Check for updates

IET Smart Grid

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

An Optimisation Algorithm for Customer Topological Paths Identification in Electrical Distribution Networks

Maurizio Vassallo[1] | Adrien Leerschool[2] | Alireza Bahmanyar[1,2] | Laurine Duchesne[2] | Simon Gerard[3] | Thomas Wehenkel[3] | Damien Ernst[1]

[1]Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium | [2]Intelligent Systems Solutions, Haulogy, Neupré, Belgium | [3]RESA, Liège, Belgium

Correspondence: Maurizio Vassallo (mvassallo@uliege.be)

Received: 12 November 2025 | **Revised:** 10 June 2026 | **Accepted:** 1 July 2026

Keywords: distribution networks | integer programming | network topology

ABSTRACT

Accurate identification of network topologies in low-voltage (LV) electrical distribution networks is critical for grid management, yet remains challenging due to incomplete geographic information system (GIS) records and limited advanced metring infrastructure (AMI) coverage. This paper proposes an ILP optimisation framework that infers customer topological paths solely from static data available in DSO databases, including the GIS coordinates of network elements and the assignment of each customer to a feeder terminal junction within the MV/LV substation. An A^* algorithm is introduced to reduce the search space to a tractable set of candidate paths, which are then optimally selected subject to network radial constraints. The framework is validated on both an academic test network and a real Belgian low-voltage network, achieving an 88% customer connection rate. A robustness analysis under corrupted feeder assignment data demonstrates a fault detection rate exceeding 95% and a false connection rate below 5% at realistic error levels.

1 | Introduction

Power distribution networks are the infrastructure responsible for delivering electricity from the transmission system to individual consumers. As distribution networks grow in complexity and scale, driven by factors, such as the accommodation of new customers and the increased integration of distributed energy resources (DERs), the need for improved network planning and operation strategies becomes essential. A significant challenge in this context is identifying the specific routes electricity takes to reach customers, a task referred to as topological path identification (TPI). Accurate knowledge of network topologies and customer paths is often unavailable to distribution system operators (DSOs) due to incomplete, outdated or erroneous data records.

Solving the TPI problem provides DSOs with a clear knowledge of the connectivity between customers and medium-voltage/low-voltage (MV/LV) substations. However, accurately identifying customer topological paths remains a challenging task due to several factors. Firstly, DSOs frequently rely on incomplete or inaccurate geographic information system (GIS) data, which may contain errors derived from outdated recordings or limitations in GPS technologies [1]. Secondly, although the integration of advanced metring infrastructure (AMI) in distribution networks has increased, many European countries still lack widespread AMI coverage [2], limiting the feasibility of measurement-based methods. Finally, scalability poses a significant obstacle; as networks grow in size and complexity, traditional approaches often fail to provide efficient and scalable solutions.

Maurizio Vassallo and Adrien Leerschool contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *IET Smart Grid* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

Despite these challenges, the accurate identification of customer topological paths is fundamental to numerous applications in modern power systems. It serves as a foundation for improving digital representations of distribution networks, supporting DSOs in performing simulations and analyses.

1.1 | Literature Review

In power distribution networks, various methodologies have been proposed to address the network topology identification or the TPI problem specifically. These methods can be characterised into two main categories: techniques that only rely on static data, such as GIS data of the elements in the network, and methods that also rely on AMI data, such as smart metre recordings. Static data-based methods generally focus on establishing relationships and connections between elements. Relevant studies include: Refs. [3–12]. For example, in refs. [3, 4] the authors identify the network topology of different networks using publicly available geographical data and connecting the elements based on their distances. In ref. [5], topology identification is performed using Euclidean distances and a breadth-first search algorithm to identify clusters, followed by refinement through connecting nodes to their most probable lines. The paper [6] presents a kd-tree algorithm for identifying clusters of elements, followed by connecting the elements within each cluster based on a predefined distance. In ref. [7], the authors present a new procedure to exploit graph theory and data structure properties to efficiently detect and correct errors in models of radial secondary systems. The paper [8] uses knowledge graphs to represent the existing relationships among elements in LV distribution networks. The knowledge graphs are then used to identify the correct network topology.

The main advantage of these methods is that they operate without relying on AMI measurements, making them suitable for DSOs with limited metring infrastructure. However, their major limitation is that GIS data can be incomplete, inaccurate or missing for certain areas of the networks.

Dynamic data-based solutions rely on AMI measurements from the network. Relevant studies include [13–24]. For example, in ref. [13], the authors use line current sensor measurements and a mixed-integer linear programming (MILP) optimisation algorithm to identify the topology of a power distribution network in California. The authors in ref. [14] introduce an iterative methodology for reconstructing network topology and cable parameters of LV three-phase networks with limited smart metre coverage. In ref. [15], the authors present a method for topology identification using a combination of knowledge graphs and graph neural networks constructed based on remote signalling and telemetry data. The authors in the paper [16] propose a similar approach using mixed-integer quadratic programming (MIQP) to identify the topology of power distribution networks using smart sensor recordings. Although AMI-based methods can estimate topology and line parameters (e.g., impedance), they often require extensive sensor coverage, limiting their applicability in networks with incomplete AMI deployment. This limitation restricts their applicability in networks with incomplete AMI coverage as evidenced by some studies using limited smart metre data [23, 25, 26].

1.2 | Paper Contribution

This paper presents an optimisation algorithm to assist DSOs in addressing the TPI problem using only static data: GIS coordinates of network elements and customer-to-feeder-terminal-junction assignments as recorded in MV/LV substation databases. Although our previous work [27] established the definition of the TPI problem and introduced the raw data transformation process, this paper introduces and evaluates a concrete computational approach for solving the problem.

Our methodology relies exclusively on GIS data of network elements and customer connections to MV/LV transformers, making it particularly suitable for networks where GIS data are incomplete or inaccurate and AMI coverage is limited. The problem is formulated as an integer linear programming (ILP) optimisation algorithm. The objective function of the optimisation problem focuses on maximising the number of customers connected to the correct MV/LV transformer, whereas some constraints ensure that the solution satisfies the DSO's expectations. The proposed methodology effectively addresses data inaccuracies and incompleteness.

Moreover, a repository with the code of the proposed methodology is released. The repository is publicly available at the following link: https://github.com/TPIproblem/OptimalTPI.

1.3 | Paper Organisation

The rest of the paper is organised as follows: Section 2 presents the definition of the power network elements considered in this work. Section 3 defines the concept of topological paths. The problem statement is presented in Section 4. Section 5 defines the methodology used to solve the problem. The methodology is applied to an academic example in Section 6 and a real Belgian power distribution network in Section 7. Section 8 concludes the paper and outlines directions for future work.

2 | Power Distribution Modelling

The following notation and definitions are consistent with those introduced in our previous paper [27].

Power distribution networks typically encompass a wide range of elements.

The set of elements $\mathcal{E}$ is defined as follows:

$$\mathcal{E} = \{e_1, \ldots, e_k, \ldots, e_{|\mathcal{E}|}\} \tag{1}$$

where $|\mathcal{E}|$ represents the total number of elements in the set $\mathcal{E}$ and a single element is denoted as e_k, with $k \in \{1, \ldots, |\mathcal{E}|\}$.

Each element can have some attributes, such as its type and coordinates. The set of all element attributes $\mathcal{A}$ is defined as follows:

$$\mathcal{A} = \{a_1, \ldots, a_{|\mathcal{A}|}\}. \tag{2}$$

A single element $e \in \mathcal{E}$ is defined as a set of attributes. Formally:

$$e = \{e.a_1, \ldots, e.a_k, \ldots, e.a_{|e|}\} \tag{3}$$

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where the dot notation is used to access the attributes of the element e, and a_k represents, for example, the type of the element e, its coordinates or other relevant properties.

Each element within the network can be categorised based on its type attribute, serving as a characteristic that distinguishes it from other elements. The set of types, denoted as $\mathcal{T}$, includes all the possible types associated with the elements in $\mathcal{E}$:

$$\mathcal{T} = \{t_1, \ldots, t_{|\mathcal{T}|}\}. \tag{4}$$

Among the typical types, there may be *customer* connection elements, MV/LV *transformer* substations.

Given an element $e \in \mathcal{E}$, its type, $t \in \mathcal{T}$, is accessed as follows:

$$t = e.\,type. \tag{5}$$

2.1 | Network Topological Functions

The *Subset*() function is used to identify all elements of a specific type. This function takes two input parameters: the element set, $\mathcal{E}$ and a type $t \in \mathcal{T}$. Formally:

$$Subset(\mathcal{E}, t) = \{e \in \mathcal{E} \mid e.\,type = t\}. \tag{6}$$

In particular, three useful subsets are defined:

$\mathtt{C}$: represents the set of customers in the network, $\mathtt{C} = Subset\,(\mathcal{E}, customer)$.

$\mathtt{T}$: represents the set of terminal elements, generally $\mathtt{T} = Subset(\mathcal{E}, transformer)$. In this paper, the terminal element is the feeder junction in the MV/LV substation, though the methodology supports any network element in this role.

$\mathtt{R}$: represents the set of all elements, excluding customers and terminal elements, $\mathtt{R} = \mathcal{E} - \mathtt{C} - \mathtt{T}$.

