Knowledge graph-based method for identifying topological structure of low-voltage distribution network

Gao Zepu1, Luo Yongjian1, Xu Ziwei1, Yu Yilan1, Zhang Lianmei1

1School of Electrical Engineering and Automation, Wuhan University, Wuhan, 430072 Hubei, People’s Republic of China

Abstract: The correct topological relationship is crucial in the low-voltage distribution network, as the actual topological structure of the low-voltage distribution network changes frequently and tremendously due to the need of operation and maintenance, and it cannot be correctly reflected upon failure in timely updating of data, low circulation, poor quality etc. Therefore, it is necessary to identify the topology. The knowledge graph technology can clearly reflect the existing relationship between data, deducing and mining hidden knowledge, suitable for topology identification of the low-voltage distribution network. In the study, the knowledge graph technology was employed for topology identification: firstly, analyse the construction method of the knowledge graph, integrate data in multiple low-voltage distribution network information systems based on the knowledge graph technology, deduce missing data, find out the relationship between data, and then build the knowledge graph of low-voltage distribution network topological structure, and finally, based on ‘Typical Design Specification of Low-Voltage Distribution Network Infrastructure Project’ and semantic segmentation, identify the user–transformer relationship in low-voltage distribution network information system. The test results of the examples were very satisfactory, showing the theoretical values and practical application values of the identification method proposed in this study.

Nomenclature

| Symbol | Definition |
|--------|------------|
| O_i    | category i ontology |
| P_i    | attribute set of category i ontology |
| R_i    | relation attribute set of category i ontology |
| P_i_e  | unique identifying entity |
| P_i_r  | relationship identification |
| E_j    | entity j of category i ontology |
| P_i_j  | attribute set of entities |
| P_j_v  | attribute value set of entities |
| P_j_s  | entity self-tagging attribute |
| P_j_f  | entity connection flag attribute |
| R_{j,m} | relationship between entities or entity and attribute |
| G      | triplet |
| E      | entity set |
| P      | relationship set |
| S      | attribute set |
| T_m    | transformer |
| B_m    | building set |
| U_m    | user set |
| N_T    | number of transformers |
| S_{SN} | transformer capacity |
| cos φ  | power factor of transformer |
| N_{user} | number of user meters |
| P_l    | typical load of low-voltage users |
| N_{BD} | number of transformers meeting specification |

1 Introduction

In a low-voltage distribution network, the correct recording of the topological structure by its information system is the premise of fine management and safe operation of the power grid [1]. In the topological structure, the most important factor is the user–transformer relationship. At present, the topological data in the low-voltage distribution network is manually entered, so the correctness of the information cannot be guaranteed. Due to the continuous expansion of the scale of the power grid, new energies and distributed power supply connected for operation and other reasons, the power distribution network topology maintains the dynamic changes, and the data volume of marketing, measurement, geographical information system (GIS) etc. increase sharply, the correctness of data cannot be guaranteed, the systems operate independently, and the circulation between data is poor, making it difficult to verify the topological structure, therefore, topology identification is urgently required [2].

Topology identification is for verification of the correctness of topological data of the low-voltage distribution network in the information system. The current topology identification methods can be divided into online methods and offline methods. The offline methods are used for onsite identification with hardware equipment as transformer area topology identification device, branch line monitor unit etc. featuring high cost, high demand for labour resources and low efficiency. The online methods are usually used for topology identification with program-based topological information in the system, including the user–transformer power frequency zero-crossing correlation analysis, for determining the correctness of topology according to the changes of the voltage waveform of users in a certain area; the historical power failure event correlation judgment, determining the correctness of topology through analysing the actual reaction for power failure events of users in an area of the information system; and methods for verifying topological correctness based on measurement of power and load data, and monitoring the real-time status of the breaker [3–8]. With the emergence of advanced metering infrastructure, a huge amount of data of different types at fixed intervals can be captured, and other methods were introduced, e.g. [9]. The method for determining the correct topology with the correlation between voltages at users; the method for determining the topological correctness with power flow algorithm by identifying the best topology and comparing it with the existing topological structure; the voltage clustering-based method for determining topological structure with data from smart meters; the method of BC Hydro (Canada), which developed an intelligent data analysis unit, embedded in the intelligent electric meter, for verifying and correcting the low-voltage power distribution network topology structure in the GIS [10, 11]. Such methods are intended for topology identification by analysing the related information in the measuring system, with low cost and...
high real-time level, and have gradually become one of the hot spots of current topology identification. However, the low-voltage power distribution network currently shows low quality in data acquisition, and such online methods have higher requirements for the quantity and quality of data, rely on data acquisition and occupy communication channels. The methods for topology identification vary from one another in the advantages and disadvantages, and scenarios and effect of application, and there is not a single method capable of solving all the problems of identifying low-voltage topological structure in the actual complicated operation scenarios, so multiple identification methods must be used in the actual applications [12].

