Integrated Framework of Multisource Data Fusion for Outage Location in Looped Distribution Systems

Liming Liu, *Graduate Student Member, IEEE*, Yuxuan Yuan, *Member, IEEE*, Zhaoyu Wang, *Senior Member, IEEE*, Yiyun Yao, *Member, IEEE*, and Fei Ding, *Senior Member, IEEE*

***Abstract*—Accurate outage location is essential for expediting post-outage power restoration, minimizing outage duration, and enhancing the resilience of distribution networks. With the advent of advanced metering infrastructure, data-driven outage location methods have significantly advanced beyond traditional approaches that rely on manual inspections. However, existing methods still face critical challenges, like reliance on single-source data, limited ability to handle partially observable systems or difficulties with loop networks. To the best of our knowledge, no single approach has comprehensively addressed all of these challenges at once. To this end, this paper proposes a comprehensive multisource data fusion framework for outage locations via probabilistic graph networks. The framework consists of three key phases. First, a novel method for reconstituting distribution networks with loops is developed, transforming looped networks into multiple radial subnetworks that retain all outage causalities of the original network. Second, Bayesian network (BN) models are established for each subnetwork, integrating multiple data sources and network structures. Finally, a joint Gibbs sampling mechanism, featuring forward and backward information flow, is designed to merge data from separate BN models and maximize the utilization of limited evidence, ensuring accurate outage location identification. The framework was validated on two modified public test systems, and comparative studies confirmed its effectiveness.**

***Index Terms*—Distribution system resilience, outage location, probabilistic graph model, multisource data fusion.**

Nomenclature

Abbreviations

AMI	Advanced Metering Infrastructure
BN	Bayesian Network
Br-Le	Branch Level
DER	Distributed Energy Resources
DFS	Depth-first Search
DN	Distribution Network
DSO	Distribution System Operator
EDR	Evidence Density Ratio
JGS	Joint Gibbs Sampling
LPM	Linear Pooling Model
MDF	Multisource Data Fusion
Obs-Le	Levels of Observability
PDF	Probability Density Function
SCADA	Supervisory Control and Data Acquisition
SM	Smart Meter
SVM	Support Vector Machine
Sys-Le	System Level

Constants

α	Information entropy-based weight
β	Fixed weight for distribution mixture
ΔT	Information collection time after outage
η	Set of all evidence types
τ	Threshold for sampling branch/customer status
K	JSP iteration number
M	Minimum number of constructed subnetworks

Indices and Sets

$\mathcal{B}$	Probabilistic graph network set
$\mathcal{B}_{\text{seq}}$	Ordered BN set
G^s	Set that collects all the reconstructed subnetworks
L_i	Path set of the node i
P_N	Path sets of all the nodes in the network

Variables

$\bar{P_\phi}$	Weighted mixture distribution from fusion operation
$\boldsymbol{C}$	Status of customer switches
$\boldsymbol{D}$	States of the primary network branches
$\boldsymbol{h}(\cdot)$	Information entropy-based metric
$\mathcal{Y}$	Distribution network topology
$\mathcal{Z}$	Multisource evidence
$\Psi(\cdot)$	Probability fusion module
$B_{\sigma(m)}$	m-th BN in the $\mathcal{B}_{\text{seq}}$
c_i^j	Status of customer i connected to branch i
$Ch(\cdot)$	Represents the child nodes of the target node
$d_i^{(k),[m]}$	State of the branch i in subnetwork m in $(k+1)_{th}$ JGS iteration
d_i	Binary state of branch i

Received 2 October 2024; revised 10 January 2025; accepted 1 February 2025. This work was supported in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DoE) under Contract DE-AC36-08GO28308; in part by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technology Office under Agreement 40385; in part by the National Science Foundation under ECCS 2042314; and in part by the Power System Engineering and Research Center (PSERC S-110). Paper no. TSG-01700-2024. *(Corresponding author: Zhaoyu Wang.)*

Liming Liu, Yuxuan Yuan, and Zhaoyu Wang are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: limingl@iastate.edu; yuanyx@iastate.edu; wzy@iastate.edu).

Yiyun Yao and Fei Ding are with the Power System Engineering Center, National Renewable Energy Laboratory, Golden, CO 80214 USA (e-mail: yiyun.yao@nrel.gov; fei.ding@nrel.gov).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2025.3540979.

Digital Object Identifier 10.1109/TSG.2025.3540979

$F_x^{-1}(\cdot)$	Inverse transform to sample states from distribution
$G^s_{\sigma(m)}$	Subnetwork utilized to build $B_{\sigma(M)}$
$Pa(\cdot)$	Represents the parent nodes of the target node
$r_e^{[i]}$	Evidence density ratio of subnetwork i
z_i^b	Grid parameters of branch i
$z_{i,j}^h$	Human-related evidenceof customer j
$z_{i,j}^m$	Meter-related information of customer j
z_i^v	Vegetation data of branch i
z_i^w	Weather conditions of branch i

I. Introduction

SEVER power outages caused by recent extreme weather events have emphasized the urgent need to improve the resilience of distribution power systems [1]. One critical yet challenging aspect of strengthening power system resilience is the accurate and efficient location of the outages, particularly within DNs, where the majority of outage events occur [2], [3]. Traditionally, outage locations are identified through customer trouble calls or manual inspections. However, relying on trouble calls alone is unreliable, as it is estimated that only one-third of customers report outages within the first hour post outage [4]. While manual inspections, combined with expert knowledge, can provide acceptable outage locations [5], this approach is labor-intensive, costly, and time-consuming, making it suboptimal for DSOs.

The recent development of AMI-based techniques has brought promising solutions to the outage location problem. Through bidirectional communication, SMs can transmit "last gasp" signals to utilities when there is a loss of power [6]. While some utilities have fully observable systems, meter malfunctions and communication delays can render it impractical for utilities from relying solely on last gasp signals to accurately assess the current state of the system and further make informed predictions regarding the location of outages in real time. In addition to SMs, other advanced sensors with real-time communication abilities (e.g., second-level line measurements) have also shown potential for solving outage detection issues and have been explored in previous studies [7]. However, due to budgetary constraints of utilities, the widespread deployment of these advanced devices, particularly among smaller utilities, remains limited [8]. Furthermore, the growing integration of DERs has added complexity to the design and operation of DNs [9], [10], raising concerns about the continued effectiveness of traditional outage location methods. These challenges highlight the needs for more practical and scalable solutions to improve outage location.

To tackle these challenges, recent studies have increasingly focused on data-driven methods for outage detection. Existing research in this area can be broadly categorized into two groups based on the data sources used: SM data-based and non-SM data-based methods.

Class I - SM data-based methods: These methods primarily leverage SM measurements and last gasp signals for outage detection. Reference [11] proposed a classification-based outage location model using a multi-label SVM, where the SMs' last gasp signals are used to pinpoint outage branches in fully observable networks. In our previous work [2], a generative adversarial network-based approach was introduced to detect outage regions, even in partially observable systems, distinguishing it from the method in [11]. Similarly, [12] introduced a probabilistic and fuzzy logic algorithm for analyzing outage data using AMI. Reference [13] developed an outage monitoring method leveraging stochastic time series analysis and SM voltage measurements, which showed significant changes post-outage. This method was validated on both radial and looped DNs, which are common in urban settings. Reference [14] proposed a spectral clustering method based on SM outage notifications, which provides accurate outage detection results, but the large outage areas identified instead of branch-level results offer limited information to operators.

Class II - non-SM data-based methods: in contrast to Class I methods, Class II approaches leverage information from various external sources to detect outages in distribution systems. Reference [15] utilized social sensors within a probabilistic framework for outage detection, while [16] integrated weather data into an ensemble learning model to identify outages in distribution systems. In [17], an outage location framework tailored for systems with tree structures is proposed. This framework integrates real-time line flow measurements with predicted loads, facilitating both efficient outage detection and optimal sensor placement. Moreover, [7] presents a mixed-integer linear programming model that utilizes line flow measurements and AMI data to identify the topology of the distribution system under various operation conditions, outages and normal situations. Similarly, [18] addresses outage identification, system state estimation, and topology error correction concurrently, through an optimization framework based on mixed-integer quadratic programming. Despite the increasing deployment of distribution system line measurements in some utilities, widely equipping such measurements remains impractical due to budget limitations. Rather than relying on new sensor installations, researchers have increasingly focused on leveraging the complementary nature of various data sources to enhance outage detection. In [19], a two-phase knowledge-based system for outage location is proposed, which fuses multiple data sources. This framework integrates traditional escalation to locate outage areas and meter polling to confirm statuses, using data from trouble calls, SCADA systems, and automated meter readings. Further advancing multisource data fusion, [20] proposes a transformer-based deep learning model that fuses operational and meteorological data to provide power outage warnings. Similarly, [21] utilizes BNs to incorporate multisource evidence and network structures, enabling accurate outage location in partially observable distribution systems. However, due to the inherent limitations of BNs, this method is not suitable for looped DNs.

