Received March 4, 2022, accepted March 21, 2022, date of publication March 24, 2022, date of current version March 31, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3161987

Hierarchical Distribution Network Topology Formulation and Dimensionality Reduction Using Homeomorphism Transformation

JINMING CHEN[1,2], (Member, IEEE), WEI JIANG[1], (Senior Member, IEEE), ZHIQI XU[1], YE CHEN[2], (Member, IEEE), HAO JIAO[2], (Member, IEEE), MINGHUA WANG[1], (Graduate Student Member, IEEE), YUBO YUAN[2], (Member, IEEE), AND WU CHEN[1], (Senior Member, IEEE)

[1]School of Electrical Engineering, Southeast University, Nanjing, Jiangsu 210096, China
[2]State Grid Jiangsu Electric Power Company Ltd., Research Institute, Nanjing, Jiangsu 210096, China

Corresponding author: Jinming Chen (cjm_nju@163.com)

ABSTRACT The scales of the power distribution networks in real-world power grids expand quickly while the network structures are becoming more and more complex. The power grid companies analyze the power distribution networks in different business scenarios with different topology models. In this work, we propose a hierarchical graph model to describe the medium-voltage distribution network (which is a typical power distribution network in power grids) based on homeomorphic transformation. The hierarchical graph model preserves the basic network topology described by the traditional Common Information Model (CIM). Firstly, the nodes in the distribution network topology are classified according to graph theory. Secondly, three typical business scenarios of distribution network topology analysis are summarized, and the original model is simplified by progressive dimensionality reduction method to meet the analysis requirements of different scenarios, the simplified method consists of three abstract levels: critical path, core path and minimal path, and can effectively reduce the space complexity of the model while maintaining the topological properties. Thirdly, a multi-level distribution network topology construction and mapping method based on the graph database is proposed. It is used to realize the rapid conversion and traceability between different levels of topology. Finally, a practical distribution network in a county is used as an example to verify the effectiveness of the proposed method in the aspects such as topology rendering and path searching. The evaluation indicates that the proposed model can visualize the distribution network intuitively. The model can also speed up the visualization and path searching significantly.

INDEX TERMS Dimensionality reduction, distribution network, homeomorphism transformation, topology formulation.

I. INTRODUCTION

A. BACKGROUND

For power grid companies, the power distribution network is the final link to the users, and its quality directly relates to the user experience. In order to improve the reliability of power supply, the medium-voltage distribution network is designed in a closed-loop manner and is operated in the open-loop topology. The power distribution network contains a large number of tie switches and segment switches, which makes the distribution network form a complex network [1], [2]. The topology structure of the power distribution network is the basis of analysis and visualization, thus it is important for distribution network operation and management [3]–[6].

During power grid management, the requirements and granularity of different business scenarios are different. The scopes and levels of the involved power grid topology are also quite different.

- Distribution network planning [7], [8] is a typical macro-level business application. It analyzes at a large spatial scale, spanning many regional distribution networks, power supply units and feeders.
- Distribution network operation mode optimization [9] is a typical meso-level business application. The analysis

The associate editor coordinating the review of this manuscript and approving it for publication was Ravindra Singh.

scope spans multiple feeders that are directly or indirectly connected, and the analysis granularity is a feeder segment.

- Distribution network inspection and maintenance [10] is a typical micro-level business application. It focuses on the feeder-level topology to locate and analyze equipment or components such as towers, ring network cabinets (ring main unit), and distribution transformers.

In order to realize the automatic management of the power grid, the existing power grid analytic systems are mostly based on the "connection point-terminal" model of IEC CIM [11] to model and store the topology of the distribution networks. Most of the existing systems use a relational database to store the topology model. Each row in the topology data table represents the connection relationship between a certain device and adjacent connection points or terminals. The topology analysis over power distribution networks requires iterative queries of the topology table.

However, the existing power grid analytic systems cannot meet the modeling and storage requirements of the more and more complex business applications. In terms of distribution network topology modeling, different power grid business applications focus on different topology ranges. They require a multi-level distribution network topology model. However, the existing CIM model can only provide the most fine-grained connection relationship between power grid equipment, which is only suitable for micro-level analysis. It hardly meets the analysis needs of meso-level and macro-level applications. In terms of distribution network model storage, the current relational database storage performance is difficult to meet the real-time requirements of complex distribution network topology analysis and visualization rendering applications. The multi-hop neighborhood and path analysis in those applications will generate a large number of traversal and join operations in the relational database, resulting in high I/O overhead and long calculation time.

B. RELATED WORK

In order to support multi-level and multi-type power grid management applications, a variety of distribution network topology modeling and storage methods have been proposed by academia and industry. Among the CIM-based topology models, single-line diagrams and contact diagrams are the mostly used topology models in the current engineering applications. The single-line diagram [12], [13] focuses on the micro-level topology and is widely used in production management systems. It describes the details and connection relationships of equipment in a single feeder, but it cannot intuitively describe the connection characteristics of the distribution network. The connection diagram [14], [15] focuses on describing the macro grid characteristics of the regional distribution network. However, the rendering speed of connection diagram visualization is slow. The dense arrangement of equipment in connection diagrams results in a poor visualization experience. It is also difficult to establish connections between the single-line diagrams and the connection diagrams. Literature [16] proposes a meso-level topology model centered around substations. Its scope is between the single-line diagram and the connection diagram. However, this model cannot intuitively describe the feeder connection relationships across substations.

To simplify the distribution network topology models, Literature [17] propose regarding on-column switches as nodes and feeders between adjacent nodes and distribution transformers as edges. Literature [18] propose a topology simplification method using feeder blocks and equivalent branches to improve the efficiency of reliability assessment calculations. Literature [19] proposes a topology simplification method for the business scenario of failover recovery. The method shrinks the topology of distribution network branch lines and merges load points. Literature [20] simplify the distribution network topology by evaluating the effect of edge partitioning on graph connectivity.

However, the above distribution network topology modeling methods fail to provide a systematic and comprehensive modeling method. The existing modeling methods focus on a single level. They cannot adapt to various power grid business applications on different levels, like transmission tower optimization between cities [21] and power grid visualization [22].

In terms of distribution network topology storage, in order to overcome the shortcomings of relational database storage, Literature [23] propose a power system database framework based on the CIM model and graph database. A graph database is a non-relational database. It stores relational information between entities based on a graph model. Through distributed parallel computing, it has unique performance advantages in dealing with massive and complex business scenarios [24]. The power grid is a natural graph, which is suitable for modeling and calculation using a graph database. Literature [25] use a simplified CIM/E model to effectively reduce the complexity of graph databases.

However, the above storage methods mainly concentrate on small-scale or medium-scale large power grids. The power distribution networks have much larger scales than large power grid. The existing storage methods do not optimize their storage design for large-scale distribution networks.

C. CONTRIBUTIONS

To meet the needs of multi-level analysis in power grid business applications on large-scale power distribution networks, in this paper, we propose a hierarchical graph model for distribution network topology based on homeomorphic transformation. In this paper, we focus on modeling the medium-voltage distribution network in power distribution networks. The medium-voltage distribution network is complex and large with multiple kinds of equipment in it. We further propose a group of techniques to construct the hierarchical graph model efficiently and store the model in a graph database. We summarize our contributions in this paper as follows:

(1) We propose a homeomorphic transformation method to reduce the dimension of the ''connection point-terminal'' model used by the CIM model. The homeomorphic transformation can preserve the electrical connectivity characteristics of the power grid. It is the basis of the simplified operation of the distribution network topology (Section 2).

