GeoTopo-Net: A GIS-Based Framework for Automated Topology Recognition and Equipment Grouping in Power Distribution Networks

Oğuz Deniz[1], Elif Rana Tekin[1], Süha Nur Arslan[1]

[1]SI GSW GIS, Siemens A.Ş., Ankara/Türkiye – oguzdeniz179@gmail.com, elifranatekin2002@gmail.com, suha.arslan@siemens.com

Keywords: GIS, Componentization, Network structuring, Graph-based equipment grouping, Topology recognition

Abstract

Power distribution networks are critical infrastructure, yet their effective management is limited by the lack of standardized, topologically coherent equipment representations. Existing research focuses on isolated tasks such as load profiling or failure detection without offering integrated frameworks for automated network structuring. This work introduces GeoTopo-Net, a novel multi-stage pipeline that automatically groups power distribution equipment into topologically meaningful components. The pipeline employs three key stages: Data Standardization converts diverse raw equipment into consistent geographic formats, distinguishing between linear network components and discrete devices; AI-Assisted Spatial Sampling uses density-based clustering to identify switch groups and generates localized analysis regions with accompanying spatial data, separating simple from complex equipment configurations; and Heuristic Grouping applies specialized algorithms tailored to equipment complexity. For simple busbar arrangements, the algorithm uses network traversal and geometric proximity to identify main sections and their connectors. For complex multi-switch configurations, a refined multi-phase approach systematically segments equipment at critical boundaries, builds connectivity relationships, and applies classification rules based on connection patterns. Simultaneously, a dedicated algorithm processes network's transmission lines by merging continuous segments while respecting critical equipment locations and identifying logical connection points. The framework transforms raw, heterogeneous network data into fully organized, spatially-referenced datasets—optimally structured for network optimization, outage simulation, and asset monitoring—while enabling topological correctness checks of network equipment, with particular focus on busbar architectures, to support integration with real-time SCADA systems. By systematically addressing power network topology complexities through automated analysis, GeoTopo-Net advances beyond existing approaches, providing a comprehensive foundation for intelligent grid management with modular design facilitating integration with existing geographic information systems.

1. Introduction

Power distribution networks represent the backbone of modern electrical infrastructure, connecting generation sources to end consumers through complex arrangements of transmission lines, transformers, switches, and busbar systems. As these networks evolve to accommodate renewable energy integration, smart grid technologies, and increasing urbanization demands, the need for accurate, automated topology recognition and equipment grouping has become essential. However, current network management practices are hindered by the lack of standardized, topologically coherent equipment representations that can effectively integrate with real-time operational systems such as SCADA.

Previous research in power systems has largely concentrated on isolated aspects of network management. For instance, machine learning algorithms have been extensively explored for tasks such as equipment failure detection and diagnosis (Doostan & Chowdhury, 2017; Wang et al., 2021; Jiang et al., 2024), while clustering techniques have been applied to consumer classification and load profiling (Grigoraș et al., 2023). Furthermore, advancements in image processing and deep learning have enabled intelligent identification and classification of power system devices from various data sources, including geospatial imagery and multi-modal data fusion (Wang & Meng, 2019; Su et al., 2021; Wang et al., 2021; Wang et al., 2023). These studies, while critical to their specific domains, often do not provide a comprehensive, integrated framework for automating the fundamental process of network topology recognition and equipment grouping in a GIS-centric environment.

The ability to automatically and accurately group power distribution equipment into topologically meaningful components is very important for several reasons: forms the foundation for robust network modeling for simulation and optimization, facilitates efficient outage management, and enhances the capabilities of asset health monitoring systems. Traditional manual methods for topology identification are labor-intensive, prone to errors, and struggle to keep pace with the dynamic nature and increasing complexity of modern grids. Recent efforts have started to bridge this gap by integrating GIS data with electrical measurements (Huang et al., 2022) and leveraging knowledge graphs combined with Graph Neural Networks (GNNs) for fault-tolerant topology identification (Wang et al., 2021). The emergence of graph databases also shows promise for improving real-time topology data governance in substations (Li et al., 2023). Despite these advances, a comprehensive, multi-stage pipeline that systematically addresses the challenges of data heterogeneity, spatial complexity, and diverse equipment configurations remains a critical need.

We introduce GeoTopo-Net, a novel multi-stage pipeline designed to automatically group power distribution equipment into topologically coherent components. Our framework goes beyond existing clustering and machine learning approaches by offering an integrated GIS-based solution for automated network structuring. We address the core challenge of transforming raw, often inconsistent, equipment data into a standardized, analyzable format. The pipeline's innovative features include AI-assisted spatial sampling for localized analysis of complex busbar regions—a technique inspired by object detection and clustering methods like DBSCAN (Ester et al., 1996)—and the application of specialized heuristic algorithms tailored to the unique geometric and topological characteristics of different equipment types. This systematic approach ensures that both simple and intricate busbar configurations, as well as extensive linear network components, are accurately identified and grouped.

