Urban Distribution Grid Topology Reconstruction via Lasso

Yizheng Liao
Dept. of Civil and Environmental Eng.
Stanford University
Stanford, CA 94305
Email: yzliao@stanford.edu

Yang Weng
Dept. of Civil and Environmental Eng.
Stanford University
Stanford, CA 94305
Email: yangweng@stanford.edu

Ram Rajagopal
Dept. of Civil and Environmental Eng.
Stanford University
Stanford, CA 94305
Email: ramr@stanford.edu

***Abstract*—The growing integration of distributed energy resources (DERs) in urban areas raises various reliability issues. To ensure robust distribution grid operation, grid monitoring tools are needed, where the topology reconstruction serves as the first step. However, the topology reconstruction is hard in distribution grid. This is because 1) the branches are difficult and expensive to monitor since most of them are underground in urban areas; and 2) the assumption of radial topology in many studies is inappropriate for meshed urban grids. To address these drawbacks, we propose a new data-driven approach to reconstruct distribution grid topology by utilizing the newly available smart meter data. Specifically, a graphical model is built to model the probabilistic relationships among different voltage measurements. With proof, the topology reconstruction problem is formulated as a regularized linear regression problem (Lasso) to deal with meshed network structures. Simulation results show highly accurate estimation in IEEE standard distribution test systems with and without loops using real smart meter data.**

I. INTRODUCTION

Providing sustainable and economic energy is one of the key missions for Smart Cities. This is because the large scale integration of distributed energy resources (DERs) not only creates more sustainable energy sources but also reduces the electricity cost and the transmission loss. To achieve this goal, many cities have proposed their plans to integrate DERs, such as photovoltaics, electric vehicles, and energy storage devices. Examples include Smart Cities San Diego Project, Amsterdam Smart City Project, and the Zurich 2000 Watts Society.

While producing new opportunities, the large-scale penetration of DERs also poses new challenges to the distribution system operation. On the generation side, the reverse power flow can render the existing protective systems inadequate. Without appropriate monitoring and controls, even a small-scale DER integration can destabilize the local grid and cause reliability issues for customers [1]. On the demand side, the frequent plug-and-play electric vehicles will impact the distribution grid power quality such as voltage unbalance and transformer overload [2]. To better control the distribution grid on both generation and demand sides, grid monitoring tools (such as state estimation) are needed, where topology information is a prerequisite. The topology information is also a basis in determining various effects caused by the integration of DERs, such as islanding and line work hazards. Therefore, a highly active and accurate topology estimation process is necessary for distribution grid operation.

In transmission grid, its topology has infrequent changes and can be identified by the topology processor [3], where topology errors are determined by a post-state estimation procedure [4]. Unfortunately, in secondary distribution grids, these methods have limited performance due to various reasons. For example, the secondary distribution grids are immense and mostly underground in many major metropolitan areas, such as New York City, Chicago, and San Francisco. According to [5], in New York City, the total length of underground power cables is about 94000 miles. This fact makes installing any topology identification device time consuming and expensive. Even worse, the reconfiguration of underground distribution grids [6] makes the methods developed for overhead transmission grids unsuitable.

These difficulties make the assumptions in past studies on topology estimation invalid. For instance, [7], [8] assume the availability of the switch connectivity map and search for the right combination. State estimation-based algorithm [9] and power flow-based algorithm [10] require the admittance matrix. These assumptions become inaccurate in urban networks due to the unavailability of switch connectivity maps and admittance matrices in newly reconfigured distribution network. Even if this information is available, it may be outdated or incorrectly documented due to human interaction without information updating. For example, as many DERs do not belong to the utilities, their connection statuses may not be reported to the utilities. Also, these approaches only focus on radial networks but many secondary distribution grids in urban areas are meshed [5]. Furthermore, the meshed structures will be ubiquitous [6] with the integration of DERs. Therefore, the capability to identify topology with loops is important in urban system topology estimation.

Fortunately, the smart meters at households enable a new opportunity to utilize the time-series data, which are unavailable previously in electric power industry [11], to solve new problems. One application of using these data is reconstructing the distribution grid topology in a data-driven manner, because branch data in secondary grids are difficult and expensive to obtain [1] for their underground wiring. Hence, in this paper, we restrict us to only use end-user data [12], which are easy to obtain. These data are measured by smart meters and include voltage magnitude and real/reactive powers. We aim at helping these devices identify the physical system in which they operate and discover their neighbor buses.

