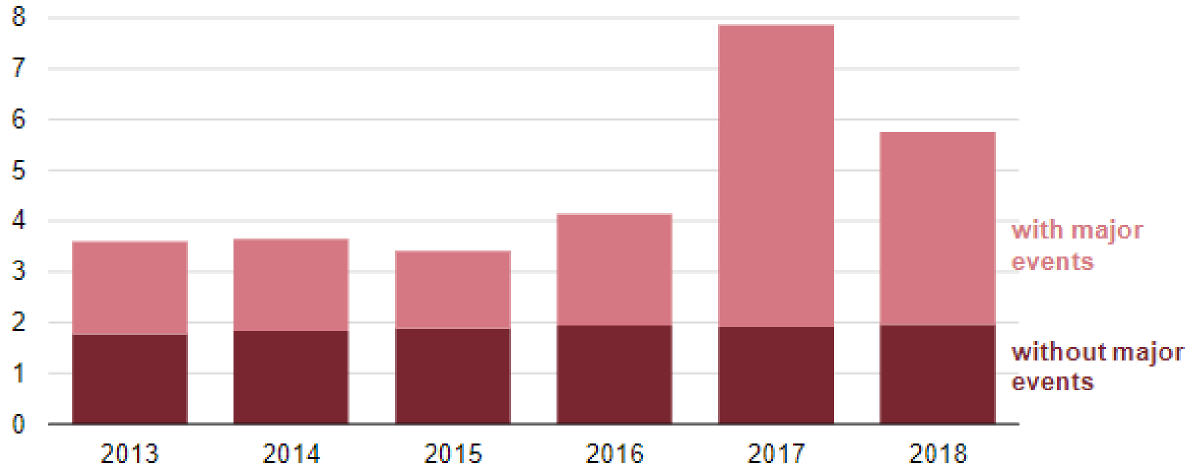
Multi-Source Data Driven Outage Detection in Distribution Systems for Decision Support using Probabilistic Graphical Models

Zhaoyu Wang
Northrop Grumman Associate Professor
Iowa State University

Motivation of Data-Driven Outage Location

- Outages can lead to a sharp decline in grid resilience with significant socio-economic losses, which cost an average of about *$18 billion to $33 billion per year* in the U.S.
- Outage detection and location is the first task after service disruptions.
- Most utilities still rely on customer reports to track outages, which can cause waste of up to *80% of the invaluable restoration time*. Hence, effective outage detection and location methods are critical to reduce outage duration.

In recent years, customers experienced longer outages. In 2018, each customer lost power for around **5.8 hours**.



1.9 million customers in Midwest were affected by 1.4 million outages between August 10 and 13.

Source: https://www.eia.gov/ https://poweroutage.us

Outage Data Sources

- While SCADA reports main feeder outages, there are multiple data sources that report lateral and grid-edge outages:

Data Source	Pros	Cons
SM last gasp signal	High accuracy, fast	Limited sensor coverage communication failures
Social media data	Generally available	Misreports, unreliable
Customer trouble call	Generally available	Low report rate, misreports
Weather data	Generally available	Lack of detailed location information

- Combing SMs with conventional outage data sources is an ideal solution for outage detection, but it is difficult to achieve because:
  - Heterogeneous characteristics: accuracy levels and reporting rates
  - Partially observable grids with **limited** sensors
  - May provide conflicting or mis-information

Combining Multiple Data Sources

- This combination means integrating evidences from different data sources as well as different customers:

$$P_{\boldsymbol{D},\boldsymbol{C}|E}(\boldsymbol{d},\boldsymbol{c}|\boldsymbol{e}) \tag{1}$$

  - ❖ ***D*** and ***C*** represent the states of primary network branches and the connections of customers
  - ❖ ***E*** is the multi-source evidence set (i.e., trouble call, last gasp signal)
  - ❖ Uppercase: random and evidence variables; lowercase: realization of variables

- Existing methods to solve Eq. 1:
  - Directly solving Eq. 1 using brute-force search over all possible combinations of branch/customer state ([1],[2])
  - ➢ **Limitation**: computationally infeasible for large systems
  - Assuming *full independency* among all data sources ([3]-[5])
  - ➢ **Limitation**: outage data sources and branches/customers are interdependent

Outage Location via Probabilistic Graph Learning

✓Leveraged the conditional independence inherent in distribution grids (rather than the assumed full independence in pooling methods) to encode the distribution network and its data into probabilistic graphs, i.e., ***Bayesian networks*** (BN) [6]

- ➢Node: states of branches/customers and outage data sources. Edge: probabilistic influence of one node on another.
- ➢For example, if the utility knows that a customer is in outage, probabilities of receiving SM last gasp signals and trouble calls from that customer will be uncorrelated.

Advantages:

- ✓ Accurately decompose and efficiently compute Eq. 1
- ✓ Address the problem of insufficient evidences, i.e., low SM coverage or low customer report rates
- ✓ Adaptable to newly added data sources

Single-source methods	Pooling models	Brute-force search	Proposed probabilistic graph
$P_{D,C\|E}(d,c\|e_1)$	$\prod P_{D_i\|E}(d_i\|e)\prod P_{C_j\|E}(c_j\|e)$	$P_{D,C,E}(\{d_i\}_{i=1}^{n},\{c_j\}_{j=1}^{m},e)$	$\frac{\prod P_{D_i\|D_{i-1}}(d_i\|d_{i-1})\prod P_{C_k\|D_i}(c_k\|d_i)\prod P_{E_{k,z}\|C_k}(e_{k,z}\|c_k)}{\sum_i\sum_k\{\prod P_{D_i\|D_{i-1}}(d_i\|d_{i-1})\prod P_{C_k\|D_i}(c_k\|d_i)\prod P_{E_{k,z}\|C_k}(e_{k,z}\|c_k)\}}$ (2)

