Creating synthetic distribution networks

September 10, 2019

I Prior Works

Previous works by Overbye et al. [1, 2] propose methods to generate synthetic transmission network which bear topological and geographical resemblance with a real power system. At first, a statistical study is performed to estimate the per capita power consumption. This, along with the population information from the US Census data is used to estimate the real/reactive power demand at each ZIP code center which is considered to be locations of load substations. This estimated value is used as the substation load and voltage is assigned based on typical values present in the geographic region. Generator location and capacity data obtained from the US Energy Information Administration (EIA) is used to identify locations of generator substations. Thereafter, a Delaunay Triangulation method is used to connect the generator and load substations while maintaining the statistical aspects of a real power system network at the same time. However, these works lack methods to generate multiple number of such realistic synthetic networks. Furthermore, the models aggregate the distribution system to a load substation with an aggregated load value. In our work, we present a method to generate realistic synthetic distribution networks in as much detail as possible.

II Preliminaries

In this work, we try to generate the distribution network for a given region (county/town/city) from different open source publicly available information. These data pertain to following sources.

- Transportation network data published by NAVTEQ [3].
- Geographical location of high voltage (HV) and extra-high voltage (EHV) substations from data sets published by U.S. Energy Information Administration (EIA) [4].
- Residential electric power demand information developed in the models by [5].

The distribution system is synthesized in a way such that it resembles a typical radial distribution feeder network. The goal of this work is to generate the primary and secondary distribution network to connect the substations to all building locations. We want a mapping between the substations and residential locations. This can be achieved by mapping the substations to nodes of the transportation (road) network and assigning a map between residential buildings and road network links. The road network can be used as a proxy for the primary distribution network. The secondary distribution network can be generated from the second map.

The NAVTEQ transportation network data is available in the form of an edge-list or a list of links. Each link e has an associated integer level l_e from the set $\{1, 2, 3, 4, 5\}$ that describes the link type (e.g., a

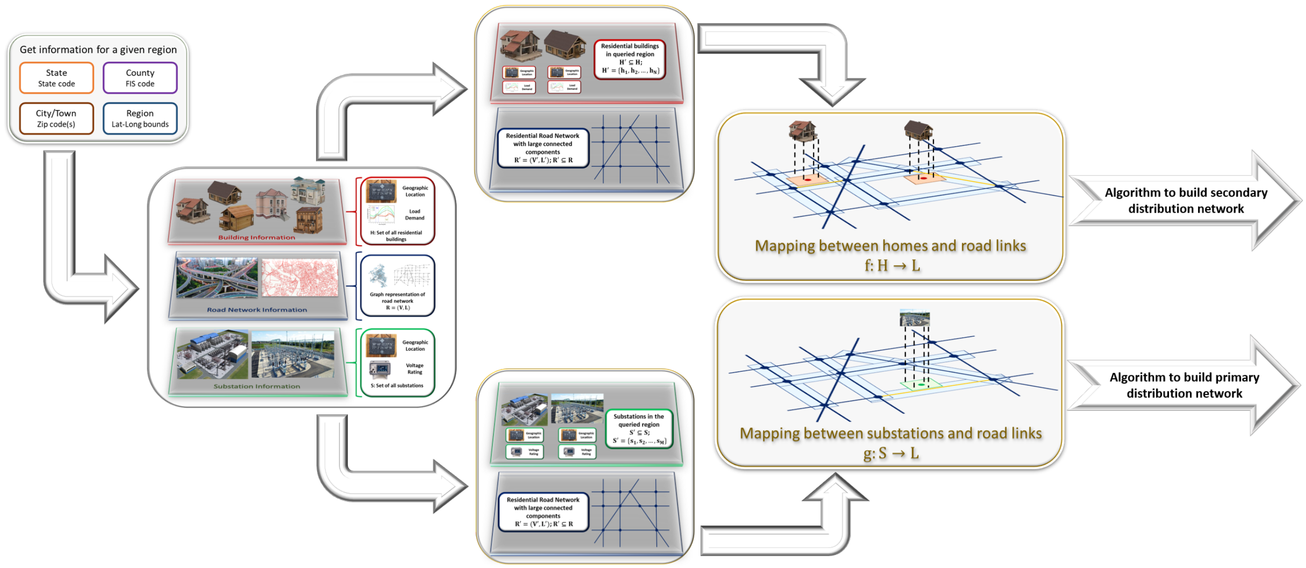

Figure 1: Flowchart showing the generation of maps between three data sets.