Despite extensive research on data-driven outage location methods, several challenges are yet to be unresolved. First, assuming full observability across all distribution systems is impractical, as not all customers have SMs, and SM signal communication failures during extreme events can undermine the model performance. Second, while some methods address partial observability, they lack the granularity needed for branch-level outage detection, as they rely on a single data source. Third, most methods are designed for radial networks, with only a few that can be extended for loop systems. However, these methods don't account for the unique characteristics of looped systems, resulting in a lack of stability. To tackle these challenges, this study introduces a comprehensive framework for integrating information from multiple data sources to pinpoint outage locations in distribution systems. The key contributions are as follows:

- This research develops a comprehensive MDF framework for outage location in distribution systems. By fully utilizing the complementary characteristics of multiple data sources, the integrated framework accommodates varying levels of system observabilities and provides stable outage location results.
- A novel network reconstitution method is developed for DNs with loops, which examines the constraints of employing BNs in outage location applications, specifically focusing on the limitations imposed by the use of directed acyclic graphs. The method serves as a foundational step for the framework, enabling our outage location framework to be applied to both radial and looped networks.
- A JGS mechanism is proposed to infer outage locations based on the multiple BNs. It addresses the high-dimensional challenges posed by multi-source data and the application of the framework in large-scale DNs. By incorporating both information forward and backward phases, the utilization of limited evidence, which is common in real scenarios, is optimized.

The remainder of this paper is organized as follows. Section II outlines the problem statement for outage location. In Section III, the DN reconstitution method is presented. Section IV details the proposed multisource data fusion framework. Numerical results are provided in Section V, and conclusions are drawn in Section VI.

II. Problem Statement of Outage Location

Outage events inherently result in topological changes within the electrical grid, making the identification of outage locations dependent on inferring the probabilities of various post-event operational topologies [21]. Effective outage location relies on comprehensive outage information. Traditional methods, which often rely solely on customer reports, are limited in their effectiveness. By contrast, integrating a broader spectrum of outage-related data, also referred to as multiple data sources or multisource data, such as SM last gasp signals, customer reports, and weather information, has the potential to significantly enhance both the accuracy and timeliness of outage detection [22]. The diverse data sources complement one another, addressing issues like low SM coverage or limited customer reporting, without the need for additional metering devices. Weather information, such as wind speed and extreme weather events, is a critical factor in the reliability of distribution systems and has been previously incorporated into outage location studies [16]. This information is typically obtainable from local weather stations and online sources. Despite industrial surveys indicating that utilities often underutilize SM data [23], last gasp signals, which are easily retrieved through direct communication with individual SMs, are already integrated into outage management systems as a critical complement to customer trouble reports. Moreover, technological advancements, particularly in natural language processing, have facilitated the extraction of valuable data from social media platforms such as Facebook and Twitter, which function as social sensors [15]. This data can be systematically processed and converted into binary evidence, making it suitable for further application in outage detection and analysis. While this study emphasizes specific data sources, it does not imply that these are the only applicable inputs for the proposed outage detection model. With the continued development of measurement technologies, devices, e.g., micro phasor measurement units [24] and fiber optic sensors [25] also offer significant potential for enhancing outage detection and localization. However, to preserve the model's general applicability across various DSOs, these sources are not extensively discussed in this work. Nonetheless, the proposed method, designed as a general framework, retains the flexibility to incorporate additional data sources as they become available.

The data from various sources discussed above can serve as evidence supporting the accurate localization of outages. We utilize Bayes theory to mathematically formulate the outage inference process based on the multisource evidence $\mathcal{Z}$. The conditional PDF of DN topology $\mathcal{Y}$, considering the post-outage evidence $\boldsymbol{\mathcal{Z}}$, is expressed as $P(\mathcal{Y} = y \mid \boldsymbol{\mathcal{Z}} = z)$. It is derived from the joint distribution of $\mathcal{Y}$ and $\boldsymbol{\mathcal{Z}}$, denoted by $P_{\mathcal{Y},\boldsymbol{\mathcal{Z}}}(g, z)$ and marginal distribution of $\boldsymbol{\mathcal{Z}}$, i.e., $P_{\boldsymbol{\mathcal{Z}}}(z)$. The outage location is identified by determining the most likely candidate topology, which is achieved by maximizing the conditional PDF as follows:

$$y^* = \arg\max_{y} P(\mathcal{Y} = y|\boldsymbol{\mathcal{Z}} = z) = \frac{P_{\mathcal{Y},\boldsymbol{\mathcal{Z}}}(y, z)}{P_{\boldsymbol{\mathcal{Z}}}(z)}, \tag{1}$$

where y^* represents the most probable DN connection following the outage. The variable $\mathcal{Y}$ is a multinomial variable that captures the states of the primary network branches $\boldsymbol{D}$ and the status of customer switches $\boldsymbol{C}$, collectively represented as $\{\boldsymbol{D}, \boldsymbol{C}\}$. The variable $\boldsymbol{D} \in \mathbb{B}$, where $\mathbb{B} = \{0, 1\}$, indicates the binary state of the primary network branches. Specifically, $d_i = 1$ if d_i with $i \in \{1, 2, \ldots, n\}$ is outaged, and $d_i = 0$ otherwise. Similarly, $\boldsymbol{C} \in \mathbb{B}$ for customers, takes the value 1 if the customer is disconnected from the network. The process of maximizing over topology candidates can be achieved by identifying the most probable states of individual branches or customers through their respective conditional PDFs, $P_{D_i|\mathcal{Z}}(d_i \mid z)$ and $P_{C_i^j|\mathcal{Z}}(c_i^j \mid z)$, which are formulated as follows.

$$P_{D_i|\mathcal{Z}}(d_i \mid z) = \sum_{\{\boldsymbol{d},\boldsymbol{c}\}\setminus d_i} P_{\boldsymbol{D},\boldsymbol{C}|\mathcal{Z}}(\boldsymbol{d}, \boldsymbol{c} \mid z) = \sum_{\{\boldsymbol{d},\boldsymbol{c}\}\setminus d_i} \frac{P_{\boldsymbol{D},\boldsymbol{C},\mathcal{Z}}(\boldsymbol{d}, \boldsymbol{c}, z)}{P_{\mathcal{Z}}(z)} \tag{2}$$

$$P_{C_i^j|\mathcal{Z}}\left(c_i^j \mid z\right) = \sum_{\{\boldsymbol{d},\boldsymbol{c}\}\setminus c_i^j} P_{\boldsymbol{D},\boldsymbol{C}|\mathcal{Z}}(\boldsymbol{d}, \boldsymbol{c} \mid z) = \sum_{\{\boldsymbol{d},\boldsymbol{c}\}\setminus c_i^j} \frac{P_{\boldsymbol{D},\boldsymbol{C},\mathcal{Z}}(\boldsymbol{d}, \boldsymbol{c}, z)}{P_{\mathcal{Z}}(z)} \tag{3}$$

where, $A \setminus B$ denotes the set of elements in A but not in B.

Generally, the primary objective of the outage location by multisource data fusion is to solve (2) and (3) after the outage occurs. To accomplish this, an outage location method based on a BN is introduced in our previous work [21]. The BN approach significantly reduces the computational complexity of outage location inference in high-dimensional spaces and effectively leverages tree-like features of DNs to facilitate outage inference. However, the basic BN model is limited in its application to DNs with loops, as it is only suitable for directed acyclic graphs. Despite this limitation, the characteristics of the BN model offer distinct advantages for outage location, particularly in weakly meshed networks, where most of the network remains radial. To overcome the limitations of the existing model, an integrated outage location framework via BN is proposed in this paper, which is designed to handle both looped and radial DNs.

Given the BNs are only applicable to directed acyclic graphs, the fundamental and direct strategy for managing weakly meshed networks is to reconstitute the complex network into multiple radial networks. These radial networks must encompass all the outage causalities present in the original meshed network. For instance, if the system network contains a loop, the target branch within the loop will have two potential paths to the power source. These two paths represent all the potential routes for energizing the target branch. By decomposing the weakly meshed network into two radial subnetworks, each containing one of the paths, the BN can be utilized to infer the status of the target branch based on the evidence from these subnetworks. Consequently, the outage inference results for the target branch in the DN can be obtained by combining the inference results from multiple BNs. This idea is clear, but it raises three essential questions: 1) How can we decompose the looped network into radial subnetworks that comprehensively capture all potential outage scenarios while ensuring the decomposition is optimal? 2) How can we extend our existing BN model to perform inference for each subnetwork? 3) How can we effectively integrate the inference results from each subnetwork to derive a reliable final outage location?

In this paper, our proposed integrated outage location framework addresses these questions separately. Detailed explanations and solutions will be provided in the following sections.

III. Distribution Network Reconstitution

To address the first question, it is essential to design a method for decomposing and reconstituting the DN with loops. This section proposes a two-step approach, consisting of a DFS-based network decomposition model and a network reconstruction model using an iterative method, serving as the first module of the framework.