By delivering a completely grouped and geo-referenced dataset, GeoTopo-Net establishes a structured foundation for a wide array of analytical and operational tasks, supporting scalable analyses and improved decision-making across GIS-based environments. Furthermore, it enables crucial topological correctness checks of network equipment, particularly for complex busbar architectures, enabling integration with real-time SCADA systems. As highlighted by recent reviews, there is a recognized gap in integrated frameworks combining spatial analysis with network optimization (Marković et al., 2023). GeoTopo-Net directly addresses this by systematically resolving power network topology complexities through automated analysis, thus advancing beyond current isolated solutions and providing a comprehensive foundation for intelligent grid management.

2. Methodology

Our GeoTopo-Net framework addresses the complex challenge of automated power distribution equipment grouping through a systematic multi-stage pipeline designed to transform heterogeneous network data into topologically coherent representations. The methodology integrates spatial analysis techniques with graph-theoretic algorithms, leveraging the complementary strengths of geometric proximity analysis and network connectivity reasoning. Each stage of the pipeline is detailed below, from initial data harmonization to the final integration of grouped components.

The proposed methodology operates on the principle that power distribution networks exhibit hierarchical spatial organization, where local equipment configurations (particularly around switching stations and busbar arrangements) require specialized analysis approaches distinct from global network connectivity patterns. The pipeline architecture reflects this insight through its modular design, enabling targeted processing strategies for different equipment types and complexity levels. Complete workflow consists of five interconnected stages: (1) Data Standardization transforms diverse input formats into consistent geometric representations; (2) AI-Assisted Busbar Sampling identifies and isolates localized analysis regions using density-based clustering; (3) Heuristic Busbar Grouping applies specialized algorithms tailored to junction complexity; (4) Additional Line Grouping processes network transmission components while respecting critical equipment boundaries; and (5) Integration and Output synthesizes results into unified, topology-aware datasets. Each stage incorporates validation mechanisms to ensure topological consistency and geometric accuracy throughout the processing chain.

2.1 Data Standardization

The foundational stage of the pipeline addresses the critical challenge of data heterogeneity inherent in raw GIS exports from utility databases. Raw data often lacks a consistent schema, with equipment represented using a variety of geometric types and attribute conventions. The primary objective of this stage is to homogenize this diverse input into a predictable and reliable format for all subsequent analyses.

2.1.1 Geometric Alignment Process:

The pipeline begins by converting diverse power network datasets—from utility asset management systems, GIS, or field surveys—into a consistent GeoJSON framework, with all geometry information stored in WKT format (Herring et al., 2011). Mixed geometry types and inconsistent attribute schemas are unified through targeted conversions: linear elements such as transmission lines, feeders and busbars are rendered as precise **LineString** features (with multipart segments preserved), while discrete devices like switches, transformers and regulators become **Point** features positioned at their true electrical connection nodes rather than geometric centroids.

This approach not only ensures spatial accuracy and topological integrity for all downstream connectivity analyses, but also automates adaptability to new or unforeseen equipment types: any future device class—whether an additional transformer subtype or an entirely novel asset like capacitors—will be classified automatically as a Point feature (or LineString if its geometry is linear), allowing users to extend the network model without altering the core algorithm.

2.1.2 Attribute Schema Normalization:

Concurrent with geometric standardization, the framework implements attribute schema normalization to establish consistent data access patterns for downstream processing stages. Equipment identifiers are validated and standardized according to utility naming conventions, while equipment types are mapped to a controlled vocabulary that enables reliable algorithmic classification. Voltage levels, operational states, and connectivity information are parsed and validated to ensure data quality throughout the pipeline. This preprocessing step reduces computational complexity in subsequent stages by consolidating clearly related components while preserving individual feature geometry for detailed analysis.

2.2 AI-Assisted Busbar Sampling

Busbar systems, particularly within substations, represent the most topologically complex regions of a power network. They are characterized by a high density of interconnected equipment in a small geographic area. To manage this complexity, we employ a spatial sampling strategy that isolates these dense regions for focused analysis, a technique analogous to region proposal networks in computer vision. Figure 1 shows the raw QGIS representation of an example substation's busbar layout.