In this paper, we prove that a distribution grid can be modeled as a probabilistic graphical model. Then, a topology estimation algorithm is proposed based on the L_1-penalized linear regression, which is known as a Lasso problem and is fitted by the historical data. Our algorithm has several advantages. Firstly, it can reconstruct a loopy network in a short time. Secondly, our approach does not allow error propagation, which occurs in other methods [13]. This is because the connectivity is recovered at each bus individually. Thirdly, the proposed algorithm can be implemented in parallel, which reduces the computational complexity significantly.

Whereafter, the performance of our data-driven method is verified by simulations on the standard IEEE 8- and 123-bus distribution test cases [14], with and without loops. Real smart meter data collected by Pacific Gas and Electric Company (PG&E) and emulated rooftop solar generation data [15] are utilized for simulation. Our algorithm outperforms the approaches in recent papers [13], [16] in meshed networks.

The rest of the paper is organized as follows: Section II introduces the modeling and the problem of data-driven topology identification. Section III uses a proof to justify that the topology can be efficiently reconstructed by regularized linear regression. A detailed algorithm is illustrated as well. Section IV evaluates the performance of the new method and Section V concludes the paper.

II. System Model

In order to formulate the topology reconstruction problem, firstly, we need to describe the distribution grid and data. A distribution grid is defined as a physical network with buses and branches that connect buses. To utilize the time series data provided by smart meters, we construct a graphical model $G = (\mathcal{S}, \mathcal{E})$ where $\mathcal{S} = \{1, 2, \ldots, p\}$ is the set of the vertices and $\mathcal{E} = \{e_{ij}, i, j \in \mathcal{S}\}$ represents the set of the edges. Since edges are unidirectional in our model, $e_{ij} = e_{ji}$. In our graphical model, a node is corresponding to a bus in the physical layer and is modeled as a random variable V_s. The edge that connects node i and j represents the statistical dependence between the measurements collected at bus i and j. The physical network and the graphical model can be visualized in Fig. 1. At time t, the noiseless voltage measurement at bus s is $v_s^t = |v_s^t|e^{j\theta_s^t} \in \mathbb{C}$, where $|v_s^t| \in \mathbb{R}$ denotes the voltage magnitude in per unit and $\theta_s^t \in \mathbb{R}$ denotes the voltage phase angle in degrees. These measurements are in the steady state and all voltages are sinusoidal signals at the same frequency.

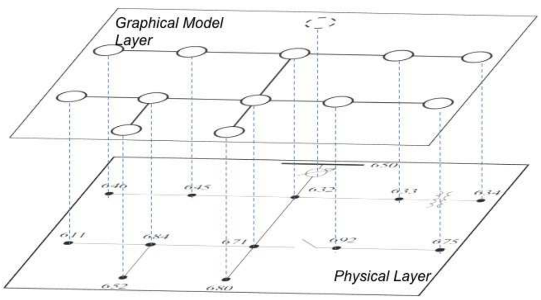

Fig. 1. Physical network with graphical model layer.

The problem of distribution grid topology reconstruction is defined as follows:

- Problem: data-driven topology reconstruction based on voltages
- Given: a sequence of historical measurements $v_s^t, s \in \mathcal{S}, t = 1, \ldots, N$ and a partially known grid topology, as shown in Fig. 2
- Find: the grid topology $\mathcal{E}$ in the dashed box in Fig. 2

III. Topology Reconstruction with Lasso

In our graphical model, buses are modeled as random variable.s Therefore, we use a joint probability distribution to represent the interdependency among buses:

$$\begin{aligned} P(\mathbf{V}_{\mathcal{S}}) &= P(V_2, V_3, \ldots, V_p) \\ &= P(V_2)P(V_3|V_2)\ldots P(V_p|V_2, \ldots, V_{p-1}). \end{aligned}$$

Bus 1 is the slack bus, which is a constant with a unit magnitude and zero phase angle. Therefore, it is omitted from the joint probability distribution above. In a distribution grid,

