experimental data are collected from the confirmed COVID-19 patients in Henan Province in China (until March 13, 2020), and have been released by the Henan Province Government. After data cleaning, the patient knowledge graph of the spatiotemporal information is constructed with 2312 entities and 5055 entity relationships. The entity labels include "patient", "train number", "place of residence", "place of departure", "destination", "hospital", "event", etc.

### B. EXPERIMENTAL PLATFORM

The experimental platform adopts the B/S architecture, which is developed and implemented based on the ECharts and Java languages. By taking advantage of both rich GIS information and the interactive visual analysis system, a visual analysis system of the spatiotemporal information knowledge graph of COVID-19 is constructed. The system interface is shown in Figure 6.

Figure 6 is the network analysis of the patients' relationships in Henan Province, which uses the following components.

A: Maps and migration maps. It mainly displays the location information, urban trajectory and migration trajectory of patients.



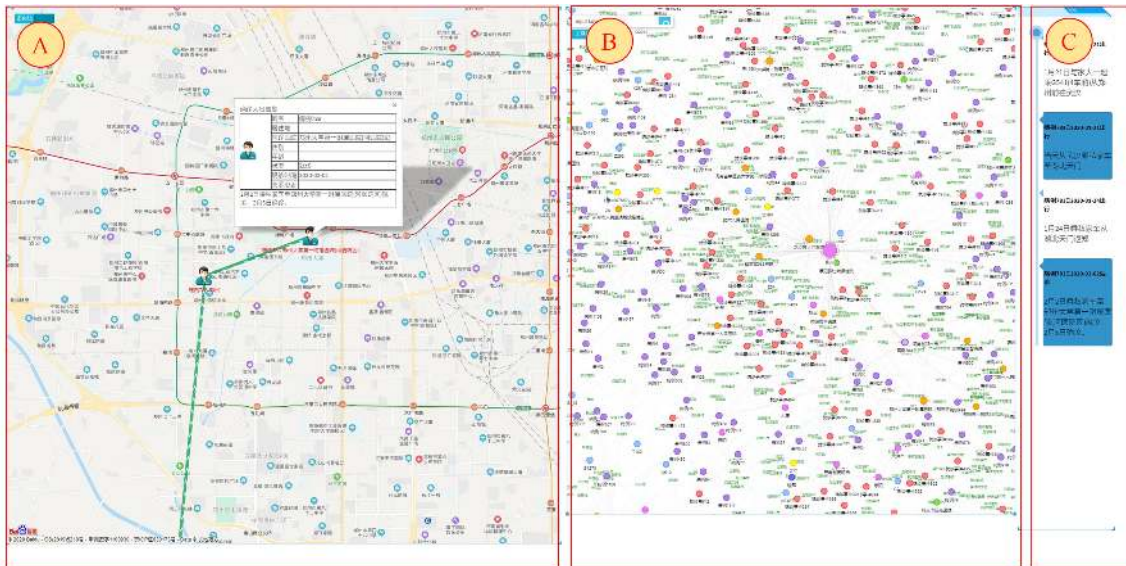


FIGURE 6. Visual analysis of the COVID-19 patient's knowledge graph.

B: Interactive visual analysis of the knowledge graph of COVID-19 patients. Through the collaborative analysis with the maps, the patient type analysis, the regional prevention and control situation analysis, the gathering infected patient analysis, high-risk group prevention and control, the city track analysis, and the missing reports and concealed cases analysis, the following functions are realized: patient relationship tracking, high-risk group prevention and control information release, event pattern analysis, etc.

C: Event axis. These lists are used to show the temporal and spatial characteristics of patients' visiting events, diagnosis events, travel events, contact events, shopping events, gathering events, quarantine events and fever events.

### C. KNOWLEDGE GRAPH SEMANTIC WEB ANALYSIS

#### 1) ANALYSIS OF PATIENTS' TYPES

After screening the "patient" and "departure place" entities, in Figure 7, it can be clearly seen that in the early stage of the epidemic, most of the infected people are related to "Wuhan", "Hankou" and other Hubei areas, belonging to the "direct input" patient type.

#### 2) ANALYSIS OF REGIONAL PREVENTION AND CONTROL SITUATION

After screening the "patient" and "location" entities, as shown in Figure 8, it can be seen clearly that the patients are mainly located in "Zhengzhou", "Xinyang", "Pingdingshan", "Nanyang", "Jiaozuo", "Shangqiu", "Xinxiang", "Kaifeng", "Hebi", "Luohe", "Zhoukou", etc. Some patients moved directly between different cities, resulting in cross-regional infections. There is no personnel movement in "Jiyuan", "Luohe", and "Sanmenxia" due to strict control measures.