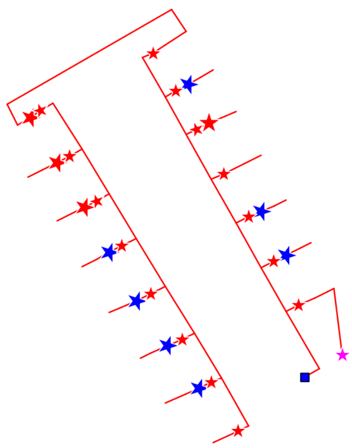

Figure 1. An example of a substation busbar layout in QGIS

The process is initiated by applying the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to the standardized point data. The clustering is performed specifically on switches that are not located directly on main transmission lines, as these are strong indicators of complex junctions or busbar configurations. DBSCAN is chosen for its ability to identify arbitrarily shaped clusters without requiring a predefined number of clusters, making it ideal for the irregular layouts of power equipment.

2.2.1 Clustering-Based Region Identification: The busbar sampling stage localizes analysis through systematic identification of switching equipment clusters that correspond to individual busbar regions. The process operates on switch locations extracted from the standardized point equipment dataset, applying DBSCAN clustering with parameters

$$\varepsilon = \frac{d}{2} \quad \text{and} \quad \text{min\_samples} = 1, \tag{1}$$

where d represents the default grid size in coordinate units. The clustering algorithm processes switch coordinates

$$\mathbf{S} = \{(x_i, y_i)\}_{i=1}^{n} \tag{2}$$

to identify coherent groupings that represent individual switching stations or busbar configurations. Each resulting cluster C_j defines a local analysis region characterized by its centroid $\bar{\mathbf{c}}_j$ and spatial extent.

2.2.2 Adaptive Tile Generation: For each identified cluster C_j, the framework generates a localized analysis window via adaptive tile extraction:

$$\text{Tile}(C_j) = \begin{cases} \text{Box}\big(\bar{c}_{x,j} \pm \frac{d}{2},\, \bar{c}_{y,j} \pm \frac{d}{2}\big), & |C_j| = 1, \\ \text{Box}\big(\bar{c}_{x,j} \pm \frac{s}{2},\, \bar{c}_{y,j} \pm \frac{s}{2}\big), & |C_j| > 1, \end{cases} \tag{3}$$

where

$$s = \max\big(\text{width}(C_j),\, \text{height}(C_j),\, d\big) + 0.1\, d \tag{4}$$

is the side length of the square tile (the larger of the cluster's bounding-box dimensions or the default grid size d, plus a 10% margin).

The tile extraction process then produces both PNG visualizations and corresponding GeoJSON vector data for each region. All equipment—switches, transformers, and line segments—whose geometries intersect the tile boundaries are retrieved via spatial indexing for further analysis.

2.2.3 Automated Sample Classification: The framework implements automated classification to route tiles to appropriate processing pipelines based on the complexity of switching layouts. A tile is classified as a ***multi-switch busbar*** if

$$|S_{\text{tile}}| > 1, \tag{5}$$

as a ***single-switch busbar*** sample if:

$$|S_{\text{tile}}| = 1 \quad \text{and} \quad \exists\, l \in L_{\text{tile}} : \text{layer}(l) = \texttt{busbar}, \tag{6}$$

and as a *vitual node (cnode) operation window* if:

$$\begin{aligned} &|S_{\text{tile}}| = 1, \\ &\big|\{\, l \in L_{\text{tile}} :\ l \cap s \neq \emptyset,\ \text{layer}(l) \neq \texttt{busbar}\}\big| \quad \geq 2. \end{aligned} \tag{7}$$

This classification enables selective application of specialized grouping algorithms, optimizing computational efficiency while ensuring appropriate analysis depth for different busbar architectures, and this modular approach also supports concurrent processing of different busbar configurations (both single-switch and multi-switch) alongside transmission line groupings, as each follows its own dedicated algorithm. Figure 2 shows the generated version of the busbar from Figure 1, produced by the AI-assisted sampling stage.

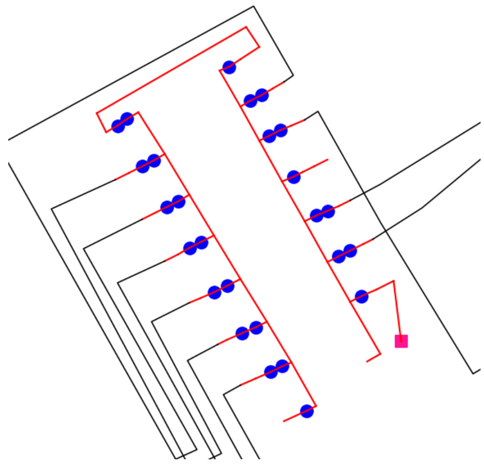

Figure 2. AI-assisted sampling of the busbar tile from Figure 1

In Figure 2, blue points indicate switch locations, the pink square marks the transformer, red lines represent busbar segments, and black lines correspond to transmission line segments. This color-coded visualization helps distinguish equipment types and network layers within the sampled busbar tile.