The *Dist*() function takes two elements $e_k, e_m \in \mathcal{E}$ as input, and it outputs a scalar value, d, representing the distance between them. The *Dist*() function is defined as follows:

$$d = Dist(e_k, e_m) = \|e_k.\,coor, e_m.coor\|_2 \tag{7}$$

where $\|\bullet\|_2$ represents the Euclidean distance and where $e_k.coor$ and $e_m.\,coor$ give the coordinates of the element e_k and e_m respectively.

The *Connections*() function identifies all elements in the set $\mathcal{E}$ that can be connected to a given element, $e \in \mathcal{E}$, considering some specific conditions. The conditions for connectivity can depend on the DSOs' requirements such as distance between the elements, the element types and so on.

Therefore, the *Connections*() function takes three inputs: $\mathcal{E}$, e and some conditions, returning a subset $\mathtt{K} \subseteq \mathcal{E}$. Formally:

$$\begin{aligned}\mathtt{K} &= \text{Connections}(\mathcal{E}, e, \text{conditions})\\ &= \{e_m \in \mathcal{E} \mid e_m \text{ can be directly connected to } e \text{ if the conditions hold}\}.\end{aligned} \tag{8}$$

3 | Topological Paths

This section serves as the foundation for understanding the concepts of hypothetical, real and estimated paths within the context of this work.

3.1 | Paths

A customer topological path is represented as an ordered list of elements:

$$p = (p_1, p_2, \ldots, p_k, \ldots, p_{|p|-1}, p_{|p|}) \tag{9}$$

where p is a generic path, p_1 represents the initial element in the path, which is a customer; p_k represents the kth element in the path, and $p_{|p|}$ represents the terminal element, generally an MV/LV transformer.

Therefore, $p_1 \in \mathtt{C}$ and $p_k \in \mathtt{R}$ with $k \in \{2, \ldots, |p| - 1\}$ and $p_{|p|} \in \mathtt{T}$.

3.2 | Hypothetical Paths

Hypothetical paths refer to routes where only the initial elements (the customers) are known, whereas the connections between intermediate elements up to the terminal point are unknown. These paths represent potential ways that electricity might follow to supply a customer.

The set of hypothetical paths, denoted as $\mathcal{H}$. Each hypothetical path, $h \in \mathcal{H}$, represents a possible supply route for a customer. The number of hypothetical paths, given the set of elements $\mathcal{E}$, is finite, and it is given by:

$$|\mathcal{H}| = |\mathtt{C}| \cdot \sum_{k=1}^{|\mathtt{R}|} {}^{|\mathtt{R}|}P_k \cdot |\mathtt{T}| \tag{10}$$

where ${}^nP_m$ is the permutation formula, which calculates the number of ways to arrange m items from a set of n items $\left({}^nP_m = \frac{n!}{(n-m)!}\right)$.

3.3 | Real Paths

The set of all real paths, denoted as $\mathcal{P}$, is a subset of the hypothetical path set $\mathcal{H}$. Each path $p \in \mathcal{P}$ represents the actual sequence of network elements that connect a customer to an MV/LV transformer.

In distribution power networks, as generally radial, the number of real paths is equal to the number of customers, since one customer is supplied by only one MV/LV transformer at any given moment. Therefore, the following condition holds $|\mathcal{P}| = |\mathtt{C}|$, where $\mathtt{C}$ represents the set of elements of type customers.

3.4 | Estimated Paths

The estimated paths, denoted by $\widehat{\mathcal{P}}$, form a subset of the hypothetical paths, $\mathcal{H}$. This set of estimated paths, $\widehat{\mathcal{P}}$, is an approximation to the set of real paths within the network, $\mathcal{P}$.

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The set $\widehat{\mathcal{P}}$ satisfies the property that the number of estimated paths in the set is equal to the number of customers considered in the network. Therefore, the following condition holds $|\widehat{\mathcal{P}}| = |\mathcal{P}| = |\mathrm{C}|$.

3.5 | Paths Visualisation

Figure 1 illustrates a simplified network, highlighting hypothetical, real and estimated paths.

As shown in Figure 1a, the data about network elements are often incomplete, making it difficult to accurately identify the customer paths.

Figure 1b shows some of the hypothetical paths. In the absence of prior information, any path could potentially represent the real one.

This highlights the challenge that, with limited information, only limited conclusions can be drawn about whether a hypothetical path corresponds to its real one. Many hypothetical paths for each customer may be available.

Figure 1c,d illustrate the real and estimated paths for each customer, respectively. Each customer is associated with exactly one real path and one estimated path, both of which are subsets of the hypothetical path set. Therefore: $\mathcal{P} \subset \mathcal{H}$ and $\widehat{\mathcal{P}} \subset \mathcal{H}$.

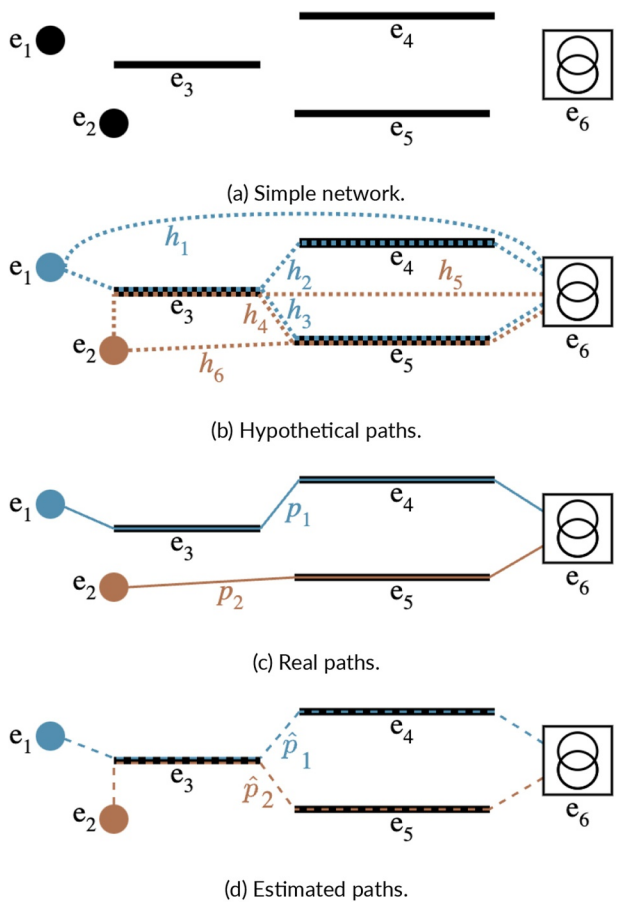


FIGURE 1 | Visualisation of the different sets of paths.

Figure 2 presents a Venn diagram illustrating the relationships among the path sets. The set of hypothetical paths, $\mathcal{H}$, encompasses all other sets, representing all possible paths within the network. Notably, the sets of real paths, $\mathcal{P}$, and estimated paths, $\widehat{\mathcal{P}}$, overlap to a certain degree, highlighting similarities and discrepancies between the estimation and ground truth. The overlap between $\mathcal{P}$ and $\widehat{\mathcal{P}}$, indicates that some estimated paths match the real ones while some others do not. For example, the paths $\widehat{p}_1$ and p_1 for customer e_1, match perfectly. However, data inaccuracies can result in discrepancies, for example, the estimated path $\widehat{p}_2$ deviates from the real path p_2 for the customer e_2.

4 | Problem Statement

Generally, DSOs possess different types of information, which we represent formally as the set $\mathcal{I}$. This data can include, for example, files about the GIS data of the network elements, customers' information such as annual power consumption, network configuration rules explaining how the elements are connected and more.

The goal of the DSOs is, given all the raw information available to them, to estimate the customer topological paths that are as close as possible to the real paths. Denoting the set of real paths as the set $\mathcal{P}$ and the set of estimated paths as $\widehat{\mathcal{P}}$, the objective is to identify two sets to be as close as possible and in the best case equal resulting in equal sets, which corresponds to a Jaccard index of 1: $J(\widehat{\mathcal{P}}, \mathcal{P}) = \frac{|\widehat{\mathcal{P}} \cap \mathcal{P}|}{|\widehat{\mathcal{P}} \cup \mathcal{P}|} = 1$.

Therefore, the problem is to identify a methodology, $\mathfrak{M}()$, that, starting from the set of raw information, $\mathcal{I}$ can identify the best approximation of the real paths $\mathcal{P}$. Formally:

$$\widehat{\mathcal{P}} = \mathfrak{M}(\mathcal{I}) \quad \text{s.t.} \quad \widehat{\mathcal{P}} \simeq \mathcal{P}. \tag{11}$$

5 | Methodology

This section provides a detailed description of the sequential procedure of the proposed methodology, $\mathfrak{M}()$, which aims to detect customer topological paths in electrical distribution networks. Figure 3 illustrates the methodology graphically.

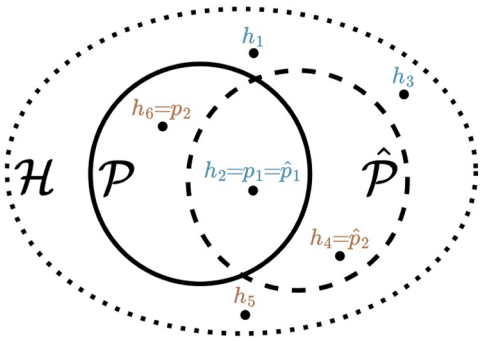


FIGURE 2 | Representation of the different sets and their relationships.