A topology identification method based on knowledge graph was first proposed in this study. The knowledge graph technology can clearly analyse the relationship between entities, and through mining and inferring the relationship, makes up the missing information, determines the correctness of existing information, and identifies the hidden internal connections in the system, fully meeting the requirements for topology identification [13]. Compared with existing methods, the proposed method could overcome the disadvantages of requiring massive high-quality operation data and occupying communication channel of online identification, and by utilising the characteristics of the knowledge graph, also reduced the dependence of topology identification on data, and showed the excellent identification results for specific scenarios [14]. The advantages and applicability of the method were also verified with the simulations.

The rest of the paper is organised as follows: Section 2 provides a brief introduction to the knowledge graph technology. Section 3 introduces knowledge graph building methods for topology structure of low-voltage power distribution network. Section 4 describes the topology identification methods based on knowledge graph. Section 5 evaluates the method proposed with simulation algorithm, and Section 6 is the summary.

2 Knowledge graph

2.1 Definition and status quo

With the rapid development of computer science and internet technology and the explosive growth of the volume of data, the knowledge graph was introduced to deal with relevant issues [15]. It consists of a series of different graphs showing the knowledge development process and the structural relationship, used to describe knowledge resources and their carriers, and also mine, analyse, construct, draw and display knowledge and their interrelations with the visualisation techniques [16]. The knowledge graph can demonstrate the internal structure of knowledge and the connection between knowledge dynamically and intuitively.

Now knowledge graphs are mostly applied in the internet field, but few are applied in the power industry, in particular, for the core functions of power grids, such as marketing, GIS, measurement etc. and there are no relevant applications of knowledge graphs [17].

2.2 Categories of knowledge graph

According to different application fields, the knowledge graph can be divided into the general knowledge graph and the industrial knowledge graph [18].

The general knowledge graph is not limited to a specific field, with wide knowledge coverage and complex networks of knowledge, usually constructed with a bottom-up approach. The most representative knowledge graph is Google Knowledge Graph, containing 500 million entity objects and 35 billion relationship information, and the typical representative knowledge graph in China includes Baidu Zhixin, Sogou Zhilifang etc. and there are no relevant applications of knowledge graphs [19].

The industrial knowledge graph is targeted at specific fields, with low breadth but deep depth of knowledge and fixed data sources and pertinence [19], usually built in a combination of top-down and bottom-up approaches. The examples of industrial knowledge graph include Geonames, containing the most comprehensive geological knowledge in the world, Linked Movie Database, the world largest movie knowledge graph etc. [20].

2.3 Construction method

The methods for building knowledge graph include the following: manual construction, or construction based on swarm intelligence, internet-linked data, machine learning or information extraction [21].

The construction process of knowledge graph mainly includes the following steps:

Knowledge extraction: Extract knowledge elements, such as entities, relationships, attributes and so on from the source data, and transform them into machine-readable knowledge.

Knowledge fusion: Disambiguate and integrate knowledge, as knowledge acquired from different sources may be contradictory.

Knowledge storage: After the fusion, link knowledge with each other, and store as triplets in the graph database.

Knowledge inference: On the basis of the knowledge graph, conduct intelligent applications such as link prediction, knowledge QA etc. for improvement and expansion of existing knowledge graph.

3 Construction of knowledge graph of low-voltage distribution network topology

3.1 Construction process

The knowledge graph of the low-voltage distribution network topology belongs to the industrial knowledge graph, which contains a large number of physical equipment, such as transformers, feeders etc. The data of these physical equipment and its attributes come from the existing distribution network information system database (marketing, measurement, GIS etc.) of the grid company, recorded in the formats of natural language, numbers and geographic coordinates etc. [22]. As revealed in the extensive surveys, the database used in the distribution network information system is usually the structured database, e.g. Oracle database. Structured databases store data in a fixed and rigorous manner. Therefore, for constructing a knowledge graph of low-voltage distribution network topology, the following improvements were made to the conventional construction methods:

(i) In the topology of the low-voltage distribution network, based on the CIM structure (common information model) and the physical devices, first establish the relationship between physical devices, then considering the attributes of physical devices, build the knowledge graph with the ‘top-down’ approach.

(ii) The structured database is adopted in the distribution network information system, and the knowledge extraction of structured data is relatively simple. While extracting the knowledge of database data, also extract the fields in the database.

(iii) There are system barriers between the information systems of the distribution network, so fuse the knowledge extracted from multiple information system databases, to improve the same entity attributes from different information systems, and distinguish the same name but different entity or the different name but the same entity.

(iv) Conduct ontology modelling for the physical devices involved in the knowledge graph of low-voltage distribution network topology. The ontology is the abstract concept of such devices and contains all the attributes. The ontology set is as small as possible, and the relationship between the ontology can be artificially defined.

(v) Based on the constructed ontology, make the knowledge extracted from the database and fused correspond with the ontology, to form the entity. Due to the fixity of structured data, construct the relationship between entities and attributes through sample data training.

(vi) Store the triplet consisting of the constructed entities, attributes and relationship (entity, relationship and entity) or form (entity, relationship and attribute) in the graph database, and construct the knowledge graph of low-voltage distribution network topology.
(vii) Display the knowledge graph of low-voltage distribution network topology with the visualisation tool, analyse the relationship between knowledge, and deduce the errors in the low-voltage distribution network topology.