Encoding Distribution Grids and Data into Probabilistic Graphs

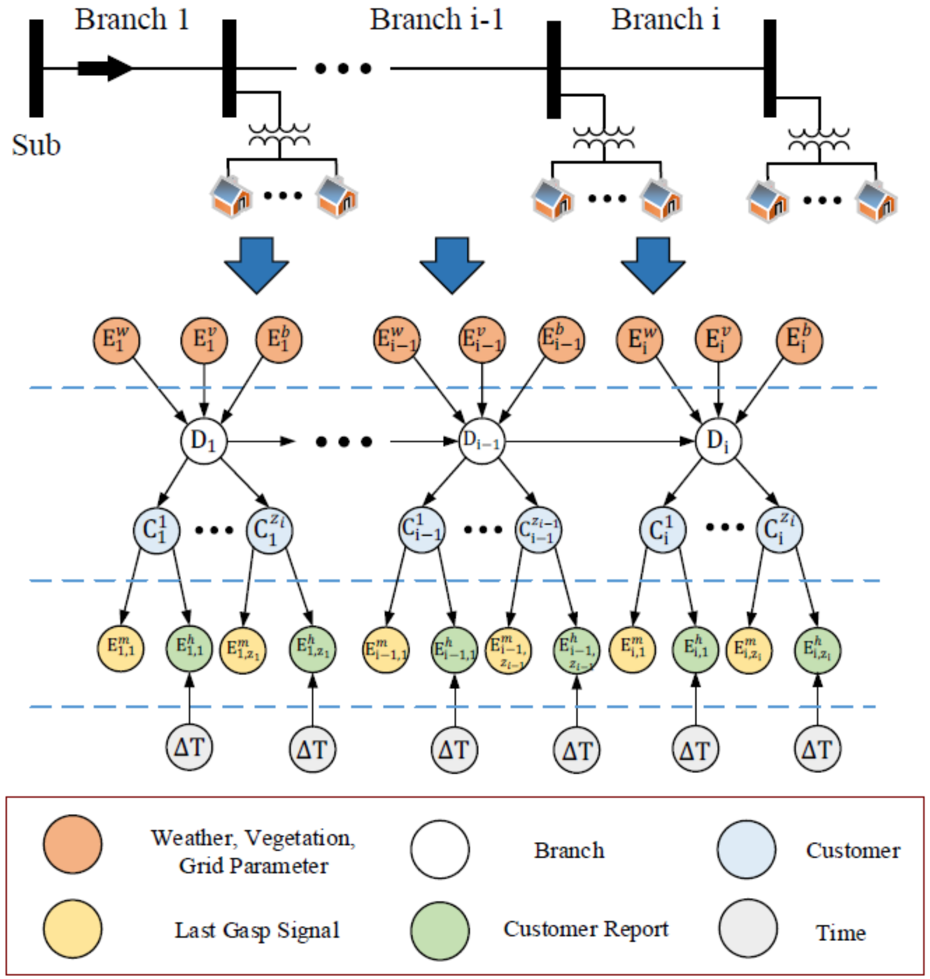
✓Following these conditional independencies, customers can be modeled as parent nodes for outage data sources in the graph. Putting together these features, a simple directed graph for a radial system can be constructed, as shown in the right figure.

✓D_i, C_i^j: states of branches/customers

✓ $E_i^w, E_i^v, E_i^b, E_{i,j}^m, E_{i,j}^h, \Delta T$: outage data sources

✓Parent variable (variable at the end of the arrow) : the immediate causal source of influence for its child variables (variables pointed by the arrow).

✓If the values of the parent variables are known, then the child variable becomes conditionally independent of variables that do not directly influence it.



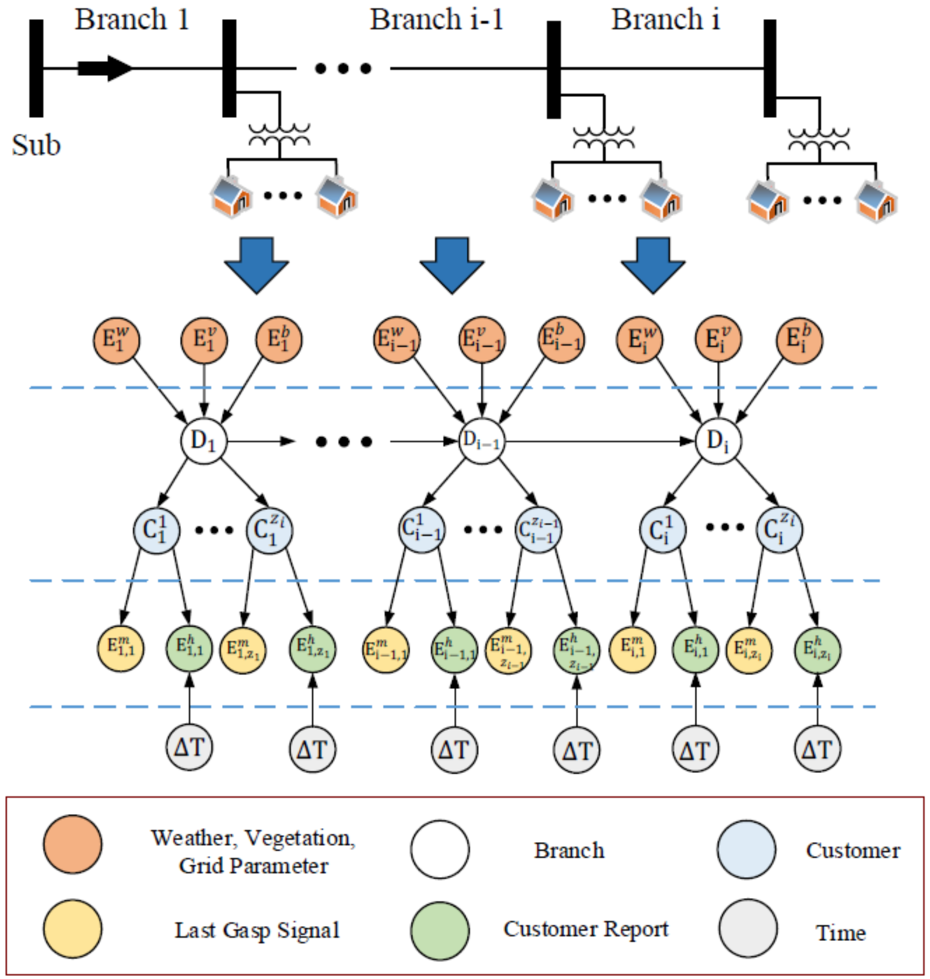
Encoding Distribution Grids and Data into Probabilistic Graphs

Salient Features of this Graphical Method:

✓It seamlessly integrates heterogeneous data sources. Different accuracy levels and reporting rates of various sources can be captured by conditional probabilities.

