Creating Realistic Synthetic Power Distribution Networks based on Interdependent Road Infrastructure

Rounak Meyur[1*] , **Madhav Marathe**[2] , **Anil Vullikanti**[2] and **Virgilio Centeno**[1]

[1]Department of Electrical and Computer Engineering, Virginia Tech
[2]Biocomplexity Institute & Department of Computer Science, University of Virginia
{rounakm8, virgilio}@vt.edu, {marathe, vsakumar}@virginia.edu

Abstract

Physical inter-dependencies between networked civil infrastructures such as transportation and power system network are well known. In order to analyze complex non-linear co-relations between such networks, datasets pertaining to such real infrastructures are required. Such data are not readily available due to their sensitive nature. This work proposes a methodology to generate realistic synthetic distribution network for a given geographical region. The generated network is not the actual distribution system, but is very similar to the real distribution network. The proposed methodology is applied to create synthetic distribution networks in Montgomery county of south-west Virginia, USA. The created networks are validated for their structural feasibility and ability to operate within acceptable voltage limits under average load demand scenario.

1   Introduction

In the present day world, human behavior, social networks and civil infrastructures are closely intertwined to each other. Recent works such as Zeng et al. [2019], Duan et al. [2019] show how decisions undertaken by human behavior impact the load on infrastructure which in turn affects the following human decisions. Recent advances in artificial intelligence and computational science has opened new opportunities to analyze complex non-linear network dynamics in interdependent networks such as Adiga et al. [2020].

In order to study coupled infrastructures, a central challenge is the lack of realistic data sets. Such datasets should have detailed representations of each network as well as the interactions across infrastructures and the interactions with the human population. Over the past few years, Gegner et al. [2016], Birchfield et al. [2017], Atat et al. [2019] have tried to address the problem of generating realistic power networks so that they are openly available to researchers for studying complex phenomena. However, these works are limited to power transmission network where the impact of individual consumer decision is not prominent.

---

*rounakm8@vt.edu

Problem This works proposes a methodology to generate realistic synthetic distribution networks for a geographical location using open source information. We further aim to generate a network which follows structural (radial configuration) and operational constraints (voltage and power flows within limits) of a real distribution system.

1.1   Related Works

Several methodologies have been proposed to generate random distribution network topologies based on statistical distributions. A graph theory based approach to construct a synthetic distribution network is studied in Schweitzer et al. [2017] using statistical distributions of graph attributes (node degree, hops from source etc.). In Atat et al. [2019] a stochastic geometry based approach is proposed to place transformers in synthetic power networks. However, these works do not consider the various power system operation constraints such as node voltage and edge flows. Furthermore, the generated networks are not optimal in terms of network loss and cost of installation which is always the primary consideration of a distribution company.

Kadavil et al. [2016] proposed a bottom-up approach is used to build a synthetic distribution network from a substation and thereafter populating it with consumer loads. A mixed integer second order conic program (MI-SOCP) based optimization problem is formulated by Trpovski et al. [2018] to generate a synthetic medium voltage network for a geographical region. However, the location of loads are considered to be the zip code centers and loads are aggregated for each zip code which results in an aggregated synthetic distribution network of the region. The generated synthetic networks do not replicate the distribution of actual consumers in a geographical region. Thus, similar networks would be generated for rural and urban locations. In this paper, we have considered actual residential locations and therefore, the generated networks accurately represent the geographical location.

One of the important aspects to generate optimal realistic synthetic distribution networks is to maintain the radial configuration Lei et al. [2019] and ensure that the proposed optimization framework does not produce isolated loops in the network. It has been revisited by multiple works on distribution network reconfiguration (DNR) Singh et al. [2019a,b], Taylor and Hover [2012] and distribution system planning

(DSP) Lavorato et al. [2012], Ponnavaikko et al. [1987], Wang et al. [2019]. However, most of these works pertain to selecting edges in medium voltage networks with fixed number of known root nodes (substations or distributed energy resources). Furthermore, the non-substation (or non-root) nodes are either considered to have positive load demand Taylor and Hover [2012], Lavorato et al. [2012], Ponnavaikko et al. [1987] or handled by additional *single commodity flow* constraints Singh et al. [2019b], Wang et al. [2019]. In the proposed work, an optimal radial distribution network topology is generated for a given geographical location (rural or urban). The network is considered to have multiple unknown root nodes which are connected to the substation feeder via long distance high voltage feeder lines and are responsible to serve remotely located consumers.

1.2 Contribution

The contribution of this paper can be summed up through the following aspects: (i) proposes a methodology to create realistic synthetic distribution network using information from other infrastructures such as transportation networks, residential data etc., (ii) creates an optimal network by minimizing the overall length of distribution lines which is a principal consideration of power companies while planning distribution networks, (iii) generates a distribution network which is particular to the geographical location of interest and hence provides a realistic representation of the actual network.