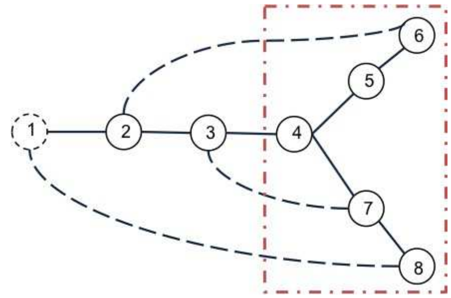

Fig. 2. A graphical model of IEEE 8-bus system with loops. The dashed edges are additional branches to form loops. The branches within the red dashed box are assumed to be unknown.

the correlation between interconnected neighboring buses are higher than that between non-neighbor buses. Therefore, a reasonable approximation is

$$P(\mathbf{V}_{\mathcal{S}}) \simeq \prod_{s=2}^{p} P(V_s|\mathbf{V}_{\mathcal{N}(s)}), \tag{1}$$

where the neighbor set $\mathcal{N}(s)$ contains the vertices of the neighbors of V_s, i.e. $\mathcal{N}(s) = \{t \in \mathcal{S}|e_{st} \in \mathcal{E}\}$. With this approximation, finding grid topology is equivalent to finding the neighbors of each bus. We are not restricted to parent nodes here because the topology in a secondary distribution grid can be both radial and loopy. We want to propose an algorithm that works for both types of network. In the following, we will show that, with an appropriate assumption, V_s only has statistical dependence with its neighbors, and explain why we can approximate $P(\mathbf{V}_{\mathcal{S}})$ in such a way. In the following context, the operator $\setminus$ denotes the complement operator, i.e. $\mathcal{A}\setminus\mathcal{B} = \{i \in \mathcal{A}, i \notin \mathcal{B}\}$.

Theorem 1. *In a distribution grid, if the current injection at each bus is approximately independent, the voltage of bus s and the voltages of all other buses that are not connected with bus s are conditionally independent, given the voltages of the neighbors of bus s, i.e.* $V_s \perp \{V_u, u \in \mathcal{S}\setminus\{\mathcal{N}(s), s\}\}|V_{\mathcal{N}(s)}$.

Proof: In this proof, we will firstly show the conditional independence for a simple example. Then we will prove the conditional independence in a generalized distribution network.

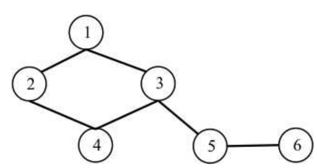

Fig. 3. A 6-bus system.

For the network in Fig. 3, the circuit equation $Y\mathbf{V} = \mathbf{I}$ is equivalent to:

$$\begin{bmatrix} y_{11} & -y_{12} & -y_{13} & 0 & 0 & 0 \\ -y_{12} & y_{22} & 0 & -y_{24} & 0 & 0 \\ -y_{13} & 0 & y_{33} & -y_{34} & -y_{35} & 0 \\ 0 & -y_{24} & -y_{34} & y_{44} & 0 & 0 \\ 0 & 0 & -y_{35} & 0 & y_{55} & -y_{56} \\ 0 & 0 & 0 & 0 & -y_{56} & y_{66} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \end{bmatrix},$$

where $y_{ij} = y_{ji}$ denotes the admittance between bus i and j, and the self-admittance is defined as $y_{ii} = \sum_{j=1, i\neq j}^{6} y_{ij}$. If $y_{ij} = 0$, there is no branch between bus i and j.

Let's pick up bus 1 as an example. For bus 1, the neighbor set $\mathcal{N}(1)$ is $\{2, 3\}$. Given $V_2 = v_2$ and $V_3 = v_3$, we have following equations:

$$I_1 + v_2 y_{12} + v_3 y_{13} = V_1 y_{11} \quad (2)$$

$$I_4 + v_2 y_{24} + v_3 y_{34} = V_4 y_{44} \quad (3)$$

$$I_5 + v_3 y_{35} = V_5 y_{55} - V_6 y_{56} \quad (4)$$

$$I_6 = -V_5 y_{56} + V_6 y_{66} \quad (5)$$

For bus 4, due to the assumption of current injection independence, i.e. $I_1 \perp I_4$, V_1 and V_4 are conditionally independent given V_2 and V_3. For bus 5 and bus 6, $V_1|\{V_2, V_3\}$ and $\{V_5, V_6\}|\{V_2, V_3\}$ are independent because of the assumption of current injections. Therefore, for this example, we prove that the conditional independence holds.