✓It is scalable and adaptive, as new data sources can be directly connected to their parent nodes in the graph without the need to re-learn the structure from scratch. The graph structure can be easily changed if there is a change in network topology.

✓It is robust with respect to misreports and inconsistencies in outage evidences, as uncertainty of each data source is explicitly modeled using graph parameters.

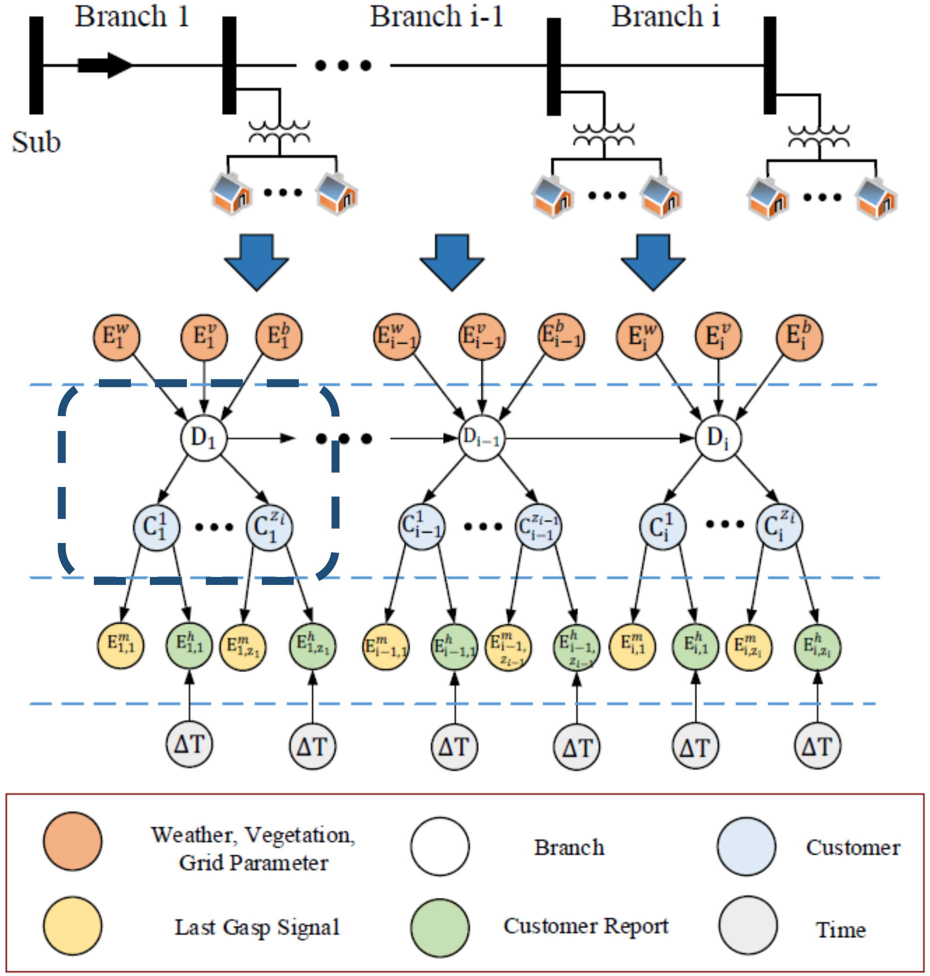


Offline Parameter Learning and Online Inference in Graphical Models

Each edge in the constructed graph has a *parameter* that quantifies the probabilistic relationship between its parent and child nodes. A parameter is the conditional probability of *effect* A, given the value of *cause* B, represented as $P_{A|B}(a|b)$.

(1) $\boldsymbol{P_{c_j^i|D_i}(c_j^i|d_i)}$: the chance of de-energization of customer if the state of its parent branch is known

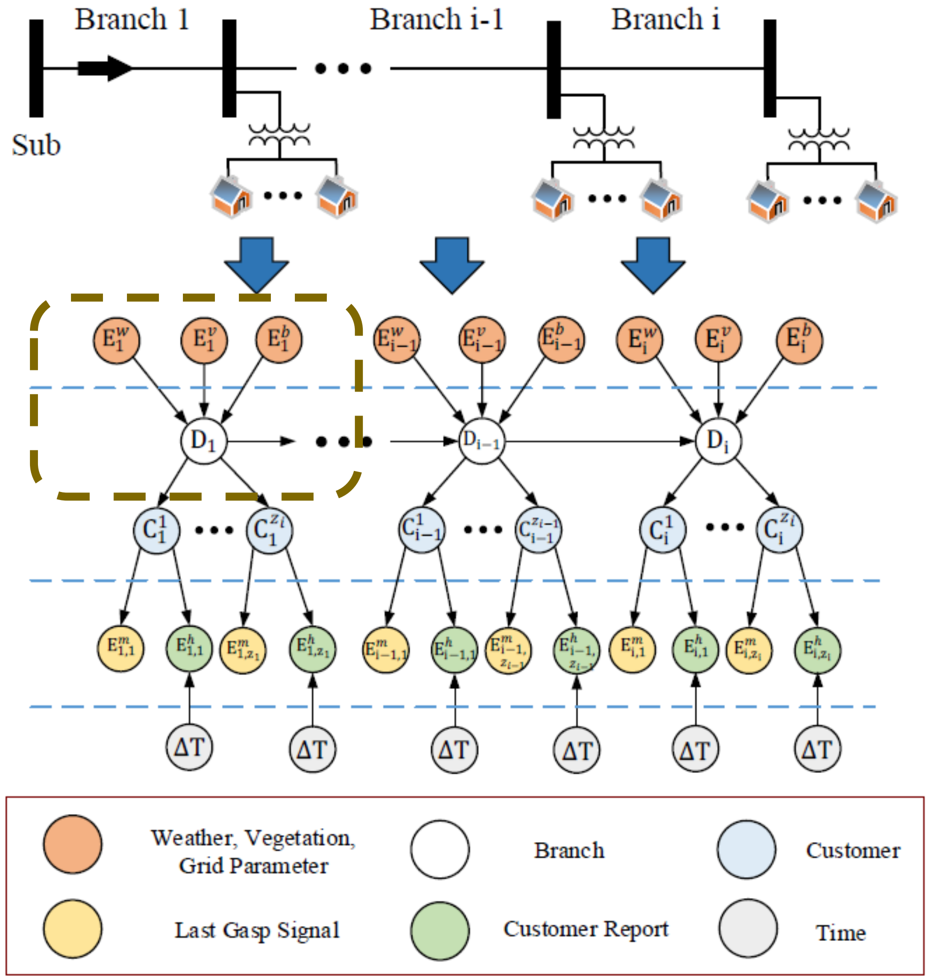
- If parent branch is outaged, then customer is certainly outaged.
- If the branch is energized, the customer could still be outaged as a result of the customer's own failures, regardless of the states of the neighbor customers.
- The parameter is learned empirically from historic data.



