Reconstruction of Low-Voltage Networks with Limited Observability

Daniele Marulli, Sébastien Mathieu, Antonio Sutera, Damien Ernst
Department of Electrical Engineering and Computer Science
University of Liège, Belgium

***Abstract*—This work addresses the problem of reconstructing topology and line parameters of three-phase low-voltage networks when no a priori information about them is provided and not all the nodes in the grid are equipped with smart meters. We present a methodology to obtain an estimation of the electrical model of each phase of the network, analyzing voltage and current time-series measurements provided by the available meters. The proposed methodology involves an iterative algorithm developed to tackle the network reconstruction problem when unmetered nodes are located reasonably far from each other. The algorithm has been tested on a 30-node network with different sets of metered nodes, providing relevant solutions in most of the scenarios with more than 80% of metered nodes.**

I. Introduction

With an ever-increasing penetration of distributed energy resources and electric vehicles, the efficient management of distribution networks becomes more complicated and distribution system operators (DSOs) might encounter serious difficulties in guaranteeing the safety of their network at Low-Voltage (LV) levels. In order to implement effective preventive or corrective measures, DSOs need to be able to assess the system's response to different realistic scenarios. This analysis is usually performed through power flow studies [1], but reliable solutions require accurate information about the topology of the network and the physical characteristics of the lines. At these voltage levels, networks are mostly operated radially and most loads are connected to the feeder through a single phase and the neutral wire. DSOs do not always know how households, feeders and other appliances are interconnected. LV networks topology can change over time because of faults, maintenance and reconfiguration and existing database can contain outdated or inaccurate information. In other words, DSOs usually lack of a reliable electrical model of their LV network. This lack of knowledge can hinder therefore an efficient management and development of the system. Network reconstruction is the mathematical process that allows to deduce this information. Effective network reconstruction methods for LV networks are thus essential for the development of *smarter* grids [2]. The goal of this work is to present a methodology that allows to infer the connections between nodes and to assess line parameters of a LV network from time-series measurements provided by a limited amount of meters in the grid. An overview of existing methods is now presented, followed by a clear description of the network reconstruction problem.

II. Literature Review

In transmission systems, topology and line parameters are usually known [2] and, thanks to a large number of measurement devices, changes can be detected through well-established state estimation methods [3]. Unfortunately, the same does not apply to LV distribution networks. With the increasing importance and complexity of distribution systems, the reconstruction problem of LV networks has gained more attention and it has been recently tackled adopting different assumptions and approaches. Recent works in literature, such as [2], [4], [5], focus on identifying the network topology when limited information is available. Algorithms that aim, as this work does, to identify both network topology and lines parameters at the same time, are presented in [6]–[9]. Authors of [6] use the evaluation of voltage sensitivities with respect to active and reactive power injections and Prũfer sequences to reconstruct the topology of small networks, assuming that only specific cables types and lengths are used for the lines. The identification problem in [7] takes the name of inverse power flow problem, where the system admittance matrix is found by solving an unconstrained least square problem. The case with non-measured nodes in the grid, also referred as *hidden nodes*, is also tackled, both for meshed and radial topologies, with the assumption that these hidden nodes have zero net current injections. The inverse power flow problem is extended to polyphase systems in [8], with the full-observability assumption. Finally, an algorithm to jointly estimate both admittances and topology, assuming that the measurements for all the non-zero power injecting nodes are available, is presented in [9]. A summary comparison table can be found in Table I. All the methods that have been examined share the assumption that every node in the network, or at least the power-injecting ones, has a meter attached to it. This paper presents a methodology to tackle the network reconstruction problem even when some power-injecting nodes in the grid are not metered.

III. Problem statement

Let us consider a three-phase four-wire LV radial distribution network. The three phases are denoted by a, b and c, the neutral wire by d. Let $\mathcal{P} = \{a, b, c\}$, $\mathcal{N}$ and $\mathcal{B}$ be the set of phase indexes, the set of nodes and the set of branches of the network, respectively. The system is observed over a finite time period, which is discretised in T intervals with time step Δt. Accessing the value of a variable at certain time $t \in \{1, \dots, T\}$ is done using an additional subscript, e.g.:

Table I
COMPARISON OF THE MAIN METHODS TO RECONSTRUCT LV NETWORKS.

	[2]	[4]	[5]	[6]	[7]	[8]	[9]	[*]
Topology identification	●	●	●	●	●	●	●	●
Line parameters estimation	○	○	○	◐	●	●	●	●
Hidden non-injecting nodes	○	○	○	●	●	○	●	●
Hidden injecting nodes	○	○	○	○	○	○	○	◐
Required measurements	V	V	E	V, S	V, I	V, I	V, S	V,I
No assumptions on cables	◐	●	●	○	●	●	●	●
Unbalanced poly-phase systems	○	○	○	○	○	●	○	●

∗Proposed methodology
○ = does not provide feature; ◐ = partially provides property;
● = provides property;
V = Voltage; I = Current; S = Power; E = Energy;

$x_{i,t}$ refers to the variable x_i at time t. The absence of such subscript denotes the entire time-series. Boldface characters, e.g. $\mathbf{x}$, are used to highlight complex values. To proceed in the problem formulation, let us consider separately the three phases of the network, as showed in Figure 1.