(2) Based on the homeomorphic transformation, we further propose three topological levels for power distribution networks. The three levels are developed around the concept of critical path, core path and minimal path, respectively. They can support the analysis of micro, meso and macro distribution network business, respectively. We connect the three levels with an inherent mapping to form a systematic hierarchical graph model for power distribution networks. Compared with the existing modeling methods, our method supports different levels of analysis at the same time and also supports cross-level business applications. (Section 3)

(3) We propose three distribution network topology simplification methods to construct the three levels in the hierarchical graph model from the original CIM model. We also propose a mapping schema to store the hierarchical graph model in a graph database. The mapping schema supports both the single-plane topology query and cross-plane topology analysis. (Section 4)

(4) We evaluate the effectiveness and efficiency of the proposed hierarchical graph model and the graph-based storage method over a real-world county-level regional distribution network. We adopt visualization analysis and transfer path search as two typical analysis application scenarios in our evaluation. The evaluation shows that the proposed hierarchical model can significantly reduce the topology complexity in medium and macro business scenarios, achieving exponential improvements in topology loading and query efficiency. The transfer path search application shows that the mapping schema that we use in the graph database can converse and trace between different topology levels in a short time. Compared with the traditional relational database storage, our method can reduce the calculation time of transfer path search by up to 3 orders of magnitude. (Section 5)

To sum up, the advantages of the modeling and storage methods proposed in this paper over the existing methods are in two aspects. On the one hand, our method supports micro, meso and macro levels of power grid analysis at the same time. On the other hand, our method can effectively improve the data querying efficiency and reduce the running time of power grid analysis applications.

The rest of the article is organized as follows. Section II introduces the definition of basic topology model of medium voltage distribution network. The typical scenarios of distribution network topology analysis and their corresponding simplified methods are introduced in section III. The application of three business scenarios in graph database and case analysis are described in Sections IV and V respectively. Finally, Section VI is reserved for conclusions and discussions.

II. DEFINITION OF THE BASIC GRAPH MODEL OF THE MVDN

The topology of the MVDN is in form of a hierarchical structure. Basically, the feeder consists of basic components like switches, transformers, breakers and power lines, and is connected to a bus at specific substation. Several feeders are interconnected with contact point. A power supply unit usually includes several related feeders. A regional distribution network includes multiple power supply units. With the development and expansion of the urban area, new power supply units are planned, designed, constructed and connected to the exits network. Thus, the modern distribution system is more complex than the transmission system. In this paper, we focus on the primary equipment in the distribution network, and do not consider secondary equipment or automation equipment, such as voltage/current transformers, remote terminal unit and fault indicator.

A. 'CONNECTIVITY NODE – TERMINAL' MODEL

The topology of a distribution network can be modeled by the graph $G = (V, E)$, where V and E refer to the node set and edge set respectively. According to the different meanings given to V and E, there are different kinds of graph models of the distribution network topology. IEC CIM is one of them. It uses an object-oriented method to define the basic components, and the 'Connective node – Terminal' model to describe the connection relationship among them, i.e., each conductive component is associated with one or more terminals, and each terminal is associated with a connectivity node.

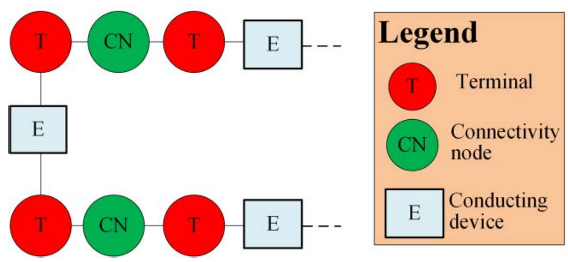

FIGURE 1. Connectivity node-terminal model.

As shown in Figure 1, the connectivity node is the point where several terminals of the conductive components are connected with zero impedance. It helps build up the connections among different conductive components.

The MVDN mainly consists of conducting components and buses. Conducting components include breakers, switches, power lines, distribution transformers, etc. Buses include the ends of the power lines, buses of the transformers and substations. In the CIM model, all these components are connected using the connectivity nodes and the terminals.

Figure 2 is a partial diagram of a single feeder, whose connectivity node-terminal model is shown in Figure 3. We denote N as the number of the conducting components (N_1: distribution transformers, N_2: switches, N_3: power lines), M as the number of the bus components, K as the number of connectivity nodes, and usually $K \gg M$.

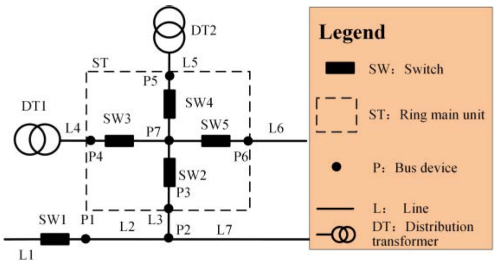

FIGURE 2. Single feeder diagram (partial).

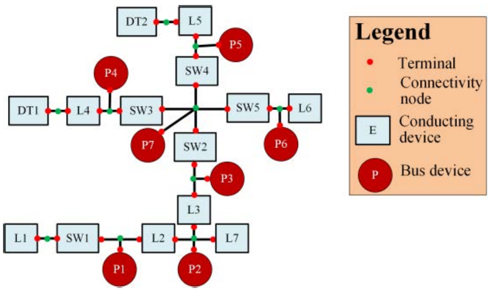

FIGURE 3. Connectivity node-terminal model of figure 2.

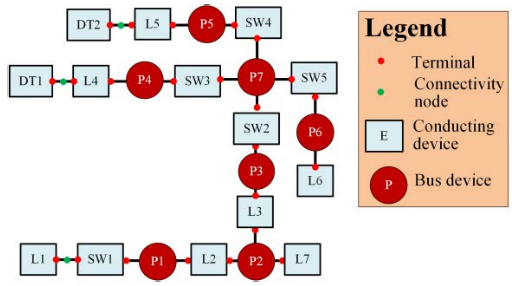

FIGURE 4. Connectivity node-terminal model after the bus contraction transformation.

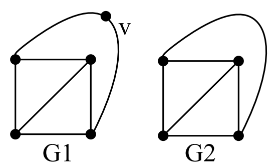

FIGURE 5. Example of homeomorphism transformation.

Since each distribution transformer, switch, power line and bus have 1, 2, 2 and 1 terminals respectively, the total number of terminals is $T_{\text{num1}} = N_1 + 2^* N_2 + 2^* N_3 + M$. Hence the total number of nodes in the graph shown in Figure 3 is

$$|V_1| = N + M + T_{\text{num1}} + K = 2N_1 + 3(N_2 + N_3) + 2M + K \quad (1)$$

It should be noted that in the CIM model, the bus component is connected to the connectivity point with the terminal. This part can be simplified as a bus contraction model. After the transformation, Figure 3 can be converted to Figure 4, where the number of terminals is $T_{num2} = N_1 + 2^*(N_2 + N_3)$, and the total number of the graph is

$$|V_2| = N + T_{\text{num2}} + K = 2N_1 + 3(N_2 + N_3) + K \quad (2)$$

B. TOPOLOGY SIMPLICATION MODEL BASED ON HOMEOMORPHIC TRANSFORMATION

In the graph theory, homeomorphism refers to the preservation of all the topological characteristics of a given space after mapping operation such as continuous extension and bending, between two topological spaces. That is, all topological properties of a given space are preserved by repeated expansion and contraction of 2-degree nodes, as defined in Appendix A. Figure 5 shows an example of homeomorphism transformation. The graph $G1$ is contracted to the graph $G2$ with the 2-degree node v, and the graph $G2$ is expanded to the graph G1 with the 2-degree node v. Therefore, the graph $G1$ and $G2$ are homeomorphic.