2.3 Heuristic Busbar Grouping

With busbar regions isolated and categorized by the AI-assisted sampling process, this stage applies two distinct heuristic algorithms to group the equipment within each sample. To optimize performance for graph-based operations, the GeoJSON data for each sample is first converted into a tabular CSV format, retaining essential attributes: **ID**, **geometric coordinates**, **type**, and **layer**.

2.3.1 Single-Switch Busbar Grouping: This algorithm is designed to process the less complex "single-switch" samples. Its logic is based on graph traversal and geometric proximity to reconstruct the busbar and its associated connectors. As detailed in Algorithm 1, the process first builds an adjacency graph from the line segments. It then systematically identifies the main busbar sections by performing a Depth-First Search (DFS) (Tarjan et al., 1972) starting from segments that are directly connected to main transmission lines (layer = 'line').

Once the primary busbar groups are established, the algorithm identifies any remaining busbar segments that are connected to the single switch but not yet grouped. These are attached to the nearest main busbar group based on a proximity threshold ϵ. Finally, any remaining segments that connect only to switches are grouped together as "connectors", representing the functional links within the junction.

Algorithm 1 Single-Switch Busbar Grouping

Require: L: set of line segments
Require: S: set of switches
Require: ϵ: distance threshold
Ensure: B: busbar groups, C: connector groups
Initialize $B \leftarrow \emptyset, C \leftarrow \emptyset$
Mark all $l \in L$ as unvisited
Build adjacency map from nodes to their incident lines
{Stage 1: Extract contiguous busbar segments}
for all $l \in L$ unvisited **such that** layer(l) = `busbar` **and** l adjacent to any non-busbar line **do**
  $G \leftarrow \emptyset$
  DFS_Busbarl, G
  $B \leftarrow B \cup \{G\}$
end for
{Stage 2: Attach stray segments to nearest busbar group}
for all $l \in L$ unvisited **such that** layer(l) = `busbar` **and** one endpoint in S **do**
  if $\min_{G \in B} \text{dist}(\text{other\_pt}(l), \text{centroid}(G)) \leq \epsilon$ **then**
    Attach l to nearest G, mark l visited
  end if
end for
{Stage 3: Identify connector groups}
for all $l \in L$ unvisited **such that** l intersects some switch in S but no non-busbar line **do**
  $G \leftarrow \emptyset$
  DFS_Connectorl, G
  $C \leftarrow C \cup \{G\}$
end for
DFS_Busbarl, G
Mark l visited; add l to G
for all neighboring $l' \in L$ unvisited **where** layer(l') = `busbar` **and** l' not touching any switch **do**
  DFS_Busbarl', G
end for
DFS_Connectorl, G
Mark l visited; add l to G
for all neighboring $l' \in L$ unvisited **where** l' does not touch any non-busbar line **do**
  DFS_Connectorl', G
end for

This combined graph-traversal and proximity analysis yields robust, topology-preserving busbar and connector groupings for downstream processing, as illustrated in Figure 3.

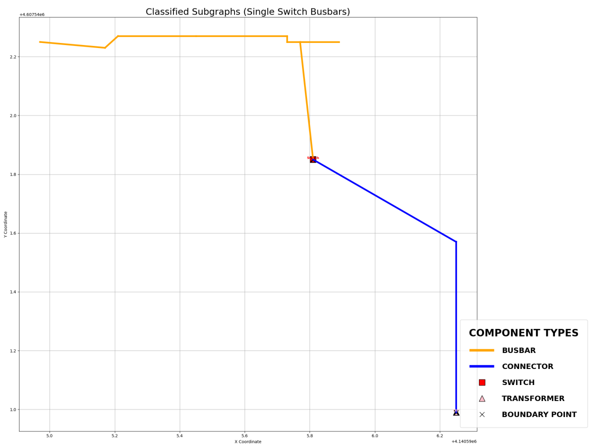

Figure 3. Example grouped output of a single-switch busbar grouping algorithm.

After each busbar and connector group is identified, unique IDs are generated and assigned—busbar IDs, connector IDs, switch IDs, and transformer IDs—via dedicated ID-generation functions to ensure traceability throughout the pipeline.

2.3.2 Multi-Switch Busbar Grouping:

For the more complex "multi-switch" samples, a more sophisticated, multi-phase heuristic is required. This algorithm, outlined in Algorithm 2, is designed to deconstruct and classify the complex web of connections.