2 Preliminaries

2.1 Distribution system

The distribution network consisting of overhead power lines, underground cables, pole top transformers is responsible to bring electrical power from high voltage (HV)(greater than 33kV) transmission system to the end residential consumers requiring a low voltage (LV) level (208-480V). This is normally done in a two-step procedure: (i) the high transmission level voltage is stepped down to medium voltage (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 ensures protection coordination among reclosures, breakers and downstream fuses in the distribution system feeders Zamani et al. [2010].

2.2 Available datasets

In this work, we try to generate the distribution network for a given geographical region (county/town/city) from different open source publicly available information. These data pertain to following sources: (i) transportation network data published by NAVTEQ [2016], (ii) Geographical location of HV substations from data sets published by U.S. Department of Homeland Security [2019] and (iii) Residential electric power

demand information developed in the models by Thorve et al. [2018].

Roads The road network represented in the form of a graph $\mathcal{R} = (\mathcal{V}_{\mathcal{R}}, \mathcal{L}_{\mathcal{R}})$, where $\mathcal{V}_{\mathcal{R}}$ and $\mathcal{L}_{\mathcal{R}}$ are respectively the sets of nodes and links of the network. Each road link $l \in \mathcal{L}_{\mathcal{R}}$ is represented as an unordered pair of terminal nodes (u, v) with $u, v \in \mathcal{V}_{\mathcal{R}}$. Each road node has a spatial embedding in form of longitude and latitude. Therefore each node $v \in \mathcal{V}_{\mathcal{R}}$ can be represented in two dimensional space as $\mathbf{p_v} \in \mathbb{R}^2$. Similarly, a road link $l = (u, v)$ can be represented as a vector $\mathbf{p_u} - \mathbf{p_v}$.

Substations The set of substations $\mathsf{S} = \{s_1, s_2, \cdots, s_M\}$, where the area consists of M substations and their respective geographical location data. Each substation can be represented by a point in the 2-D space as $\mathbf{p_s} \in \mathbb{R}^2$.

Residences The set of residential buildings with geographical coordinates $\mathsf{H} = \{h_1, h_2, \cdots, h_N\}$, where the area consists of N home locations. Each residential building can be represented by a point in the 2-D space as $\mathbf{p_h} \in \mathbb{R}^2$.

3 Proposed Approach

The goal of this work is to generate a realistic synthetic distribution network to connect the substations to all residential building locations. Therefore, the problem of synthesizing such networks is considered to be a two-step bottom-up procedure: the first step constructs the secondary distribution network connecting the residential buildings to pole-top/pad-mounted transformers and 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.

To summarize the overall procedure, the tasks would be the following: (i) Evaluate a mapping between sets of residential buildings and road network links such that each residence is mapped to the nearest road link. Additionally, 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) Identify subset of road network links which connects distribution substation transformers to the local pole-top/pad-mounted transformers in the form of a radial feeder network.

3.1 Mapping residences to road links

This section details the proposed methodology to identify subsets of residential buildings near each road network link. This information would be used in the successive steps to generate the secondary distribution network. The points along road network would serve as local distribution transformers delivering power to the residential buildings mapped to the link.

Algorithm 1 is used to compute the nearest road network link to a given point with associated spatial embedding. First, a bounding region of suitable size is evaluated for each road network link. This is done such that any point in the region is within a radius r from any internal point of the road network

link l. The bounding region for link $l = (u, v) \in \mathcal{L}_\mathcal{R}$ is

$$\mathbf{B_l} = \{\mathbf{p} \big| \|\mathbf{p} - \mathbf{p_l}\|_2 \le r, \forall \mathbf{p_l} = \theta \mathbf{p_u} + (1 - \theta) \mathbf{p_v}, \theta \in [0, 1]\} \quad (1)$$

Similarly, a bounding region is considered for a residential building $h \in \mathsf{H}$

$$\mathbf{B_h} = \{\mathbf{p} \big| \|\mathbf{p} - \mathbf{p_h}\|_2 \le r\} \quad (2)$$

The intersections between the bounding region of the building and those for the links are stored and indexed in a *quad-tree* data structure Finkel and Bentley [1974]. This information is retrieved to identify the k selected links which are comparably nearer to the residential building than the others. This algorithm reduces the computational burden of evaluating the distance between all road links and residential buildings. Now, the geodesic distance of point $\mathbf{p_h}$ can be computed from the k links and the nearest link can be identified. The

---

Algorithm 1 Find the nearest link in $\mathcal{L}_\mathcal{R}$ to a given point $\mathbf{p}$.

Require: Radius for bounding boxes r.
1: **for** each link $\mathbf{l} \in \mathcal{L}_\mathcal{R}$ **do**
2:      evaluate bounding box $\mathbf{B_l}$ for each link $\mathbf{l}$ using Eq. 1.
3: Evaluate bounding box $\mathbf{B_p}$ for point $\mathbf{p}$ using Eq. 2.
4: 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 $\mathbf{B_p}$.

5: Find the link $\mathbf{l}^\star$ among k short-listed links, which is nearest to point $\mathbf{p}$.

Algorithm 2 Nodeset for secondary network generation.