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

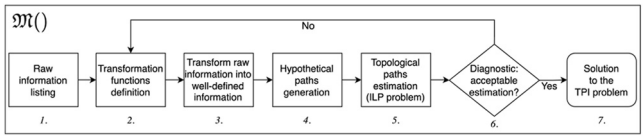


FIGURE 3 | Flowchart of the steps proposed by our methodology, $\mathfrak{M}()$, to identify the customer topological paths in electrical distribution networks.

5.1 | Raw Information Listing

Raw information refers to available information or data that comes from various sources. The set of raw information is listed as follows: $\mathcal{I} = \{i_1, ..., i_k, ..., i_{|\mathcal{I}|}\}$, with i_k $(k \in \{1, ..., |\mathcal{I}|\})$ denoting a single piece of raw information.

For this paper, we use only static data, and in particular, the data needed are the GIS coordinates of the different elements in the network, their types and the feeder terminal junction to which each customer is connected. Generally, this kind of information is available to the DSOs, even if it may be incomplete and/or inaccurate. Formally, the data needed for this paper are as follows:

- The set of elements, $\mathcal{E}$.
- The set of attributes, $\mathcal{A}$, contains at least the coordinates of the elements, the types and the feeder terminal junction to which each customer is connected.
- The set of types $\mathcal{T}$ containing at least customer, line, junction and transformer.

Therefore, *e. coor*, *e. type* must be known for all the elements, $e \in \mathcal{E}$ and *c. junction* is known for all the customers, $c \in \mathtt{C}$.

It is important to note that this formulation relies on the assumption that the feeder connection (*c.junction*) provided in the static dataset are correct. A detailed discussion of this assumption, explicitly contrasted with AMI-based methodologies, and the framework's diagnostic behaviour when these records are erroneous, is provided in Section 7.6.

5.2 | Transformation Functions Definition

Transformation functions are responsible for transforming the available raw information into clear knowledge that is relevant within the specific context of a power network. The set of transformation functions is denoted as follows: $\mathcal{F} = \{f_1, ..., f_m, ..., f_{|\mathcal{F}|}\}$, with f_m $(m \in \{1, ..., |\mathcal{F}|\})$ denoting a single transformation function.

Mathematically, the transformation functions take one piece of raw information as input and return the transformed information as output.

5.3 | Well-Defined Information

The well-defined information forms the basis for providing a structured foundation to identify customer paths that are as close as possible to reality. The set of well-defined information is denoted as follows: $\mathcal{I}' = \{i'_1, ..., i'_n, ..., i'_{|\mathcal{I}'|}\}$, with i'_n $(n \in \{1, ..., |\mathcal{I}'|\})$ denoting a single piece of well-defined information. Each element in the set $\mathcal{I}'$ is the result of applying a transformation function on a piece of raw information, that is, $i'_n = f_m(i_k)$.

5.4 | Hypothetical Paths Generation

Given the set of well-defined information, $\mathcal{I}'$, the set of hypothetical path compatible with the set $\mathcal{I}'$, denoted as $\mathcal{H}^{\mathcal{I}'}$, is constructed. Each piece of information in $\mathcal{I}'$ allows to reduce the size of the hypothetical path set $\mathcal{H}^{\mathcal{I}'}$.

The idea is that the more data are available, the closer to reality the paths inside the set $\mathcal{H}^{\mathcal{I}'}$ are.

5.4.1 | *Search Space Reduction*

Given the set of well-defined information, $\mathcal{I}'$, the framework constructs the set of compatible hypothetical paths, denoted as $\mathcal{H}^{\mathcal{I}'}$. However, enumerating the complete set $\mathcal{H}$ becomes intractable for large-scale distribution networks due to the combinatorial growth in the number of possible routing configurations.

To ensure computational tractability, we introduce an A^* algorithm as a search space reduction strategy [28]. The algorithm uses information of the starting point (the customer $c \in \mathtt{C}$) and the final point (the feeder terminal junction $t \in \mathtt{T}$) to explore valid paths.

For a given customer c and target junction t, the A^* algorithm evaluates the potential next element $e \in \mathcal{E}$ by minimising a cost function:

$$f(e) = g(e) + h(e) \tag{12}$$

where $g(e)$ represents the accumulated path length from the customer c to the current element e, defined as follows:

$$g(e) = \sum_{k=1}^{|p|-1} Dist(p_k, p_{k+1}) \tag{13}$$

where $p_1 = c$ is the customer, $p_{|p|} = e$ is the current element and each consecutive pair (p_k, p_{k+1}) is a directly connected element pair along the explored path.

The heuristic function $h(e)$ is defined as the Euclidean distance from the current element e to the target terminal junction t:

$$h(e) = Dist(e, t) \tag{14}$$

with $Dist()$ as defined in Equation (7).

During the search process, the paths are discarded by the constraints defined in $\mathcal{I}'$ (e.g., the maximum connection distance D and the maximum overall path length L). Moreover, the algorithm terminates upon identifying a maximum of N valid candidate paths per customer. By applying this search strategy, the framework generates a feasible subset of hypothetical paths $(\mathcal{H}^{\mathcal{I}'})$, enabling the ILP solver to converge efficiently.

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dgo, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5.5 | Topological Paths Estimation

After generating the hypothetical paths compatible with the well-defined information, $\mathcal{H}^{\mathcal{I}'}$, the next step of the methodology is to identify the paths in the set $\mathcal{H}^{\mathcal{I}'}$ that are an approximation of the real paths in the actual network, therefore identifying the set of estimated paths $\widehat{\mathcal{P}}$.

Although various methods can solve the TPI problem, we present an ILP approach designed to find the solution $\widehat{\mathcal{P}}$ and to meet the DSO requirements.

5.5.1 | *Matrices Generation*

The following sections describe the main components of our ILP formulation, including the matrices and the optimisation problem.

- Hypothetical paths matrix:
  The set of hypothetical paths, $\mathcal{H}^{\mathcal{I}'}$ can be represented as a binary matrix with as many rows as $|\mathcal{H}^{\mathcal{I}'}|$ and as many columns as $|\mathtt{C}| + |\mathtt{R}| + |\mathtt{T}|$.
  Therefore, the hypothetical paths matrix, denoted as $\mathbf{H}$, can be written as follows:

$$\mathbf{H} = \begin{array}{c} \\ h_1 \\ h_2 \\ \vdots \\ h_k \\ \vdots \\ h_{|\mathcal{H}^{\mathcal{I}'}|} \end{array} \begin{array}{c} \begin{array}{cccccccccccc} c_1 & c_2 & \cdots & c_{|\mathtt{C}|} & r_1 & r_2 & \cdots & r_{|\mathtt{R}|} & t_1 & t_2 & \cdots & t_{|\mathtt{T}|} \end{array} \\ \begin{pmatrix} 0 & 1 & \cdots & 0 & 1 & 0 & \cdots & 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 1 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \\ \mathbf{H}_{\mathtt{C}} \qquad\qquad \mathbf{H}_{\mathtt{R}} \qquad\qquad \mathbf{H}_{\mathtt{T}} \end{array} \tag{15}$$

  Each row h_k in the matrix $\mathbf{H}$ represents a hypothetical path compatible with the well-defined information. A value of 1 or 0 in this row indicates whether the corresponding element, $e \in \mathtt{C} \cup \mathtt{R} \cup \mathtt{T}$, is present (1) or not (0) in path h_k. A single element of the matrix is accessed as $\mathbf{H}_{(k,\,m)}$ where k represents the row $\left(k \in \left\{1,\, ...,\, |\mathcal{H}^{\mathcal{I}'}|\right\}\right)$ and m represents the column $(m \in \{1,\, ...,\, |\mathtt{C}| + |\mathtt{R}| + |\mathtt{T}|\})$.

  For clarity, we divide the $\mathbf{H}$ matrix into three sub-matrices
  - The $\mathbf{H}_{\mathtt{C}}$ matrix, denoted as customers' elements matrix, indicates the presence of a customer $c \in \mathtt{C}$ in any path, $h \in \mathcal{H}^{\mathcal{I}'}$. Therefore, the $\mathbf{H}_{\mathtt{C}}$ matrix has dimension $|\mathcal{H}^{\mathcal{I}'}| \times |\mathtt{C}|$.
    Since one and only one customer can be present in one path, the following condition must be satisfied:

$$\sum_{k}^{|\mathtt{C}|} \mathbf{H}_{\mathtt{C}\,(m,\,k)} = 1, \qquad \forall m \in \left\{1,\, ...,\, |\mathcal{H}^{\mathcal{I}'}|\right\}. \tag{16}$$

  - The $\mathbf{H}_{\mathtt{R}}$ matrix, denoted as the remaining elements matrix, indicates the presence of elements $r \in \mathtt{R}$ in the paths, $h \in \mathcal{H}^{\mathcal{I}'}$. Therefore, the $\mathbf{H}_{\mathtt{R}}$ matrix has dimension $|\mathcal{H}^{\mathcal{I}'}| \times |\mathtt{R}|$.
    No particular condition is imposed on the matrix $\mathbf{H}_{\mathtt{R}}$.
  - The $\mathbf{H}_{\mathtt{T}}$ matrix, referred to as the terminal elements matrix, indicates the presence of a terminal node, $t \in \mathtt{T}$, in the paths, $h \in \mathcal{H}^{\mathcal{I}'}$. Therefore, the $\mathbf{H}_{\mathtt{T}}$ matrix has dimension $|\mathcal{H}^{\mathcal{I}'}| \times |\mathtt{T}|$.
    Since one and only one terminal element can be present in one path, the following condition must be satisfied:

$$\sum_{k}^{|\mathtt{T}|} \mathbf{H}_{\mathtt{T}\,(m,\,k)} = 1, \quad \forall m \in \left\{1,\, ...,\, |\mathcal{H}^{\mathcal{I}'}|\right\}. \tag{17}$$

