Creating Realistic Power Distribution Networks using Interdependent Road Infrastructure

Rounak Meyur, Madhav Marathe, Anil Vullikanti,
Henning Mortveit, Samarth Swarup
Biocomplexity Institute, University of Virginia
Charlottesville, Virginia
{*rm5nz, marathe, vsakumar, henning.mortveit, swarup*}*@virginia.edu*

Virgilio Centeno, Arun Phadke
ECE Department, Virginia Tech
Blacksburg, Virginia
{*virgilio, aphadke*}*@vt.edu*

Abstract—It is well known that physical interdependencies exist between networked civil infrastructures such as transportation and power system networks. In order to analyze complex non-linear correlations between such networks, datasets pertaining to such real infrastructures are required. However, such data are not readily available due to their proprietary nature. This work proposes a methodology to generate realistic synthetic power distribution networks for a given geographical region. A network generated in this manner is not the actual distribution system, but its functionality is very similar to the real distribution network. The synthetic network connects high voltage substations to individual residential consumers through primary and secondary distribution networks. Here, the distribution network is generated by solving an optimization problem which minimizes the overall length of the network subject to structural and power flow constraints. This work also incorporates identification of long high voltage feeders originating from substations and connecting remotely situated customers in rural geographic locations while maintaining voltage regulation within acceptable limits. The proposed methodology is applied to the state of Virginia and creates synthetic distribution networks which are validated by comparing them to actual power distribution networks at the same location.

Index Terms—synthetic distribution networks, radial networks, Mixed Integer Linear Programming

I. Introduction

Networked infrastructures are vital components of society and instrumental for its proper functioning [1]. A common structural attribute of such networked infrastructures is that they are located close to human settlements. Examples of some of these networked infrastructures include power system network, communication network, transportation network etc. Furthermore, the interdependencies among these networked infrastructures are well known [2].

The US government has termed these networks as critical infrastructures since they provide *enabling functions* across other infrastructure sectors [3]. A failure in one network can lead to cascading events in other dependent networked infrastructures and eventually cause an immense impact on the nation's economy. In recent times, there has been an increased interest in studying the resilience and reliability of such networks during extreme events. The complex non-linear correlations between different networks have drawn the attention of many researchers [2]. However, a fundamental barrier to progress in this area is the lack of data pertaining to such networked infrastructures.

On another front, emerging technologies and policies are being proposed to make these networks resilient to failures. For example, in the case of a power distribution system, the role of distributed energy resources (DERs) is being widely acknowledged as an efficient means to enhance the network resiliency. The focus of power systems research has moved towards efficient deployment and control of DERs in the distribution system [4]–[8]. However, there remains a lack of realistic test case scenarios to implement the proposed methodologies for their validation. Most of the algorithms and technologies are validated against standard IEEE test feeders or small-sized distribution networks (e.g. 13, 37, 123 bus feeders etc. [9]). The unavailability of appropriate network data is primarily due to the proprietary nature of the data as well as its enormous volume [10].

Problem The abstract problem we study can be formally stated as: *Given a set of residences and electric substations with their respective geographical coordinates, construct a power distribution network connecting these points that resembles a real operating distribution system network.* This work proposes a methodology to generate realistic synthetic distribution networks for a geographical region using open source information. The generated network connects substations to individual residential buildings through medium voltage (MV) and low-voltage (LV) networks. The synthetic networks are not exact replicas of the actual distribution network; rather they resemble the actual network in terms of structural properties. At the same time, the generated network satisfies the operational constraints (voltage and power flows limits) of an actual operating distribution system.

Contribution Our main contributions are: (i) We develop a *first principles*-based methodology to create a realistic synthetic distribution network using information from other infrastructures such as transportation networks, residential data, etc.; (ii) Our approach results in an optimal network by minimizing the overall length of distribution lines which is a principal consideration of power companies while planning distribution networks; (iii) Our method generates a distribution network which is particular to the geographical location of interest and hence provides a realistic representation of the

Table I: Major contributions of proposed work

Aspect	Previous Works	Present work
Network type	Synthetic transmission networks with generators and aggregated loads [11]–[15]	Synthetic distribution networks are created which connect high voltage substations to individual residential consumers.
Realism of generated networks	Statistical distribution of network attributes are used to generate synthetic power grids [16]	Realistic distribution networks comprising of primary and secondary levels are generated for a given geographical location. The generated network resembles an optimal network designed by power distribution companies. It follows the usual structural and power flow constraints of a typical distribution system.
	Stochastic geometry based approach to place transformers in distribution networks [17]	
	Heuristic approach to synthesize distribution networks from substations and populate them with consumer loads [13]	
Radiality of network	Radiality is ensured by avoiding isolated cycles or considering single commodity flow model [8], [18]–[20].	Power balance constraint is proved to be a sufficient condition to ensure radiality of the generated LV network.
Network attributes	Small sized networks such as standard IEEE test systems are used [4], [5].	An optimal radial network is identified for an unknown number of root nodes (high voltage feeders) and for large-sized networks with more than 20000 nodes.
	The number of root nodes (substations) are known beforehand in the problem definition.	

actual network; (iv) The nodes and edges of the generated network are labeled with all necessary attributes required for power flow analysis and therefore, can be used as suitable networks to test distribution system planning and operation algorithms.

Our approach combines real world data such as building information and road network data with general power engineering practices to construct realistic distribution systems. Instead of considering aggregated loads at zipcode centers, our method generates a detailed distribution network for a given geographical area. The economic aspect of minimizing the total length of the network (which is indicative of the investment required for building a network) has been considered as the primary objective. We include constraints motivated by power engineering practices such as maintaining a tree structure (for protection coordination), placing local transformers (pole-top and pad-mounted) along the road network and avoiding branching in the secondary network (to reduce voltage sag at the leaf nodes). We also use detailed synthetic load demand profiles from a recent model in the literature [21]. The inclusion of all these aspects has allowed us to validate power flow characteristics for the created synthetic networks.

The geographical attributes of the region are retained intact in our methodology. For example, in a rural region with multiple remote localities, a single substation feeder serving all of them would often suffer from voltage sag issues at the extreme ends. In this paper, we consider several substation feeders serving such remote localities. In this way, we ensure that the node voltages and edge flows are within acceptable limits while the radial structure of the network is maintained. The key differences of our method with other related previous methods to generate synthetic distribution networks are listed in Table I.

