Genetic Algorithm-Enhanced Power-Based Branch Refinement for Voltage-Based Topology Identification in Low-Voltage Grids

Franziska Tischbein, Nina Štumberger, Maik Stürmer, Corvin Starkulla, Andreas Ulbig
Institute of High Voltage Equipment and Grids, Digitalization and Energy Economics (IAEW) at RWTH Aachen University
Aachen, Germany
f.tischbein@iaew.rwth-aachen.de

***Abstract*—Accurate digital topology models are a key enabler for advanced monitoring and planning in low-voltage grids. Previous work combined voltage-correlation-based topology identification with a node-degree-constrained Kruskal algorithm and a randomly permuted power-based branch refinement. While effective under ideal measurements, this method degrades under typical smart-meter uncertainties. This paper introduces a Genetic Algorithm (GA) for power-based branch refinement, enabling stochastic optimization of node-to-branch assignments. The GA improves robustness to measurement uncertainties and achieves full reconstruction across four representative low-voltage grids with standard smart meter measurement uncertainties.**

***Index Terms*—grid topology identification, genetic algorithm, smart meter, measurement uncertainty**

I. Introduction

In the context of the energy transition, distribution grids are increasingly coming into focus. In particular, the growing integration of new types of consumers and decentralized generation units at the low-voltage level is expected to result in higher asset utilization and possible violations of voltage limits. To address these challenges, distribution grid operators must gradually transition toward active grid operation, for instance, by controlling grid-connected assets based on state estimation and analysis, or by reinforcing the grids in line with the evolving supply requirements.

To support these measures with real-time information on grid loading, the rollout of intelligent metering systems (smart meters) was mandated under the EU Directive 2006/32/EC. However, both grid state estimation and grid expansion planning rely on accurate digital grid models as a prerequisite. Many distribution grid operators, however, currently lack sufficiently precise digital representations of their grids, since switching operations are not consistently updated in digital form, line parameters are missing, or, in some cases, no digital model exists at all [1], [2]. Hence, current research focuses on developing measurement-based topology identification methods to create accurate digital grid models [3].

A. Related Work

According to the categorization proposed by Deka, Kekatos and Cavraro [3], this study focuses on passive topology identification methods, as these allow the reconstruction of a complete grid model based solely on passively collected smart meter data—without actuating grid resources or observing the feeder's voltage response.

Due to their good scalability and straightforward applicability in real-world scenarios, voltage time-series–based approaches have become a common choice for topology identification in the literature. These methods allow deriving an initial estimation of the grid structure even from a limited set of measurement data. In the literature a variety of methods have been proposed that infer the topology of the grid from statistical dependencies between voltage measurements, commonly employing voltage correlation analysis and maximum spanning tree (MST) algorithms [4], [5], [6], [7].

While voltage correlation–based methods can provide a rough initial estimate of the grid structure, several studies propose two-stage approaches that refine this estimate using physical consistency. Reference [8] proposes a two-stage topology identification method that extends voltage-correlation-based identification by incorporating power measurements. While the concept provides a consistent framework, it neglects real-world effects such as reactive power variation, grid losses, and uncorrelated load behavior. Reference [9] introduces a refinement approach based on random permutation of node assignments while preserving approximate power balance. However, this strategy becomes computationally demanding for large grids and may not converge due to the exponentially growing number of binary variables.

As an alternative to random permutation, the use of a Genetic Algorithm (GA) offers a structured and efficient optimization mechanism for complex, discrete search spaces. GAs have already been applied in the field of topology identification. Reference [10] employed a GA to infer missing line parameters by minimizing the deviation between simulated and measured voltages under the assumption of a known grid topology. Refernce [11] utilized a GA for phase identification in low-voltage grids by minimizing the deviation between measured and aggregated energy values. Reference [12] applied a GA for full topology identification, performing a global search for the configuration that minimizes active power residuals, starting from a completely random initial grid model.