- Terminal association matrix:
  We define a binary matrix, referred to as the terminal association matrix, denoted as $\mathbf{T}_{\mathtt{R}}$, which indicates the association of elements $r \in \mathtt{R}$ to a specific terminal element $t \in \mathtt{T}$.
  The $\mathbf{T}_{\mathtt{R}}$ matrix has dimension $|\mathtt{T}| \times |\mathtt{R}|$ and can be represented as follows:

$$\mathbf{T}_{\mathtt{R}} = \begin{array}{c} \\ t_1 \\ t_2 \\ \vdots \\ t_k \\ \vdots \\ t_{|\mathtt{T}|} \end{array} \begin{array}{c} \begin{array}{cccccccc} r_1 & r_2 & \cdots & r_m & \cdots & r_{|\mathtt{R}|-1} & r_{|\mathtt{R}|} \end{array} \\ \begin{pmatrix} 0 & 1 & \cdots & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 1 \end{pmatrix} \end{array} \tag{18}$$

  where 0 and 1 represent, respectively, whether an element $r \in \mathtt{R}$ is assigned to a terminal element $t \in \mathtt{T}$ or not.
- Estimated paths matrix:
  The solution of the TPI problem, the set $\widehat{\mathcal{P}}$, can also be conceptualised as a binary matrix that represents the hypothetical paths estimated to be the real paths. This path matrix, denoted as $\widehat{\mathbf{P}}$, has dimension $1 \times |\mathcal{H}^{\mathcal{I}'}|$.

$$\widehat{\mathbf{P}} = \begin{array}{c} \begin{array}{ccccc} h_1 & \cdots & h_k & \cdots & h_{|\mathcal{H}^{\mathcal{I}'}|} \end{array} \\ \begin{pmatrix} 1 & \cdots & 1 & \cdots & 0 \end{pmatrix} \end{array} \tag{19}$$

  where 0 and 1 represent whether the path $h_k \in \mathcal{H}^{\mathcal{I}'}$ is an estimated optimal path or not.

5.5.2 | *ILP Optimisation Problem*

The ILP formulation aims to identify the subset of hypothetical paths that best approximates the real network topology. The optimisation problem is presented in Equations (20a–20f).

$$\max_{\widehat{\mathbf{P}},\mathbf{T}_{\mathtt{R}}} \; \omega \cdot \sum_{k}^{|\mathcal{H}^{\mathcal{I}'}|} \widehat{\mathbf{P}}_{(k)} - \sum_{m}^{|\mathtt{T}|} \sum_{n}^{|\mathtt{R}|} \mathbf{T}_{\mathtt{R}\,(m,\,n)} \tag{20a}$$

$$\text{s.t.} \quad \sum_{k}^{|\mathtt{R}|} \mathbf{H}_{\mathtt{R}\,(m,\,k)} \cdot \widehat{\mathbf{P}}_{(m)} \cdot \mathbf{H}_{\mathtt{T}\,(m,\,n)} \leq \left(\left(\mathbf{T}_{\mathtt{R}} \times \mathbf{H}_{\mathtt{R}}^{\mathsf{T}}\right) \cdot \mathbf{H}_{\mathtt{T}}^{\mathsf{T}}\right)_{(n,\,m)} \tag{20b}$$

$$\forall m \in \left\{1,\, ...,\, |\mathcal{H}^{\mathcal{I}'}|\right\},\ \forall n \in \{1,\, ...,\, |\mathtt{T}|\},$$

$$\left(\widehat{\mathbf{P}} \times \mathbf{H}_{\mathtt{C}}\right)_{(k)} \leq 1 \qquad \forall k \in \{1,\, ...,\, |\mathtt{C}|\}, \tag{20c}$$

$$\sum_{k}^{|\mathtt{T}|} \mathbf{T}_{\mathtt{R}\,(k,\,m)} \leq 1 \qquad \forall m \in \{1,\, ...,\, |\mathtt{R}|\}, \tag{20d}$$

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\widehat{\mathbf{P}}_{(k)} \in \{0,1\} \qquad \forall k \in \left\{1, \ldots, |\mathcal{H}^{\mathcal{I}'}|\right\}, \tag{20e}$$

$$\mathbf{T}_{\mathtt{R}(m,n)} \in \{0,1\} \forall m \in \{1, \ldots, |\mathtt{T}|\}, \qquad \forall n \in \{1, \ldots, |\mathtt{R}|\}. \tag{20f}$$

The decision variables are as follows:

- $\widehat{\mathbf{P}}_{(k)} \in \{0,1\}$, indicating whether the hypothetical path $k \in \left\{1, \ldots, |\mathcal{H}^{\mathcal{I}'}|\right\}$ is selected as part of the estimated set $\widehat{\mathcal{P}}$.
- $\mathbf{T}_{\mathtt{R}(m,n)} \in \{0,1\}$, representing the assignment of an element $n \in \{1, \ldots, |\mathtt{R}|\}$ to a terminal element $m \in \{1, \ldots, |\mathtt{T}|\}$.

The components of the ILP formulation are interpreted as follows:

- Objective function (Equation 20a). The objective function is designed to achieve two competing goals. First, it aims to maximise the number of customers correctly connected to their terminal with $\sum_{k}^{\mathcal{H}^{\mathcal{I}'}} \widehat{\mathbf{P}}_{(k)}$. Second, it aims to penalise overly complex solutions that assign network elements to feeders without necessity (minimise the number of elements with value 1 in the matrix $\mathbf{T}_{\mathtt{R}}$). The weighting factor ω controls the trade-off between maximising customer connectivity and minimising redundant assignments.
- Path validity constraint (Equation 20b). This constraint verifies the validity of the paths by enforcing consistency among the elements of each selected path. It ensures that, for every estimated path $h \in \widehat{\mathcal{P}}$, all intermediate elements $e \in h$ are assigned to the same terminal element $t \in \mathtt{T}$. In simple terms, it ensures that all intermediate elements belonging to a path are consistently assigned to the same feeder (terminal element).
- Unique customer-path constraint (Equation 20c). Each customer $c \in \mathtt{C}$ can be associated with at most one selected path in $\widehat{\mathbf{P}}$. This prevents a customer from being linked to multiple estimated paths.
- Unique element-terminal constraint (Equation 20d). This guarantees that each element $r \in \mathtt{R}$ is associated with at most one specific terminal element $t \in \mathtt{T}$. Therefore, this constraint is designed to prevent the same elements from being associated with multiple terminal elements.
- Binary constraints (Equations 20e and 20f). These ensure that the decision variables $\widehat{\mathbf{P}}$ and $\mathbf{T}_{\mathtt{R}}$ take only binary values.

In summary, the ILP formulation identifies the optimal set of estimated paths, $\widehat{\mathcal{P}}$. It achieves this by ensuring that customers are connected to their transformers, whilst preventing different feeders from overlapping or sharing lines.

5.6 | Diagnostic Function

The set of estimated paths, $\widehat{\mathcal{P}}$, is validated with a *Diagnostic*() function. This abstract function can be customised to implement various validation criteria and methodologies, serving as a critical tool for assessing the quality of the obtained solution. The *Diagnostic*() function evaluates the found paths by checking any possible discrepancies from reality.

If inconsistencies are detected, the corresponding transformation functions can be adjusted to correct the identified issues. The methodology is then re-executed with the updated transformations until satisfactory results are achieved.

5.7 | Solution of the TPI Problem

If no discrepancies are identified or the DSO considers the solution acceptable, the set $\widehat{\mathcal{P}}$ is regarded as the final solution to the TPI problem.

6 | Academic Example

To illustrate the methodology used to solve the TPI problem, an academic example of a power distribution network with a limited number of elements is considered.

6.1 | Raw Information

We assume the set of raw information available to the DSO, denoted as $\mathcal{I}$, is provided by some pieces of information:

i_1: Documents containing the DSO's list of elements, their coordinates and their types. However, some elements may be missing, and some coordinates may be inaccurate.

i_2: MV/LV transformer cabin manuals provide instructions on cabin setup. This includes details about the organisation of connection boards and how feeder lines connect to specific terminal elements or feeder terminal junctions on these boards.

i_3: Customers' information about which feeder terminal junction in the MV/LV transformer they are connected to.

i_4: DSO's general practice is to reduce energy losses by minimising the total path length of each customer.

i_5: Information that LV networks are operated radially.

6.2 | Transformation Functions

The set $\mathcal{F}$ enumerates the transformation functions. Given that there are five pieces of raw information, the number of transformation functions is also five ($|\mathcal{F}| = |\mathcal{I}| = 5$).

6.3 | Well-Defined Information

The set of well-defined information is given by:

$i'_1 = f_1(i_1)$: Set of elements, $\mathcal{E}$ and their attributes, $\mathcal{A}$ are known.