Algorithm 2 Multi-Switch Busbar Grouping

Require: E: equipment segments, S, T: switch and transformer locations, α: size threshold
Ensure: Groups $\{H_i\}$ labeled {`busbar`, `cnode`, `connectork`, `Point`}
Build graph $G = (V, E)$ on segment endpoints; let C_0 be the largest component restricted to layer=`busbar`
$V_b \leftarrow \{v \in V : v$ incident to any $s \in S \cup T\}$
Split C_0 at each $v \in V_b$ into subgraphs $\{H_i\}$
Build intersection graph I on nodes H_i, linking H_i, H_j if they share any $v \in V_b$
for all H_i **do**
  $k_i \leftarrow |\{\text{switch-endpoints in } H_i\}|$
  if $k_i = 2$ **then**
    label$(H_i) \leftarrow$ `cnode`
  else
    label$(H_i) \leftarrow$ `unclassified`
  end if
end for
for all H_i with label=`unclassified` and $k_i \geq 2$ **do**
  label$(H_i) \leftarrow$ `potential busbar`
end for
for all H_i with label=`cnode`, $\deg_I(H_i) = 2$, and both neighbors are potential busbars **do**
  label$(H_i) \leftarrow$ `connectork`
end for
for all H_i with label=`potential busbar` **do**
  label$(H_i) \leftarrow$ `busbar`
end for
$m \leftarrow$ mean size of all busbar H_i
for all H_i with label=`cnode` and $|H_i| > \alpha\, m$ **do**
  label$(H_i) \leftarrow$ `busbar`
end for
Label any remaining single-segment unclassified H_i as `Point`

return $\{H_i\}$ with assigned labels

The multi-switch grouping algorithm adopts a multi-phase, graph-based strategy to segment and classify complex busbar configurations. It begins by constructing a graph of all busbar-layer segments in the analysis region, focusing on the largest connected component to target the primary structure. Segment endpoints and their associations with switches and transformers are explicitly tracked to support accurate segmentation at these boundary points. The network is then split at each boundary vertex, producing subgraphs that potentially represent busbar sections, connectors, or specialized components.

To analyze interactions between these subgraphs, an intersection graph is built where nodes correspond to subgraphs and edges indicate shared boundary equipment. This enables classification based on connectivity and topological patterns. The classification proceeds in multiple phases, first labeling simple connector nodes (with exactly two switch endpoints), then identifying multi-connected busbar candidates and refining roles using intersection graph analysis. Size-based rules are applied last to reclassify large cnodes as busbar segments, reflecting their operational significance. The result of this multi-phase classification process is illustrated in Figure 4, showing a grouped output of the multi-switch busbar algorithm corresponding to the configuration in Figure 1.

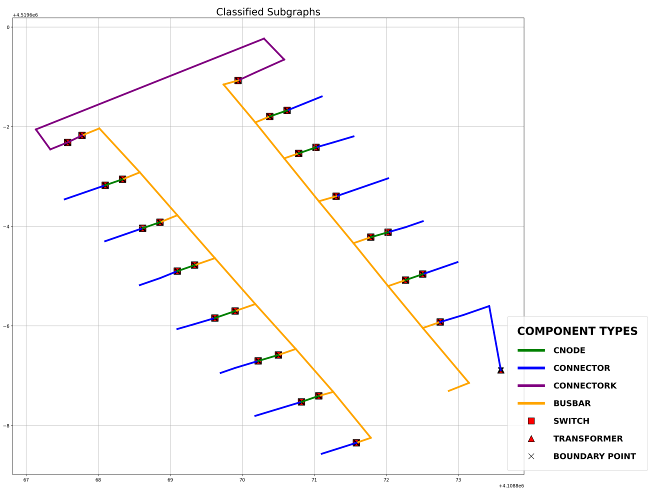

Figure 4. Example output of the multi-switch busbar grouping algorithm, corresponding to the busbar shown in Figure 1.

In cases where multiple distinct busbar structures fall within the same sampled grid, our connected component–based logic is capable of accurately distinguishing and grouping them into separate units. This ensures that each busbar is processed independently, preserving topological correctness even in dense or overlapping configurations. Figure 5 illustrates such a scenario, where multiple busbars within a single grid cell are successfully identified and grouped.