979-8-3315-8138-1

B. Contributions of this work

In contrast to the literature presented, this work introduces a two-stage topology identification process combining voltage-correlation-based topology identification with a GA-based refinement step, ensuring a physically meaningful initialization for the optimization. A recursive, branch-level refinement is introduced to correct node-to-branch assignments efficiently, reducing computational complexity compared to global approaches. In addition, measurement uncertainties and reactive power variations are explicitly considered, ensuring robust topology identification under realistic conditions.

II. Methodology and Genetic Algorithm Application

This section describes the method for the topology identification algorithm and the application of the GA. First, synthetic measurement data are generated as input for the proposed method, as described in section II-A. Based on this data, an initial grid model is reconstructed using voltage correlation analysis and a maximum spanning tree algorithm, following the procedure in section II-B. In the next step, GA-enhaced power-based branch refinement is applied, according to section II-C, to validate and adjust the reconstructed structure. Finally, the resulting grid topology is evaluated according to the methodology described in section II-D.

A. Generation of Synthetic Measurement Data

Given the limited availability of real-world measurement and grid data, this work is entirely based on synthetic datasets. The synthetic measurement time series were produced following the procedures outlined in [7], [13], while the generation of representative low-voltage grid models corresponds to the concepts presented in [14], [15].

To reflect the structural diversity of German low-voltage grids, a clustering analysis using k-means, with the number of clusters being identified using the elbow method, was conducted to identify characteristic grid representatives with varying topological and operational properties. This approach ensures that the subsequent evaluation covers a broad spectrum of real-world grid configurations. The feature set used for clustering comprises parameters such as the substation power, total line length, number of grid connections, average node degree, maximum node degree, average distance between grid connections, proportion of single-family homes, and proportion of trade and services. Thus, four representative grid areas were identified, covering typical grid types with varying branching structures, connection densities, total line lengths, and customer counts ranging from 25 to 64. Fig. 1 shows a simplified example grid structure for terminology clarification; it does not reflect any of the grids used in this study.

For each representative grid, georeferenced load and generation profiles for households, commercial trade and services, electric charging stations, heat pumps and photovoltaic units were assigned, comprising three-phase active and reactive power time series with a one-minute resolution according to [13]. Photovoltaic units were modeled using a $Q(V)$ control strategy, which is expected to become the predominant approach in future low-voltage operation.

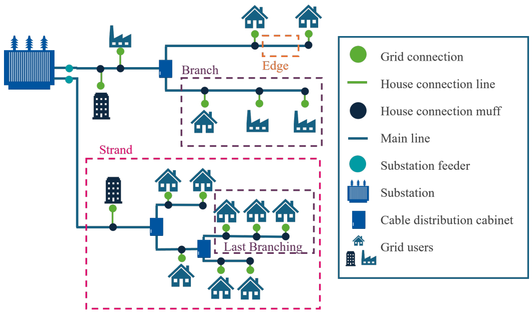


Fig. 1: Exemplary grid structure with explanations of terms Edge, Branch, Strand, Last Branching.

TABLE I: Measurement uncertainties and resolutions according to [17].

Quantity	Uncertainty (σ)	Resolution
Voltage (V)	1 %	0.1 V
Current (I)	1 %	0.01 A
Active power (P)	1 %	0.1 W
Reactive power (Q)	2 %	0.1 Var

Finally, three-phase power flow simulations were performed to generate synthetic voltage, current, and power measurements at one-minute intervals for all nodes within each grid.

To realistically represent the characteristics of measurement equipment, the corresponding measurement uncertainties were included. Gaussian-distributed noise with a standard deviation σ was added to the synthetic measurement data to model statistical uncertainty. According to [16], the manufacturer-specified measurement uncertainty corresponds to the maximum permissible error, which represents a 3σ interval. Consequently, the standard deviation σ was derived from one third of the specified maximum error. In addition, the synthetic measurement values were rounded to the instrument-specific resolution. In compliance with German regulatory requirements, the uncertainties and accuracies according to Table I are applied for smart meters. Thus, a complete data set is available, including the original grid model, for topology identification.