$i'_2 = f_2(i_2)$: The elements of two types *junction* and *transformer* are directly connected with no intermediate element. Since each feeder terminal junction has a known associated transformer, we can simplify the path identification problem by considering the path from the customer to the feeder junction terminal.

$i'_3 = f_3(i_3)$: For each element of type *customer*, $e \in \mathcal{E}$, its attribute *junction* and *c.junction*, where it is connected to is known.

$i'_4 = f_4(i_4)$: An element can be connected to another only if its distance is less than a given distance D. Moreover, the total length of a path must not exceed a maximum length L.

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$i'_5 = f_5(i_5)$: Since the network has a radial structure and does not have loops, each element can only appear once in each path.

The set of elements known to the DSO is given by:

$$\mathcal{E} = \{e_1, \ldots, e_{18}\}. \tag{21}$$

The set of attributes, $\mathcal{A}$, is given by:

$$\mathcal{A} = \{type,\ coordinate,\ junction\}. \tag{22}$$

The set of types, $\mathcal{T}$, is given by: *customer* (elements e_1–e_6), *line* (e_7–e_{12}), *junction* (e_{13}–e_{16}) and *transformer* (e_{17} and e_{18}) as shown in Figure 4a.

6.4 | Hypothetical Paths Compatible With the Well-Defined Information

The first two pieces of information are used to generate an initial set of hypothetical paths, $\mathcal{H}^{i'_1,i'_2}$, containing all the paths from a customer to a terminal junction. Figure 4b illustrates the customer-to-feeder terminal junction colours, where matching colours indicate elements that are supposed to be connected to the same terminal junction. Therefore, the size of the set of hypothetical paths, compatible with the first two pieces of the well-defined information, i'_1 and i'_2, is calculated using Equation (10):

$$|\mathcal{H}^{i'_1,i'_2}| = 61607. \tag{23}$$

The calculation is performed with the set of elements considered $\mathcal{E}$, the customer set $\mathtt{C} = Subset(\mathcal{E}, customer)$, the terminal points set $\mathtt{T} = Subset(\mathcal{E}, junction)$, the transformer set $\mathtt{F} = Subset(\mathcal{E}, transformer)$ and the remaining elements set $\mathtt{R} = \mathcal{E} - \mathtt{C} - \mathtt{T} - \mathtt{F}$.

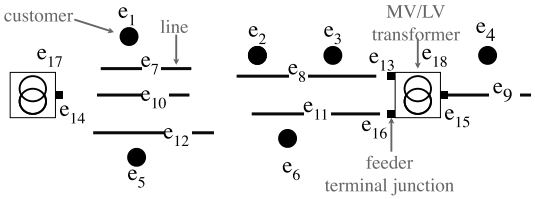
Each remaining piece of well-defined information is used to exclude the paths in $\mathcal{H}^{i'_1,i'_2}$ that are not compatible with the well-defined information. Figure 4c highlights some of the hypothetical paths for customer e_3 that satisfy the distance and connectivity constraints defined by i'_1 and i'_2. Therefore, considering all the remaining information, i'_3 to i'_5, the total number of hypothetical paths compatible with the well-defined information is $|\mathcal{H}^{\mathcal{I}'}|$. For example, Figure 4d shows some of the hypothetical paths for customer e_3 that satisfy the distance and connectivity constraints defined by i'_4 and i'_5.

Detailed information about the elements in each hypothetical path can be found in the online repository (see Section 1.2).

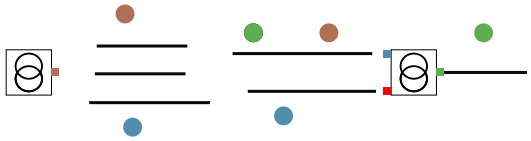
6.5 | Results Paths Estimation

The matrices $\mathbf{H}_C$, $\mathbf{H}_R$, $\mathbf{H}_T$, $\widehat{\mathbf{P}}$ and $\mathbf{T}_R$ are used in the optimisation problem as shown in Equations (20a–20f).

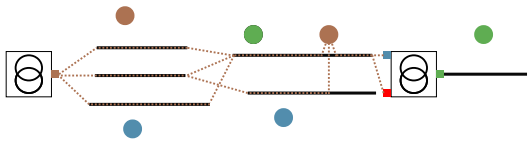
In particular, the matrices $\mathbf{H}_C$, $\mathbf{H}_R$ and $\mathbf{H}_T$ are used as the parameters of the ILP problem, whereas $\widehat{\mathbf{P}}$ and $\mathbf{T}_R$ are the decision variables and represent a proposed solution to the optimisation problem.



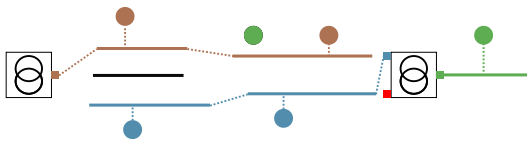
(a) Representation of the academic test network.



(b) Visualization of customer-to-feeder terminal junction assignments, where matching colors indicate elements supposed to be connected together.



(c) Image showing some hypothetical paths for the customer e_3.



(d) Final network configuration after customer path identification.

FIGURE 4 | Illustration of the process of identifying customer paths in the academic network considered.

The matrix $\widehat{\mathbf{P}}$ contains the estimated paths, which are considered as approximations of the real paths; whereas the matrix $\mathbf{T}_R$ represents which elements belong to each feeder terminal junction.

The estimated optimal paths for this academic network are as follows:

$$\widehat{\mathbf{P}} = \begin{array}{c} \\ \end{array} \begin{array}{c} \begin{matrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 & h_{10} \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array} \tag{24}$$

The matrix $\mathbf{T}_R$ is given by:

$$\mathbf{T}_R = \begin{array}{c} \\ e_{13} \\ e_{14} \\ e_{15} \\ e_{16} \end{array} \begin{array}{c} \begin{matrix} e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{25}$$

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6.6 | Results Explanation

We will now examine the results obtained from applying the optimisation method to the academic example. This analysis will clarify the meaning and role of each component in Equations (20a–20f).

6.6.1 | *Maximisation Term*

The maximisation term, Equation (20a), contains two operations:

- Maximising estimated paths: the first term, $\sum_{k}^{|\mathcal{H}^{\mathcal{I}'}|}\widehat{\mathbf{P}}_{(k)}$, focuses on maximising the number of estimated optimal paths, therefore paths whose have a value of 1.
- Penalty for unnecessary assignments: the second term, $\sum_{m}^{|\mathrm{T}|}\sum_{n}^{|\mathbb{R}|}\mathbf{T}_{\mathrm{R}\,(m,\,n)}$, introduces a penalty term related to the terminal association matrix. This term discourages assigning elements to a feeder terminal junction when unnecessary.

The value of ω is important to determine the relative importance of finding more valid paths over minimising the number of elements assigned to the path. A larger value of ω encourages the optimisation to find more valid paths, whereas a smaller value puts more emphasis on minimising the elements used. In this case, ω is set to 10.

6.6.2 | *Validity Path Constraint*

The left-hand side of the constraint term in Equation (20b) is calculated in three operations:

i. $\sum_{k}^{|\mathbb{R}|}\mathbf{H}_{\mathrm{R}\,(\bullet,\,k)}$. This operation counts the number of elements present in each hypothetical path, $h \in \mathcal{H}^{\mathcal{I}'}$. The resulting matrix, of dimensions $|\mathcal{H}^{\mathcal{I}'}| \times 1$, is presented in transposed form for space efficiency as follows:

$$\begin{array}{c} \begin{array}{cccccccccc} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 & h_{10} \end{array} \\ \begin{pmatrix} 1 & 2 & 2 & 2 & 3 & 3 & 2 & 1 & 2 & 1 \end{pmatrix} \end{array}. \tag{26}$$

ii. The result, Equation (26), is multiplied element-wise by $\widehat{\mathbf{P}}$. This operation assures the verification of the constraint only for the estimated optimal paths. The resulting matrix, still of dimensions $|\mathcal{H}^{\mathcal{I}'}| \times 1$, is transposed and presented as follows:

$$\begin{array}{c} \begin{array}{cccccccccc} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 & h_{10} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 2 & 1 \end{pmatrix} \end{array}. \tag{27}$$

iii. Finally, the result, Equation (27) is multiplied by the matrix $\mathbf{H}^{\mathrm{T}}$. The result of this operation still has dimensions $|\mathcal{H}^{\mathcal{I}'}| \times |\mathrm{T}|$ and the transposed form is shown as follows:

$$\begin{array}{c} \begin{array}{cccccccccc} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 & h_{10} \end{array} \\ \begin{array}{c} e_{13} \\ e_{14} \\ e_{15} \\ e_{16} \end{array} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}. \tag{28}$$

The element-wise multiplication of matrices in Equation (27) and $\mathbf{H}^{\mathrm{T}}$ can be formally represented using diagonal matrix multiplication, ensuring dimensional consistency between the operands. Let $\mathbf{A} \subseteq \mathbb{R}^{M\times 1}$ denote the column vector obtained from Equation (27), and let $\mathbf{B} = \mathbf{H}^{\mathrm{T}} \subseteq \mathbb{R}^{M\times N}$, with $M, N \in \mathbb{N}^{+}$. The resulting matrix $\mathbf{C} \subseteq \mathbb{R}^{M\times N}$ is computed as follows:

$$\mathbf{C} = \mathrm{diag}(\mathbf{A}) \cdot \mathbf{B}, \tag{29}$$

where $\mathrm{diag}(\mathbf{A})$ is a diagonal matrix whose diagonal entries correspond to the elements of $\mathbf{A}$.

