

Fast Single-phase Fault Location Method Based on Community Graph Depth-first Traversal for Distribution Network

Jian Dang, *Member, IEEE, Member, CSEE*, Yunjiang Yan, Rong Jia, *Member, IEEE*, Xiaowei Wang, *Member, IEEE*, and Hui Wei

Abstract—With the increasing complexity of distribution network structures originating from the high penetration of renewable energy and responsive loads, fast and accurate fault location technology for distribution networks is a prerequisite for rapid isolation of faults and restoration of the power supply. In this paper, a fault location method based on community graph depth-first traversal is proposed for fast location of single-phase ground faults in distribution networks. First, this paper defines the fault graph weight of the vertices in the distribution network graph model, which can be used to reflect the topology of the vertices and fault points as well as the fluctuation of the vertices' currents. Then, the vertices on the graph model are clustered by using an improved parallel louvain method (IPLM). Finally, the community formed by IPLM is used as the smallest unit for depth-first traversal to achieve fast and accurate location of the fault section. The paper develops a distribution network graph model of IEEE 33-bus system on the graph database for testing. And three other methods are selected for comparison with IPLMDF. The test results show that IPLMDF can achieve fast and accurate fault location when half of the nodes in the distribution network are equipped with D-PMUs. When some of the D-PMUs lose time synchronization, it is still possible to locate the fault section, and at the same time, the locating results can be avoided by falling into local optimal solutions.

Index Terms—Depth-first traversal, fault location, graph computation, single-phase fault.

I. INTRODUCTION

NOWADAYS, with the growing load and the penetration of large amounts of renewable energy, the challenge of power supply reliability in the distribution network is becoming significantly important. It is crucial for electric utilities to minimize outage time [1]. Accuracy and rapidity have become two major metrics in fault location research. Therefore, a fast and accurate fault location method is a prerequisite for repairing faults, rapidly restoring power supply,

and improving the reliability of power supply [2]. Single-phase ground fault accounts for more than 70% of all faults in the distribution network and is the most common fault in the distribution network [3]. Impedance-based methods [4], [5], traveling wave-based methods [6], [7], and S-injection methods [8] are commonly used in fault location research.

Impedance-based methods are widely used in fault location research because they are easy to implement. Reference [4] uses an impedance-based method to screen out all possible fault locations, and determines the actual fault point by simulating the same fault. Reference [5] uses an algorithm for iterative state estimation to overcome measurement data inaccuracies. The accuracy of the impedance-based methods is susceptible to a variety of factors such as fault resistance, distributed power access, etc. Compared with the impedance-based methods, the traveling wave-based methods are not affected by the fault resistance and system operation mode, but are more sensitive to line branches. Reference [6] proposes a method based on traveling wave arrival computational fault time difference (CFTD), which is not affected by traveling wave speed inaccuracies. In reference [7], the PCA-SVM model is constructed to extract the curve cluster characteristics of the fault current traveling wave. In addition to the methods based on impedance and traveling wave, an active method is proposed in reference [8]. For a resonant grounded system, whose neutral point grounds are via an electromagnetic hybrid Petersen coil (EHPC), the identification basis of fault feeder and fault section can be provided by adjusting the EHPC complex impedance [8]. This method requires supplementary devices and extra operations, which increase potential risks and costs [9]. Although the above three types of methods have been improved, they are all constrained to varying degrees by the imperfection of the distribution network monitoring system.

By virtue of the synchronized, low-latency and high-resolution measurement data, the application of the distribution-level phasor measurement unit (D-PMU) in distribution networks continues to increase, and many scholars have carried out research on distribution network fault location based on D-PMU [9]–[14]. Reference [9] identifies the fault section by calculating the transient energy of the zero-sequence current in selected frequency bands for each section. Reference [10] proposes a method to achieve fault location without line parameters, which solves the effect of inaccurate line parameters

Manuscript received August 31, 2020; revised December 15, 2020; accepted March 18, 2021. Date of online publication July 9, 2021; date of current version July 23, 2022. This work was supported by the National Natural Science Foundation of China (Grant Nos. 52009106, 51779206), the National Key R&D Program of China (No. 2018YFB1500800), and the Natural Science Fund Youth Project of Shaanxi Province (2019J-130).

J. Dang, Y. Yan, R. Jia (corresponding author, email: jiarong@xaut.edu.cn), X. Wang and H. Wei are with Xi'an Key Laboratory of Intelligent Energy, Xi'an University of Technology, Shaanxi 710048, China.

DOI: 10.17775/CSEEJPES.2020.04650

in distribution networks. However, this method cannot be applied to single-phase ground faults in neutral ungrounded systems. There are also neural network-based fault location methods [11], [12]. Reference [11] proposes an adaptive convolutional neural network-based method, but this type of method usually faces the problem of insufficient sample data.

References [9]–[11] focus too much on the study of D-PMU measurement information and do not make full use of the topology information of the distribution network. Reference [12] proposes a fault location method based on a graph convolutional neural network, which incorporates the network topology and is more adaptable to changes in the network topology. Reference [13] proposes a minimum spanning tree-based method, which identifies the fault section based on the network topology and the phase of zero sequence current. However, the methods in references [12], [13] need to configure D-PMUs on each node in the network to ensure the accuracy of the localization results, which reduces the feasibility of these methods. Reference [14] proposes a fault location method with only installed D-PMU at the head and end of the feeder, but this method cannot be used to locate a single-phase ground fault.

Power network topology is a type of natural graph structure, yet this feature is underutilized in existing studies. Graph databases and graph computation have good application prospects in mining the value of network topology information, and currently there is no fault location method based on graph computation. To address the above problems, this paper proposes a fast single-phase fault location method based on community graph depth-first traversal. The community graph formed by the improved parallel louvain method ensures that the faulted section can be identified in 2–3 queries, regardless of where in the network the fault occurs. The proposed method can rapidly and accurately locate the fault section in the case of high resistance grounding. Moreover, the proposed method only requires D-PMUs to be configured on half of the nodes in the distribution network model. And it can still rapidly and accurately locate the fault section when some of the D-PMUs are disturbed and lose their time synchronization.

The rest of this paper is organized as follows. The graph database and graph computation are introduced and the distribution network graph modeling is described in Section II. The definition of fault graph weight is given and a depth-first traversal based on fault graph weight is proposed in Section III. The parallel louvain method is introduced and the community graph depth-first traversal is proposed in Section IV. The accuracy, robustness and calculation speed of the proposed method are verified on the IEEE 33-bus system graph model in Section V. Conclusions are provided in Section VI.

II. DISTRIBUTION NETWORK GRAPH MODELING

A. Graph Database and Graph Computation

A graph databases has emerged in recent years. It is one type of non-relational database that can be used to process massive data in parallel. A graph database is formed based on the idea of graph theory. Vertices and edges are the two elements making up the graph database. Data can be stored in the

vertices and the relationship between data can be represented by edges. The edges can send computing messages between the connected vertices [15].

The grid is a natural graph structure. The devices in the grid can be abstracted as vertices. The information of devices can be stored in the attributions of the vertex. Transmission lines can be described as edges, and even the various relationships between devices can be abstracted as edges. When the data of grid is stored in the form of graph data, the topological structure of the power system can be preserved and easily expanded [16].

Graph computation is a type of node-based parallel computing, which means each vertex has its own calculation logic and all calculations on each vertex are independent to other vertices [17]. Pregel, developed by Google, is the earliest graph computation system. Pregel proposed a vertex-centric programming model and its calculation mode is a bulk synchronous parallel computing model (BSP).

