Grid-Edge Technologies to Enhance Distribution Grid Modeling and Operation

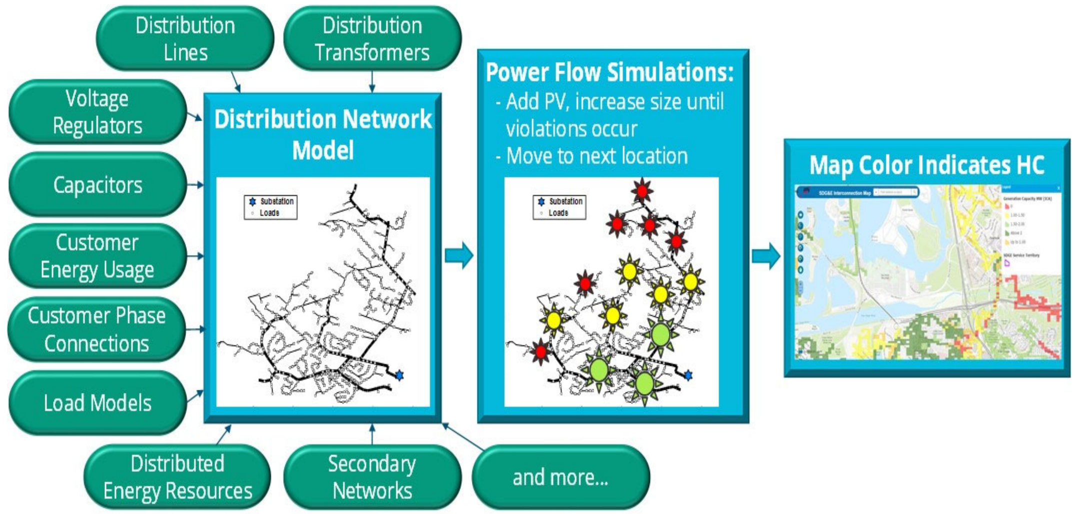
Zhaoyu Wang
Professor
Iowa State University
(wzy@iastate.edu)

Project Background

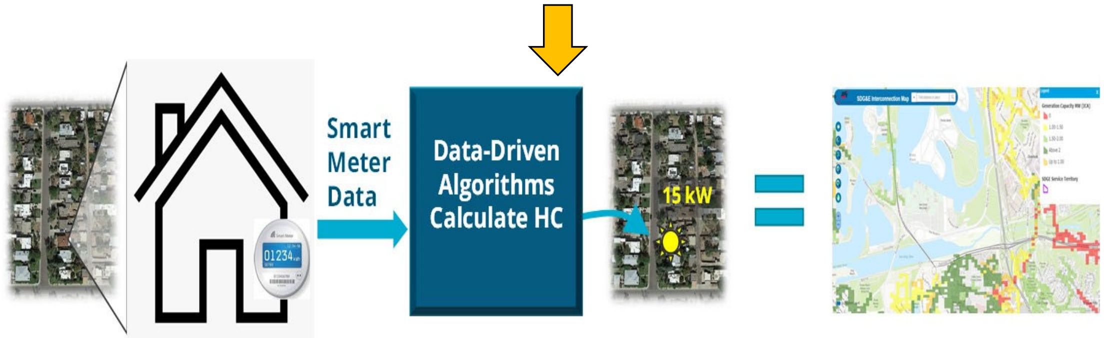
Smart Meter Data: A Gateway for Reducing Solar Soft Costs with ***Mo***del-Free ***H***osting ***Ca***pacity (MoHCa) Maps

- The objective of the project is to develop scalable algorithms for estimating the ***voltage-constrained*** and ***thermal-constrained*** HC at smart meter locations through ***Data-driven Model-free HCA method***.

- The developed algorithms will be validated on utility datasets and incorporated into ***Open Modeling Framework (OMF.coop)*** for over 260 utilities and vendors throughout the US to use.



Conventional model-based HCA



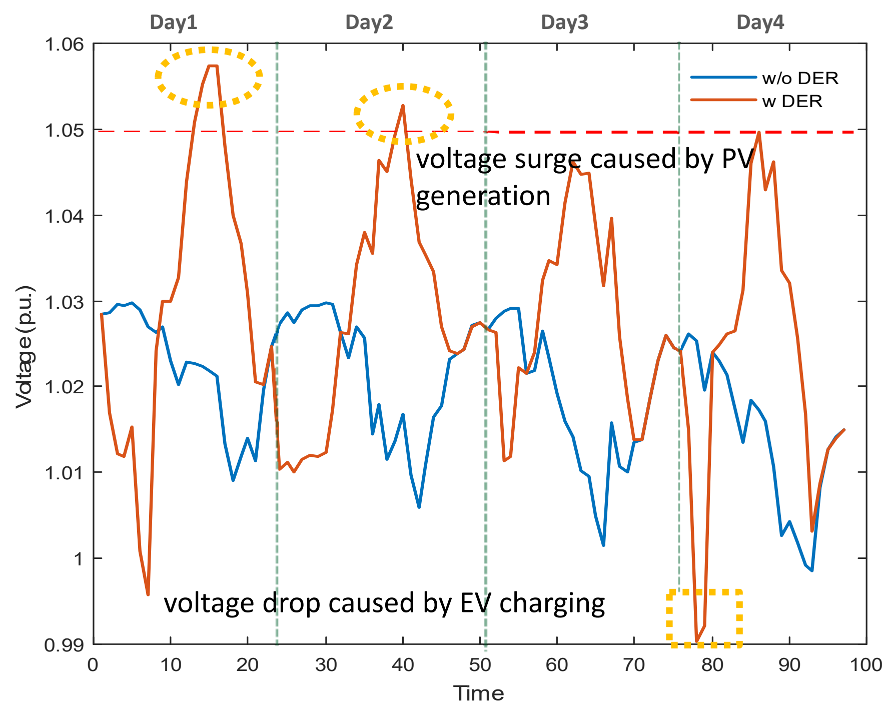
Data-driven Model-free HCA

Secondary Network Voltage Calculation

➢Integrating DERs into distribution networks introduces ***voltage issues***.

➢Model-free voltage calculation is a promising approach, yet existing research still presents certain limitations:

- ***Overlooking low-voltage secondary distribution network (SDNets)*** [8][9]
- ***Performing poorly for high-impact, low-probability extreme voltage scenarios*** [10][11]
- Typically black-box, ***lacking physics-informed interpretability*** [12][13][14]



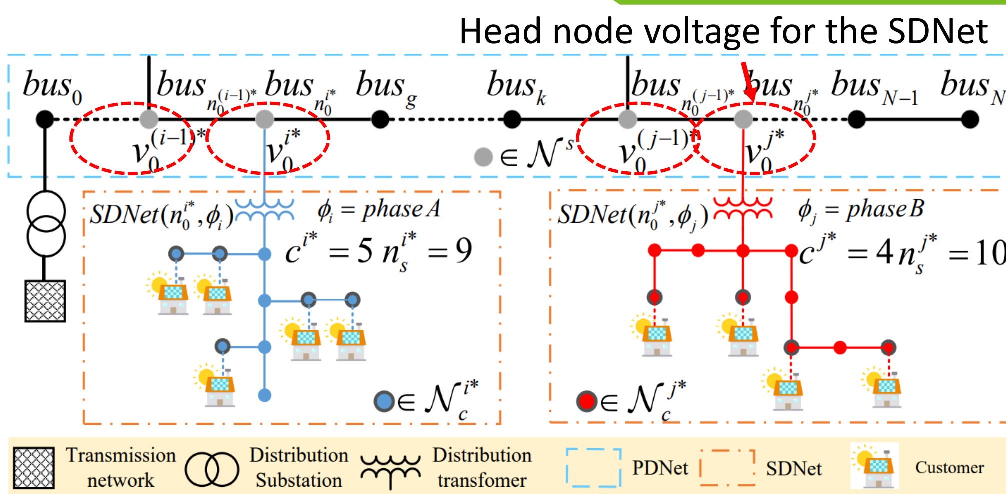
The illustration of DER influence on customer voltage

➢We propose a model-free voltage calculation method for ***integrated primary-secondary networks*** based on a customized ***physics-inspired neural network*** (PINN) by using only ***smart meter data***.