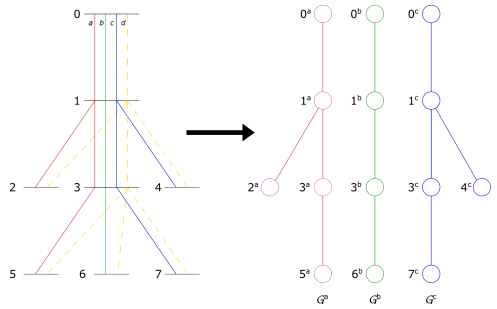

Figure 1. Graphical representation of $\mathcal{G}^a$, $\mathcal{G}^b$, $\mathcal{G}^c$ for a 8-node network.

Phase $p \in \mathcal{P}$ of the network is modeled as a rooted tree $\mathcal{G}^p = (\mathcal{N}^p, \mathcal{E}^p)$. $\mathcal{N}^p$ represents the subset of nodes connected to phase p of the substation while the set of edges $\mathcal{E}^p$ represents the phase p of the lines connecting those nodes. The root node of $\mathcal{G}^p$ is phase p of the MV/LV substation. Let $e = (i, j) \in \mathcal{E}^p$, with $i, j \in \mathcal{N}^p$, represent the branch connecting i to j, where i is the parent node and j is the child node. For every edge $e = (i, j) \in \mathcal{E}^p$, let $\mathcal{D}^p_e$ be the set of descendant nodes of j. Let $\mathbf{V}^p_{n,t}$ and $\mathbf{I}^p_{n,t}$ be the complex voltage and the complex current injection of node $n \in \mathcal{N}^p$ at phase p and time $t \in \{1, \dots, T\}$. Let $\mathcal{M}^p \subseteq \mathcal{N}^p$ be the set of nodes in $\mathcal{G}^p$ for which a meter is available. We assume that meters provide both voltage and current time-series measurements of the nodes to which they are connected. We refer to $\mathcal{H}^p = \mathcal{N}^p \setminus \mathcal{M}^p$ as the set of hidden nodes.

The voltage drop associated to an edge $e = (i, j) \in \mathcal{E}^p$ is:

$$\Delta\mathbf{V}^p_{e,t} = \mathbf{V}^p_{i,t} - \mathbf{V}^p_{j,t} \qquad \forall t \in \{1, \dots, T\} \tag{1}$$

Defining $\mathbf{Z}^p_e$ as the branch self-impedance, the voltage drop can be written as:

$$\Delta\mathbf{V}^p_{e,t} = \mathbf{Z}^p_e \mathbf{I}^p_{e,t} + \boldsymbol{\varepsilon}_{e,t} \qquad \forall t \in \{1, \dots, T\} \tag{2}$$

where $\mathbf{I}^p_{e,t}$ is the current flowing in the branch and the term $\boldsymbol{\varepsilon}_{e,t}$ is the contribution given by mutual and shunt impedances to the voltage drop. We assume that $|\boldsymbol{\varepsilon}_{e,t}| \ll |\mathbf{Z}^p_e \mathbf{I}^p_{e,t}|$.

The problem we want to address is the inference of $\mathcal{G}^a$, $\mathcal{G}^b$ and $\mathcal{G}^c$ assuming that:

- no *a priori* information about how many nodes are in the grid, how they are connected and cable characteristics is given;
- the topology does not change during the observation period;
- voltage and current injection time-series are known only for the subset of nodes $\mathcal{M}^a$, $\mathcal{M}^b$, $\mathcal{M}^c$;
- each phase of the substation is metered.

Obtaining an accurate estimation of $\mathcal{G}^a$, $\mathcal{G}^b$ and $\mathcal{G}^c$ allows to get, at the same time, information about the number of nodes in the grid, the topology of the network and the impedance of the lines.

IV. IDENTIFICATION ALGORITHM

The identification algorithm consists of three main parts named topology estimation, topology validation and hidden node detection, respectively, that are applied to each phase $p \in \mathcal{P}$ independently. Let $\hat{\mathcal{N}}^p \subseteq \mathcal{N}^p$ be the pool of nodes made available to the identification algorithm. $\hat{\mathcal{N}}^p$ initially corresponds to the set of observed nodes $\mathcal{M}^p$.

Topology Estimation: The first step of the algorithm consists in the estimation of a rooted tree $\hat{\mathcal{G}}^p = (\hat{\mathcal{N}}^p, \hat{\mathcal{E}}^p)$, the topology of which aims to be as close as possible to that of $\mathcal{G}^p = (\mathcal{N}^p, \mathcal{E}^p)$. This operation is carried out using correlation analysis on voltage measurements to infer proximity between nodes. Correlation-based approaches have already been proved successful both in phase [10] and topology [2] identification methods. Load and production profiles at different nodes of the LV network can present similar patterns due to comparable residential occupancy profiles and weather conditions. To be sure that this does not affect the results of the voltage correlation analysis, voltage time-series are pre-processed applying a high pass filter, as suggested in [2]. Let R_{ij} be the Pearson Correlation Coefficient between the filtered voltage time-series of nodes i and j. Let $\hat{\mathcal{G}}^p_w$ be a weighted complete graph built on $\hat{\mathcal{N}}^p$, where the edge weight of branch (i, j) is equal to R_{ij}. The estimated topology $\hat{\mathcal{G}}^p = (\hat{\mathcal{N}}^p, \hat{\mathcal{E}}^p)$ is obtained computing the *maximum spanning tree* on $\hat{\mathcal{G}}^p_w$.