Offline Parameter Learning and Online Inference in Graphical Models

(2) $P_{D_i|D_{i-1},E_i^w,E_i^v,E_i^b}(d_i|d_{i-1},e_i^w,e_i^v,e_i^b)$: the chance of de-energization for a child branch if the state of its parent branch is known

- If the feeder is interrupted at any arbitrary node before node i, we can automatically conclude that D_i= 1, regardless of the values of the other variables.
- If the parent branch is energized, the child branch may still be de-energized due to the branch's own failure (i.e., P_i^f).
- To deal with data scarcity, a fragility model is utilized to estimate P_i^f based on wind speed $\boldsymbol{e_i^w}$, vegetation information $\boldsymbol{e_i^v}$, and grid parameters $\boldsymbol{e_i^b}$ [7].
- **Fragility mode**: a series model with the fragility analysis of each pole and conductor within the individual branch.



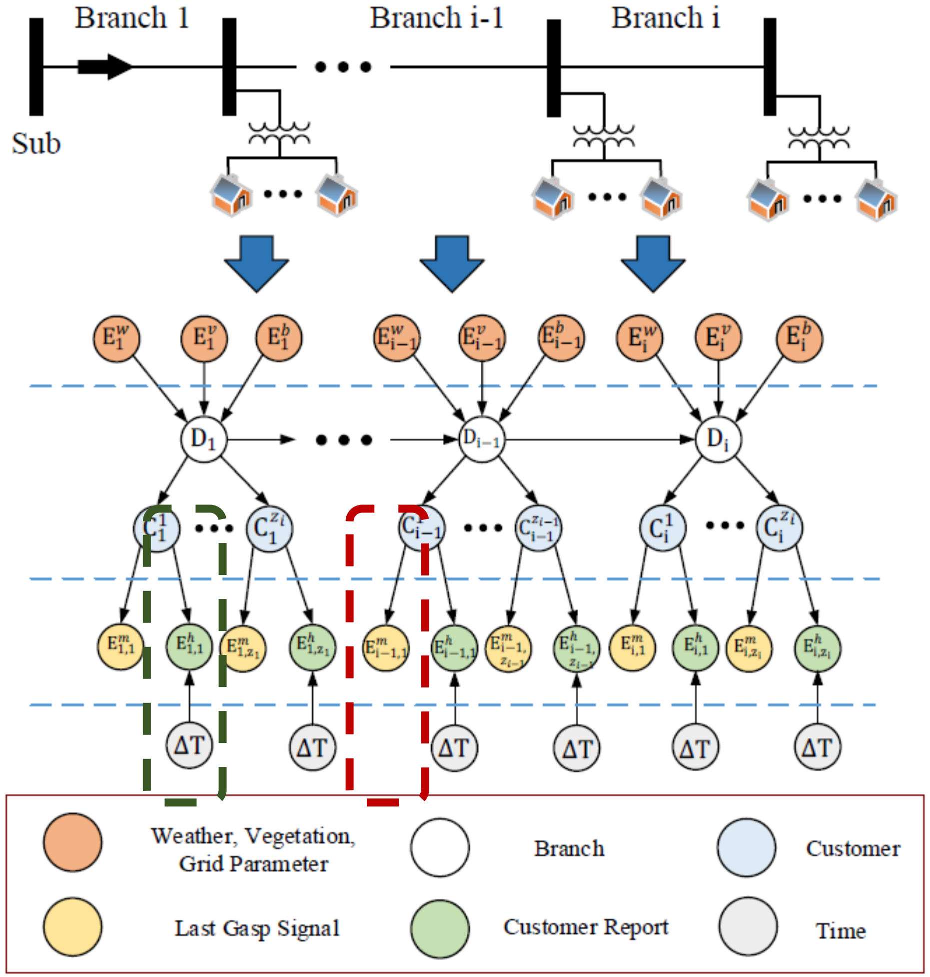
Offline Parameter Learning and Online Inference in Graphical Models

(3) $P_{E^h_{i,j}|C^j_i,\Delta T}\left(e^h_{i,j}\middle|c^j_i,\Delta t\right)$: the probability of receiving human-based evidence (i.e., customer call and social media) if the state of the corresponding customer is known

- ΔT: waiting time of outage location inference (i.e., 10 minutes).
- When ΔT increases, utilities can receive more human-based evidence.
- This CDF is formulated using an exponential distribution. The parameter is learned empirically from historic data.

(4) $P_{E^m_{i,j}|C^j_i}\left(e^m_{i,j}\middle|c^j_i\right)$: the probability of receiving meter-evidence (i.e., last gasp signal) if the state of the corresponding customer is known, which depends system observability and SM accuracy levels.

- Delivered instantaneously.
- The parameter is learned empirically from historic data.