II. Preliminaries

A. Distribution system

The distribution network consisting of overhead power lines, underground cables, pole top transformers is responsible to

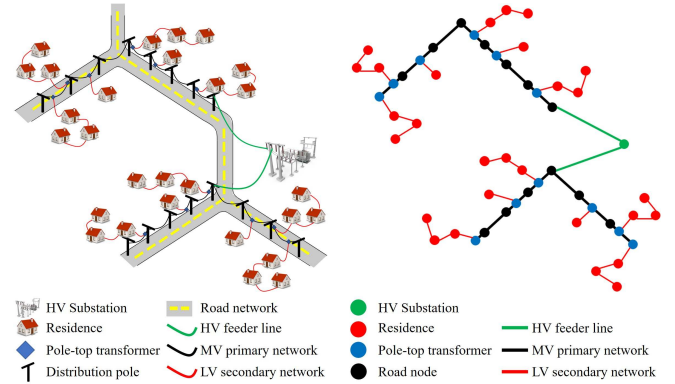|			
HV Substation	Road network	HV Substation	HV feeder line
Residence	HV feeder line	Residence	MV primary network
Pole-top transformer	MV primary network	Pole-top transformer	Road node
Distribution pole	LV secondary network	Road node	LV secondary network

Figure 1: A schematic of synthetic distribution network. The substation feeds the distribution network through high voltage feeder lines. The primary network originates from the feeder lines and follows the road network as much as possible. The secondary network originates from transformers along primary network and connects individual residences. The single line diagram for the same network with its individual elements is shown on the right figure.

bring electrical power from high voltage (HV; greater than 33kV) transmission system to the end residential consumers requiring a LV level (208-480V). This is normally done in a two-step procedure: (i) the high transmission level voltage is stepped down to MV level at distribution substations and distributed to local transformers (pole-top/pad-mounted) through *primary distribution network*, (ii) the voltage is further stepped down to LV at the local distribution transformers and distributed to individual customers through *secondary distribution network*.

Distribution systems (primary and secondary) are usually configured in a *radial* structure where power flow is unidirectional from source to consumers. Such radial structure facilitates protection coordination among reclosures, breakers and downstream fuses in the distribution system feeders [6].

Fig. 1 shows a schematic of the structure of synthetic distribution network. The network connects substation to individual residential buildings through primary and secondary networks. The substation first feeds the primary network through high voltage feeder lines. It is assumed that the primary network follows the available road network to the maximum extent since distribution poles are generally placed along the road network to facilitate maintenance. The secondary network originates from the pole top transformers which are placed along primary network. The secondary network connects individual residences through short chains. Branching in these chains is avoided in order to reduce voltage drop at the leaf nodes.

B. Available datasets

We use open source publicly available information regarding several infrastructures to generate the synthetic distribution networks. These data pertain to following sources: (i) transportation network data published by [22], (ii) Geographical location of HV substations from data sets published by [23] and (iii) Residential electric power demand information developed in the models by [21].

Table II: Datasets and related attributes used to generate synthetic distribution network

Dataset	Source	Attributes
Substation	Electric substation data published by US Department of Homeland Security [23]	• substation ID • longitude • latitude
Road network	GIS and electronic navigable maps published by NAVTEQ [22]	• node ID • node longitude • node latitude • link ID • link importance level
Residences	Synthetic population and electric load demand profiles generated by [21]	• residence ID • longitude • latitude • average hourly load demand

Roads The road network represented in the form of a graph $\mathcal{R} = (\mathcal{V}_R, \mathcal{L}_R)$, where $\mathcal{V}_R$ and $\mathcal{L}_R$ are respectively the sets of nodes and links of the network. Each road link $l \in \mathcal{L}_R$ is represented as an unordered pair of terminal nodes (u, v) with $u, v \in \mathcal{V}_R$. Each road node has a spatial embedding in form of longitude and latitude.
Substations The set of M substations $\mathcal{S} = \{s_1, s_2, \cdots, s_M\}$ with their respective geographical location data.
Residences The set of N residential buildings with geographical coordinates $\mathcal{H} = \{h_1, h_2, \cdots, h_N\}$ and hourly load profile.
Formal Problem Statement. Given a set of residence and electric substation locations, construct a power distribution network connecting these points such that (i) the network

graph is a forest of trees originating from substations, (ii) covers all residential buildings, (iii) the edge flows and node voltages are within acceptable limits and (iv) overall length of the network is minimized.

III. PROPOSED APPROACH

The problem of creating a synthetic distribution network can be considered to be a combinatorial optimization problem, where we identify the optimal edges from a possible edge set connecting residences to substations. Due to the large number of points to be connected, the combinatorial problem is computationally challenging. As discussed in Section II, a typical distribution system consists of primary and secondary networks. Therefore, the synthesis of such networks is considered to be a two-step bottom-up procedure: (i) the first step constructs the secondary distribution network connecting the residential buildings to pole-top/pad-mounted transformers and (ii) the second step involves connecting these local transformers to distribution substations through feeders and laterals. In our work, the road network is used as a proxy for the primary distribution network. Therefore, the locations of local pole-top transformers are considered to be internal points on the road network links.

The overall algorithm to generate synthetic distribution networks is outlined through the following points: (i) Evaluate a mapping between residential buildings and road network links such that each residence is mapped to the nearest road link (III-A). Thereafter, we identify probable locations of local distribution transformers along these road network links. (ii) Connect the local distribution transformers to mapped residences in a radial configuration to resemble a typical secondary distribution network (III-B). (iii) Connect the local transformers to distribution substation in a forest of trees with the road network as a proxy (III-C).

A. Mapping residences to road links

This section details the methodology to identify subsets of residential buildings near each road network link. This information would be used in the succeeding step to generate the secondary distribution network.

Algorithm 1 lists the steps to compute the nearest road network link to a given point. Fig 2 provides a visual representation of the same. First, we draw bounding boxes around each road link and residences. The intersections between the bounding box of a residence with those of the links are stored and indexed in a *quad-tree* data structure. This information is retrieved to identify the links (green links in Fig.2) which are comparably nearer to the residential building than the others. Now, the geodesic distance of the residence point can be computed from these short-listed links and the nearest link can be identified henceforth. This approach reduces the computational burden of evaluating the distance between all road links and residential buildings.

Following this step, we can compute an inverse map which represents the nearby residences mapped to a given road link. It is assumed that the residences would be fed by local pole
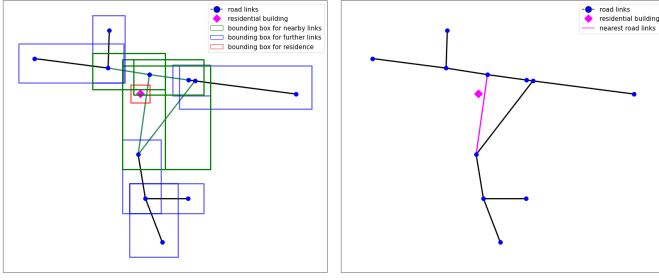
Figure 2: Steps showing procedure of Algorithm 1. Bounding boxes are drawn around each road link. The nearby links to a given residence point are short-listed by identifying the boxes (green) which intersect with the one around the residence point (red). The nearest link is the closest among these shortlisted links.

Algorithm 1 Find the nearest link in $\mathcal{L}_R$ to a given point **p**.

Require: Dimensions of bounding box.
1: Draw bounding box $\mathbf{B_l}$ around each link $\mathbf{l} \in \mathcal{L}_R$.
2: Draw bounding box $\mathbf{B_p}$ around point **p**.
3: Find the bounding boxes $\mathbf{B_{l_1}}, \mathbf{B_{l_2}}, \cdots, \mathbf{B_{l_k}}$ corresponding to the links $\mathbf{l_1}, \mathbf{l_2}, \cdots, \mathbf{l_k}$ which intersect with $\mathbf{B_p}$.
4: Find the link $\mathbf{l}^\star$ among k short-listed links, which is nearest to point **p**.

---

top transformers located along the mapped link. However, their locations are not known beforehand. Therefore, we consider their probable locations along the link by interpolating multiple points at equal separation along it (based on standard engineering practice [24]).