Topology Validation: The second process checks each edge $e = (i, j) \in \hat{\mathcal{E}}^p$ and suggests where, if any, missing nodes and wrong connections are. This process is performed evaluating the estimated impedance time-series $\hat{\mathbf{Z}}^p_e = \{\hat{\mathbf{Z}}^p_{e,1}, \dots, \hat{\mathbf{Z}}^p_{e,T}\}$ for each edge $e \in \hat{\mathcal{E}}^p$ as:

$$\hat{\mathbf{Z}}^p_{e,t} = \frac{\Delta\hat{\mathbf{V}}^p_{e,t}}{\hat{\mathbf{I}}^p_{e,t}} \qquad t \in \{1, \dots, T\} \tag{3}$$

$\Delta\hat{\mathbf{V}}^p_{e,t}$ is the voltage drop associated to e. The estimated current $\hat{\mathbf{I}}^p_{e,t}$ flowing in e is obtained as the sum of the current injections of the set of nodes detected by the topology estimation as $\hat{\mathcal{D}}^p_e$:

$$\hat{\mathbf{I}}^p_{e,t} = \sum_{n \in \hat{\mathcal{D}}^p_e} \mathbf{I}^p_{n,t} \qquad t \in \{1, \dots, T\} \tag{4}$$

Exploiting Equation 2, we assume that, if edge $e = (i, j) \in \hat{\mathcal{E}}^p$ corresponds to a branch in the real network, the values in $\hat{\mathbf{Z}}_e^p$ tend all to be close to the same constant value. Edge e is labelled as valid if:

$$RSD(\hat{\mathbf{Z}}_e^p) \leq \lambda \tag{5}$$

where $RSD(\hat{\mathbf{Z}}_e^p)$ is the *relative standard deviation* of time-series $\hat{\mathbf{Z}}_e^p$, while λ is an arbitrary threshold, e.g.: determined by statistical tests, for the largest accepted relative standard deviation. Let $\hat{\mathcal{U}}^p \subseteq \hat{\mathcal{E}}^p$ be the set of rejected edges for which the inequality in (5) is not satisfied.

Hidden Node Detection: The third process of the algorithm exploits the results provided from the previous steps to find the location of hidden nodes. This step is based on the assumption that unmetered nodes are far enough from each other, formalized as follows:

Assumption 1: The distance on $\mathcal{G}^p$, defined as the minimum number of edges that connect two nodes, between any pair of hidden nodes i and j is greater than three.

Assumption 1 allows to locate each hidden node in the actual topology analyzing the metered nodes that surround it. Consider a hidden node $X \in \mathcal{H}^p$. If X is a node with a non-zero net power injection, the current flowing through the path that connects it to the root computed by the topology validation step is not the correct one, since the contribution of the unobserved $\mathbf{I}_X^p$ is missing. This causes the rejection of the edges in that path, as showed in Figure 2. Note also that, since node 5 is hidden and it is not a terminal node, the algorithm mistakenly detects an edge connecting node 3 to node 6. This reasoning suggests that, whenever there is a path of unvalidated edges in $\hat{\mathcal{G}}^p$, a missing node X is near the leaf node of such path. The algorithm focuses on one of the

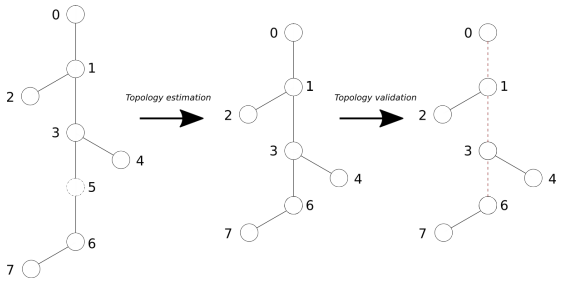

Figure 2. First two steps of the identification algorithm with node 5 as hidden node. Dashed edges are rejected by the validation step.

hidden nodes at a time to estimate its location and, if possible, its voltage and current injection. Then, we use these results to add such node to $\hat{\mathcal{N}}^p$ and to recompute again topology estimation and validation steps.

Let $A \in \hat{\mathcal{N}}^p$ be the node, among all the nodes that are connected to a rejected edge, with the longest path to the root (node 6 in Figure 2). Let B and C, with $B, C \in \hat{\mathcal{N}}^p$, be the parent and grandparent node of A, respectively. Using the previous reasoning, we assume that an unobserved node $X \in \mathcal{H}^p$ is adjacent to A in the original $\mathcal{G}^p$. This can occur in three topological configurations, as showed in Figure 3.