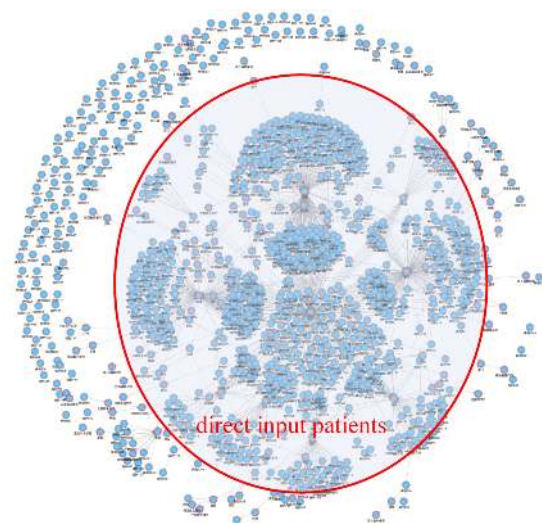


FIGURE 7. Analysis of patient types.

#### 3) PATIENTS ANALYSIS OF FAMILY GATHERING INFECTION

After the character relationship is screened and the free node is removed, the high risk of infection from families has been proved by the Louvain group analysis, as shown in Figure 9.

Therefore, it is of necessity to implement quarantine measures for families.

#### 4) HIGH-RISK GROUP PREVENTION AND CONTROL ANALYSIS

There are several patients who are infected during family gatherings. For example, as shown in Figure 10, based on the betweenness centrality analysis, "case 450", "case 572",

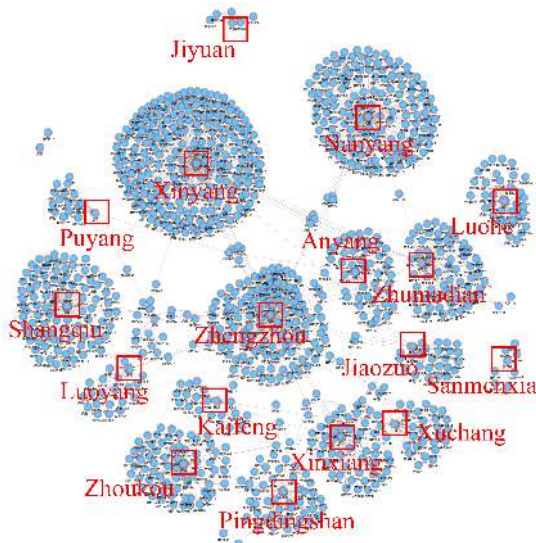


FIGURE 8. Analysis of regional prevention and control situation.

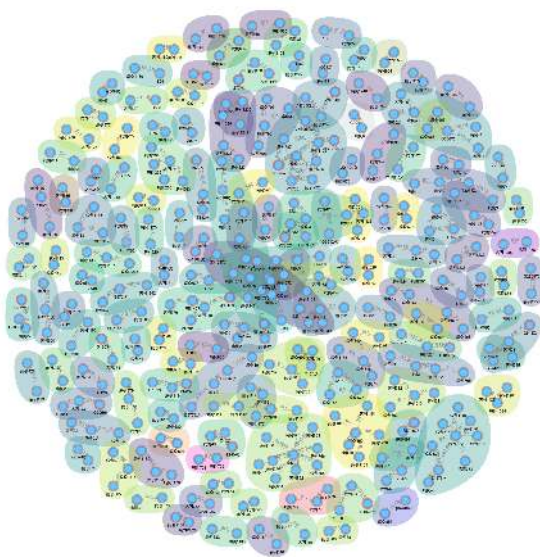


FIGURE 9. Patients' analysis of family gathering infection.

"case 758", "case 1270" and "case 1106" all get infected by their family members who they have contacted. These cases are typical super spreaders (as shown in Figure 11(a)), and thus it is necessary to remind the personnel who have close contact with them.

As shown in Figure 11(b), "Case 49", "case 786", "case 424", "case 922" and "case 787" took the same flight to travel abroad on the same day. It proves the necessity of the release the key flight information in time for early warnings, and the passengers on this flight should be quarantined for particular observations.

Node prediction analysis may discover further information. For example, as shown in Figure 11(c), the data show that case 41 and case 758 took the same train on the same day.