A. DFS-Based Network Decomposition Model

The problem of ascertaining the energization status of a branch or customer can be formulated as identifying feasible paths between the concerned branch or customer and power supply sources, herein represented by substations. Consider a DN represented by an undirected graph $\mathcal{G} = (V, E)$, where $V = \{v_1, \ldots, v_n\} \cup \{v_s\}$ denotes the set of primary bus nodes and $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$ denotes the set of primary network branches. During a system outage, branches that lose power are categorized into the set E^F, while those remaining energized constitute the set E^N, satisfying $E = E^N \cup E^F$. For each node v_i in the network, there exists an ensemble of paths, denoted $L_i = \{l_i^1, \ldots, l_i^b\}$, each of which establishes connectivity from v_i to the source node v_s. A path l_i^b is classified as "active" if all its constituent edges are members of E^N, i.e., $\forall e_j \in l_i, e_j \in E^N$. The aggregation of all active paths of node v_i is denoted as L_i^a. Outage areas within the network can be precisely identified by locating nodes for which the set L_i^a is empty. Hence, to deduce the status of variable $\boldsymbol{C}$ or $\boldsymbol{D}$, the initial step is to identify the ensemble set of L_i for each node i. This process decomposes the original network into multiple sets of paths. While manual decomposition is feasible for simple networks, it becomes impractical for networks with thousands of buses. Consequently, we introduce a DFS-based model for network decomposition to efficiently handle complex network structures.

DFS is an algorithm used for traversing or searching tree or graph data structures. Starting from a root node, it explores each branch as deeply as possible before retracing its steps. Essentially, it dives deep into a graph, visiting a node and proceeding to its adjacent, unvisited nodes sequentially until it reaches a dead-end. Then, it backtracks and explores other unvisited nodes in a similar manner. In our case, for each node v_i, we use the DFS method to find all paths from the node v_i to v_s, i.e., L_i, then the path sets of all the nodes form the network path set $P_N = \{L_i | i = 1, \ldots, n\}$, which will be the input of second step. The details of the DFS-based network decomposition method are provided in **Algorithm 1**.

Algorithm 1 DFS-Based Network Decomposition Algorithm

Require: Adjacency matrix of the network $\boldsymbol{A}$, target node index i, power source node index s.

initialize path set of node i as $L_i = \{\}$; current directed path $\boldsymbol{c}$ and its end node v_e; visited status as $\boldsymbol{q} \in \mathbb{R}^{1\times n}$
execute the procedure ***DFS***
procedure ***DFS***($\boldsymbol{A}$, i, s, $\boldsymbol{q}$, $\boldsymbol{c}$)
    if $i == s$ **then**
        $L_i \leftarrow L_i \cup \{\boldsymbol{c}\}$
        return
    end if
    $\boldsymbol{q}_i \leftarrow 1$; $B \leftarrow \{j | \boldsymbol{A}_{i,j} > 0\}$; $\boldsymbol{c} \leftarrow \boldsymbol{c} + (v_i, v_e)$
    for k from B **do**
        if $\boldsymbol{q}_k > 0$ **then**
            DFS ($\boldsymbol{A}$, k, s, $\boldsymbol{q}$, $\boldsymbol{c}$)
        end if
    end for
    $\boldsymbol{q}_i \leftarrow 0$
    return
end procedure
return L_i

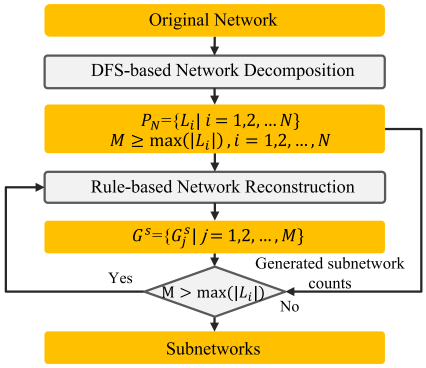


Fig. 1. The flowchart of the network reconstruction via iterative method.

B. Network Reconstruction via Iterative Method

The status of each path from L_i, whether active or inactive, can be inferred either directly from multisource information or derived from the previous BN-based outage location method [21]. However, implementing these ideas presents practical challenges. Firstly, the computational feasibility of these approaches is compromised by the exponential increase in the number of paths within L_i, especially in complex meshed DNs. Secondly, a high degree of redundancy exists among the models, as paths in L_i may share a majority of nodes or branches, leading to the construction of similar models for overlapping paths. Thirdly, treating the status of paths in isolation impedes the flow of observable information, such as trouble calls, between models, resulting in suboptimal data utilization and, ultimately, diminishing the model performance. To address these issues, we propose a network reconstruction method that employs an iterative method. The reconstructed networks, referred to as subnetworks below, enhance the efficiency and accuracy of downstream inference models in pinpointing outage locations. The flowchart of the method is shown in Fig. 1. Based on the network path set P_N, the rule-based network reconstruction is performed, with the primary rule being the avoidance of looped subnetworks during path merging. The objective is to create equivalent networks without loops, ensuring they are suitable for BN construction. The algorithm proceeds by first selecting one of the active paths l_m^b of node m that has the largest active path set L_m to serve as the initial subnetwork G_j^s. This is because the number of final subnetworks cannot be smaller than the maximum size of any path set, as discussed in more detail below. Another active path l_u^q is randomly selected from the set $P_N \setminus \{L_m, G_j^s\}$. If merging G_j^s and l_u^q does not create loops, the merge proceeds, and G_j^s is updated accordingly. This merging process is repeated until no more paths can be combined without creating loops. Once merging is complete for G_j^s, it is stored, and the process begins anew with another path l_m^{b+1} from L_m. This continues until all active paths from L_m are incorporated into subnetworks. If active paths remain unprocessed, the algorithm restarts the merging process until all paths are considered. The stopping criterion here is based on reaching the theoretical minimum number of subnetworks. Without this criterion, the random selection of paths during the merging process could lead to variations in the composition of G^s and the total number of subnetworks. To minimize the computation burden of subsequent BN modeling and outage inference, it is advantageous to find the G^s with the smallest possible size $|G^s|$, which has been proven to correspond to the maximum size of the path set L_i across all nodes $i = 1, 2, \ldots, N$. The details are outlined in **Proposition 1**. All the subnetworks G_j^s form the subnetwork set $G^s = \{G_j^s \mid j = 1, \ldots, M\}$. The resulting subnetworks capture all outage causalities of the original network, and the proposed method ensures the minimum number of subnetworks, thereby reducing the complexity of the subsequent outage inference models.

Proposition 1: The lower bound for the number of constructed subnetworks, denoted as M, is at least equal to the maximum size of the path set L_i across all nodes $i = 1, 2, \ldots, N$. Formally, this is expressed as $M \geq \max(|L_i|)$.

Proof: Assume, for the sake of contradiction, that $M < \max(|L_i|)$. Let L_k be the path set with the largest number of paths, i.e., $k = \mathrm{argmax}_i |L_i|$. In this case, more than one path from L_k would need to be merged into the same subnetwork. Consider any two distinct paths l_k^r and l_k^t from L_k. Both paths include the target node v_k and the source node v_s. Merging these paths into the same subnetwork would result in a loop, violating the network reconstruction rule. Thus, our assumption that $M < \max(|L_i|)$ leads to a contradiction. Therefore, $M \geq \max(|L_i|)$, proving the proposition.

C. Discussions on Network Reconstitution

Regarding the proposed DN reconstitution method, several points should be further clarified. Firstly, computation efficiency is an important consideration for making the method practical. However, since the network reconstruction is conducted offline, it has a higher tolerance for computation time compared to online methods. Therefore, computing efficiency is not our primary concern. Additionally, as previous work has shown [26], [27], the looped distribution power system is usually weakly meshed with a small number of cycles, resulting in a lower computational burden. Consequently, the computation cost is not a significant issue for our method. Furthermore, distribution network reconfiguration is common in operational practice. By handling the topology reconfiguration process offline, frequent topology changes do not significantly impact the efficiency of outage location inference. For each topology configuration scenario, a dedicated network reconstitution is performed, and the corresponding subnetwork sets G^s are precomputed and stored. During real-time operations, the appropriate model is selected based on the current topology, enabling efficient and seamless integration into the outage location framework. Additionally, the proposed framework is designed to operate in three-phase systems. The network reconstitution process and the outage detection framework are executed independently for each phase. The results from the three phases are subsequently integrated to determine the final outage location outcomes. For simplicity and ease of demonstration, this paper focuses on single-phase networks to illustrate the performance of the proposed model.