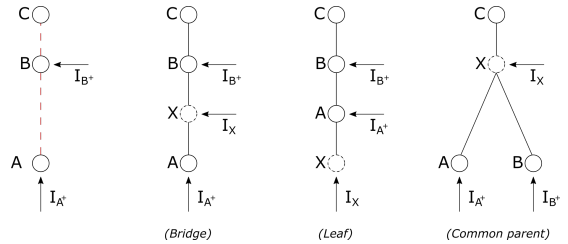

Figure 3. Possible locations for a hidden node X adjacent to A.

The hidden node detection step examines these three configurations, named bridge, leaf and common parent, to assess which suits best. Let $\hat{\mathbf{I}}_{A^+}^p$ and $\hat{\mathbf{I}}_{B^+}^p$ be defined as:

$$\hat{\mathbf{I}}_{A^+}^p = \hat{\mathbf{I}}_{(AB)}^p \tag{6}$$

$$\hat{\mathbf{I}}_{B^+}^p = \hat{\mathbf{I}}_{(AB)}^p - \hat{\mathbf{I}}_{(BC)}^p \tag{7}$$

$\hat{\mathbf{I}}_{A^+}^p$ is the contribution to the estimated $\hat{\mathbf{I}}_{(BC)}^p$ flowing through (BC) of A and its descendant nodes, while $\hat{\mathbf{I}}_{B^+}^p$ is the rest of the current. In order to detect the correct location of X, the algorithm solves three optimization problems:

Configuration "Bridge":

$$\begin{aligned} \underset{\hat{\mathbf{i}}_X^p, \hat{\mathbf{Z}}_{(XA)}^p, \hat{\mathbf{Z}}_{(BX)}^p, \hat{\mathbf{Z}}_{(CB)}^p}{\text{minimize}} \quad & \delta^b = \sum_{t=0}^{T} |\mathbf{V}_{C,t} - \mathbf{V}_{B,t} - \hat{\mathbf{Z}}_{(CB)}^p(\hat{\mathbf{I}}_{A^+,t}^p + \hat{\mathbf{I}}_{B^+,t}^p + \hat{\mathbf{I}}_{X,t}^p)| \\ \text{subject to} \quad & \mathbf{V}_{B,t} - \mathbf{V}_{A,t} = \hat{\mathbf{Z}}_{(BX)}^p(\hat{\mathbf{I}}_{A^+,t}^p + \hat{\mathbf{I}}_{X,t}^p) + \hat{\mathbf{Z}}_{(XA)}^p \hat{\mathbf{I}}_{A^+,t}^p \qquad t = 0, \ldots, T \end{aligned} \tag{8}$$

Configuration "Leaf":

$$\begin{aligned} \underset{\hat{\mathbf{i}}_X^p, \hat{\mathbf{Z}}_{(BA)}, \hat{\mathbf{Z}}_{(CB)}}{\text{minimize}} \quad & \delta^l = \sum_{t=0}^{T} |\mathbf{V}_{C,t} - \mathbf{V}_{B,t} - \hat{\mathbf{Z}}_{(CB)}^p(\hat{\mathbf{I}}_{A^+,t}^p + \hat{\mathbf{I}}_{B^+,t}^p + \hat{\mathbf{I}}_{X,t}^p)| \\ \text{subject to} \quad & \mathbf{V}_{B,t} - \mathbf{V}_{A,t} = \hat{\mathbf{Z}}_{(BA)}^p(\hat{\mathbf{I}}_{A^+,t}^p + \hat{\mathbf{I}}_{X,t}^p) \qquad t = 0, \ldots, T \end{aligned} \tag{9}$$

Configuration "Common parent":

$$\underset{\hat{\mathbf{Z}}_{(AX)}^p, \hat{\mathbf{Z}}_{(BX)}^p}{\text{minimize}} \quad \delta^{cp} = \sum_{t=0}^{T} |(\mathbf{V}_{A,t}^p - \hat{\mathbf{Z}}_{(AX)}^p \hat{\mathbf{I}}_{A^+,t}^p) - (\mathbf{V}_{B,t}^p - \hat{\mathbf{Z}}_{(BX)}^p \hat{\mathbf{I}}_{B^+,t}^p)| \tag{10}$$

Such problems are formulated by exploiting the currents and voltages relationship occurring in each configuration. Once problems (8 - 10) have been solved, the algorithm selects the configuration associated to the smallest $\delta \in \{\delta^b, \delta^l, \delta^{cp}\}$. Depending on the predicted location of X, the pool of nodes $\hat{\mathcal{N}}^p$ is updated accordingly.