B. Generation of an Initial Grid Model via Voltage Correlation Analysis

Based on the synthetic voltage measurement data, an initial estimate of the grid topology is obtained through voltage correlation analysis combined with a modified maximum spanning tree approach, as described in [9]. In this reconstruction step, only nodal voltage measurements and node degree constraints are used.

Pairwise correlations between nodes are quantified using the Kendall rank correlation coefficient, which is particularly suitable for detecting non-linear relationships and exhibits robustness against outliers. For two nodes u and v with voltage time series $x_i(u)$ and $x_i(v)$, the Kendall coefficient τ_{uv} is determined according to (1), where n denotes the total number of measurement samples:

$$\tau_{uv} = \frac{2}{n(n-1)} \sum_{i<j} \operatorname{sgn}\big(x_i(u) - x_j(u)\big) \operatorname{sgn}\big(x_i(v) - x_j(v)\big). \tag{1}$$

The resulting correlation matrix forms the basis for constructing the initial grid model using a modified maximum spanning tree algorithm. To ensure physically plausible grid structures, a modified Kruskal algorithm according to [9] is used that enforces lower and upper bounds on the node degree. These limits reflect known structural constraints, such as the typical number of outgoing feeders at substations, the number of branches in cable distribution cabinets, or the single connection at terminal nodes (e.g., households). Edges that violate degree limits or create cycles are omitted. Where necessary, local tree segments are iteratively adjusted to meet the remaining degree constraints.

C. Genetic Algorithm Enhanced Power-Based Branch Refinement

After the initial grid topology has been reconstructed from voltage correlations, the model is further validated and refined using power measurement data. The procedure follows the general concept introduced in [9], which compares the aggregated apparent power of all nodes within a branch to the corresponding measured feeder power at the substation or cable distribution cabinet. In contrast to the previous approach relying on random node permutations, the refinement process is here replaced by a GA to achieve an efficient optimization.

A GA is a population-based stochastic optimization method inspired by natural evolution, evolving candidate solutions through selection, crossover, and mutation [18]. Each individual corresponds to a feasible assignment of nodes to branches within the grid. The goal is to minimize the mismatch between the measured feeder power, at the cable distribution cabinet or substation supplying the respective branch, and the sum of the apparent powers of all nodes assigned to that branch.

The quality of each individual $\mathcal{X}$ is quantified by a fitness function, which measures the mean apparent power deviation per branch. For each branch s, the measured feeder power $S^{(s)}_{\text{feeder}}(t)$ at time t is compared with the aggregated nodal power $\sum_{i\in s(\mathcal{X})} S_i(t)$ of all nodes assigned to this branch. The overall objective is therefore formulated according to II-C, where T is the number of time samples considered, $S_i(t)$ denotes the apparent power of node i at time t, and S is the set of all branches.

$$f(\mathcal{X}) = \sum_{s\in S} \left| \frac{1}{T} \sum_{t=1}^{T} \left(S^{(s)}_{\text{feeder}}(t) - \sum_{i\in s(\mathcal{X})} S_i(t) \right) \right| \tag{2}$$

The vector $\mathcal{X} = [x_1, \ldots, x_{N_{\text{nodes}}}]$ defines the branch assignment of each node, with x_i representing the branch to which node i belongs. The aim of the GA is to find the assignment $\mathcal{X}$ that minimizes $f(\mathcal{X})$, thereby ensuring consistency between the reconstructed topology and the measured power flows.

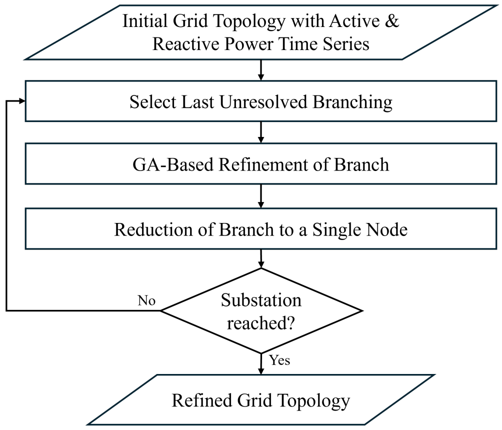


Fig. 2: Overview of the wrapper function.