Offline Parameter Learning and Online Inference in Graphical Models

- In general, Eq. 1 can be directly solved by applying chain rule to graph parameters, as follows:

$$P_{\boldsymbol{D},\boldsymbol{C}|\boldsymbol{E}}(\boldsymbol{d},\boldsymbol{c}|\boldsymbol{e}) = \frac{\prod P_{\boldsymbol{D_i}|\boldsymbol{D_{i-1}},\boldsymbol{E_i^w},\boldsymbol{E_i^v},\boldsymbol{E_i^b}}(\boldsymbol{d_i}|\boldsymbol{d_{i-1}},\boldsymbol{e_i^w},\boldsymbol{e_i^v},\boldsymbol{e_i^b}) \prod P_{\boldsymbol{C_j^i}|\boldsymbol{D_i}}(\boldsymbol{c_j^i}|\boldsymbol{d_i}) \prod P_{\boldsymbol{E_{i,j}^h}|\boldsymbol{C_i^j},\boldsymbol{\Delta T}}(\boldsymbol{e_{i,j}^h}|\boldsymbol{c_i^j},\boldsymbol{\Delta t}) \prod P_{\boldsymbol{E_{i,j}^m}|\boldsymbol{C_i^j}}(\boldsymbol{e_{i,j}^m}|\boldsymbol{c_i^j})}{\sum_i \sum_k \{\prod P_{\boldsymbol{D_i}|\boldsymbol{D_{i-1}},\boldsymbol{E_i^w},\boldsymbol{E_i^v},\boldsymbol{E_i^b}}(\boldsymbol{d_i}|\boldsymbol{d_{i-1}},\boldsymbol{e_i^w},\boldsymbol{e_i^v},\boldsymbol{e_i^b}) \prod P_{\boldsymbol{C_j^i}|\boldsymbol{D_i}}(\boldsymbol{c_j^i}|\boldsymbol{d_i}) \prod P_{\boldsymbol{E_{i,j}^h}|\boldsymbol{C_i^j},\boldsymbol{\Delta T}}(\boldsymbol{e_{i,j}^h}|\boldsymbol{c_i^j},\boldsymbol{\Delta t}) \prod P_{\boldsymbol{E_{i,j}^m}|\boldsymbol{C_i^j}}(\boldsymbol{e_{i,j}^m}|\boldsymbol{c_i^j})\}} \quad (3)$$

- This approach takes advantage of the conditional independence inherent in the graphs, rather than simply assuming full independence as with the pooling method.

- Instead of a brute-force enumeration of all possible {D,C} values from scratch, the proposed graphical technique relies on graph parameters to decompose $P_{\boldsymbol{D},\boldsymbol{C}|\boldsymbol{E}}(\boldsymbol{d},\boldsymbol{c}|\boldsymbol{e})$, which leads to an exponential reduction in outage detection time.

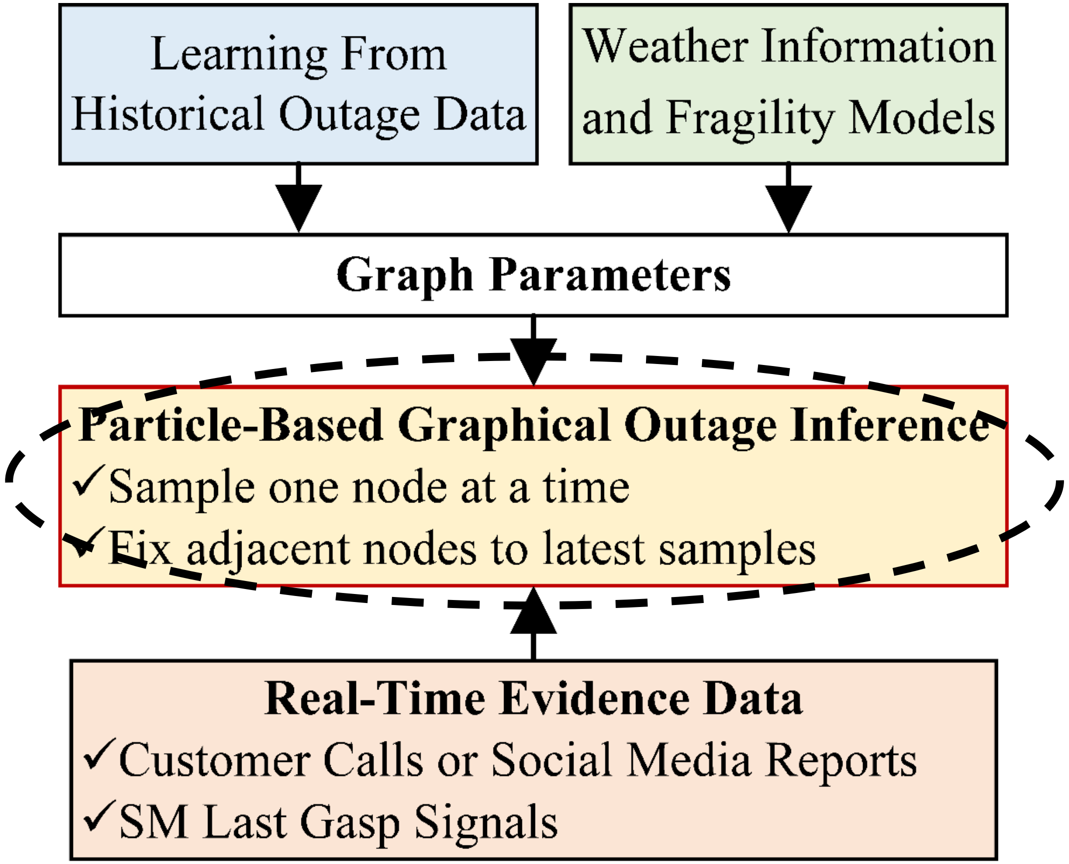
Offline Parameter Learning and Online Inference in Graphical Models

- **Challenge**: solving Eq. 3 requires running computationally expensive summation operations over all nodes of the graph simultaneously, which is not scalable for large distribution grids.

- **Solution**: Using fast particle-based inference methods, such as Gibbs sampling [7], to perform graphical inference efficiently.