Graph computation can be applied to many aspects within the power system. Reference [18] realizes power flow calculation of the power system by graph computation. Reference [19] uses graph computation to do the real-time power system network topology analysis. They use the idea of graph partitioning, and propose the parallel topology analysis algorithm based on the graph computing framework. Reference [20] presents a graph-computation based EMS (Energy Management System, EMS) which takes advantage of both nodal and hierarchical parallelism to achieve high-speed graph-based state estimation, power flow, and contingency analysis applications. In this paper, we take graph computation to calculate the voltage and current of bus without D-PMU installed.

B. Distribution Network with D-PMU Graph Modeling

In the graph database, Schema specifies the types and attributes of the vertices and edges of the graph data, as well as the relationship between various vertices. In order to fully describe the state of the distribution network in the event of a fault, this paper establishes a Schema as shown in Fig. 1 on the graph database TigerGraph, which contains different types of vertices and edges. Each vertex contains six attributes: ID, upstream vertex, downstream vertex, fault graph weight, D-PMU state, and community ID. Each edge contains two attributes: NAME and fault graph weight. Among the above attributes, the role of ID is to ensure the uniqueness of each vertex. The role of upstream and downstream nodes is to determine the location of the vertices on the network topology. The fault graph weight is used to describe the strength of the fault on the nodes and lines. The definition of the fault graph weight will be introduced in the next section. The D-PMU state is a two-dimensional array of two elements with 0 or 1 to describe whether a node is configured with D-PMU and whether D-PMU is synchronized. The NAME of each edge is named as “Upstream Vertex ID - Downstream Vertex ID.”

Based on the of set attributes, the relevant information in the distribution model is converted into data text for vertices and edges, respectively. By loading them into the Schema, the corresponding distribution network graph model can be obtained. The detailed flow is shown in Fig. 1.

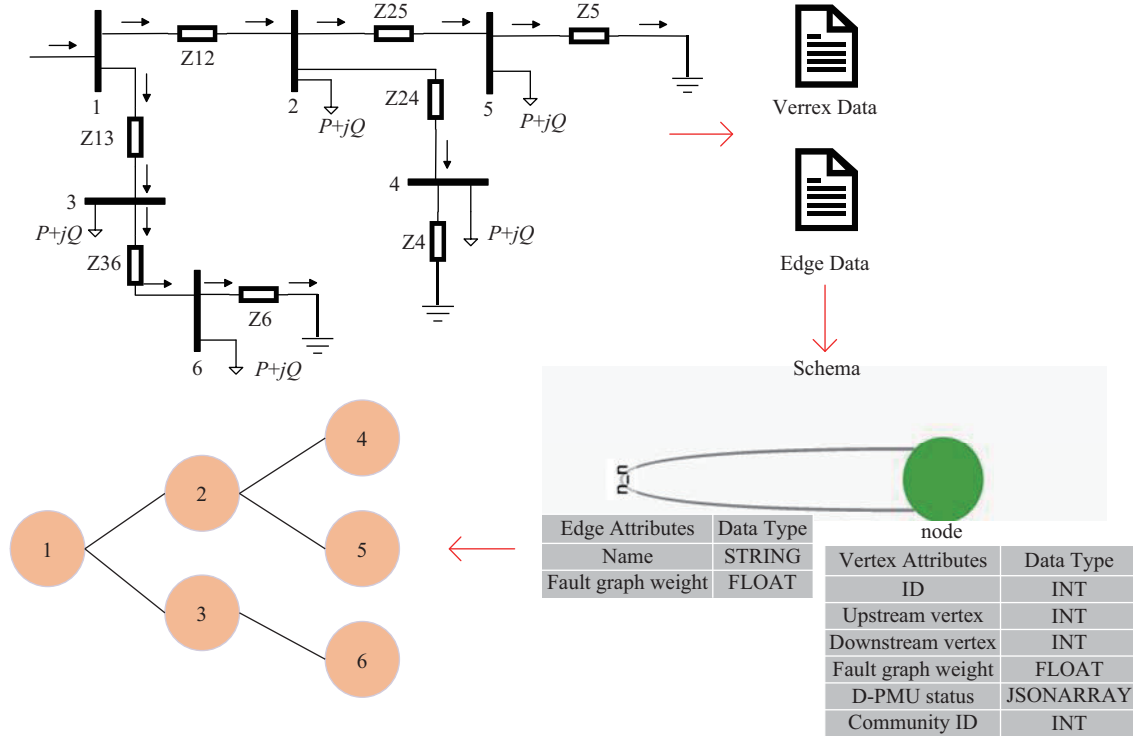


Fig. 1. Distribution network diagram modeling process.

In this paper, D-PMUs are configured as follows: no D-PMU is configured at the terminal node, and only one of any two connected nodes in the distribution model is configured with a D-PMU. A configuration principle ensures that nodes h and j on both sides of each node i that is not configured with D-PMU are configured with D-PMU. If the D-PMU configuration state attribute of vertex i is represented by k_i , it can be expressed as:

$$k_i = \begin{cases} 1 & \text{Node with D-PMU Configured} \\ 0 & \text{Node without D-PMU Configured} \end{cases} \quad (1)$$

For each edge e_{ij} of the distribution network graph model, the following equation is satisfied.

$$k_i + k_j = 1 \quad (2)$$

Equation (2) ensures that each line in the distribution network contains a D-PMU, and the information of unconfigured D-PMU nodes can be calculated from two adjacent configured D-PMU nodes to satisfy the global observable condition.

Based on the above configuration principles, the distribution model can be simplified to the set of minimum monitoring units shown in Fig. 2. Node i is the unconfigured D-PMU node, and node h and node j are the nodes upstream and downstream of node i that are configured with D-PMU, respectively.

Figure 2(a) corresponds to the case where node i is not a node at the end of the line, in which case the computation of node i information is divided into two parts: forward solving voltage and backward solving current, as shown in (3) and (4).

$$\dot{U}_i = \dot{U}_j + \dot{I}_j \cdot Z_{ij} \quad (3)$$

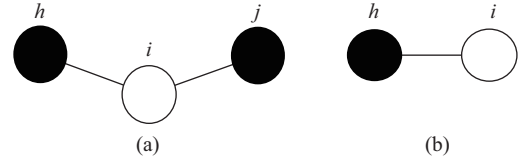


Fig. 2. Minimum monitoring unit.

$$\dot{I}_i = \frac{\dot{U}_h - \dot{U}_i}{Z_{hi}} \quad (4)$$

Figure 2(b) corresponds to the case where node i is a branch end node, and node i has no downstream node j . The voltage information of node i is calculated from (3) to (5), and the current information is calculated in the same way.

$$\dot{U}_i = \dot{U}_h - \left(\dot{I}_h - \left(\frac{P_h + jQ_h}{U_h} \right)^* \right) Z_{hi} \quad (5)$$

In this paper, a modified IEEE 33-bus system graph model is built on the graph database as shown in Fig. 8, which is more intuitive than using the adjacency matrix to reflect the connection between nodes, and the topology of the graph model is consistent with the actual topology to ensure that the graph calculation can be performed directly on the actual topology of the grid. The green nodes are the installed D-PMU nodes and the black nodes are the unconfigured D-PMU nodes.