Therefore, the resulting matrix in Equation (28) has dimensions $|\mathcal{H}^{\mathcal{I}'}| \times |\mathrm{T}|$.

Similarly, the right-hand side of the constraint term in Equation (20b) is calculated in two operations:

i. The matrix multiplication of $\mathbf{T}_{\mathrm{R}}$ and $\mathbf{H}_{\mathrm{R}}^{\mathrm{T}}$. This operation calculates the number of elements that are present at the same time in a feeder terminal junction $t \in \mathrm{T}$ and in a path $h \in \mathcal{H}^{\mathcal{I}'}$. The resulting matrix has dimensions $|\mathrm{T}| \times |\mathcal{H}^{\mathcal{I}'}|$ and is given as follows:

$$\begin{array}{c} \begin{array}{cccccccccc} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 & h_{10} \end{array} \\ \begin{array}{c} e_{13} \\ e_{14} \\ e_{15} \\ e_{16} \end{array} \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 2 & 1 \\ 1 & 1 & 1 & 1 & 2 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}. \tag{30}$$

For example, considering the value at position (e_{14}, h_7) of the matrix in Equation (30). The value is 2 since the same elements e_7 and e_8 are at the same time assigned to the feeder terminal junction e_{14} (second row of the junction association matrix in Equation 25) and are present in the path h_7 (hypothetical path h_7 composed by the elements: $h_7 = (e_3, e_7, e_8 \text{ and } e_{14})$).

ii. Similarly, for the left-hand side, the result of Equation (30) is multiplied by the matrix $\mathbf{H}_{\mathrm{T}}$. This assures that the comparison on both sides of the inequality is performed only on the elements that are assigned to the same feeder terminal junction. The result of this operation has dimensions $|\mathrm{T}| \times |\mathcal{H}^{\mathcal{I}'}|$ and is expressed as follows:

$$\begin{array}{c} \begin{array}{cccccccccc} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 & h_{10} \end{array} \\ \begin{array}{c} e_{13} \\ e_{14} \\ e_{15} \\ e_{16} \end{array} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}. \tag{31}$$

Finally, by comparing Equations (28) and (31), the condition Equation (28) $\leq$ Equation (31) is verified for each corresponding value of the matrices.

In general, the constraint Equation (20b) assures that all the elements of an estimated path belong to the same feeder terminal junction. This is guaranteed when the number of elements in each estimated path is less than or equal to the total number of elements in the corresponding feeder terminal junction.

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6.6.3 | *Unique Estimated Path for Customer Constraint*

The constraint in Equation (20c) is given as follows:

$$\left(\hat{\mathbf{P}} \times \mathbf{H}_{\mathbb{C}}\right) = \begin{matrix} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{matrix} \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}, \tag{32}$$

$$\left(\hat{\mathbf{P}} \times \mathbf{H}_{\mathbb{C}}\right)_{(k)} \leq 1, \qquad \forall k \in \{1, ..., |\mathbb{C}|\}. \tag{33}$$

Equation (32) is a matrix of dimensions $1 \times |\mathbb{C}|$ and in this academic example, all the values are 1 except for e_2 whose value is 0 since no path was identified for that customer.

6.6.4 | *Elements Assigned to a Unique Feeder Terminal Junction Constraint*

The constraint in Equation (20d) is given as follows:

$$\sum_{k}^{|\mathbb{T}|} \mathbf{T}_{\mathbb{R}(k,\bullet)} = \begin{matrix} e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \end{matrix} \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}, \tag{34}$$

$$\sum_{k}^{|\mathbb{T}|} \mathbf{T}_{\mathbb{R}(k,m)} \leq 1, \qquad \forall m \in \{1, ..., |\mathbb{R}|\}. \tag{35}$$

Equation (34) is a matrix of dimensions $1 \times |\mathbb{R}|$ and in this academic example, all the values are 1 except for e_{10} whose value is 0 since the element e_{10} is not used by any path. Therefore, it is not assigned to any feeder terminal junction.

A possible explanation for why e_{10} is not used is that it is a redundant element in the dataset or that the optimisation algorithm may have chosen that alternative paths provide a better solution.

6.7 | Example Evaluation

Because the academic network is a controlled example, the real paths P are known, allowing a direct quantitative assessment of the estimated paths $\widehat{P}$. From Equation (32), $\widehat{P} \times H_C = (1\ 0\ 1\ 1\ 1\ 1)$, so a valid path is recovered for five of the six customers: e_1, e_3, e_4, e_5 and e_6, giving a per-customer connection accuracy of $5/6 = 83.3\%$ and, using the Jaccard index introduced in Section 4, to:

$$J(\widehat{P}, P) = \frac{|P \cap \widehat{P}|}{|P \cup \widehat{P}|} = \frac{5}{6} \approx 0.83. \tag{36}$$

The single discrepancy, customer e_2, is analysed in the diagnostic evaluation below.

6.8 | Diagnostic Function

The proposed solution, the matrices $\widehat{\mathbf{P}}$ and $\mathbf{T}_{\mathbb{R}}$, are then evaluated using the *Diagnostic*() function to detect any possible issues.

For example, the solution proposed does not find the paths for each customer, since it is not possible to find a path for customer e_2 given the data available to the DSO. This issue is reported to the DSO, who can perform some further analysis on the case and understand why no path was found.

For this controlled example the cause can be identified precisely. The A^* search generates candidate paths for e_2, but none can be selected: choosing any of them would require assigning a line element to two different feeder terminal junctions, violating the path-validity constraint (Equation 20b). This behaviour is consistent with an erroneous recorded *c.junction* for e_2. Therefore, the *Diagnostic*() function flags e_2 for verification rather than forcing an inconsistent connection.

7 | Belgian Network

We apply our methodology to a real Belgian LV network characterised by incomplete GIS data and missing customer connections to the network. For this use case, we assume that the set of raw information, $\mathcal{I}$, the transformation functions, $\mathcal{F}$ and the well-defined information, $\mathcal{I}'$, are the same as those described in the academic example in Section 6.

For this real case, the set of elements known to the Belgian DSO is given by:

$$\mathcal{E} = \{e_1, ..., e_{1089}\}. \tag{37}$$

The set of attributes, $\mathcal{A}$, is given by:

$$\mathcal{A} = \{type,\ coordinate,\ junction\}. \tag{38}$$

The number of elements for the different types is given in Table 1.

7.1 | Hypothetical Paths Compatible With the Well-Defined Information

The total number of hypothetical paths in the set $\mathcal{H}$ is generally determined using Equation 10. Following the methodology $\mathfrak{M}()$, the hypothetical paths not compatible with the well-defined information are excluded, keeping only those compatible with each piece of information, obtaining the set $\mathcal{H}^{I'}$. For the real Belgian network considered, given the large number of elements in the network, it is impractical to explicitly represent the set $\mathcal{H}$ as a set of all the possible hypothetical paths.

For this reason, to efficiently identify the hypothetical paths that are compatible with the well-defined information, the A^* algorithm mentioned in Section 5.4.1 is used. The implementation of the $A^*()$ function is available on the online repository (see Section 1.2). Below, we present the general concept.

After the execution of the $A^*()$ algorithm, considering a maximum number of paths for each customer of $N = 15$, the total number of hypothetical paths is $|\mathcal{H}^{\mathcal{I}'}| = 17335$. The value of N is selected as a balance between computational efficiency and the likelihood of identifying an optimal solution.

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Algorithm performance, hyperparameters and execution statistics for the Belgian network.

Category	Metric	Value
Network profile	Total elements ($\mathcal{E}$)	1089
	Customers (C)	526
	Lines (R)	441
	Feeder junctions (T)	96
	MV/LV transformers (F)	26
Hyperparameters	Max candidate paths per customer (N)	15
	Max connection distance (D)	50 m
	Max path length (L)	500 m
	Objective weighting factor (ω)	10
Statistics	Candidate paths generated $\left(\lvert\mathcal{H}^{\mathcal{I}'}\rvert\right)$	17,335
	A^* generation time	264 s
	ILP optimisation time	5.8 s
Results	Customers connected	88%

7.2 | Matrix Generation

The set $\mathcal{H}^{\mathcal{I}'}$ is decomposed into the three sub-matrices, $\mathbf{H}_{\mathrm{C}}$, $\mathbf{H}_{\mathrm{R}}$ and $\mathbf{H}_{\mathrm{T}}$ as explained in Section 5.5.1.

These matrices are used as input parameters of the optimisation problem.

The matrices $\widehat{\mathbf{P}}$ and $\mathbf{T}_{\mathrm{R}}$ are used as output decision variables of the optimisation algorithm.