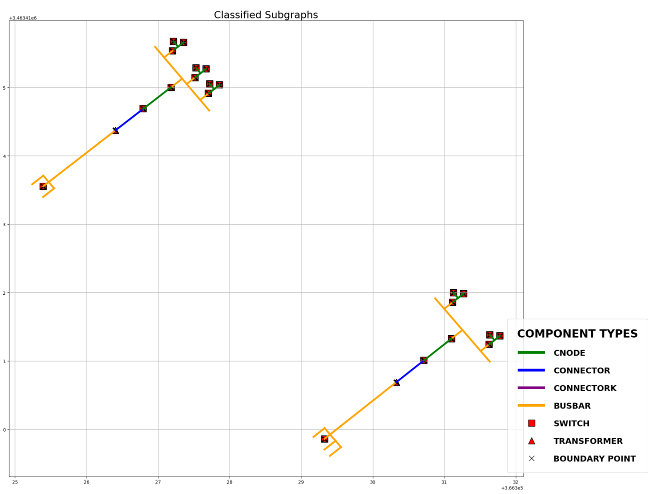

Figure 5. Grouped output of multiple busbars within a single grid

2.3.3 Transmission Line Grouping:

While previous stages focus on dense busbar regions, this stage addresses the extensive linear components of the network—the transmission and distribution lines that were not part of the busbar samples. The goal is to merge fragmented line segments into continuous, electrically contiguous groups while respecting the topological boundaries imposed by network equipment.

The algorithm organizes transmission and distribution segments outside localized busbar regions by merging geometrically continuous lines into coherent operational units while respecting logical boundaries created by network equipment. A central aspect is identifying "blocking points"—such as switches or transformers—where line merging is halted to maintain the network's operational integrity.

Our heuristic uses an iterative merging strategy, prioritizing segments with shared endpoints and compatible attributes, while preventing merges across blocking points or complex junctions. Once merging is complete, the system classifies connection points into two types: CNODEL, which represent intersections involving critical equipment, and CLINE, which represent purely topological junctions. These virtual nodes ensure a clear and analyzable representation of both infrastructure and its operational boundaries. This algorithm is detailed in Algorithm 3.

Algorithm 3 Transmission Line Grouping

Require: L: line segments, B: blocking equipment (e.g., switches, transformers), optional flag: `allowed_to_merge`
Ensure: Grouped lines G, connection points: CNODEL and CLINE
  Extract blocking coordinates P_b from B
  Mark all $l \in L$ as `unmerged`
  Initialize spatial index on L
  $G \leftarrow \emptyset$
  // Phase 1: Line Grouping new merges occur
  for all $l_i \in L$ where l_i is `unmerged` **do**
    for all l_j sharing an endpoint with l_i **do**
      if GIS_ID(l_i) = GIS_ID(l_j) **then**
        Merge l_i, l_j into group g; mark merged shared endpoint $\notin P_b$ **and** degree at point = 2
        Merge l_i, l_j into group g; mark merged `allowed_to_merge` **and** shared endpoint $\notin P_b$
        Merge l_i, l_j into group g; mark merged
      end if
    end for
  end for
  Generate new IDs and add merged groups to G
  // Phase 2: Connection Point Identification
  for all endpoints p shared by ≥ 2 lines in G **do**
    if $p \in P_b$ **then**
      Create 2 CNODEL features at p
      Connect each intersecting line and blocking equipment ID to CNODELs
    else
      Create 1 CLINE feature at p
      Connect each intersecting line to CLINE
    end if
  end for

  return G with associated CNODEL and CLINE features

As the final component of our grouping framework, this modular approach enables the entire power network to be grouped systematically by layers and by individually clustered equipment. Such modular grouping facilitates the localization and diagnosis of potential data errors, improving both the accuracy and maintainability of the network model.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W18-2025-103-2026 |

2.4 Integration and Output

The GeoTopo-Net pipeline's final stage integrates all data into a unified, comprehensive dataset, prioritizing geometric accuracy and topological coherence. This involves validating grouped components and cross-referencing connections to accurately represent the network. The integrated data is then exported as standardized GeoJSON with detailed attribute schemas, including group type and processing metadata. This ensures compatibility and maintains geometric precision for downstream applications.

Finally, a quality assurance process validates topological consistency, accuracy, and completeness. Automated checks identify issues, generating reports that document results and aid future optimization.

3. Results

To evaluate the effectiveness of the GeoTopo-Net pipeline, we conducted comprehensive testing on three power distribution networks of varying scales from different countries. Due to confidentiality requirements, specific geographic locations and utility details cannot be disclosed; however, all test networks represent middle-voltage power distribution systems containing diverse busbar configurations and substation architectures typical of modern electrical infrastructure.