Note: Primary Distribution Network (PDNet) Secondary Distribution Network (SDNet) Power Flow (PFlw)

PDNet-SDNets Coupled Power Flow Model

- To assist the design of PINN model structure, we develop a coupled power flow (PFlw) model for integrated primary-secondary networks.
- The PFlw model explicitly captures the SDNets PFlw using linearized power flow equations, and implicitly considers the influence from PDNet PFlw changes.



The structure of integrated primary-secondary distribution networks

PDNet-SDNets coupled PFlw model

$$v_c = Ep_c + Hq_c - m_c^s + \Psi(p_c, q_c) + \chi(p_c, q_c, v_c)$$

Customers' active power and reactive power recorded by smart meters

Squared voltages recorded by customers' smart meters

$Ep_c + Hq_c - m_c^s$ → *Model the SDNets using linearized power flow equations.*

$\Psi(p_c, q_c)$ → *Consider the voltage variances at customer nodes caused by the PFlw changes in the PDNet.*

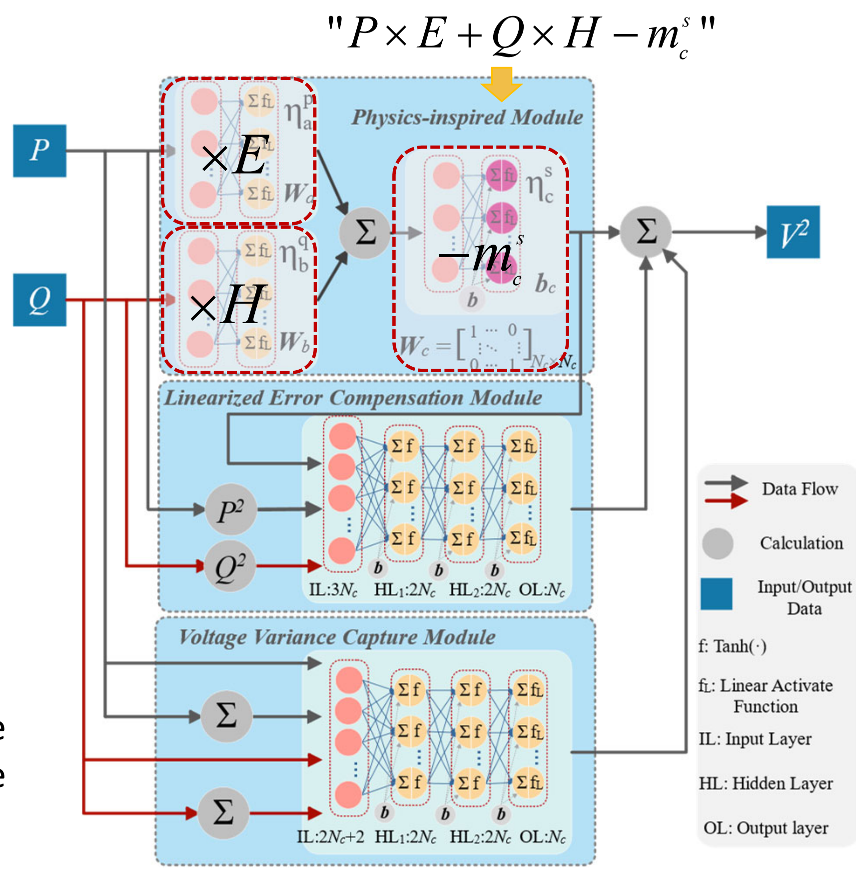
$\chi(p_c, q_c, v_c)$ → *Compensate for the linearization error caused by lines' and secondary xfrms' losses.*

Physics-inspired Model-Free Voltage Calculation

- Inspired by the coupled PFlw model, we design a PINN model

$$v_c = Ep_c + Hq_c - m_c^s + \chi(p_c, q_c, v_c) + \Psi(p_c, q_c)$$

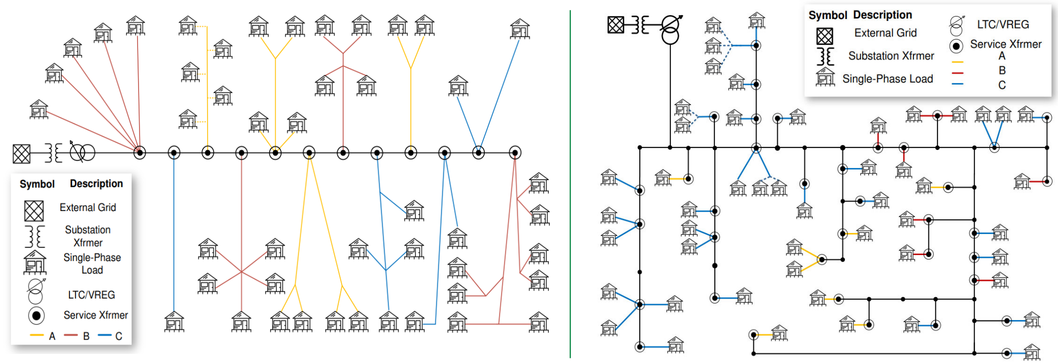
- ***Physics-inspired module (PIM)***
  - ✓We design the PIM module to capture the linearized power flow equations of the SNDets.
  - ✓The inputs of the PIM module are P,Q data for all customers.
- ***Linearized error compensation module (LECM)***
  - ✓We use a fully connected neural network to capture the linearization error caused by lines' and xfrms' losses.
  - ✓Squared P, Q data and squared approximate V of each customer are considered as inputs.
- ***Voltage variance capture module (VVCM)***
  - ✓We utilize a fully connected neural network to capture the customer voltage variance caused by PFlw changes in the PDNet.
  - ✓The P, Q data of each customer and total P, Q loads of all customers are the inputs.



The structure of the proposed PINN model

Test Circuits and Simulation Setting

- Three distribution feeder models are used to perform case studies, and each model integrates secondary xfrms and SDNets.
  - Two public testing circuits, namely, ***EPRI Secondary Topology Model*** marked as "EPRI12Bus" and ***EPRI Ck5*** circuit.
  - One real utility feeder marked as "Real40Bus".



Topologies of EPRI12Bus model (left) and Real40Bus model (right)

- We test the proposed model in five scenarios, denoted as S1 to S5 in Table I.

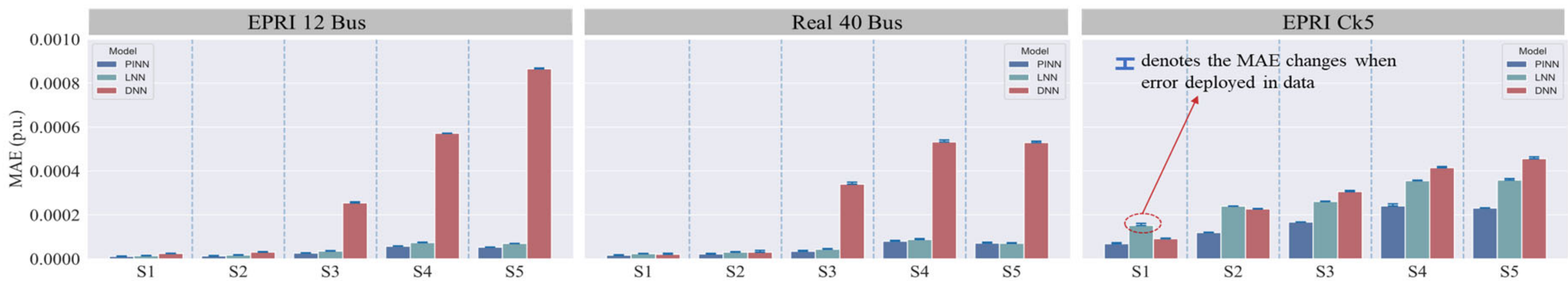
- The PV load data comes from over 300 solar inverters at 4-10 kW in Midwest U.S. The EV data is collected from various real datasets and has charging capacities at 3-10 kW.