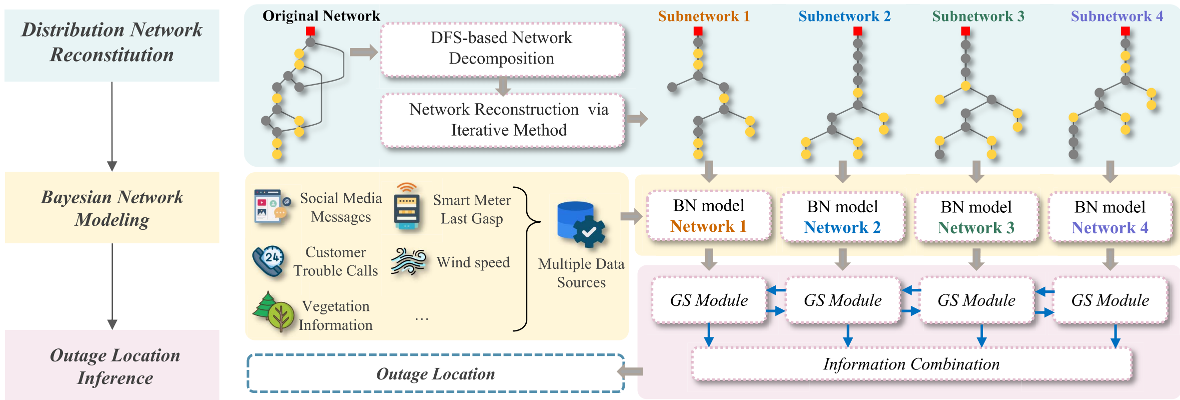


Fig. 2. The structure of the integrated framework of multisource data fusion for outage location.

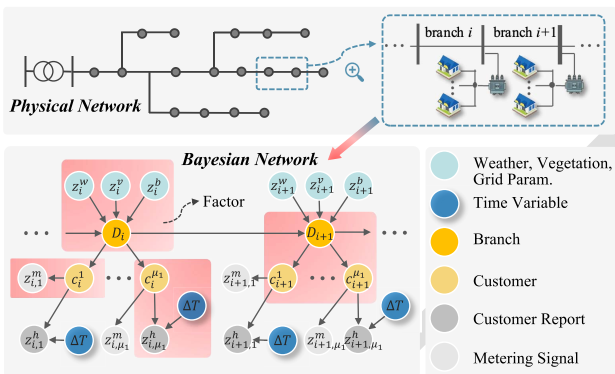


Fig. 3. Illustration of the BN model structure and four main factors.

IV. Probabilistic Graph Inference-Based Outage Location Framework

Based on the DN reconstitution method, subnetworks capturing all causal information are prepared for outage location. A probabilistic graph model, specifically a BN model, is applied to each subnetwork. The BN-based approach addresses computational complexity and prevents overfitting in outage location inference. The key strength of this approach lies in its ability to seamlessly integrate diverse data sources by exploiting the conditional independencies present in both the grid and data. These independencies allow for a scalable and efficient graphical representation, improving the accuracy and efficiency of outage inference. All BN models are combined and inferred through the JGS mechanism. These procedures form the outage location framework, as shown in Fig. 2.

A. Outage Location Based on BN Model

Specifically, the proposed method decomposes the joint PDF $P_{\boldsymbol{D},\boldsymbol{C},\boldsymbol{\mathcal{Z}}}(\boldsymbol{d},\boldsymbol{c},\boldsymbol{z})$ into a series of smaller, more manageable factors. As discussed in Section II, multisource data can be collected post-outage to form the evidence set $\boldsymbol{\mathcal{Z}}$. To maintain generality, in this work, we select outage evidence from the customer side, including trouble calls and social media messages gathered within ΔT after the outage, which is categorized as human-related evidence and denoted by $z^h_{i,j}$. The last gasp signals from SMs are classified as meter-related information, denoted by $z^m_{i,j}$. Additionally, weather conditions, vegetation data, and grid parameters are considered, denoted by $\{z^w_i, z^v_i, z^b_i\}$, respectively. Then, the construction of the BN, based on the structure of the physical DN and the dependencies between variables and various forms of evidence, is illustrated in Fig. 3. In this BN-based representation, four main factors, encoded in a graph structure and marked in red, compactly break down the original high-dimensional joint PDFs. The factor $P_{D_i|Pa(D_i)}(d_i \mid Pa(d_i))$ represents the relationship of branch d_i with four parent variables, denoted as $Pa(d_i) = \{d_{i-1}, z^w_i, z^v_i, z^b_i\}$, which has direct causal influences. Factor $P_{C^j_i|Pa(C^j_i)}(c^j_i \mid Pa(c^j_i))$ represents the conditional PDF of the status of customer j given parent variables. The evidence $z^h_{i,j}$ is determined by the two parent variables: customer status and time after the outage. This relationship is captured by the factor $P_{\mathcal{Z}^h_{i,j}|Pa(\mathcal{Z}^h_{i,j})}(z^h_{i,j} \mid Pa(z^h_{i,j}))$. Considering the SM signals will be delivered to utilities almost instantaneously after the outage, the parent of evidence $z^m_{i,j}$ will be only customer status. Therefore, the final factor is constructed as $P_{\mathcal{Z}^m_{i,j}|Pa(\mathcal{Z}^m_{i,j})}(z^m_{i,j} \mid Pa(z^m_{i,j}))$. By utilizing this computationally efficient BN-based method, the conditional PDF for the state of each primary branch can be quickly inferred based on the data from multiple sources, enabling rapid identification of outage locations. More details about the BN structure development and parameterization can be found in [21].

Following the above procedure, a BN can be constructed for each subnetwork in G^s. These BNs collectively form the model set $\mathcal{B}$. In the next section, we will present an advanced joint inference method for $\mathcal{B}$, enabling the identification of final outage locations. It is important to note that for the original network, the unknown state variables $\{\boldsymbol{D}, \boldsymbol{C}\}$ represent the universal set. For each subnetwork, the state variables of the customers and branches will form a subset of this universal set, marked as $\{\boldsymbol{D}^{[i]}, \boldsymbol{C}^{[i]}\}$ for B_i, and $(\cdot)^{[i]}$ represents the relationship with B_i.

B. Outage Location Inference Using Joint Gibbs Sampling

Through network reconstitution and the construction of the outage inference models, the original outage problem has been significantly simplified. However, directly solving equations (2)-(3) remains impractical due to the computationally expensive summation operations $P_{\boldsymbol{\mathcal{Z}}}(z)$ overall nodes of the graph simultaneously, particularly in large-scale DNs [28]. To address this issue, the Gibbs sampling (GS) algorithm can be introduced to conduct the outage inference over the BN model. GS is a Markov Chain Monte Carlo method that samples from complex distributions by iteratively using simpler conditional distributions [29]. It efficiently handles high-dimensional spaces, making it ideal for large-scale BNs [30], where exact methods are computationally infeasible. However, different from a single BN model for the radial network, the probabilistic graph network set $\mathcal{B}$ with multiple BNs is obtained here to be further inferred. A common approach for addressing this challenge is to apply the basic GS algorithm to perform inference for each $B_i \in \mathcal{B}, \forall i$ individually, and then combine the results to obtain the final outage inference. While this approach is theoretically feasible, it has two significant limitations. First, the outage inference results from different BNs may conflict due to the varying and incomplete information received by the different subnetworks, introducing challenges in resolving these conflicts. Moreover, while multi-source data fusion enhances outage location accuracy, it also increases the risk of misinformation from data collection errors, such as erroneous "last gasp" signals from smart meters or natural language processing inaccuracies in social media messages. Separate inference in each subnetwork exacerbates these issues, as the misinformation can propagate and lead to inconsistent or inaccurate conclusions that are difficult to reconcile. Second, in real-world applications, DNs are often partially observable, leading to limited information for outage location inference [2]. While there may be some overlap of customers among subnetworks, separately inferring these BN models risks underutilizing the available information, particularly in emergency scenarios such as large-scale outages caused by severe weather, where communication channels may be blocked or damaged [31]. To address these limitations, we propose the JGS method as the final step of the framework, designed to enhance the accuracy and reliability of outage location inference.

To enable simultaneous inference across the BNs, we designed and integrated a two-phase information forward-backward mechanism into the basic GS method, achieving joint sampling. The core idea of this mechanism is to facilitate the transfer of information between different subnetworks during the iterative process, thereby maximizing the utilization of limited data and ultimately enhancing inference accuracy. The first challenge in implementing this mechanism is determining the direction of information flow, specifically how to sequence the BNs from $\mathcal{B}$. The information collected from both the customer side and the grid side during ΔT after the outage trigger point is converted into evidence and serves as input for the outage location inference task. Due to variations among customers, the evidence received by each subnetwork may differ. Naturally, information should flow from the subnetwork with more evidence, and thus more information, to those with less. To quantify these differences in information, we define an EDR, denoted as r_e, which represents the level of information received by each subnetwork. This ratio can be expressed as follows:

$$r_e^{[i]} = \frac{\sum_{\kappa \in \eta, j=1}^{|\boldsymbol{C}^{[i]}|} z_{i,j}^{\kappa}}{|\boldsymbol{C}| \cdot |\eta|} \tag{4}$$

where $|\boldsymbol{C}^{[i]}|$ denotes the total customer number in subnetwork i; η represents the set of all evidence types, such as trouble calls, Twitter posts, and SM last gasp signals; and $|\eta|$ indicates the number of evidence types. The $r_e^{[i]}$ measures the proportion of evidence collected during the outage event relative to the theoretical maximum amount of evidence. Ideally, if all customers in the subnetwork experience an outage and successfully transmit last gasp signals through their SMs while also reporting the outage via phone and social media, the ratio would be 1. However, in practice, $r_e^{[i]}$ tends to be lower due to the restricted outage area, partial SM coverage and limited customer interaction during the outage. Additionally, $r_e^{[i]}$ will vary across subnetworks, reflecting differences in SM installation and customer feedback. Consequently, the BNs in $\mathcal{B}$ can be organized in descending order of $r_e^{[i]}$, forming a sequence $\mathcal{B}_{\text{seq}} = (B_{\sigma(1)}, \ldots, B_{\sigma(M)})$.