B. Creation of Secondary Network

The goal of this step is to create the secondary distribution network connecting local transformers along a given road network link l to the set of $\mathcal{V}_H$ residences mapped to it. The set of probable local transformers interpolated along the link is $\mathcal{V}_T$. We need to connect the set of residences $\mathcal{V}_H$ to actual transformer nodes $\mathcal{V}_T^\star \subseteq \mathcal{V}_T$ in a forest of *starlike* trees with each tree rooted at a transformer node.

We start with a candidate set $\mathcal{E}_D$ of edges between the residence and probable transformer nodes. The secondary network edges are to be selected from this candidate set. We construct an undirected graph $\mathcal{G}_D := (\mathcal{V}, \mathcal{E}_D)$ with node set $\mathcal{V} = \mathcal{V}_T \bigcup \mathcal{V}_H$ and edge set $\mathcal{E}_D$. The aim of this step can be formalized through the following problem statement.

Problem III.1 (Secondary network creation problem). *Given the undirected graph $\mathcal{G}_D(\mathcal{V}, \mathcal{E}_D)$, find $\mathcal{E}_D^\star \subseteq \mathcal{E}_D$ such that the induced subgraph network $\mathcal{G}_D^\star(\mathcal{V}^\star, \mathcal{E}_D^\star)$ with $\mathcal{V}^\star = \mathcal{V}_T^\star \bigcup \mathcal{V}_H$ is a forest of starlike trees, $\mathcal{V}_T^\star \subseteq \mathcal{V}_T$ is the set of root nodes and the overall length of network is minimized.*

Note that a complete graph composed of the residence and transformer nodes can always be considered as the candidate set $\mathcal{E}_D$. In this paper, a Delaunay triangulation [25] of the residential nodes is considered to reduce size of the problem.

Edge weight assignment An edge $e : (i, j) \in \mathcal{E}_D$ is assigned a weight $w(i, j)$

$$w(i, j) = \begin{cases} \infty, & \text{if } i, j \in \mathcal{V}_T \\ \text{dist}(i, j) + \lambda \mathsf{C}(i, j), & \text{otherwise} \end{cases} \quad (1)$$

where dist $: \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ denotes the geodesic distance between the nodes i, j. The function $\mathsf{C}(u, v)$ penalizes the cost function if the edge connecting nodes u, v crosses the road link and is defined as

$$\mathsf{C}(u, v) = \begin{cases} 0, & \text{if } u, v \text{ are on same side of link} \\ 2, & \text{if } u, v \text{ are on opposite side of link} \\ 1, & \text{if } u \in \mathcal{V}_R \text{ or } v \in \mathcal{V}_R \end{cases}$$

λ is a weight factor to penalize multiple crossing of edges over the road links. It also penalizes multiple edges emerging from the root node. The weights are stacked in $|\mathcal{E}_D|$-length vector **w**. Note that an edge between two probable transformer nodes is assigned a weight of *infinity* which is equivalent of not considering them as candidate edges.

Edge variables We introduce binary variables $\{x_e\}_{e \in \mathcal{E}_D} \in \{0, 1\}^{|\mathcal{E}_D|}$. Variable $x_e = 1$ indicates that the edge is present in the optimal topology and vice versa. Each edge $e := (i, j)$ is assigned a flow variable f_e (arbitrarily) directed from node i to node j. The binary variable and flows can be stacked in $|\mathcal{E}_D|$-length vectors **x** and **f** respectively.

Node variables The average hourly load demand at the ith residence node is denoted by p_i and is strictly positive. We stack these average hourly load demands at all residence nodes in a $|\mathcal{V}_H|$ length vector **p**.

Degree constraint. Statistical surveys on distribution networks in [10], [26] show that residences along the secondary network are mostly connected in series with at most 2 neighbors. This is ensured by (2) which limits degree of residence nodes to 2.

$$\sum_{e:(h,j)} x_e \leq 2, \quad \forall h \in \mathcal{V}_H \quad (2)$$

Power flow constraints. For the connected graph $\mathcal{G}_D(\mathcal{V}, \mathcal{E}_D)$, we define the $|\mathcal{E}_D| \times |\mathcal{V}|$ branch-bus incidence matrix $\mathbf{A}_{\mathcal{G}_D}$ with the entries as

$$\mathsf{A}_{\mathcal{G}_D}(e, k) := \begin{cases} 1, & k = i \\ -1, & k = j \\ 0, & \text{otherwise} \end{cases} \quad \forall e = (i, j) \in \mathcal{E}_D \quad (3)$$

Since the order of rows and columns in $\mathbf{A}_{\mathcal{G}_D}$ is arbitrary, we can partition the columns as $\mathbf{A}_{\mathcal{G}_D} = \begin{bmatrix} \mathbf{A_T} & \mathbf{A_H} \end{bmatrix}$, without loss of generality, where the partitions are the columns corresponding to transformer and residence nodes respectively. We define $\mathbf{A_H}$ as the reduced branch-bus incidence matrix.

Assuming no network losses, (4a) represents the power balance equations at all residence nodes. Note that the optimal network is obtained from $\mathcal{G}_D$ after removing the edges for which $x_e = 0$. Therefore, we need to enforce zero flows f_e for non-existing edges. The constraint (4b) performs this task

along with constraining the flows f_e for existing edges to be within pre-specified capacities $\overline{f}$.

$$\mathbf{A_H}^T\mathbf{f} = \mathbf{p} \tag{4a}$$
$$-\overline{f}\mathbf{x} \leq \mathbf{f} \leq \overline{f}\mathbf{x} \tag{4b}$$

Ensuring radial topology The radiality requirement of $\mathcal{G}_D^\star$ can be enforced from a known graph theory property: *a forest with n nodes and m root nodes has $n-m$ edges.* In our case, $|\mathcal{V}_H| + |\mathcal{V}_T^\star|$ nodes need to be covered in a forest of trees with $|\mathcal{V}_T^\star|$ root nodes which leads us to the following constraint.

$$\sum_{e \in \mathcal{E}_D} x_e = |\mathcal{V}_H| \tag{5}$$

However, we need to ensure that there are no disconnected cycles in the optimal network. This can be done by ensuring that the residence points are connected to a transformer node [4], [19]. In our case, this condition is satisfied by the node power flow condition in (4a) if all the residential nodes consume power. This is an extension of the proposition in [4] which considers renewable generation at the nodes. However, contrary to the previous work, the following proposition is a special case where every non-root node has only positive load demand. Using the power balance constraints to ensure radial topology reduces the number of constraints when dealing with large sized networks.

The optimal secondary network graph $\mathcal{G}_D^\star(\mathcal{V}^\star, \mathcal{E}_D^\star)$ with $\mathcal{E}_D^\star = \mathcal{E}_D \setminus \{e : e \in \mathcal{E}_D, x_e = 0\}$ and $\mathcal{V}^\star = \mathcal{V}_T^\star \bigcup \mathcal{V}_H$ has at least $|\mathcal{V}_T^\star|$ components since local transformer nodes can only be root nodes and there is no candidate edge between them.