level-1 ($l_e = 1$) link could correspond to an Interstate road segment, while a level-5 ($l_e = 5$) link could correspond to a residential driveway). All links with level $l_e \leq 2$ are removed from the dataset. These links are dropped since components of the distribution network (e.g., homes) are typically not located along these links. Furthermore, connected components in the road network graph of considerable size (number of nodes greater than 10) are only considered. This is done to remove small connected components along which a radial primary distribution feeder network cannot exist.

The synthetic population data and the substation information is available for the entire United States and stored in a database together with the NAVTEQ data. The synthetic distribution network is generated for a particular county. For this purpose, the required data set is selected from the master database based on the zip code information. The available data set is therefore listed as below.

(i) The road network represented in the form of a graph $\mathsf{R} = (\mathsf{V}, \mathsf{L})$, where V and L are respectively the sets of nodes and links of the network. Each node in the graph has an associated spatial embedding in form of longitude and latitude.

(ii) The set of substations $\mathsf{S} = \{s_1, s_2, \cdots, s_M\}$, where the county consists of M substations and their respective geographical location data.

(iii) The set of residential building locations $\mathsf{H} = \{h_1, h_2, \cdots, h_N\}$, where the county consists of N home locations. Each building is associated with longitude-latitude information.

III Algorithm

A Constructing the mapping

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

Require: Radius for bounding boxes r, a mapping $\mathsf{dist} : \mathbb{R}^2 \times \mathsf{L} \to \mathbb{R}^+$ such that $\mathsf{dist}(\mathbf{p}, \mathbf{l})$ is the shortest distance between point $\mathbf{p} \in \mathbb{R}^2$ and link segment $\mathbf{l} = \mathbf{u} - \mathbf{v}$ for $\mathbf{l} \in \mathsf{L}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$.

for each link $\mathbf{l} \in \mathsf{L}$ **do**
    evaluate bounding box $\mathbf{B_l}$ such that $\mathbf{B_l} = \left\{\mathbf{x} \middle| ||\mathbf{x} - \mathbf{x_l}||_2 \leq r, \forall \mathbf{x_l} = \theta\mathbf{u} + (1-\theta)\mathbf{v}, \theta = [0,1]\right\}$.
Evaluate bounding box $\mathbf{B_p}$ for point $\mathbf{p}$ such that $\mathbf{B_l} = \left\{\mathbf{x} \middle| ||\mathbf{x} - \mathbf{p}||_2 \leq r\right\}$.
Find the bounding boxes $\mathbf{B_{l_1}}, \mathbf{B_{l_2}}, \cdots, \mathbf{B_{l_m}}$ which intersect with $\mathbf{B_p}$.
Find the link $\mathbf{l}^\star \in \mathsf{L}'$, where $\mathsf{L}' = \{\mathbf{l_1}, \mathbf{l_2}, \cdots, \mathbf{l_m}\} \subseteq \mathsf{L}$ such that $\mathbf{l}^\star = \arg\min_{\mathbf{l} \in \mathsf{L}} \mathsf{dist}(\mathbf{p}, \mathbf{l})$.

Algorithm 2 Generate mapping between building and road network data sets.

Require: Road network graph $\mathsf{R} = (\mathsf{V}, \mathsf{L})$, set of residential buildings H, a mapping $\mathsf{dist} : \mathsf{H} \times \mathsf{V} \to \mathbb{R}^+$ such that $\mathsf{dist}(h, v)$ is the Euclidean distance between building $h \in \mathsf{H}$ and road network node $v \in \mathsf{V}$.