Require: Road network graph $\mathcal{R} = (\mathcal{V_R}, \mathcal{L_R})$, set of residential buildings H, minimum distance between local transformers d.
1: **for** each building $h \in \mathsf{H}$ **do**
2:      find mapping $f : \mathsf{H} \to \mathcal{L_R}$ using Algorithm 1 to generate the nearest link $e \in \mathcal{L_R}$.
3: **for** each link $l = (u, v) \in \mathcal{L_R}$ **do**
4:      find the inverse mapping $f^{-1} : \mathcal{L_R} \to \mathsf{H}$ which generates the set of buildings H_l associated with l.
5:      Interpolate m points $\mathsf{T}_l = \{t_1, t_2, \cdots t_m\}$ on link l between road nodes u and v which are d distance apart from each other.

---

internal points along the road links which are mapped to at least one residence are assumed to be probable locations of local distribution transformers. Therefore, an additional objective is to identify points along the road network link where local distribution transformers may be placed. In this paper, we interpolate multiple points along the link which are a definite length apart from each other. This is depicted in Algorithm 2.

3.2 Creation of secondary network

This section details the generation of secondary distribution networks connecting set of residential buildings H_l mapped to a road network link $l \in \mathcal{L_R}$ with set of local distribution

transformer nodes T_l located along l. Without loss of generality, we only consider non-trivial road links which are mapped to at least one residential building, i.e, $f^{-1}(l) = \mathsf{H}_l \ne \emptyset$. The generated network should be a *forest* of disconnected trees with each tree rooted at one of the transformer nodes $t \in \mathsf{T_l}$. The forest should cover all residential buildings in such a way that the overall length of distribution lines is minimized and intersections/overlaps of road link (proxy for primary) and secondary networks are avoided.

Graph generation In the proposed methodology, for a given non-trivial road network link $l \in \mathcal{L_R}$, a fully connected undirected graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ is constructed with node set $\mathcal{V}$ and edge set $\mathcal{E}$ that are incident on $\mathcal{V}$. The node set comprises of n_h residences mapped to the link l and n_t transformers along l (i.e., $\mathcal{V} = \mathsf{H}_l \cup \mathsf{T}_l$). Any edge $e \in \mathcal{E}$ is defined by incident nodes (i, j) where $i, j \in \mathcal{V}$. Since $\mathcal{G}$ is fully connected, the edge set $\mathcal{E}$ consists of all pairs of entries (i, j) for $i, j \in \mathcal{V}$. The aim of this step is to select an optimal subgraph $\tilde{\mathcal{G}} := (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$ which is a forest of disconnected trees.

Edge variables In order to identify which edges are required to be connected in the optimal topology, we introduce binary variables $\{x_e\}_{e \in \mathcal{E}} \in \{0, 1\}^{|\mathcal{E}|}$. 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 directed from the source node i to destination node j. The binary variable and flows can be stacked in $|\mathcal{E}|$-length vectors $\mathbf{x}$ and $\mathbf{f}$ respectively.

Node variables Each node $i \in \mathcal{V}$ either consumes power (if it is a residential building) or it injects power into the network (if it is a pole-top/pad-mounted transformer). Since each residence is associated with an hourly load demand profile, the average hourly load can be computed and be denoted by $p_i > 0$ which denotes its power consumption. For transformer nodes, power flows into the network which requres $p_i < 0$. The nodal power consumption at all nodes can be stacked in $n_t + n_h$ length vector $\mathbf{p}$. Note that $\mathbf{p} = \begin{bmatrix} \mathbf{p_T}^T & \mathbf{p_H}^T \end{bmatrix}^T$ with $\mathbf{p_H}$ denoting the power consumption at the home nodes stacked in a n_h length vector.

Degree constraint Statistical surveys on distribution networks in Rotering et al. [2011] show that residences along the secondary network are mostly connected in series with at most 2 neighbors. This is ensured by (3) which limits degree of residence nodes to 2.

$$\sum_{e:(h,j)} x_e \le 2, \quad \forall h \in \mathsf{H}_l \quad (3)$$

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

$$\mathsf{A}_{\mathcal{G}\,k,e} := \begin{cases} +1, & k = i \\ -1, & k = j \\ 0, & \text{otherwise} \end{cases} \quad \forall e = (i, j) \in \mathcal{E} \quad (4)$$

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

The *linearized distribution flow* (LDF) model has been extensively used for several grid optimization tasks in order to relate power consumption and flows to voltages. By ignoring network losses, the LDF model gives the relation between power consumption at the nodes and power flow along edges as (5a). Note that the optimal network is obtained from $\mathcal{G}$ after removing the edges for which $x_e = 0$. Therefore, (5a) is accompanied by (5b) which forces flows f_e to be zero for non-existing edges. The value of $\overline{f}$ in (5b) can be chosen to be the power flow limits in the edges.

$$\mathbf{A_H f} = \mathbf{p_H} \tag{5a}$$

$$-\overline{f} x_e \le f_e \le \overline{f} x_e, \quad \forall e \in \mathcal{E} \tag{5b}$$

Ensuring radial topology In our case, this condition is satisfied by the node power flow condition in (5a) if all the residential nodes consume power which is a reasonable assumption.