In the model shown in Figure 4, the connectivity nodes and terminals are virtual nodes without physical meaning. For this reason, the model can be further simplified using the homeomorphism transformation. The steps are shown as follows.

Step1-1: With the contraction of a graph with a 2-degree node, the two-degree virtual nodes are merged into the virtual edges. Only the nodes that represent components are kept in the graph.

Step1-2: With the homeomorphic transformation, the nodes that represent bus components and the virtual edges near them are merged into 'physical edges'.

After the two steps, Figure 4 is converted into Figure 6. In this model, the number of the nodes in the graph is

$$|V_3| = N + M = N_1 + N_2 + N_3 + M \quad (3)$$

because K ≫ M, we have K > 2M, and then

$$|V_2| = 2\,N_1 + 3(N_2 + N_3) + K > 2(N_1 + N_2 + N_3 + M) \quad (4)$$

According to equation (3) and (4), we have $|V_2|/|V_3| > 2$. This means that the number of nodes in the compacted model is less than 1/2 of the one of the original model. Moreover, no physical information is lost in the contraction. This paper uses this compacted model as the graph model of the MV distribution network.

III. DISTRIBUTION NETWORK TOPOLOGY ANALYSIS TYPICAL SCENARIOS AND CORRESPONDING TOPOLOGY LEVELS

The MVDN has a large volume of equipment, complex connection relationships, and its topology changes frequently. Topology is important basic information about the distribution network. The difference in application scenarios leads to differential topology resolution requirements. This paper divides the application scenarios into three levels: micro, meso and macro levels, which correspond to the three topology levels: feeder level, power supply unit level and regional distribution network level. The power supply unit is an independent power supply area composed of multiple interconnected feeders, which is defined in Appendix B.

These service scenarios involve a wide range of power supply areas, a difference of 3∼5 orders of magnitude in device volumes, and different analysis characteristics.

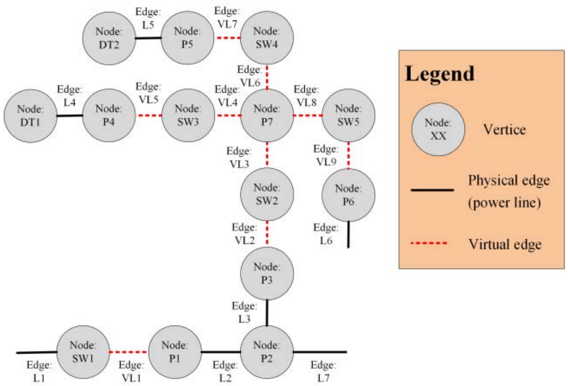

FIGURE 6. Simplified model based on shrinkage of terminals and connectivity nodes.

TABLE 1. Three levels of the scenarios.

Hierarchy	Typical scenarios	Topological level	Analysis granularity	Negligible topology details
Micro	Power line inspection	Single feeder	component	-
Meso	Load transfer analysis	Contact unit	feeder segment	details inside the feeder segment
Marco	Network planning	Regional distribution network	feeder	details not related to the connection relationship in feeders

Table 1 lists the three scenarios and topology analysis characteristics, including analysis granularity and negligible topology details. In order to improve the performance requirements of various scene topology analysis and visual rendering, hierarchical topology abstraction method can be used to reduce dimension compression.

In this section, a three-level paths (critical path, core path and minimal path) dimensionality reduction method is proposed. It takes a single feeder as the basic object, and the three-level paths are all based on the corresponding abstract model of feeders.

A. FEEDER-LEVEL MICRO SCENARIO (ORIGINAL PATH)

Micro scenarios are generally limited to single feeders, and typical micro applications include equipment asset management, daily line inspection, and line high loss and theft analysis. Therefore, they can be analyzed independently feeder by feeder. The granularity of analysis is at the equipment level, requiring the components and connection relationships within the feeder to be as complete and detailed as possible.

The distribution network topology corresponding to the micro business scenario is the basic graph model based on the contraction and simplification of terminals and connection points proposed in Section II, which is called "original path".

B. POWER-SUPPLY-UNIT-LEVEL MESO SCENARIO (CRITICAL PATH)

Meso scenarios can generally be limited to the inside of the power supply unit of the distribution network. Typical meso applications include planned maintenance of distribution

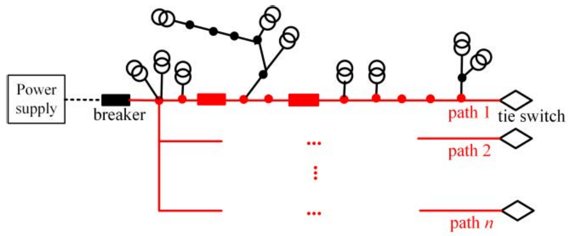

FIGURE 7. An example of the main path of the feeder.

networks, fault outage recovery, and optimization of heavy-haul line operation, etc. These applications involve cross-feeder analysis, so the model of a single feeder cannot satisfy the requirement. The analysis range should be expanded to multiple feeders connected by tie switches. In these applications, the analysis resolution is the feeder segment controlled by tie/section switches, because the details inside the feeder segment are negligible.

1) RELATED TOPOLOGY NODE DEFINITIONS

Definition 1 Main path of the feeder:

The set of paths formed by circuit breaker nodes connected to feeder trees and tie switch nodes is called main paths of the feeder.

As is shown in Figure 7, $\text{path}_1 \sim \text{path}_n$ are formed by accessing n tie switches from the breaker. The main path $G_{\text{f}-\text{main}}$ is obtained by taking the union of these paths, that is

$$G_{\text{f-main}} = \bigcup\nolimits_{i=1}^{n} path_i \tag{5}$$

The main path is the principal structure of a feeder. Obviously, the main feeder path $G_{\text{f}-\text{main}}$ is a subtree of the original topology model G_{f}. The root node of $G_{\text{f}-\text{main}}$ is the breaker, and the leaf nodes of it are the tie switches. Feeder main path is an important intermediate form in the simplified evolution of distribution network topology from original path to critical path.

Definition 2 Bifurcation vertex (BV):

The node which is on the main path of the feeder, and of which the degree is greater than or equal to 3, is the bifurcation vertex.

Definition 3 Load subtree, load vertex (LV):

For the node N_d on the main path of the feeder, if there is a subtree with this node as the root and one or more subtrees with leaf nodes on the original path, and the other nodes in the subtree are not on the main path, the subtree is called a load subtree, and the node N_d is a load node (LV). If a load node is also a bifurcation node, the node is called a bifurcation node with load instead of an ordinary load node.

Definition 4 relay vertex (RV):

The non-load, non-switch 2-degree vertex on the main path of the feeder, such as cable intermediate joints, cable terminal heads, ordinary relay towers, are defined as the relay vertex (RV).

Definition 5 Feeder segment:

The path between any two adjacent switch (breakers, section switch and tie switch) vertices or bifurcation vertices is defined as the feeder segment.