Fig. 1 shows the method for improving the construction of the knowledge graph of low-voltage distribution network topology. According to the characteristics of the data source, the characteristics of the low-voltage distribution network topology were analysed, and the conventional construction method of knowledge graph was improved, making it more targeted. After construction, the knowledge graph of low-voltage distribution network topology is more professional and practical, and the specific implementation methods of each step will be described below.

3.2 Ontology construction $O_i$

Ontology is the formal representation of concepts, the structure of concepts, and the relationship between concepts, used as the general template for building entities. The ontology library is used in the knowledge graph to store ontologies. The ontology construction structure is shown in Fig. 2.

By analysing the physical equipment involved in the low-voltage distribution network topology and its information, the ontologies of the knowledge graph of low-voltage distribution network topology constructed in this paper can be divided into four categories: transformer, transformer area (the master meter of the transformer area), feeder and the meters of low-voltage users. The ontologies to be constructed were extracted from the database of the low-voltage distribution network information system. The fields of various data types were extracted, and made corresponding to the four ontology categories, and ontological attributes not found in the database can be added artificially. This kind of ontology is formed by connecting and combining the data type with its ontology.

The expression of ontology construction is shown in formula (1):

$$O_i = (P_i, R_i)$$  \hspace{1cm} (1)

where $i$ is the ontology SN established for user–transformer relationship verification, $i = \{1, 2, 3, 4\}$, $O_1$ is the transformer ontology, $O_2$ is the transformer area ontology, $O_3$ is the feeder ontology, $O_4$ is the meter of low-voltage user, $P_i$ is the attribute set of ontology, and $R_i$ is the relationship attribute set of ontology $O_i$.

Each ontology $O_i$ has an attribute $P_{ix} = \text{ID}$ for uniquely identifying the entity, as well as an attribute $P_{iy} = \text{RID}$, the relation identity.

In this paper, Protégé software was employed to assist in building ontologies, including: (i) four types of physical devices and their common attributes and private attributes, i.e. transformers, transformer area, feeders, and the meter of low-voltage user; (ii) versatility relationship between devices, device attributes and attributes. The constructed ontology library is versatile, available to save storage space. The relationship between ontology of the low-voltage distribution network topologies is shown in Table 1, and the ontology construction results are shown in Fig. 3. As shown in Fig. 3, the four categories of constructed ontologies and their relationship can be seen, the attribute type of each ontology is hidden in the ontology, and the attribute of an ontology can be viewed.

3.3 Entity construction $E_{ij}$

Knowledge was extracted from the structured database of marketing, measurement, GIS and other information systems. The fields in the structured database represent the data type of each column of the database sheet. Each record (row) of the database sheet is the description of an entity, and each record is divided into multiple entity attributes (values) by each column. The extracted knowledge was classified, and the knowledge is associated with the ontology according to the constructed ontology to form an entity in the low-voltage distribution network topology, e.g. 1# transformer (length, starting point, end point, related transformer area etc.), 1# transformer area (number of users, name, geographical location etc.); 1# transformer area (line loss, load).

The expression of entity construction is shown in formula (2):

$$E_{ij} = (P_{ij}, \text{PV}_{ij})$$  \hspace{1cm} (2)

where $i$ is same as that in formula (1), $E_{ij}$ is the $j$th entity of ontology $O_i$, $P_{ij}$ is the attribute set of the $j$th entity of ontology $O_i$, and $\text{PV}_{ij}$ is the attribute value set of the $j$th entity of ontology $O_i$.

There is a one-to-one correspondence between $P_{ij}$ and $\text{PV}_{ij}$.

Each entity $E_{ij}$ has a self-tagging attribute $P_{ijx} = \text{ID}$, and a connection flag attribute $P_{ijy} = \text{RID}$, and $\text{PV}_{ijx}$ is unique in the ontology $E_{ij}$. For example, under the low-voltage user meter ontology $O_4$, all the low-voltage user meter entities contain a transformer connection flag attribute $P_{ijx} = \text{TransferID}$, which

| Table 1 Ontological relations of low-voltage distribution network topology |
|---------------------------------------------------------------|
| **Ontology** | **Relationship** | **Ontology** |
| transformer | separation line | feeder |
| transformer | power supply range | transformer area |
| transformer area | belong | transformer user |
| transformer area | contain | meter of low-voltage user |
| feeder | connect | transformer |
| feeder | power transmission | transformer area |
| meter of low-voltage user | belong | transformer area |

![Fig. 1 Improved method for constructing knowledge graph of low-voltage distribution network topology](image)

![Fig. 2 Ontology building structure](image)
indicate that it belongs to the entity $O_1$; and the low-voltage user meter ontology $O_2$ contains the attribute $P_{10} = \text{Address}$, which indicates the power address of the low-voltage user meter.

The above-mentioned information system databases are independent of each other, and different attributes of the same entity may come from multiple independent databases, so the entity of each distribution network information system was built separately, the self-tagging attribute of each entity was compared, and the entities with the same self-tagging attribute $P_{10}=\text{ID}$ were fused into one. The step is advantageous in reducing the workload and completing the entity disambiguation and data fusion at the same time.