TABLE I
SIMULATION SCENARIO GENERATION SETTING

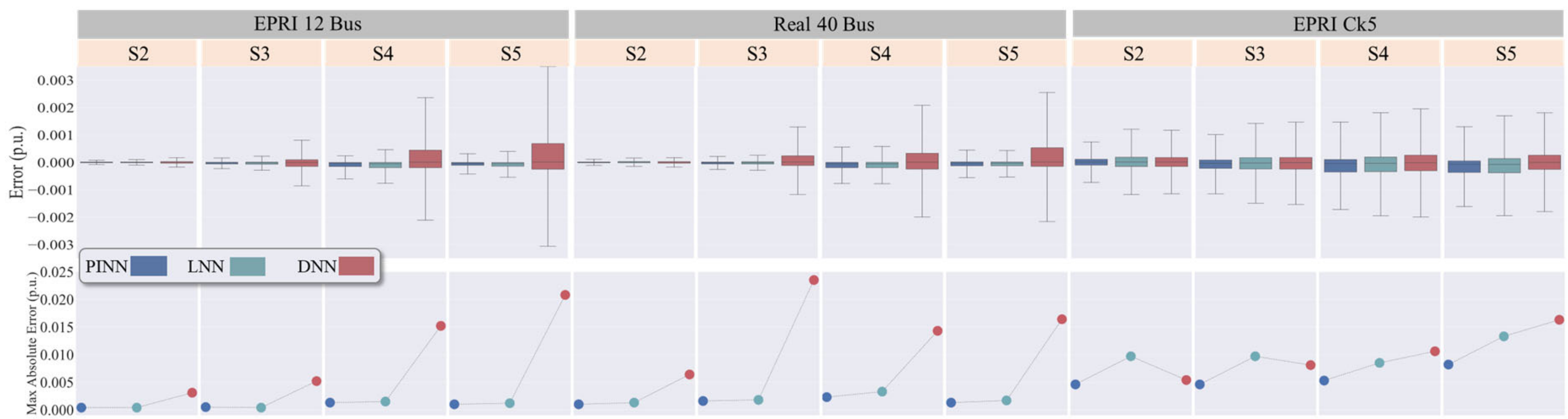
Scenario	Training	Testing	PV Penetration(%) EPRI12Bus/Real40Bus/EPRICk5
S1	basic	basic	0%/ 0%/ 0%
S2	25%PV	25%PV	39%/ 56%/ 57%
S3	basic	25%PV	39%/ 56%/ 57%
S4	basic	50%PV	114%/ 108%/ 93%
S5	basic	50%PV + 20%EV	114%/ 108%/ 93%

Results of Voltage Calculation

➢ Models are built on one-year smart meter data, where 80% is used for training and 20% is used for validating, and then tested on another one-year data. The errors of voltage calculation are shown below.

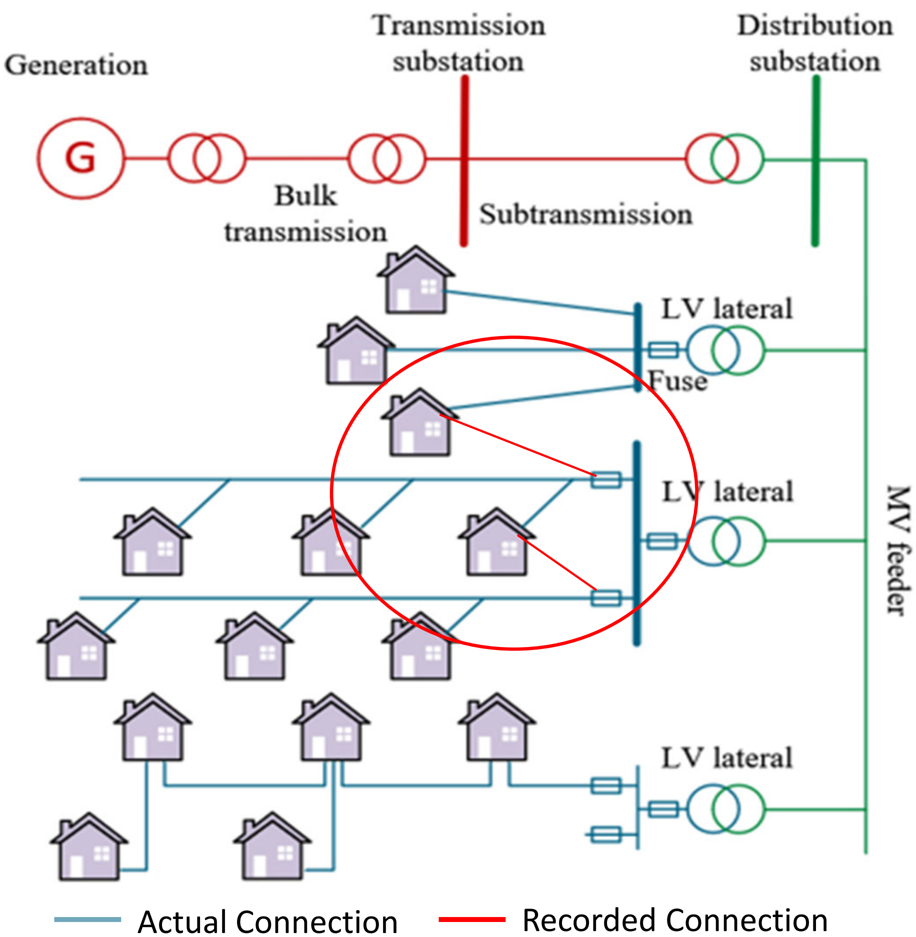


The MAE of three models over different scenarios based on accurate data and noise-added data



The error distribution of three models over different scenarios during daytime (6 a.m. to 6 p.m.) period and max absolute error in each scenarios

Introduction to T-C Connectivity Grouping



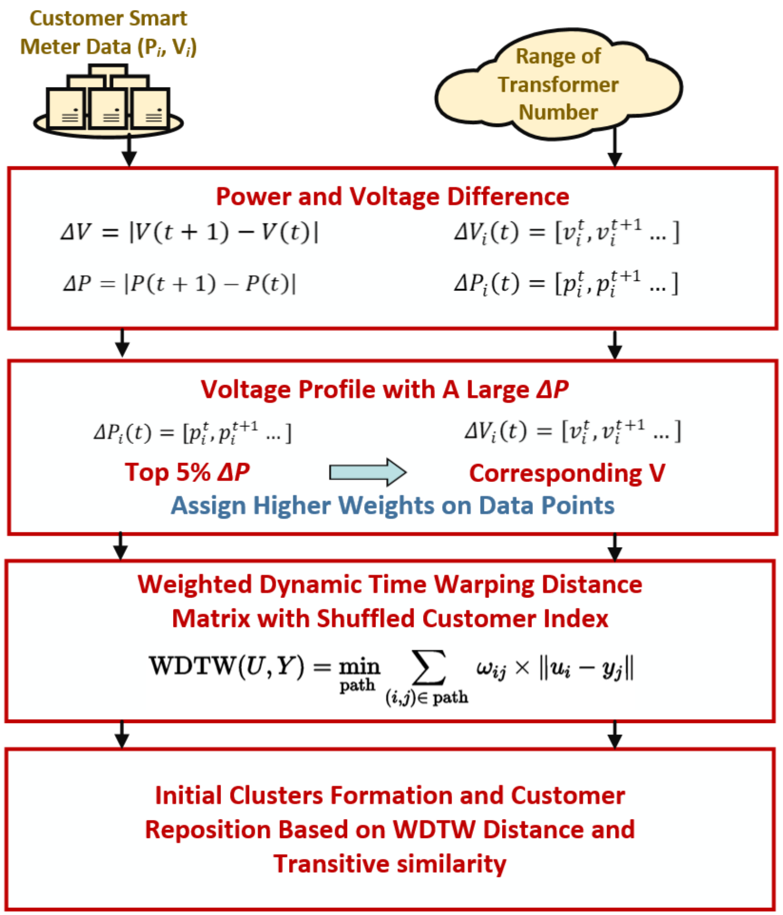
- **Problem Statement:** Transformer-customer (T-C) connectivity grouping refers to the determination of the physical connections between customers and the corresponding transformers.
- **Challenges and Difficulties:**
  - The numbers of transformers and customers are large.
  - Lack of measurements at distribution transformers.
  - Outdated and erroneous connectivity models.