III. DEPTH-FIRST TRAVERSAL BASED ON FAULT GRAPH WEIGHT

A. Definition of Fault Graph Weight

Single-phase ground fault is the most common fault type in distribution networks. When a single-phase ground fault oc-

curs in a small-current grounding system (neutral ungrounded system or resonant grounding system), the symmetry of the three-phase line to ground is broken and the fault point forms a circuit only through the capacitance to ground. Taking the neutral point ungrounded system as an example, when a single-phase grounding fault occurs, the simplified zero-sequence network is as shown in Fig. 3.

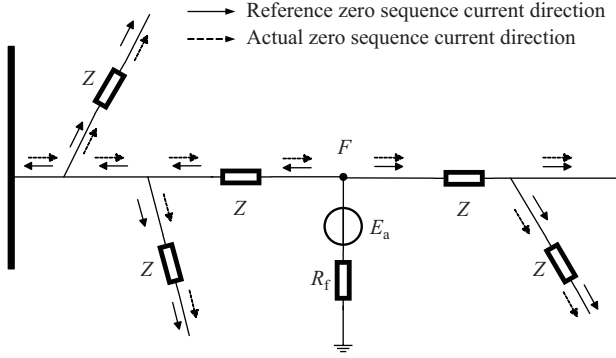


Fig. 3. Simplify zero sequence network diagram.

With the zero-sequence network shown in Fig. 3, it can be concluded that the zero-sequence current of the node located before the fault point on the fault branch lags the zero-sequence voltage by 90° , and the zero-sequence current of the other nodes leads the zero-sequence voltage by 90° [15]. Therefore, by judging the phase relationship between the zero-sequence current and zero-sequence voltage of each node, the position relationship between each node and the fault point can be derived.

Let $\Delta\theta$ be the phase difference between the zero-sequence current and zero-sequence voltage. Consider the interference of background noise, set $\pm 10^\circ$ as the reasonable deviation that may occur between $\Delta\theta$ and the theoretical value when rounding $\Delta\theta$. (7) is the rounding function of $\Delta\theta$.

$$\Delta\theta = \theta_{I_0} - \theta_{U_0} \quad (6)$$

$$\sigma_i(\Delta\theta) = \begin{cases} 1 & \Delta\theta \in (-80^\circ, -100^\circ) \\ 0 & \Delta\theta \in (-10^\circ, 10^\circ) \\ -1 & \Delta\theta \in (80^\circ, 100^\circ) \end{cases} \quad (7)$$

Set the line between nodes 11 and 12 in the IEEE 33-bus system to have an a-phase short circuit to ground at 0.2 s with a fault resistance of 10Ω and excise the fault at 0.4 s. Fig. 4(a) and (b) show the variation of the three-phase currents recorded by the D-PMU at node 1 and node 17 at the time of the fault. After the occurrence of the fault, the three-phase current at node 1 rapidly increases, with the fault phase current significantly increasing. The non-fault phase current at node 17 rapidly increases, while its fault phase current rapidly drops and is close to zero. At the first node, the phase with the largest sudden change in current value is the fault phase, and the fault phase is recorded as k .

Set a time window $T = 0.05$ s and calculate the standard deviation of the current of the fault phase at each node within a time window.

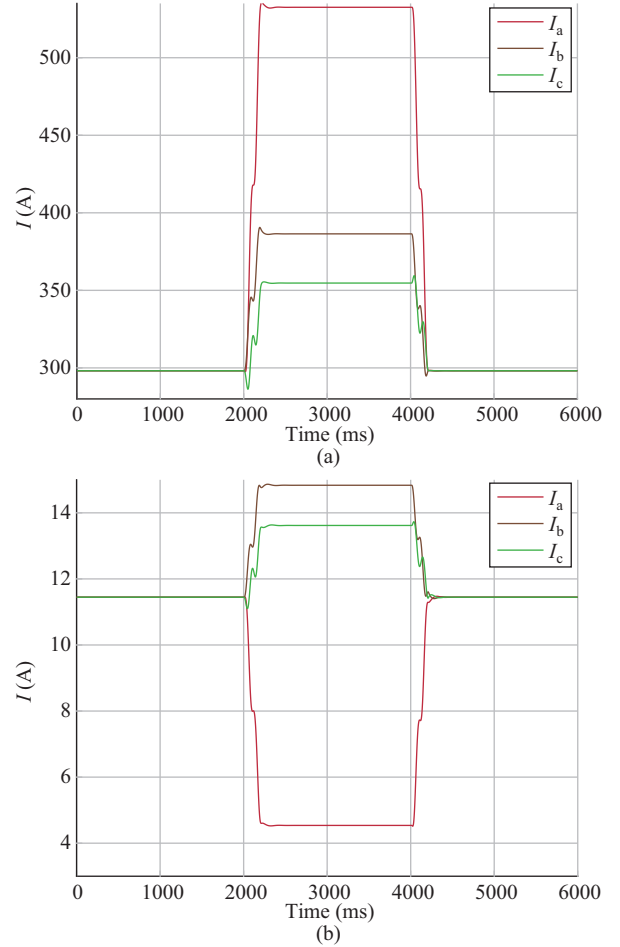


Fig. 4. Three-phase current variation.

$$S_{i,k} = \sqrt{\sum_{t=t_0}^{t_0+T} \frac{(I_{t,i,k} - \bar{I}_{i,k})^2}{Tf - 1}} \quad (8)$$

In (8), k is the fault phase in the three phases, f is the adopted frequency of the D-PMU, T is the time window of 0.05 s, t_0 is the starting moment, and $I_{t,i,k}$ is the amplitude of the fault phase current at node i at moment t .

The fault graph weight of node i contains not only the phase difference of the zero-sequence current and voltage, but also the amplitude fluctuation of the fault phase current, which is denoted as $S'_{i,k}$.

$$S'_{i,k} = S_{i,k} \cdot \sigma_i(\Delta\theta) \quad (9)$$

B. Depth-first Traversal based on Fault Graph Weight

Most of the existing studies on fault location in distribution networks, whether based on transient or steady-state quantities, rely on a large number of measurement points in the network, lacking the use of network topology information and requiring high time synchronization of measurement points in the network. The long power supply radius of the distribution network, many line branches and high background noise will not only increase the cost but also make it difficult to guarantee the time synchronization of measurement points in the network when a large number of measurement points are arranged.

Under these conditions, the accuracy of the fault location will be affected.

To address the above issues, this paper proposes a depth-first traversal based on fault map weights for fault location. The fault graph weight, as an attribute of the vertices in the distribution network graph model, reflects both the fluctuation of node currents after a fault and the relationship between nodes and fault points in the topology. Depth-first traversal on the distribution network graph model can make full use of the topology information of the network.

In this paper, starting from the first node, the branch where the node $S'_{i,k} > 0$ is located is traversed in priority, and the traversal is stopped when a node j is visited with $S'_{j,k} < 0$. The line connected by node j and its upstream node i is the fault section.

In the test, it was found that the fault graph weight of the node without D-PMU installed on the fault section was always greater than 0. Therefore, when the last visited node i is an uninstalled D-PMU node, further judgment is required using (10).

$$\delta = \frac{S'_{i,k}}{S'_{h,k}} \quad (10)$$

In (10), node h is the upstream node of node i . When $\delta < 0.9$, the fault section is between node h and node i , otherwise, the fault section is between node i and node j .