### 3.4 Relationship construction $R_{ij}$

The relationship is the core of the knowledge graph, and also the key for solving the problem with the knowledge graph. Relationship construction includes the relationship between entities, and the relationship between entities and attributes.

The relationship between entities can directly inherit the relationship between the corresponding ontologies, e.g. the ‘power supply range’ in the ‘transformer–power supply range – transformer area’ is the relationship between the ontology of the transformer and that of the transformer area, and after inheritance, the relationship between entities is ‘transformer A – power supply range – transformer area B’.

The relationship between the entity and the attribute can be formed with the fields extracted according to the knowledge and relevant modifiers, e.g. ‘low-voltage user meter A – voltage value is – 220 V’, in which ‘voltage value is’ is the relationship between the entity ‘low-voltage user meter A’ and its attribute value ‘220 V’, to be obtained by extracting fields from the database with knowledge plus the relevant modifiers.

The expression of relationship construction is shown in formula (3):

$$ R_{ij,xmn} = \{E_{ij},E_{mn}\} $$

where $i$ has the same meaning as in formula (1), $m$ has the same meaning as $i$, $n$ and $j$ are the location of data in the category, $E_{ij}$ and $E_{mn}$ are the entities or attributes, and $R_{ij,xmn}$ is the relationship between the entities or attributes, also including the relationship attribute set $R_i$ between ontologies artificially created in Section 2.2.

### 3.5 Completing knowledge graph construction

After construction of the ontology, entity and relationship, the entity and entity, entity and attribute were connected with the relationship, combined into a triplet and stored in the graph database, to form the knowledge library of the low-voltage distribution network topology. The form of the triplet is shown in formula (4):

$$ G = \{E,R,S\} $$

where $G$ is the triplet set, $E$ is the set of entities, which includes $|E|$ kinds of different entities, and each entity has a unique ID, $R$ is the set of relationships, which includes $|R|$ kinds of relationships, and $S$ is the set of attributes.

The user–transformer relationship between the low-voltage user meter and the transformer can be inferred from the relationship between the transformer and the transformer area and the relationship between the transformer area and the low-voltage user meter, expressed as: the low-voltage user meter entity $E_{1n}$ ‘belongs to’ the ‘power supply area’ of the transformer entity $E_{1n}$.

The triplet in the knowledge library, including two forms, i.e. (Entity 1, relationship and Entity 2) and (entity, relationship and attribute), displays the complicated relationship diagram between entities in a visual way to form the knowledge graph of the low-voltage distribution network topology.

### 4 Identification method for low-voltage distribution network topology

#### 4.1 Flow for topology identification

In topology identification of low-voltage distribution network, the most important goal is to verify the user–transformer relationship between the low-voltage user meter and the transformer, select the low-voltage user meter with the wrong user–transformer relationship, and conduct the manual inspection and confirmation at the site. At present, there are two main types of errors in the user–transformer relationship of the low-voltage distribution network:

(i) The user–transformer relationship recorded in the low-voltage distribution network information system does not match the actual user–transformer relationship.

(ii) The user–transformer relationship information of some low-voltage user meter in the low-voltage distribution network information system is partly missing.

By analysing the relationship between entities in the low-voltage distribution network topology, the knowledge graph can identify any error of the user–transformer relationship in the distribution network information system; and through knowledge inference and mining, add the missing user–transformer relationship information of the low-voltage user meter in the distribution network information system.

In the completed knowledge graph of the low-voltage distribution network topology mentioned above, each transformer area, transformer, feeder and low-voltage user meter were linked with the respective relationships, with a close storage distance for the low-voltage user meter and the transformers of the same transformer area in the graph database, convenient for data retrieval, viewing and analysis. Based on the constructed knowledge graph and the existing data, the power addresses of the low-voltage user meters in the transformer area were standardised with the address standard, followed by the semantic segmentation. Comparative analysis was made to the results of semantic segmentation of the power addresses of all the low-voltage user meters in the same transformer area, and the low-voltage user meters not meeting ‘Typical design specification of the low-voltage distribution network infrastructure project’ were picked out, to achieve the goal of topology identification. The relevant flowchart is shown in Fig. 4.

#### 4.2 Semantic segmentation and adding attributes

The power address of the low-voltage user meter is manually entered by the power grid personnel, not accurate or standard. According to ‘Rules for Describing Standard Address of National Geographic Information’ [23], the power address was divided into ten layers (Levels 1–10) in the study, with the territory gradually reducing from Level 1 to Level 10. The meaning of each layer is shown in Table 2. The power address of all low-voltage user meters in the same transformer area, and the low-voltage user meters not meeting ‘Typical design specification of the low-voltage distribution network infrastructure project’ were picked out, to achieve the goal of topology identification. The relevant flowchart is shown in Fig. 4.

![Fig. 3 Example of ontology construction](https://example.com/fig3.png)
In the knowledge graph, there is a correlation between the low-voltage user meter of the same transformer area or residential area. For a non-standard power address, it can be standardised by comparing with the standard power address in the same transformer area or residential area. Take Fig. 5 as an example.