Proposition III.1. *The graph $\mathcal{G}_D^\star(\mathcal{V}^\star, \mathcal{E}_D^\star)$ with reduced branch-bus incidence matrix $\mathbf{A_H}$ (corresponding to columns of $\mathcal{V}_H$) and node power demand vector $\mathbf{p} \in \mathbb{R}^{\mathcal{V}_H}$, with strictly positive entries, has exactly $|\mathcal{V}_T^\star|$ connected components if and only if there exists $\mathbf{f} \in \mathbb{R}^{|\mathcal{E}_D^\star|}$, such that (4a) is satisfied.*

Proof. The proof follows along similar lines to proof of Proposition 1 in [4]. Proving by contradiction, suppose $\mathcal{G}_D^\star(\mathcal{V}^\star, \mathcal{E}_D^\star)$ has more than $|\mathcal{V}_T^\star|$ connected components and there exists $\mathbf{f} \in \mathbb{R}^{|\mathcal{E}_D^\star|}$ satisfying the proposed equality. Therefore, there exists a connected component $\mathcal{G}_S(\mathcal{V}_S, \mathcal{E}_S)$ which is a maximal connected subgraph with $\mathcal{V}_S \subset \mathcal{V}_H$ and $\mathcal{V}_S \bigcap \mathcal{V}_T^\star = \emptyset$. Let $\mathbf{A}_S$ denote the bus incidence matrix of $\mathcal{G}_S$. By definition, it holds that $\mathbf{A}_S\mathbf{1} = \mathbf{0}$.

Since graph $\mathcal{G}_S(\mathcal{V}_S, \mathcal{E}_S)$ is a maximal connected subgraph of $\mathcal{G}_D^\star$, there exists no edge (i,j) with $i \in \mathcal{V}_S$ and $j \in \mathcal{V}_{\overline{S}}$, where $\mathcal{V}_{\overline{S}} = \mathcal{V}^\star \setminus \mathcal{V}_S$. Since the order of rows and columns of $\mathbf{A_H}$ are arbitrary, we can partition without loss of generality as

$$\mathbf{A_H} = \begin{bmatrix} \mathbf{A}_{\overline{S}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_S \end{bmatrix}$$

We can partition vectors $\mathbf{f}$ and $\mathbf{p}$ conformably to $\mathbf{A_H}$ to get the following equality

$$\begin{bmatrix} \mathbf{A}_{\overline{S}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_S \end{bmatrix}^T \begin{bmatrix} \mathbf{f}_{\overline{S}} \\ \mathbf{f}_S \end{bmatrix} = \begin{bmatrix} \mathbf{p}_{\overline{S}} \\ \mathbf{p}_S \end{bmatrix}$$

where $\mathbf{p}_S, \mathbf{p}_{\overline{S}}$ are respectively the vectorized power demands for node sets $\mathcal{V}_S$ and $i \in \mathcal{V}_{\overline{S}}$. From the second block, it implies that

$$\mathbf{A}_S^T\mathbf{f}_S = \mathbf{p}_S \Rightarrow \mathbf{1}^T\mathbf{p}_S = \mathbf{1}^T\mathbf{A}_S^T\mathbf{f}_S = \mathbf{0} \tag{6}$$

Since all entries of $\mathbf{p}$ are strictly positive from the initial assumption, we have $\mathbf{1}^T\mathbf{p}_S \neq 0$ which contradicts (6) and completes the proof. $\square$

Generating optimal network topology The optimal secondary network is obtained after solving the optimization problem.

$$\min_{\mathbf{x}} \mathbf{w}^T\mathbf{x}$$
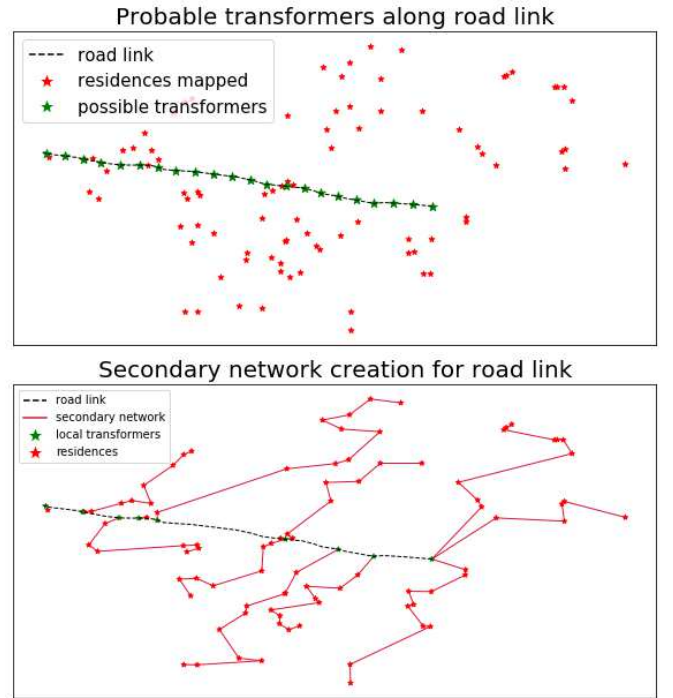$$\text{s.to. } (2), (4), (5) \tag{7}$$

Figure 3: The probable transformer locations are identified along the road network link and the secondary network is created by solving the optimization problem. The network originates from the road link and connects residences mapped to it in a forest of starlike trees rooted at the transformers.

Example. The secondary network connecting residences along a road network link is shown in Fig. 3. The first figure shows the residences (red) mapped to the road link and the probable transformers (green points) along the link. A Delaunay triangulation is considered to connect the points and obtain the set of possible edges. Thereafter, the optimization problem is solved which identifies the optimal set of edges as shown in the bottom figure. We observe that all residences are connected in star-like trees with roots as one of the transformers along the road link.

C. Creation of Primary Network

In this step, the goal is to create the primary distribution network which connects the substations to the local transformers. We aggregate the load at residences to the local transformer locations (obtained as an output from the preceding step). In this paper, we have aggregated the average demand at the residences; however, *diversity factor* may be used in the aggregation to appropriately rate the transformers [27]. The objective of the current step is to connect all these local transformer locations to the substation.
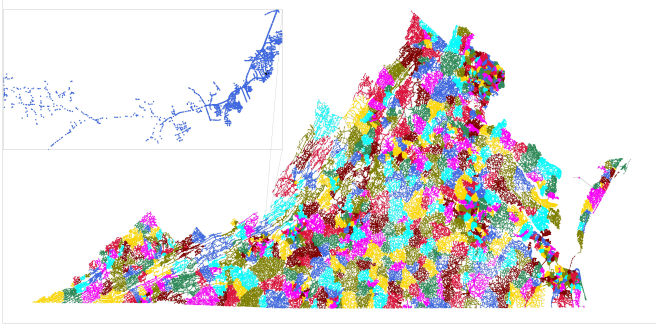
Figure 4: Voronoi partitioning of the road and transformer nodes in Virginia based on the shortest path distance (along road network) to the substations.

Partitioning the problem. As mentioned earlier in Section II, we use the road network as a proxy for the synthetic primary network. The problem of primary network generation is to select the optimal edge set from the candidate road links. However, the combinatorial problem becomes challenging to solve when considering a large geographical area. Therefore, the problem of generating the primary network is partitioned into a number of sub-problems (one problem corresponding to each distribution substation). First, we include the local transformer nodes (located along road links) in the original road network graph. Thereafter, we partition the nodes in an intermediate step and use the road network subgraph composed of these partitioned nodes to construct the primary distribution network.