In the problem of time synchronization of measurement units, each node in the distribution network graph model has the D-PMU state attribute. This attribute is a custom two-dimensional array [D-PMU configuration, D-PMU synchronization], and the various types of this variable are shown in Fig. 5.

[0,0]	~~[0,1]~~
[1,1]	[1,0]

Fig. 5. All cases of D-PMU status.

In the above figure, [0,0] and [1,1] are two normal states, there is no [0,1] state because the node which is not configured with D-PMU will not have a D-PMU time synchronization [1,0] is the state where the node D-PMU loses time synchronization, and the judgment condition is as follows:

$$\begin{cases} k_i = 1 \\ \sigma_i(\Delta\theta) = 0 \\ \sigma_j(\Delta\theta) \neq 0 \end{cases} \quad \begin{matrix} i \neq j \\ v_i, v_j \in G(v, e) \end{matrix} \quad (11)$$

From the previous analysis, we can conclude that the node located in front of the fault point on the fault branch has a fault graph weight greater than 0, and the rest of the nodes have a fault graph weight less than 0. When the time synchronization of some measurement units fails, the fault location results will not be affected as long as the positive and negative signs of the fault graph weights of all nodes in the network can be guaranteed to be unchanged.

Suppose that when a failure occurs in the network of Fig. 8, after the D-PMU of node 9 loses its time synchronization, the

node becomes non-viewable, and node 8 and node 10 are only viewable through the D-PMUs of nodes 7 and 11. The fault graph weight of node 9 can be estimated using the fault graph weights of node 8 and node 10, which is estimated by taking the average of the fault graph weights of neighboring nodes.

$$S'_{s,k} = AVG \left(\sum S'_{t,k} \right) \quad e_{st} \in G(v, e) \quad (12)$$

Equation (12) is the formula of estimation of the fault graph weight, where s represents the node whose fault graph weight is 0 due to the loss of synchronization of D-PMU, and node t is the vertex adjacent to node s . The graph model built in this paper is an undirected graph, so node t represents the upstream and downstream nodes of node s .

Since nodes 8–11 are all nodes located in front of the fault point, the estimated fault graph weight of node 9 will be numerically different from the original case, but the positive and negative attributes remain the same and will not affect the result of fault location. If the D-PMU of node 7 and node 11 also lose their temporal synchronization, the estimated fault graph weights of node 6 to node 12 can only rely on the information of node 5 and node 13. Since both $S'_{5,k}$ and $S'_{13,k}$ are greater than 0, the accuracy of the fault location results can still be guaranteed.

When the D-PMU at node 13 also loses temporal synchronization, it is necessary to use the fault graph weight of node 15 for estimation. In this case there will be $S'_{5,k} > 0$ and $S'_{15,k} < 0$. The positivity and negativity of the fault map weights of all nodes in the interval from node 6 to node 14 are difficult to determine, and the fault location results will be affected.

Therefore, the minimum constraint for accurate fault location is that the temporal accuracy of the D-PMU at the connecting fault branch on the non-fault branch, the D-PMU at the first node on the fault branch, and the D-PMU at the two nodes closest to the fault point must not fail.

IV. COMMUNITY GRAPH DEPTH-FIRST TRAVERSAL

In distribution network fault location, real-time is an important requirement. Fig. 6 shows the fault section search process based on depth-first traversal. This method can locate the fault section after a few traversal steps when the fault point is close to the power node. If the fault point is near the end of the line and the fault branch is long, multiple traversals are required to locate the fault section, which makes it difficult to guarantee the real-time nature of the algorithm.

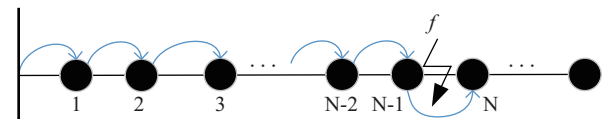


Fig. 6. Depth-first traversal query process.

To ensure fast location, this paper combines depth-first traversal with graph clustering. The louvain method is a locally greedy graph clustering algorithm that divides the graph into multiple communities. Fault section search based on a community graph ensures that fault sections can be located

by a few traversal steps even when the fault point is near the end of the line.

A. Parallel Louvain Method

The Louvain method was proposed in 2008 by Blondel *et al.* of Louvain University to process graph data [21]. This algorithm performs the exploration of community discovery on the graph model based on the gain of modularity, dividing the graph into several communities according to the edge weights, and its objective function is to maximize the modularity of the whole network. Community is a type of subgraph formed after clustering. The modularity is a measure of the strength of connections between nodes, usually denoted by the letter Q , which is calculated as follows:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (13)$$

Equation (13) is the formula of Modularity Q , where A_{ij} represents the weight of the edge between node i and node j , $m = \frac{1}{2} \sum_{i,j} A_{ij}$ is half the sum of weights of all edges, and $k_j = \sum_j A_{ij}$ is the sum of the weights of the edges connected to node j , c_j is the community to which vertex j is assigned, when $c_i = c_j$, $\delta(c_i, c_j) = 1$ otherwise 0.

It is easy to see by the above equation that if all the vertices in the network are independent communities, the modularity of the whole network is 0. If the modularity of the whole network is to be maximized, it is necessary to maximize A_{ij} within the same community in (13).

If the weight of an edge is viewed as the closeness of the connection between two vertices, then this process can be understood from the perspective of the null domain as each vertex tends to be assigned to the community where the vertex with which it is most closely connected is located. The core of Louvain method is based on this process, so that the vertices with strong ties are in the same community, and the weak ties between vertices are diluted to achieve the clustering of complex networks.

The Louvain method is divided into two phases, in each of which the modularity gain resulting from the vertex movement is calculated iteratively. The first stage of Louvain method calculates the modularity gain due to individual node movement, and the second stage calculates the modularity gain due to community merging.

$$\Delta Q = \left[\frac{\sum_{in} + 2k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (14)$$

Equation (14) is the formula for calculating the modularity gain ΔQ , where $\sum_{in}$ is the sum of the weights of all edges in C_j , $\sum_{tot}$ is the sum of the edge incident to vertices in C_j , $k_{i,in}$ is the weight of the edge connected vertex i and C_j . By simplifying the above equation, we can obtain the simplified modularity formula as follows:

$$\Delta Q = \frac{1}{2m} \left(k_{i,in} - \frac{\sum_{tot} k_i}{2m} \right) \quad (15)$$

In references [22], parallel computing is added to the louvain method, and the parallel louvain method (PLM) is proposed. PLM initially treats each node as an independent community. In each iteration, all vertices in the network are visited in parallel, the modularity gain ΔQ generated by moving each vertex to an adjacent community is calculated, and the vertex is moved to the community with the largest ΔQ . Each time a vertex is moved from its original community to another community, it must be guaranteed that the modularity gain is positive, i.e., $\Delta Q > 0$. Node-parallel computation improves the computational speed of the Louvain method in each iteration, but may produce negative modularity gains during vertex community attribution changes. This is due to the fact that all vertices calculate the modularity gain of moving themselves to an adjacent community in parallel, and by the time a vertex is moved to an adjacent community, the structure of that community may have changed and the actual modularity gain may be negative. The negative modularity gain can be corrected by multiple iterations. The community attribution of each vertex no longer changes after one iteration indicates that the vertices have been moved to the best community at this point and the first phase of PLM is complete.

The second phase of the Louvain's algorithm is computed in the same way as the first phase, but the object of computation is changed from the vertices in the network to the community. Each community is considered as a "super node," and the connection weights among the "super nodes" are calculated, and the first stage of PLM is executed with the "super node" as the smallest unit.