Initialize: A mapping from road nodes to set of sets of buildings $p : \mathsf{V} \to \mathcal{H}$ such that $p(v) = \emptyset, \forall v \in \mathsf{V}$
for each building $h \in \mathsf{H}$ **do**
    find mapping $f : \mathsf{H} \to \mathsf{L}$ using Algorithm 1 to generate the nearest link $e \in \mathsf{L}$.
for each link $e = (u, v) \in \mathsf{L}$ **do**
    Initialize four sets: $\mathsf{H_{uA}}, \mathsf{H_{vA}}, \mathsf{H_{uB}}, \mathsf{H_{vB}} = \emptyset$.
    find the inverse mapping $f^{-1} : \mathsf{L} \to \mathsf{H}$ which generates the set of buildings $\mathsf{H_e}$ associated with e.
    find the set of buildings $\mathsf{H_{eA}}, \mathsf{H_{eB}} \subseteq \mathsf{H_e}$ with $\mathsf{H_{eA}} \cup \mathsf{H_{eB}} = \mathsf{H_e}$ which are on opposite sides of e.
    for each building $h \in \mathsf{H_{eA}}$ **do**
        if $\mathsf{dist}(h, u) < \mathsf{dist}(h, v)$ **then**
            Add this building to set $\mathsf{H_{uA}}$: $\mathsf{H_{uA}} \leftarrow \mathsf{H_{uA}} \cup \{h\}$
        else
            Add this building to set $\mathsf{H_{vA}}$: $\mathsf{H_{vA}} \leftarrow \mathsf{H_{vA}} \cup \{h\}$
    for each building $h \in \mathsf{H_{eB}}$ **do**
        if $\mathsf{dist}(h, u) < \mathsf{dist}(h, v)$ **then**
            Add this building to set $\mathsf{H_{uB}}$: $\mathsf{H_{uB}} \leftarrow \mathsf{H_{uB}} \cup \{h\}$
        else
            Add this building to set $\mathsf{H_{vB}}$: $\mathsf{H_{vB}} \leftarrow \mathsf{H_{vB}} \cup \{h\}$
    Add the sets $\mathsf{H_{uA}}, \mathsf{H_{uB}}$ to the mapping $p(u)$: $p(u) \leftarrow p(u) \cup \{\mathsf{H_{uA}}, \mathsf{H_{uB}}\}$.
    Add the sets $\mathsf{H_{vA}}, \mathsf{H_{vB}}$ to the mapping $p(v)$: $p(v) \leftarrow p(v) \cup \{\mathsf{H_{vA}}, \mathsf{H_{vB}}\}$.

Algorithm 3 Generate mapping between substations and road network nodes.

Require: Road network graph $\mathsf{R} = (\mathsf{V}, \mathsf{L})$, set of substations S, a mapping $\mathsf{dist} : \mathsf{S} \times \mathsf{V} \to \mathbb{R}^{+}$ such that $\mathsf{dist}(s, v)$ is the Euclidean distance between substation $s \in \mathsf{S}$ and road network node $v \in \mathsf{V}$, mapping $p : \mathsf{V} \to \mathcal{H}$ obtained from Algorithm 2, minimum number of road network nodes mapped to a substation N_{min}.

for each road network node $v \in \mathsf{V}$ **do**
  if $p(v) = \emptyset$ **then**
    Remove node v from set of all road network nodes $\mathsf{V} \leftarrow \mathsf{V} \setminus \{v\}$.
  else
    Find a mapping $g : \mathsf{V} \to \mathsf{S}$ which identifies the nearest substation to the road node v such that $g(v) = \arg\min_{s \in \mathsf{S}} \mathsf{dist}(s, v)$.
for each substation $s \in \mathsf{S}$ **do**
  Find inverse mapping $g^{-1} : \mathsf{S} \to \mathsf{V}$ which generates the set of road network nodes $\mathsf{V_s}$ associated with s.
  if $|g^{-1}(s)| < N_{min}$ **then**
    Remove the substation from set of all substations, $\mathsf{S} \leftarrow \mathsf{S} \setminus \{s\}$
Recompute mapping $g : \mathsf{V} \to \mathsf{S}$ which identifies the nearest substation to the road node v such that $g(v) = \arg\min_{s \in \mathsf{S}} \mathsf{dist}(s, v)$.