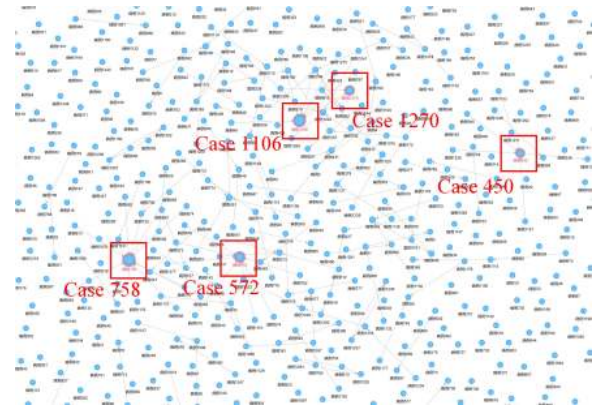


FIGURE 10. Analysis of super spreaders based on betweenness centrality

The similarity analysis of node information is conducted, and a similarity value of 1.5 is obtained. This outcome indicates that case 41, who has not been listed as one of the contactees in the released details, might have contacted case 17. The value of the associated forecast between case 17 and case 41 is 1.649, as shown in Figure 12. Therefore, it is suspected that case 41 (or case 17) is unreported or concealed.

5) ANALYSIS OF PATIENT TRAJECTORIES IN THE CITY  
The analysis of the patient trajectories in urban areas could help to take particular measures to report and control the places where cases have visited and the public transportation they have taken. In Figure 13, the left graph in the following shows the patient trajectories in the city, the middle graph is the sequential events network of "confirmed diagnosis–confirmed diagnosis–confirmed diagnosis–confirmed diagnosis", and the right is the event axis.

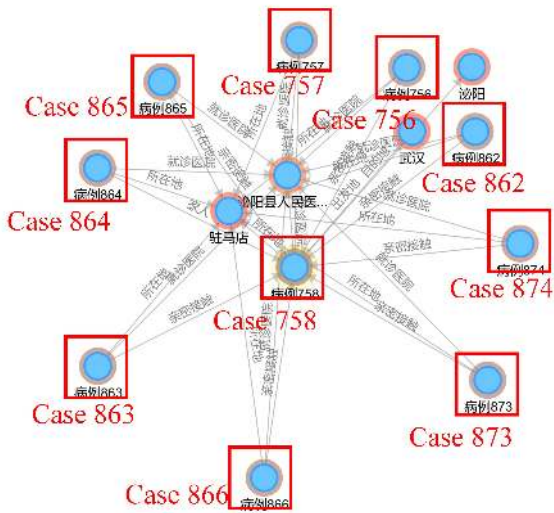
## VI. DISCUSSION

Experiments have proved that the knowledge graph of the spatiotemporal information of COVID-19 based on various heterogeneous data could take advantage of both semantic web analysis and the visual analysis of the spatiotemporal information for analysis and application to measures that are used to realize "macro situation control" of the epidemic outbreak and "precise prevention and control for patients". Such measures help to improve the accuracy of locating the epidemic transmission path and preventing the epidemic from further spreading.

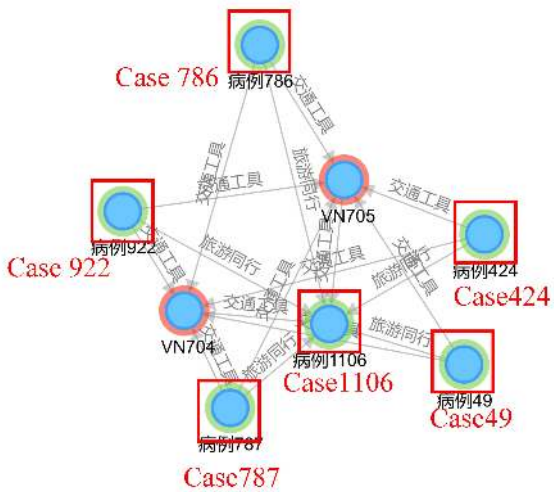
However, the database of this paper is still relatively weak. The proliferation of the data scale and semantic web scale would lead to low analysis efficiency and weak robustness. The knowledge graph of the spatiotemporal information for epidemic prevention and control in the future can be further studied from the following two aspects.