Literature Review and Limitations

Ref.	Main Approach	Data source and Prior Knowledge	Limitations
[1]	Cubic SVM classifier model	Customer smart meter V, transformer numbers, partial knowledge of T-C pairs, drop line parameters	(-) Need partial connectivity knowledge to train a classifier. (-) Require prior knowledge on transformer numbers and must be accurate. (-) Not robust to bad/missing data. (-) Only for model calibration not identification.
[2]	K-means clustering, t-SNE, and self-organizing map		
[3]	Pearson correlation coefficient		
[4]	Hausdorff similarity assessment	Customer smart meter P and V, DER P and V, transformer numbers, measurements on transformers	(-) Transformer-side measurements are not always available. (-) Cluster stage is highly dependent on prior knowledge of transformer numbers. (-) Assumptions and data processing technique are not always applicable on US distribution systems.
[5]	Markov random field		
[6]	Silhouette coefficient and power loss coefficient		
[7]	Weighted convolution power optimization model		

- Major limitations in existing works:
    - Require accurate transformer numbers and/or partial T-C relationships.
    - Require measurements on transformers.
    - Can only do model calibration not identification.

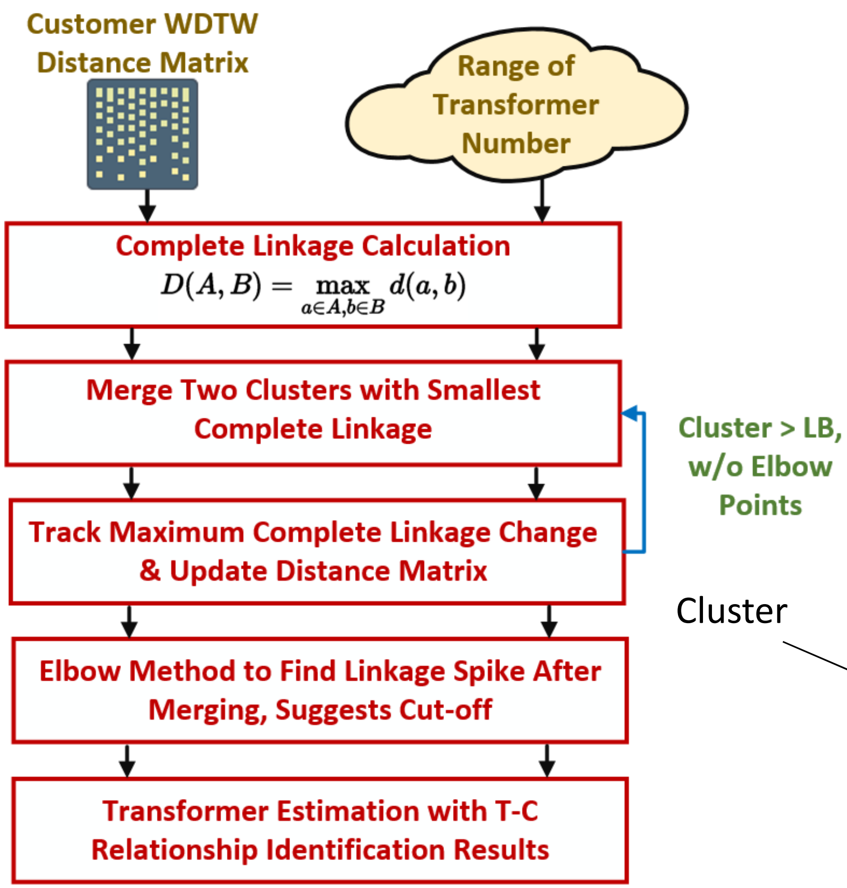
Methodology - Initial Clustering Phase



Data Processing and Initial Clustering Phase:

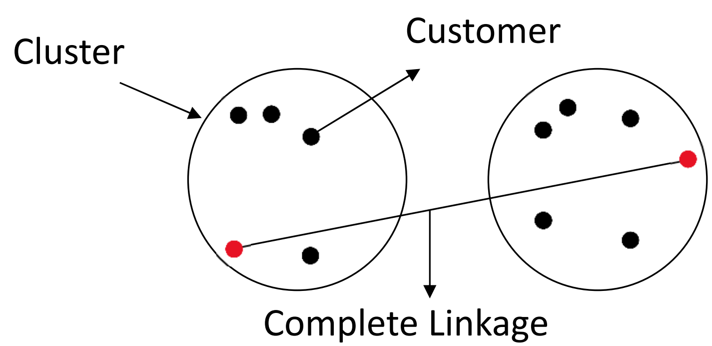
- ✓ Find customer voltage measurements with a large ΔP. Those voltage data are assigned with larger weights.
- ✓ Customer voltage data is used to calculate the dynamic warping distance between any two customers. Shorter distance = higher correlations.
- ✓ The distance matrix is then used to generate initial clusters. For each customer, top x close distance customers are first grouped together. x = total customers / upper range of transformer number.
- ✓ One customer can be moved to another cluster if the total DTW distance is shorter in the new cluster.
- ✓ Each customer can only exist in one cluster.

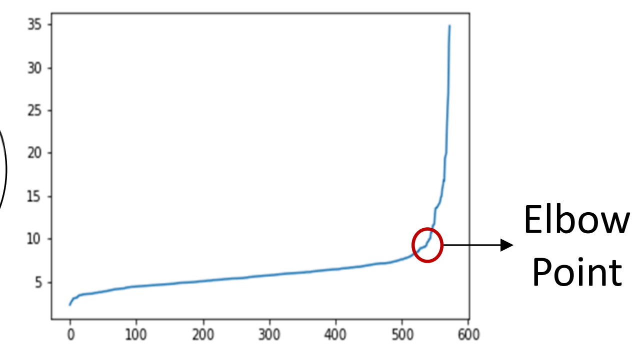
Methodology - Cluster Adjustment Phase



Cluster Adjustment Phase:

- ✓ **Customer DTW distance matrix** is used to calculate the maximum Euclidean distance (complete linkage) between every two clusters.
- ✓ Iteratively merge two clusters with the smallest complete linkage into one cluster to reduce the total number of clusters to the lower bound of the transformer range.
- ✓ Track the **maximum complete linkage** at each merging step.
- ✓ Use the **elbow method** to find a linkage spike after merging, which suggests a cut-off point where the two clusters being merged are substantially different from each other.