B. Community Graph Depth-first Traversal

In the graph model of this paper, the weight w_{ij} of each edge is the fault graph weight of vertex j downstream of this edge, i.e.,

$$w_{ij} = S'_{j,k} \quad \forall e_{ij} \in G(v, e) \quad (16)$$

In this paper, we run PLM on the distribution network diagram model to obtain the community graph. Then the depth-first traversal is performed based on the community graph.

In this paper, we first visit the first vertex of the network and record the ID of the community where the first vertex is located, noted as fid . Then, the most downstream vertex in the community with community ID fid that has a fault graph weight greater than 0 is located on the topology which is visited, and denoted as s . Visit the downstream vertex of vertex s , denoted as t . If the fault graph weight of t is greater than 0, assign the community ID where vertex t is located to fid . The above steps are repeated until the failure graph weight of vertex t is less than 0 after a certain iteration, the loop ends and vertex s and vertex t are output. At this point, if a D-PMU is configured on vertex s , the fault section is the line represented by the edge connecting vertex s and vertex t . If vertex s is a vertex without D-PMU configured, it is necessary to use (10) to further determine the exact fault section.

Based on the results of the graph clustering in PLM, the fault location process in this paper is improved from

vertex-based depth-first traversal to community-based depth-first traversal, which reduces the number of iterations of depth-first traversal. Therefore, the effect of PLM clustering will directly affect the speed of fault location in this paper. In testing, it was found that PLM graph clustering is not always ideal, and at the most, two or three vertices with fault graph weights greater than 0 will form a community. In this paper, we expect the community formed by vertices with fault graph weights greater than 0 to be as large as possible, so that the number of iterations in the fault localization process will be further reduced.

To address the above situation, this paper analyzes the reasons for the unsatisfactory clustering effect of PLM. There is $\Delta Q > 0$ constraint for each vertex community attribution change in PLM. Since the distribution network graph model is built with reference to the radial distribution network topology in this paper, this results in nearly 80% of the vertices with degree 2, and the connection relationship between the vertices is relatively simple. This means that for 80% of the vertices, the only community attribution available is to the community where the upstream and downstream vertices are located. With strict adherence to the $\Delta Q > 0$ constraint, PLM forms many small communities with limited improvement in query speed.

Based on the above analysis, this paper improves the PLM and proposes an improved parallel louvain method (IPLM). In the first phase of PLM, the graph model has been changed in a round of community attribution in terms of vertices to form a preliminary community distribution. The second phase of PLM is based on the “super node” unit and then the community attribution change, compared with the first phase, the second phase has fewer variables and the results are more easily converged. In this paper, the constraint on the change of vertex community attribution in PLM Phase II is modified and its constraint is changed from $\Delta Q > 0$ as shown in (17).

$$\Delta Q > 0 \text{ or } \sum S'_{i,k} = \sum |S'_{i,k}|, v_i \in C_i \quad (17)$$

In this paper, a constraint is added to PLM Phase II: the sum

of the fault graph weights of all vertices in the community is equal to the sum of the absolute values of the fault graph weights. The two constraints are “or” relationships, and either one of the two constraints can be satisfied to fuse two neighboring communities into one community. This improvement significantly increases the size of the community, reducing the number of communities and the number of iterations of the fault section query process.

In this paper, phase A ground fault occurs at f_3 on the distribution network model shown in Fig. 8, with a fault resistance of 100 Ω . The weights of each edge in the distribution network graph model are obtained according to (17), and the direct depth-first traversal query, PLM-based depth-first traversal query, and IPLM-based depth-first traversal query are used respectively, and the positioning results and the number of iterations are shown Table I.

TABLE I
COMPARISON OF THREE SEARCH METHODS

Search method	Location result	Iterations	Maximum community size
Direct depth-first traversal query	14,15	14	1
PLM-based depth-first traversal query	14,15	5	5
IPLM-based depth-first traversal query	14,15	2	13

V. CASE STUDIES

A. Accuracy Verification of the Proposed Method

In this paper, the IEEE 33-bus system graph model is built on the graph database, and the green nodes in Fig. 8 are D-PMU nodes installed and the black nodes are unconfigured D-PMU nodes. Phase A ground fault was set on six lines in the distribution network diagram model, as shown in Fig. 7.

In order to consider the influence of fault resistance on the accuracy of fault location, this paper sets the fault resistance with resistance values of 100 Ω , 500 Ω and 1000 Ω on all six fault lines. Since the fault occurs at different locations

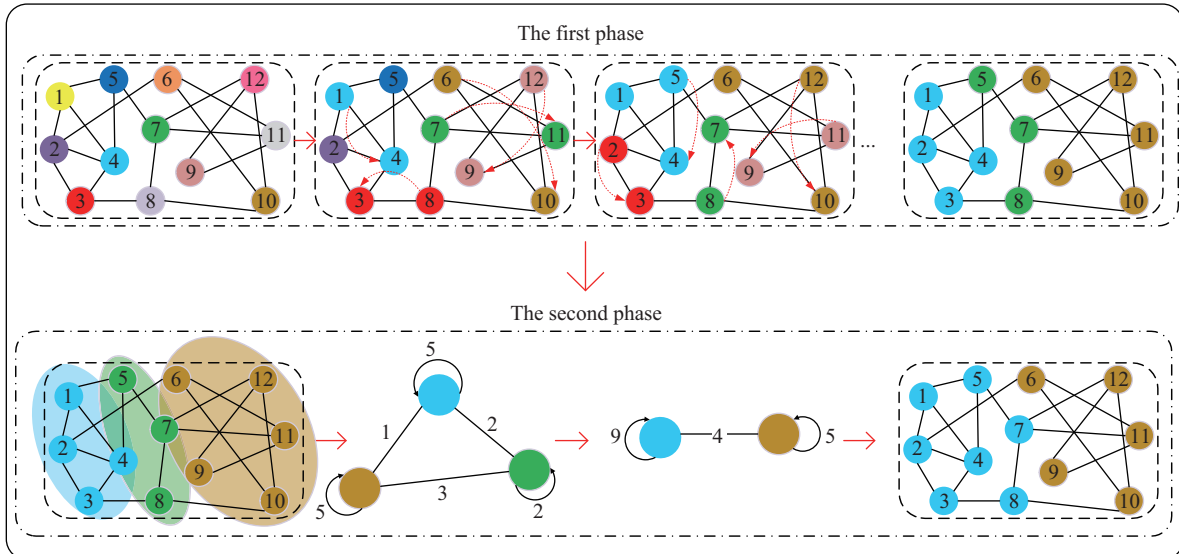


Fig. 7. PLM process.