If the algorithm picks configuration "Bridge", an additional node $\hat{X}$ is added to $\hat{\mathcal{N}}^p$. The estimated current injection $\hat{\mathbf{I}}_X^p$ is extracted from solution of (8), along with the values of $\hat{\mathbf{Z}}_{(XA)}^p$ and $\hat{\mathbf{Z}}_{(BX)}^p$. The estimation of the voltage time-series $\mathbf{V}_X^p$ is computed as:

$$\hat{\mathbf{V}}_{X,t}^p = \mathbf{V}_{A,t}^p - \hat{\mathbf{Z}}_{(XA)}^p \hat{\mathbf{I}}_{A^+,t}^p \qquad \forall t \in \{1, \ldots, T\} \tag{11}$$

If configuration "Leaf" is selected, the current injection $\hat{\mathbf{I}}_X^p$ is extracted from solution of (9), while the voltage time-series of X can not estimated. To proceed with the identification process, node $A \in \hat{\mathcal{N}}^p$ is substituted by an auxiliary node X' with the same voltage of A. The current injection is given by:

$$\hat{\mathbf{I}}_{X'}^p = \hat{\mathbf{I}}_A^p + \hat{\mathbf{I}}_X^p \tag{12}$$

This allows to correct the current flowing in the path to root of A for the next validation step and to proceed in the identification process.

When configuration "Common parent" is selected, the estimation of $\hat{\mathbf{V}}^p_X$ is constructed using the values of $\hat{\mathbf{Z}}^p_{(XA)}$ and $\hat{\mathbf{Z}}^p_{(XB)}$ obtained from solution of (10). We have:

$$\hat{\mathbf{V}}^p_{X,t} = \begin{cases} \mathbf{V}^p_{A,t} - \hat{\mathbf{Z}}^p_{(XA)}\hat{\mathbf{I}}^p_{A^+,t} & \text{if } \hat{\mathbf{Z}}^p_{(XA)}\hat{\mathbf{I}}^p_{A^+,t} \leq \hat{\mathbf{Z}}^p_{(XB)}\hat{\mathbf{I}}^p_{B^+,t} \\ \mathbf{V}^p_{B,t} - \hat{\mathbf{Z}}^p_{(XB)}\hat{\mathbf{I}}^p_{B^+,t} & \text{if } \hat{\mathbf{Z}}^p_{(XA)}\hat{\mathbf{I}}^p_{A^+,t} > \hat{\mathbf{Z}}^p_{(XB)}\hat{\mathbf{I}}^p_{B^+,t} \end{cases} \quad t \in \{1,\dots,T\} \tag{13}$$

A node $\hat{X}$ is added to $\hat{\mathcal{N}}^p$ with voltage $\hat{\mathbf{V}}^p_X$ and a zero net current injection, since (10) does not provide an estimation of the missing current injection. If X is indeed a net zero-power injecting node, edge (CX) is accepted in the topology by the validation step. Otherwise, the correct $\hat{\mathbf{I}}^p_{(CX)}$ is obtained solving (9) in the next iteration.

Whenever Assumption 1 is not valid, the hidden node detection step might fail to find a node X that corresponds to a hidden node in the actual network, since none of the topological configurations exploited to formulate (8 - 10) may be correct. Before updating the node pool with X, the algorithm checks if this makes the edge connecting C to its child node satisfy Equation (5). If not, the wrong X is not added to the network, since it would compromise the rest of the identification process. In this case, the algorithm proceeds considering the edge connecting A to B as a valid branch.

Once the pool of nodes $\hat{\mathcal{N}}^p$ has been updated, topology validation and estimation steps are processed again. The algorithm is performed until all the edges in $\hat{\mathcal{G}}^p$ are labeled as valid. Finally, the estimation of line impedances is given by the mean value of $\hat{\mathbf{Z}}^p_e$ in Equation (3). The complete identification algorithm is detailed in Algorithm 1.

V. Case Study

The developed algorithm is tested on a three-phase four-wire 30-node network whose topology is showed in Figure 4. Node S0 is the MV/LV substation, nodes from H1 to H24

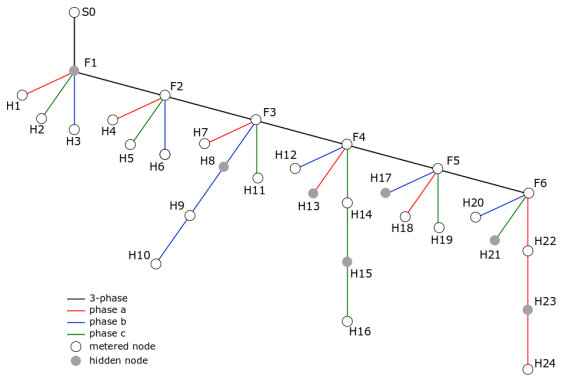

Figure 4. Topology of the test network used as the case study. Gray dots represent hidden nodes in the main test case.

are single-phase customers and nodes from F1 to F6 are their connection point to the main feeder. Household active power profiles are constructed from time-series measurements of residential Belgian smart-meters. Reactive power consumption

Algorithm 1 Identification algorithm for phase $p \in \mathcal{P}$

Input:

- Set of nodes $\mathcal{M}^p$, time-series synchronized measurements of $\mathbf{V}^p_n$ and $\mathbf{I}^p_n$, $\forall n \in \mathcal{M}^p$.