Proposition 1. *A graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with reduced branch-bus incidence matrix $\mathbf{A_H}$ and residential node power consumption vector $\mathbf{p_H}$, with strictly positive entries, is connected if and only if there exists a vector $\mathbf{f} \in \mathbb{R}^{|\mathcal{E}|}$, such that (5a) is satisfied.*

Proof. Proving by contradiction, suppose $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is not connected and there exists $\mathbf{f} \in \mathbb{R}^{|\mathcal{E}|}$ satisfying the proposed equality. If $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is not connected, then there exists a connected component $\mathcal{G}_\mathcal{S}(\mathcal{V}_\mathcal{S}, \mathcal{E}_\mathcal{S})$ which is a maximal connected subgraph with $\mathcal{V}_\mathcal{S} \subset \mathcal{V}$ and $\mathsf{T}_l \bigcap \mathcal{V}_\mathcal{S} = \emptyset$. Let $\mathbf{A}_\mathcal{S}$ denote the bus incidence matrix of $\tilde{\mathcal{G}}_\mathcal{S}$. By definition, it holds that $\mathbf{1}^T \mathbf{A}_\mathcal{S} = \mathbf{0}$.

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

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

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

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

where $\mathbf{p}_\mathcal{S}, \mathbf{p}_{\overline{\mathcal{S}}}$ are respectively the power consumptions at residence nodes $i \in \mathcal{V}_\mathcal{S}$ and $i \in \mathcal{V}_{\overline{\mathcal{S}}}$ stacked together. From the second block, it implies that

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

Since all entries of $\mathbf{p_H}$ are positive from the initial assumption, we have $\mathbf{1}^T \mathbf{p}_\mathcal{S} \ne 0$ which contradicts (6) and completes the proof. □

Once the connectivity of the subgraph $\tilde{\mathcal{G}}$ is ensured the radiality requirement can be enforced by the following constraint

$$\sum_{e \in \mathcal{E}} x_e = n_h \tag{7}$$

Generating optimal network topology The edges $(u, v) \in \mathcal{E}$ for all $u, v \in \mathcal{V}$ are assigned weights $w(u, v)$

$$w(u, v) = \begin{cases} \infty, & \text{if } u, v \in \mathsf{T_e} \\ \text{dist}(u, v) + \lambda \mathsf{C}(u, v), & \text{otherwise} \end{cases} \tag{8}$$

where $\text{dist} : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ denotes the geodesic distance between the residential locations u, v. The function C denotes if the edge crosses the nearest road link and is defined as

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

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 $|\mathcal{E}|$-length vector $\mathbf{w}$. Thereafter, a forest with trees rooted at the probable transformer nodes and spanning all the nodes of G is considered as the secondary network. This is obtained after solving the optimization problem.

$$\min_{\mathbf{x}} \mathbf{w}^T \mathbf{x} \tag{9}$$
$$\text{s.to. } (3), (5), (7)$$

3.3 Creation of primary network

The previous optimization framework is used to create secondary distribution network for all links $l \in \mathcal{L_R}$ for which $f^{-1}(l) \ne \emptyset$ holds true. Now, our goal is to connect the local transformer nodes (roots of tree components in secondary distribution network) to the substations with the road network as a proxy. The primary distribution network should also maintain a radial configuration. Additionally, the substation might be located far away from remote local transformers (particularly in rural feeders). In such a case, a high voltage feeder line (higher than the primary distribution level) is used to deliver power over a long distance to the remotely located customers. Thereafter, a step down transformer with a voltage regulator is used to bring the voltage down to the primary voltage level.

Graph generation The first step to construct the primary network is to identify the set of possible edges which might be a part of the primary network. Note that the transformer nodes are internal points along the road network links. We follow Algorithm 3 to obtain the possible set of edges.

Clustering of nodes for each substation An optimization problem can be solved to identify the optimal set of edges which minimizes the overall length of the network and at the same time satisfies all structural constraints. However, the scale of such a problem can be as high as dealing with 15000 nodes when an entire county is considered.

The problem of creating primary network for a region with multiple substations can be solved for individual substation. To this end, the road network nodes and transformers are clustered so that each node (road network or transformer) is mapped to the nearest substation. The graph $\mathcal{G_P}(\mathcal{V_P}, \mathcal{E_P})$ is partitioned into M subgraphs $\{\mathcal{G}_{s_1}, \mathcal{G}_{s_2}, \cdots, \mathcal{G}_{s_M}\}$ corresponding to each of the substation. This is depicted in Fig 1 where each color represents a partition of road and transformer nodes. These partitions are known as Voronoi cells

Algorithm 3 Generation of graph to create primary network.

Require: Road network graph $\mathcal{R} = (\mathcal{V}_\mathcal{R}, \mathcal{L}_\mathcal{R})$, set of road links $\mathcal{L}_\mathcal{T} = \{l_k | l_k \in \mathcal{L}_\mathcal{R}, f^{-1}(l_k) \neq \emptyset\}$.