1) Information Forward Phase: Given the collected evidence $\boldsymbol{\mathcal{Z}}$ and the ordered BN set $\mathcal{B}_{\text{seq}}$, initial samples are randomly generated across all the unknown state variables $\{\boldsymbol{D}^{(0)}, \boldsymbol{C}^{(0)}\}$, and the initial state of the samples in each subnetwork can be initiated as $\{\boldsymbol{D}^{(0),[i]}, \boldsymbol{C}^{(0),[i]}\}, \forall i \in \{1, \ldots, M\}$. The sampling process begins with the first BN model, $B_{\sigma(m)}$ (where $m = 1$) in the sequence $\mathcal{B}_{\text{seq}}$, with a randomly chosen state variable $d_i^{[m]}$ designated as the starting point. In the $(k+1)_{th}$ iteration of the JGS, based on the structure of the $B_{\sigma(m)}$, the designated samples to the parents and children of $d_i^{[m]}$ are fed into the local Bayesian estimator. The conditional PDF of $d_i^{[m]}$ is then approximated based on the latest sample as shown in equation (5):

$$\begin{aligned} P_{\Phi}\Big(d_i \mid d_{-\boldsymbol{i}}^{\mathfrak{K}}\Big) \\ = \frac{P_{D_i|Pa(D_i)}(d_i \mid Pa(d_i))P_{Ch(D_i)|D_i}(Ch(d_i) \mid d_i)}{\sum_{d_i} P_{D_i|Pa(D_i)}(d_i \mid Pa(d_i))P_{Ch(D_i)|D_i}(Ch(d_i) \mid d_i)} \end{aligned} \tag{5}$$

here, $d_{-\boldsymbol{i}}^{\mathfrak{K}}$ represents the latest updated sample excluding d_i, where $\mathfrak{K}$ denotes the variable script $(k), [m]$ due to space constraints; Similar to the denotation of $Pa(\cdot)$, $Ch(d_i)$ represents the child nodes of d_i; and:

$$P_{D_i|Pa(D_i)}(d_i \mid Pa(d_i)) = P_{D_i|D_{i-1}, \boldsymbol{\mathcal{Z}}^{\{w,v,b\}}}\Big(d_i \mid d_{i-1}^{\mathfrak{K}}, z_i^w, z_i^v, z_i^b\Big)$$

$$\begin{aligned} &P_{Ch(D_i)|D_i}(Ch(d_i) \mid d_i) \\ &= P_{D_{i+1}|D_i, \boldsymbol{\mathcal{Z}}^{\{w,v,b\}}}\Big(d_{i+1}^{\mathfrak{K}} \mid d_i, e_{i+1}^w, e_{i+1}^v, e_{i+1}^b\Big) \prod_{j=1}^{\mu_i} P_{C_i^j|D_i}\Big(c_i^{j,\mathfrak{K}} \mid d_i\Big). \end{aligned}$$

Considering the $P_{\Phi}(d_i \mid d_{-\boldsymbol{i}}^{(k),[m]})$ represents the PDF over a single random variable $d_i^{[m]}$, it can be efficiently calculated. The new sample of $d_i^{(k),[m]}$ is then drawn from this PDF using the inverse transform method [28]. This sample is subsequently used to update the current state of $B_{\sigma(m+1)}$,[1] i.e., $d_i^{(k),[m+1]}$. The newly generated samples $d_i^{(k),[m]}$ contains the information from the evidence provided by the customers in $B_{\sigma(m^-)}$, for $m^- \leq m$. Updating the state of network $m+1$ represents

[1]Only the states of common variables between $B_{\sigma(m)}$ and $B_{\sigma(m+1)}$ are updated, considering subnetwork differences.

the flow forwards of data from the previous BNs with higher EDR, allowing for a better approximation of the PDF. This process of PDF calculation, sample generation, and data flow continues until the variable states of the final network $B_{\sigma(M)}$ are updated, completing the information forward step.

2) Information Backward Phase: While the forward phase facilitates the transfer of information from the first to the last ordered BN, it does not support backward information flow. This limitation is addressed in the backward phase. During the backward phase, information is transferred back to each BN by merging the approximated PDFs $P_{\Phi}^{[m]}$ generated during the forward process across all networks. To achieve this PDF combination, an information entropy-based merging method is employed within the probability fusion module, with its specific formulation provided in equation (6):

$$\bar{P_{\Phi}} = \sum_{m=1}^{M} P_{\phi}^{[m]} \left(\frac{1 - \boldsymbol{h}(D^{[m]})}{M - \sum_{i=1}^{M} \boldsymbol{h}(D^{[m]})} \alpha + \beta \right) \tag{6}$$

where, $\bar{P_{\phi}}$ denotes the weighted mixture distribution from the fusion operation; β represents the fixed weight, constrained by $\beta \leq \frac{1}{M}$, and $\alpha = 1 - M\beta$ denotes the dynamic component of the weights based on entropy; the $\boldsymbol{h}(D^{[m]})$ is the entropy-based metric shown as:

$$\boldsymbol{h}(D^{[m]}) = -\sum_{i=1}^{|\boldsymbol{D}|} P_{\Phi}\Big(d_i = 1 \mid d_{-i}^{\mathfrak{R}}\Big) \cdot \ln\Big(P_{\Phi}\Big(d_i = 1 \mid d_{-i}^{\mathfrak{R}}\Big)\Big). \tag{7}$$

In information theory, entropy quantifies the uncertainty or randomness within a dataset, measuring the unpredictability and information content of a message or data stream. Here, we design an entropy-based metric to assign appropriate weights to each subnetwork's results based on their respective levels of uncertainty. BNs with lower uncertainties, indicated by higher $\boldsymbol{h}(D^{[m]})$ values, are given greater weight in the probability fusion process. The fusion module dynamically assesses the reliability of inference results for each BN according to this metric. Subsequently, a new sample is drawn from the fused distribution $\bar{P_{\Phi}}$, which is then used as the initial state for $B_{\sigma(1)}$ in the next iteration. This marks the end of the backward phase, initiating another forward phase.

The information flow of the forward and backward phases is illustrated in Fig. 4. This two-phase JSP process is repeated until a specific number of random samples, e.g., K, are obtained for the unknown variables. The target conditional PDFs, $P_{D_i|\mathcal{Z}}(d_i \mid z)$ and $P_{C_i^j|\mathcal{Z}}(c_i^j \mid z)$, can then be approximated by tallying up the samples produced by the GS algorithm, a method that has been theoretically validated [28]. After completing the JSP iteration, the most probable values for each branch and customer state are determined based on the estimated conditional PDFs to resolve equations (2) and (3). Since the state variables are binary, a threshold value τ is applied to determine whether a branch i is energized. Once the states of all branches and customers are obtained, the location of the outage events can be easily identified. The detailed procedure for outage location inference using JGS is presented in **Algorithm 2**.

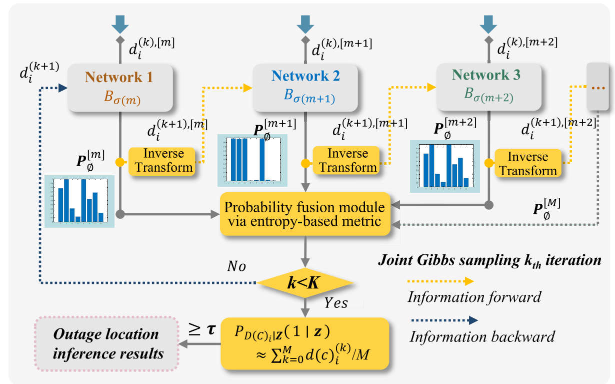


Fig. 4. Illustration of the joint Gibbs sampling mechanism, demonstrated with three subnetworks. Additional networks can be integrated similarly.