Output:

- A rooted tree $\hat{\mathcal{G}}^p = \{\hat{\mathcal{N}}^p, \hat{\mathcal{E}}^p\}$ with a set of impedances associated to $\hat{\mathcal{E}}^p$.

Procedure:

1: Let $\hat{\mathcal{N}}^p = \mathcal{M}^p$ and $\hat{\mathcal{U}}^p = \emptyset$.
2: Apply an high pass filter on $|\mathbf{V}^p_i|$ to obtain a filtered time series $\mathbf{V}^{p,f}_n$, $\forall n \in \hat{\mathcal{N}}^p$.
3: Evaluate the Pearson correlation coefficient $R_{i,j}$, $\forall i,j \in \mathcal{M}_p$ with $i \neq j$.
4: Let $\hat{\mathcal{G}}^p_w$ be the weighted complete graph on $\hat{\mathcal{N}}^p$ where each branch (i,j) has R_{ij} as its weight.
5: Let $\hat{\mathcal{G}}^p = \{\hat{\mathcal{N}}^p, \hat{\mathcal{E}}^p\}$ the maximum spanning tree computed on $\hat{\mathcal{G}}^p_w$.
6: Let $\hat{\mathcal{U}}^p = \{e \in \hat{\mathcal{E}}^p \mid RSTD(\hat{\mathbf{Z}}^p_e) > \lambda\}$.
7: **while** $\hat{\mathcal{U}}^p \neq \emptyset$:
  a: Let A be the node in with the maximum depth $\hat{\mathcal{G}}^p$ among the nodes connected to an edge $e \in \hat{\mathcal{U}}^p$.
  b: Let B and C be the parent and grandparent node of A, respectively.
  c: Find δ^b, δ^l, δ^{cp} from (8) — (10).
  d: **if** $argmin\{\delta^b, \delta^l, \delta^{cp}\} = \delta^b$:
    (i) Evaluate $\hat{\mathbf{V}}^p_X$ from (11), $\hat{\mathbf{I}}^p_X$ from (8).
    elif $argmin\{\delta^b, \delta^l, \delta^{cp}\} = \delta^l$:
    (i) Let $\hat{\mathbf{V}}^p_X = \mathbf{V}^p_A$, evaluate $\hat{\mathbf{I}}^p_X$ from (9).
    (ii) $\hat{\mathcal{N}}^p = \hat{\mathcal{N}}^p \setminus \{A\}$.
    elif $argmin\{\delta^b, \delta^l, \delta^{cp}\} = \delta^{cp}$:
    (i) Evaluate $\hat{\mathbf{V}}^p_X$ from (13), let $\hat{\mathbf{I}}^p_X = \mathbf{0}$.
  e: Update $\hat{\mathcal{N}}^p = \hat{\mathcal{N}}^p \cup \{X\}$ and repeat steps 2 — 6.
8: Evaluate $\mathbf{Z}^p_e$ as the mean value of $\hat{\mathbf{Z}}^p_e = \{\hat{\mathbf{Z}}^p_{e,1}, \dots, \hat{\mathbf{Z}}^p_{e,T}\}$ from Equation (3), $\forall e \in \hat{\mathcal{E}}^p$.

is obtained associating a different power factor ranging from 0.93 to 0.97 to each of the households. Some of the households are also equipped with PV units. Peaks and time patterns in power profiles vary widely from a household to another, therefore unbalanced conditions are expected in the network. Two different cable types and various cable length have been selected for the modeling of the main feeder and the laterals. Voltages and currents in the grid are computed through a detailed unbalanced load flow algorithm. The dynamics of the network are evaluated over 15 days, with a resolution of one simulation every 15 minutes, resulting in 1440 time steps. Time-series of voltage and current injection at each node and phase of the network are extrapolated from the solutions of the load flow analysis to create pseudo-measurements data. The tolerated relative standard deviation for impedances in Equation (5) is arbitrarily set to 0.1.

To assess the performances of the algorithm, a total of

830 configurations have been examined, each of which is associated to a different set of hidden nodes. Among them, 30 configurations represented the cases with one hidden node. The remaining 800 configurations were equally divided in 8 groups with an increasing number of hidden nodes that ranges from 2 to 9. Note that when a a node is considered as hidden, the identification process is provided with no information about their presence, voltages and current injections in the input. If an hidden node is a feeder node, e.g.: F1, none of the measurements for three phases is provided to the algorithm. The sets of hidden nodes have been randomly selected, having the effect that Assumption 1 is not valid in most of configurations. In particular, none of the scenarios in with more than 7 hidden nodes satisfied such assumption.

Let us first focus on a main test case to analyze the results provided by the algorithm in configurations where Assumption 1 is satisfied. The configuration where nodes F1, H7, H13, H15, H17, H21 and H23 are hidden nodes has been selected. In this scenario the algorithm was able to produce an exact representation of the network topology for each phase, as can be seen in Figure 5. All the missing nodes have been correctly detected and placed in the right location. The three single-graphs can be merged together to obtain the topology of the three-phase network.