3.1 Test Datasets

The evaluation encompassed three distinct network scales to assess scalability and robustness across different operational contexts. Small network covers 132.6 km² and comprises 1,588 equipment assets, including 1 multi-switch busbar and 81 single-switch busbars. The medium network spans 1359.251 km² with 93,308 equipment assets, containing 3,280 multi-switch busbars and a corresponding number of single-switch configurations. The large network covers 53732.433 km² and contains 97,967 equipment assets, including 3,209 multi-switch busbars and 11,923 single-switch busbars. This range of network sizes enables comprehensive assessment of the pipeline's performance across varying infrastructure complexities and geographic scales.

3.2 Assessment Method

Accuracy measurements rely on systematic rule-based validation rather than manual inspection, ensuring consistent and scalable assessment criteria. For single-switch busbars, validation confirms the presence of required components: busbar segments, connector elements, and switch components, with transformers being optional. Additionally, the validation process verifies that virtual busbar components are generated with geometries matching the grouped busbar to ensure proper intra-busbar connectivity representation.

Connectivity validation forms a critical component of the assessment process, ensuring no interruptions occur in both intra-busbar connections and busbar-to-transmission line interfaces. The same validation framework applies to multi-switch busbars, with accuracy measurements calculated as the average performance across all identified components and connectivity relationships.

3.3 Performance Metrics and Evaluation

GeoTopo-Net demonstrated high accuracy across all major equipment categories. For single-switch busbar configurations, pipeline achieved **100% accuracy** in both grouping and component identification across all test networks. Similarly, transmission line grouping maintained **100% accuracy** for all tested systems, successfully merging contiguous line segments while respecting topological boundaries imposed by network equipment.

Multi-switch busbar processing achieved greater than **90% accuracy** across all test networks, representing robust performance given the inherent complexity of these configurations. However, the analysis revealed specific limitations for overcomplicated stations containing more than 3-4 interconnected busbars with coupling connectors and exceeding 40 switches. Such configurations, while representing a very small fraction of typical middle-voltage network infrastructure, occasionally require manual validation to ensure complete topological correctness.

Network	Area (km^2)	Assets	Multi/Single-switch Busbars
Small	132.60	1 588	1 / 81
Medium	1 359.25	93 308	3280 / 0
Large	53 732.43	97 967	3209 / 11923

Table 1. Network characteristics of the three test systems

Network	Single/Line grouping (%)	Multi-switch (%)	Time
Small	100/100	>95	46s
Medium	100/100	>92	253s
Large	100/100	>90	983s

Table 2. Accuracy and runtime of GeoTopo-Net across the three networks

Our pipeline demonstrates favorable runtime performance, exhibits near-linear scaling, processing the small, medium, and large networks in 46 seconds, 4 minutes 13 seconds, and 16 minutes 38 seconds, respectively. All tests were conducted on an **Intel Core i5-10310U CPU @ 2.2GHz with 16GB RAM (Windows 11)**. As an example of the method's precision, Figure 6 shows a multi-switch busbar from the large network that was grouped with 100% accuracy.

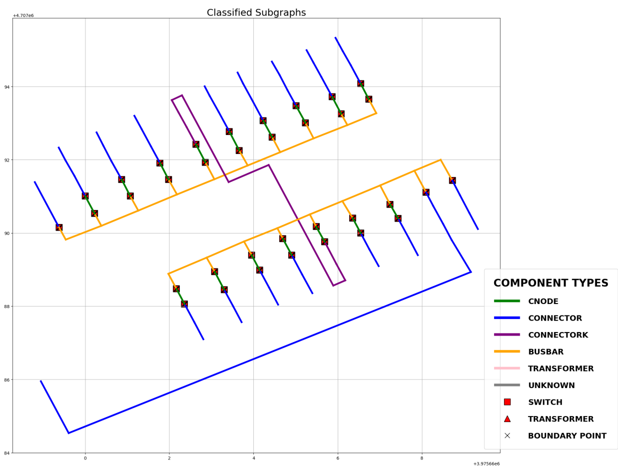

Figure 6. Example of a randomly chosen multi-switch busbar from the large network, grouped and identified with 100% accuracy.

While GeoTopo-Net performs robustly across varied networks, rare edge cases—such as highly interconnected multi-switch stations—can challenge its heuristic grouping and point to opportunities for future refinement. Likewise, the rule-based validation ensures consistency but may miss specialized equipment nuances, although such deviations are infrequent in MV systems. Figure 7 illustrates one overcomplicated multi-switch busbar from the large network, where sub-optimal grouping performance is observed.