Algorithm 2 Joint Gibbs Sampling Algorithm

Require: Networks G^s, BN set $\mathcal{B} = \{B_m \mid m = 1, \ldots M\}$, iteration number K, evidence set $\mathcal{Z}$, cutoff threshold τ

initialize samples $x^{(0)} \leftarrow \{d_i^{(0)}, c_i^{j,(0)} \mid \forall i, j\}$ randomly generated from Binomial distribution of order 2; then, update $x^{(0)} \leftarrow x^{(0)} \cup \mathcal{Z}$.

order the BN set as sequence $\mathcal{B}_{\text{seq}} = (B_{\sigma(1)}, \ldots, B_{\sigma(M)})$, ensuring $r_e(G^s_{\sigma(m-1)}) \leq r_e(G^s_{\sigma(m)})$ for all $m \leq M$.

for $k = 0, \ldots, K$ **do**
  for $m = 1, \ldots, M$ **do**
    $x^{(k),[m]} \leftarrow x^{(k)}$
    execute procedure $P_{\Phi}^{[m]} \leftarrow \boldsymbol{GS}(B_{\sigma(m)}, x^{(k),[m]})$
    draw $x^{(k),[m]}$ from $P_{\Phi}^{[m]}$, $x^{(k),[m]} \leftarrow F_x^{-1}(P_{\Phi}^{[m]})$
  end for
  $\bar{P_{\Phi}} \leftarrow \Psi(P_{\Phi}^{[1]}, \ldots, P_{\Phi}^{[M]})$ ▷ $\Psi(\cdot)$ Prob. fusion module
  $x^{(k+1)} \leftarrow F_x^{-1}(\bar{P_{\Phi}})$ ▷ $F_x^{-1}(\cdot)$ Inverse transform
end for

procedure $\boldsymbol{GS}(B_{\sigma(m)}, \boldsymbol{x})$
  for $i = 1, \ldots, |D + C|$ **do**
    select one random variable x_i
    update $\boldsymbol{x_{-i}} \leftarrow \boldsymbol{x} \setminus x_i$
    $P_{\Phi} \leftarrow \frac{P_{X_i|Pa(X_i)}(x_i|Pa(x_i))P_{Ch(X_i)|X_i}(Ch(x_i)|x_i)}{\sum_{x_i} P_{X_i|Pa(X_i)}(x_i|Pa(x_i))P_{Ch(X_i)|X_i}(Ch(x_i)|x_i)}$
  end for
  return
end procedure

return P_{Φ}

$P_{D_i|\mathcal{Z}}(d_i = 1 \mid z) \leftarrow \frac{\sum_{k=0}^{K} d_i^{(k)}}{K}, \forall i$

$P_{C_i^j|\mathcal{Z}}(c_i^j = 1 \mid z) \leftarrow \frac{\sum_{k=0}^{K} c_i^{j,(k)}}{K}, \forall i, j$

inference state by τ: $d_i \leftarrow 1, \forall i$ if $P_{D_i|\mathcal{Z}}(d_i = 1 \mid z) \geq \tau$; τ: $c_i^j \leftarrow 1, \forall i, j$ if $P_{C_i^j|\mathcal{Z}}(c_i^j = 1 \mid z) \geq \tau$

return $c_i^j, d_i \forall i, j$

V. Numerical Results

To evaluate the practical performance of the proposed model, a series of numerical case studies are carried out in this section. The two testing systems employed are derived from a widely utilized IEEE 123-bus DN testing system [32], [33] and a real-world 51 bus distribution feeder [34], referred to as the IEEE system and Real system, respectively. Both systems

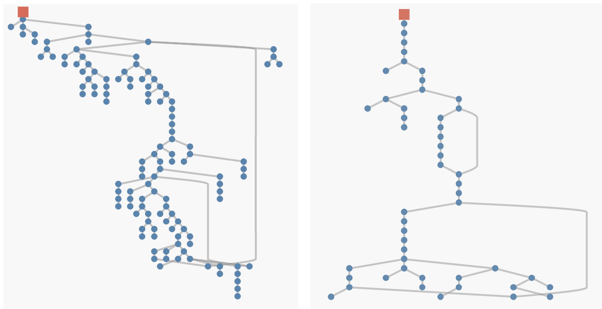


Fig. 5. Topological information of the testing systems: modified IEEE 123 bus model (left) with two loops and modified real 51 bus system (right) with two loops.

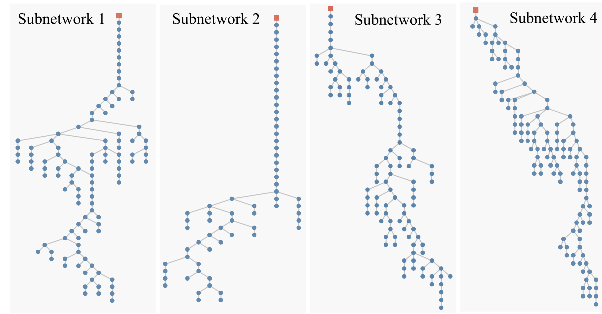


Fig. 6. Topology of subnetworks from modified IEEE 123 bus model by network reconstitution.

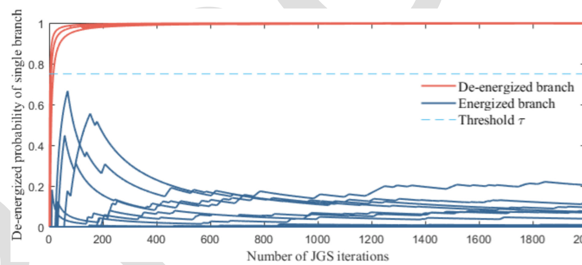


Fig. 7. Iteration of joint Gibbs sampling process. This case is illustrated by the outage in the IEEE system with 75% observability.

are publicly available online. To create weakly meshed DNs, two loops were added to each testing system, with detailed topology information presented in Fig. 5. To simulate partially observable scenarios akin to real-world applications, we define three levels of observability, i.e., 25%, 50%, and 75%, which correspond to the percentage of nodes equipped with SMs. A total of 1,000 outage scenarios were generated using the Monte Carlo method, with outage locations and SM locations selected randomly, making the testing results not influenced by any specific scenario setting. Among these scenarios, 500 involve a single outage block, 300 involve two outage areas, and 200 involve more than two outage areas. Typically, one or two outage areas occur simultaneously in normal weather, covering most scenarios. However, we added the scenarios with over two outage areas to account for severe events during extreme weather. For each scenario, evidence information was generated based on the outage location and system observability. Ideally, SMs are designed to transmit last gasp signals immediately following an outage. However, considering the reliability of AMI devices and the associated communication infrastructure, only a portion of these signals is ultimately received by utilities for outage localization. Based on historical data, we set the signal collection ratio at 82%. Additionally, customer trouble calls and social media messages are assumed to be collected within ΔT (e.g., 15 minutes post-outage). In the real application, this time can be adjustable according to the customer reporting time tallied up from outage reports. Following the approach discussed in [21], human-related evidence is modeled using an exponential PDF with time ΔT. To prevent information leakage from evidence generation to outage inference, parameters substantially different from the true values were chosen, reflecting the reality that actual parameters are unknown. Furthermore, during real outages, customers may report issues unrelated to the outage or request other services, resulting in misinformation in trouble calls. Errors in natural language processing and communication failures may also reduce the accuracy of the evidence. To account for these factors, we introduce around 10%, 10%, and 5% erroneous information into the generated evidence.

A. Network Reconstitution

Using the DN reconstitution method, the original looped networks are transformed into multiple radial subnetworks. For both the IEEE system and Real system, four subnetworks are generated, consistent with $\max(|L_i|)$ for each system, achieving the minimum number of subnetworks. The topology of the four subnetworks for the IEEE system is illustrated in Fig. 6. As we can see, not all load nodes are included in each subnetwork; Subnetwork 4 contains the most nodes, matching the original network, while Subnetworks 1 to 3 include only partial load nodes–107, 65, and 109, respectively. This disparity in node numbers results in varying levels of information. Based on the EDR, the subnetworks are then ranked accordingly for BN construction.

B. Performance of the Multisource Data Fusion Framework

After constructing BNs for each subnetwork, the obtained $\mathcal{B}$ is inferred using the designed JPS method. An example of an iteration process for the JBS method on the IEEE system with 75% observability is illustrated in Fig. 7. As observed, for branches in outage-affected areas (marked in red), the probabilities converge to significantly higher values compared to unaffected branches (marked in blue). By applying the threshold τ, de-energized branches can be identified, allowing for accurate outage location detection. This demonstrates the effectiveness of the proposed framework. To further assess the framework's performance under different scenarios and various outage events, four commonly used metrics, i.e., accuracy, precision, recall, and F1-score, are employed to present the branch level (Br-Le) accuracy. Detailed formulations of these metrics can be found in [21]. Besides the four metrics, system level (Sys-Le) accuracy is defined as the ratio between the fully identified case number to the total outage events number, to measure the system level performance.

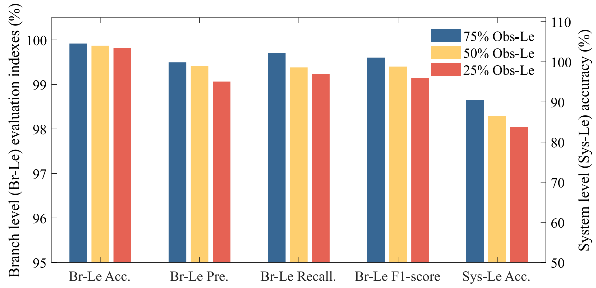


Fig. 8. Performance of the proposed framework under various evidence scenarios on the IEEE system.