Due to the definition of the fitness function as described above, the prior consideration of the node degree is essential to ensure a correct fundamental structure of the initial grid topology, particularly with respect to the number of branches. Preliminary studies have shown that, without taking the node degree into account, a correct fundamental structure of the reconstructed grid topology cannot be ensured [9].

Wrapper Function for Recursive Refinement: Before the GA-based optimization is executed, a hierarchical wrapper is applied to determine where the algorithm should be used within the grid structure. An overview of the hierarchical wrapper is shown in Fig. 2 This wrapper controls the recursive correction process and ensures that the GA operates locally and efficiently in the relevant parts of the grid.

First, all strands exhibiting significant deviations between the measured feeder power and the sum of the assigned nodal powers are identified as containing potentially misassigned nodes. Within each affected strand, the last branching is detected and analyzed first. If incorrect node assignments are present within this branch, the GA is applied locally to determine an improved configuration. The candidate nodes considered for reassignment include all nodes belonging to branches that currently contain misallocated nodes.

The GA is executed iteratively until the correct assignment for the current branch is found, with each iteration using the improved result from the previous one as the new initial solution. Once a branch has been successfully corrected, it is aggregated into a single equivalent node, reducing the structural complexity of the grid. Subsequently, the next higher branch is examined, and if any misassigned nodes are detected, the GA is applied again in the same manner. This recursive process continues until all branches within a strand have been analyzed and corrected. After a strand has been fully validated, the procedure is repeated for all remaining strands in the grid.

Refinement Process for Single Branches: The GA starts from an initial population that is derived from the initial estimated topology. Within the branch under consideration, targeted random variations are introduced: entire branches can be exchanged, individual nodes can be swapped between branches, or single nodes can be moved to different branches.

During each generation, all individuals are evaluated using the fitness function. During selection, the fittest individuals are automatically preserved for the next generation, whereas the remaining parent pairs are chosen probabilistically according to their fitness values to balance exploitation and exploration. This mechanism ensures that fitter individuals have a higher likelihood of being selected for reproduction, while weaker ones still contribute to maintaining diversity in the population.

New individuals are created through crossover and mutation. The crossover operator recombines structural information from two parents: either entire branches are exchanged or individual nodes are swapped between parents. The mutation operator introduces random variations to prevent premature convergence. Mutations are applied either by exchanging nodes between branches or moving nodes to other branches. Both operators are applied with adaptive probabilities: when no improvement is observed over several generations, the mutation rate is increased to encourage exploration; when fitness improves steadily, crossover is favored to exploit good configurations.

The iterative process continues until either a predefined maximum number of generations $k_{\max}$ is reached or the fitness value $f(\mathcal{X})$ falls below a threshold corresponding to the expected average technical grid losses. The latter criterion is considered sufficient, since a perfect match between measured and reconstructed power values is not physically achievable due to measurement uncertainties and inherent electrical losses. The refinement process thereby ensures that reconstructed topologies remain physically consistent and robust, even under realistic measurement uncertainty levels. The recursive, branch-wise application of the GA limits the search space to locally inconsistent strands, resulting in a significantly lower computational effort than global permutation-based refinement.