The nodes can be partitioned such that each of them is assigned to the geographically nearest substation. However, such partitioning gives rise to multiple disconnected components (of varied sizes) in the induced subgraph. In such a situation, the transformers cannot be connected through the road network. To this end, the transformers are clustered so that each of them is mapped to the nearest substation along the road network. This is depicted in Fig 4 where each color represents a partition of transformer nodes. These partitions are known as Voronoi cells which are centered at the substation location. The partitioning is done based on the *shortest path distance* metric [28] which ensures that each node is mapped to the nearest substation along the road network and each induced subgraph has a single connected component.

Requirements Once we obtain the *road network subgraph* $\mathcal{G}_R(\mathcal{V}, \mathcal{E}_R)$ with $\mathcal{V} = \mathcal{V}_R \bigcup \mathcal{V}_T$ for each partition, the goal of

the current step is to select edges $\mathcal{E}_R^\star \subseteq \mathcal{E}_R$ such that (i) all local transformer nodes ($i \in \mathcal{V}_T$) are connected, (ii) the road network nodes ($i \in \mathcal{V}_R$) can be covered only to connect local transformer nodes (i.e., the road nodes cannot be leaf nodes in the created primary network) and (iii) the created primary network is a forest of trees with each tree being rooted at a road node and connects local transformer nodes through the road links in a tree structure.

The last condition follows the assumption that a substation can have multiple feeder lines connecting several root nodes feeding local transformers in remote localities. The formal problem statement is provided here.

Problem III.2 (Primary network creation problem). *Given the connected network $\mathcal{G}_R(\mathcal{V}, \mathcal{E}_R)$, find $\mathcal{E}_R^\star \subseteq \mathcal{E}_R$ such that the induced subgraph network $\mathcal{G}_R^\star(\mathcal{V}^\star, \mathcal{E}_R^\star)$, with $\mathcal{V}^\star = \mathcal{V}_R^\star \bigcup \mathcal{V}_T$, is a forest of trees with each tree rooted at some $r \in \mathcal{V}_R^\star \subseteq \mathcal{V}_R$ and the overall network length is minimized.*

Node variables. The node set $\mathcal{V} = \mathcal{V}_R \bigcup \mathcal{V}_T$ comprises of road and transformer nodes respectively. The residential average hourly demands are aggregated at the transformer locations and stacked in a $|\mathcal{V}_T|$-length vector $\mathbf{p}$. Note that the power demand at road nodes is zero since they are dummy nodes with a purpose to connect transformer nodes. Let v_i represent the voltage at the node i. The node voltages can be stacked in $|\mathcal{V}_R| + |\mathcal{V}_T|$ length vectors $\mathbf{v}$.

We assign binary variables $\{y_r, z_r\}_{r \in \mathcal{V}_R} \in \{0, 1\}$. Variable $y_r = 1$ indicates that road network node r is part of the primary network and vice versa. Variable $z_r = 0$ indicates that road network node r is a root node (connected to substation through a high voltage feeder line) in the primary network and $z_r = 1$ indicates otherwise. The binary variables are stacked in $|\mathcal{V}_R|$-length vectors $\mathbf{y}$ and $\mathbf{z}$ respectively.

Edge variables In order to identify which edges comprise of the primary network, each edge e is assigned a binary variable $\{x_e\}_{e \in \mathcal{E}_R} \in \{0, 1\}^{|\mathcal{E}_R|}$. Variable $x_e = 1$ indicates that the edge is part of the optimal primary network and vice versa. The binary variable for all edges and edge power flows can be stacked in a $|\mathcal{E}_R|$-length vectors $\mathbf{x}$ and $\mathbf{f}$ respectively.

Connectivity constraint Let $e = (r, j) \in \mathcal{E}_R$ denote an edge which is incident on the road node $r \in \mathcal{V}_R$. $\sum_{e:(r,j)} x_e$ is the degree of node r in graph $\mathcal{G}_R$.

$$\sum_{e=(r,j)} x_e \leq |\mathcal{E}_R| y_r, \qquad \forall r \in \mathcal{V}_R \qquad (8a)$$

$$\sum_{e=(r,j)} x_e \geq y_r, \qquad \forall r \in \mathcal{V}_R \qquad (8b)$$

$$1 - z_r \leq y_r, \qquad \forall r \in \mathcal{V}_R \qquad (8c)$$

$$\sum_{e:(r,j)} x_e \geq 2(y_r + z_r - 1), \qquad \forall r \in \mathcal{V}_R \qquad (8d)$$

If r is not included ($y_r = 0$) in the primary network, (8a) and (8b) ensures that there are no incident edges on r since $\sum_{e=(r,j)} x_e = 0$. Further, (8c) ensures $z_r = 1$ which implies that an unselected road node is treated as a non-root node.

6

If r is included in primary network and is not a root node ($y_r = 1, z_r = 1$), it has to be a transfer node with a minimum degree of 2 which is ensured by (8d) through $\sum_{e=(r,j)} x_e \geq 2$. If r is included in primary network and is a root node ($y_r = 1, z_r = 0$), (8a), (8b) and (8d) ensures that the degree of the road node is positive.

Ensuring radiality constraint. The primary network is a forest of trees with $|\mathcal{V}_T| + \sum_{r \in \mathcal{V}_R} y_r$ nodes and $\sum_{r \in \mathcal{V}_R}(1 - z_r)$ root nodes. Therefore, from graph theory, we have the radiality constraint as

$$\sum_{e \in \mathcal{E}_R} x_e = |\mathcal{V}_T| + \sum_{r \in \mathcal{V}_R} y_r - \sum_{r \in \mathcal{V}_R} (1 - z_r) \qquad (9)$$

However, this is not a sufficient condition for radiality since it does not avoid formation of disconnected components with cycles. Therefore, we need to enforce constraints to avoid formation of cycles in individual trees.

Power balance and flow constraints We can define the branch-bus incidence matrix $\mathbf{A}_{\mathcal{G_R}} \in \mathbb{R}^{|\mathcal{E}_R| \times |\mathcal{V}|}$. Since the order of rows and columns in $\mathbf{A}_{\mathcal{G_R}}$ is arbitrary, we can partition the rows without loss of generality as $\mathbf{A}_{\mathcal{G_R}} = \begin{bmatrix} \mathbf{A_T} & \mathbf{A_R} \end{bmatrix}$. Here, the partitions $\mathbf{A_T}$ and $\mathbf{A_R}$ are obtained by stacking the columns of $\mathbf{A}_{\mathcal{G_R}}$ corresponding to transformer and road nodes respectively.

$$\mathbf{A_T}^T \mathbf{f} = \mathbf{p} \qquad (10a)$$

$$-\overline{s}(1 - \mathbf{z}) \leq \mathbf{A_R}^T \mathbf{f} \leq \overline{s}(1 - \mathbf{z}) \qquad (10b)$$

$$-\overline{f}\mathbf{x} \leq \mathbf{f} \leq \overline{f}\mathbf{x} \qquad (10c)$$

(10a) ensures that a path exists between each transformer node and a root node. If a road node is not a root node (with $z_r = 1$), (10b) enforces that the consumption at the node is 0. If a road node is a root node (with $z_r = 0$), the power consumption/injection at the node is limited by the feeder capacity $\overline{s}$. (10c) ensures that if an edge is selected ($x_e = 1$), the power flow is limited by the line capacity $\overline{f}$.