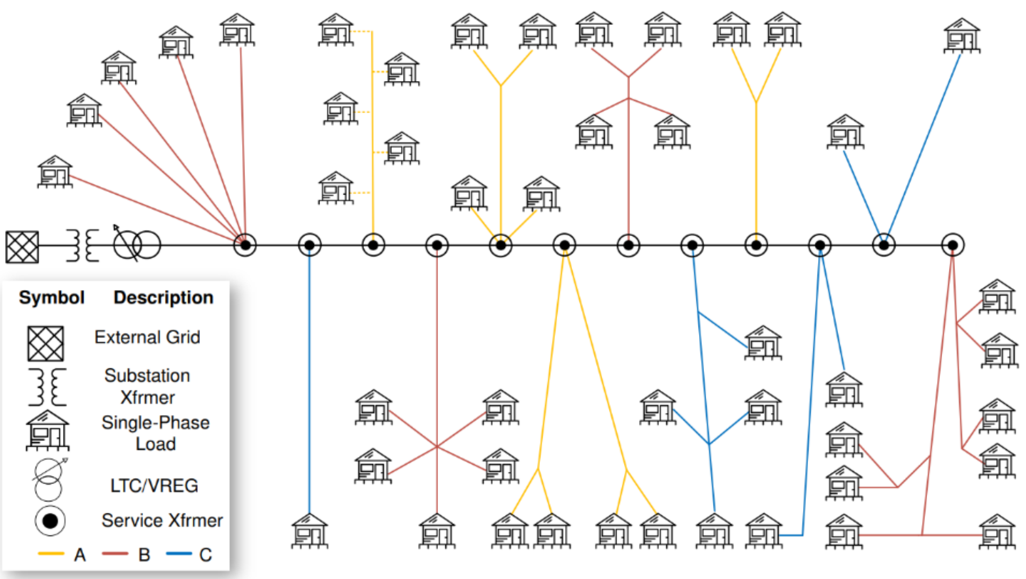




Case Study and Numerical Results

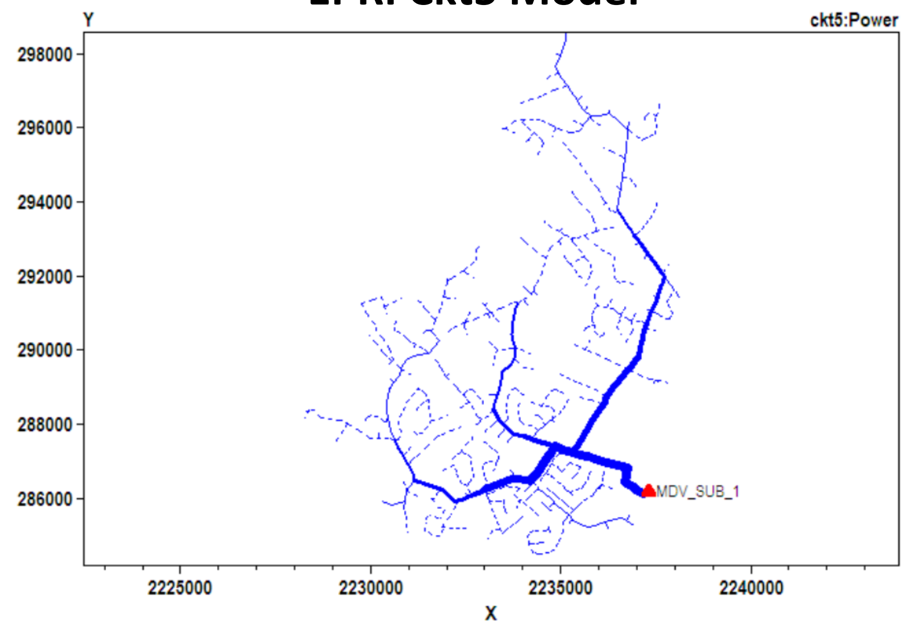
- Model test dataset: ***EPRI Secondary Topology (ST) Model*** and ***EPRI Ckt5 Model***
- Input data: One year of customer smart meter voltage measurements at 15-min resolution.

EPRI Secondary Topology Model



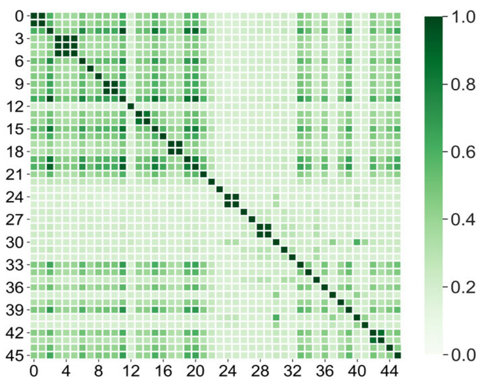
Transformers: 12; Customers: 46

EPRI Ckt5 Model

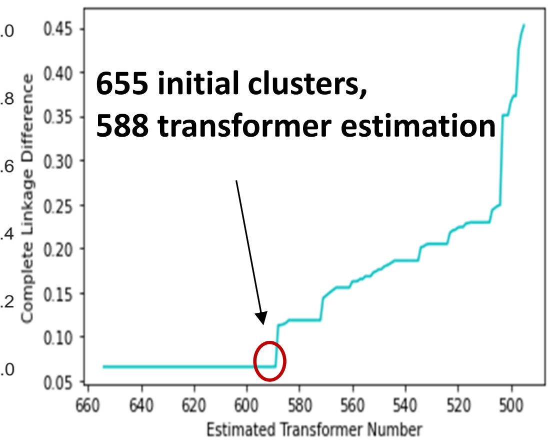


Transformers: 591; Customers: 1379

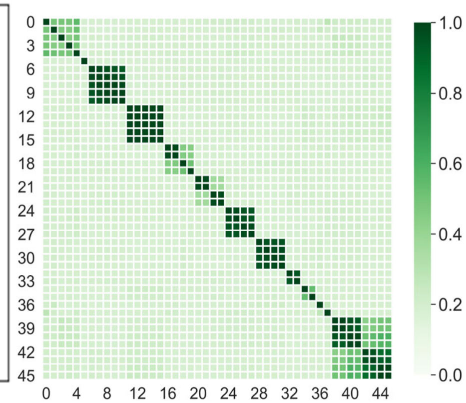
Case Study and Numerical Results



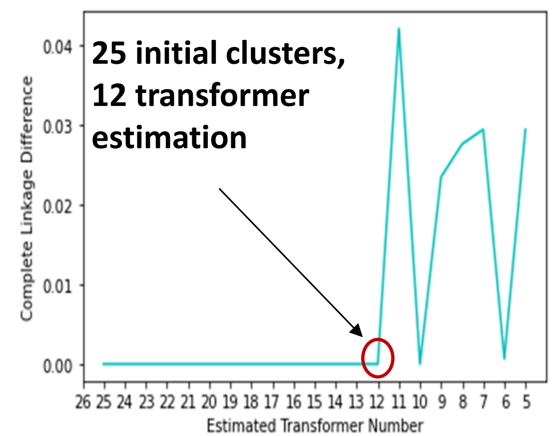
Customer DTW distance heatmap for EPRI ckt5 model (partial). Darker green = shorter distance



Maximum complete linkage changes by cluster merging iteration (ckt5)