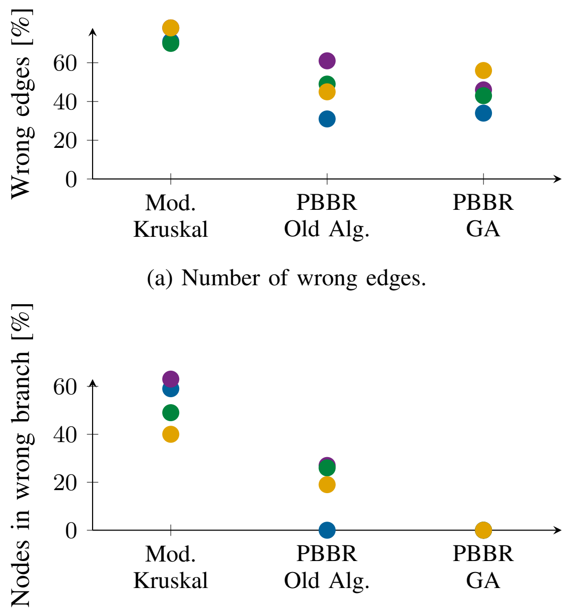
D. Evaluation and Definition of the Error Measure

In order to subsequently evaluate the reconstructed topology, the error rate of the incorrectly reconstructed edges is taken into account. The second evaluation measure is the share of nodes that are assigned to a wrong branch. This provides information about the quality of the general reconstruction of the grid structure and the determination of switching states in cable distribution cabinets.

III. Results

In the following, the results of the GA-enhanced topology identification are presented. The analysis is based on four representative grid areas, each equipped with supply tasks corresponding to a typical summer week. This setup captures the potential influence of photovoltaic units and their reactive power behavior. The synthetic measurement values were generated with measurement uncertainties according to the standard case for smart meters.

To evaluate the performance of the GA-enhanced power-based branch refinement, the results are compared with those obtained from the modified Kruskal algorithm, based solely on voltage measurements and node degree constraints, as well as from the previous power-based branch refinement approach, which was based on random permutation of nodes.



(a) Number of wrong edges.

(b) Number of nodes assigned to wrong branch.

Fig. 3: Evaluation measures of the topology identification using the modified Kruskal algorithm, the power-based branch refinement (PBBR) using the old algorithm, based on random permutation of grid nodes, and the GA-based approach. The error rates of the four representative grids are marked in color.

Fig. 3 shows the topology identification error rates obtained at different algorithmic stages, distinguishing between the share of incorrectly reconstructed edges and the number of nodes assigned to incorrect branches. The results in Fig. 3a show a clear improvement when the power-based branch refinement step is applied after the modified Kruskal algorithm. However, no significant difference can be observed between the old and the GA-enhanced algorithm in terms of the number of wrongly reconstructed edges. Since node order within a branch is randomized, the edge-based error rate is less informative. Therefore, Fig. 3b shows the number of nodes assigned to the wrong branch. Here, a distinct improvement can be observed when including the power-based branch refinement step. In particular, the GA-enhanced algorithm achieves a complete correction of the node-to-branch assignments in all four test grids, whereas the old approach fails to reach full consistency, especially in grids with a high number of branches. These results indicate that the GA-enhanced refinement effectively compensates for local misassignments caused by correlation-based initialization, particularly in grids with similar voltage profiles. Moreover, due to its population-based optimization strategy, the GA shows higher robustness against measurement uncertainties.

979-8-3315-8138-1

IV. Conclusion and Outlook

This paper presented a two-stage topology identification framework for low-voltage grids, combining voltage correlation based identification with a power-based branch refinement using a recursive genetic algorithm. The proposed method achieves a physically consistent and robust identification of node-to-branch assignment, even under realistic smart meter measurement uncertainties. Compared to randomly permuted refinement approaches, the recursive GA application significantly improves reconstruction accuracy while maintaining computational efficiency.

Future work will adress the integration of geocoordinates and current measurements to reliably determine both the order of nodes within each branch and the corresponding line parameters.

Acknowledgment

Federal Ministry for Economic Affairs and Energy

This project received funding from the German Federal Ministry for Economic Affairs and Energy under the agreement no. 03EI4048Q (Quirinus Control).

References

[1] D. Deka, M. Chertkov, and S. Backhaus, "Topology estimation using graphical models in multi-phase power distribution grids," *IEEE Transactions on Power Systems*, vol. 35, no. 3, May 2020. DOI: 10.1109/tpwrs.2019.2897004.