Let's extend the proof to a more general case. For a grid with p buses, the current and voltage relationship of bus s can be written as $I_s = V_s y_{ss} - \sum_{i \in \mathcal{N}(s)} V_i y_{si}$ with $y_{ss} = \sum_{i \in \mathcal{N}(s)} y_{si}$. Given $V_i = v_i$ for all i in $\mathcal{N}(s)$, the equation above becomes to

$$I_s + \sum_{i \in \mathcal{N}(s)} v_i y_{si} = V_s y_{ss}. \quad (6)$$

For bus u, which is not connected with bus s, i.e. $u \in \mathcal{S} \backslash \{\mathcal{N}(s), s\}$, we have a similar equation:

$$I_u + \sum_{j \in \mathcal{N}(u)} V_j y_{uj} = V_u y_{uu}. \quad (7)$$

The relationship between the neighbor sets $\mathcal{N}(s)$ and $\mathcal{N}(u)$ can be divided into two scenarios and we will discuss each of them individually.

If $\mathcal{N}(u) \cap \mathcal{N}(s) = \emptyset$, (7) remains the same. Therefore, V_s and $\{V_u, \mathbf{V}_{\mathcal{N}(u)}\}$ are conditionally independent given $\mathbf{V}_{\mathcal{N}(s)}$. This is similar to V_6 in (5).

If $\mathcal{N}(u) \cap \mathcal{N}(s) \neq \emptyset$, (7) becomes

$$I_u + \sum_{j \in \mathcal{N}(s) \cap \mathcal{N}(u)} v_j y_{uj} = V_u y_{uu} - \sum_{j \in \mathcal{N}(u) \backslash \mathcal{N}(s)} V_j y_{uj}.$$

Hence, with the same reason, V_s and $\{V_u, \mathbf{V}_{\mathcal{N}(u) \backslash \mathcal{N}(s)}\}$ are independent conditional on $\mathbf{V}_{\mathcal{N}(s)}$. Therefore, the conditional independence is proved in this case. This condition is similar to V_5 in (4). Also, V_4 in (3) is a special case where $\mathcal{N}(u) = \mathcal{N}(s)$. In conclusion, given $\mathbf{V}_{\mathcal{N}(s)}$, V_s is conditionally independent with $\mathbf{V}_{\mathcal{S} \backslash \{\mathcal{N}(s), s\}}$.

■

With Theorem 1, we can claim that with the assumption of current injection independence, the joint distribution in (1) can be rewritten as $P(\mathbf{V}_{\mathcal{S}}) = \prod_{s=2}^{p} P(V_s | \mathbf{V}_{\mathcal{N}(s)})$. Therefore, finding the topology is equivalent to finding the neighbor set $\mathcal{N}(s)$ for each bus s. In following section, we will propose an efficient algorithm to find $\mathcal{N}(s)$.

A. Topology Reconstruction via Conditional Inference

For a given vertex s, $\mathbf{V}_{\mathcal{S} \backslash \{s\}}$ denotes the collection of all other random variables in the graph. If we assume $\mathbf{V}_{\mathcal{S}}$ follows a multivariate Gaussian distribution, the conditional distribution of V_s given $\mathbf{V}_{\mathcal{S} \backslash \{s\}}$ is also Gaussian distribution. Therefore, based on the Gaussian probability density function, V_s can be described by a linear equation based on $\mathbf{V}_{\mathcal{S} \backslash \{s\}}$ and an error term, as shown below:

$$V_s = \mathbf{V}_{\mathcal{S} \backslash \{s\}}^T \beta^s + W_{\mathcal{S} \backslash \{s\}}, \quad (8)$$

where β^s denotes the parameter vector. In this linear equation, $W_{\mathcal{S} \backslash \{s\}}$ is a zero-mean Gaussian variable with $\text{Var}(W_{\mathcal{S} \backslash \{s\}}) = \text{Var}(V_s | \mathbf{V}_{\mathcal{S} \backslash \{s\}})$ and is independent with $\mathbf{V}_{\mathcal{S} \backslash \{s\}}$. As shown in [17], the nonzero coefficient in β^s indicates the statistical dependence between two nodes. Therefore, the connectivity identification problem becomes a linear regression problem. For each bus s, we can use the estimated parameter vector $\widehat{\beta^s}$ to find its neighbors.