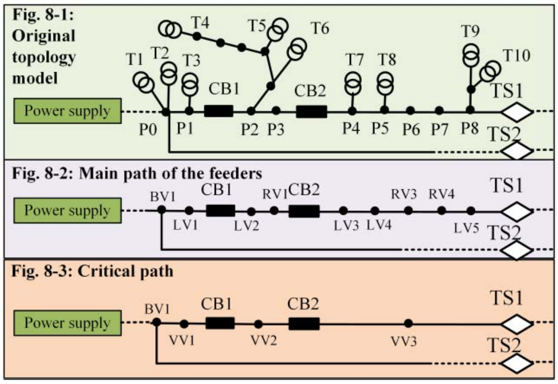

FIGURE 8. Critical path abstraction process.

Definition 6 Virtual Load Vertex (VV):

When the feeder segment with a load vertex is abstracted into an edge, a vertex will be added to the edge to represent the loads in the feeder segment. This vertex is defined as the virtual load vertex.

2) CRITICAL PATH SIMPLIFICATION METHOD

The meso-level scenario focuses on the analysis of the transfer capacity between feeders, and the load characteristics of each segment on the feeder. Therefore, the internal structure of the feeder segment can be ignored. The goal of the critical path abstraction is to prune the original path, that is, neglect the load vertex, load subtrees, and relay vertex in it. The steps of generating the core path from the original topology model are as follows:

Step 2-1: Extract the main path and load subtrees from the original path, and label BV, LV and RV, and remove all load subtrees.

Step 2-2: Traverse the main path to identify all feeder segments, and combine the load vertices and relay vertices on each feeder segment into virtual load vertices.

Figure 8 shows the process of generating the feeder main path from the original path and eventually simplifying it to a critical path. In the part of original path shown in Figure 8-1, there are 9 vertices (P0∼P8) and 10 distribution transformers. First, the main path is identified, and P0 is labeled as BV1, P3, P6, P7 are labeled as RV1∼RV3, P1, P2, P4, P5, P8 are labeled as LV1∼LV5, and all the load subtrees are removed, the main path of the feeder is displayed, as shown in Figure 8-2. Then, the feeder segments {BV1,CB1}, {CB1,CB2}, {CB2,TS1} are identified, and the nodes in these segments is combined into the virtual load vertices VV1∼VV3 to obtain the critical path, as is shown in Figure 8-3.

It can be seen that the critical path is transformed from the main path of the original model through homeomorphism transformation. This process effectively simplifies the original topology without losing any critical information. The number of vertices in the topology graph is also reduced significantly. If the original path $G_{\text{Prim}} = (\text{V}_{\text{p}}, \text{E}_{\text{p}})$ has N vertices, including k tie switches, x section switches and y bifurcation vertices, its space complexity is $\text{O}(N)$. After $\text{G}_{\text{Prim}} = (V_{\text{p}}, E_{\text{p}})$ is abstracted as $G_{\text{Key}} = (V_{\text{k}}, E_{\text{k}})$, its space complexity decreases to $\text{O}(k+x+y)$, obviously $N \gg k+x+y$. Therefore, the space complexity of the critical path is much lower than that of the original path. The space complexity of the critical path is proved in Appendix C.

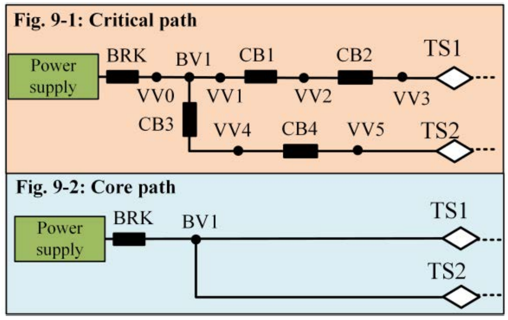

FIGURE 9. Core path abstraction process.

C. REGIONAL-LEVEL MACRO SCENARIO (CORE PATH/MINIMAL PATH)

A typical macro application scenario is the investment planning in distribution networks, including network reliability analysis, new substation positioning, upgrading of the weakness area in the distribution network. The whole MVDN, which consists of buses and feeders in each substation in the area, needs to be analyzed. The analysis granularity is power supply units or feeders, which means the analysis focuses on the feeder connection relationship instead of other details inside the feeders.

Note that some scenarios require the macro topology at first, and then degrade into meso and micro topology step by step. Therefore, the switching between topologies with different abstraction levels is necessary for analysis and visualization.

1) CORE PATH SIMPLIFICATION METHOD

The objective of core path abstraction is to shield feeder segment information in the critical path, and only preserve the core nodes involved in the feeder profile, namely, circuit breakers, bifurcation nodes and tie switches. The path formed by adjacent core nodes is called the core feeder segment. The core feeder segment is identified by traversing the critical path of the feeder, and the segmented switches and virtual load nodes in each core feeder segment are shrunk to generate the core path of the feeder.

Figure 9 shows the process of generating the core path from a simplified abstraction of the critical path. The critical path shown in Figure 9-1 contains three core feeder segments, which are{BRK,BV1},{BV1,TS1}, {BV1,TS2}. The core path in Figure 9-2 is obtained by neglecting VV0 in {BRK,BV1}, VV1∼VV3, CB1, CB2 in {BV1,TS1}, and VV4, VV5, CB3, CB4 in {BV1,TS2}.

It can be seen that the core path is transformed from the critical path through homeomorphism transformation. The number of vertices is reduced while retaining the original topological properties. If the critical path $G_{Key} = (V_k, E_k)$ has k tie switch, x section switch and y bifurcation vertices, its space complexity is O($k + x + y$). After GKey $= (V_k, E_k)$ is abstracted as $G_{Core} = (V_c, E_c)$, its space complexity decreases to O(k). Usually, the number of section switches in the distribution network is several times the number of the tie switches. The spatial complexity proof of the core path is shown in Appendix D.

2) MINIMAL PATH SIMPLIFICATION METHOD

The complexity of topology is significantly decreased after simplifying the original topology to the critical path and to the core path. However, the urban distribution network often adopts high reliability network structures, such as the double-loop network connection. In these network structures, there are many parallel-connected tie switches, which increase the complexity of topology analysis and visualization. Based on the core path, topology complexity can be further reduced by simplifying the connection switch nodes.

Definition 7 Equivalent tie switch, virtual tie switch:

In the critical path, if two bifurcation vertices on different feeders are connected by several parallel tie switches, these switches are the equivalent tie switch. The connection between the two bifurcation vertices can be simplified as an edge with a virtual tie switch on it.

The steps of generating the minimal path are as follows:

Step 3-1: Check the core path of the feeders in a power supply unit

Case 1: Bifurcation vertex BV1 is in the feeder GCore1, and bifurcation vertex BV2 is in the feeder GCore2. There are q parallel tie switches between BV1 and BV2. They are equivalent tie switches, and are combined into a virtual tie switch, and go to Step 3-2;

Case 2: If the situation in case 1 doesn't exist, the abstraction process is finished.

Step 3-2: The degree of BV1 and BV2 decreases after the combination of the equivalent tie switches:

$$BV1.degree = BV1.degree - q + 1$$
$$BV2.degree = BV2.degree - q + 1$$

If the degree of any of the vertices decreases to 2, it becomes a rely vertex, and can be neglected by using the steps in Section III.B.

Step 3-3: Go back to Step 3-1 to look for any other equivalent tie switches again.