- 1) The expansion of the data association scale. The data could be improved by adding associated multi-source, cross-domain data types, such as social network data, spatiotemporal location information, epidemic surveillance information and geographic environment data, to

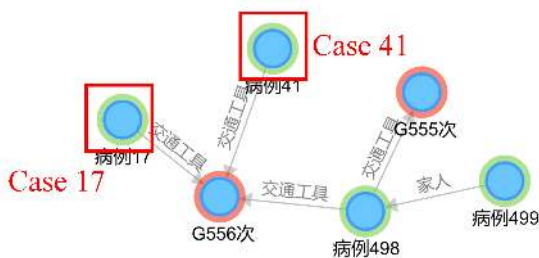




(a) super-spreader warning



(b) warning of infection in travels



(c) prediction of patients' similarity nodes

FIGURE 11. Analysis of typical cases



FIGURE 12. Node attribute similarity calculation

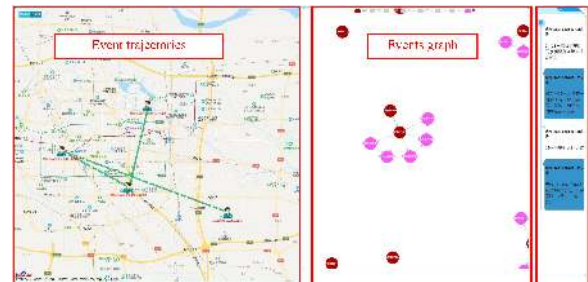


FIGURE 13. Analysis of patients' trajectories in urban areas

build knowledge models and representations suitable for large-scale temporal features. This improvement would help to form a large-scale spatiotemporal information semantic web of the COVID-19 epidemic situation and build up the relationships among epidemic information cross domains and disciplines. For example, the relationship between characters, in addition to "family", "relatives", there are "friends", "colleagues", "teacher-student" relations, etc. After expanding the character relationship type, a complex character relationship network can be established based on large-scale data such as company human resource data, government household registration data, etc., which would help to identify people suspected of contacting confirmed cases.

- 2) Use a graph neural network to learn the spatiotemporal knowledge graph. COVID-19 patients' spatiotemporal information knowledge graph is essentially a large-scale directed graph. A graph neural network model can be used to design a highly dynamic spatial-temporal neural network that meets the sequential characteristics of directed graphs. This design could be used for large-scale node prediction, link prediction and community clustering prediction algorithms in large-scale complex networks to realize the real-time prediction and simulation of the epidemic situation of the COVID-19 spatiotemporal information complex network. For example, we could preset the weights of spatial locations, patients' influence, hidden relations, and patients' preferences to study and construct a graph neural network model to implement community detection, epidemic control area classification and analysis of infection patterns.

## VII. CONCLUSION

In response to the outbreaks, geographers took the initiative to continuously integrate the measures related to space-time big data such as COVID-19 epidemic maps, the distribution of communities with diagnosed residents, and the dynamic trajectory of patients. Their efforts in such areas like the spatial transmission of the epidemic, the dynamic analysis of the disease's condition and the allocation of emergency resources, might be help of differentiated prevention and control and thus cause social and economic impacts. They have constructed spatial interaction models of population migration and spatial interaction models of epidemics. In general, spatiotemporal big data played a particular role in the emergency stage of the epidemic prevention and control.

In view of the lack of data association in spatiotemporal information analysis and the lack of spatiotemporal situation analysis in knowledge graphs, this paper combines the semantic web of the geographic knowledge graph with the visual analysis model of spatial information and puts forward the comprehensive utilization of related technologies of the geographic knowledge graph and big data visual analysis. This realizes the situation analysis of COVID-19 and the exploration of patient relationships through interactive collaborative analysis.

Compared to previously conducted studies, there are seven innovations in this paper.

- 1) It constructs the COVID-19 patients' knowledge graph based on the information of the COVID-19 patients in Henan Province.
- 2) It conducts patient type analysis, regional prevention, and control situation analysis, gathering infected patients' analysis, high-risk group prevention and control analysis, city trajectory analysis, and missing report and concealed case analysis to trace the relationships between patients, publish information for high-risk group prevention and control, and analyse event patterns.
- 3) It designs curve graphs to present the epidemic situation in China and in Henan Province based on the released COVID-19 information that has been analysed using multi-view collaborative analysis technology.
- 4) It designs a bubble chart showing the development trend of COVID-19 in all provinces and cities in China.
- 5) It designs a migration chart that is mainly based on the train passenger numbers published on people.com.cn that have reported confirmed cases.
- 6) It designs a network node chart showing the relationship network in the patient knowledge graph.
- 7) It designs the distribution map showing the distribution of the daily confirmed cases in Henan Province.