Customer DTW distance heatmap for EPRI ST model. Darker green = shorter distance

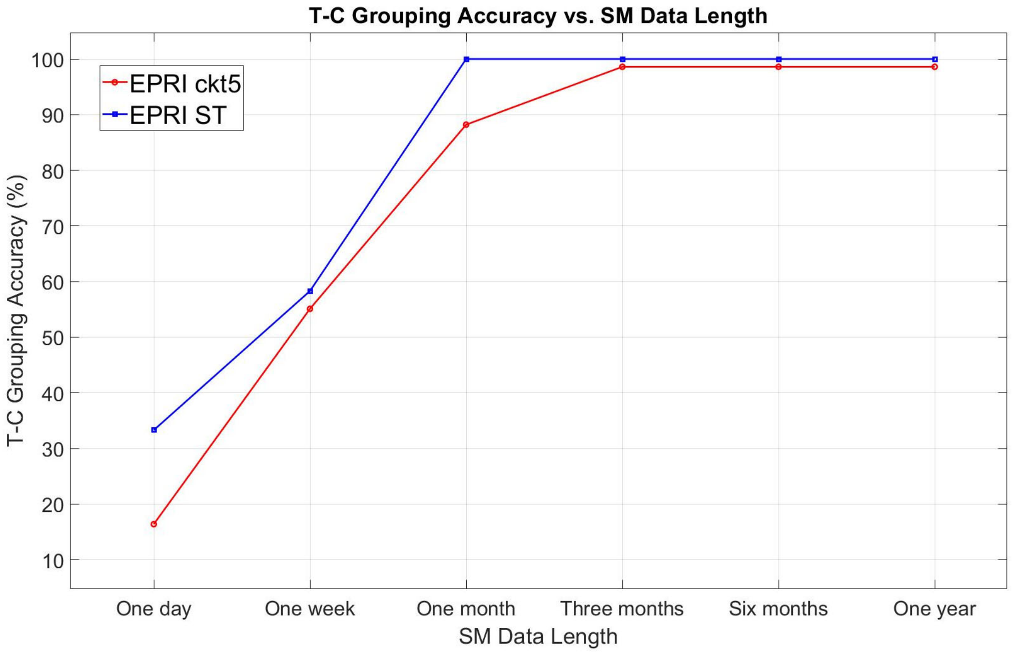


Maximum complete linkage changes by cluster merging iteration (ST)

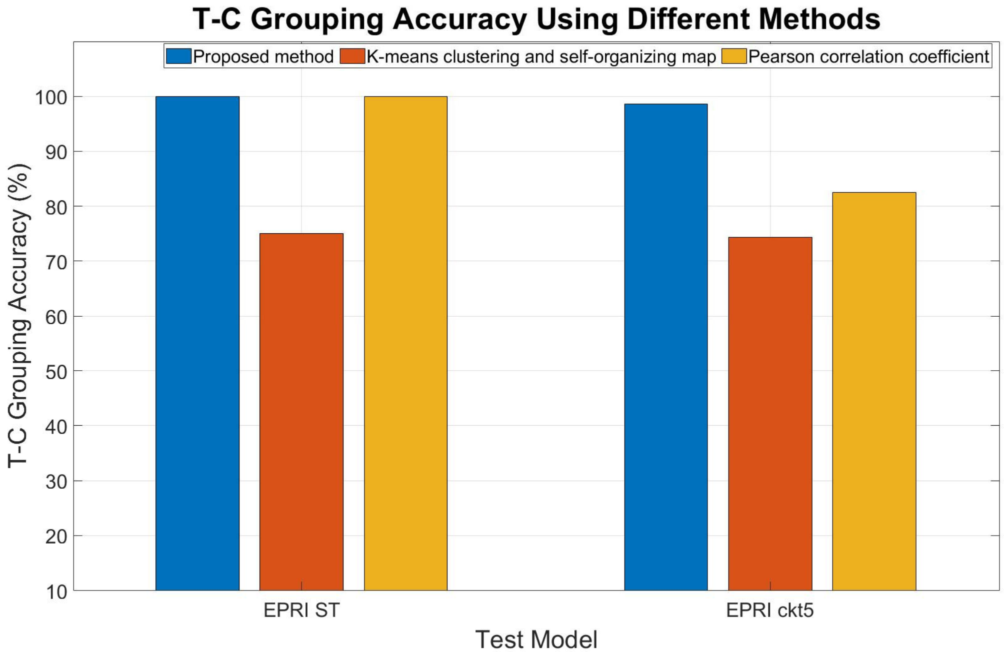
Model	Actual Transformers	Transformer Range	Estimated Transformers	Transformer Error Margin	Transformer with Correct Customer Grouping	Accuracy
EPRI ST	12	5-25	12	0	12	100%
EPRI ckt5	591	495-655	588	-3	583	98.6%

- A larger transformer range is selected to test the algorithm's robustness.
- Accuracy = Number of transformers with correct customer groupings / actual transformer number.

Sensitivity and Comparison Analysis



Sensitivity analysis on customer SM data length vs. T-C relationship identification accuracy



Comparison test using proposed method vs. existing works; data length = three months

- One month (EPRI ST) and three months (EPRI ckt5) of SM data is sufficient to obtain good T-C grouping results.
- Comparison test indicates the proposed method outperformed the existing works on the larger system.

Conclusion and Future Work

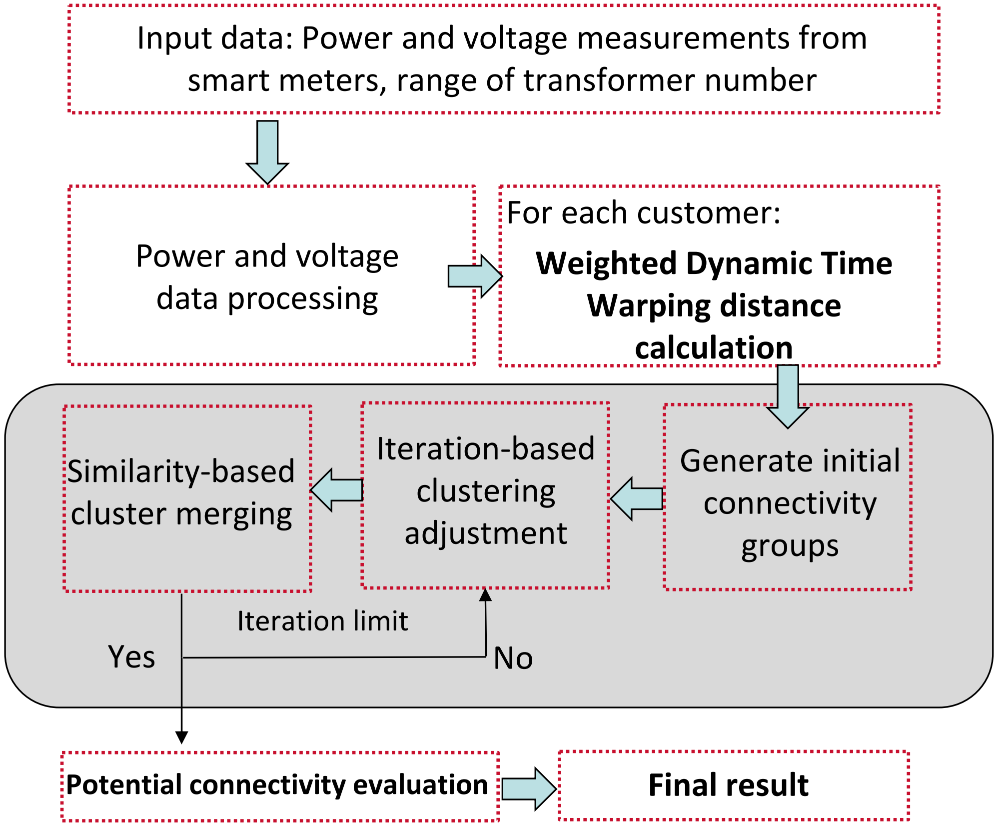
- It is possible to identify the T-C connectivity using only smart meter data without any prior knowledge of transformer numbers and transformer-side measurements.
- Leveraging the structure inspired by the PDNet-SDNets coupled PFlw, the PINN model shows potentials for extrapolation and capturing physical characteristics of the electrical network.
- Evaluations using two public testing systems and a real utility feeder model confirmed the effectiveness of the model in voltage calculation. The testing results also prove the proposed model's extrapolation, which is the ability to handle unseen scenarios.
- Future work will assess the model's adaptability to topology change.