In Figure 10, there are 4 Ring Main Units (RMUs) at the terminals of feeder A and feeder B. The buses of the RMUs are A3, A4, B3 and B4. These buses are connected by tie switches TS1∼TS6. According to the definition 11, TS1∼TS3 are the equivalent switches between A3 and B3, and TS4∼TS6 are the equivalent switches between A4 and B4.

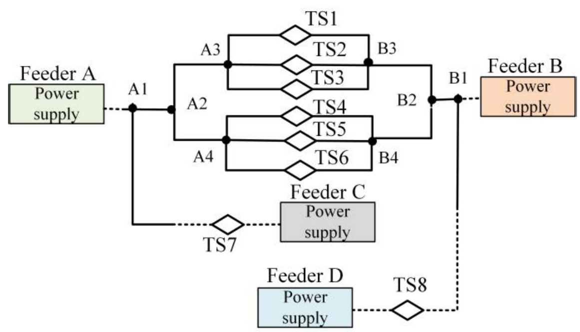

FIGURE 10. Two feeders with equivalent switches between them.

TABLE 2. Abstract method and node selection in different topology layers.

Topology level	Abstraction level	Retained vertices	Neglected vertices	Abstraction method
Single feeder (micro)	Original path	All	-	-
Power supply unit (meso)	Critical path	breakers, bifurcation vertices, tie switches, section switches, new virtual load vertices	Load vertex, load subtree, relay vertex	Load aggregation, homeomorphism transformation
Regional distribution network (macro)	Core path/ mininal path	breaker, bifurcation vertices tie switch	Section switch, virtual load vertex	Homeomorphism transformation, duplicate edge combination

The process of generating the minimal path of Figure 10 is illustrated in Figure 11. First, the equivalent switches TS1∼TS3 and TS4∼TS6 are simplified as virtual tie switches TSV1 and TSV2, as is shown in Figure 11-1. The degrees of A3, A4, B3 and B4 decrease from 4 to 2, so they are neglected, as is shown in Figure 11-2. Then, TSV1 and TSV2 become equivalent tie switches, so they are simplified as TSV3, as is shown in Figure 11-3. Meanwhile, the degree of A2 and B2 decreases from 3 to 2, so they are neglected, as is shown in Figure 11-4. Finally, there are no equivalent tie switches and 2-degree vertices in the path, and the abstraction process is completed.

If one feeder is connected to other h feeders with k tie switches, and the tie switches on the minimal path is k', then obviously $h <= k' <= k$. It means that the space complexity of the path decreases from O(k) to O(k‘).

D. TOPOLOGY DIMENSION REDUCTION SUMMARY

The service characteristics of different scenarios are introduced in the previous section. The spatial complexity of the distribution network topology is gradually reduced through continuous dimension reduction abstractions.

Table 2 lists the corresponding abstract methods and node selection at different topological levels: The original path corresponds to the microscopic topological level, including all nodes in the distribution network topology; The critical path abstract method corresponds to the meso topological level. Since the structure in the feeder segment can be ignored

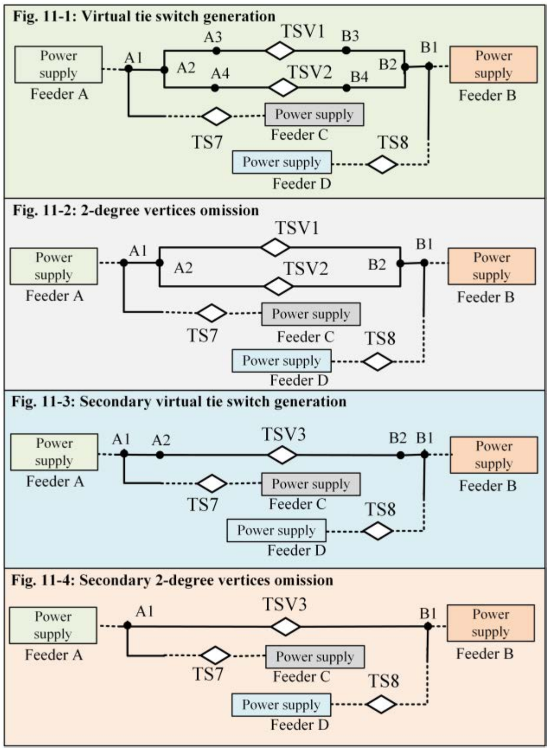

FIGURE 11. Minimal path abstraction process.

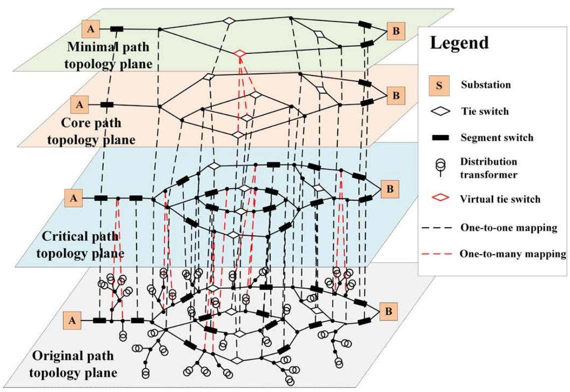

FIGURE 12. Four topology planes with mapping relationships.

in the meso scene, the load nodes, load subtrees and relay nodes in the original path can be removed. The core path (including the minimal) abstract method corresponds to the macroscopic topological level, and can ignore the structures in the feeder that have nothing to do with the contact characteristics, so only three types of nodes such as feeder circuit breaker, bifurcation node and tie switch are reserved.

IV. APPLICATION OF THE HIERARCHICAL MODELS USING NEO4J GRAPH DATABASE

Neo4j is one of the most popular graph databases. According to the above-mentioned hierarchical models, we transform the CIM model of each feeder in the regional distribution network to a vertice-edge based model and store in the Neo4j graph database.

Compared to the traditional relational database, the Neo4j graph database has significantly improved the storage and

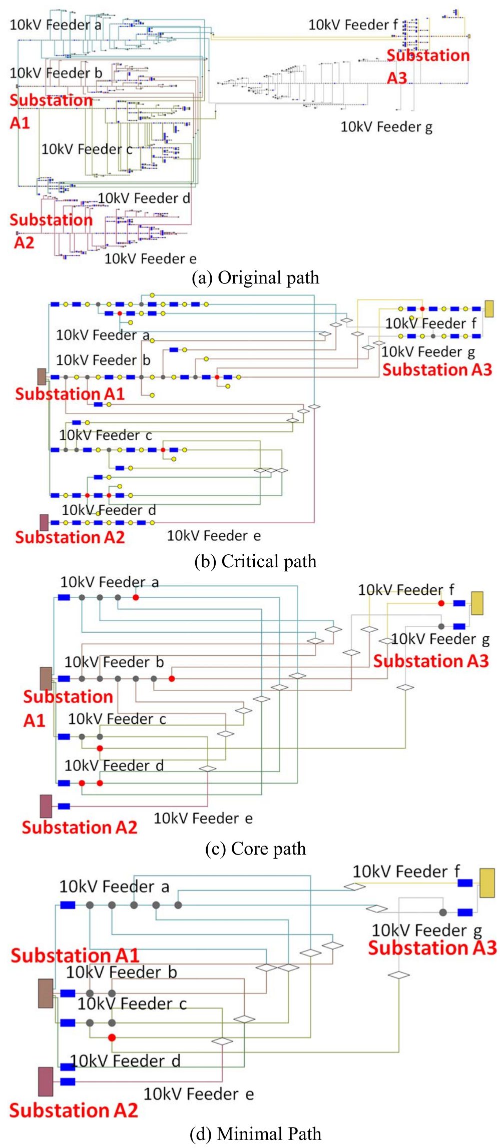

FIGURE 13. Visualized topology of the power supply unit CB003.

analysis efficiency on the distribution network topology. In the relational database, the information of each device is stored as an independent record in the table, and is bridged by connection points and terminal records to each other. Therefore, in topology analysis, it is necessary to recursively query all the related vertices and load their information into RAM. When multiple feeders are involved, the queried information need to be spliced. However, it does not require such complex operation when using Neo4j database, since there is a natural connecting relationship among the vertices in Neo4j. Based on the Cypher language, the topology of the entire network can be searched and analyzed efficiently. In addition, the Neo4j database can store data attribute values for the vertices and edges through key-value pairs. Moreover, personalized labels can also be defined on the vertices and edges, so the Neo4j database has rich description capability on topology.