1: Initialize graph $\mathcal{G}_\mathcal{P}(\mathcal{V}_\mathcal{P}, \mathcal{E}_\mathcal{P}) \leftarrow \mathcal{R}$
2: **for** each road link $l_k = (u_k, v_k) \in \mathcal{L}_\mathcal{T}$ **do**
3:     Solve optimization problem in Section 3.2.
4:     Identify set of transformers $\mathsf{T}_k = \{t_1, t_2, \cdots, t_{m_k}\}$
5:     Construct path $\mathcal{P}_k = (u_k, t_1, t_2, \cdots, t_{m_k}, v_k)$
6:     Remove link $l_k = (u_k, v_k)$ from graph, $\mathcal{E}_\mathcal{P} \leftarrow \mathcal{E}_\mathcal{P} \setminus \{l_k\}$
7:     **for** each edge $p_i \in \mathcal{P}_k$ **do**
8:         Add edge to graph, $\mathcal{E}_\mathcal{P} \leftarrow \mathcal{E}_\mathcal{P} \cup \{p_i\}$
9:     Update nodes with transformer nodes $\mathcal{V} \leftarrow \mathcal{V} \cup \mathsf{T}_k$
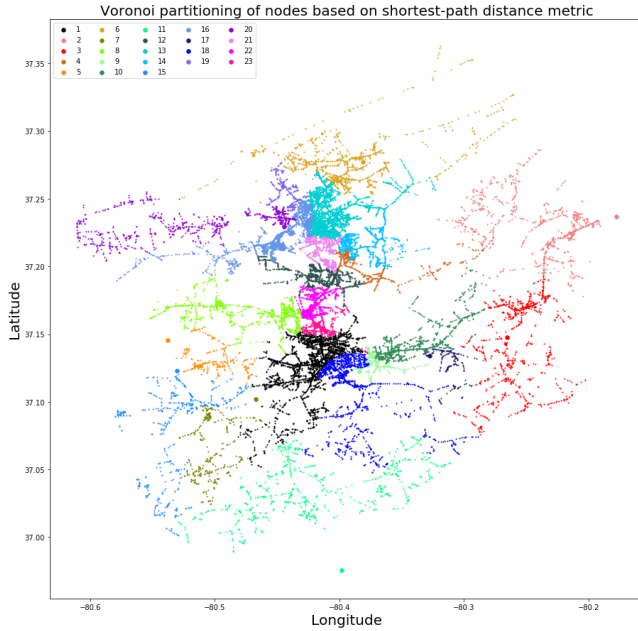10: Return graph $\mathcal{G}_\mathcal{P}$

Figure 1: Road network and transformer nodes clustered based on the shortest-path distance to different substations.

which are centered at the substation location. The partitioning is done based on the shortest path distance metric which ensures that each node is mapped to the nearest substation and each induced subgraph $\mathcal{G}_s(\mathcal{V}_s, \mathcal{E}_s)$ corresponding to substation s has a single connected component.

Our goal is to identify optimal primary network $\mathcal{P}_s(\tilde{\mathcal{V}}_s, \tilde{\mathcal{E}}_s)$ from the set of edges in $\mathcal{G}_s(\mathcal{V}_s, \mathcal{E}_s)$. Note that $\mathcal{V}_s$ consists of local transformer nodes as well as road network nodes ($\mathcal{V}_s = \mathcal{V}_{s_T} \cup \mathcal{V}_{s_R}$). The optimal primary network is a forest of trees covering all nodes in the set of transformer nodes $\mathcal{V}_{s_T}$. However, not all road network nodes need to be covered as they serve as *transfer* nodes which has no power consumption and are present to transfer power from preceding to succeeding node. Furthermore, the road network nodes can also act as root node connecting HV feeder lines from the substation with the primary network. That is, the nodes in set $\mathcal{V}_{s_R}$

may either be chosen to be included in $\mathcal{P}_s$ or not. If a road node is chosen it can either be a root node for a tree or can be a *transfer* node with a degree of at least 2.

Node variables $\mathcal{V}_s$ comprises of n_t transformer and n_r road nodes. Let p_i denote the power consumption at the ith transformer node. This is obtained by summing up the total load demand of residences connected to the transformer. These power consumptions can be stacked in a n_t-length vector $\mathbf{p}$. Let v_i represent the voltage at the node i. The nodal voltages at all nodes can be stacked in $n_t + n_r$ length vectors $\mathbf{v}$.

We assign binary variables $\{y_r, z_r\}_{r \in \mathcal{V}_{s_R}} \in \{0, 1\}$. 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 chosen and is a root node in the optimal primary network. $z_r = 1$ indicates that the road node is a non-root node. Furthermore, we enforce that if a road node is not chosen ($y_r = 0$), it is to be treated as a non-root node ($z_r = 1$).

Edge variables In order to identify which edges comprise of the primary network, each edge e is assigned a binary variable $\{x_e\}_{e \in \mathcal{E}_s} \in \{0, 1\}^{|\mathcal{E}_s|}$. 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 $|\mathcal{E}_s|$-length vectors $\mathbf{x}$ and $\mathbf{f}$ respectively.