A typical distribution grid is not fully connected. Therefore, the graph is sparse and many coefficients in β^s are zero. A widely used constraint to ensure the sparsity is L_1 norm because it leads to a convex optimization problem and can be solved efficiently. This type of problem is also known as Lasso [18]. It minimizes the sum of squared errors with a bound on the sum of the absolute values of coefficients (L_1 norm). With L_1 norm penalty, the linear regression in (8) is formulated as:

$$\widehat{\beta^s} = \underset{\beta^s(1)=0}{\text{argmin}} \left\{ \sum_{t=1}^{N} (v_s^t - (\mathbf{v}_{\mathcal{S} \backslash \{s\}}^t)^T \beta^s)^2 + \lambda \|\beta^s\|_1 \right\}, \quad (9)$$

where $\lambda \geq 0$ denotes the regularization parameter and $\|.\|_1$ is L_1 norm. The term $\|\beta^s\|_1$ is called the penalty term. If $\lambda = 0$, the optimization problem in (9) is a standard least squares problem. After solving the Lasso problem, we can find the neighbor set $\widehat{N(s)}$ which contains the indices of all non-zero coefficients in $\widehat{\beta^s}$.

An edge between bus i and j is visited twice because each bus has an individual estimate $\widehat{\beta}$. A simple way to combine both estimates is AND rule, $\hat{e}_{ij} = \widehat{\beta_j^i} \wedge \widehat{\beta_i^j}$, where $\wedge$ denotes the logical "and" operator.

The AND rule provides reliable results but also has a high probability to miss an edge. This is because an edge is decided only when both estimates have an agreement. If one estimate has an error, an edge will be missed. To address this drawback, we propose another rule called OR rule, $\hat{e}_{ij} = \widehat{\beta_j^i} \vee \widehat{\beta_i^j}$, where $\vee$ denotes the logical "or" operator.

A disadvantage of both AND rule and OR rule is that they cannot guarantee the estimated graph is a connected graph. It means that some buses form a small graph and are isolated from other buses. To overcome this drawback, we use both rules to decide an edge. We propose a new rule, AND-OR rule, to merge the results of AND and OR rules. This new rule has multiple steps. Firstly, the AND rule is applied and produces an estimated edge set $\widehat{\mathcal{E}_{\text{AND}}}$. Then we apply a physical law to diagnosis the estimated edges. In power system, differences in voltage magnitudes drive the flow of reactive power [19]. Therefore, for bus s, unless it is a substation, it should have at least one neighbor with higher voltage magnitude because the reactive power needs to be supplied from a bus with higher voltage magnitude. If none of the neighbors of bus s is satisfied, we can conclude that bus s is not connected to the main grid. Therefore, we apply the OR rule with some modifications to find a new neighbor bus j, as shown below:

$$\hat{e}_{sj} = (\hat{\beta}_j^s \vee \hat{\beta}_s^j) \wedge \neg(\hat{\beta}_j^s \wedge \hat{\beta}_s^j) \wedge (\hat{E}(V_j) > \hat{E}(V_s)), \quad (10)$$

where $\neg$ denotes the logic "not" operator and $\hat{E}$ denotes the empirical mean. The estimated edge set is denoted as $\widehat{\mathcal{E}_{\text{AND-OR}}}$. These steps are summarized in a flow chart in Fig. 4. In Section IV, we will show that the AND-OR rule is more robust than the AND and OR rules.

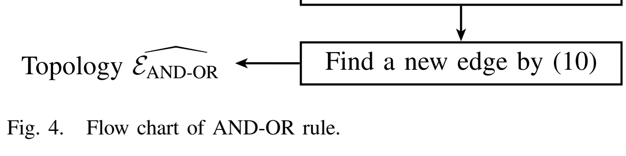

Fig. 4. Flow chart of AND-OR rule.

In distribution grid, phase angles are hard to obtain because the phasor measurement units (PMUs) have not been widely installed. Since the phase angles have small variation within a distribution grid, we can only use voltage magnitude for topology reconstruction. The steps of our proposed algorithm for topology reconstruction are summarized in Algorithm. 1.