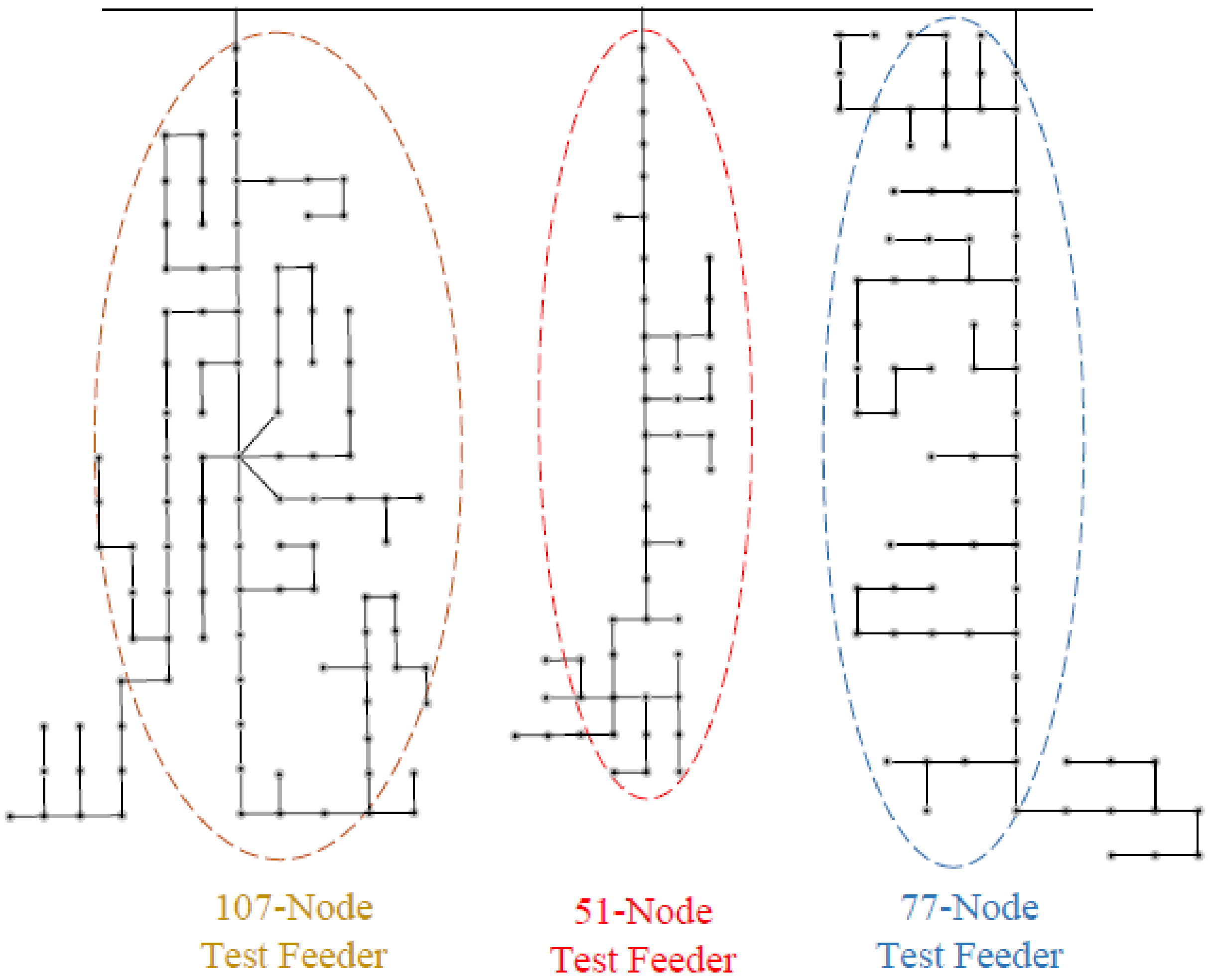
- Particle-based methods sample from individual nodes in the graph repeatedly at each iteration while fixing all the others to their latest samples.

- The key idea is to limit computation to a single node at each step while still considering nodal interdependence, which enables immense acceleration of outage inference in large grids.



Numerical Results

- Tested on three real distribution feeders, 51-, 77-, and 107-node test feeders.
- Evaluated the proposed method under three different observability levels, 25%, 50%, 75% for each test feeder.
- Generate 1500 outages for each case (a total of 9 cases).
- In each outage, the outage location is randomly chosen.
- The amount and location of meter-based evidence in each scenario is determined system observability.
- The human-based evidence is generated using a pre-defined exponential PDF (different from (3)) given ΔT=10 minutes.