Starting from the analysis of the semantic web and the spatiotemporal distribution, a multi-level analysis and application from "macro situation control" to "precision patient prevention and control" has been implemented, which can be used to improve the scientificity and accuracy of the surveillance, prediction and responses related to new coronavirus

pneumonia epidemics. It will provide support for the further promotion of the intelligent prevention and control of major epidemic outbreaks and will help to improve the accuracy of locating epidemic transmission path and preventing it from further spreading.

## ACKNOWLEDGMENT

The authors are thankful for the work of different scholars in knowledge graphs that are cited in the references provided in this paper. The authors would also thank the developers of ECharts for their excellent open-source visualization tools.

## REFERENCES

- [1] X. Huang, J. Zhang, D. Li, and P. Li, "Knowledge graph embedding based question answering," in *Proc. 12th ACM International Conference on Web Search and Data Mining*, Melbourne, VIC, Australia, Jan. 2019, pp. 105–113, doi: 10.1145/3289600.3290956.
- [2] Y. Hao, Y. Zhang, K. Liu, S. He, Z. Liu, H. Wu, and J. Zhao, "An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, vol. 1, Aug. 2017, pp. 221–231, doi: 10.18653/v1/P17-1021.
- [3] H. Wang, F. Zhang, X. Xie, and M. Guo, "DKN: deep knowledge-aware network for news recommendation," in *Proceedings of the 2018 World Wide Web Conference*, Lyon, France, April. 2018, pp. 1835–1844, doi: 10.1145/3178876.3186175.
- [4] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T. Chua, "Explainable reasoning over knowledge graphs for recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Honolulu, Hawaii, USA, Feb. 2019, pp. 5329–5336, doi: 10.1609/aaai.v33i01.33015329.
- [5] P. James, "Knowledge graphs," *Linguistic Instruments in Knowledge Engineering*. Amsterdam: Elsevier, 1992, pp. 97–117.
- [6] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, G.J. Labra, S. Kirrane, S. Neumaier, A. Polleres, R. Navigli, A.N. Ngonga, S. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, and A. Zimmermann, "Knowledge graphs," *arXiv preprint arXiv:2003.02320*, 2020.
- [7] N. Creswick, J.I. Westbrook, "Social network analysis of medication advice-seeking interactions among staff in an Australian hospital," *International journal of medical informatics*, vol. 79, no. 6, pp. e116–e125, 2010, doi: 10.1016/j.ijmedinf.2008.08.005.
- [8] M. Small, P. Shi, and C.K. Tse, "Plausible models for propagation of the SARS virus," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 87, no. 9, pp. 2379–2386, 2004, doi: 10.1109/TVLSI.2004.833400.
- [9] W. Guan, Z. Ni, Y. Hu, W. Liang, and N. Zhong, "Clinical Characteristics of Coronavirus Disease 2019 in China," *New England Journal of Medicine*, 2020, doi: 10.1056/NEJMoa2002032.
- [10] F. He, Z. Hu, W. Zhang, L. Cai, G. Cai, and K. Aoyagi, "Construction and evaluation of two computational models for predicting the incidence of influenza in Nagasaki Prefecture, Japan," *Scientific reports*, vol. 7, 2017, doi:10.1038/s41598-017-07475-3.
- [11] M. Moradi Kordmahalleh, M. Gorji Sefidmazgi, A. Homaifar, D. B. Kc, and A. Guiseppi-Elie, "Time-series forecasting with evolvable partially connected artificial neural network," in *Proceedings of the Companion Publication of the 2014 Annual Conference on Genetic and Evolutionary Computation*, Vancouver BC, Canada, July. 2014, pp. 79–80, doi: 10.1145/2598394.2598435.
- [12] P. M. Sloot, S. V. Ivanov, A. V. Boukhanovsky, D. A. van de Vijver, and C. A. Boucher, "Stochastic simulation of HIV population dynamics through complex network modelling," *International Journal of Computer Mathematics*, vol. 85, no. 8, pp. 1175–1187, 2008, doi: 10.1080/00207160701750583.
- [13] A.H. Auchincloss, A.V. Diez Roux, "A new tool for epidemiology: The usefulness of dynamic-agent models in understanding place effects on health," *American Journal of Epidemiology*, vol. 168, 2008, doi:10.1093/aje/kwn118.
- [14] S. Mei, P. M. Sloot, R. Quax, Y. Zhu, and W. Wang, "Complex agent networks explaining the HIV epidemic among homosexual men in Am-