Voltage constraints. The created primary network should also satisfy the operational voltage constraints. The long HV lines from the substation to the root nodes in the optimal primary distribution network end in voltage regulators which ensure that the root nodes have a voltage of 1pu. This is ensured by (11a). (11b) limits the voltage at all nodes within the acceptable limits of $[\underline{v}, \overline{v}]$.

$$(1 - v_r) \leq z_r \ \forall r \in \mathcal{V}_{s_R} \qquad (11a)$$

$$\underline{v}\mathbf{1} \leq \mathbf{v} \leq \overline{v}\mathbf{1} \qquad (11b)$$

$$-(1 - x_e)M \leq v_i - v_j - r_e f_e \leq (1 - x_e)M, \ \forall e \in \mathcal{E}_s \qquad (11c)$$

If r_e denotes the resistance of the line $e : (i,j) \in \mathcal{E}_R$, the Linearized Distribution Flow (LDF) model relates the squared voltage magnitude to power flows linearly as $v_i^2 - v_j^2 = 2r_e f_e$ where f_e is the entry from the vector $\mathbf{f}$ corresponding to edge e. The squared voltage can be approximated as $v_i^2 \approx 2v_i - 1$ which leads to the relation $v_i - v_j = r_e P_e$ (see [29]). Notice that (11c) enforces this constraint on only those edges $e \in \mathcal{E}_R$

for which $x_e = 1$. Here, a tight bound of M can be selected as $M = \overline{v} - \underline{v}$.

Generating optimal primary network Each edge $e = (u,v) \in \mathcal{E}_R$ is assigned a weight $w_e = w(u,v) = \text{dist}(u,v)$ which is the geodesic distance between the nodes. Additionally, for every road node $r \in \mathcal{V}_R$, we compute its geodesic distance from the substation s and is denoted by d_r. The optimal primary network topology is obtained by solving the optimization problem.

$$\min_{\mathbf{x,y,z}} \sum_{e \in \mathcal{E}_s} x_e w_e + \sum_{e \in \mathcal{V}_{s_R}} (1 - z_r) d_r \qquad (12)$$
$$\text{s.to. } (8), (9), (10), (11)$$

Handling large optimization problems. Note that for the optimization problem in (10), each road node corresponds to two binary variables and each edge corresponds to one. For a Voronoi partition with a large number of road and transformer nodes, the optimization problem to create the primary network is required to be scaled for faster computation. For this purpose, we further partition the nodes based on *Girvan-Newman* hierarchical community detection algorithm [30]. We continue partitioning until all the partitions reach a specified condition (for example, have at most some specified number of nodes, have at most a specified load demand etc.). Such a partitioning leads to connected components in the graph and the optimization problem in (10) can be solved for each partition separately.

Example. We synthesize the primary distribution network originating from a substation in south-west Virginia. The induced subgraph of the road network composed of nodes in the partition consists of 2990 nodes and 3178 edges. We apply the Girvan-Newman hierarchical community detection algorithm to further partition the road network subgraph. Fig. 5 shows the partitions in the network which are of comparable sizes (less than 700 nodes in each partition). Thereafter, we solve the optimization problem (10) for each of these sub-partitions. Note that the substation connects these partitions through one or more high voltage feeder lines (blue lines). The optimal edges are selected from the road links such that all the transformer nodes (green nodes) are covered in a tree structure and the overall length of the network is minimized. Finally, the associated secondary distribution network is appended with the primary to obtain the complete synthetic distribution network.

The synthetic network is generated following power engineering operational constraints such as the node voltages and edge flows. Considering that all residential customers are consuming average hourly load, a linearized distributed power flow program is run on the network. The operating voltage at each node is recorded and shown in the left figure of Fig 6 through a colored map. The ANSI C84.1 standard mandates a $\pm 5\%$ voltage regulation for normal operating condition. It is observed that the voltage profile is within acceptable limits of 0.95 to 1.0 pu. The voltage profile at the extreme ends can be further improved by optimally placing capacitors along the network. The loading level of each line is also depicted in the right figure of Fig 6. The loading level of lines near the
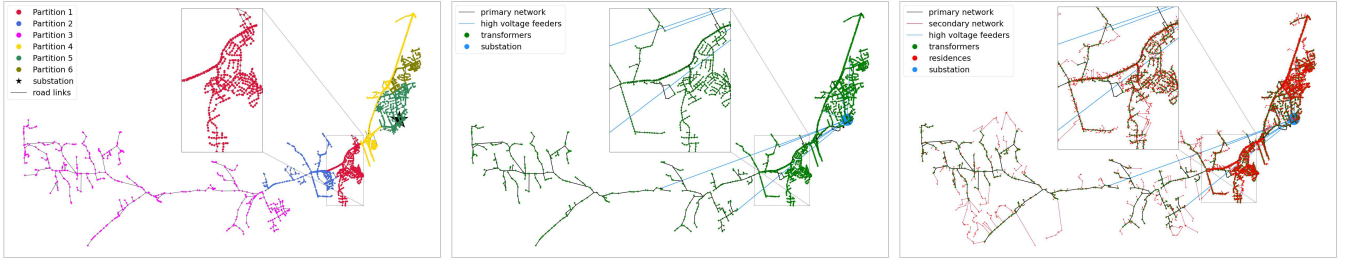
Figure 5: Synthetic primary network generation for a substation in south-west Virginia. First, the Girvan-Newman algorithm is applied to partition the large road network graph. Thereafter, the optimization problem is solved to obtain the optimal primary network. Finally the complete synthetic network is created by adding the secondary network to the primary.
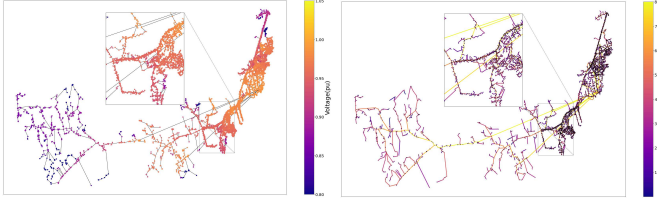

Figure 6: Node voltages and edge flows in the created synthetic network. The generated synthetic distribution network satisfies the usual power engineering operational constraints.

residences are small where as those for the feeders are high, which is expected to occur for a radial distribution network. This completes the operational validation of the network where we can conclude that when all residential customers consume average hourly load, the power distribution network operates in a *secure* state with voltages and line flows within acceptable operational limits.

Table III: Residence and road network statistics

		Mean	Median	Minimum	Maximum
Residences		24907	11454	1940	404109
Road	Nodes	6320	4090	505	71126
Network	Edges	7438	4675	610	86514

Computational scalability. We generate synthetic distribution network for the entire state of Virginia by solving two MILP optimization problems. There are 134 counties and independent cities and residence and road network information is extracted for each of them individually. The county-wise statistics of this extracted data is provided in Table III. In this section, we discuss about the factors affecting the computational speed of the generation procedure.