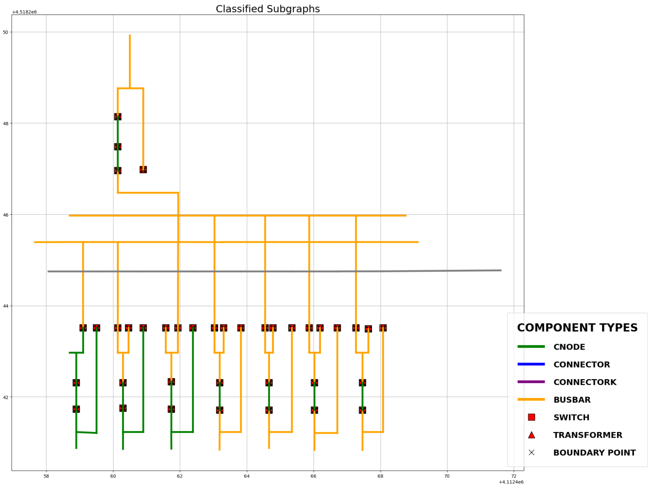

Figure 7. Sub-optimal grouping of an overcomplicated large-network multi-switch busbar

4. Conclusion

We introduced GeoTopo-Net, a novel multi-stage pipeline that successfully automates the topological grouping of power distribution equipment. By systematically integrating data standardization, AI-assisted spatial sampling, and specialized heuristic algorithms, our framework transforms raw, heterogeneous GIS data into a standardized, analyzable network model. This automates a critical and traditionally manual task, providing a foundational data layer for advanced applications like network optimization, outage simulation, and SCADA integration.

Our evaluation on real-world networks confirmed the framework's robustness and efficiency. The pipeline achieved 100% accuracy on transmission lines and single-switch busbars and over 90% on complex multi-switch configurations, all while demonstrating scalable performance on large datasets. While highly effective, we acknowledge that the heuristic could be refined for a small number of "overcomplicated" busbar systems. We plan to investigate these edge cases through the incorporation of Graph Neural Networks (GNNs), and to extend the framework to low-voltage networks or real-time data integration. In summary, GeoTopo-Net represents a significant advancement in automated power network structuring. It provides a reliable, scalable, and accurate method for establishing network topology, laying the groundwork for more intelligent, resilient, and efficient power systems.

5. References

1. Doostan, Milad, and Badrul H. Chowdhury. "Power distribution system equipment failure identification using machine learning algorithms". *2017 IEEE Power Energy Society General Meeting*. IEEE.

2. Ester, Martin, et al. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 226-231.

3. Grigoraș, Gheorghe, et al. "Contributions to power grid system analysis based on clustering techniques." *Sensors*, 23(4): 1895.

4. Huang, Xubo, et al. "Topology Identification of Low-voltage Distribution Network Based on GIS and Voltage Waveform Correlation Analysis." *2022 2nd International Conference on Electrical Engineering and Mechatronics Technology (ICEEMT)*. IEEE.

5. Jiang, Xuebao, et al. "Busbar fault diagnosis method based on multi-source information fusion." *Frontiers in Energy Research*, 12: 1443570.

6. P. Li, Y. Tang, W. Huang, M. Duan, J. Hou and J. Zhang, "Graph-Based Automatic Identification and Correction of Bus Topology in 500kV Substations," *2023 IEEE International Conference on Advanced Power System Automation and Protection (APAP)*, Xuchang, China, pp. 184-188.

7. Marković, Marija, Matthew Bossart, and Bri-Mathias Hodge. "Machine learning for modern power distribution systems: Progress and perspectives." *Journal of Renewable and Sustainable Energy*, 15(3): 032301.

8. Ni, Qingzhong, and Hui Jiang. "Topology identification of low-voltage distribution network based on deep convolutional time-series clustering." *Energies*, 16(11): 4274.

9. Su, Shi, et al. "Intelligent power equipment identification model based on grid topology analysis."*E3S Web of Conferences*, Vol. 260. EDP Sciences.

10. Wang, Haiyan, and Fanwei Meng. "Research on power equipment recognition method based on image processing." *EURASIP Journal on Image and Video Processing*, 2019(1): 57.

11. Wang, WenBin, et al. "Power distribution equipment and defect identification technology based on deep learning."*Journal of Physics:*, Conference Series. Vol. 2030. No. 1. IOP Publishing, 2021.

12. Wang, Changgang, Jun An, and Gang Mu. "Power system network topology identification based on knowledge graph and graph neural network." *Frontiers in Energy Research*, 8: 613331.

13. Wang, Zhecheng, Arun Majumdar, and Ram Rajagopal. "Geospatial mapping of distribution grid with machine learning and publicly-accessible multi-modal data." *Nature Communications*, 14(1): 5006.

14. Herring, John. "Opengis® implementation standard for geographic information-simple feature access-part 1: Common architecture [corrigendum]." (2011).

15. Tarjan, R. E. (1972). Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2), 146–160.