Thank You!
Q&A

wzy@iastate.edu

Backup Slides

Proposed Methodology - Overview

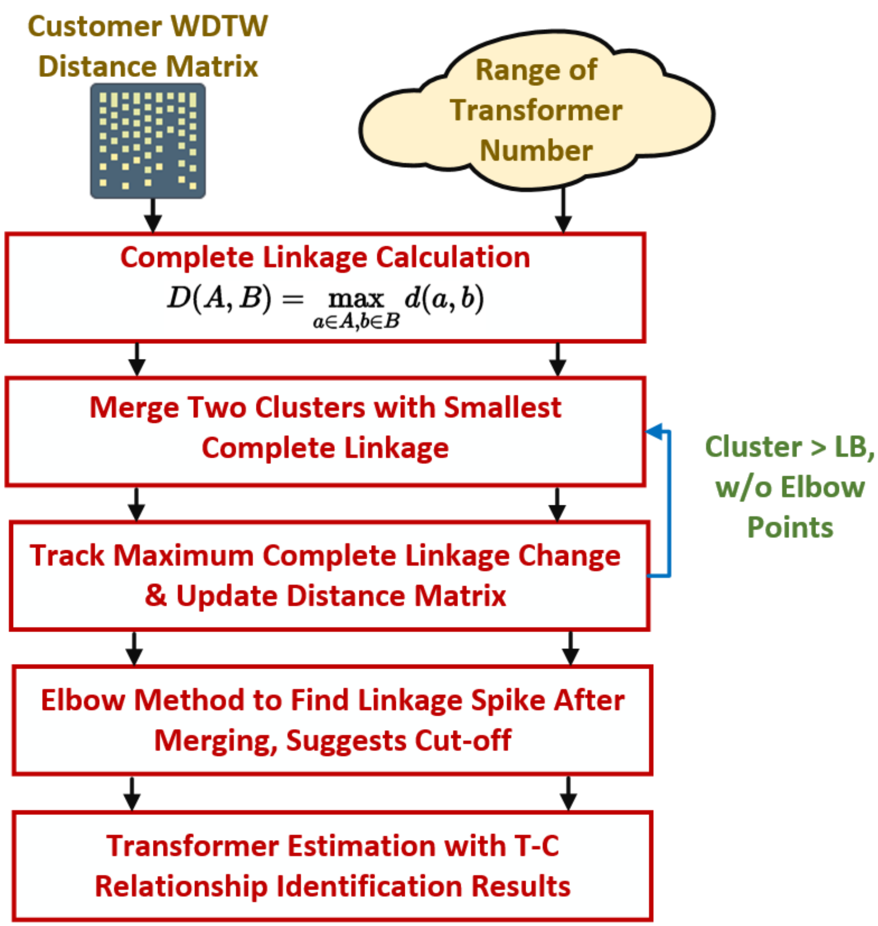


Flow chart of the proposed method

We propose a multi-stage framework to identify the T-C connectivity:

- ✓ Use voltage correlations between customers to develop a customer clustering algorithm.
- ✓ The accurate number of service transformers in the system is **not known** as a priori.
- ✓ The range of transformer number is estimated.
- ✓ An initial cluster set is formed based on the voltage correlations using weighted dynamic time warping (WDTW).
- ✓ Estimate the transformer number by a similarity-based cluster merging algorithm.

Methodology - Cluster Adjustment



Cluster Adjustment Phase:

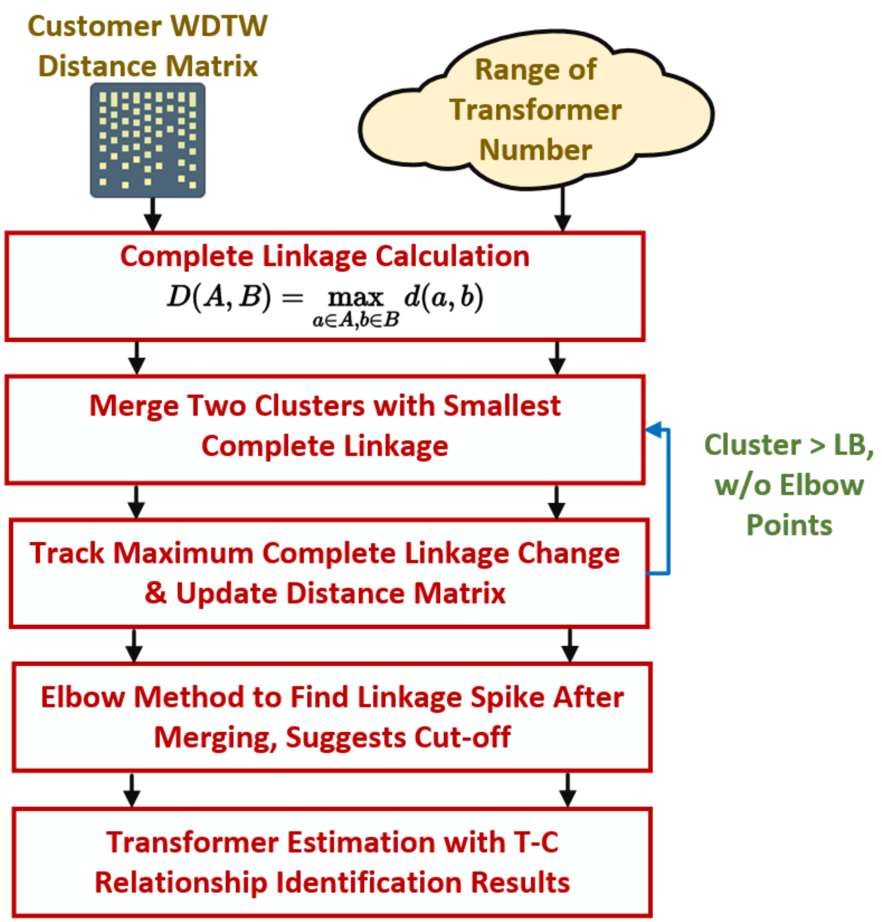
- ✓ Based on the **hierarchical clustering** idea, we merge two clusters into one iteratively until the total number of clusters reaches the transformer estimation LB.
- ✓ The distance between two clusters is the maximum distance between any point in cluster A and any point in cluster B, noted as **complete linkage**:

$$Distance(A,B) = \max(dist(a,b)) \quad (2)$$

where a is a point in cluster A, and b is a point in cluster B.

LB: Lower bound; UB: Upper bound

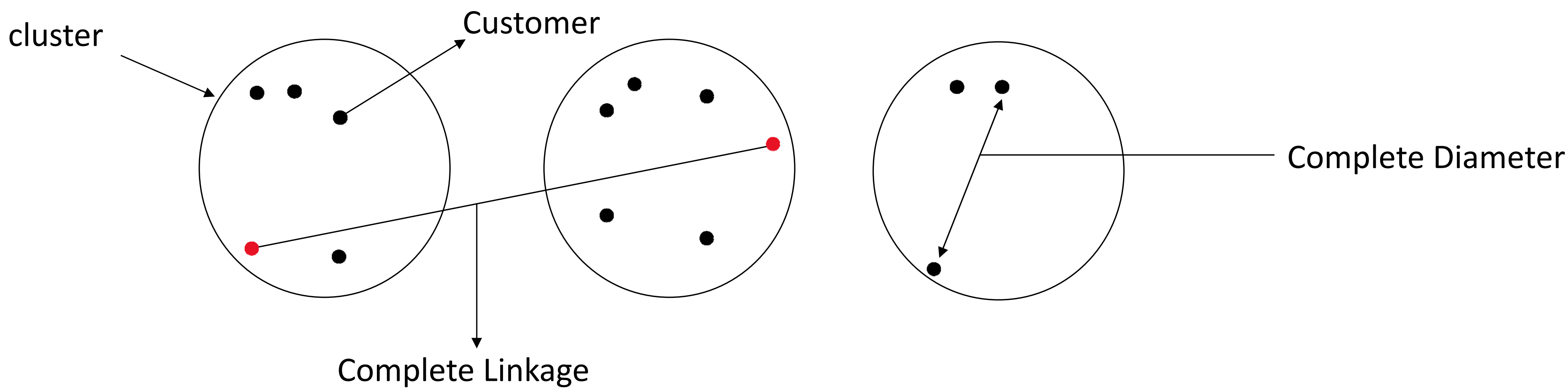
Methodology - Cluster Adjustment



Cluster Adjustment Phase (con'd):

- ✓ From the calculated linkage distances, identify the smallest linkage distance. The two clusters associated with this smallest distance are the ones that will merge next.
- ✓ After merging, update the distance matrix to reflect the distances between the new merged cluster and all other remaining clusters.
- ✓ Monitoring complete linkage changing in each process, using the **elbow method** to find a linkage spike after a merging process.
- ✓ The "elbow" of this curve is the point where the linkage distances start to increase substantially. It represents a cut-off point where clusters being merged are becoming substantially different from each other.