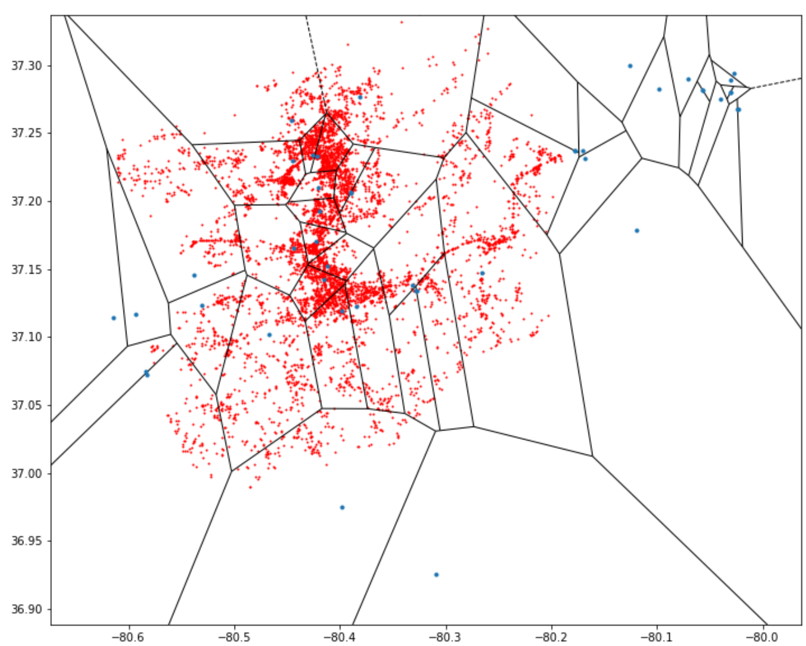

Figure 2: Voronoi regions formed by substations (blue points) and the road nodes (red points) mapped within each region. A number of substations have very road nodes associated with it. Therefore, the Voronoi regions are recomputed after removing the unmapped substations.

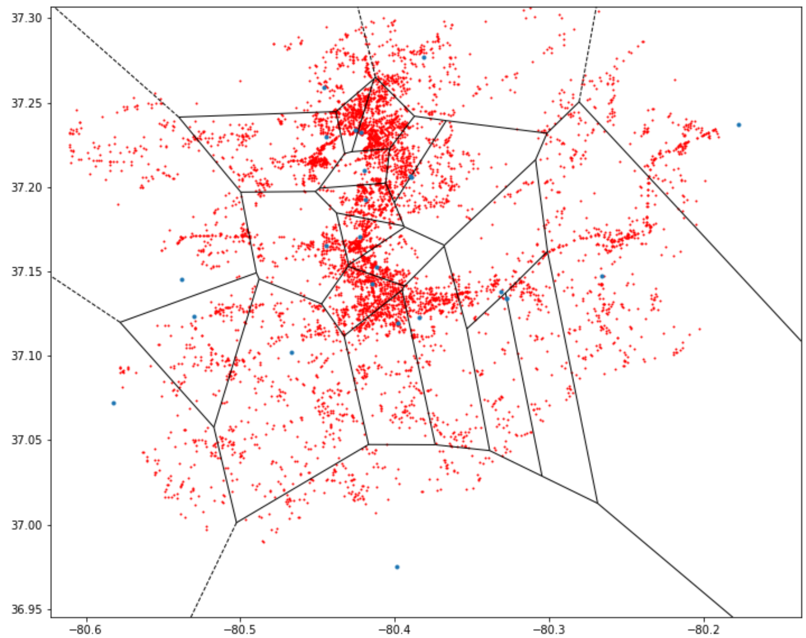

Figure 3: Voronoi regions formed by substations (blue points) and the road nodes (red points) mapped within each region. Almost all the regions are densely distributed by road network nodes. The boundary regions seem to be empty. However, they would be filled by road nodes from the neighboring counties.

References

[1] 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.

[2] 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.

[3] NAVTEQ. Title of NAVTEQ data set webpage. https://www.navteq.com. Accessed:.

[4] U.S. Department of Homeland Security. Electric Substations. https://hifld-geoplatform.opendata.arcgis.com/datasets/electric-substations. Updated: 2019-07-09.

[5] 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.