Algorithm 1 Distribution Grid Topology Reconstruction via Lasso

---

Require: $|v_s^t|$ for $s = 2, \dots, p$, $t = 1, \dots, N$

1: Normalize and standardize $|V_s|$ such that it has zero mean and unity variance.
2: **for** $s = 2, \dots, p$ **do**
3: Solve the Lasso optimization problem in (9) for bus s and find the parameter vector estimate $\widehat{\beta^s}$.
4: **end for**
5: Combine the parameter vectors $\{\widehat{\beta^s}, s \in \mathcal{S}\}$ via the AND-OR rule to form an edge estimate $\widehat{\mathcal{E}_{\text{AND-OR}}}$.

An efficient algorithm to solve Lasso problem is Least Angle Regression (LAR) [20]. For a network with p buses and N observations, the computational complexity of LAR at each bus is $\mathcal{O}(p^3 + Np^2)$, which is as same as the least squares regression [20]. Therefore, the overall computational complexity of the proposed algorithm is $\mathcal{O}(p^4 + Np^3)$. Since the for loop between Step 2 and Step 4 can be computed in parallel, thus, we can maintain the same complexity as a single linear regression problem.

B. Choice of Regularization Parameter

The choice of λ is critical in Lasso because it affects to the number of non-zero coefficients in β^s. When λ is small, the penalty term has no effect and the solution is close to the least squares solution. When λ is large, some coefficients of $\widehat{\beta^s}$ are zeros. To control the probability of false alarm α, we introduce Lemma 1 to find the optimal λ for a given false alarm rate.

Lemma 1. *Given the probability of false alarm $0 \le \alpha \le 1$, the regularization parameter λ for bus s is*

$$\lambda^s(\alpha) = \frac{2\hat{\sigma}_s}{\sqrt{N}} \tilde{\Phi}^{-1}\left(\frac{\alpha}{2p^2}\right), \tag{11}$$

where $\tilde{\Phi} = 1 - \Phi$, Φ is the cumulative distribution function of $\mathcal{N}(0,1)$, and $\hat{\sigma}_s^2$ is the empirical variance of V_s [17].

IV. Simulation and Results

The simulations are implemented on the IEEE PES distribution networks for IEEE 8-bus and 123-bus networks [14]. In order to validate the performance of the proposed approach on loopy networks, we add several branches in both networks to create loops. The loopy 8-bus system is shown in Fig. 2. In each network, Bus 1 is selected as the slack bus. The historical data have been preprocessed by the MATLAB Power System Simulation Package (MATPOWER) [21]. To simulate the power system behavior in a more practical pattern, we use the load profile from Pacific Gas and Electric Company (PG&E) in the subsequent simulation. This profile contains anonymized and secure hourly smart meter readings over $110,000$ PG&E residential customers for a period of one year spanning from 2011 to 2012. The reactive power q_s^t at bus s and time t is computed according to a randomly generated power factor pf_s^t, which follows a uniform distribution, e.g. $pf_s^t \sim \text{Unif}(0.85, 0.95)$. To obtain voltage magnitude at time t, i.e. $|v_s^t|$, we run a power flow to generate the states of the power system. To obtain time-series data, we run the power flow to generate hourly data over a year.

A. Error Rate

To summarize performances in various simulation cases, we define the error rate (ER) as

$$\text{ER} = \left(\frac{\sum_{e_{ij} \in \mathcal{E}} \mathbb{I}\left(e_{ij} \notin \hat{\mathcal{E}}\right) + \sum_{e_{ij} \in \hat{\mathcal{E}}} \mathbb{I}\left(e_{ij} \notin \mathcal{E}\right)}{|\mathcal{E}|} \right) \times 100\%$$

where $\hat{\mathcal{E}}$ denotes the estimated set of branches and $|\mathcal{E}|$ denotes the size of the set $\mathcal{E}$. The first term of ER counts the number of missing edges. The second term captures the number of falsely connected edges. Table I summarizes ER on 8-bus and 123-bus systems with $\alpha = 0.05$.

TABLE I. Error Rate, $\alpha = 0.05$

Rule	8-Bus	8-Bus with Loops	123-Bus	123-Bus with Loops
AND	0%	10%	3.28%	3.21%
OR	14.29%	0%	1.64%	1.61%
AND-OR	0%	0%	0%	0%

In Table I, the AND-OR rule can perfectly recover the topologies in all four networks. For the OR rule, it incorrectly detects one edge in 8-bus networks. The AND rule makes no mistake in 8-bus radial network but one error in 8-bus loopy network. For 123-bus networks, the OR rule outperforms the AND rule. But both have an error rate with than 5%, which is the false alarm rate.