- sterdam," *Mathematics and Computers in Simulation*, vol. 80, no. 5, pp. 1018-1030, 2010, doi: 10.1016/j.matcom.2009.12.008.
- [15] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21484-21489, 2009, doi:10.1073/pnas.0906910106.
- [16] E.M. Rodríguez, F. Díaz, and M.V. Pérez, "Spatio-temporal clustering of American Cutaneous Leishmaniasis in a rural municipality of Venezuela," *Epidemics*, vol. 5, no. 1, pp. 11-19, 2013, doi:10.1016/j.epidem.2012.10.002.
- [17] A. Wesolowski, N. Eagle, A.J. Tatem, D.L. Smith, A.M. Noor, R.W.Snow, and C.O. Buckee, "Quantifying the impact of human mobility on malaria," *Science*, vol. 338, no. 6104, pp. 267-270, 2012, doi:10.1126/science.1223467.
- [18] E. Wu, W. Liu, and S. Chawla, "Spatio-temporal outlier detection in precipitation data," *International Workshop on Knowledge Discovery from Sensor Data*, Springer, Berlin, Heidelberg, pp. 115-133, 2008, doi:10.1007/978-3-642-12519-5\_7.
- [19] C.A. Donnelly, A.C. Ghani, G.M. Leung, A.J. Hedley, C.R. Fraser, S. Riley, L.J. Abu-Raddad, L.M. Ho, T.Q. Thach, P. Chau, K.P. Chan and T.H. Lam, "Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong," *The Lancet*, vol. 361, no. 9371, pp. 1761-1766, 2003, doi: 10.1016/S0140-6736(03)13410-1.
- [20] E. Kalipeni, L.C. Zulu, "HIV and AIDS in Africa: a geographic analysis at multiple spatial scales," *GeoJournal*, vol. 77, no. 4, pp. 505-523, 2012, doi: 10.1007/s10708-010-9358-6.
- [21] Y. Wu, G. Lee, X. Fu, H. Soh, and T. Hung, "Mining weather information in dengue outbreak: predicting future cases based on wavelet, SVM and GA," in *Advances in Electrical Engineering and Computational Science*. Springer, Dordrecht, 2009, pp. 483-494, doi: 10.1007/978-90-481-2311-7\_41.
- [22] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in *Proceedings of the 13th international conference on World Wide Web*, New York, NY, USA, 2004, pp. 491-501, doi: 10.1145/988672.988739.
- [23] S.C. Jin, Y.B. Tao, Y.Y. Yan, and H.R. Dai, "Infectious disease patterns analysis based on visualization of multidimensional space-time data," *Journal of Computer-Aided Design & Computer Graphics*, vol. 31, no. 2, pp. 241-255, 2019, doi: 10.3724/SP.J.1089.2019.17653.
- [24] G. Zhou, F. Chen, "Urban congestion areas prediction by combining knowledge graph and deep spatio-temporal convolutional neural network," in *2019 4th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*, Guilin, China, Apr. 2019, pp. 105-108, doi: 10.1109/ICECTT.2019.00031.
- [25] B. Jiang, G. Wan, J. Xu, F. Li, and H. Wen, "Geographic knowledge graph building extracted from multi-sourced heterogeneous data," *Acta Geodaetica Cartographica Sinica*, vol. 47, no. 8, pp. 1051-1061, Aug. 2018, doi:10.11947/j.AGCS.2018.20180113.
- [26] M. Marrero, J. Urbano, S. Sánchez-Cuadrado, J. Morato, and J.M. Gómez-Berbís, "Named entity recognition: Fallacies, challenges and opportunities," *Computer Standards & Interfaces*, vol. 35, no. 5, pp. 482-489, 2013, doi: 10.1016/j.csi.2012.09.004.
- [27] D. Clare, "Reading geography between the lines: Extracting local place knowledge from text," *Spatial Information Theory*, vol. 8116, pp. 320-337, 2013, doi: 10.1007/978-3-319-01790-7\_18.
- [28] L. Yu, F. Lu, and X. Liu, "A bootstrapping based approach for open geo-entity relation extraction," *Acta Geodaetica et Cartographica Sinica*, vol. 45, no. 5, pp. 616-622, 2016, doi: 10.11947/j.AGCS.2106.20150181.
- [29] C. Loglisci, D. Ienco, M. Roche, M. Teisseire, and D. Malerba, "Toward geographic information harvesting: extraction of spatial relational facts from web documents," in *Proceedings of the 12th International Conference on Data Mining Workshops (ICDMW 2012)*, NW Washington, DC, United States, Dec. 2012, pp. 789-796, doi: 10.1109/ICDMW.2012.20.
- [30] K. Kyzirakos, M. Karpathiotakis, and M. Koubarakis, "Strabon: A Semantic Geospatial DBMS," in *Proceedings of the 11th International Semantic Web Conference (ISWC 2012)*, Boston, MA, USA, Nov. 2012, pp. 295-311.
- [31] H. Duan, L. Meng, C. Huang, Y. Li, and J. Li, "A method for geo semantic spatial index on SPARQL query," *Acta Geodaetica et Cartographica Sinica*, vol. 42, no. 2, pp. 193-199, 2014, doi: 10.13485/j.cnki.11-2089.2014.0028.
- [32] N. Gür, T.B. Pedersen, E. Zimányi, and K. Hose, "A foundation for spatial data warehouses on the semantic web," *Semantic Web*, vol. 9, no. 5, pp. 557-587, 2018, doi: 10.3233/sw-170281.