In Fig. 5, the low-voltage user meters A and B belong to the same transformer area or residential area, and the low-voltage user meter B power address is not standard, with only the address layers of Levels 6–10. By comparing and analysing with the power address of meter A, the power address of meter B can be completed.

For the low-voltage user meter with missing power address, the power addresses of all low-voltage user meters under the same transformer were called, the number of floors and users were inferred and each user on each floor was corresponded with the transformer and area or residential area, and the low-voltage user meter B power address is not standard, with only the address layers of Levels 6–10. By comparing and analysing with the power address of meter A, the power address of meter B can be completed.

According to the ‘Data specification for geo-entity, geographic name and address’, the standardised power addresses were grammatically segmented. Same as the levels for standardising addresses, the address segmentation is also divided into ten levels. Levels 1–10 addresses after segmentation were added to the knowledge graph of the low-voltage distribution network topology as the new attribute of the low-voltage user meter. For example: {User1, Level1, Value1_1} means that the value of the attribute Level 1 of the low-voltage user meter is Value1_1, and so on for the attributes and values of other address levels.

### 4.3 Identification process

Based on the knowledge graph, the building set \( B_m \) (\( B_m = \{ B_m1, \ldots, B_mn \} \)) involved in all users under the transformer \( T_m \) in a transformer area were extracted, and all the users \( U_m = \{ UB_m1, L, UB_mn \} \) included in \( B_m \) were identified. When the attribute value of Levels 1–6 is the same, they are the meters of low-voltage users of the same residential area; when the attribute value of Levels 1–7 is the same, they are the meters of the low-voltage users of the same building; when the attribute value of Levels 1–8 is the same, they are the meters of the low-voltage users of the same unit of the same building; and when the attribute value of Levels 1–9 is the same, they are the meters of the low-voltage users of the same floor in the same unit of the same building.

When it is the low-voltage user meter of the same building, the number of transformers \( N_B \) for this building can be counted according to the relationship in the knowledge graph, as shown in formula (5):

\[
N_B = \begin{cases} 
1 & \text{if } BD \geq 1 \\
2 & \text{otherwise} \\
\text{others} 
\end{cases} 
\]

According to the transformer capacity \( S_{KN} \), the transformer design power factor \( \cos \phi \), the number of user meters \( N_{user} \), and the typical load \( P_{DN} \) of low-voltage users specified in ‘Typical design of distribution network infrastructure at 10 kV and below’ [24], it can be concluded that: there are few users in the middle and lower floors, so one transformer can meet the demand for electricity, and the low-voltage user meters of the same building in the same residential area belong to a same transformer. The low-voltage user meters that do not meet the design rules were picked out and the following corrections were made: change the transformer of the low-voltage user meter with error to the TransformerID with more correct user–transformer relationship of the same building or same floor. The address of the low-voltage user meter under the same transformer is the same building or the same floor, and the number of user meters of each floor in the same building should be the same. Comparing the number of user meters in each floor, the floor with fewer user meters is the low-voltage user meter in the distribution network information system with missing user–transformer relationship information. The user–transformer relationship information of this low-voltage user meter is then changed to the Transformer ID of other low-voltage user meters with correct user–transformer relationship in the same floor.

### 4.4 Application scenarios

The topology identification method based on knowledge graph described in this paper is applicable to the following application scenarios:

| Layer | Meaning |
|-------|---------|
| Level 1 | province, autonomous region, etc. |
| Level 2 | city etc. |
| Level 3 | county, district etc. |
| Level 4 | town, township etc. |
| Level 5 | street, road, avenue etc. |
| Level 6 | residential area, garden etc. |
| Level 7 | building etc. |
| Level 8 | unit number etc. |
| Level 9 | floor number etc. |
| Level 10 | room etc. |

### 4.3 Identification process

Based on the knowledge graph, the building set \( B_m \) (\( B_m = \{ B_m1, \ldots, B_mn \} \)) involved in all users under the transformer \( T_m \) in a transformer area were extracted, and all the users \( U_m = \{ UB_m1, L, UB_mn \} \) included in \( B_m \) were identified. When the attribute value of Levels 1–6 is the same, they are the meters of low-voltage users of the same residential area; when the attribute value of Levels 1–7 is the same, they are the meters of the low-voltage users of the same building; when the attribute value of Levels 1–8 is the same, they are the meters of the low-voltage users of the same unit of the same building; and when the attribute value of Levels 1–9 is the same, they are the meters of the low-voltage users of the same floor in the same unit of the same building.

When it is the low-voltage user meter of the same building, the number of transformers \( N_B \) for this building can be counted according to the relationship in the knowledge graph, as shown in formula (5):

\[
N_B = \begin{cases} 
1 & \text{if } BD \geq 1 \\
2 & \text{otherwise} \\
\text{others} 
\end{cases} 
\]
(i) **Low-voltage users and transformer area of urban residents:** In the transformer areas consisting of urban residents, the power addresses are usually closely correlated, with certain patterns to follow, so the results are highly accurate after inference with the knowledge graph.