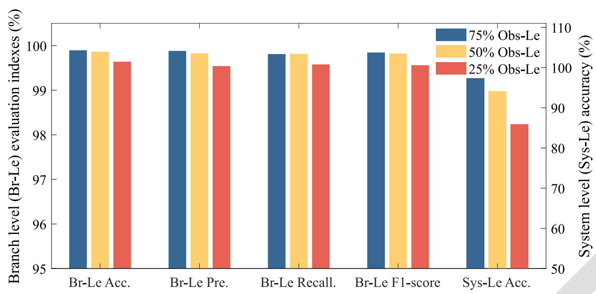
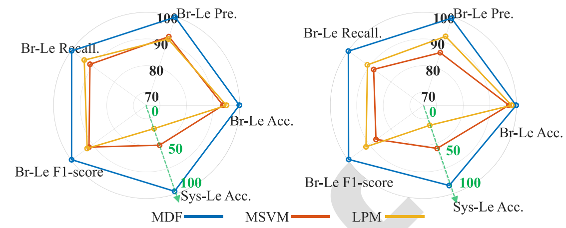
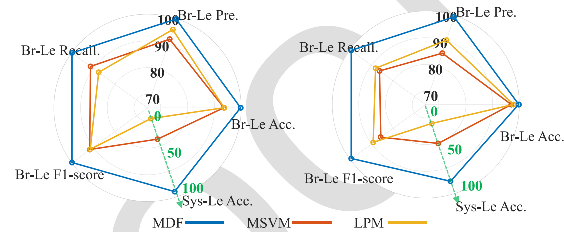


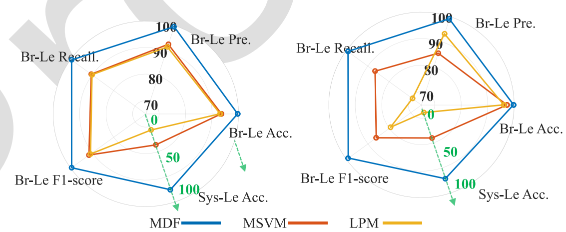
Fig. 9. Performance of the proposed framework under various evidence scenarios on the Real system.



(a) performance of the three models in 75% Obs-Le scenarios



(b) performance of the three models in 50% Obs-Le scenarios



(c) performance of the three models in 25% Obs-Le scenarios

Fig. 10. Comparative results of the three models on two testing systems across various observability levels. The green axis representing system accuracy follows a different scale, ranging from 0% to 100%.

Fig. 8 and 9 illustrate the five performance indicators under varying levels of observability (Obs-Le) for both testing systems. The results indicate that the proposed outage location framework delivers exceptional performance, with the four primary classification metrics exceeding 98% in all three scenarios. Although Sys-Le accuracy is somewhat lower due to its more rigorous criteria, the worst-case scenario, with 25% observability in the IEEE system, still achieves over 85%. These findings showcase that while reduced observability does affect accuracy, the framework remains highly effective in low-Obs-Le situations, maintaining strong overall performance.

C. Outage Location Model Comparison

To further assess the performance of the proposed MDF framework, a comparative analysis was conducted against two previously established methods: the SVM-based approach [11] and the LPM [1]. The SVM-based method applies a multi-label SVM (MSVM) classification scheme to identify line outages using SM data. The LPM aims to approximate the global posterior probability of the line outages by linearly combining local posterior probabilities from multiple data sources. Consistent with prior work, a Br-Le evaluation was used to ensure a fair comparison. The results, displayed in Fig. 10, illustrate the performance of the three models across various observability levels on two test systems.

As shown, the proposed framework consistently outperforms the other models in all scenarios, with Sys-Le accuracy exhibiting the most significant differences. LPM shows the lowest accuracy, as it neglects the dependencies between system components, making it vulnerable to misinformation and limited evidence. Despite using multiple data sources, its performance lags behind that of the proposed framework. On the other hand, the MSVM model captures the relationship between branch status and evidence data, but its reliance on single-source meter data poses challenges, especially at low observability levels. This is reflected in the decline in performance metrics as observability level decreases. In contrast, the proposed MDF framework fully accounts for the interdependencies between system components and effectively integrates data from both metered and non-metered sources. This comprehensive approach enables more accurate and stable outage location results, regardless of the system Obs-Le.

D. Sensitivity Analysis of Prior Information Bias

In field applications of the framework, certain prior information, such as the SM last gasp collection ratio and customer trouble call report ratio, can not be directly available. While estimated values can be derived from historical data, errors are inevitable. To evaluate the impact of deviations in prior information, a sensitivity analysis was conducted to assess the model's performance under varying levels of parameter bias. Using the previously described outage scenarios, prior probabilities were perturbed with error levels of 10%, 20%, and 30%. The resulting outage location accuracy across different observabilities and testing systems is summarized in Fig. 11. The results indicate that while system-level accuracies decrease with increasing error levels, the overall accuracy

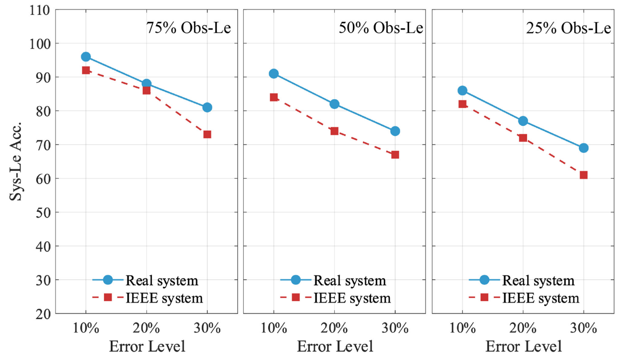


Fig. 11. Results of the sensitivity analysis on the BNs' prior information bias. The Sys-Le Acc. for two testing systems under different observabilities is presented. Br-Le accuracies are not displayed due to space limitations but remain consistently above 90%.

remains within an acceptable range. At the 30% error level, the Sys-Le Acc. experiences a more prominent decline than the 10% error level due to the inconsistency of the prior parameters. However, other Br-Le accuracy metrics remain above 90%, showcasing satisfactory performance. It is important to note that, although the case study intentionally introduced significant error levels, the accuracy of prior information is expected to improve over time as utilities accumulate more outage records and related information. This enhancement in prior information will ensure the framework's reliability in practical applications.

E. Framework Computational Complexity

A standard PC equipped with an Intel Xeon E-2224 CPU (3.40 GHz) and 16.0 GB of RAM was used to perform a comprehensive computational complexity analysis. Both the IEEE system and a real-world system were analyzed to evaluate performance. For the distribution network reconstitution task, the average computation time across 20 repetitions was 0.765 seconds for the real-world system and 2.174 seconds for the IEEE system, indicating slightly higher computational requirements for the latter. In the outage location inference step, the average computation times were 74.53 seconds and 107.67 seconds for the real-world and IEEE systems, respectively. Notably, as the proposed outage location framework operates the feeder-wise application, parallel computation of different feeders can mitigate the computational impact of large feeder numbers, enhancing its practicality for real-world distribution networks.

VI. Conclusion

This paper proposed an integrated multisource data fusion framework for outage location detection using a probabilistic graph network. Specifically, a DN reconstitution method was developed to manage DNs with loops by converting the original looped networks into multiple subnetworks. These subnetworks capture all outage causalities in the original network and serve as a foundational step. By embedding multiple sources of evidence and subnetwork structures, BN models were established for each subnetwork. To maximize the use of limited evidence, the JGP mechanism was designed to enable interactive inference among the BN models, ultimately producing the outage location results. The framework was validated through simulations on two testing systems, and a comparative study with prior works confirmed its effectiveness in identifying outage locations in DNs with loops. In future work, we plan to explore and integrate a physics-embedded module that incorporates system protection mechanisms into the framework to enhance its accuracy and efficiency. Additionally, leveraging the proposed framework, we aim to investigate methods for simultaneously addressing outage location and network configuration challenges in radial distribution systems.

Acknowledgment

The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting this article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work or allow others to do so, for U.S. Government purposes.

References

[1] C. Chen, J. Wang, and D. Ton, "Modernizing distribution system restoration to achieve grid resiliency against extreme weather events: An integrated solution," *Proc. IEEE*, vol. 105, no. 7, pp. 1267–1288, Jul. 2017.

[2] Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "Outage detection in partially observable distribution systems using smart meters and generative adversarial networks," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5418–5430, Nov. 2020.

[3] R. A. Jacob, S. Paul, S. Chowdhury, Y. R. Gel, and J. Zhang, "Real-time outage management in active distribution networks using reinforcement learning over graphs," *Nat. Commun.*, vol. 15, no. 1, p. 4766, 2024.

[4] W. R. Cassel, "Distribution management systems: Functions and payback," *IEEE Trans. Power Syst.*, vol. 8, no. 3, pp. 796–801, Aug. 1993.

[5] M. U. Sumagayan, C. Premachandra, R. B. Mangorsi, C. J. Salaan, H. W. H. Premachandra, and H. Kawanaka, "Detecting power lines using point instance network for distribution line inspection," *IEEE Access*, vol. 9, pp. 107998–108008, 2021.