Connectivity constraint The first set of constraints are defined to ensures the structural feasibility of road nodes in the optimal primary network.

$$\sum_{e:(r,j)} x_e \leq |\mathcal{E}_s| y_r, \quad \forall r \in \mathcal{V}_{s_R} \tag{10a}$$

$$\sum_{e:(r,j)} x_e \geq y_r, \quad \forall r \in \mathcal{V}_{s_R} \tag{10b}$$

$$\sum_{e:(r,j)} x_e \geq 2(y_r + z_r - 1), \quad \forall r \in \mathcal{V}_{s_R} \tag{10c}$$

$$1 - z_r \leq y_r, \quad \forall r \in \mathcal{V}_{s_R} \tag{10d}$$

Let $e : (r, j)$ denote the set of all edges $e \in \mathcal{E}_s$ which are incident on the road node $r \in \mathcal{V}_{s_R}$. Therefore, $\sum_{e:(r,j)} x_e$ is the degree of node r in graph $\mathcal{G}_s$. There are three possibilities for each road node r: (i) it is not chosen, (ii) it is chosen and is not a root node and (iii) it is chosen and is a root node.

If r is not chosen ($y_r = 0$), (10d) ensures $z_r = 1$. (10a), (10b)and (10c) ensures that there are no incident edges on r. If r is chosen and is a root node ($y_r = 1, z_r = 0$), (10a), (10b)and (10c) ensures that the degree of the road node is positive. Finally, if r is chosen 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 (10c). The binary variables y_r, z_r can be stacked in n_r-length vectors $\mathbf{y}$ and $\mathbf{z}$ respectively.

Ensuring radiality constraint Since all the road network nodes need not be covered by the generated radial topology, we have the following equivalent of Eq. 7.

$$\sum_{e \in \mathcal{E}_s} x_e = |\mathcal{V}_s| - \sum_{r \in \mathcal{V}_{s_R}} (1 - y_r) - \sum_{r \in \mathcal{V}_{s_R}} (1 - z_r) \tag{11}$$

The total number of edges in the optimal primary network is equal to the difference between total number of nodes to

be covered and number of root nodes. The last term in (11) denotes the number of root nodes and the first two terms together indicates the total number of nodes to be covered.

Power balance and flow constraints We can define the branch-bus incidence matrix $\mathbf{A}_{\mathcal{G}_s} \in \mathbb{R}^{|\mathcal{V}_s| \times |\mathcal{E}_s|}$ similar to Eq. 4. Since the order of rows and columns in $\mathbf{A}_{\mathcal{G}_s}$ is arbitrary, we can partition the rows without loss of generality as $\mathbf{A}_{\mathcal{G}_s} = \begin{bmatrix} \mathbf{A}_{\mathbf{s_T}}^T & \mathbf{A}_{\mathbf{s_R}}^T \end{bmatrix}^T$. Here, the partitions $\mathbf{A}_{\mathbf{s_T}}$ and $\mathbf{A}_{\mathbf{s_R}}$ are obtained by stacking the rows of $\mathbf{A}_{\mathcal{G}_s}$ corresponding to transformer nodes and road nodes respectively.

$$\mathbf{A}_{\mathbf{s_T}} \mathbf{f} = \mathbf{p} \tag{12a}$$
$$-\overline{s}(1 - \mathbf{z}) \leq \mathbf{A}_{\mathbf{s_R}} \mathbf{f} \leq \overline{s}(1 - \mathbf{z}) \tag{12b}$$
$$-\overline{f}\mathbf{x} \leq \mathbf{f} \leq \overline{f}\mathbf{x} \tag{12c}$$

(12a) ensures that a path exists between each transformer node and a root node following Proposition 1. If a road node is not a root node (with $z_r = 1$), (12b) 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 substation power capacity limits $[-\overline{s}, \overline{s}]$. (12c) ensures that if an edge $e \in \mathcal{E}_s$ is selected ($x_e = 1$), the power flow is limited by the line capacity limits $[-\overline{f}, \overline{f}]$.

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 (13a). (13b) limits the voltage at all nodes within the acceptable limits of $[\underline{v}, \overline{v}]$.

$$(1 - v_r) \leq z_r \; \forall r \in \mathcal{V}_{s_R} \tag{13a}$$
$$\underline{v}\mathbf{1} \leq \mathbf{v} \leq \overline{v}\mathbf{1} \tag{13b}$$
$$-(1 - x_e)M \leq v_i - v_j - r_e f_e \leq (1 - x_e)M, \; \forall e \in \mathcal{E}_s \tag{13c}$$

If r_e denotes the resistance of the line $e : (i, j) \in \mathcal{E}_s$, the 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 $v_i - v_j = r_e P_e$. Notice that this constraint is only activated for those edges $e \in \mathcal{E}_s$ for which $x_e = 1$. Therefore, we can enforce this constraint as (13c). If an edge is not selected ($x_e = 0$), (12c) ensures that $f_e = 0$. Therefore, M in (13c) can be selected to be $M = \overline{v} - \underline{v}$.