- [33] F. Lu, L. Yu, and P. Qiu, "On geographic knowledge graph," *Journal of Geo-information Science*, vol. 19, no. 6, pp. 723-734, 2017, doi: 10.3724/SP.J.1047.2017.00723.
- [34] [34]J. Chen, S. Deng, and H. Chen, "Crowdgeo: Crowdsourced geo-knowledge graph," in *China Conference on Knowledge Graph and Semantic Computing*, Springer, Singapore, Aug. 2017, pp. 165-172, doi: 10.1007/978-981-10-7359-5\_17.
- [35] S. Wang, X. Zhang, P. Ye, M. Du, Y. Lu, and H. Xue, "Geographic Knowledge Graph (GeoKG): A Formalized Geographic Knowledge Representation," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 4, p. 184, 2019, doi:10.3390/ijgi8040184.
- [36] Y. Sun, M. Sarwat, "A generic database indexing framework for large-scale geographic knowledge graphs," in *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle, Washington, USA, Nov. 2018, pp. 289-298, doi: 10.1145/3274895.3274966.
- [37] Y. Sun, J. Yu, and M. Sarwat, "Demonstrating spindra: a geographic knowledge graph management system," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, Macau, China, Apr. 2019, pp. 2044-2047, doi: 10.1109/ICDE.2019.00235.
- [38] X. Wu, X.P. Zhao, and J.K. Yu, "Trajectory graph: a method of trajectory information identifying based on knowledge graph," *Application Research of Computers*, vol.37, no. 11, to be published, doi: 10.19734/j.issn.1001-3695.2019.07.0270.
- [39] Y.P. Xiao, D.Q. Chen, S.H. Wei, Q. Li, H.H. Wang, and M. Xu, "Rumor propagation dynamic model based on evolutionary game and anti-rumor," *Nonlinear Dynamics*, vol.95, no. 1, pp. 523-539, 2019, doi: https://doi.org/10.1007/s11071-018-4579-1.
- [40] Y. Xiao, Q. Yang, C. Sang and Y. Liu, "Rumor Diffusion Model Based on Representation Learning and Anti-Rumor," *IEEE Transactions on Network and Service Management*, doi: 10.1109/TNSM.2020.2994141.
- [41] Y.P. Xiao, C.G. Song, and Y.B. Liu, "Social hotspot propagation dynamics model based on multidimensional attributes and evolutionary games," *Communications in Nonlinear Science and Numerical Simulation*, vol.67, pp.13-25, 2019, doi: https://doi.org/10.1016/j.cnsns.2018.06.017.
- [42] Y. Gao, S. Gao, R. Li, and Y. Liu, "A semantic geographical knowledge wiki system mashed up with google maps," *Science China Technological Sciences*, vol. 53, pp. 52-60, 2010, doi: 10.1007/s11431-010-3208-0.
- [43] Y. Liu, R. Li, K. Chen, Y. Yuan, L. Huang, and H. Yu, "KIDGS: A geographical knowledge-informed digital gazetteer service," in *Geoinformatics 2009: The 17th International Conference on Geoinformatics*, Fairfax, Virginia, USA, vols 1 and 2, Aug 2009, doi:10.1109/GEOINFORMATICS.2009.5293495.
- [44] H. Lin, M. Chen, "Managing and sharing geographic knowledge in virtual geographic environments (VGEs)," *Annals of GIS*, vol. 21, no. 4, pp. 261-263, 2015, doi: 10.1080/19475683.2015.1099568.
- [45] L. You, H. Lin, "Towards a research agenda for knowledge engineering of virtual geographical environments," *Annals of GIS*, vol. 21, no. 3, pp. 163-171, 2016, doi: 10.1080/19475683.2016.1199594.
- [46] L. Robert, "A conceptual framework for geographic knowledge engineering," *Journal of Visual Languages & Computing*, vol. 25, no. 1, pp. 2-19, 2014, doi: 10.1016/j.jvlc.2013.10.004.
- [47] B. Jiang, L. Tan, Y. Ren, and F. Li, "Intelligent interaction with virtual geographical environments based on geographic knowledge graph," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 10, p. 428, Sep. 2019, doi: 10.3390/ijgi8100428.
- [48] J. Liu, C. Chen, and J. Xu, "Spatiotemporal analysis of trajectories of guqin celebrities based on crowdsourcing data," *Journal of Geo-information Science*, vol. 21, no. 6, pp. 844-853, 2019, doi: 10.12082/dqxkx.2019.180575.
- [49] J.N. Liu, H.Y. Liu, X.H. Chen, X. Guo, W.Y. Guo, X.M. Zhu, Q.B. Zhao, and J. Li, "Terrorism Event Model by Knowledge Graph," *Geomatics and Information Science of Wuhan University*, to be published. doi: 10.13203/j.whugis20190428.
- [50] L.C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35-41, 1977, doi: 10.2307/3033543.
- [51] Z. Song, H. Duan, Y. Ge, and X. Qiu, "A novel measure of centrality based on betweenness," in *2015 Chinese Automation Congress (CAC)*, Wuhan, China, Jan. 2016, pp. 174-178, doi: 10.1109/CAC.2015.7382491.
- [52] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks," in *2010 international conference on advances in social networks analysis and mining*, NW Washington, DC, United States, Aug. 2010, pp. 176-183, doi:10.1109/ASONAM.2010.17.