(ii) **High building:** With the development of economy, the height of buildings increases continuously. The buildings with >30 or 40 floors are usually powered with 2–4 transformers, and it is prone to cause the wrong user–transformer relationship in the floors where the power transformers are interfaced. The knowledge graph-based topology identification method can be used to identify and modify the error information.

(iii) **Urban residential area:** The power addresses of low-voltage user meters in the residential area are highly standardised, so it is easy to standardise the non-standard addresses. With the knowledge graph-based topology identification, the user–transformer relationship of each user of each building in the whole residential area can be identified, with the good identification result.

### 4.5 Advantages of the algorithm

The knowledge graph-based topology identification method described in this paper features high accuracy, low cost, low data volume, no occupation of the communication channel, high practicability, low dependence on data volume and data quality, and availability for adding data in the information system of the low-voltage distribution network. It can be widely applied in urban areas, and also used for other applied studies based on the established knowledge graph.

### 5 Analysis of examples

#### 5.1 Basic information of calculation examples

In order to test the effect of the topology identification method based on knowledge graph, the simulation data and the actual data of a power supply company were used for testing in the study.

The basic information of the calculation examples is as follows:

(i) **Simulation data:** Totally, the simulation involved 3 transformer areas, 6 transformers, 7 feeders and several low-voltage user meters, with about 400 entities. Some of the low-voltage user meter information was artificially set as error information, for testing the effect of the knowledge graph-based topology identification method.

(ii) **Data from grid company:** The actual data in the GIS, marketing, measurement and other information systems of a power supply company was used, and three typical transformer areas were chosen, including 432 low-voltage user meters. As it was impossible to conduct the field visit to verify the correctness of existing grid data, the existing information data was deemed as correct data. The user in a certain transformer area of the existing data was artificially adjusted to another transformer area, as the targets of the topology identification method based on the knowledge graph.

The standardisation rate of power addresses and the success rate of topological identification were both 100%. In case the standard power address only accounts for 9.5%, the standardisation rate of the power address could be increased to 100% through conducting semantic

### 5.2 Analysis of simulation data examples

In the simulation example, some representative attributes were selected from each ontology as the simulation template, with the specific attributes shown in Table 3. To verify the application scope of the method used, the height of floors were set differently, i.e., low floor (Floor 6), middle floor (Floor 12) and high floor (Floor 24).

The simulation data was imported into the database as the original data. According to the knowledge graph construction method described above, the low-voltage distribution network topology knowledge graph was built, including 387 entity nodes, 1935 attribute nodes and 2814 relationships. Then the power addresses of the low-voltage users in the knowledge graph were standardised, with the results shown in Table 4. The standardised addresses were hierarchically stored in the knowledge graph as a newly added attributes of the low-voltage user meter. Then the topology identification analysis was performed on the target transformer or transformer area, with the identification results shown in Fig. 6 and Table 5.

As it can be seen from Fig. 6 and Table 5, the standardisation rate of the power address and the success rate of topology identification were both 100%. In case the standard power address only accounts for 9.5%, the standardisation rate of the power address could be increased to 100% through conducting semantic

### Table 3 Representative attributes of ontology

| Ontology/Transformer area | Transformer model | Feeder name | Low-voltage user meter number | attribute | name | name | name | name |
|--------------------------|------------------|-------------|-----------------------------|-----------|------|------|------|------|
| type (public/ special transformer) | rated primary voltage, V | rated secondary voltage, V | rated capacity, kVA | equipment model | transformer area | transformer electricity consumption, voltage |
| operation state | consumption | transformer id | address | transformer address | transformer area | electricity consumption |

### Table 4 Results of standardisation of power address (simulated)

| Condition | Number of standard addresses | Number of non-standard addresses | Standardisation rate of power addresses, % |
|-----------|------------------------------|---------------------------------|------------------------------------------|
| before    | 36                           | 342                             | 9.52                                     |
| after     | 378                          | 0                               | 100                                      |

![Fig. 6 Topology identification result diagram](http://creativecommons.org/licenses/by-nc/3.0/)
5.3 Analysis of actual data calculation examples of power grid companies

The test results of simulation data in Section 5.2 were satisfactory, with the significant and complete distribution network characteristics, and no interference factors. However, the actual conditions in the distribution network are more complicated and variable, therefore, the actual data of the power grid was used to further verify the practicability and adaptability of the knowledge graph topology identification method, and the comparison and analysis were also made with the Pearson's correlation coefficient method.

According to the actual data of the distribution network information system, the low-voltage distribution network topology knowledge graph was constructed. In the actual applications, the power addresses are entered at will, with poor standardisation level. The invalid address means that the low-voltage user meter does not have the power address attribute, and it can only be supplemented according to the semantic relationship between the power addresses of the entire building. As seen from Table 6, actually the non-standard address can be converted into a standard address using the knowledge graph and semantic analysis; and for the invalid address, the only way is to make up the missing part of the account number in the user address of the entire building, unit or residential area, so the standardisation rate of the address cannot reach 100%.