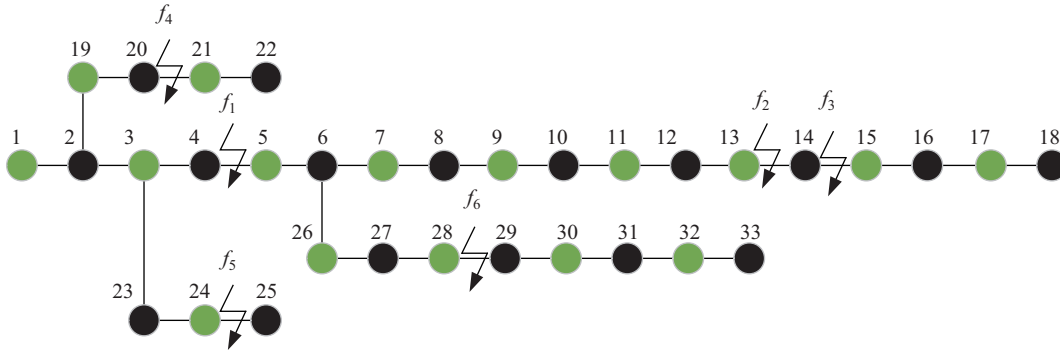


Fig. 8. The modified IEEE 33-bus system.

on the line, the fault current at the time of the fault will be different; therefore, this paper sets up single-phase grounding and earthing faults at the first 10%, 50% and last 10% of the total length of the above six lines to verify the fault location accuracy of the algorithm in the whole section of the line.

For different fault situations on the above six lines, the depth-first traversal method (DF), PLM-based depth-first traversal method (PLMDF), IPLM-based depth-first traversal method (IPLMDF) and cuckoo algorithm-based fault location (CS) mentioned in the reference [23] are used in this paper, respectively. The first three methods are implemented on the graph database TigerGraph and are based on the network topology for fault section location. In the CS, assuming that all the above fault currents can be detected by the D-PMU, (7) is modified to (18), and the result generated by (18) is used as the basis for locating the CS.

$$\sigma_i(\Delta\theta) = \begin{cases} 1 & \Delta\theta \in (-80^\circ, -100^\circ) \\ 0 & \Delta\theta \notin (-80^\circ, -100^\circ) \end{cases} \quad (18)$$

Table II shows the results of the tests. From the results in Table II, it can be seen that the three fault localizations based on the graph calculation proposed in this paper can directly and accurately locate the faults occurring at f_1 , f_2 , f_3 , f_4 and f_6 . When a fault occurs at f_5 , the output of the first three methods in this paper is 25,0. Because 25 vertices are line end vertices in the distribution network diagram model and there are no ID 0 vertices in the model, the locating result is still accurate.

Comparing CS with the first three methods, only the fault occurring at f_5 in the output result is accurate, and the rest of the output results contain both fault and normal sections, with a larger localization range. This is due to the fact that the CS relies only on the overcurrent signal with the measurement unit for fault location, and the smallest unit of output is in the section between the two measurement units. Therefore, with the same number of D-PMUs configured, the DF, PLMDF, and IPLMDF proposed in this paper have better positioning effects than CS.

B. Robustness Verification of the Proposed Method

The conclusions in the previous subsection have verified the positioning accuracy of DF, PLMDF, and IPLMDF in all D-PMU normal operating cases.

TABLE II
ACCURACY VERIFICATION OF THE PROPOSED METHOD

The actual location of fault	Fault location on the line	Ground resistance	DF	PLMDF	IPLMDF	CS
4,5	Top 10%	100 Ω	4,5	4,5	4,5	3,5
4,5		500 Ω	4,5	4,5	4,5	3,5
4,5		1000 Ω	4,5	4,5	4,5	3,5
4,5	50%	100 Ω	4,5	4,5	4,5	3,5
4,5		500 Ω	4,5	4,5	4,5	3,5
4,5		1000 Ω	4,5	4,5	4,5	3,5
4,5	Last 10%	100 Ω	4,5	4,5	4,5	3,5
4,5		500 Ω	4,5	4,5	4,5	3,5
4,5		1000 Ω	4,5	4,5	4,5	3,5
13,14	Top 10%	100 Ω	13,14	13,14	13,14	13,15
13,14		500 Ω	13,14	13,14	13,14	13,15
13,14		1000 Ω	13,14	13,14	13,14	13,15
13,14	50%	100 Ω	13,14	13,14	13,14	13,15
13,14		500 Ω	13,14	13,14	13,14	13,15
13,14		1000 Ω	13,14	13,14	13,14	13,15
13,14	Last 10%	100 Ω	13,14	13,14	13,14	13,15
13,14		500 Ω	13,14	13,14	13,14	13,15
13,14		1000 Ω	13,14	13,14	13,14	13,15
14,15	Top 10%	100 Ω	14,15	14,15	14,15	13,15
14,15		500 Ω	14,15	14,15	14,15	13,15
14,15		1000 Ω	14,15	14,15	14,15	13,15
14,15	50%	100 Ω	14,15	14,15	14,15	13,15
14,15		500 Ω	14,15	14,15	14,15	13,15
14,15		1000 Ω	14,15	14,15	14,15	13,15
14,15	Last 10%	100 Ω	14,15	14,15	14,15	13,15
14,15		500 Ω	14,15	14,15	14,15	13,15
14,15		1000 Ω	14,15	14,15	14,15	13,15
20,21	Top 10%	100 Ω	20,21	20,21	20,21	19,21
20,21		500 Ω	20,21	20,21	20,21	19,21
20,21		1000 Ω	20,21	20,21	20,21	19,21
20,21	50%	100 Ω	20,21	20,21	20,21	19,21
20,21		500 Ω	20,21	20,21	20,21	19,21
20,21		1000 Ω	20,21	20,21	20,21	19,21
20,21	Last 10%	100 Ω	20,21	20,21	20,21	19,21
20,21		500 Ω	20,21	20,21	20,21	19,21
20,21		1000 Ω	20,21	20,21	20,21	19,21
24,25	Top 10%	100 Ω	25,0	25,0	25,0	24,25
24,25		500 Ω	25,0	25,0	25,0	24,25
24,25		1000 Ω	25,0	25,0	25,0	24,25
24,25	50%	100 Ω	25,0	25,0	25,0	24,25
24,25		500 Ω	25,0	25,0	25,0	24,25
24,25		1000 Ω	25,0	25,0	25,0	24,25
24,25	Last 10%	100 Ω	25,0	25,0	25,0	24,25
24,25		500 Ω	25,0	25,0	25,0	24,25
24,25		1000 Ω	25,0	25,0	25,0	24,25
28,29	Top 10%	100 Ω	28,29	28,29	28,29	28,30
28,29		500 Ω	28,29	28,29	28,29	28,30
28,29		1000 Ω	28,29	28,29	28,29	28,30
28,29	50%	100 Ω	28,29	28,29	28,29	28,30
28,29		500 Ω	28,29	28,29	28,29	28,30
28,29		1000 Ω	28,29	28,29	28,29	28,30
28,29	Last 10%	100 Ω	28,29	28,29	28,29	28,30
28,29		500 Ω	28,29	28,29	28,29	28,30
28,29		1000 Ω	28,29	28,29	28,29	28,30

In this paper, a phase A ground fault with the fault resistance of $1000\ \Omega$ is set at $f_1 - f_6$ 50% of the line, and some D-PMUs in the network lose time synchronization. We test the fault location effectiveness of the above four methods when some D-PMUs lose time synchronization. Table III shows the results of the tests.

Table III compares the accuracy of the four fault-setting unit methods in the event that some of the D-PMUs in the network lose their time synchronization.

Among the 13 cases mentioned above, DF and PLMDF can accurately locate the fault section in the first 9 cases, and IPLMDF can accurately locate the fault section in the first 10 cases. CS does not converge the output results in the 5th and 9th cases and outputs incorrect results in the 3rd case compared to the first three methods. In the 10th case, the results of DF do not converge, PLMDF falls into local optimum, IPLMDF accurately locates the fault, and CS does not accurately locate the fault section.