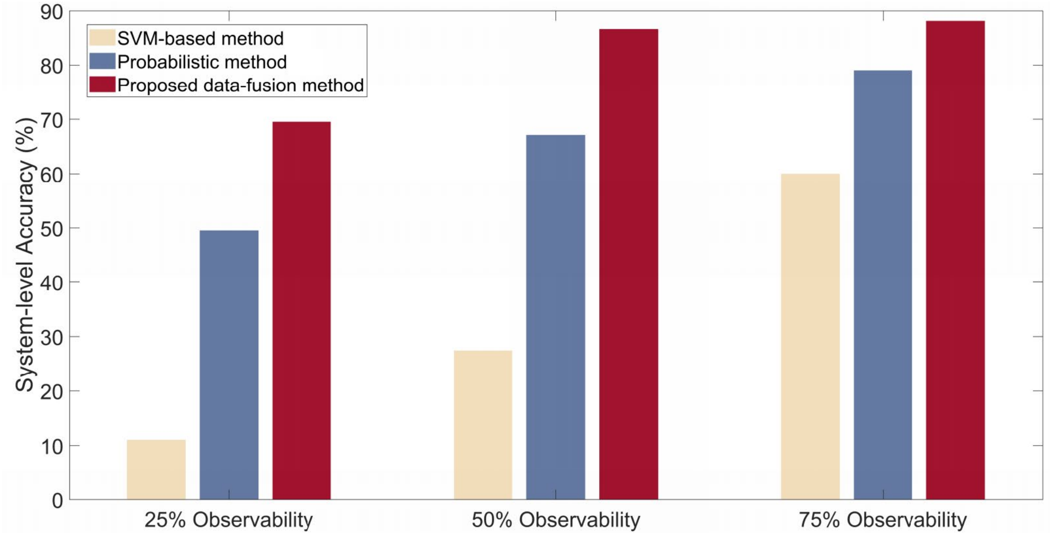


Numerical Results

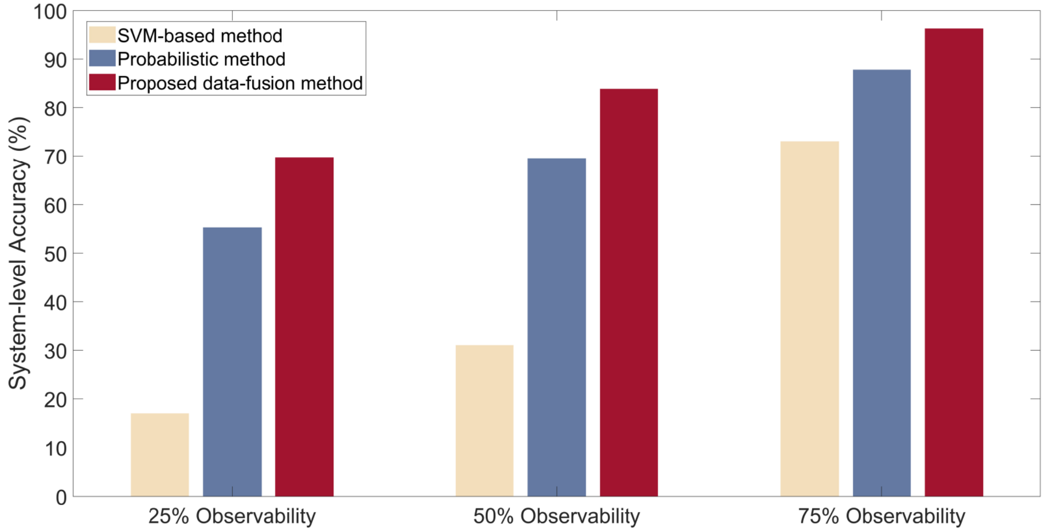
System Name	Observability	Branch-level Accuracy	Branch-level Precision	Branch-level Recall	Branch-level F_1	Location Accuracy
51-Node Test Feeder	25%	99.05%	86.48%	99.56%	90.65%	69.73%
	50%	99.65%	92.77%	99.82%	95.07%	83.93%
	75%	99.89%	98.38%	100%	98.93%	96.33%
77-Node Test Feeder	25%	98.7%	83.47%	98.88%	88.05%	69.5%
	50%	99.41%	92.43%	98.86%	94.32%	86.6%
	75%	99.60%	92.82%	99.89%	95.24%	88.1%
106-Node Test Feeder	25%	98.92%	83.91%	99.05%	88.61%	69.6%
	50%	99.58%	91.11%	99.54%	94.1%	80.9%
	75%	99.92%	98.19%	100%	98.88%	92.6%

- ✓ Observability is determined by the number of SMs.
- ✓ The performance of the proposed outage location method improves as the observability increases, due to the high confidence levels of meter-based evidence.
- ✓ The proposed algorithm shows almost the same level of performance over the different test networks.