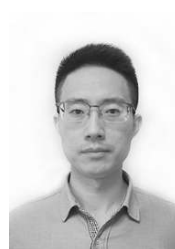
- [53] M.E.J. Newman, M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 26113, pp. 1-16, 2004, doi: 10.1103/PhysRevE.69.026113.



of spatiotemporal big data and Geographic knowledge graphs.

BINGCHUAN JIANG received B.S. and M.S. degrees in cartography and geographic information engineering from the Institute of Geospatial Information, Information Engineering University, China, in 2005 and 2009, respectively, and a Ph.D. degree from the Institute of Geospatial Information, Strategic Support Force Information Engineering University, China, in 2013. He is currently a lecturer at Information Engineering University.

His research interests include the Visual analysis



ographical Environments.

XIAOJUN ZHOU received B.S. and M.S. degrees in cartography and geographic Information engineering from the Institute of Geospatial Information, Information Engineering University, China, in 2004 and 2008, respectively, and a Ph.D. degree from the Institute of Geospatial Information, Strategic Support Force Information Engineering University, China, in 2014. He is currently a lecturer at Information Engineering University. His research interests include 3D GIS and Virtual Ge-



virtual Reality, and Scientific Visualization.

LIHENG TAN received a B.S. in Cartography and Geographic Information Systems from the School of Resource and Environmental Sciences, Wuhan University, in 2009, and an M.E in Control Science and Engineering from the College of Information Systems and Management, National University of Defense Technology, in 2011. He is currently a lecturer at Information Engineering University. His research interests include Computer Graphics, 3D GIS, Virtual Geographical Environments, Vir-

...



and geographic information systems.

XIONG YOU received B.S. and M.S. degrees in cartography and geographic information engineering from the Institute of Surveying and Mapping, China, in 1983 and 1991, respectively, and a Ph.D. degree from the Institute of Surveying and Mapping, China, in 1994. He is currently a professor and Doctoral Supervisor with the Institute of Geospatial Information, Strategic Support Force Information Engineering University. His research interests include virtual geographic environments



systems, deep learning, and remote sensing image analysis.

KE LI received B.S. and M.S. degrees in cartography and geographic information engineering from the Institute of Surveying and Mapping, China, in 1999 and 2002, respectively, and a Ph.D. degree from the Institute of Surveying and Mapping, China, in 2008. He is currently an Associate Professor and a Master's Supervisor with the Institute of Geospatial Information, Strategic Support Force Information Engineering University. His research interests include geographic information



TINGTING LI received bachelors and master's degrees in English Translation and English Literature from Wuhan University and Zhengzhou University in 2017 and 2019, respectively. She is now a teaching assistant at Information Engineering University. Her research interest is the application of translation variation strategies.