In practical applications, by adding labels such as 'original path', 'critical path', 'core path', and 'minimal path' to the

TABLE 3. The number of vertices in different topology plane.

Network range	Number of feeders	original path	critical path	core path	minimal path
10kV Feeder X	1	195	15	9	-
Power Supply Unit CB002	3	1027	47	22	22
Power Supply Unit CB003	7	3552	115	36	28
North Zone	57	22481	894	313	220
Whole county	146	65236	2195	733	528

vertices and edges in each abstraction level, four topology planes which are logically independent can be constructed. Afterwards, as shown in Figure 12, the mapping relationship between adjacent topology planes is established. The physical vertices, such as switches, bifurcation points, etc. in the higher-level topology plane and the corresponding physical vertices in the lower-level topology plane are connected by one-to-one mapping edges. The virtual vertices, such as the virtual load, virtual tie switch, etc. in the higher-level topology plane and the corresponding physical vertices in the lower-level topology plane are connected by one-to-many mapping edges. In this way, fast switching between different topology planes can be realized.

V. CASE STUDY

The proposed algorithm is verified using the case of a practical MVDN in a limited scale county-level area in Jiangsu province, China. The distribution network contains 146 10kV feeders and 5800 distribution transformers, and is divided into four zones: East Zone, West Zone, North Zone and South Zone. Each of them consists of several power supply units. The relational database software and graph database software used in this test are MySQL 5.2.4 and Neo4j 3.5.17. The test is conducted on a PC workstation with a 4-core 3.3GHz CPU and 16GB RAM.

A. SCENARIO: VISUALIZATION

In the visualization interface of the test distribution network, the density of vertex (devices) and edges (power lines) directly affect the loading latency and the user's experience on the displayed topology.

This section carries out the analysis of the vertex quantity and the query time of the distribution network in different topology scales. Single feeder (10kV Feeder X), small power supply unit CB002 (3 feeders), medium power supply unit CB003 (7 feeders), North Zone (57 feeders), and the distribution network of the whole county (146 feeders) are analyzed respectively. Their numbers of vertices in different topology scales are shown in Table 3.

It can be seen that the number of vertices becomes smaller and smaller from original path to the critical path. The compression ratio of the abstraction is more than 20:1. The number of vertices in the core path/minimal path is further reduced on the basis of the critical path. Figure 13 shows the

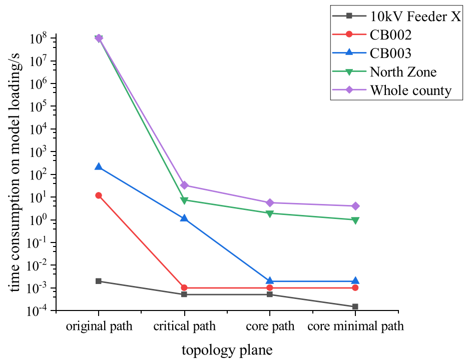

FIGURE 14. Loading time of the networks in different topology planes.

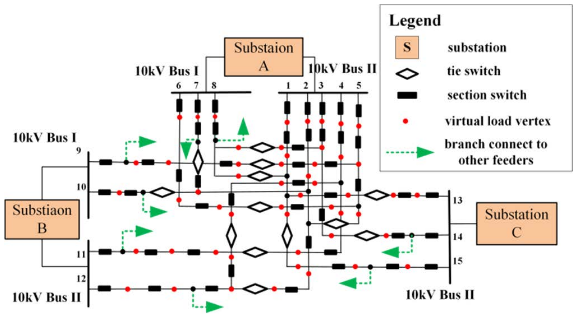

FIGURE 15. Contact diagram of the feeders.

visualized original path, critical path, core path and minimal path of the power supply unit CB003.

The loading performances of the networks in different topology planes are tested using the visual debugging tool in Neo4j. The result is shown in Figure 14. Note that a logarithmic transformation is applied on the Y-axis.

It can be seen that the loading time may exceeds 10 seconds if the original path is used. Such a long time will influence the user's experience in application. If the scale of the network becomes larger, Timeout Error may occur. The step-by-step abstraction, i.e., original path → critical path → core path → minimal path, can tackle this problem by reducing the loading time to second-level.

B. SCENARIO: LOAD TRANSFER PATH SEARCHING

The load transfer is the most frequent operation in the MVDN. When the maintenance is carried out or the failure occurs, the load transfer between feeders can be performed through a series of actions of the section switches and tie switches.

The case used in this section is shown in Figure 15. There are two main transformers in Substation A supplying 12 feeders on its 10kV Bus I and 10kV Bus II. When the maintenance is going to be carried out on one of the main transformers, the other transformer will undertake all the loads. Therefore, a load transfer for the five feeders (feeder *1*-feeder *5*) connected to 10kV Bus II is required. In order to design the load transfer strategy, all the possible load transfer paths need to be found out and analyzed to form the optimal scheme.

The task is accomplished by using the deep first search (DFS) algorithm on the original path and the critical

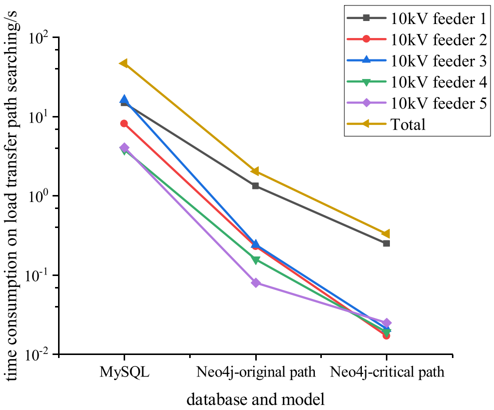

FIGURE 16. **Comparison on the searching performance.**

TABLE 4. **The number of possible transfer paths between each feeder and the directly connected feeder.**

Feeder	6	7	8	9	10	11	12	13	14	15	Total
1	-	-	2	-	-	-	-	2	-	2	6
2	-	-	-	-	2	-	3	-	-	-	5
3	-	-	3	-	-	-	-	-	3		6
4	-	-	-	-	-	3	4	-	-	-	7
5	4	-	-	4	-	-	-	-	-	-	8

path in Neo4j, and on the model stored in the relational database MySQL for comparison. The time consumptions of the three models are shown in Figure 16. It can be seen that after the path search on the critical path scale is completed, the original path needed in the transfer can be obtained through the reverse mapping introduced in section IV.