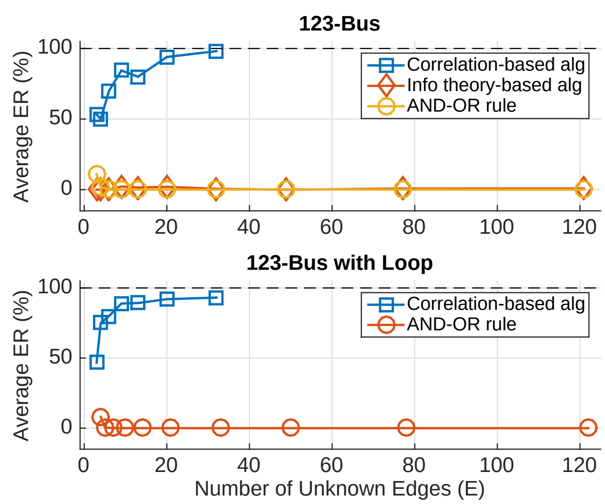

Fig. 5. ER comparison of the proposed algorithm and other existing methods in IEEE 123-bus system, with and without loop.

In Fig. 5, we compare the proposed algorithm with a correlation-based algorithm [13] and an information theory-based algorithm [16]. The optimization problem in [13] is solved by CVX, a package for specifying and solving convex programs [22]. The x-coordinate represents the number of edges that are needed to be identified. The y-coordinate represents the error rate. Our algorithm consistently has an approximate zero error rate, which is comparable with the approach in [16] in radial network, while the detection ability of [13] decreases. The information theory-based algorithm is not included in the loopy network because it is proposed for radial networks.

Fig. 6 compares the computational time of our approach with the correlation-based algorithm in loopy networks. The

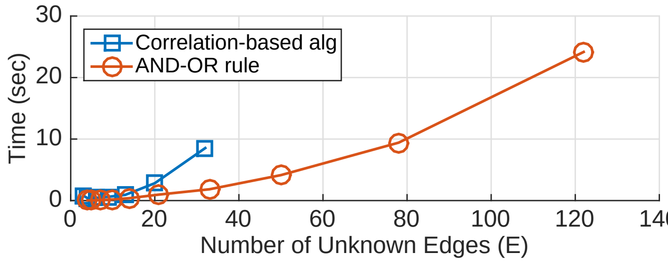

Fig. 6. Comparison of computational time of the proposed algorithm and other existing methods in IEEE 123-bus system with loops.

time is averaged over 20 iterations. All computations were done on a personal computer with Intel(R) Core(TM) i7 3615QM CPU @ 2.30GHz. Although our method is not implemented in parallel, it is consistently faster than [13]. For 8-bus network with loops in Fig. 2, our approach requires 0.3 seconds. Although [16] mainly focuses on the radial topology, it can be extended to find one loop by running the tree-based algorithm for each possible 3-tuple of buses, which is $8 \times 7 \times 6 = 336$ combinations. With 0.1 seconds per run, the time to reconstruct a 8-bus system with one loop is about 33.6 seconds, which is much slower than our method. Even worse, the system in Fig. 2 has three loops and the required time of [16] will grow up exponentially. Therefore, our approach is the most efficient one in meshed networks.

B. Integration of DERs

Every bus in 8-bus networks and 12 buses randomly chosen from 123-bus networks are integrated with solar panels. The hourly power generation profile is computed by PVWatts Calculator, an online application developed by the National Renewable Energy Laboratory (NREL) that estimates the power generation of a photovoltaic system based on weather and physical parameters [15]. The data are computed based on the weather history in North California and the physical parameters of a 5kW solar panel. The renewable power generator is modeled as a negative load.

TABLE II. Error Rate with DER, $\alpha = 0.05$

Rule	8-Bus	8-Bus with Loops	123-Bus	123-Bus with Loops
AND	14.29%	0%	1.64%	1.59%
OR	14.29%	0%	0%	0%
AND-OR	0%	0%	0%	0%

Table II illustrates the error rates with the integration of DERs. The AND-OR rule still has the lowest error rate and perfectly recovers the topologies in all three networks. In 8-bus network with loops, both the AND and OR rules have zero error as well. Unlike the case without DERs, both AND rule and OR rule have better performances in 123-bus networks.