Generating optimal primary network Each edge $e = (u, v) \in \mathcal{E}_s$ is assigned a weight $w_e = w(u, v) = \mathsf{dist}(u, v)$ which is the geodesic distance between the nodes. Additionally, for every road node $r \in \mathcal{V}_{s_R}$, 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.

$$\min_{\mathbf{x}, \mathbf{y}, \mathbf{z}} \sum_{e \in \mathcal{E}_s} x_e w_e + \sum_{e \in \mathcal{V}_{s_R}} (1 - z_r) d_r \tag{14}$$
$$\text{s.to. } (10), (11), (12), (13)$$

4 Results and Discussion

The proposed methodology is used to generate synthetic distribution networks connecting substations to individual resi-
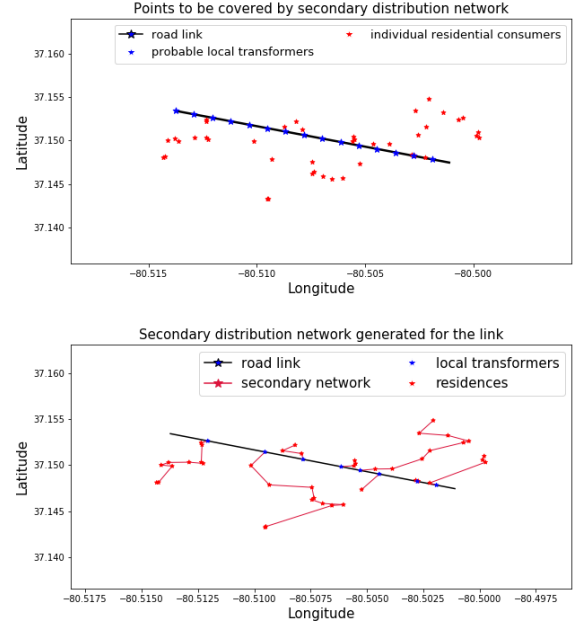

Figure 2: Secondary distribution network generated for a residences along a road link in Montgomery county of south-west Virginia, USA. The top figure shows the probable local transformers and residences to be connected. The bottom figure shows the generated optimal secondary distribution network.

dences in Montgomery county of south-west Virginia, USA. The geographical location consists of 37223 residential locations and the associated road network has 8500 nodes and 10115 edges. There are 23 substations from which the distribution network originate.

The first task is to map individual residential location to the nearest road network link. This is followed by the creation of secondary distribution network for each link. For example, Fig. 2 shows the creation of secondary distribution network for a set of residences mapped to a road link. The blue points along the road link indicate the probable locations of local transformers along the link and the red points are the residences which are required to be connected. The optimal secondary network is shown in the bottom figure which covers all the residences mapped to the road link and rooted at the local transformer nodes.

The secondary network is generated for all road links mapped to at least one residence. The number of local transformers along the road links is about 14000. The final task is to generate the primary distribution network which connects these local transformers to the substations. For this purpose, we divide the entire network into separate regions as shown in Fig. 1. The optimal primary network originating from one substation is shown in Fig. 3. The first figure shows road network consisting of possible edges and the local transformers which are to be connected. The second figure shows the generated optimal primary network. Fig. 4 shows the entire synthetic distribution network (primary and secondary) generated for the same substation. The generated network is val-
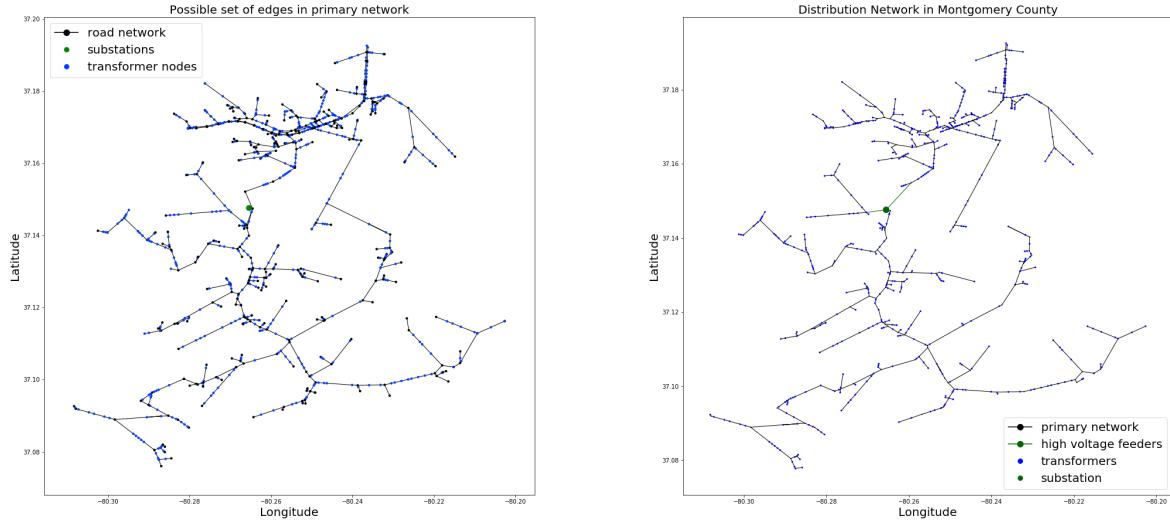
Figure 3: Primary distribution network generated for a substation in Montgomery county of south-west Virginia, USA. The left figure shows the set of possible road network edges from which the primary network is selected and the local transformers which are required to be connected. The right figure shows the optimal primary distribution with three feeders (green edges) originating from the substation.

idated for the structural and operational constraints. The network has a radial structure with no loops and the operating voltage at different points in the network is shown in the second figure. The operating voltages are calculated when the residential customers are consuming average load. It is observed that the voltage profile is within acceptable limits of 0.95 to 1.0 pu. The voltage profile can be further improved by optimally placing capacitors along the network.