7.3 | Optimisation Problem Results

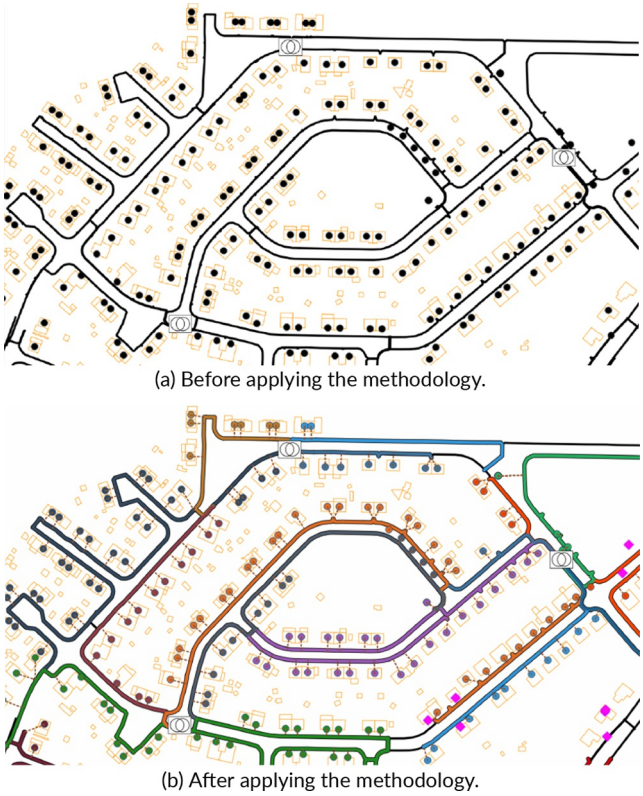
After executing the optimisation algorithm, a solution is obtained in the form of matrices $\widehat{\mathbf{P}}$ and $\mathbf{T}_{\mathrm{R}}$.

Displaying the matrices directly might be challenging for visualisation and comprehension. Therefore, we summarise and present the results in a more comprehensible way.

Therefore, given the matrix, $\widehat{\mathbf{P}}$, it is possible to represent the connections among the elements of the estimated paths graphically. In particular, each element of an estimated optimal path is assigned a specific colour, with black reserved for unassigned elements.

The colour that is assigned to each element depends on the matrix $\mathbf{T}_{\mathrm{R}}$. In particular, the same colour is assigned to the elements that belong to the same feeder terminal junction.

Figure 5a shows a part of the Belgian network considered with three MV/LV transformers, some lines and customers. In Figure 5b, it is possible to see the network after the identification of the paths: colours have been assigned to customers and lines belonging to the same feeder terminal junction. Dashed lines represent the connections between the elements, for example, customer-line connections.



(a) Before applying the methodology.

(b) After applying the methodology.

FIGURE 5 | Part of the real network considered.

7.4 | Diagnostic Function

Moreover, a few purple squares are present. These are customers for whom no path was identified, and these are discovered with the *Diagnostic*() function.

Possible reasons for failing to identify a path include incorrect or missing data for the feeder terminal junction associated with a customer or the inability to assign a line to its corresponding feeder terminal junction.

These issues are reported to the DSO for further investigation.

One of the main challenges in this TPI problem is the absence of a well-defined ground truth as no explicit reference data is available. Consequently, quantitative criteria cannot be established for a rigorous evaluation of algorithmic performance. Instead, assessment must rely on qualitative evaluation through human interpretation. For example, in this case, the DSO could access only the schematic images of customer paths, and a human-in-the-loop process was employed to compare the identified paths against the schematics to assess their accuracy.

7.5 | Parameterisation and Computational Scalability

The performance and scalability of the proposed TPI framework depend on a defined set of hyperparameters, summarised alongside execution metrics in Table 1. The spatial threshold $D = 50$ m and maximum path length $L = 500$ m were selected based on the DSO's network characteristics and operational experience, effectively restricting the search space to electrically

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

plausible routes. To control combinatorial growth, candidate paths per customer were capped at $N = 15$, ensuring sufficient routing diversity whilst maintaining tractability. Within the ILP formulation (Equation 20a), the weighting factor $\omega = 10$ balances the maximisation of identified customer paths against the penalisation of too long paths.

The methodology was implemented in Python using the Gurobi optimiser and evaluated on a standard machine (Intel i7, 8GB RAM). Applied to the 1089-element Belgian network, the A^* algorithm evaluated the spatial constraints and generated 17,335 candidate paths in 264 s. Thanks to the highly pruned search space, the resulting matrix sparsity enabled the ILP solver to subsequently reach global optimality in just 5.8 s. As detailed in Table 1, this runtime disparity highlights that the candidate-path generation phase is the dominant bottleneck, making it the primary target for future scalability improvements.

7.6 | Robustness Analysis

The proposed framework assumes that customer-to-feeder terminal junction assignments (*c.junction*) provided in DSO records are correct. This separates it from AMI-based identification methods, which can infer or verify such assignments directly from smart metre measurements and are therefore less sensitive to errors in static records. However, AMI-based approaches require sufficient sensor coverage—a condition that many European DSOs cannot yet meet [2].

Therefore, in data-scarce settings, the accuracy of *c.junction* records becomes the critical input quality concern. To evaluate the robustness of the proposed framework against DSO data inaccuracies, a robustness analysis is performed on a part of a sub-network of the real-world case study (the same area depicted in Figure 5). The considered sub-network comprises 211 customers, 3 transformers and 17 feeder terminal junctions.

We changed the input data by randomly corrupting the customer terminal junction (*c.junction*) assignments for a varying percentage of customers (from 10% up to 100%) among one of the terminal junction of the 2 other transformers. To ensure statistical significance, 10 independent iterations are performed for each corruption level, whereas the spatial hyperparameters (D and L) and the objective penalty (ω) are kept constant.

The objective is to measure the algorithm's ability to correctly connect unchanged customers and avoiding connect corrupted customers. The performance is evaluated across three metrics illustrated in Figure 6:

- Fault detection rate: The percentage of customers with corrupted feeder assignments for which the framework failed to identify a valid path, correctly avoiding a false connection (true positive rate).
- False connection rate: The percentage of customers with corrupted feeder assignments for which the framework identified a valid path to the wrong feeder (false positive rate).
- Network connectivity rate: The percentage of customers that are assigned a path to their correct feeder terminal junction, where ground truth is defined by the original uncorrupted assignment (accuracy).

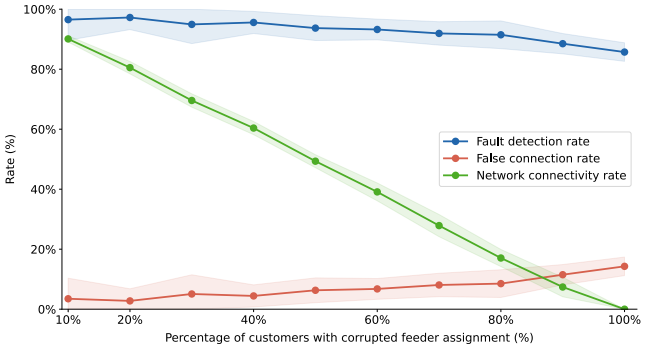


FIGURE 6 | Sensitivity analysis of the TPI framework under varying levels of corruption. Shaded regions represent the ±2 standard deviation interval computed over 10 independent runs per corruption level.

The results show that under realistic error levels (between 10% and 30% corrupted customer-feeder assignments), the fault detection rate exceeds 95%, meaning that the framework correctly refuses to assign the vast majority of corrupted customers to an inconsistent feeder, leaving them unconnected for further DSO investigation. At the same time, the false connection rate remains below 5%, showing that wrong connections are rarely accepted. The network connectivity rate decreases approximately linearly with the corruption level, reflecting the expected reduction in the number of customers for which a feasible path can be identified as database inconsistencies increase.

It should be noted that corrupted assignments are drawn from feeder terminal junctions associated with other transformers. As these transformers typically serve different geographic areas, many corrupted assignments are geographically distant, causing violations of the distance (D) or path-length (L) constraints and rejection before the ILP stage. Although this is an intended feature of the framework, the reported fault detection rate partly reflects the spatial implausibility of the injected errors. A more challenging scenario is assigning a customer to a nearby but incorrect junction, particularly within the same MV/LV substation. In such cases, feasible short paths exist, spatial constraints are less effective, and detection becomes harder. Therefore, the reported fault detection rate should be regarded as an upper bound for the studied corruption model; evaluation under spatially adjacent, same-substation corruption is left for future work.

This confirms that the proposed TPI framework prioritises physical grid realities over data errors, with the potential to act as a reliable anomaly detection filter for DSOs.

8 | Conclusion

This paper presented an optimisation algorithm to address the TPI problem in LV power distribution networks. Using only static GIS data and customer-to-transformer associations provided by DSOs, the methodology first constructs a set of hypothetical paths via an A^* search strategy, then selects the optimal

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

subset through an ILP formulation that maximises the number of customers assigned to physically feasible paths. By operating without AMI data, such as smart metre recordings, the approach is broadly applicable across distribution networks, including those in regions with limited metring infrastructure, and avoids the cost and complexity of integrating dynamic measurement sources.

The effectiveness of the methodology was validated on both an academic test network and a real Belgian low-voltage network comprising 1089 elements, achieving an 88% customer connection rate. A robustness analysis further demonstrated that, under realistic levels of corrupted customer-feeder assignments (10%–30%), the framework achieved a fault detection rate above 95% whilst maintaining a false connection rate below 5%. This methodology is fundamental for constructing an accurate digital reconstruction of power distribution networks, ultimately aiding DSOs in network management and optimisation.

Future work will focus on extending the optimisation to backup paths, integrating AMI data to further improve accuracy, and improving computational scalability for larger networks as proposed in ref. [29].