As for the assessment of impedances, they are evaluated for all the lines except for F6-H21, F5-H17 and F4-H13. The estimation of the impedance for branches F6-H21, F5-H17 and F4-H13 is not determined because nodes H21, H17 and H13 were hidden leaf nodes and therefore the voltage drop of such edges can not be computed. The estimated impedance for some of the branches is showed in Table II. The estimated resistive part is close to the exact value for all the branches, while the evaluation of the reactive part is less precise for branches connecting the feeder to the households. This error is probably due a stronger influence of the modeled shunt admittances on the voltage drop on those connection. If the ratio between the resistive and reactive part of the impedance is known, e.g.: databases contain information on the type of cables used, such error can be reduced.

Table II
IMPEDANCE ESTIMATION FOR SOME BRANCHES IN THE MAIN TEST CASE.

Branch	Value (Ω)	Estimated (Ω)
S0-F1	0.014 + 0.046j	0.013 + 0.042j
F1-F2	0.015 + 0.052j	0.015 + 0.048j
F1-H1	0.067 + 0.033j	0.066 + 0.003j
F2-H6	0.030 + 0.015j	0.029 + 0.002j
F3-H7	0.025 + 0.012j	0.024 + 0.002j
H8-H9	0.039 + 0.019j	0.038 + 0.002j
F4-H12	0.034 + 0.016j	0.033 + 0.002j
F4-F5	0.017 + 0.058j	0.017 + 0.053j
F5-H18	0.044 + 0.026j	0.043 + 0.003j
F6-H22	0.046 + 0.022j	0.045 + 0.003j
H23-H24	0.032 + 0.015j	0.031 + 0.002j

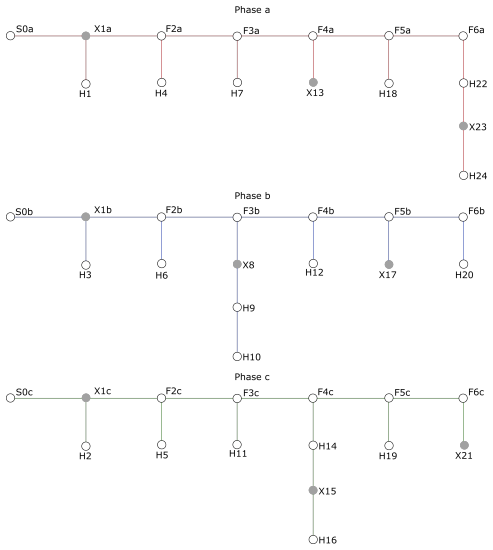

Figure 5. Estimated topology in the main test case.

Two metrics are considered to analyze the solutions provided by the algorithm in all 830 configurations. The first metric is the percentage of the network branches that have been correctly detected by the algorithm. This value provide an assessment of the quality for the estimation of the network topology. Note that for this evaluation, a three-phase branch is considered as detected only if the nodes at its ends are directly connected by an edge in all $\hat{\mathcal{G}}^a$, $\hat{\mathcal{G}}^b$ and $\hat{\mathcal{G}}^c$.

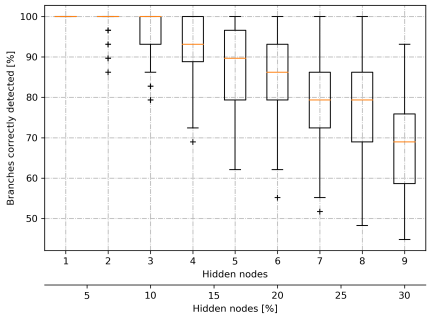

Figure 6. Percentage of branches correctly detected by the algorithm.

As showed in Figure 6, the algorithm have identified the correct topology in most of the configurations with less than 10% of hidden nodes nodes in the network. With sets of hidden nodes corresponding from 10% to 20% of the entire set of nodes, Assumption 1 is very likely to be not valid. Nevertheless, the algorithm still performed well, accurately detecting more than the 80% of network branches in most of the cases. Test cases with 7 and 8 hidden nodes still produced the exact topology in some of the configurations, but the average accuracy decreases in the rest of the cases. The algorithm was not able to find the exact topology in any configuration with 9 hidden nodes, corresponding to the 30% of the nodes, but it was still able to correctly identify more than half of the network's branches in most cases.