[6] W. Luan and W. Li, "Smart metering and infrastructure," in *Smart Grids*. Boca Raton, FL, USA: CRC Press, 2017, pp. 399–420.

[7] A. Gandluru, S. Poudel, and A. Dubey, "Joint estimation of operational topology and outages for unbalanced power distribution systems," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 605–617, Jan. 2020.

[8] T. A. Short, *Electric Power Distribution Equipment and Systems*. Boca Raton, FL, USA: CRC Press, 2018.

[9] H.-C. Seo, "New protection scheme in loop distribution system with distributed generation," *Energies*, vol. 13, no. 22, p. 5897, 2020.

[10] Z. Wang, B. Chen, J. Wang, and C. Chen, "Networked microgrids for self-healing power systems," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 310–319, Jan. 2016.

[11] Z. S. Hosseini, M. Mahoor, and A. Khodaei, "AMI-enabled distribution network line outage identification via multi-label SVM," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5470–5472, Sep. 2018.

[12] K. Sridharan and N. N. Schulz, "Outage management through AMR systems using an intelligent data filter," *IEEE Trans. Power Del.*, vol. 16, no. 4, pp. 669–675, Oct. 2001.

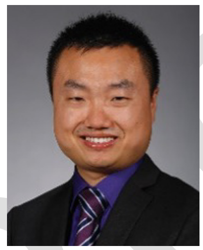
[13] Y. Liao, Y. Weng, C.-W. Tan, and R. Rajagopal, "Urban distribution grid line outage identification," in *Proc. Int. Conf. Probab. Methods Appl. Power Syst. (PMAPS)*, 2016, pp. 1–8.

[14] S. Bajagain and A. Dubey, "Spectral clustering for fast outage detection and visualization in power distribution systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, 2022, pp. 1–5.

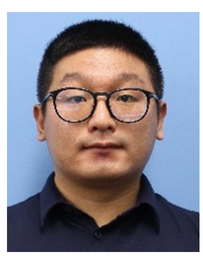
[15] H. Sun, Z. Wang, J. Wang, Z. Huang, N. Carrington, and J. Liao, "Data-driven power outage detection by social sensors," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2516–2524, Sep. 2016.
[16] P. Kankanala, S. Das, and A. Pahwa, "AdaBoost $^{+}$: An ensemble learning approach for estimating weather-related outages in distribution systems," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 359–367, Jan. 2014.
[17] R. A. Sevlian, Y. Zhao, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Outage detection using load and line flow measurements in power distribution systems," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2053–2069, Mar. 2018.
[18] Z. Soltani, S. Ma, M. Khorsand, and V. Vittal, "Simultaneous robust state estimation, topology error processing, and outage detection for unbalanced distribution systems," *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2018–2034, May 2023.
[19] Y. Liu and N. N. Schulz, "Knowledge-based system for distribution system outage locating using comprehensive information," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 451–456, May 2002.
[20] J. He, S. Lei, and L. Wu, "Multi-source data fusion for power outage warning based on transformer model," *J. Phys., Conf. Ser.*, vol. 2816, no. 1, 2024, Art. no. 12030.
[21] Y. Yuan, K. Dehghanpour, Z. Wang, and F. Bu, "Multisource data fusion outage location in distribution systems via probabilistic graphical models," *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1357–1371, Mar. 2022.
[22] A. Bahmanyar, S. Jamali, A. Estebsari, and E. Bompard, "A comparison framework for distribution system outage and fault location methods," *Electric Power Syst. Res.*, vol. 145, pp. 19–34, Apr. 2017.
[23] S. Grafelman and F. Small. "Smart meters are made smart by people and processes." 2015. [Online]. Available: https://electricenergyonline.com/energy/magazine/900/article/Smart-Meters-Are-Made-Smart-by-People-and-Processes.htm
[24] Y. Zhang, J. Wang, and M. E. Khodayar, "Graph-based faulted line identification using micro-PMU data in distribution systems," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 3982–3992, Sep. 2020.
[25] A. Raghavan et al., "Low-cost embedded optical sensing systems for distribution transformer monitoring," *IEEE Trans. Power Del.*, vol. 36, no. 2, pp. 1007–1014, Apr. 2021.
[26] I. Džafić and R. A. Jabr, "Real time multiphase state estimation in weakly meshed distribution networks with distributed generation," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4560–4569, Nov. 2017.
[27] G. Celli, F. Pilo, G. Pisano, V. Allegranza, R. Cicoria, and A. Iaria, "Meshed vs. radial MV distribution network in presence of large amount of DG," in *Proc. IEEE PES Power Syst. Conf. Expo.*, 2004, pp. 709–714.
[28] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT press, 2009.
[29] G. Casella and E. I. George, "Explaining the Gibbs sampler," *Amer. Stat.*, vol. 46, no. 3, pp. 167–174, 1992.
[30] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Boca Raton, FL, USA: Chapman Hall/CRC, 1995.
[31] Z. Bie, Y. Lin, G. Li, and F. Li, "Battling the extreme: A study on the power system resilience," *Proc. IEEE*, vol. 105, no. 7, pp. 1253–1266, Jul. 2017.
[32] N. Shi, R. Cheng, L. Liu, Z. Wang, Q. Zhang, and M. J. Reno, "Data-driven affinely adjustable robust volt/VAr control," *IEEE Trans. Smart Grid*, vol. 15, no. 1, pp. 247–259, Jan. 2024.
[33] Z. Wang and J. Wang, "Self-healing resilient distribution systems based on sectionalization into microgrids," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3139–3149, Nov. 2015.
[34] F. Bu, Y. Yuan, Z. Wang, K. Dehghanpour, and A. Kimber, "A time-series distribution test system based on real utility data," in *Proc. North Amer. Power Symp. (NAPS)*, 2019, pp. 1–6.

Liming Liu (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from North China Electric Power University in 2016 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Iowa State University. His research interests encompass power distribution systems, distribution system modeling, and the application of optimization and machine learning techniques to power systems.

Yuxuan Yuan (Member, IEEE) received the B.S. and Ph.D. degrees in electrical and computer engineering from Iowa State University, Ames, IA, USA, in 2017 and 2022, respectively. He is currently a Data Analysis Engineer with Electric Power Engineers. He is the technical lead for a multitude of projects funded by multiple major U.S. utilities, the Department of Energy, and NYSERDA. He has extensive experience in power distribution systems specializing in power data analysis, real-time event identification, integrated grid planning, renewable energy and EV integration, and distribution system state estimation using advanced AI techniques. He was the recipient of the 2023 Outstanding Doctoral Dissertation from IEEE Power and Energy Society.



Zhaoyu Wang (Senior Member, IEEE) received the B.S. and first M.S. degrees in electrical engineering from Shanghai Jiao Tong University, and the second M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology. Since 2015, he has been an Assistant, an Associate, and a Full Professor with Iowa State University. His research interests include optimization and data analytics in power distribution systems and microgrids. He was the recipient of the National Science Foundation CAREER Award, the IEEE Power and Energy Society Outstanding Young Engineer Award, the Northrop Grumman Endowment, College of Engineering's Early Achievement in Research Award, and the Harpole-Pentair Young Faculty Award Endowment. He is the lead Principal Investigator for over $24M projects funded by the National Science Foundation, the Department of Energy, National Laboratories, PSERC, and Iowa Economic Development Authority. He is an Associate Editor of IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY, IEEE POWER ENGINEERING LETTERS, and *IET Smart Grid*. He was an Associate Editor of IEEE TRANSACTIONS ON POWER SYSTEMS and IEEE TRANSACTIONS ON SMART GRID, and the Chair of PSOPE Awards Subcommittee. He is the Secretary and a Technical Committee Program Chair of IEEE Power System Operation, Planning and Economics Committee, the Vice Chair of IEEE Distribution System Operation and Planning Subcommittee, the Secretary of IEEE Task Force on IEEE P3102 Standard for Conservation Voltage Reduction Data Collection and Management Procedures, and the Vice Chair of IEEE Task Force on Advances in Natural Disaster Mitigation Methods.



Yiyun Yao (Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2015 and 2019, respectively. He is currently a Researcher with the National Renewable Energy Laboratory, Golden, CO, USA, where he focuses on power system resilience, distributed energy resource integration, and optimization techniques for grid operations. He has authored many publications in leading IEEE journals and conferences. He is a recipient of multiple Best Paper Awards and the outstanding reviewer awards from IEEE.

Fei Ding (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Case Western Reserve University. She is a Distinguished Member of Research Staff and the Manager of the Grid Automation and Controls Group, National Renewable Energy Laboratory (NREL). She leads NREL's Autonomous Energy Systems Strategic Public and Private Partnership Program. She has extensive research experience in DER grid integration, grid resilience, and distribution automation. As a PI/Co-PI, she has led over 15 research projects funded by different funding agencies to develop and demonstrate advanced DER technologies to achieve autonomous energy systems and enhance grid reliability, resilience, and cybersecurity.