The topology identification results are shown in Fig. 7 and Table 7. When there are few low-voltage user meters under the transformer area, the success rate of identification is low, as the amount of data is low, and the degree of correlation between the relationships is relatively low, therefore the success rate of identification is relatively low. In addition, the power address cannot be 100% standardised, which is also the reason why the success rate of topology identification in Areas A and C did not reach 100%.

An example of the topology identification process with the method of the study is shown in Fig. 8. The power addresses of the low-voltage user meters under No. 301 transformer in transformer Area C are all located in Building 14 of Funing Park, and the transformer recorded in the distribution information system of No. 503 in the 5th floor of Building 14 and No. 902 in the 9th floor of Building 14 was No. 602 transformer; and by acquiring the power addresses of all low-voltage user meters under No. 602 transformer, they are all the users in Building 3 of Funing Park, so it can be judged that the low-voltage user meters of No. 503 in the 5th floor of Building 14 and No. 902 in the 9th floor of Building 14 have the wrong user-transformer relationship, and it should be modified to the same Transformer ID = Transformer 301 as most of the low-voltage user meters in Building 14.

To further verify the effect of topology identification under specific scenarios with the proposed method, we extracted the data of 30 transformer areas from a certain grid company, all of the modern residential areas or high buildings, with each transformer area including at least 40 low-voltage user meters. Error messages were artificially set for the 30 transformer areas, and topology identification was conducted with the proposed method, with the identification results shown in Table 8.

Table 8 shows the results for verifying the effect of topology identification by the proposed method with a great deal of data from transformer areas, and the success rate of topology identification was all over 80%, mostly over 90%; and the more the number of low-voltage user meters was in the transformer area, the higher the success rate was, demonstrating the satisfactory identification results under specific scenarios with the method.

For the specific scenarios and the scenarios applicable to the study, the test was conducted using the actually measured voltage data (no power interruption, sampling once every 15 min, with the data period of 14 days) according to the Pearson’s correlation coefficient method [25], and the results of identification are shown in Table 9.

The success rate of topology identification with Pearson's correlation coefficient method largely relied on the specific wrong topological scenarios. In the scenarios with short electrical distance...
in residential areas, buildings etc. the success rate of topology identification decreased significantly. The voltage data used in Pearson's correlation coefficient method was the data volume of each user measured 96 times per day for 14 days, with all the data transmitted to the identification centre via the communication channel, while the power address data of low-voltage user meters used in the study could be directly read from the database by the identification centre, without occupying the communication channel.

In the actual applications, multiple methods must be adopted for identification. The identification results with the method of this paper can be verified with those obtained with the Pearson's correlation coefficient method, and the reliability in identifying the wrong user with two methods is higher than that with a single method.

6 Conclusion

In this paper, knowledge graph technology was introduced into the low-voltage distribution network, and applied to topology identification of low-voltage distribution network for the first time, and a topology identification method based on knowledge graph was proposed. The contributions made in the study include: (i) detailed description of an improved construction process of the knowledge graph of low-voltage distribution network topology; (ii) study on semantic segmentation method of power addresses; (iii) a knowledge graph topology identification method based on the above method.

As seen in the calculation examples of simulation data and actual data, this method could basically achieve >90% success rate in topology identification, demonstrating the high standardisation rate and high accuracy for typical high building scenarios; and it can improve the data quality of low-voltage distribution network information system, and be widely used in high buildings, residential areas and other range of urban residents. For typical high building scenarios, the proposed method features the high success rate and low cost in topology identification, low reliance on data volume and data quality and no occupation of communication channel, with the significant superiority and practicability, as compared with existing topology identification methods.

If more rules on topology identification are developed, studying attribute relationships other than the power address can further improve the integrity of the knowledge graph and the accuracy of topology identification, and the knowledge graph can also be applied in other aspects of the low-voltage distribution network to solve problems, which is also the focus of future studies.