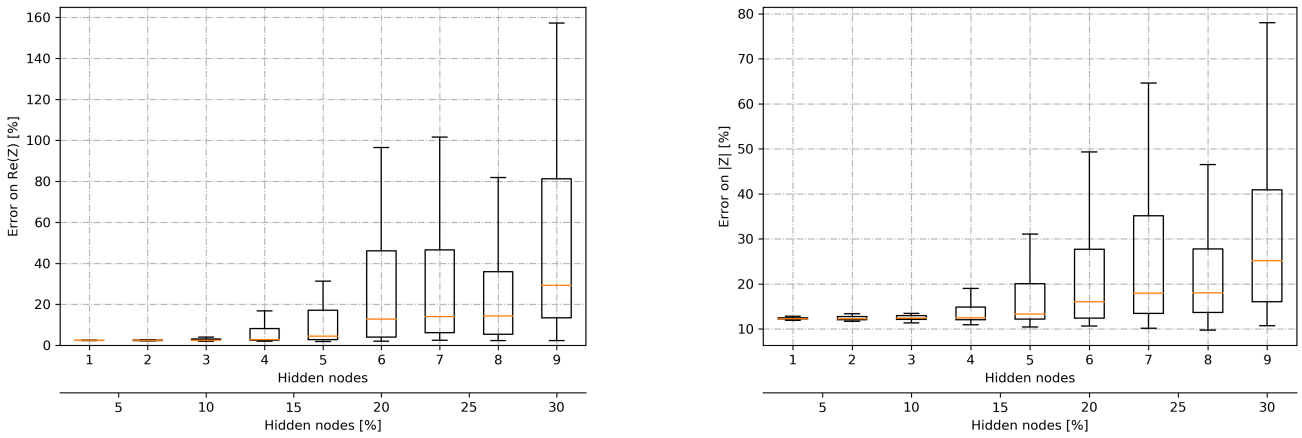

Figure 7. Mean absolute percentage error on the cumulative impedance of the path to the root of metered nodes.

To evaluate the quality of line parameters assessment, we compared the estimation of the impedance between a metered node and the substation, evaluated as the sum of the self impedances in its path to the root, with its exact value. The mean absolute percentage error, evaluated both on the magnitude and on the real part of the impedance, has been used as a second error metric. Results are presented in Figure 7. In all cases with less than 10% of hidden nodes, accuracy in the assessment of the impedances was comparable to that observed in the main test case. As already stated, the error on the estimation of the self impedances that is observed in these solutions is probably due to the approximation made for Equation (2). For configurations with more hidden nodes, the algorithm provided similar results only in few cases. However it has been observed that, whenever Assumption 1 was not satisfied, the algorithm was not able to identify location and current injection of some of the hidden nodes. When this happens, the estimated topology and currents flowing in some sections of the network are miscalculated, and therefore the error on the computation of the impedance largely increases.

VI. Conclusion

In this work, the problem of reconstructing a model for LV networks, identifying topology and line parameters, has been addressed. We propose an algorithm to tackle the network reconstruction problem by analyzing available meter data when some power injecting nodes in the network are not metered, assuming that such nodes are located reasonably far from each other. The algorithm detects the location of hidden nodes from an approximate estimation of the topology and branch impedances, then it reconstructs the variables associated to hidden nodes, one node per iteration. A case study on a 30-node network has been described to test the algorithm in different scenarios. A main test case with 7 hidden nodes where the condition on the minimum distance between such nodes was valid has been presented, showing that the algorithm provided an accurate estimation of topology and line parameters. Additional configurations have been then investigated to analyze the performances of the proposed methodology and highlight its limitations when the condition on the minimum distance between hidden nodes is not satisfied. When applied to these test cases, the algorithm provided relevant solutions in most of the scenarios with 20% or less of hidden nodes. Future work could focus on the reconstruction of networks where a larger section of the grid can not be observed, exploiting additional information available to DSOs, such as GIS data and cable characteristics. Additional effort could be put into a more exhaustive estimation of the line parameters, taking into account the evaluation of mutual and shunt impedances. The case when meters provide only the magnitude of the complex variables could also be considered.

VII. Acknowledgments

The authors thank RESA for providing data and constructive discussions that greatly contributed to this work.

References

[1] J. J. Grainger and W. D. Stevenson Jr., “Power System Analysis”, McGrawHill, Inc., 1994.

[2] S. Bolognani, N. Bof, D. Michelotti, R. Muraro and L. Schenato, “Identification of power distribution network topology via voltage correlation analysis,” IEEE 52nd Conference on Decision and Control, 2013.

[3] A. Monticelli, “Electric power system state estimation,” Proceedings of the IEEE, 2000.

[4] Y. Weng, Y. Liao and R. Rajagopal, “Distributed energy resources topology identification via graphical modeling,” IEEE Transactions on Power Systems, 2016.

[5] S.J. Pappu, N. Bhatt, R. Pasumarthy and A.Rajeswaran, “Identifying topology of low voltage distribution networks based on smart meter data,” IEEE Transactions on Smart Grid, 2017.

[6] K. Soumalas, G. Messinis and N. Hatziargyriou, “A data driven approach to distribution network topology identification,” 2017 IEEE Manchester PowerTech, 2017.

[7] Y. Yuan, O. Ardakanian, S. Low And C. Tomlin, “On the inverse power flow problem,” arXiv preprint arXiv:1610.06631, 2016.

[8] O. Ardakanian, Y. Yuan, V. Wong, R. Dobbe, S. Low, A. von Meier and C. J. Tomlin, “On identification of distribution grids,” IEEE Transactions on Control of Network Systems, 2019.

[9] J. Yu, Y. Weng and R. Rajagopal, “PaToPa: A data-driven parameter and topology joint estimation framework in distribution grids,” IEEE Transactions on Power Systems, 2017.

[10] F. Olivier, A. Sutera, P. Geurts, R. Fonteneau and D. Ernst, “Phase identification of smart meters by clustering voltage measurements,” 2018 Power Systems Computation Conference (PSCC), 2018.