Complete Diameter and Complete Linkage

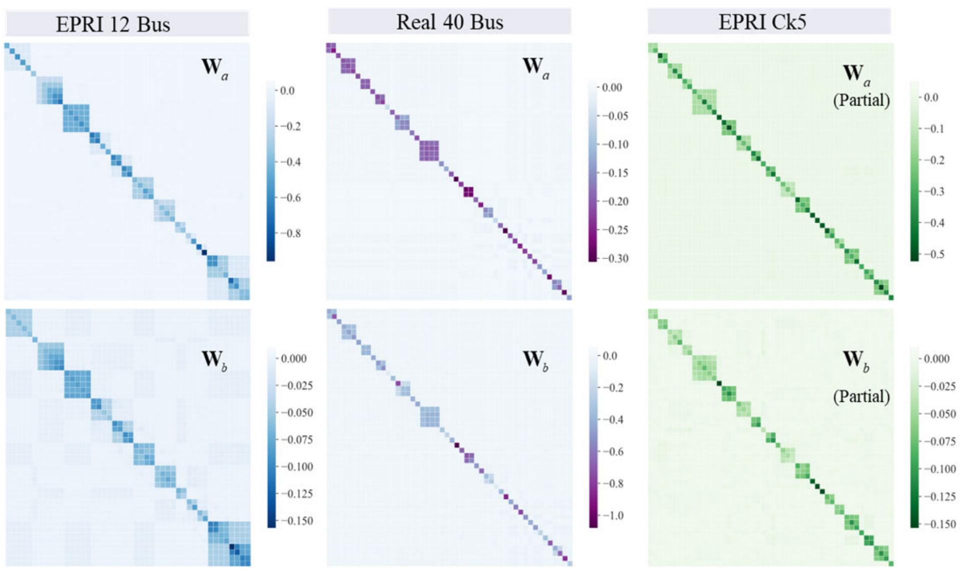


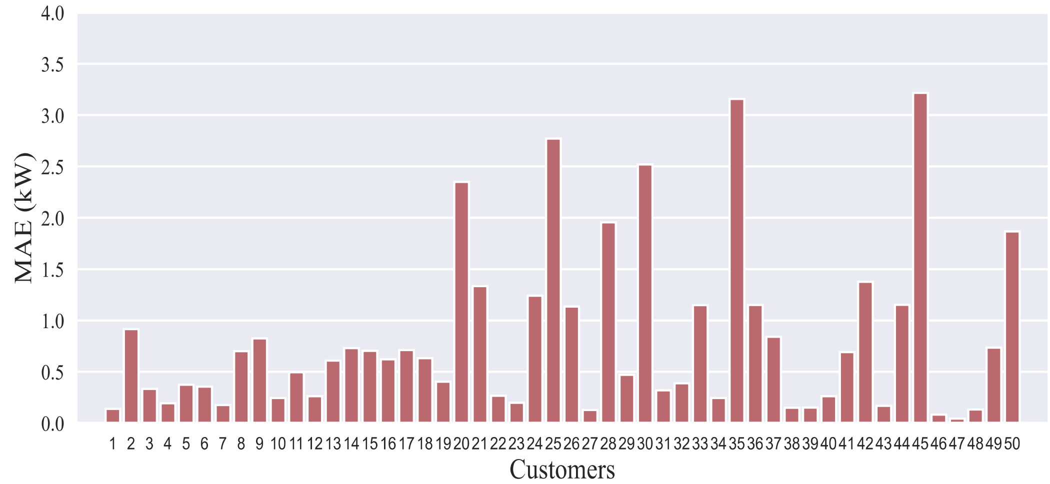
- The **Complete Diameter** of a cluster is the maximum distance between any two points within that cluster. This metric is often used to characterize the extent or size of a cluster in terms of how spread out its data points are.
- **Complete linkage** is a method for hierarchical clustering. When determining which two clusters should be combined, the complete linkage method considers the distance between the two most distant points (or farthest neighbors) in the different clusters.

Applications of PINN-based Voltage Calculation Model

- The designed model demonstrates excellent potential extrapolation capabilities due to the special structure, making it suitable for calculating voltages in high-penetration PV scenarios.

- The Real40Bus model is selected to complete the HC analysis. The performance of our model is competitive compared to previous locational HC work[7].





Average MAE of maximum accessible PV power for all customers

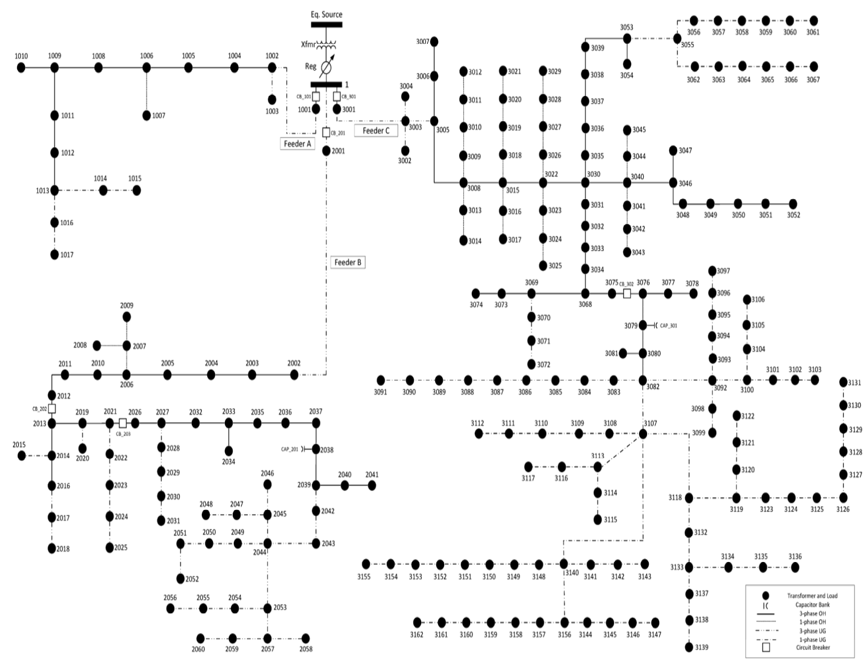
- ✓ The proposed PINN model, featuring a well-designed physics-inspired module, offers novel perspectives on solving transformer-customer (TC) connectivity problems.

- ✓ The designed method leverages the abundant physical information contained in W{a,b}

Data Sharing

With permission from our utility partner, we share a real distribution grid model with one-year smart meter measurements. This dataset provides an opportunity for researchers and engineers to perform validation and demonstration using real utility grid models and field measurements.

- The system consists of 3 feeders and 240 nodes and is located in Midwest U.S.
- The system has 1120 customers and all of them are equipped with smart meters. These smart meters measure hourly energy consumption (kWh). We share the one-year real smart meter measurements for 2017.
- The system has standard electric components such as overhead lines, underground cables, substation transformers with LTC, line switches, capacitor banks, and secondary distribution transformers. The real system topology and component parameters are included.



Test system diagram

You may download the dataset at: http://wzy.ece.iastate.edu/Testsystem.html , including system description (in .doc and .xlsx), smart meter data (in .xlsx), OpenDSS model, and Matlab code for quasi-static time-series simulation.