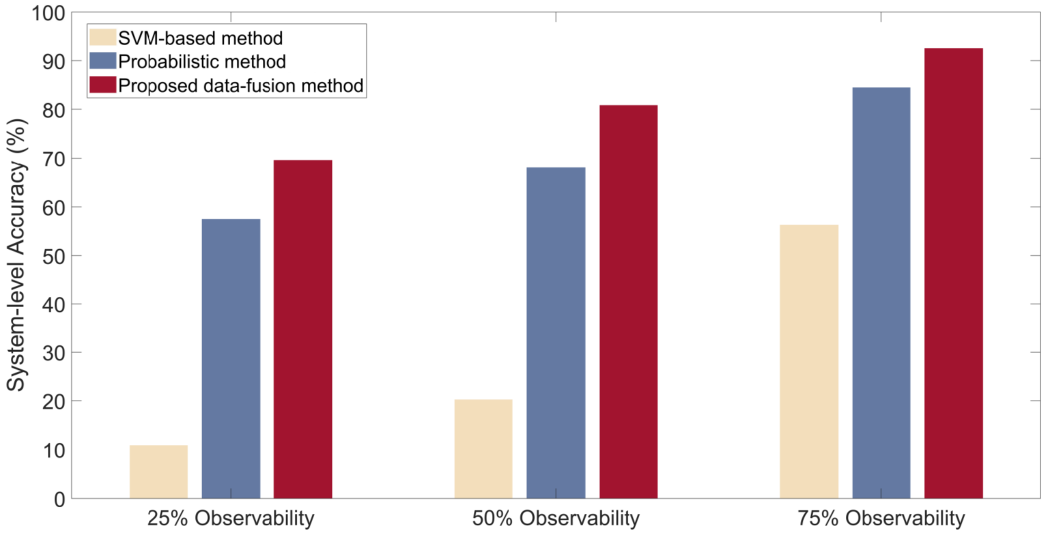
Numerical Results



Comparison results of the 77-node test system



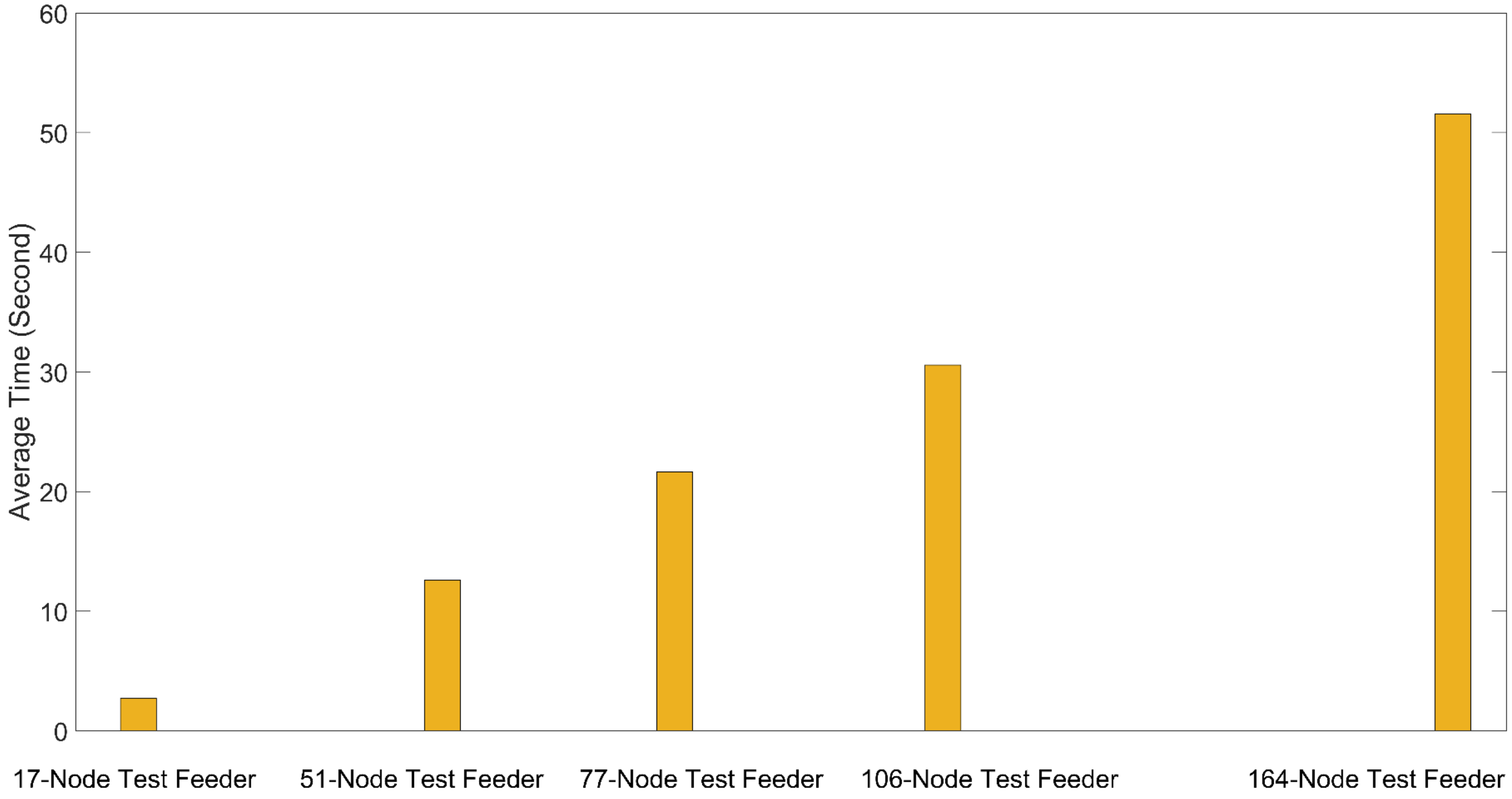
Comparison results of the 51-node test system



Comparison results of the 106-node test system

- ✓ Compared with two existing outage location methods, a SVM-based approach [8] and a probabilistic approach [9].
- ✓ [9] and our method generally outperform [8], especially for system with low observability.
- ✓ Among the data fusion-based methods, our method performs better than [9] because the proposed method not only uses data from smart meters, but also effectively combines data from non-metered data sources.

Numerical Results



- ✓ Conducted on two additional real-world distribution feeders: a 17-node and 164-node feeders to provide a comprehensive computational complexity analysis.
- ✓ Using our method, the average computational time for outage location has an approximately linear, rather than exponential, relationship with the size of the distribution grid.

Conclusions

- Combining heterogenous data sources can significantly improve outage detection accuracy.
- Outage data sources are conditionally independent.
- Our method encodes the network's topology and the causal relationship between outage evidence and branch states into BNs by leveraging the conditional independence inherent in distribution grids.
- The proposed graphical method, by the merit of its multi-source nature, is highly robust against low observability, while at the same time maintaining high detection speed.
- Future study will seek to extend the proposed method in meshed grids with high penetration distributed energy resources.

References

- [1] Mahoor, M., Salmasi, F.R. and Najafabadi, T.A., 2017. A hierarchical smart street lighting system with brute-force energy optimization. IEEE Sensors Journal, vol. 17, no. 9, pp.2871-2879.
- [2] Enqvist, Olof, Fangyuan Jiang, and Fredrik Kahl. "A brute-force algorithm for reconstructing a scene from two projections." CVPR 2011. IEEE, 2011.
- [3] 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, 2020.
- [4] R. Moghaddass and J. Wang, “A hierarchical framework for smart grid anomaly detection using large-scale smart meter data,” IEEE Trans. Smart Grid, vol. 9, no. 6, pp. 5820–5830, Nov. 2018.
- [5] A. M. Salman, Y. Li, and M. G. Stewart, “Evaluating system reliability and targeted hardening strategies of power distribution systems subjected to hurricanes,” Rel. Eng. Syst. Safety, vol. 144, pp. 319–333, Dec. 2015.
- [6] C. Fu, Z. Yu, D. Shi, H. Li, C. Wang, Z. Wang, and J. Li, “Bayesian estimation-based parameter estimation for composite load,” 2019. [Online]. Available: https://arxiv.org/abs/1903.10695.
- [7] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin., Bayesian Data Analysis. CRC Press, 2013.
- [8] 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.
- [9] Y. Jiang, “Data-driven probabilistic fault location of electric power distribution systems incorporating data uncertainties,” IEEE Trans. SmarT Grid, vol. 12, no. 5, pp. 4522–4534, 2021.

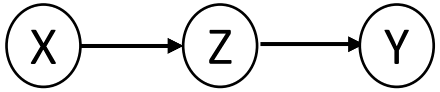
Thank you!

Q&A

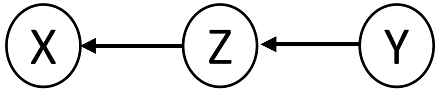
Encoding Distribution Grids and Data into Probabilistic Graphs

✓The edge directions are important.

- ➢ Variable Z is cause of Y and effect of X. In (b), variable Z is a common effect for both X and Y.



(a) Causal Chain



(b) Common Effect

- ➢ For (a), X cannot influence Y via Z if Z is observed.
- ➢ For (b), X can influence Y via Z, but only if Z is not observed.