Table IV: Summary of time complexity

Algorithm Steps	Step 1 Mapping	Step 2 Secondary Network	Intermediate step Voronoi partitioning	Step 3 Primary Network
Worst time complexity	$\mathcal{O}(n)$	$\mathcal{O}(2^n)$	$\mathcal{O}(n \log n)$	$\mathcal{O}(2^n)$
Maximum computation time	5 mins	23 hours	3 mins	28 hours

All the experiments (county level secondary network generation, primary network generation for Voronoi partition) are conducted on a node with CentOS 7.6.1810 x8664 having 40 cores - each core running Intel Xeon Gold 6148 CPU@ 2.40 GHz. The total available memory on every node is 150GB, shared by the 40 cores on the node. Table IV summarises the worst time complexity and worst computation time for each step in the algorithm.

For the secondary network generation, we observe that the time to create the connection between the residences and local transformers increases exponentially with the number of residences to be connected. The box plot shows significant level of variation for same range of residences. This is due to the pattern in which they are distributed around the road network link. A congested distribution of residences at particular spots along a road link requires comparably more time than the cases where residences are uniformly placed along the link. For the primary network generation, the computation time also exponentially increases with the number of binary variables involved in the problem.
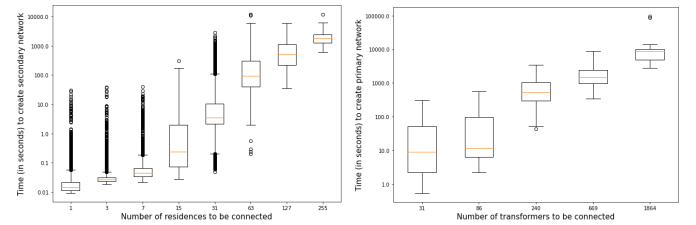

Figure 7: Box plots showing time required to create the secondary (left) and primary (right) distribution networks. The time required increases exponentially with the number of binary variables. The number of binary variables is proportional to number of residences (for secondary network) and transformer nodes (for primary network) to be connected.

IV. VALIDATION AND VERIFICATION

The primary goal of the paper is to generate realistic synthetic distribution networks which resemble real-world networks. We obtained real world power distribution networks for town of Blacksburg in south-west Virginia. However, the actual network has been incrementally built over a long period

of time. Therefore, it is expected that there would be structural differences between the two networks. In this section, we attempt to validate the proposed methodology for generating networks

Variation in network structure. Fig. 8 shows the structural comparison between the actual distribution network of a region and the synthetic network generated for the same. A zoomed in view (bottom right) of a portion of the networks shows that they are similar to each other. While the synthetic network almost retraces the road network, the actual network is adjacent to it. This validates our primary assumption that the primary distribution network follows the road network. The actual network has multiple redundancies in its structure which is present to increase the resiliency of the distribution network. Our work does not integrate such attributes while generating the synthetic networks which results in a different structure from the real network.

Feeder selection. The significant structural difference between the two networks is the substation feeder to which it is connected. The actual network is connected to the nearest geographically located substation, where as the synthetic network is connected to a distant feeder. It is to be noted that the synthetic distribution network is generated using first principles with the assumption that it follows the road network to the maximum extent (it is easier to place distribution poles along the road network). The zoomed in view on the center right of Fig 8 provides the explanation for the structural difference in the two networks. It is seen that the nearest substation (from which the actual network is fed) has no road links connecting itself to the neighborhood. On the contrary the other substation, though located at a further distance geographically, has road links connecting itself to the neighborhood under consideration. The intermediate step in our algorithm assigns nodes to substations through a shortest network distance based Voronoi partitioning. Therefore, the synthetic network generation algorithm assigns the nodes in the neighborhood to the latter substation.

Note that in the proposed algorithm, the substation feeders need not have equal sharing of load which is similar to the real world situation wherein substations have different capacities. The maximum capacity of each feeder (or feeder rating) can be pre-defined or randomly selected and used as a constraint in the network generation algorithm.

Variation in node voltage profile and edge flows. From the above discussion, it is evident that though the actual and synthetic network are similar for most part, the feeders connecting the network to the substation are often different which leads to the observed differences. We now compare the voltages at the residences when they are connected to the actual and synthetic network in Fig. 9. We call this *operational validity*: The basic idea is that if we substitute the actual network with the synthetic network, we should see minimal node voltage differences between the two. Here, the green lines signify $\pm 1\%$ deviation from the voltages if the residences had been connected to the actual network. We observe that the residence voltages in the synthetic network remain within
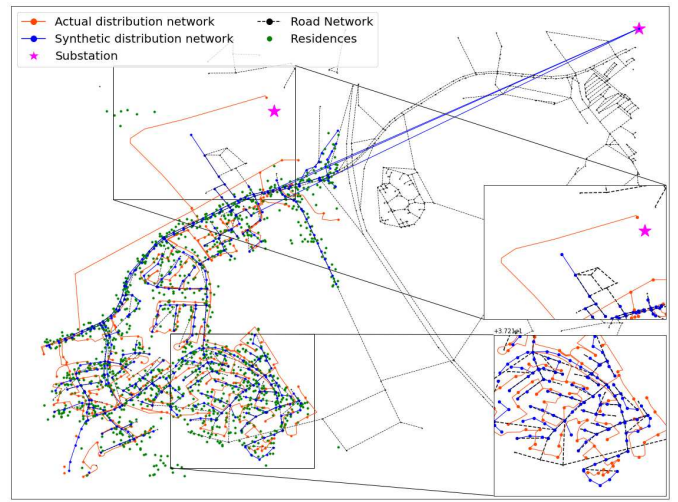

Figure 8: Comparing the structure of actual distribution network (red) with the synthetic distribution network (blue) generated using first principles from the knowledge of road network (black). Inset figure on bottom right shows structural similarities between the pair of networks. The inset figure on center right shows the absence of road links near the geographically nearest substation which feeds the actual network. Our synthetic network is fed from the substation which is closest along the road network but is geographically further.
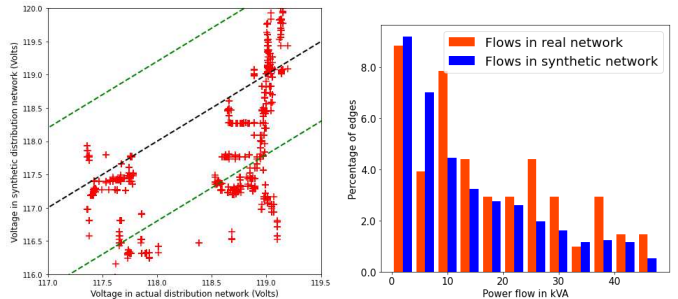

Figure 9: Comparing the residential node voltages and edge power flows for the two networks. The voltages in the synthetic network are within $\pm 1\%$ voltage regulation of the voltages in the actual network. The observable differences in the edge flow distribution of the two networks are primarily due to their structural dissimilarities.

this $\pm 1\%$ regulation, which validates our generated synthetic network. We also compare the edge flows in the two networks through the histogram in Fig. 9. Since the networks are radial in structure, the edge flows is proportional to the number of children nodes. The structural differences in the two networks explain the observed deviation in the edge flows of the networks. Note that our methodology aims to provides a generic framework to create multiple synthetic distribution networks which resemble actual networks but do not exactly mimic them. One can fine-tune the parameters/constraints in the optimization problem to obtain a different structure.