Table 4 shows the number of possible transfer paths among feeders and the contact lines. As mentioned in section II, the transfer path search belongs to the meso topology analysis scenario, and the structure within the feeder segment can be ignored. The test results also verify that the critical path based reverse mapping can obtain paths that are identical to the ones obtained from the original path search.

It can be seen that, in the scenario of searching the load transfer path, the performance is improved dozens of times by using the original path rather than the traditional relational database (MySQL). Moreover, the performance is further improved by using the critical path rather than the original path. It reveals that the graph-model-based topology abstraction and dimensionality reduction can significantly improve the path search performance in meso or macro scenarios.

VI. CONCLUSION

In this paper, the simplification method of distribution network model based on homeomorphic variation is proposed, which can significantly reduce the topology complexity in medium and macro business scenarios, and realize the exponential improvement of topology loading and query efficiency. At the same time, the application of graph database realizes the fast conversion and traceability of different topologies, which lays a data foundation for the efficient calculation and analysis of distribution network. Finally, the effectiveness of the algorithm in topology rendering and routing search is verified by an example of county level regional distribution network.

However, this paper does not consider the dynamic characteristics of the distribution network topology, so there are still some limitations, how to effectively describe, store and efficiently analyze the historical multitemporal topology of the distribution network based on the graph model remains to be further studied. In addition, the research of this paper is limited to medium voltage distribution network, and the topology modeling and analysis of low voltage distribution network is also worth further study.

APPENDIX

A. DEFINITION OF HOMEOMORPHISM TRANSFORMATION

Homeomorphism refers to the preservation of all topological properties of a given space by mapping between two topological spaces through continuous extension and bending. The basic definition of homeomorphism transformation is as follows:

Definition A-1 Expansion and contraction of a graph with a 2-degree node:

Inserting a 2-degree node in an edge of graph G will divide it into two edges, which is called the expansion of a graph with a 2-degree node. Deleting a 2-degree node will combine the two edges connected to the node into one edge, which is called the contraction of a graph with a 2-degree node.

Definition A-2 Isomorphism:

If two graphs are completely structural equivalent, they are isomorphic.

Definition A-3 Homeomorphism:

If the graphs $G1$ and $G2$ are isomorphic originally or after expansions and contractions with 2-degree nodes, they are homeomorphic.

B. DEFINITION OF A POWER SUPPLY UNIT

A partial distribution network consisting of a set of feeders with direct or indirect contact relationships is called a power supply unit. Therefore, the power supply unit is an area which is independent in the dispatch, operation and maintenance. It is suitable for a topological level in meso scenario analysis.

C. SPACE COMPLEXITY PROOF OF CRITICAL PATH

If the original path $G_{\text{Prim}} = (V_{\text{p}}, E_{\text{p}})$ has N vertices, including k tie switches, x section switches and y bifurcation vertices. Assume that the virtual load node set is VVSet and the virtual load node set is VLSet.

Let GKeyS $= (V_{\text{ks}}, E_{\text{ks}})$ be a subtree whose critical path contains only egress circuit breaker, contact switch, segmental switch, bifurcation point and connection relation of each node. The total number of nodes in the subtree is:

$$|V_{ks}| = k + x + y + 1 \tag{C-1}$$

By the characteristics of the tree, the tree of the total number of edges $|E_{ks}|$ equals to the total number of nodes minus one tree, namely:

$$|E_{ks}| = |V_{ks}| - 1 = k + x + y \quad \text{(C-2)}$$

According to definition 5, each edge of GKeyS is a feeder segment. Each feeder segment contains at most one virtual load node. Therefore, the number of virtual load nodes meets the following requirements:

$$|\text{VVSet}| <= |E_{ks}| = k + x + y \quad \text{(C-3)}$$

The total number of GKey nodes is as follows:

$$|V_{\text{k}}| = |V_{\text{ks}}| + |\text{VVSet}| \quad \text{(C-4)}$$

According to (C-1), (C-3) and (C-4),

$$|V_{\text{k}}| < = (k + x + y + 1) + (k + x + y) < \\ = 2(k + x + y) + 1 \quad \text{(C-5)}$$

Therefore, the space complexity required to store the GKey is O($k + x + y$).

D. SPACE COMPLEXITY PROOF OF CORE PATH

Assuming that the feeder critical path GKey $=$ (V_{k}, E_{k}) contains k contact switches, x segmental switches and y bifurcation points, the total number of nodes in the corresponding core path tree is:

$$|V_{\text{c}}| = k + y + 1 \quad \text{(D-1)}$$

According to the characteristics of the tree, the total number of edges of the core path is equal to the total number of nodes minus 1, namely:

$$|E_{\text{c}}| = |V_{\text{c}}| - 1 = k + y \quad \text{(D-2)}$$

In addition, since the degree of all bifurcation nodes is greater than or equal to 3, the total degree d_cnt of all nodes in the core path satisfies:

$$\text{d\_cnt} >= k^{*}1 + y^{*}3 + 1 \quad \text{(D-3)}$$

Since the tree has the property that the total degree of nodes is twice the number of edges, the number of edges of the core path satisfies:

$$|E_{\text{c}}| = \text{d\_cnt}/2 >= (k + 3y + 1)/2 \quad \text{(D-4)}$$

To sum up, it can be concluded from (D-2) and (D-4):

$$k + y > = (k + 3p + 1)/2 \quad \text{(D-5)}$$
$$k > = y + 1 \quad \text{(D-6)}$$

According to formula (D-1) and (D-6), the total number of nodes in the feeder core path $|V\text{c}| <= 2k$, so the storage complexity required by the core path is O(k).

REFERENCES

[1] J. Xiao, T. Zhang, C.-S. Wang, P. Zhang, X.-D. Wang, and S.-Q. Yuan, "TSC-based planning: A new method to exploit asset efficiency of distribution systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2013, pp. 1–5, doi: 10.1109/PESMG.2013.6672403.

[2] N. Zhou, L. Luo, G. Sheng, and X. Jiang, "Power distribution network dynamic topology awareness and localization based on subspace perturbation model," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1479–1488, Mar. 2020, doi: 10.1109/TPWRS.2019.2943695.

[3] M.-G. Choi, S.-J. Ahn, J.-H. Choi, S.-M. Cho, and S.-Y. Yun, "Adaptive protection method of distribution networks using the sensitivity analysis for changed network topologies based on base network topology," *IEEE Access*, vol. 8, pp. 148169–148180, 2020, doi: 10.1109/ACCESS.2020.3015517.

[4] S. J. Pappu, N. Bhatt, R. Pasumarthy, and A. Rajeswaran, "Identifying topology of low voltage distribution networks based on smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5113–5122, Sep. 2018, doi: 10.1109/TSG.2017.2680542.

[5] J. Li, C. Wei, Y. Wang, P. Zhang, H. Zhang, and G. Li, "Study on visualization of distribution topology based on intelligent algorithms," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2013, pp. 1–5, doi: 10.1109/NAPS.2013.6666942.

[6] C. Ten, E. Wuergler, H. Diehl, and H. B. Gooi, "Extraction of geospatial topology and graphics for distribution automation framework," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1776–1782, Nov. 2008, doi: 10.1109/TPWRS.2008.2004835.