5 Conclusion

This paper proposes a methodology to generate synthetic distribution networks for a particular geographical location based on available road network data. The generated network connects individual residential customers to substations while maintaining a radial configuration. Additionally, the network is created such that the overall length of overhead lines is minimized which is similar to planning methodologies undertaken by distribution companies. This ensures that generated synthetic distribution network is realistic and can be used to represent the network of the geographic location accurately.

References

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.

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)*, pages 1–12, Oct 2019.

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*, 32(2):1502–1510, Mar 2017.

D. Duan, C. Lv, S. Si, Z. Wang, D. Li, J. Gao, S. Havlin, H. E. Stanley, and S. Boccaletti. Universal Behavior of Cascading Failures in Interdependent Networks. *Proceedings of the National Academy of Sciences*, 116(45):22452–22457, 2019.

R. A. Finkel and J. L. Bentley. Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 4:1–9, 1974.

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)*, pages 1–6, Feb 2016.

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 Power and Energy Society General Meeting (PESGM)*, pages 1–5, Jul 2016.

M. Lavorato, J. F. Franco, M. J. Rider, and R. Romero. Imposing Radiality Constraints in Distribution System Opti-
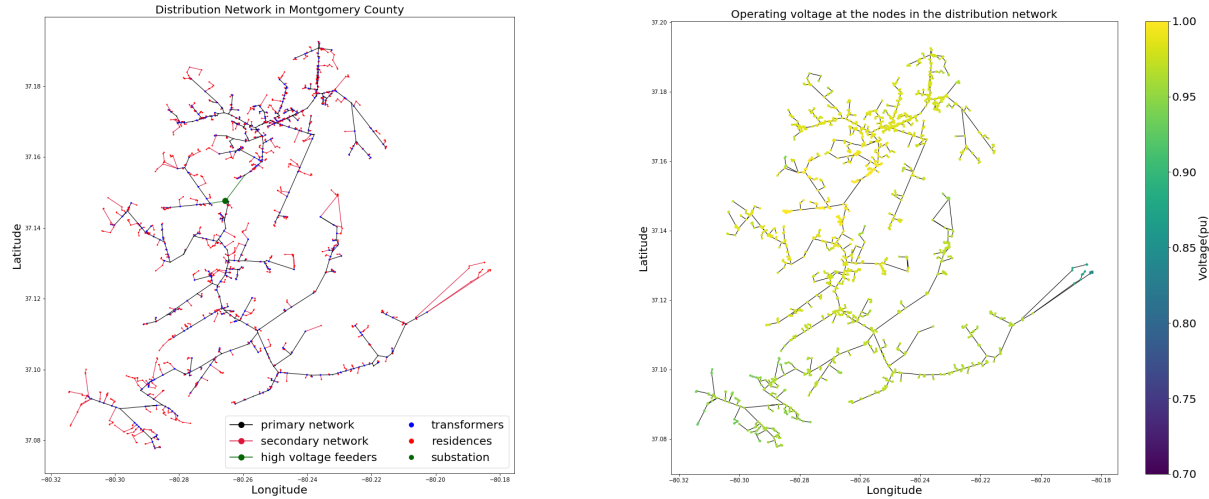
Figure 4: The left figure shows the primary and secondary distribution network for a substation in Montgomery county of southwest Virginia, USA. The right figure shows the operating voltage at different points in the network when individual customers are consuming average load demand.

mization Problems. *IEEE Transactions on Power Systems*, 27(1):172–180, Feb 2012.

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

NAVTEQ. Title of NAVTEQ data set webpage. https://www.navteq.com, 2016.

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*, 2(4):1157–1163, Oct 1987.

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*, pages 1–8, Jul 2011.

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*, 7(2):271–284, Jun 2017.

M. K. Singh, V. Kekatos, and C.-C. Liu. Optimal Distribution System Restoration with Microgrids and Distributed Generators. In *2019 IEEE Power and Energy Society General Meeting (PESGM)*, 2019a.

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

J. A. Taylor and F. S. Hover. Convex Models of Distribution

System Reconfiguration. *IEEE Transactions on Power Systems*, 27(3):1407–1413, Aug 2012.

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)*, pages 548–559, Dec 2018.

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)*, pages 1–6, Sep 2018.

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

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

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

G. Zeng, D. Li, S. Guo, L. Gao, Z. Gao, H. E. Stanley, and S. Havlin. Proceedings of the National Academy of Sciences. *Proceedings of the National Academy of Sciences*, 116(1):23–28, 2019.