V. Conclusion

We present a methodology to generate synthetic distribution networks for a given geographic location based on the available road network data, which maintains a radial structure and minimizes the total length. We have generated synthetic distribution network for the state of Virginia and proposed methodologies to scale our optimization algorithms for regions of large size. Such realistic synthetic networks are useful in a number of applications when studying power systems engineering. Finally we provided a structural and operational validation of the generated networks by comparing them with real distribution networks.

The consideration of economic, engineering, behavioral and geographical aspects for inferring the distribution network has allowed us to choose certain parameters (e.g., feeder capacities, maximum number of feeders etc.)—these parameters can be pre-defined by user's requirement or can be randomly selected to create ensembles of networks, which can be a direction for further research. Another important aspect of real distribution networks is the existence of multiple redundancies to increase the resiliency of the distribution system, which is a future research direction.

References

[1] D. A. Reed, K. C. Kapur, and R. D. Christie, "Methodology for Assessing the Resilience of Networked Infrastructure," *IEEE Systems Journal*, vol. 3, no. 2, pp. 174–180, June 2009.

[2] A. Adiga, C. J. Kuhlman, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, R. E. Stearns, and A. Vullikanti, "Learning Coalition-Based Interactions in Networked Social Systems," in *Association for the Advancement of Artificial Intelligence Conference 2020*, Apr 2020.

[3] Presidential Policy Directive, "Critical Infrastructure Security and Resilience (PPD-21)," Feb 2013.

[4] M. K. Singh, V. Kekatos, S. Taheri, K. P. Schneider, and C.-C. Liu, "Enforcing radiality constraints for der-aided power distribution grid reconfiguration," 2019.

[5] M. K. Singh, V. Kekatos, and C.-C. Liu, "Optimal Distribution System Restoration with Microgrids and Distributed Generators," in *2019 IEEE PES General Meeting*, 2019.

[6] A. Zamani, T. Sidhu, and A. Yazdani, "A strategy for protection coordination in radial distribution networks with distributed generators," in *IEEE PES General Meeting*, July 2010, pp. 1–8.

[7] P. P. Barker and R. W. De Mello, "Determining the impact of distributed generation on power systems. I. Radial distribution systems," in *2000 Power Engineering Society Summer Meeting*, vol. 3, July 2000, pp. 1645–1656.

[8] M. Lavorato, J. F. Franco, M. J. Rider, and R. Romero, "Imposing Radiality Constraints in Distribution System Optimization Problems," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 172–180, Feb 2012.

[9] K. P. Schneider, B. A. Mather, B. C. Pal, C. . Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. R. Pereira, L. F. Ochoa, L. R. de Araujo, R. C. Dugan, S. Matthias, S. Paudyal, T. E. McDermott, and W. Kersting, "Analytic considerations and design basis for the ieee distribution test feeders," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3181–3188, 2018.

[10] F. Postigo, C. Mateo, T. Gómez, B. Palmintier, B. Hodge, V. Krishnan, F. de Cuadra, and B. Mather, "A Review of Power Distribution Test Feeders in the United States and the Need for Synthetic Representative Networks," *Energies*, vol. 10, no. 11, pp. 1896.1–14, Nov 2017.

[11] K. M. Gegner, A. B. Birchfield, Ti Xu, K. S. Shetye, and T. J. Overbye, "A Methodology for the Creation of Geographically Realistic Synthetic Power Flow Models," in *2016 IEEE Power and Energy Conference at Illinois (PECI)*, Feb 2016, pp. 1–6.

[12] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye, and T. J. Overbye, "Statistical Considerations in the Creation of Realistic Synthetic Power Grids for Geomagnetic Disturbance Studies," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1502–1510, Mar 2017.

[13] A. Trpovski, D. Recalde, and T. Hamacher, "Synthetic Distribution Grid Generation Using Power System Planning: Case Study of Singapore," in *2018 53rd International Universities Power Engineering Conference (UPEC)*, Sep 2018, pp. 1–6.

[14] R. Kadavil, T. M. Hansen, and S. Suryanarayanan, "An Algorithmic Approach for Creating Diverse Stochastic Feeder Datasets for Power Systems Co-simulations," in *2016 IEEE PES General Meeting (PESGM)*, Jul 2016, pp. 1–5.

[15] 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*, vol. 26, no. 1, pp. 190–197, Feb 2011.

[16] E. Schweitzer, A. Scaglione, A. Monti, and G. A. Pagani, "Automated Generation Algorithm for Synthetic Medium Voltage Radial Distribution Systems," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, no. 2, pp. 271–284, Jun 2017.

[17] R. Atat, M. Ismail, M. F. Shaaban, E. Serpedin, and T. Overbye, "Stochastic Geometry-Based Model for Dynamic Allocation of Metering Equipment in Spatio-Temporal Expanding Power Grids," *IEEE Transactions on Smart Grid (Early Access)*, pp. 1–12, Oct 2019.

[18] N. Ponnavaikko, K. S. P. Rao, and S. S. Venkata, "Distribution System Planning through a Quadratic Mixed Integer Programming Approach," *IEEE Transactions on Power Delivery*, vol. 2, no. 4, pp. 1157–1163, Oct 1987.

[19] S. Lei, C. Chen, Y. Song, and Y. Hou, "Radiality Constraints for Resilient Reconfiguration of Distribution Systems: Formulation and Application to Microgrid Formation," 2019.

[20] Y. Wang, Y. Xu, J. Li, J. He, and X. Wang, "On the Radial Topological Constraints for Distribution System Restoration and Reconfiguration Problems," 2019.

[21] S. Thorve, S. Swarup, A. Marathe, Y. Chungbaek, E. K. Nordberg, and M. V. Marathe, "Simulating Residential Energy Demand in Urban and Rural Areas," in *2018 Winter Simulation Conference (WSC)*, Dec 2018, pp. 548–559.

[22] "Here streetmap premium for the u.s." 2020, last accessed 13 Feb 2020. [Online]. Available: www.here.com

[23] U.S. Department of Homeland Security, "Electric Substations," https://hifld-geoplatform.opendata.arcgis.com/datasets/electric-substations, 2019.

[24] A. J. Pansini, *Guide to Electrical Power Distribution Systems*, 6th ed. CRC Press, Jun 2005.

[25] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*. Berlin, Heidelberg: Springer-Verlag, 1985.

[26] N. Rotering, C. Schröders, J. Kellermann, and A. Moser, "Medium-Voltage Network Planning with Optimized Power Factor Control of Distributed Generators," in *2011 IEEE Power and Energy Society General Meeting*, Jul 2011, pp. 1–8.

[27] W. H. Kersting, *Distribution System Modeling and Analysis*. Abingdon: CRC Press, Jan 2012.

[28] M. Erwig, "The graph voronoi diagram with applications," *Networks*, vol. 36, no. 3, pp. 156–163, Sept 2000.

[29] S. Bolognani and F. Dörfler, "Fast power system analysis via implicit linearization of the power flow manifold," in *53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep 2015, pp. 402–409.

[30] M. Girvan and M. E. J. Newman, "Community Structure in Social and Biological Networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.