7 References

[1] Liu, Y., Zhang, N., Kang, C.: ‘A review on data-driven analysis and optimization of power grid’, Autom. Electr. Power Syst., 2018, 42, (6), pp. 157–167
[2] Lang, Y., Li, J., Luo, Y., et al.: ‘Large power grid topology analysis based on graph partitioning’, Power Syst. Prot. Control, 2017, 45, (23), pp. 108–115
[3] Li, Y., Xu, L., Li, Y.: ‘Topology connecting strategy of the modular multilevel matrix converter for optimal switching times’, Electr. Power Eng. Technol., 2018, 33, (2), pp. 34–39
[4] Liao, Y., Wang, J., Wu, M., et al.: ‘Distribution grid topology reconstruction: an information theoretic approach’, North American Power Symp., Charlotte, NC, USA., 2015
[5] Luan, W., Peng, J., Maras, M., et al.: ‘Distribution network topology error correction using smart meter data analytics’, 2013 IEEE Power and Energy Society General Meeting (PES), Vancouver, BC, Canada, 2013
[6] Ying, J., Mei, J., Wang, Y., et al.: ‘Research on verification method of feeder topology model for distribution main station’, Power Syst. Prot. Control, 2018, 46, (7), pp. 83–89
[7] Pappu, S.J., Bhatt, N., Paumary, R., et al.: ‘Identifying topology of low voltage (LV) distribution networks based on smart meter data’, IEEE Trans. Smart Grid, 2018, 9, pp. 5111–5122
[8] Lourenço, E.M., Coelho, E.F.R., Pal, B.C.: ‘Topology error and bad data processing in generalized state estimation’, IEEE Trans. Power Syst., 2015, 30, pp. 3910–3920
[9] Yan, Q., Wang, J., Huang, K., et al.: ‘Analysis of equivalent topology and power- flow calculation method of unipolar power load’, Electr. Power Eng. Technol., 2018, 33, (23), pp. 5523–5531
[10] Li, Y., Feng, B., Li, G., et al.: ‘Optimal distributed generation planning in active distribution networks considering integration of energy storage’, J. Energy, 2018, 210, pp. 1073–1081
[11] Sharon, Y., Annaswamy, A., Legbedji, M.A., et al.: ‘Topology identification in distribution network with limited measurements’, 2012 IEEE PES Innovative Smart Grid Technologies, IGIT 2012, Washington, DC, USA., 2012, pp. 1–6
[12] Zhou, D., Zhang, M., Zhu, H., et al.: ‘Research and application of intelligent distribution network dispatching interactive under the background of electric power system reformation’, Electr. Power Eng. Technol., 2018, 37, (2), pp. 89–94
[13] Liu, Q., Li, Y., Duan, H., et al.: ‘Knowledge graph construction techniques’, J. Comput. Res. Dev., 2016, 53, (3), pp. 582–600
[14] Liu, Z., Wang, H.: ‘Retrieval method for defect records of power equipment based on knowledge graph technology’, Autom. Electr. Power Syst., 2018, 42, (14), pp. 158–164
[15] Zhu, G., Iglesias, C.A.: ‘Sematch: semantic similarity framework for knowledge graphs’, Knowl.-Based Syst., 2017, 130, pp. 30–32
[16] Singh, M.: ‘Protection coordination in distribution systems with and without distributed energy resources – a review’, Prot. Control Mod. Power Syst., 2017, 2, (1), p. 27
[17] Uyar, A., Aliyu, F.M.: ‘Evaluating search features of Google knowledge graph and Bing satori’, Online Inf. Rev., 2015, 39, (2), pp. 197–213
[18] Shi, B., Weninger, T.: ‘Discriminative predicate path mining for fact checking in knowledge graphs’. Knowl.-Based Syst., 2016, 104, pp. 123–133
[19] Li, X., Xu, J., Guo, S.: ‘Construction and application of knowledge graph of power dispatching automation system’, Electr. Power, 2019, 52, pp. 1–8
[20] Wang, C., Ma, X., Chen, J., et al.: ‘Information extraction and knowledge graph construction from geoscience literature’, Comput. Geosci., 2018, 112, pp. 112–120
[21] Shi, Z., Zhang, L., Hu, X., et al.: ‘Power system transient stability rules extraction based on multi-attribute decision tree’, Trans. China Electrotech. Soc., 2019, 34, pp. 1–12. Available at https://doi.org/10.19958/j. cjeet.1000-6753.tcn.180646
[22] Yang, S., Hao, R.: ‘A visual analysis of the status quo and trend of knowledge mapping research’, Intell. Data Work, 2012, 33, (4), pp. 22–28
[23] GBT 23505-2009, Coding Rules for Place Names/Addresses of Digital City Geographic Information Public Platform.
[24] Typical design of infrastructure project for distribution network of 10 kV and below in Guangzhou Power Supply Bureau.2017-9-30
[25] Xiono, Y., Zhao, Y., Tu, Z., et al.: ‘Topology checking method for low voltage distribution network based on improved Pearson correlation coefficient’, Power Syst. Prot. Control, 2019, 47, (11), pp. 37–43

Table 8 Topology identification results of multiple transformer areas

| Success rate of topology identification, % | Number of low-voltage user meters included | Number of transformer areas |
|------------------------------------------|-----------------------------------------|-----------------------------|
| over 95                                   | 40–70                                   | 2                           |
|                                          | 70–100                                  | 5                           |
|                                          | over 100                                | 6                           |
| 90–95                                     | 40–70                                   | 1                           |
|                                          | 70–100                                  | 8                           |
|                                          | over 100                                | 6                           |
| 80–90                                     | 40–70                                   | 2                           |
|                                          | 70–100                                  | 0                           |
|                                          | over 100                                | 0                           |

Table 9 Identification results of Pearson's correlation coefficient method for voltage

| Specific scenarios of the study | Number of messages set wrongly | Number of identifications with Pearson's correlation coefficient method | Success rate of topology identification, % |
|---------------------------------|-------------------------------|------------------------------------------------------------------|-------------------------------------------|
|                                 | 12                            | 11                                                               | 91.67                                      |

This is an open access article published by the IET under the Creative Commons Attribution -NonCommercial License (http://creativecommons.org/licenses/by-nc/3.0/)