The 10th case community diagram is shown in Fig. 9. Vertex 6 is a branch vertex on the graph model with downstream vertices 7 and 26. until the D-PMU of vertex 7 loses synchronization, the fault graph weight of vertex 7 is less than 0 and does not cause misjudgment. After the loss of synchronization of the D-PMU at vertex 7, the estimated fault map weight obtained is greater than 0. At this point, the failure graph weights of both vertices downstream of vertex 6 are greater than 0, so the results of DF do not converge. However, vertex 6 and vertex 7 are in the same community, vertex 26 is in another community, which causes the PLMDF to fall into a local optimum during the query process, and the output fault section is 7,8. The community is larger in IPLMDF, vertex 6, vertex 7 and vertex 26 are merged into the same community and the output results in the global optimal solution, i.e., 28,29.

In the last three cases, the two D-PMUs closest to the

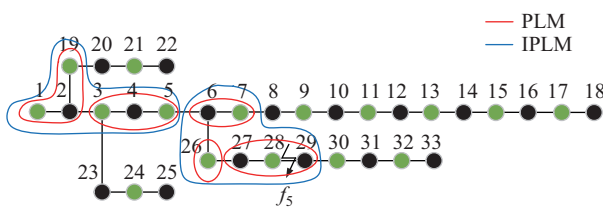


Fig. 9. The 10th case community graphs.

fault point lose synchronization, and none of the above four methods can locate the exact fault section.

In summary, when some D-PMUs lose time synchronization in the network, IPLMDF has the best positioning effect, followed by DF and PLMDF, and CS has the worst.

C. Comparison of the Calculation Speed of the Proposed Method

In the previous two sections, the accuracy of DF, PLMFDF, and IPLMDF were verified in the normal case and in the partial D-PMU out-of-step case. In this section, we want to compare the computational speed of the above four methods.

Since the number of iterations of the above four methods is primarily affected by the location of the fault section, the results of the midpoint of the line where the fault occurred and the fault resistance value of $1000\ \Omega$ are selected as being representative in this paper.

In Fig. 10, the horizontal coordinates are the fault locations $f_1 - f_6$ and the vertical coordinates are the number of iterations. Fig. 9 shows the number of iterations for each case in Table III, the horizontal coordinate is the serial number corresponding to each case, and the vertical coordinate is the

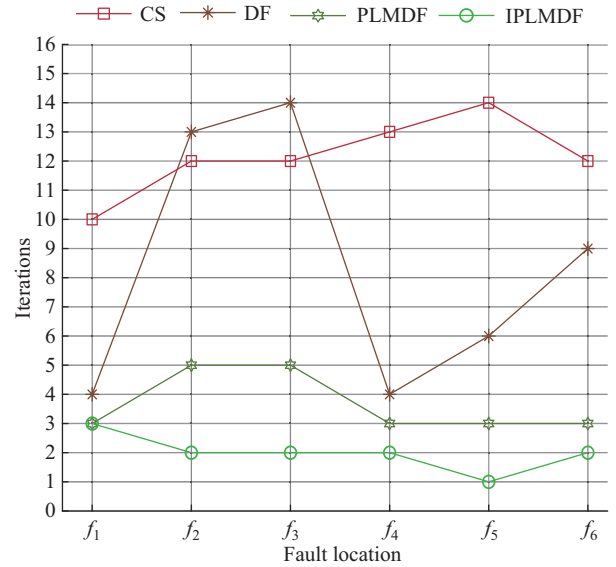


Fig. 10. Comparison of calculation speed of four methods under normal conditions.

TABLE III
ROBUSTNESS VERIFICATION OF THE PROPOSED METHOD

Serial number	The actual location of fault	ID of the D-PMU out of step	DF	PLMDF	IPLMDF	CS
1	4,5	28	4,5	4,5	4,5	3,5
2	13,14	9	13,14	13,14	13,14	13,15
3	13,14	9,11	13,14	13,14	13,14	7,9
4	14,15	9,17	14,15	14,15	14,15	13,15
5	14,15	3,7,11,21	14,15	14,15	14,15	5,7,26; 9,11; 13,15; 5,24
6	20,21	5,17,28	20,21	20,21	20,21	19,21
7	24,25	3	25,0	25,0	25,0	24,25
8	24,25	3,21,30	25,0	25,0	25,0	24,25
9	28,29	3,26	28,29	28,29	28,29	28,30; 5,12; 1,3
10	28,29	7,21	7,8; 28,29	7,8	28,29	28,30
11	4,5	3	1,2	1,2	1,2	1,3
12	14,15	13,15	13,14	13,14	13,14	11,13
13	20,21	19	19,20	19,20	19,20	1,3,19

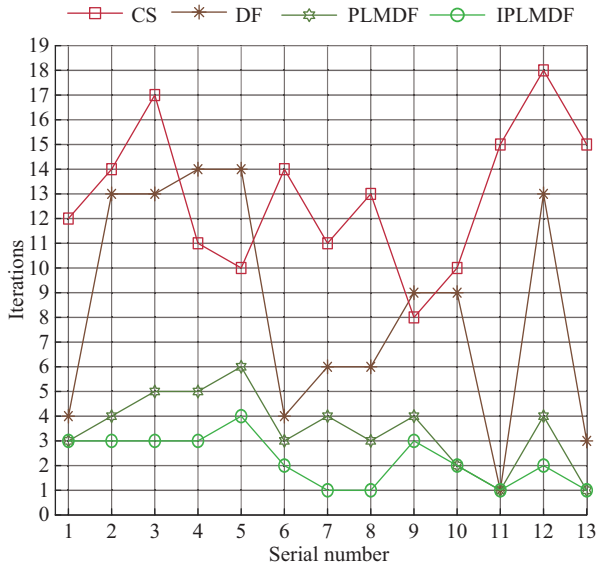


Fig. 11. Comparison of calculation speed of four methods under some D-PMUs out of step.

number of iterations.

From Figs. 10 and 11, we can conclude that: CS has unstable performance due to the random process, and the number of iterations is generally more than 10; DF is most sensitive to the topological location of the fault section, and it can quickly locate the fault section when it is close to the power node, but requires more iterations to locate the fault section if it is close to the end of the line; PLMDF is less affected by the topological location of the fault section than DF. Compared with DF, PLMDF is less affected by the topological location of the fault section, but still requires 5 or more iterations to locate the fault if the fault section is near the end of the line; IPLMDF has the most stable performance and usually requires only 2–3 iterations to directly and accurately locate the fault section.

By comparing the above three aspects, it can be intuitively determined that IPLMDF outperforms the other three methods both in terms of fault tolerance and computation speed.

VI. CONCLUSION

In this paper, a distribution network graph model is built on the graph database TigerGraph, and the fault graph weight of the vertices are defined to reflect the topological relationship between the vertices and the fault points, as well as the degree of fault impact on the vertex currents. In this paper, a fault location method based on community graph depth-first traversal is proposed for fast location of single-phase ground faults in distribution networks. The main idea of this approach is the use of an improved PLM to cluster the vertices in the graph model to form communities of different sizes. And then perform a depth-first traversal with the community as the smallest unit. The positioning accuracy, robustness, and computational speed of the four methods, DF, PLMDF, IPLMDF, and CS, were compared through simulation tests. With half of the nodes in the distribution network model configured with D-PMUs, IPLMDF can quickly and accurately determine the localization

of the fault sections. In addition, IPLMDF is applicable to the whole section of the line, independent of the fault resistance, and has a certain degree of robustness. In the case of partial D-PMU losing time synchronization, it is still possible to quickly and accurately locate the fault section, while avoiding the situation of falling into local optimal solutions.