[7] P. Shi, Y. Li, Y. Du, Q. Cai, and H. Chen, "Distribution network planning considering uncertainty of incremental distribution network access," in *Proc. IEEE 4th Conf. Energy Internet Energy Syst. Integr. (EI2)*, Oct. 2020, pp. 2244–2247, doi: 10.1109/EI250167.2020.9347026.

[8] Z. Li, W. Wu, X. Tai, and B. Zhang, "A reliability-constrained expansion planning model for mesh distribution networks," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 948–960, Mar. 2021, doi: 10.1109/TPWRS.2020.3015061.

[9] C. He and D. Li, "Multi-objective operation mode optimization of medium voltage distribution networks based on improved binary particle swarm optimization," in *Proc. 4th Int. Conf. Intell. Green Building Smart Grid (IGBSG)*, Sep. 2019, pp. 639–644, doi: 10.1109/IGBSG.2019.8886212.

[10] D. Yan, H. Cao, T. Wang, R. Chen, and S. Xue, "Graph-based knowledge acquisition with convolutional networks for distribution network patrol robots," *IEEE Trans. Artif. Intell.*, vol. 2, no. 5, pp. 384–393, Oct. 2021, doi: 10.1109/TAI.2021.3087116.

[11] J. P. Britton and A. N. deVos, "CIM-based standards and CIM evolution," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 758–764, May 2005, doi: 10.1109/TPWRS.2005.846202.

[12] D. Wei, M. Zhaoyong, C. Yaomin, and Z. Boxi, "Automatic generation for single-line diagram of distribution network," in *Proc. IEEE 11th Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2016, pp. 348–353, doi: 10.1109/ICIEA.2016.7603607.

[13] U. Corbellini, "Power systems single line diagrams standardization proposal for a international horizontal standard," in *Proc. AEIT Annu. Conf. From Res. Ind., Need More Effective Technol. Transf. (AEIT)*, Sep. 2014, pp. 1–6, doi: 10.1109/AEIT.2014.7002032.

[14] L. Chen, L. Chen, and W. Ouyang, "Ring structure based layout algorithm for distribution network contact-graph," (in Chinese), *Automat. Electr. Power Syst.*, vol. 40, no. 24, pp. 151–156, 2016.

[15] X. Sun, D. Lv, H. Wang, M. Hou, and Z. Zhang, "Topology identification method based on multi-agent for outage areas in multi-contact distribution network," in *Proc. IEEE Int. Conf. Power Renew. Energy (ICPRE)*, Oct. 2016, pp. 456–459, doi: 10.1109/ICPRE.2016.7871252.

[16] *IEC Energy Management System Application Program Interface (EMS-API)—Part 301: Common Information Model (CIM) Base*, IEC Standard 61970-301, 2nd ed., Apr. 2009.

[17] J. Liu, H. Cheng, and P. Bi, "A simplified model for distribution system," (in Chinese), *Proc. CSEE*, vol. 21, no. 12, pp. 78–83, 2001.

[18] Z. Wei, F. Zhou, C. Xiao, H. Feng, and H. Dai, "A new algorithm of reliability evaluation for complex MVDNs based on implified network model," (in Chinese), *Power Syst. Technol.*, vol. 30, no. 15, pp. 72–75, 2006.

[19] H. Q. Liao, D. Liu, and Y. H. Huang, "Analysis on transfer capability of distribution network based on CIM topological contraction," (in Chinese), *Power Syst. Technol.*, vol. 36, no. 6, pp. 51–55, 2012.

[20] M. Sun, S. Dong, Z. Cao, H. Wu, and S. Xia, "A path-based method for feeder partition available capability evaluation under N-1 security criterion," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2017, pp. 1–5, doi: 10.1109/PESGM.2017.8273816.

[21] S. Poulik, S. Das, and G. Ghorai, "Randic index of bipolar fuzzy graphs and its application in network systems," *J. Appl. Math. Comput.*, vol. 67, no. 1, pp. 1–25, Aug. 2021, doi: 10.1007/S12190-021-01619-5.

[22] P. C. Wong, K. Schneider, P. Mackey, H. Foote, G. Chin, Jr., R. Guttromson, and J. Thomas, "A novel visualization technique for electric power grid analytics," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 3, pp. 410–423, May/Jun. 2009, doi: 10.1109/TVCG.2008.197.

[23] P. Shi, G. Fan, S. Li, and D. Kou, "Big data storage technology for smart distribution grid based on neo4j graph database," in *Proc. IEEE 4th Int. Conf. Electron. Technol. (ICET)*, May 2021, pp. 441–445, doi: 10.1109/ICET51757.2021.9451128.

[24] A. Bhattacharyya and D. Chakravarty, "(Graph database: A survey)," in *Proc. Int. Conf. Comput., Electr. Commun. Eng. (ICCECE)*, Jan. 2020, pp. 1–8, doi: 10.1109/ICCECE48148.2020.9223105.

[25] Z. Zhou, C. Yuan, Z. Yao, J. Dai, G. Liu, R. Dai, Z. Wang, and G. M. Huang, "CIM/E oriented graph database model architecture and parallel network topology processing," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2018, pp. 1–5, doi: 10.1109/PESGM.2018.8586367.

JINMING CHEN (Member, IEEE) was born in China, in 1985. He received the B.S. and M.S. degrees in computer science and technology from Nanjing University, Nanjing, China, in 2006 and 2009, respectively. He is currently pursuing the Ph.D. degree with Southeast University. He is also with the Electric Power Research Institute, State Grid Jiangsu Electric Power Company Ltd. His current research interests include big data and AI applications in smart distribution.

WEI JIANG (Senior Member, IEEE) was born in China, in 1982. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2004, 2008, and 2012, respectively. He is currently working as a Professor of electrical engineering with the School of Electrical Engineering, Southeast University. His research interests include application of power electronics in distributed generation systems, energy storage systems, and power quality control.

ZHIQI XU was born in China, in 1997. He received the B.S. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2019. He is currently pursuing the Ph.D. degree in electrical engineering with Southeast University, Nanjing, China. His research interest includes situational awareness in distribution networks.

YE CHEN (Member, IEEE) was born in China, in 1994. He received the M.S. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2018. He is currently with the Electric Power Research Institute, State Grid Jiangsu Electric Power Company Ltd. His current research interest includes application of statistical methods in distribution power grid.

HAO JIAO (Member, IEEE) was born in China, in 1991. He received the M.S. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2017. He is currently with the Electric Power Research Institute, State Grid Jiangsu Electric Power Company Ltd. His current research interests include AI applications in smart distribution, evaluation, and optimization of distributed photovoltaic.

MINGHUA WANG (Graduate Student Member, IEEE) was born in China, in 1998. He received the B.S. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2020. He is currently pursuing the M.S. degree in electrical engineering with Southeast University, Nanjing, China. His research interest includes reliability analysis of distribution networks.

YUBO YUAN (Member, IEEE) was born in China, in 1975. He received the Ph.D. degree from the School of Electrical Engineering, Southeast University, Nanjing, China, in 2004. He is currently with the Electric Power Research Institute, State Grid Jiangsu Electric Power Company Ltd. His current research interests include power system relay protection and dc power transmission and distribution control protection.

WU CHEN (Senior Member, IEEE) was born in China, in 1981. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2003, 2006, and 2009, respectively. He is currently working as a Professor of electrical engineering with the School of Electrical Engineering, Southeast University. His research interests include soft-switching converters, power delivery, and power electronic system integration.