[2] Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, "Bericht zum zustand und ausbau der verteilernetze 2022," Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, Bonn, Germany, Tech. Rep., 2023.

[3] D. Deka, V. Kekatos, and G. Cavraro, "Learning distribution grid topologies: A tutorial," *IEEE Transactions on Smart Grid*, May 2023. DOI: 10.1109/tsg.2023.3271902.

[4] C. Deng and X. Chu, "Topology estimation of power distribution grids using graphical model," in *2020 IEEE/IAS Industrial and Commercial Power System Asia (I& CPS Asia)*, 2020.

[5] H. Liang, L. Tong, and X. Zou, "A data-driven topology estimation for distribution grid," in *2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA)*, 2021. DOI: 10.1109/ICIEA51954.2021.9516040.

[6] A. Benzerga, D. Maruli, A. Sutera, A. Bahmanyar, S. Mathieu, and D. Ernst, "Low-voltage network topology and impedance identification using smart meter measurements," in *2021 IEEE Madrid PowerTech*, 2021, pp. 1–6.

[7] F. Tischbein, D. Kubon, M. Stroot, and A. Ulbig, "Influence of measurement uncertainties on the quality of grid topology identification in three-phase systems," in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, 2024, pp. 1–6.

[8] R. Leenknecht, H. Ponsaerts, M. Numair, and D. Van Hertem, "Distribution network topology identification via variance analysis of smart meter data," in *2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, 2025, pp. 1–5. DOI: 10.1109/ISGTEurope64741.2025.11305291.

[9] F. Tischbein, N. Štumberger, and A. Ulbig, "Voltage-based topology identification in low-voltage grids using a node-degree-constraint kruskal algorithm and power-based branch refinement," in *2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, 2025, pp. 1–5. DOI: 10.1109/ISGTEurope64741.2025.11305327.

[10] Á. Rodríguez del Nozal, R. Carmona-Pardo, J. M. Mauricio, and E. Romero-Ramos, "An evolutionary computational approach for the identification of distribution networks models," *Engineering Applications of Artificial Intelligence*, vol. 137, p. 109 184, Nov. 2024, ISSN: 0952-1976. DOI: 10.1016/j.engappai.2024.109184. [Online]. Available: http://dx.doi.org/10.1016/j.engappai.2024.109184.

[11] E. Zhu and K. Chen, "Research on topology identification method for low voltage distribution area based on genetic algorithm," in *2023 International Conference on Computer Science and Automation Technology (CSAT)*, 2023, pp. 588–592. DOI: 10.1109/CSAT61646.2023.00158.

[12] B. Liu, D. Wang, Y. Li, L. Qiao, and S. Chen, "Topology identification method of distribution network based on branch active power," *Journal of Physics: Conference Series*, vol. 2108, no. 1, p. 012 062, Nov. 2021. DOI: 10.1088/1742-6596/2108/1/012062.

[13] F. Tischbein, C. Hölscher, F. Klein-Helmkamp, and A. Ulbig, "Influence of local reactive power control on the quality of grid topology identification," in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, 2024, pp. 1–6.

[14] M. Trageser et al., "Automated routing of feeders in electrical distribution grids," *Electric Power Systems Research*, vol. 211, p. 108 217, Oct. 2022. DOI: 10.1016/j.epsr.2022.108217.

[15] M. Trageser, C. M. Vertgewall, T. Offergeld, and A. Ulbig, "Analysis of future electric mobility scenarios based on synthetic, georeferenced grid models," in *CIRED Porto Workshop 2022: E-mobility and power distribution systems*, 2022. DOI: 10.1049/icp.2022.0764.

[16] *DIN EN 60359:2002-09, Electrical and Electronic Measurement Equipment - Expression of the Performance*, Norm, Sep. 20002.

[17] FNN VDE, *Specification Basic Meters - Functional Properties*, Germany, May 2018.

[18] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st. USA: Addison-Wesley Longman Publishing Co., Inc., 1989, ISBN: 0201157675.