Currently, the IPLMDF proposed in this paper can only be used to locate single-phase ground faults. In our next study, it is planned to extend the application of IPLMDF to all types of short-circuit faults.

REFERENCES

- [1] M. Gholami, A. Abbaspour, M. Moeini-Agtaie, M. Fotuhi-Firuzabad, and M. Lehtonen, "Detecting the location of short-circuit faults in active distribution network using PMU-based state estimation," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1396–1406, Mar. 2020.
- [2] M. M. Saha, J. Izykowski, and E. Rosolowski, "Fault Location on Power Networks," *New York: Springer*, 2010, pp. 1–4.
- [3] J. Liu, X. Q. Zhang, W. Shen, L. Quan, and Z. H. Zhang, "Performance testing of single phase to earth fault location for distribution network with neutral point non-effectively grounded systems," *Automation of Electric Power Systems*, vol. 42, no. 1, pp. 138–143, Jan. 2018.
- [4] M. Daisy and R. Dashti, "Single phase fault location in electrical distribution feeder using hybrid method," *Energy*, vol. 103, pp. 356–368, May 2016.
- [5] S. Jamali and A. Bahmanyar, "A new fault location method for distribution networks using sparse measurements," *International Journal of Electrical Power & Energy Systems*, vol. 81, pp. 459–468, Oct. 2016.
- [6] R. Chen, X. Yin, X. G. Yin, Y. L. Li, and J. Y. Lin, "Computational fault time difference-based fault location method for branched power distribution networks," *IEEE Access*, vol. 7, pp. 181972–181982, Feb. 2019.
- [7] H. C. Shu, J. L. Liu, X. C. Tian, "Fault Location of Radial Distribution Network Based on Sudden Changes of Fault Traveling Wave Along Line and Model Matching," *Automation of Electric Power Systems*, vol. 44, no. 9, pp. 158–163, May. 2020.
- [8] P. Wang, B. C. Chen, H. Zhou, C. H. Tian, and B. Sun, "Fault location in resonant grounded network by adaptive control of neutral-to-earth complex impedance," *IEEE Transactions on Power Delivery*, vol. 33, no. 2, pp. 689–698, Apr. 2018.
- [9] X. W. Wang, Wang, H. X. Zhang, F. Shi, Q. W. Wu, V. Terzija, W. Xie, and C. Fang, "Location of single phase to ground faults in distribution networks based on synchronous transients energy analysis," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 774–785, Jan. 2020.
- [10] C. B. Wang, Z. H. Yun, H. X. Zhang, F. Shi, P. Ling, and B. P. Xie, "Parameter-free fault location algorithm for multi-terminal overhead transmission line of distribution network based on μ MPMU," *Power System Technology*, vol. 43, no. 9, pp. 3202–3209, Sep. 2019.
- [11] J. F. Liang, T. J. Jing, H. N. Niu, and J. B. Wang, "Two-terminal fault location method of distribution network based on adaptive convolution neural network," *IEEE Access*, vol. 8, pp. 54035–54043, Mar. 2020.
- [12] K. J. Chen, J. Hu, Y. Zhang, Z. Q. Yu, and J. L. He, "Fault location in power distribution systems via deep graph convolutional networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 119–131, Jan. 2020.
- [13] L. Zhang, B. Zhao, L. J. Wang, and X. Q. Fan, "On-line fault location method for ungrounded system based on graph theory," *Power System Technology*, vol. 43, no. 8, pp. 3025–3031, Aug. 2019.
- [14] K. Jia, L. Li, Z. Yang, G. K. Zhao, and T. S. Bi, "Research on distribution network fault location based on Bayesian compressed sensing theory," *Proceedings of the CSEE*, vol. 39, no. 12, pp. 3475–3485, Jun. 2019.
- [15] Z. X. Zhou, C. Yuan, Z. Y. Yao, J. P. Dai, G. Y. Liu, R. C. Dai, Z. W. Wang, and G. M. Huang, "CIM/E oriented graph database model architecture and parallel network topology processing," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 2018, pp. 1–5.
- [16] H. Huang, J. P. Dai, Y. Wang, X. Shan, M. Y. Zhai, and G. Y. Liu, "Graph database based construction and network topology of CIM/E for power grid," *Automation of Electric Power Systems*, vol. 43, no. 22, pp. 122–129, Nov. 2019.

- [17] C. Yuan, Y. Q. Zhou, G. F. Zhang, G. Y. Liu, R. C. Dai, X. Chen, and Z. W. Wang, "Exploration of graph computing in power system state estimation," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 2018, pp. 1–5.
- [18] G. Y. Liu, K. W. Liu, D. Shi, W. D. Zhu, Z. W. Wang, and X. Chen, "Graph computation and its applications in smart grid," in *2017 IEEE International Congress on Big Data (BigData Congress)*, 2017, pp. 507–510.
- [19] J. P. Dai, Z. Y. Yao, G. F. Zhang, G. Y. Liu, R. C. Dai, and Z. W. Wang, "Graph computing-based real-time network topology analysis for power system," in *2019 IEEE Power & Energy Society General Meeting (PESGM)*, 2019, pp. 1–5.
- [20] R. C. Dai, G. Y. Liu, Z. W. Wang, B. W. Kan, and C. Yuan, "A novel graph-based energy management system," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 1845–1853, May 2020.
- [21] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, pp. 10008, Oct. 2008.
- [22] C. L. Staudt and H. Meyerhenke, "Engineering parallel algorithms for community detection in massive networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 1, pp. 171–184, Jan. 2016.
- [23] X. L. Huang, Z. Y. Xie, and X. Y. Huang, "Fault location of distribution network base on improved cuckoo search algorithm," *IEEE Access*, vol. 8, pp. 2272–2283, Dec. 2019.



Rong Jia received the Ph.D. degree from Xi'an University of Technology, Shannxi, China, in 1999. He is currently a professor with the Department of Electrical Engineering, Xi'an University of Technology, Shannxi, China. His research interest includes power system operations and control, and power equipment status monitoring.



Xiaowei Wang received the Ph.D. degree from Xi'an Jiaotong University, Shannxi, China, in 2019. He is currently an Associate Professor with the Department of Electrical Engineering, Xi'an University of Technology, Shaanxi, China. His research interest includes small current grounding system and fault line selection.



Jian Dang received the Ph.D. degree from Xi'an University of Technology, Shannxi, China, in 2018. He is currently an Associate Professor with the Department of Electrical Engineering, Xi'an University of Technology, Shaanxi, China. His research interests include fault location of distribution networks, artificial Intelligence and graph applications in power systems.



Hui Wei received the B.S. degree in Electrical Engineering at Xi'an University of Technology, Shannxi, China, in 2015. She is currently pursuing the M.S. degree in Electrical Engineering at Xi'an University of Technology. Her research interests include knowledge graph applications in power systems, power equipment state estimation.



Yunjiang Yan received the B.S. degree in Electrical Engineering at Xi'an University of Technology, Shannxi, China, in 2015. He is currently pursuing the M.S. degree in Electrical Engineering at Xi'an University of Technology. His research interests include graph computation application in power system, distribution network fault location.