Author Contributions

Maurizio Vassallo: conceptualisation, formal analysis, methodology, software, validation, writing – original draft. **Adrien Leerschool:** conceptualization, formal analysis, methodology, software, validation, writing – review and editing. **Alireza Bahmanyar:** conceptualisation, validation, writing – review and editing. **Laurine Duchesne:** conceptualisation, validation, writing – review and editing. **Simon Gerard:** resources, validation. **Thomas Wehenkel:** resources, validation. **Damien Ernst:** conceptualisation, resources, supervision, validation.

Funding

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are not publicly available due to proprietary restrictions. A minimal working example demonstrating the methodology is available on GitHub at https://github.com/TPIproblem/OptimalTPI.

References

1. F. Geth, M. Vanin, and D. V. Hertem, "Data Quality Challenges in Existing Distribution Network Datasets," in *IET Conference Proceedings CP823* Vol, 2023, no. 6 (2023): 2421–2425, https://doi.org/10.1049/icp.2023.1175.

2. M. Kochański, K. Korczak, and T. Skoczkowski, "Technology Innovation System Analysis of Electricity Smart Metering in the European Union," *Energies* 13, no. 4 (2020): 916, https://doi.org/10.3390/en13040916.

3. C. Mateo Domingo, T. Gomez San Roman, A. Sanchez-Miralles, J. P. Peco Gonzalez, and A. Candela Martinez, "A Reference Network Model for Large-Scale Distribution Planning With Automatic Street Map Generation," *IEEE Transactions on Power Systems* 26, no. 1 (2011): 190–197, https://doi.org/10.1109/TPWRS.2010.2052077.

4. A. Seack, J. Kays, and C. Rehtanz, "Generating Low Voltage Grids on the Basis of Public Available Map Data," CIRED Workshop (2014), https://api.semanticscholar.org/CorpusID:38279545.

5. A. Navarro-Espinosa and L. F. Ochoa, "Reconstruction of Low Voltage Distribution Networks: From GIS Data to Power Flow Models," CIRED (2015), https://api.semanticscholar.org/CorpusID:55344363.

6. A. Guzmán, J. Quirós-Tortós, and G. Valverde, "Efficient Connectivity Identification of Large-Scale Distribution Network Elements in GIS," *IEEE Manchester PowerTech* (2017): 1–6, https://doi.org/10.1109/PTC.2017.7981064.

7. A. Guzmán, A. Argüello, J. Quirós-Tortós, and G. Valverde, "Processing and Correction of Secondary System Models in Geographic Information Systems," *IEEE Transactions on Industrial Informatics* 15, no. 6 (2019): 3482–3491, https://doi.org/10.1109/TII.2018.2876356.

8. G. Zepu, L. Yongjian, X. Ziwei, Y. Yilan, and Z. Lianmei, "Knowledge Graph-Based Method for Identifying Topological Structure of Low-Voltage Distribution Network," *Journal of Engineering* 2020, no. 12 (2020): 1177–1184, https://doi.org/10.1049/joe.2019.1319.

9. X. Li, X. Feng, Z. Zeng, X. Xu, and Y. Zhang, "Distribution Feeder One-Line Diagrams Automatic Generation From Geographic Diagrams Based on GIS," in *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies* (2008), 2228–2232, https://doi.org/10.1109/DRPT.2008.4523781.

10. L. Wu, Y. Lin, and W. Pang, "Distribution Network Topology Modelling and Automatic Mapping Based on CIM and GIS," in *IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)* (2018), 668–671, https://doi.org/10.1109/ITOEC.2018.8740683.

11. L. Shang, R. Hu, H. Ci, W. Zhang, and G. Ouyang, "Automatic Generation Algorithm of Distribution Network Topology Map Based on GIS Drawing," *IOP Conference Series: Earth and Environmental Science* 384, no. 1 (2019): 012231, https://doi.org/10.1088/1755-1315/384/1/012231.

12. Z. Yin, J. Luo, L. Wang, and R. Ye, "Research on Network Topology Analysis Method of Distribution Management Based on GIS," in *International Conference on Intelligent Transportation, Big Data and Smart City (ICITBS)* (2021), 145–148, https://doi.org/10.1109/ICITBS53129.2021.00045.

13. M. Farajollahi, A. Shahsavari, and H. Mohsenian-Rad, "Topology Identification in Distribution Systems Using Line Current Sensors: An MILP Approach," *IEEE Transactions on Smart Grid* 11, no. 2 (2020): 1159–1170, https://doi.org/10.1109/TSG.2019.2933006.

14. D. Marulli, S. Mathieu, A. Benzerga, A. Sutera, and D. Ernst, "Reconstruction of Low-Voltage Networks With Limited Observability," in *IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)* (2021), 1–5, https://doi.org/10.1109/ISGTEurope52324.2021.9640163.

15. C. Wang, J. An, and G. Mu, "Power System Network Topology Identification Based on Knowledge Graph and Graph Neural Network," *Frontiers in Energy Research* 8 (2021): 613331, https://doi.org/10.3389/fenrg.2020.613331.

16. Z. Soltani and M. Khorsand, "Real-Time Topology Detection and State Estimation in Distribution Systems Using Micro-PMU and Smart Meter Data," *IEEE Systems Journal* 16, no. 3 (2022): 3554–3565, https://doi.org/10.1109/JSYST.2022.3153449.

17. P. A. Parikh and T. D. Nielsen, "Transforming Traditional Geographic Information System to Support Smart Distribution Systems," in *2009 IEEE/PES Power Systems Conference and Exposition* (2009), 1–4, https://doi.org/10.1109/PSCE.2009.4839979.

18. J. Peppanen, S. Grijalva, M. J. Reno, and R. J. Broderick, "Distribution System Low-Voltage Circuit Topology Estimation Using Smart

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Metering Data," in *IEEE/PES Transmission and Distribution Conference and Exposition* (2016), 1–5, https://doi.org/10.1109/TDC.2016.7519985.

19. T. Morrell, V. Venkataramanan, A. Srivastava, A. Bose, and C. C. Liu, "Modeling of Electric Distribution Feeder Using Smart Meter Data," in *IEEE/PES Transmission and Distribution Conference and Exposition (T&D)* (2018), 1–9, https://doi.org/10.1109/TDC.2018.8440540.

20. S. Cui, P. Zeng, C. Song, Z. Wang, and G. Li, "Low-Voltage Distribution Network Topology Identification Based on Constrained Least Square and Graph Theory," *Soft Computing* 26, no. 17 (2022): 8509–8519, https://doi.org/10.1007/s00500-022-07151-3.

21. H. Li, W. Liang, Y. Liang, Z. Li, and G. Wang, "Topology Identification Method for Residential Areas in Low-Voltage Distribution Networks Based on Unsupervised Learning and Graph Theory," *Electric Power Systems Research* 215 (2023): 108969, https://doi.org/10.1016/j.epsr.2022.108969.

22. X. Yu, J. Zhao, H. Zhang, X. Wang, and X. Bian, "Data-Driven Distributed Grid Topology Identification Using Backtracking Jacobian Matrix Approach," *IEEE Transactions on Industrial Informatics* 20, no. 2 (2024): 1711–1720, https://doi.org/10.1109/TII.2023.3280936.

23. A. B. Pengwah, Y. Z. Gerdroodbari, R. Razzaghi, and L. L. H. Andrew, "Topology Identification of Distribution Networks With Partial Smart Meter Coverage," *IEEE Transactions on Power Delivery* 39, no. 2 (2024): 992–1001, https://doi.org/10.1109/TPWRD.2024.3354292.

24. D. Liu, J. S. Giraldo, P. Palensky, and P. P. Vergara, "Topology Identification and Parameters Estimation of LV Distribution Networks Using Open GIS Data," *International Journal of Electrical Power & Energy Systems* 164 (2025): 110395, https://doi.org/10.1016/j.ijepes.2024.110395.

25. 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), 1–6, https://doi.org/10.1109/PowerTech46648.2021.9495093.

26. A. Hoogsteyn, M. Vanin, A. Koirala, and D. V. Hertem, "Low Voltage Customer Phase Identification Methods Based on Smart Meter Data," *Electric Power Systems Research* 212 (2022): 108524, https://doi.org/10.1016/j.epsr.2022.108524.

27. M. Vassallo, A. Bahmanyar, L. Duchesne, et al., "A Systematic Procedure for Topological Path Identification With Raw Data Transformation in Electrical Distribution Networks," in *Proceedings of International Conference on Energy, Electrical and Power Engineering (CEEPE)* (2024), 707–715, https://doi.org/10.48550/arXiv.2409.09075.

28. P. E. Hart, N. J. Nilsson, and B. Raphael, "A Formal Basis for the Heuristic Determination of Minimum Cost Paths," *IEEE Transactions on Systems Science and Cybernetics* 4, no. 2 (1968): 100–107, https://doi.org/10.1109/TSSC.1968.300136.

29. T. Théate, L. Duchesne, A. Leerschool, et al., "Smart Meters Phase Identification for Topology Verification: Practical Challenges and Insights From a Case Study," *CIRED* 2025, no. 14 (2025): 2260–2264, https://doi.org/10.1049/icp.2025.2414.

25152947, 2026, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/stg2.70106 by Thirion Paul - Dge, Wiley Online Library on [16/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License