V. Conclusion

In this paper, we propose a data-driven algorithm that reconstructs the topology of secondary distribution grid. Unlike existing approaches, our method does not require the knowledge about the switch connectivity maps or admittance matrices and only utilizes the newly available smart meter data. Also, many past studies only focus on the radial topology but our method can reconstruct a meshed network. We prove that a distribution grid can be formulated as a probabilistic graphical model and be reconstructed by regularized linear regression. We verify the proposed algorithm on both IEEE 8- and 123-bus systems with and without loops. Finally, our algorithm can perfectly reconstruct the topology in a short time, with and without the integration of DERs.

References

[1] S. Dey, A. Jessa, and L. Gelbien, "Urban grid monitoring renewables integration," in *IEEE Conference on Innovative Technologies for an Efficient and Reliable Electricity Supply*, 2010, pp. 252–256.

[2] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid," *Power Systems, IEEE Transactions on*, vol. 25, no. 1, pp. 371–380, 2010.

[3] J. Huang, V. Gupta, and Y.-F. Huang, "Electric grid state estimators for distribution systems with microgrids," in *IEEE 46th Annul Conference on Information Sciences and Systems*, 2012, pp. 1–6.

[4] R. Lugtu, D. Hackett, K. Liu, and D. Might, "Power system state estimation: Detection of topological errors," *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-99, no. 6, pp. 2406–2412, Nov 1980.

[5] C. Rudin, D. Waltz, R. Anderson, A. Boulanger, A. Salleb-Aouissi, M. Chow, H. Dutta, P. Gross, B. Huang, S. Ierome, D. Isaac, A. Kressner, R. Passonneau, A. Radeva, and L. Wu, "Machine learning for the new york city power grid," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 2, pp. 328–345, Feb 2012.

[6] R. E. Brown, "Impact of smart grid on distribution system design," in *IEEE PES General Meeting: Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, pp. 1–4.

[7] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning in power distribution networks," *arXiv preprint arXiv:1501.04131*, 2015.

[8] Y. Sharon, A. M. Annaswamy, A. L. Motto, and A. Chakraborty, "Topology identification in distribution network with limited measurements," in *IEEE PES General Meeting: Innovative Smart Grid Technologies*, 2012, pp. 1–6.

[9] M. Baran, J. Jung, and T. McDermott, "Topology error identification using branch current state estimation for distribution systems," in *IEEE Transmission & Distribution Conference & Exposition: Asia and Pacific*, 2009, pp. 1–4.

[10] R. Arghandeh, M. Gahr, A. von Meier, G. Cavraro, M. Ruh, and G. Andersson, "Topology detection in microgrids with micro-synchrophasors," *arXiv preprint arXiv:1502.06938*, 2015.

[11] Y. Weng, R. Negi, and M. Ilic, "Historical data-driven state estimation for electric power systems," in *IEEE International Conference on Smart Grid Communications*, Oct 2013, pp. 97–102.

[12] Y. Weng and R. Rajagopal, "Probabilistic baseline estimation via gaussian process," in *IEEE PES General Meeting*, July 2015, pp. 1–5.

[13] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," in *IEEE 52nd Annual Conference on Decision and Control*, 2013, pp. 1659–1664.

[14] W. H. Kersting, "Radial distribution test feeders," in *IEEE Power Engineering Society Winter Meeting*, vol. 2, 2001, pp. 908–912.

[15] A. P. Dobos, "Pvwatts version 5 manual," *National Renewable Energy Laboratory, September*, 2014.

[16] Y. Liao, Y. Weng, M. Wu, and R. Rajagopal, "Distribution grid topology reconstruction: An information theoretic approach," in *IEEE 47th North American Power Symposium*, Oct 2015.

[17] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *The Annals of Statistics*, pp. 1436–1462, 2006.

[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[19] C. Coffrin and P. Van Hentenryck, "A linear-programming approximation of ac power flows," *INFORMS Journal on Computing*, vol. 26, no. 4, pp. 718–734, 2014.

[20] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani *et al.*, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[21] R. D. Zimmerman and C. E. Murillo-Sanchez, "Matpower, a matlab power system simulation package," *http://www.pserc.cornell.edu/ matpower/manual.pdf*, 2010.

[22] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.